

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DERİN ÖĞRENME İLE KİŞİ KİMLİĞİNİN YENİDEN
TANIMLANMASI

Sara AKAN

DOKTORA TEZİ

Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Programı

Danışman
Prof. Dr. Songül VARLI

Eylül, 2023

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**DERİN ÖĞRENME İLE KİŞİ KİMLİĞİNİN YENİDEN
TANIMLANMASI**

Sara AKAN tarafından hazırlanan tez çalışması 12.09.2023 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, Bilgisayar Mühendisliği Programı **DOKTORA TEZİ** olarak kabul edilmiştir.

Prof. Dr. Songül VARLI
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Prof. Dr. Songül VARLI, Danışman
Yıldız Teknik Üniversitesi

Prof. Dr. Gökhan BİLGİN, Üye
Yıldız Teknik Üniversitesi

Doç. Dr. Tevfik AYTEİN, Üye
Bahçeşehir Üniversitesi

Prof. Dr. Çiğdem EROĞLU ERDEM, Üye
Marmara Üniversitesi

Dr. Öğr. Üye. Oğuz ALTUN, Üye
Yıldız Teknik Üniversitesi

Danışmanım Prof. Dr. Songül VARLI sorumluluğunda tarafımda hazırlanan Derin Öğrenme ile Kişi Kimliğinin Yeniden Tanımlanması başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Sara AKAN

İmza



Aileme
ve
Kıymetli eşime

TEŞEKKÜR

Hazırlamış bulunduğum bu naçizane doktora tez çalışmasında, doğrudan veya dolaylı olarak katkısı olan kişilere teşekkürü bir borç bilirim. Bu bağlamda öncelikli olarak, tez çalışmam süresince yaptığım araştırmalarda beni her zaman destekleyen, cesaretlendiren, bana yol gösteren, yönlendiren, bilgi ve tecrübesini benimle paylaşan değerli hocam ve tez danışmanım Prof. Dr. Songül Varlı'ya en içten teşekkürlerimi ve saygılarımı sunarım. Kendisinin rehberliği ve deneyimli tavsiyeleri, tezimin kalitesini ve bütünlüğünü önemli ölçüde artırmıştır. Ayrıca, teknik yardımları ve güncel bilgileriyle sürekli yanımda olan arkadaşım Dr. Öğr. Üyesi Saeid Alp'e minnettarlığımı sunmak istiyorum. Bu süreç boyunca bana inanmayı hiç bırakmayan, öncelikle annem ve babam olmak üzere tüm aileme, her an yanımda olan sevgili eşim Taymaz Akan'a en içten şükranlarımı sunuyorum. Onların sevgi dolu destekleri ve inancı, zorlu tez sürecinde moral ve motivasyon kaynağım olmuştur. Sizler olmadan bu başarıyı elde etmek mümkün olmazdı. Son olarak, tez çalışmalarımnda yardımlarını esirgemeyen tüm meslektaşlarıma, dostlarıma ve diğer destekçilere de teşekkür etmek istiyorum.

Sara AKAN

İÇİNDEKİLER

SİMGE LİSTESİ	viii
KISALTMA LİSTESİ	x
ŞEKİL LİSTESİ	xiii
TABLolar LİSTESİ	xv
ÖZET	xvi
ABSTRACT	xviii
1 GİRİŞ	1
1.1 Kişi Yeniden Tanımlaması.....	5
1.1.1 Tanım	5
1.1.2 Kişinin Yeniden Tanımlanması Uygulamaları	6
1.1.3 Kişinin Yeniden Tanımlanmasının Zorlukları	8
1.2 Futbol Takımlarında Sporcu Yeniden Tanımlaması	10
1.2.1 Futbolcu Yeniden Tanımlama Uygulamaları	12
1.2.2 Futbolcunun Yeniden Tanımlanmasının Zorlukları	13
1.3 Motivasyon ve Araştırma Sorunu	15
1.4 Araştırma Hedefleri ve Katkıları.....	18
1.5 Tezin Genel Bakışı.....	18
2 TEMEL KAVRAMLAR	20
2.1 Kişi Yeniden Tanımlanmasının Mevcut Durumuna Genel Bir Bakış	20
2.1.1 Geleneksel El Yapımı Kişi Re-ID Yöntemleri	21
2.1.2 Derin Öğrenmeye Dayalı Kişi Yeniden Tanımlama Yöntemleri	23
2.2 Derin Evrişimli Sinir Ağları.....	24
2.2.1 Evrişimli Sinir Ağı Katmanları.....	25
2.2.2 Aktivasyon fonksiyonları.....	28
2.2.3 SoftMax Sınıflandırıcı ve Kayıp Fonksiyonu	30
2.2.4 Derin Evrişimli Sinir Ağlarının Eğitimi	31
2.2.5 Derin Evrişimli Sinir Ağlarında Aşırı Uyumla Başa Çıkma	33
2.2.6 Derin Evrişimli Sinir Ağı Mimarileri	34
2.3 Transfer Öğrenme ile Evrişimli Sinir Ağı Modelleri Geliştirme	37
2.3.1 Transfer Öğrenme Kavramı	38
2.3.2 Önceden Eğitilmiş Derin Modellerin İnce Ayar ile Transfer Öğrenme ...	39
2.4 Transformer’ler	41
2.4.1 İlk Transformer	42

2.4.2	Vision Transformer	45
2.4.3	Swin Transformer	46
2.5	Kişinin Yeniden Tanımlanmasında Kayıp Metrikleri.....	51
2.6	Füzyon Terminolojisi ve Stratejileri: Birleştirme Yöntemleri ve Stratejileri ..	51
2.7	Kişi Kimliğinin Yeniden Belirleme Benchmark Veri Setleri ve Değerlendirme Metrikleri.....	54
2.7.1	Kişi Re-ID Karşılaştırma veri kümeleri.....	54
2.7.2	Kişi Re-ID Değerlendirme Metrikleri.....	58
2.8	Bölüm Özeti	60
3	DERİN ÖĞRENMEYE DAYALI KİŞİ KİMLİĞİNİN YENİDEN BELİRLEME YÖNTEMLERİNİN TAKSONOMİSİ	61
3.1	Tanımlama Modeline Dayalı Yöntemler	61
3.2	Doğrulama Modeline Dayalı Yöntemler.....	64
3.3	Mesafe-Metrik Tabanlı Modele Dayalı Yöntemler.....	66
3.4	Pozisyon Yönlendirmeli ve Parça Tabanlı Modele Dayalı Yöntemler.....	68
3.5	Trensformer’ler Modeline Dayalı Yöntemler	71
3.6	Takım Sporları Yayın Videolarında Oyuncuların Kimliklerinin Yeniden Belirlenmesi	73
3.7	Bölüm Özeti	73
4	METHODOLOJİ	75
4.1	Önerilen Yöntem.....	78
4.2	Futbolcu Kimliği Yeniden Tanımlama için Temel Model Oluşturma.....	79
4.2.1	Temel Model Oluşturma	79
4.2.2	RFES-ReID Modeli	80
4.3	Öznitelikleri Çıkarma ve Mesafe Metrik Öğrenme	87
4.4	Kayıp Hesaplaması.....	87
4.5	Model optimization	90
4.6	Deneyisel Çalışma.....	92
4.6.1	Benchmark Veri Setleri	92
4.6.2	Veri Ön İşleme.....	93
4.6.3	Performans Değerlendirme Metrikleri.....	94
4.6.4	Deneyisel Kurulum	94
5	DENEYSEL SONUÇLAR VE TARTIŞMA	95
5.1	Önerilen Yöntemin Eğitim Sonuçları.....	96
5.2	Önerilen Yöntemin Değerlendirme Sonuçları	97
5.3	Kişi Yeniden Tanımlamanın Niteliksel Analizi.....	101
5.4	Son Teknoloji Yöntemlerle Karşılaştırma	102

6 SONUÇ	106
6.1 Sonuçlar.....	106
6.2 Öneriler	108
KAYNAKÇA	110
TEZDEN ÜRETİLMİŞ YAYINLAR	128



SİMGE LİSTESİ

ε	Adım boyutu
$w^{(i)}$	Ağırlıklar
$h^{(i)}$	Aktivasyon fonksiyonu
K	Aktivasyon sayısı
z	Aktivasyon vektörü
p	Arama görüntüsü
m	Bernoulli rasgele değişkeni
L_{CE}	Çapraz entropi
N_{gt}	Doğru değerlerin sayısı
$q(r)$	Doğru şekilde yeniden tanımlanmış sorguların sayısı
τ	Eşik değer
G	Galeri görüntüsü
x	Giriş değeri
i	Gizli katman düğümlerinin sayısı
m_t	Gradyan
β_1	Gradyanın 1'inci momenti
β_2	Gradyanın 2'inci momenti
∇_{θ_i}	Gradyanı temsili
(P, P)	Her görüntü parçasının boyutları
$h_{\theta}(x)$	Hipotez fonksiyonu
n_i	i'inci eğitim örneği
$ \cdot $	İlgili kümedeki aday sayısı
v_t	Karesi alınmış gradyanın hareketli ortalaması
L_{θ_i}	Kayıp fonksiyonu
pos	Kelimenin cümle içindeki konumunu
k	k en yakın komşuluk
n_{id}	Kişi kimliklerinin toplam kategori sayısı

$R(p, k)$	k-karşılıklı en yakın komşular
$* l$	l-genişletilmiş evrişim
M	Mesafe metriği
θ_i	Mevcut parametreler
d_{model}	Model içindeki alt katmanların çıktı boyutu
$\hat{y}_{n_i}^i$	n_i 'inci eğitim örneğinin tahmin edilen kişi sınıflandırma puanı
L_F	Odak kaybı
γ	Odaklanma parametresi
α	Öğrenme hızı
(H, W)	Orijinal görüntünün yüksekliği ve genişliği
$N(p, k)$	p arama görüntüsünün k-en yakın komşuları
a	Referans görüntü
a_n	Referans görüntüsü ile aynı kimliğe sahip olmayan görüntü
a_p	Referans görüntüsü ile aynı kimliğe sahip olan görüntü
C	Renk kanallarının sayısı
r	Sıra indeksi
$P(k)$	Sıralama listesinde kesim k'teki kesinlik
i^*	Sorgu görüntüsünün kimliği
Q	Sorgu görüntüsü
$\Phi(x)$	Standart normal dağılımın birikimli dağılım fonksiyonu
L_T	Tahmin edilen olasılık
n	Test görüntülerin sayısı
$k(t)$	Uygulanan filtre
L_T	Üçlü kayb
N	Yama sayısı
θ_{i+1}	Yeni öğrenilen parametreler

AAformer	Auto-Aligned Transformer
AAMGN	Attention-Aware Multiple Granularities Network
ABDNet	Attentive but Diverse Network
Adam	Adaptive Moment Estimation
ASA	Adaptive Split Self-Attention
CFS	Catastrophic Forgetting Score
CMC	Cumulative Match Characteristic
CNN	Convolutional Neural Networks
Conv	Convolutional
DeiT	Data-efficient Image Transformers
DenseNet	Densely-Connected Convolutional Network
DL	Deep Learning
DPM	Deformable Part Model
ELU	Exponential Linear Unit
FAN	Feature Aggregation Network
FC	Fully Connected
FFN	Fusion Feature Network
FPNN	Filter Pairing Neural Network
GAN	Generative Adversarial Network
GELU	Gaussian Error Linear Unit
GPS	Global Positioning System
HA-CNN	Harmonious Attention Convolutional Neural Network
HSV	Hue, Saturation, Value
IDE	ID Discriminative Embedding

ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IPT	Image Processing Transformer
KISSME	Keep It Simple and Straightforward Metric
LDA	Linear Diskriminant Analizi
LN	Layer Normalization
LSTM	Long Short-Term Memory
mAP	Mean Average Precision
MGN	Multiple Granularity Network
MLP	Multi-Layer Perceptron
MSA	Multi-Head Self Attention
NLP	Natural Language Processing
NN	Neural Networks
OIM	Online Instance Matching
OSNet	Omni-Scale Network
PAN	Pedestrian Alignment Network
PCA	Principal Component Analysis
PCB	Part-Based Conv Baseline
PGFA	Pose-Guided Feature Alignment
PReLU	Parametric Rectified Linear Unit
PSD	Positive Semi Definite
QDA	Quadratic Discriminant Analysis
RCN	Recurrent Comparative Network
Re-ID	Re-Identification
ReLU	Rectified Linear Unit
ResNet	Residual Convolutional Neural Network
RFES	Regional Feature Extraction Swin Transformer

RGA-SC	Relation-Aware Global Attention sequential spatial-channel
RGB	Red, Green, Blue
RNN	Recurrent Neural Networks
RPP	Refined Part-Pooling
RR	Re-Ranking
RRI	Restraint and Relaxation Iteration
SAN	Semantics Aligning Network
SSL	Self-Supervised Learning
SVD	Singular-Vector Decomposition
SVM	Support Vector Machine
SW-MAS	Shifted Windows Multi-Head Self Attention
ViT	Vision Transformer
VGG	Very Deep Convolutional Networks
W-MSA	Windows Multi-Head Self Attention
XQDA	Cross-view Quadratic Discriminant Analysis

ŞEKİL LİSTESİ

Şekil 1.1	Standard bir kişi Re-ID sisteminin blok diyagramı.....	6
Şekil 1.2	Bazı kişi Yeniden Tanımlama örnekleri.....	10
Şekil 1.3	Futbolcu yeniden tanımlama örnekleri.....	11
Şekil 1.4	Futbolcu Re-ID sisteminin blok diyagramı.	12
Şekil 1.5	Çeşitli zorlukların bazı örnekleri: (a) Giyim benzerliği; (b) Örtüşme ve kapatma; (c) ve (d) düşük çözünürlük ve farklı vücut hareketleri.	14
Şekil 2.1	Genel CNN yapısı [50].....	25
Şekil 2.2	Conv katmanının işlemlerini gösteren bir çizim[52].....	26
Şekil 2.3	2×2 filtre ve adım=2 ile maksimum ve ortalama havuzlama örnekleri.....	27
Şekil 2.4	İki FC katmanı ile basit ağ.....	28
Şekil 2.5	Çoğunlukla NN'lerde kullanılan aktivasyon fonksiyonları.	29
Şekil 2.6	Tipik residual öğrenme yapı taşları. a) ResNet-18/34 yapı bloğu (sol) ve b) ResNet-50/101/152 yapı bloğu (sağ).....	35
Şekil 2.7	ResNet-50 Mimarisi. Her bir rezidüel bloktan önce, adım boyutu 2 olan bir örnekleme işlemi gerçekleştirilir. Çıkış katmanı dışındaki tüm katmanlar ReLU aktivasyonunu kullanır. Çıkış katmanı ise SoftMax kullanır [73].	36
Şekil 2.8	ResNet-50'nin tipik bir rezidüel bloğunun görsel temsili.....	37
Şekil 2.9	Geleneksel makine öğrenimi ve transfer öğrenimi yaklaşımları arasındaki fark.	38
Şekil 2.10	Transfer öğreniminin genel akışının şematik bir diyagramı.	39
Şekil 2.11	İnce ayar prosedürü için CNN modelinin yapısı.	41
Şekil 2.12	RNN genel mimarisi[81].	42
Şekil 2.13	Orijinal Transformer mimarisi [80].....	43
Şekil 2.14	ViT'nin görsel temsili [84]	45
Şekil 2.15	Swin Transformer modeli.....	47
Şekil 2.16	Swin transformer blok modülleri.	48
Şekil 2.17	Yama birleştirme modülleri.....	49
Şekil 2.18	Kaydırılan pencere bölmelerinde öz dikkat için kullanılan hesaplama yöntemi.	50
Şekil 2.19	Farklı füzyon stratejileri için model mimarisi.....	52
Şekil 2.20	Market-1501 veri kümesi görüntülerinin örnekleri. Sütunların her biri bir kimliği temsil etmektedir.	56
Şekil 2.21	SoccerNet-ReID veri kümesi görüntülerinin örnekleri. Satırların her biri bir kimliği temsil etmektedir.	58
Şekil 2.22	CMC eğrisinin Rank 1, 5 ve 10'a dönüştürülmesi.....	59

Şekil 3.1 Derin öğrenme tabanlı tanımlama modelinin temel mimarisi [1].	62
Şekil 3.2 Derin öğrenme tabanlı doğrulama modelinin temel mimarisi.	64
Şekil 3.3 Üçlü derin model mimarisi.	67
Şekil 3.4 PCB derin modelinin yapısı [124].	70
Şekil 4.1 RFES-ReID yönteminin genel blok diyagramı	79
Şekil 4.2 a) Temel mimari, b) Swin Transformer bloğu.	80
Şekil 4.3 Yerel algı Swin transformerin (RFES) mimarisi. (a) Yerel algı bloğunun ayrıntılı yapısı; (b) Swin transformer bloğunun ayrıntılı yapısı.	81
Şekil 4.4 Kaydırılan pencerelerin etki mekanizması. (a) Giriş görüntüsü; (b) Girdi görüntüsünün W-MSA aracılığıyla pencere segmentasyonu (pencere boyutu 7 olarak ayarlanmıştır); (c) Kaydırılan pencerelerin eylemi; (d) Kaydırılan pencereler SW-MSA aracılığıyla farklı bir pencere bölümlleme yöntemi.	83
Şekil 4.5 Normal Evrişim ve Genişletilmiş Evrişim	84
Şekil 4.6 GELU ($\mu=0$, $\sigma=1$), ReLU ve ELU ($\alpha=1$) [159].	86
Şekil 4.7 Öznitelik çıkarma ve mesafe metriği öğrenme.	87
Şekil 4.8 Birleşmiş kayıp hesaplaması.	88
Şekil 4.9 Futbolcu Re-ID'si için RR prosedürü.	91
Şekil 5.1 RR uygulamasından önce ve sonra rank-5 sonuçları örneği.	101
Şekil 5.2 Önerilen yöntemin uygulanması yoluyla Market 1501 veri kümesine ait belirli sorgular için alınan ilk 20 sonuç. Sorgular en soldaki sütunda gösterilirken, galeriden alınan görseller benzerlik puanlarına göre soldan sağa doğru sıralanır. Her görüntünün üst kısmındaki T (yeşil) ve F (kırmızı) harfleri, doğru ve yanlış eşdeğerlerini göstermektedir.	102

TABLolar LİSTESİ

Tablo 2.1 ResNet Mimarisi [71].	36
Tablo 2.2 Kişi RE-ID'si için yaygın olarak kullanılan veri kümelerinin özeti.	55
Tablo 4.1 Bu tez boyunca deneylerde kullanılan veri kümelerine ilişkin ayrıntılar.	92
Tablo 5.1 Bu tez boyunca deneylerde kullanılan veri kümelerine ilişkin ayrıntılar.	95
Tablo 5.2 Önerilen yöntemin hiper parametreleri.	96
Tablo 5.3 Doğrulama hassasiyeti	97
Tablo 5.4 Değerlendirme sonuçları	98
Tablo 5.5 Kayıp fonksiyonlarının etkinliği	98
Tablo 5.6 Optimizasyon etkinliği	99
Tablo 5.7 Market-1501 veri kümesinde, en son derin öğrenme yaklaşımlarıyla performansın karşılaştırılması	103
Tablo 5.8 SoccerNet-V3 veri kümesinde, en son derin öğrenme yaklaşımlarıyla performansın karşılaştırılması	104

Derin Öğrenme ile Kişi Kimliğinin Yeniden Tanımlanması

Sara AKAN

Doktora Tezi

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Danışman: Prof. Dr. Songül VARLI

Bu tez çalışması temel olarak, bir futbol maçında herhangi bir anında farklı kamera açılarından çekilen görüntülerden futbolcunun kimliğini tanımlamaya odaklanmıştır. Bu görev, kişinin kimliğini yeniden belirleyen uygulamalara kıyasla farklı zorluklar içermektedir. Derin öğrenme teknolojisi söz konusu işlemin performansını büyük ölçüde iyileştirmiştir. Ancak aynı teknoloji futbolcular üzerinde uygulandığında yukarıda bahsi geçen zorluklardan dolayı beklenen performansı gösterememektedir. Dolayısıyla bu zorlukların üstesinden gelmek için güçlü özniteliklere ihtiyaç duyulmaktadır. Evrişimli sinir ağları yerel komşuluklara odaklandığından ilişkileri göz ardı ederken, aşağı örnekleme operatörleri de detaylı bilgi kaybına neden olur. Bu nedenle, farklı açılardan futbolcu görüntüleriyle çalışırken, mevcut mimarinin performansı genellikle tatmin edici değildir.

Bu çalışmada, böylesine zorlu bir sorunu çözmek için, RFES-ReID adlı bir yöntem önerilmiştir. Derin öğrenme teknolojisine dayalı futbolcunun kimliğinin yeniden belirlenmesi yöntemi olarak Swin Transformer ile birlikte yerel algı bloğu yöntemi önerilmiştir. Bu yöntem, futbolcunun kimliğinin yeniden belirlenmesine ilişkin temel sorunu, yani eğitim için güçlü öznitelikler eksikliğini gidermeye amaçlamaktadır.

Ardından, kişinin kimliğinin yeniden belirlenmesi eğitim sürecini daha etkili ve dengeli hale getirmek için birleştirilmiş kayıp fonksiyonlar yaklaşımı kullanılmıştır. Ayrıca, yeniden sıralama, önerilen temel model ile birleştirilerek önemli performans iyileştirmesi elde edilmiş ve aşırı uyum önlenmiştir. RFES-ReID kullanılarak futbolcu veri seti üzerinde güçlü ve ayırt edici öz nitelikler elde edilmiştir. Bununla birlikte birleştirilmiş kayıp fonksiyon kullanılarak dengeli bir öğrenme süreci sağlanmıştır. Söz konusu kazanımlarla bu alandaki boşluk doldurulmaya çalışılmıştır. Farklı veri kümeleri kullanılarak yapılan deneysel analiz ve sonuçlar, kişinin kimliğinin yeniden belirlenmesi konusunda RFES-ReID yönteminin diğer teknoloji metotlarına göre daha rekabetçi bir performans sağladığını göstermektedir.

Anahtar Kelimeler: Kişi Kimliğinin Yeniden Tanımlaması; Futbolcu Kimliğinin Yeniden Tanımlaması, Derin öğrenme; Swin Transformer; Yerel algı blok.

Person Re-identification Using Deep Learning

Sara AKAN

Doctor of Philosophy Thesis

Department of Computer Engineering

Supervisor: Prof. Dr. Songül VARLI

The proposed thesis basically focused on re-identifying the same soccer player in images taken from different camera perspectives at any moment of a soccer match. This task differs from traditional applications of person re-identification in a several important ways. Deep learning technology has greatly improved the performance of person re-identification. However, when the same technology is applied to soccer player re-identification, it does not show the same performance. Therefore, extracting strong and distinctive features to overcome the difficulties is necessary, which means that a strong architecture is required. Convolutional neural networks, which focus on local neighborhoods, tend to overlook relationships between features, and downsampling operators result in the loss of detailed information. Therefore, when soccer player images are captured from a different perspective, the performance of the model is often not satisfactory.

In this study, a method called RFES-ReID is proposed to address the challenging task of soccer player re-identification. The method utilizes a regional feature extraction block in conjunction with the Swin Transformer, leveraging deep learning technology. The main objective is to overcome the lack of strong features for training in soccer player re-identification within the deep learning framework. Then, a fusion loss function is used to make the person re-identification training process more effective and balanced. Moreover, re-ranking is combined with the baseline model proposed herein to achieve

significant performance improvement further and avoid overfitting. extracting strong and distinctive features on the soccer player dataset by using RFES-ReID and providing a balanced learning process using the fusion loss function is the most important contribution of this study to this field. Experimental analysis using various person re-identification datasets demonstrates that the proposed approach achieves competitive performance compared to state-of-the-art methods.

Keywords: Person Re-identification; Soccer Player Re-identification; Deep Learning; Swin Transformer; Regional feature extraction block.



1 GİRİŞ

Günümüzde video gözetim sistemleri, bilgisayar bilimi ve mühendisliğinde aktif kullanılan araştırma alanlarından biridir. Bu alanda, kamera ağında elde edilen videoları işlemek ve anlamlandırmak adına kamera ağ videolarını izleme ve analiz etme otomasyonu için makine öğrenimi tekniklerinin yanı sıra bilgisayarlı görü de büyük talep görmektedir. Bu araştırma alanı, farklı amaçlara göre izleme araçlarına yardımcı olabilmek için birçok farklı metot içermektedir. Bu araçlar, kişileri algılamak ve izlemek, kamera ağından olağandışı eylemleri/davranışları tanımak amacıyla çevrimiçi uygulamaları kullanmaktadır. Aynı zamanda çeşitli kameralarla elde edilen video karelerinden ilgili kişinin görüntülerini toplamak için çevrimdışı uygulamalar kullanmaktadır [1]. Birbiriyle örtüşmeyen birçok kamera tarafından çekilen videolarda hedef kişiyi aramak, video gözetimi alanlarında sorun teşkil etmektedir. Bu nedenle, kişinin kimliğinin yeniden belirlenmesi, kişi arama görevinde önemli rol oynayan bir tekniktir.

Spor oyuncusunun kimliğinin yeniden belirlenmesi, bir spor müsabakasında birden fazla kamera görüntüsünde bireysel sporcuları tanımlamayı ve izlemeyi içeren bir bilgisayarlı görü işlemidir. Amaç, oyun boyunca her sporcunun hareketlerinin sürekli ve doğru bir şekilde kaydını oluşturmaktır. Aynı oyuncuyu farklı kamera kayıtlarında duruş farklılıkları, ışıklandırma gibi etkenlerden dolayı eşleştirmek veya farklı oyuncuları ise benzer vücut tipi ve hareketlerinden dolayı ayırt etmek zorlaşmaktadır. Sporcunun kimliğinin yeniden belirlenmesi için kullanılabilecek tekniklerden birkaçı: öznitelik tabanlı yöntemler, derin öğrenme yöntemleri ve hibrit yaklaşımlardır. Öznitelik tabanlı yöntemler, sporcuları kameralar arasında eşleştirmek için renkli histogramlar veya doku tanımlayıcıları gibi el yapımı öznitelikleri kullanmaktadır. Derin öğrenme yöntemleri ise, öznitelikleri doğrudan verilerden öğrenmek için sinir ağlarına başvurmaktadır. Hibrit yaklaşımlar, doğruluğu artırmak için hem öznitelik tabanlı hem de derin öğrenme yöntemlerini birleştirmektedir. Spor oyuncusunun kimliğinin yeniden belirlemenin, spor analitiği ve yayıncılığında birçok pratik uygulaması bulunmaktadır. Ekipler, bireysel sporcuları doğru bir şekilde izleyerek oyuncu performansı hakkında önemli bilgiler elde edebilir veya stratejileri ve antrenman hakkında veriye dayalı kararlar alabilirler. İzleyiciler'e daha sürükleyici bir deneyim sağlamak adına Radyo ve Televizyon

programları bu teknolojiyi aktif olarak kullanmaktadır. Böylece favori oyuncularını yayın boyunca daha yakından takip edebilmelerine olanak tanımaktadır.

Sporda takip için en yaygın kullanılan iki yaklaşım, giyilebilir takip sistemleri ve optik takip sistemleridir. Giyilebilir takip sistemleri, takip edilen oyuncuların transponder takmasını gerektirmektedir. Bir dizi transponder izleme teknolojisi mevcuttur. Bazıları uyduları (GPS, Galileo gibi) kullanırken, bazıları da radyo frekansını kullanmaktadır. Giyilebilir sistemler taşınabilir olmakla beraber farklı ortamlarda kullanılabilirler. Giyilebilir cihazlar kullanılarak oyuncuların kimlikleri otomatik olarak tespit edilmektedir. Söz konusu cihazlar ayrıca ivme ölçer, jiroskoplar, kalp atış hızı monitörleri gibi çeşitli sensör türlerini içerebilmektedir.

Öte yandan optik izleme, oyuncunun ekipmanında herhangi bir değişiklik veya bir transponder yönetimi gerektirmeyen ve müdahaleci olmayan bir yöntemdir. Optik izleme sistemi tarafından çekilen videolardan, oyuncu pozisyonuna ek olarak top pozisyonu, oyuncu duruşu ve video destekli hakemlik için gerekli çekimlerde gibi ayrıntılı bilgiler elde edilebilmektedir. Optik izlemenin sonuçları, izlenen hedefleri görüntüye yansıtarak doğrulanabilmektedir. Oyuncuların kimlikleri genellikle manuel olarak atanmakta ve oyun boyunca yazılım kullanılarak korunmaktadır. Örneğin TRACAB [2], bir optik izleme sistemidir. Örtüşmeler, kapanmalar ve kalabalık durumlar bu sistem için zorlayıcıdır. Kimlikler değişken olduğundan bir insan operatörünün kimliklerin doğruluğunu kontrol etmesi ve gerektiğinde bunları manuel olarak ayarlaması beklenmektedir. Bu hataları tanımak ve doğru kimlikleri otomatik olarak yeniden atamak için kişinin kimliğinin yeniden belirlenmesi tekniklerini kullanmak çözüm olarak önerilmektedir.

Bir kişinin bir kameradan çekilmiş görüntüsü verildiğinde, kişiyi yeniden tanımlamak için, bu kişinin diğer kameralar tarafından çekilmiş bir galeri setinden tanımlanmasını gerektirmektedir. Bu zorlu bir görevdir. Kişinin görünümü, kapatma, aydınlatmadaki farklılıklar, bireyin pozlarındaki değişiklikler, bakış açısı ve arka planda meydana gelen kaos gibi çeşitli zorlukların bir sonucu olarak kameradan kameraya çok farklı görünebilmektedir [3]. Derin öğrenme teknikleri ortaya çıkmadan önce, kişiyi yeniden tanıma üzerine yapılan ilk araştırmalar, manuel olarak daha iyi görsel öznitelikler tasarlamaya ve daha iyi benzerlik ölçümlerini öğrenmeye odaklanmaktaydı. Son yıllarda, derin öğrenme teknikleri, derin öğrenmenin gelişmesiyle birlikte kişiyi yeniden tanımlama da yaygın olarak kullanılmaktadır. Derin öğrenme yöntemleri, geleneksel

yöntemlerden farklı olarak, kişinin görüntü özniteliklerini otomatik ve daha iyi bir şekilde çıkararak, benzerlik ölçümlerini öğrenebilir. Nesne yeniden tanımlama, belirli nesneleri sahneler ve kamera görünüşleri arasında ilişkilendirmeyi amaçlamaktadır. Doğru ve ayırt edilebilir özniteliklerin çıkarılması, yeniden tanımlamanın (Re-ID) ayrılmaz bir parçasıdır ve uzun süredir CNN tabanlı yaklaşımların hakimiyetindedir. Bugüne kadar bu konuda yapılan çalışmaların birçoğunda evrişimli sinir ağları kullanılmıştır. Bu mimari her ne kadar başarılı olsa da bazı dezavantajları bulunmaktadır. CNN'ler görüntüleri yapılandırılmış piksel dizileri olarak ele alıp öznitelik haritaları oluştururlar. Ancak tüm pikseller eşit derecede önem değerine sahip değildir. CNN'ler, önem derecelerine bakılmaksızın bir görüntüdeki tüm piksellere aynı filtreleri uygulamaktadır. Bununla birlikte, ön plan pikselleri genellikle arka plan piksellerinden daha önemlidir. Diğer bir taraftan bütün bilgiler görüntüler arasında paylaşılmaz. Yani her bir evrişim filtresi belirli kavramları (örn. Kenarlar, şekiller) tanımaktadır. Tüm görüntülerin kenarları olduğundan bir "kenar filtresi" veya "şekil filtresi" her zaman değerli olacaktır. Ancak örneğin insan gözü gibi insan özniteliklerini tespit eden bir filtre her zaman kullanışlı değildir. Çünkü insan içermeyen herhangi bir resim üzerinde uygulandığında filtreyi işe yaramaz hale getirecektir. Ayrıca her dönüştürme filtresi, yalnızca yakındaki küçük bir piksel grubu üzerinde çalışabilir. Basit bir değişle, CNN'ler, görüntüdeki yerel ilişkileri yakalamak için bu dönüşüm filtrelerini kullanır. Ancak, CNN daha geniş bağlamları anlamak ve küresel ilişkileri modellemekte zorlanmaktadır. CNN'ye daha fazla derinlik ekleyerek veya filtre boyutunu büyütürken, sorunu çözmeye çalışsa da sorunu fiilen çözmeden modeli ve hesaplamaları daha karmaşık hale getirir. Dolayısıyla, CNN modellerin performansı çok tatmin edici değildir. Öte yandan, dikkat mekanizmasının ortaya çıkışı, CNN'lerdeki detay kaybı sorununu çözmektedir. Vision Transformer'ler (ViT), bu sorunları olmayan yeni bir yöntemdir. Çünkü "küresel" bir ölçekte çalışan öz-dikkat bir mekanizmaya sahiptirler. Dikkat mekanizması, uzak bölgeler arasında anlamsal bir bağlantı kurarak daha iyi bir görüntü algısı sağlar. Ayrıca önemli olmayan pikseller için düşük dikkat puanları hesaplanmakta ve bu da temsillerin daha verimli olduğunu göstermektedir. ViT modelinin içindeki yama boyutu sabittir ve tekdüze bir ölçektir. Doğal dil işleme alanı için ölçek aynıdır. Ancak bilgisayarlı görü alanında görüntünün yama boyutu sabit değildir ve büyük veya küçük olabilir. Bazı sınıflandırma görevlerini gerçekleştirirken yamalar daha önemli olabilir. Ancak bilgisayarlı görüdeki aşağı yönlü görevler için, örneğin hedef tespiti ve piksel seviyesinde segmentasyon gibi işlemler için, yama boyutunun genellikle ayarlanması gerekmektedir. Bu, ViT için hesaplama sorunlarına yol

açabilmektedir. Yamaların boyutu sabit olduğundan dolayı aşağı yönlü görevler için iyi bir çözüm değildir. Swin Transformer'ın ortaya çıkışı, ViT'in bu tarz sorunlarına bir çözümdür. Swin Transformer, sürgülü pencereler kullanan bir ViT modifikasyonudur. Dolayısıyla, kayan pencere mekanizması sayesinde farklı pencereler arasındaki bilgi alışverişini gerçekleştirir. Diğer taraftan ViT pencere boyutu sabit olduğu için ve bu pencereler arasında ilişkiyi yakalayamadığı için sadece genel öznitelikler elde etmektedir. Swin Transformer, bir görüntünün hem genel hem de yerel özniteliklerini yakalayabilen hiyerarşik bir mimari kullanmaktadır. Spesifik olarak, görüntüyü daha küçük parçalara bölerek yerel öznitelikleri yakalamak için bu yamalara öz-dikkat dikkat mekanizması uygulamaktadır. Ardından, genel öznitelikleri yakalamak için farklı yamalardaki öznitelikleri birden çok ölçekte toplar. Bu hiyerarşik yaklaşım, karmaşık uzamsal yapılara sahip görüntüleri işlemede etkili olmasını sağlamaktadır.

Bu çalışmada futbolcu yeniden tanımlama ile ilgili bazı sorunları gidermek adına tüm görüntülerden güçlü görsel öznitelik haritaları oluşturmak için Swin Transformer modeli kullanılmıştır. Ancak, özniteliklerini sınırlamaya devam eden bir sorun bulunmaktadır. Bir futbolcunun küçük takım rozeti veya forma modelindeki ince farklılıklar gibi ayrıntılarını tam olarak yakalayamayabilir. Bu sorunu ele almak için Swin Transformer'ı, bir görüntünün yerel özniteliklerini daha iyi yakalamak adına transformer'lerin ve evrişimli sinir ağlarının (CNN'ler) avantajlarını birleştirerek Swin transformer iyileştirilmiştir. Burada Swin Transformer yapısını geliştirmek için yerel bir algı bloğu kullanılmıştır. RFES-ReID olarak geliştirdiğimiz mimari ağın bir futbolcunun görünümünün küçük ölçekli özniteliklerini tespit etme yeteneğini geliştirebilir. Bu, futbolcuyu yeniden tanımlama görevleri için son derece önemli olabilir. RFES, yerel öznitelik haritalarını birleştirerek, bir futbolcunun görünümündeki ince ayrıntıları ve ince farklılıkları daha iyi yakalayabilir ve bu da futbolcunun yeniden tanımlanmasında daha iyi doğruluk sağlamaktadır. Öğrenmeye dayalı yaklaşımlar genellikle veri dengesizliği ve bir kişiyi güçlü görünüm benzerliğine sahip diğerlerinden ayırt etme konusunda zorluklarla karşılaşır. Nesne yeniden tanımlama performansını iyileştirmek için, yanlış pozitifler ve yanlış negatifler, kayıp fonksiyonunun tasarımındaki ayrılmaz faktörlerin bir parçası olmalıdır. Dolayısıyla bu çalışmada, veri dengesizliği konusu ele alınarak ve bir odak kaybı dahil edilerek modelin, zor örneklerden etkili bir şekilde öğrenmesi kolaylaştırılmıştır. Son olarak test adımında nihai tahminin doğruluğunu artırmak için yeniden sıralama (RR) optimizasyon yöntemi kullanılmıştır. Aşırı uyumu önlemek için,

yeniden sıralama yöntemi önerilen temel modellerle birleştirilerek önemli performans iyileştirmeleri elde edilmiştir. Bu çalışmada gerçekleştirilen deneyler, önerilen yöntemin futbolcu yeniden tanımlama görevinin performansını önemli ölçüde artırabileceğini göstermiştir.

1.1 Kişi Yeniden Tanımlaması

1.1.1 Tanım

Kişiyi yeniden tanımlama görevi, bir kişinin tek bir kameradan alınmış bir görüntüsü verildiğinde, Re-ID, muhtemelen örtüşmeyen görüş alanlarına ve farklı zaman aralıklarına sahip diğer kamera görüntüleri tarafından yakalanan galeri setinden aynı kişinin tanımlanma sürecidir [4]. Yani kişinin daha önce farklı kameralardan çekilmiş tüm görüntülerini elde etmektir. Kişilerin görünümü farklı kameralar arasında önemli ölçüde değişebileceğinden zorlu bir problemdir. Kişiyi yeniden tanımlama gerekliliği, havalimanları, üniversiteler ve alışveriş merkezleri gibi halka açık yerlerde kamu güvenliğini sağlamak amacı ile ihtiyaç olarak görülmektedir.

Kişiyi yeniden tanımlama görevi, klasik tanımlama ve algılama görevlerinden farklıdır. Tanımlama görevi, kim olduğunu belirlemek için bir kişinin kimliğini bir görüntüde belirlemekten oluşmaktadır. Algılama görevi ise, bir kişi olup olmadığını belirtmek için kimliklerini bilmeden arka plandan insanları ayırt etmekten oluşmaktadır. Buna karşılık, Re-ID, bir sorgu görüntüsü alındığında, bu kişinin ne zaman ve nerede görüldüğünü belirleyerek, verilen bir kameraya ilişkin bilgi sunmaktadır. Ayrıca birden fazla kamera kullanarak, kişinin belirli bir zaman aralığı boyunca tahmini izini takip etme olasılığı sağlamaktadır. Re-ID aynı kişiye ait görüntüyü sorgu görüntüsüyle eşleşip eşleşmediğini yanıtlamaktadır.

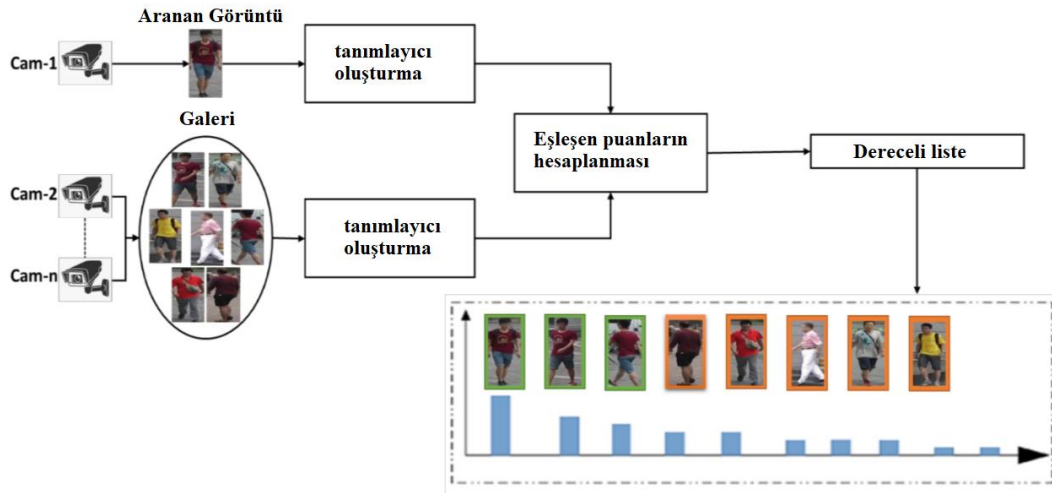
Tipik olarak, bir kişi yeniden tanımlama sistemi üç aşamadan oluşmaktadır: kişi algılama, takip etme ve geri alma (retrieval). Birinci ve ikinci aşamalarda bağımsız bilgisayarlı görüş görevleri olduğu için yürütülen kişi yeniden tanımlama çalışmalarının çoğu üçüncü aşamaya odaklanmıştır. Bu nedenle, burada yürütülen mevcut araştırma, kişinin geri getirilmesi olan üçüncü aşamaya odaklanacaktır. Kişi geri alma süreci şu şekilde formüle edilebilir:

$$i^* = \underset{i \in 1,2,\dots,N}{\operatorname{argmax}} M(Q, G_i) \quad (1.1)$$

Burada i^* , Q sorgu görüntüsünün kimliğidir ve M, Q sorgu görüntüsü ile G galeri görüntüsü arasındaki benzerliği ölçmek için bir mesafe metrik işlevidir. Bu nedenle, kişi yeniden tanımlama görevi genellikle aşağıdaki iki ana adımdan oluşmaktadır:

1. Sorgu görüntüsü ve galeri setindeki her bir görüntü için kişinin görünümünü tanımlayabilen, ayırt edebilen ve aydınlatma, kapanma, poz değişikliği ve diğer zorluklara karşı etkili bir tanımlayıcı veya temsil oluşturma.
2. Uygun bir mesafe metriği kullanılarak görüntüler arasında benzerlik eşleştirmesinin gerçekleştirilmesi ve son olarak, ölçülen benzerlik puanına göre en iyi eşleşen görüntülerin gösterilmesi.

Şekil 1.1, standart kişi Re-ID sisteminin bir tasvirini göstermektedir. Birinci adımda, sorgu ve galeri setindeki her bir görüntü için bir açıklama veya tanım oluşturulur. Ardından, tüm galeri görüntüleri ile sorgu görüntüsü arasındaki benzerlik veya eşleşme puanları hesaplanır. Son olarak, eşleşen tüm puanları azalan şekilde sıraladıktan sonra, en iyi eşleşen görüntülerin sıralı listesi oluşturulur.



Şekil 1.1 Standard bir kişi Re-ID sisteminin blok diyagramı.

Birçok araştırmada, kişi Re-ID sistemlerinin performansında iyileşme sağlamak için uzaklık metriği öğrenme ve öznitelik çıkarımı üzerinde yoğunlaşmıştır. Ancak son yıllarda, derin öğrenme (DL) tabanlı yaklaşımlar, görüntü sınıflandırmadaki başarılarının bir sonucu olarak, kişi Re-ID görevlerinde giderek daha popüler hale gelmiştir.

1.1.2 Kişinin Yeniden Tanımlanması Uygulamaları

Kişi yeniden tanımlama, çeşitli akıllı video gözetim uygulamalarında kullanılmaktadır. Kullanıldığı en tipik senaryo, güvenliği sağlamak için alışveriş merkezi, havaalanı,

otopark, üniversite veya başka herhangi bir alanda belirli bir konumun etrafına birden fazla kameranın yerleştirildiği durumdur. Belirli bir kişinin görüntüsü (sorgu) verilerek farklı kameralardan (galeri) elde edilen bir dizi görüntüde o kişi ile eşleşenler aranmaktadır. Böylece bu kameralar üzerinden o kişiyi takip etmek mümkün olmaktadır. Kişi yeniden tanımlama yöntemleri, aşağıdakiler dahil birçok pratik uygulamada uygulanabilir:

- **Çoklu kamerada kişi takibi:** Bir kişi bir kameranın bakış açısından çıkıp başka kameranın bakış açısına geçtiğinde, takip sisteminin farklı kameralarından alınan bu bağlantısız izler arasındaki ilişkiyi belirlemek için kişi yeniden tanımlamayı kullanmaktadır. Bu, bir kişinin yörüngesinin yeniden oluşturularak bir noktadan diğerine hareket yolunu takip etmektedir. Örneğin, kayıp çocukların, yaşlıların veya suçluların bulunabilmesi için güvenlik kameraları tarafından farklı açılardan alınmış görüntülerin kullanılması olarak ifade edilebilir.
- **Kişi bulma (Person retrieval):** Burada Re-ID, bir tanıma görevi ile ilişkilendirilir. Önce bir hedef kişiyi içeren belirli bir sorgu sağlanır ve ardından ilgili tüm örnekleri büyük bir veri tabanında bulmak için bir arama yapılır. Bu nedenle, Re-ID görevi görüntü bulma için kullanılır ve tipik olarak sıralanmış listeler ve ilgili öğeler sağlanmaktadır. Adli tıp veri tabanlarında kullanılan kişilerin görüntülerini geri alımı örnek olarak verilebilmektedir.
- **İnsan davranışı tanımlama ve analizi:** Re-ID, insanların değişik davranışlarını tanımlamaya ve analiz etmeye yardımcı olmaktadır. Örneğin, kamu güvenliğinin değerlendirilmesi için suçlu veya diğer şüpheli davranışları belirlemek veya müşterilerin alışveriş trendlerini gözlemlemek için müşterinin dokunarak inceleyerek ve farklı güvenlik kameraları altında mağazalarda ürünleri deneyerek analiz etmeyi içermektedir.
- **İnsan-makine etkileşimi:** Son zamanlarda insan-makine etkileşiminde kişi Re-ID uygulamalarının sayısının hızla artması bu sorunun ne kadar önemli olduğunu göstermektedir. Robotik gibi bir alanda yeniden kimlik belirleme sorununu çözmek, işbirlikçi olmayan hedef tanıma olarak adlandırılır. Muhatabın kimliği gizli tutularak, robotun sürekli olarak etrafındaki insanlar hakkında bilgi sahip olması sağlanır.
- **Spor analizi:** Basketbol ve futbol da dahil olmak üzere büyük ölçekli sporlar için kişi Re-ID'si, daha iyi bir analiz sağlamak için oyuncuların davranışlarını otomatik olarak takip edip kaydedebilir.

1.1.3 Kişinin Yeniden Tanımlanmasının Zorlukları

Bir kişinin görünümü, farklı kameraların kullanılması nedeniyle büyük ölçüde değişebildiği için bazı zorluklarla karşılaşmaktadır. Bu durumda, bir kişinin görüntü temsilindeki görsel öznitelikler çeşitlilik gösterebilir. Kişiyi yeniden tanımlama göreviyle ilgili olarak Şekil 1.2'de gösterilen bazı önemli zorluklar aşağıda açıklanmıştır.

- **Aydınlatma varyasyonları:** Gün ışığı miktarı, iç ve dış mekân arasındaki renk farkı, gölgelerin etkisi gibi bir dizi nedenden dolayı aydınlatma koşulları bir kamera sahnesinden diğerine değişebileceğinden, kişinin görünümü ve rengi değişebilmektedir. Bu nedenle, kişinin görünümü ve renginde değişiklikler meydana gelebilir. Dolayısıyla aynı kişinin görüntüleri bazen farklı olabilir, bu da eşleştirme sürecinde belirsizliği artırmaktadır.
- **Kapanma:** Büyük ve kalabalık halka açık yerlerde, bir kişinin tüm vücudunun engel olmadan net bir görüntüsünü elde etmek zor olabilir. Bu tarz mekanlarda, insanlar bölgedeki diğer insanlar veya yapılar tarafından kısmen veya tamamen kapanmış olabileceğinden, kameraların doğrudan yayaları çekemeyebilirler. Dolayısıyla kişinin bazı önemli veya benzersiz kısımları görülemiyorsa eşleştirme işlemi başarısız olabilir.
- **İnsan vücut pozisyon değişiklikleri:** Poz varyasyonları, kişinin vücut bölgesi görünürlüğünün birden fazla kamera sahnesinde değiştiği anlamına gelmektedir. Poz varyasyonlarının bir kişinin görsel görünümü üzerindeki büyük etkisi nedeniyle, sonuçta elde edilen görüntüleri tahmin etmek ve eşleştirmek son derece zorlu bir iştir. Bu görüntüler genellikle düşük çözünürlüklüdür ve kalitesi düşüktür. Ayakta duran bir pozda eğilen, koşan veya oturan bir kişiyi tespit etmek için öğrenilen bir model başarısız olacaktır.
- **Kıyafet veya aksesuar varyasyonları:** Re-ID algoritmaları, genellikle bir kişinin genel görünümüne ve farklı sahnelerde aynı giysileri giydiği varsayımına dayanmaktadır. Ancak gerçekçi bir ortamda, bir kişinin farklı kamera sahnelerinde farklı kıyafetler giymesinin veya farklı nesneler taşımasının olasılığı yüksektir. Örneğin bir kişi sırtından çantasını çıkarıp elinde taşıdığı anda, şapkasını veya ceketini çıkardığında ya da giydiğinde kamera sahnelerinde genel görünümünü değiştirmiş gibi algılanır. Ayrıca, farklı insanların benzer giysiler giyme olasılığı yüksektir. Bu, farklı insanların görsel görünümünün benzerliğine yol açar ve eşleştirme sürecinde

belirsizlięi artırır. Bu durumda kiřinin farklı grsel grnmlerinden ayırt edici znitelikler bulmak daha zordur.

- **Kamera bakıř aısı deęiřimi:** Kamera yksekliklerindeki farklılıklar, kiři ile kameralar arasındaki mesafeler ve kiřilerin kameralara ynelik farklı bakıř aıları, kiřinin řekil ve jestlerinin farklı grntleme aıları altında byk lde deęiřmesine neden olabilir. Dolayısıyla, aynı bakıř aısından ekilen farklı kiřilerin grntleri, farklı grntleme aılarından ekilmiř aynı kiřinin iki grntsne gre daha fazla benzerlik gsterebilir. Bakıř aısı deęiřkenlięi, aynı zamanda sınıf ii deęiřkenlięi ve sınıflar arası karıřıklıęı artıran en zorlu problemlerden biridir.
- **Arka plan daęınıklıęı:** Farklı kameraların kullanımı benzersiz arka plan bilgisi retmektedir. Arka plan bilgisini ortadan kaldırmak amacıyla, kiři yeniden tanımlama yntemleri genellikle kiři algılama algoritmalarının dndrdę kırpılmıř kiři grntlerinde alıřmaktadır. Ancak, mevcut kiři algılama algoritmalarının performansı Re-ID amacı iin yeterince doęru deęildir, yani algılamalar ok fazla arka plan ierebilir veya sadece kiřinin bir kısmını ierebilir. Bu nedenle, insan vcut blgeleri grntler arasında iyi bir řekilde hizalanmamıř olur ve bu durum, mevcut yntemlerin oęunda Re-ID'nin performansını ciddi řekilde etkilemektedir.
- **Grnt znrlę:** Genellikle byk ve geniř alanlarda gvenlik ve gzetim amacıyla kullanılan kameralar duvarlara yksek yerleřtirilir ve yayalar genellikle kameralardan uzaktadır. Yksek znrlkl kameralar olsa bile, bir kiři iin grnt grece dřk znrlę sahip olabilir. Dřk grnt znrlę ayrıca bir kiřinin grnmn farklı hale getirebilir, bu da eřleřtirme srecinde belirsizlięi artırır.
- **Etiketli veri eksiklięi:** Mevcut kiři Re-ID yntemlerinin oęunun derin ęrenme modellerine dayanmaktadır. Her trl varyasyona karřı dayanıklı iyi bir modeli denetimli bir řekilde eęitmek, yeterli miktarda etiketli veri olmadan mmkn deęildir. Kameradan veri toplamak ve bunları el ile etiketlemek olduka maliyetli olduęu iin, genellikle kiři Re-ID veri setlerinde her kiřinin etiketli grnt sayısı sınırlıdır. Bu, iyi bir derin modeli eęitmek iin yetersizdir. Bu nedenle, iyi bir genelleme yeteneęi elde etmek iin karmařık algoritmaların, byk miktarda etiketli veri kullanılarak ęretilmesi gerekmektedir.



Şekil 1.2 Bazı kişi Yeniden Tanımlama örnekleri.

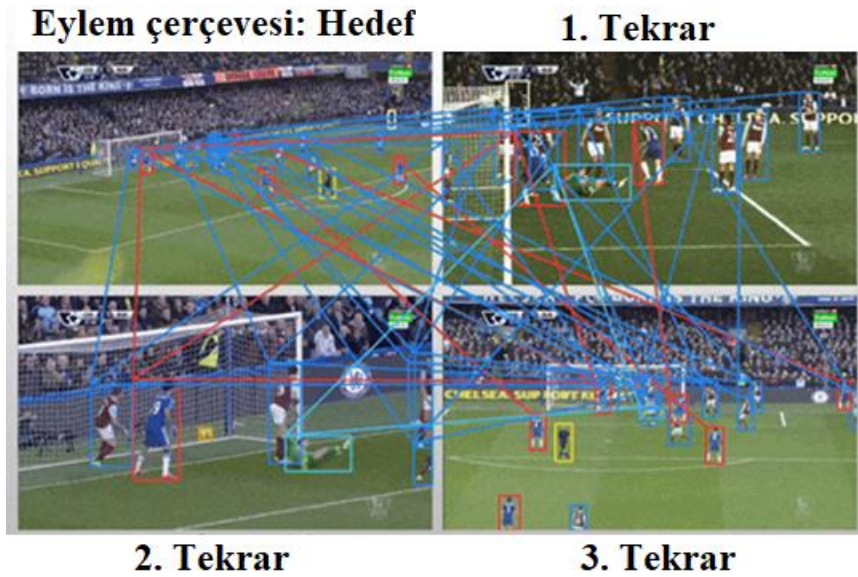
1.2 Futbol Takımlarında Sporcu Yeniden Tanımlaması

Bir futbolcunun yeniden tanımlanması, daha önce kimliği tespit edilmiş ancak oyun sırasında ilerleyen zamanda tekrardan tanıma ve aynı kimlik atama sürecini ifade etmektedir. Oyuncunun farklı müsabakada veya liglerde oynaması, bir oyuncunun görünüşünü değiştirmesi, formasında farklı bir numara giymesi gibi çeşitli durumlar gözlemlenebilir. Bir futbolcunun yeniden tanımlanması, aşağıdakiler de dahil olmak üzere çeşitli yöntemler kullanılarak yapılabilir: 1. Yüz tanıma teknolojisi: Futbolcunun yüzünün bir görüntüsünü bilinen bir görüntü veri tabanı ile karşılaştırmayı içermektedir. Ancak bu yöntem çok gerçekçi olmayabilir çünkü oyunu sırasında alınan görüntüler çok düşük çözünürlüğe sahip veya çok uzaktan olduğu için futbolcunun yüzü belli olmayabilir. 2. Forma numarası: Futbolcunun forma numarası kullanarak tanınması yöntemi, bazı zorluklarla karşılaşabilir. Futbolcular numaralarını değiştirebilir, farklı formalar giyebilir, diğer futbolcular tarafından kapatılabilir veya farklı bakış açılarından görüntü alınabilir. Bu nedenle, forma numarası bazında tanıma yöntemi her durumda güvenilir olmayabilir. Özetle, bir futbolcunun kimliğinin yeniden belirlenmesi, bir

futbolcunun kimliğinin yüz tanıma teknolojisi veya forma numaraları gibi yöntemler ile çok başarılı olmayabilir.

Genel olarak bir futbol müsabakasının yayınlanan bir videosunda, müsabakadaki her ilgi anı bir "eylem" olarak adlandırılır. Yayınlanan videoda hareketin ilk görüldüğü kareye "eylem karesi" denilmektedir. Bir yayın videosu genellikle diğer kamera bakış açılarından yakalanan tekrarları içermektedir. Bu, aynı eylemin yayınlanan videoda daha sonraki karelerde görünebileceği anlamına gelmektedir. Aynı eylemi anlatan ve yayınlanan videoda daha sonra görünen diğer kareleri "tekrar kareleri" olarak adlandırılır. Bu nedenle, bir eylem karesi ve ona karşılık gelen tekrar kareleri yayınlanan videoda farklı zamanlarda görünse de müsabakayla ilgili yaklaşık olarak aynı zamanı temsil etmektedirler. Örneğin, bir müsabakanın 5'inci dakikasında meydana gelen bir eylem, 100 numaralı karede eylem karesine ve 1000 numaralı karede bir tekrar karesine sahip olabilir.

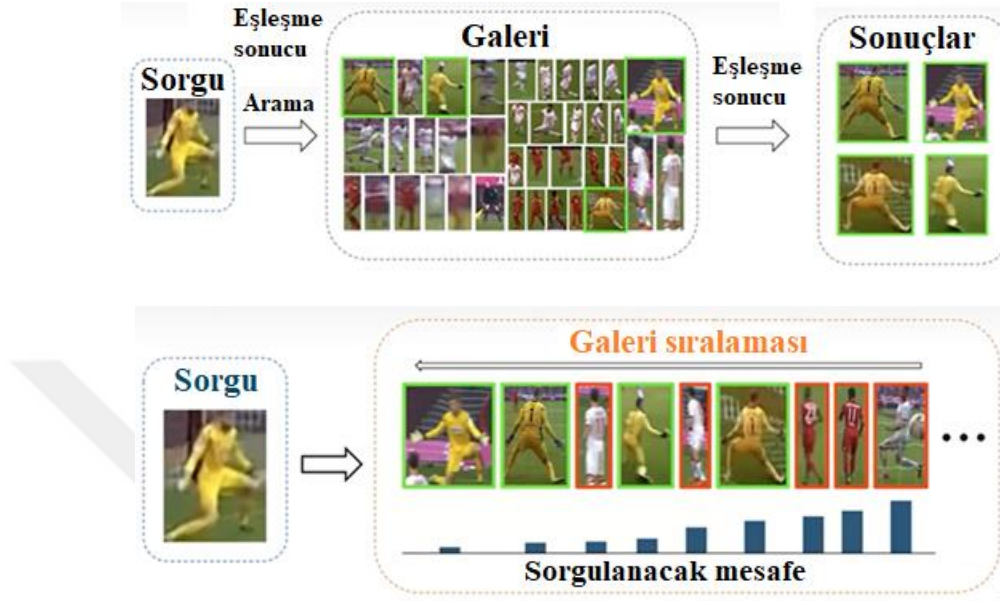
Şekil 1.3'de görüldüğü gibi amaç, eylem karesindeki bir futbolcuyu farklı tekrar karelerinde yeniden tanımlamaktır. Bu, birden fazla kamerada futbolcuları takip etmek, öne çıkan videolar oluşturmak, hakemler için sanal yardım vb. için oldukça kullanışlıdır. Futbolcu algılama bu çalışmanın kapsamı dışındadır. Bu tez çalışmasında eylem ve tekrar kareleri için futbolcu tespitlerine zaten sahip olduğu varsayılmıştır.



Şekil 1.3 Futbolcu yeniden tanımlama örnekleri.

Şekil 1.1, standart kişi Re-ID sisteminin bir tasvirini göstermektedir. Yani mantık futbolcu Re-ID'si içinde geçerlidir. Birinci adımda, galeri ve sorgu setindeki her bir

görüntü için bir tanımı veya öznitelikleri oluşturulur. Ardından Şekil 1.4’de olduğu gibi tüm galeri görüntüleri ile sorgu görüntüsü arasındaki benzerlik veya eşleşme puanları hesaplanır. Son olarak, eşleşen tüm puanları azalan şekilde sıraladıktan sonra, en iyi eşleşen görüntülerin sıralı listesi oluşturulur.



Şekil 1.4 Futbolcu Re-ID sisteminin blok diyagramı.

1.2.1 Futbolcu Yeniden Tanımlama Uygulamaları

Futbolcu yeniden tanımlama uygulamaları, temel olarak futbolcu performans analizi, futbolcu takip etme ve futbol stadyumlarında güvenlik gibi futbolla ilgili süreçlerin doğruluğunu ve verimliliğini artırmakla ilgilidir. Futbolcu yeniden tanımlamanın bazı özel uygulamaları aşağıda açıklanmıştır:

- **Performans analizi:** Futbolcuların yeniden tanımlanması, birden fazla maç ve sezon boyunca performanslarının doğru bir şekilde izlenmesine ve analiz edilmesine yardımcı olmaktadır. Bu veriler oyuncuların güçlü ve zayıf yönlerini belirlemek, oyuncu profilleri oluşturmak ve takım performansını iyileştirmek için kullanılabilir.
- **Futbolcu takip etme:** Futbol kulüpleri ve takımları, diğer takımlara ödünç verilen veya alt liglerde futbolcuların ilerleyişlerini izlemek için yeniden tanımlamayı kullanırlar. Bu, işe alım için potansiyel futbolcuların belirlemelerine ve takımlarına uygunluklarını değerlendirmelerine yardımcı olmaktadır.
- **Güvenlik:** Futbolcuların ve seyircilerin güvenliğini tehdit edebilecek kişileri belirlemek ve izlemek için de kullanılmaktadır. Tanınmış hooliganları takip etmek,

kalabalık davranışlarını gözlemlemek ve yetkisiz kişilerin sınırlı alanlara erişimini engelleme gibi faaliyetleri içermektedir.

- **Pazarlama ve hayran etkileşimi:** Taraftarların favori futbolcularını kolayca tanıyıp izlemelerine olanak tanıyarak, hayran katılımını artırabilir. Ayrıca kulüpler ve sponsorlar, futbolcuların performans ve popülerlikleri doğrultusunda özelleştirilmiş pazarlama kampanyaları oluşturmak için bu teknolojiden faydalanabilirler.

Özet olarak, futbolcu yeniden tanımlama uygulamaları arasında performans analizi, futbolcu takip etme, güvenlik ve pazarlama ve taraftar etkileşimi yer almaktadır. Bu uygulamalar futbolla ilgili süreçlerin doğruluğunu ve verimliliğini artırmaya ve futbolcuların, kulüplerin ve taraftarların genel deneyimini geliştirmeye yardımcı olmaktadır.

1.2.2 Futbolcunun Yeniden Tanımlanmasının Zorlukları

Futbolcuyu yeniden tanımlamanın birçok potansiyel faydası olsa da doğru ve güvenilir sonuçlara ulaşmak için ele alınması gereken çeşitli zorluklar da mevcuttur. Tabi bu zorluklar normal kişi yeniden tanımlamada bazı açılardan farklı ve daha zordur. Futbolcuyu yeniden tanımlamanın ana zorluklarından bazıları aşağıdaki gibidir:

- **Giyim ve ekipman:** Futbolcular genellikle benzer forma ve ekipman giyerler, bu da özellikle kalabalık stadyumlarda veya düşük kaliteli çekimlerde onları birbirinden ayırt etmeyi zorlaştırabilir.
- **Çeşitli vücut hareketi yönelimleri:** Çeşitli vücut hareketleri yeniden tanımlamanın doğruluğunu etkileyebilir. Koşma, zıplama, dönme, tekme atma veya kayma gibi farklı vücut hareketleri genel olarak futbolcularda yeniden tanımlamanın doğruluğunu önemli ölçüde etkileyebilir. Bu zorlukların üstesinden gelmek için spor analitiği şirketleri, futbolcuları takip etmek ve onları hareket halindeyken bile doğru bir şekilde tanımlamak için genellikle gelişmiş bilgisayar algoritmaları ve makine öğrenimi teknikleri kullanırlar.
- **Futbolcular arasında örtüşme ve kapanmalar:** Tıkanma veya kapanma, bir futbolcunun diğer bir futbolcunun yüzünü ve ayırt edici özelliklerini görünmez hale getirdiği durumlarda meydana gelmektedir. Örneğin bir futbolcu başka bir futbolcunun arkasında durursa yüzü kısmen veya tamamen gizlenebilir. Bu da doğru bir şekilde tanımlamayı zorlaştırır. Kalabalık alanlar, özellikle de ceza sahası gibi birden fazla futbolcunun pozisyon için itişip kakışması, birbirlerinin görüşünü

engelleyebilir. Genel olarak futbolcular arasındaki temas ve kapanma, futbolcuların yeniden tanımlanmasını daha zor hale getirebilir. Ancak gelişmiş makine öğrenme teknikleri, bu zorlukların üstesinden gelmeye ve ayırt edici öznitelikleri kısmen veya tamamen gizlenmiş olsa bile futbolcuları doğru bir şekilde tanımlamaya yardımcı olabilir.

- **Aydınlatma ve kamera açıları:** Zayıf aydınlatma koşulları ve sıra dışı kamera açıları, futbolcuların net görüntülerini elde etmeyi zorlaştırabilir ve bu da tanımlama yöntemlerinin doğruluğunu etkileyebilir.
- **Sınırlı veri:** Futbolcuları doğru bir şekilde yeniden tanımlamak için, karşılaştırılacak geniş bir görüntü ve veri tabanı olması gerekmektedir. Ancak bazı durumlarda, öznitelikle alt düzey ligler veya daha eski müsabakalar için sınırlı veriler bulunmaktadır.
- **Veri gizliliği:** Yüz tanıma teknolojisinin ve diğer tanımlama yöntemlerinin kullanılması, veri gizliliği ve güvenliği ile ilgili endişeleri artırmaktadır. Kişisel verileri korumak ve kötüye kullanılmamasını sağlamak için uygun önlemlerin alınması gerekir. Dolayısıyla bu alanda veri toplamak çok daha zor hale gelmektedir.



(a)



(b)



(c)



(d)

Şekil 1.5 Çeşitli zorlukların bazı örnekleri: (a) Giyim benzerliği; (b) Örtüşme ve kapanma; (c) ve (d) düşük çözünürlük ve farklı vücut hareketleri.

Özetle, futbolcunun yeniden tanımlanması, Şekil 1.5’de gösterildiği gibi görünüm, giyim ve ekipman, aydınlatma ve kamera açıları, çeşitli vücut hareketleri, oyuncular arasında örtüşme ve kapanma, sınırlı veri ve veri gizliliğindeki değişiklikler dahil olmak üzere çeşitli zorluklarla karşı karşıyadır. Doğru ve güvenilir sonuçlara ulaşmak için bu zorlukların ele alınması gerekmektedir.

1.3 Motivasyon ve Araştırma Sorunu

Futbolcuların yeniden tanımlanması, araştırmacılar, antrenörler ve spor organizasyonları için büyük ilgi gören bir görevdir. Çoğu durumda, mesafe, kapanma veya düşük görüntü kalitesi gibi faktörler nedeniyle oyunlar veya antrenman seansları sırasında futbolcuları doğru bir şekilde belirlemek zor olabilir. Futbolcuları doğru bir şekilde yeniden tanımlayabilmek, performansları hakkında değerli bilgiler sağlayabilir ve antrenörlerin futbolcu seçimi, antrenman ve oyun stratejisi hakkında daha bilinçli kararlar almasına yardımcı olabilir. Ek olarak futbolcuların yeniden kimlik tespiti video izleme, güvenlik ve takip gibi alanlarda da pratik uygulamalara sahip olabilir. Örneğin futbol maçları gibi büyük ölçekli etkinliklerde futbolcuların davranışlarını izlemek ve hatalı futbolcuları belirlemek çok önemlidir. Bugün var olan futbol video gözetim sistemlerinin çoğu yalnızca video yakalayabilir, kaydedebilir. Bu nedenle, analiz görevi yalnızca insan operatörlere bırakılmıştır. Bu yüzden büyük ölçekli bir takip sisteminde, bu tür insan tabanlı takip etme çok zaman alıcıdır ve ayrıca çoğu zaman kısıtlı doğruluk sağlamaktadır.

Futbol oyunlarının karmaşıklığı ve sahadaki çok sayıda oyuncu, oyuncuların görünümündeki ve kıyafetlerindeki değişkenlik gibi çeşitli faktörler, futbolcuların benzersiz özelliklerini belirlemeyi ve doğru bir şekilde eşleştirmeyi zorlaştırır. Bu nedenle, futbolcuların yeniden tanınması, bu zorluklarla başa çıkabilen özel algoritmalar ve teknikler gerektiren karmaşık bir görevdir. Öznitelikle, kapanma ve kamera görüş açısındaki değişiklikler, zaman içinde oyuncuların doğru bir şekilde takip edebilmesini zorlaştırabilir. Bu zorlukların üstesinden gelmek için, alandaki araştırmacılar, öznitelik tabanlı yöntemler, derin öğrenme tabanlı yöntemler ve mekânsal-zamansal (spatiotemporal) modeller dahil olmak üzere çeşitli yaklaşımlar kullanmaktadırlar. Bununla birlikte, sağlam ve doğru yeniden tanımlama sistemlerinin geliştirilmesi aktif bir araştırma alanı olmaya devam etmektedir. Bu alandaki temel araştırma sorularından bazıları şunlardır:

- Aydınlatma, poz ve giyimdeki değişikliklere karşı değişmeyen sağlam öznitelik temsillerini nasıl geliştirebiliriz?
- Kapanma ve kamera bakış açısındaki değişiklikleri nasıl etkili bir şekilde ele alabiliriz?
- Doğruluğu artırmak için geçici bilgileri yeniden tanımlama modellerine nasıl dahil edebiliriz?
- Futbol maçları veya video izleme senaryoları gibi gerçek dünya ortamlarında yeniden tanımlama sistemlerinin performansını nasıl değerlendirebiliriz?

Futbolcu yeniden tanımlama, spor video gözetim sistemlerinde kullanılan kritik bir araç haline gelmiştir. Bu yöntem, bir yayın ağındaki birçok kamera tarafından yakalanan videolarda hedef bir futbolcuyu aramaktan ibarettir ve bunu yaparken futbolcunun tam vücut görünümünden yararlanır. Diğer Re-ID türleriyle karşılaştırıldığında örneğin, forma veya yüz tanıma gibi, düşük çözünürlüklü kameralarla çalışabilme avantajına sahiptir. Bir futbolcunun görünümü farklı kameralar arasında önemli ölçüde değişebileceğinden, futbolcu yeniden tanımlama daha zor bir problem haline gelmiştir. Kişi Re-ID'sinin derin öğrenme yöntemleri kullanılarak önemli ölçüde geliştirilmesine rağmen, bahsedilen ek zorluklar nedeniyle hala çözülmemiş birçok sorunun bulunmaktadır. Bu durum, futbolcu Re-ID'nin geliştirilmesi için büyük potansiyel taşıdığını ve araştırmacılar için heyecan verici bir alan olduğunu göstermektedir.

Futbolcu yeniden tanımlaması, kişi yeniden tanımlamasının bir alt kümesi olarak düşünülebilir. Genelde bugüne kadar yapılan bütün çalışmalar kişi Re-ID'sine odaklanmıştır.

Futbolcu Re-ID'nin birçok sorunu ve zorluğu vardır. İlgili araştırma problemleri şu şekilde sıralanabilir:

- Bilgisayar ve GPU tabanlı sistemlerdeki gelişmelerin sonucu olarak, kişi Re-ID problemlerini çözmek için daha sofistike ve güçlü yöntemler kullanma olasılığı artmıştır. Derin öğrenme tabanlı yöntemler, özellikle CNN'ler, görüntü sınıflandırma görevlerinde başarılı bir şekilde uygulandıkları için giderek daha popüler hale gelmiştir. DL tabanlı çözümlerin başarısı, eğitim için mevcut olan veri hacmi ile ilişkilidir. CNN'lerin yüksek doğruluk sağlaması için büyük miktarda veriye ihtiyaçları vardır. Özellikle nadir olaylar için veri bulmak zor olabilir. CNN'lerin eğitimi için veri dengesi önemlidir. Veriler arasında dengesizlik olduğunda, CNN'ler

belirli sınıflara aşırı uyum sağlayabilir. CNN'lerin öznetelik çıkarmada bazı zorlukları vardır. Öznetilikle belirli durumlarda sadece yakın komşuluklara bakıldığı için yetersiz öznetelikler üretebilirler ve bunun sonucunda düşük doğruluk oranları elde etmektedir. Bu engeller, derin modelleri eğitmek ve futbolcu Re-ID görevinde daha iyi sonuçlar üretmek için alternatif bir öğrenme yaklaşımı bulma ihtiyacını ortaya koymaktadır.

- Son zamanlarda bazı araştırmacılar Transformer'ler kullanımına dayalı yeni yöntemler önermişler. Transformer'ler, CNN'lerin aksine daha az sayıda katman ile daha iyi sonuçlar verebilmektedirler. Ayrıca transformer'ler, birçok farklı ölçekteki öznetelikleri daha iyi öğrenebilir ve bu nedenle daha iyi bir öznetelik çıkarma yeteneğine sahiptirler. Transformer'ler, "küresel" bir ölçekte çalışan öz-dikkat bir mekanizmaya sahiptirler. Dikkat mekanizması, birbirinden uzaktaki görüntü bölgelerini anlamsal olarak bağlayarak daha iyi görüntü algısı sunmaktadır. Ayrıca önemli olmayan pikseller için düşük dikkat puanları hesaplanmakta ve bu da temsillerin daha verimli olduğunu göstermektedir. Ancak tüm görüntüyü dikkate alarak daha çok genel öznetelikler elde etmektedir. Bu nedenle, eğitim sürecinde kullanılmak üzere yerel ve küresel öznetelikleri içerecek şekilde daha fazla öznetelik çıkarmanın bir yolunun bulunmasına ihtiyaç vardır.
- Swin Transformer, görüntünün genel ve yerel özneteliklerini yakalamasına izin veren hiyerarşik bir mimari kullanılmaktadır. Yerel özneteliklerin yakalanması için görüntüyü daha küçük parçalara bölerek parçalar üzerinde öz-dikkat mekanizması uygulamaktadır. Bu sayede, her parça kendi içindeki önemli öznetelikleri vurgulamaktadır. Genel öznetelikler ise farklı parçalardaki öznetelikleri birden çok ölçekte toplamak suretiyle elde edilir. Bu da görüntünün daha büyük yapılarını ve genel özneteliklerini temsil etmektedir. Bu hiyerarşik yaklaşım, karmaşık uzamsal yapılara sahip görüntülerin işlenmesinde etkili olmasını sağlamaktadır. Ancak çıkarılan öznetelikleri sınırlamaya devam eden bir sorun vardır; bir futbolcunun görünümünün küçük takım rozetleri veya forma modellerindeki farklılıklar gibi ince detayları ve ayrıntılarını tam olarak yakalayamayabilir.
- Önceki çalışmalarda, Transformer'ler ve CNN'ler gibi yöntemler kullanılarak elde edilen özneteliklere dayalı olarak kişi Re-ID performansı iyileştirilmiştir. Ancak, futbol videolarında bahsedilen zorluklar nedeniyle bu yöntemler yeterince başarılı olmayabilir. Dolayısıyla Kapanma ve kamera bakış açısındaki değişikliklerin etkisini ele almak, poz ve giyimdeki değişikliklere karşı değişmeyen sağlam öznetelik

temsillerini elde etmek ve doğruluğu artırmak hala futbolcu Re-ID'si için büyük bir problemdir.

1.4 Araştırma Hedefleri ve Katkıları

Bu tezin ana amacı, futbolcu yeniden tanımlama için derin öğrenme tabanlı çözümleri araştırmak ve önermektedir. Önerilen yapı, futbolcu yeniden tanımlama performansını iyileştirmek ve bazı zorlukları ele almak için tasarlanmıştır.

Bu araştırmanın sağladığı temel katkılar aşağıda verildiği gibi özetlenmiştir:

1. Nesne sınıflandırmasında kullanılmak üzere geliştirilmiş temel bir Swin transformer modelini futbolcu Re-ID'sinde kullanılmak üzere güçlü ve ayırt edici öznetelikler çıkarmak için uyarlamak.
2. Ardından futbolcu yeniden tanımlama problemi alanına daha uygun olması için RFES-ReID adlı yöntem olarak yerel algı blok yaklaşımını kullanarak Swin transformer modelini geliştirmek.
3. Önerilen SoccerNet-v3 veri seti üzerinde futbolcu Re-ID için sınırlı sayıda çalışmanın olduğunu belirterek, bu veri seti üzerinde yapılan çalışmalar arasında RFES-ReID modelinin en iyi sonuçları verdiğini vurgulama.
4. Öğrenmeye dayalı yaklaşımların genellikle veri dengesizliği ve güçlü görünüm benzerliği gibi zorluklarla karşılaştığını belirtme. Bu zorlukları ele almak için modelin etkili bir şekilde zor örneklerden öğrenmesini sağlamak için veri dengesizliği konusuna odaklanma ve birleştirilmiş kaybı kullanma.
5. Performansı artırmak ve aşırı uyumu önlemek için yeniden sıralama (RR) yöntemini önerilen temel modelle birleştirme.
6. Önerilen yöntemin etkinliğini değerlendirmek ve performansını doğrulamak için standart bir değerlendirme protokolü olan kümülatif eşleştirme öznetelikleri (CMC) eğrisi ve ortalama hassasiyet (mAP) ölçümünü kullanma. Bu ölçütler, önerilen yaklaşımın diğer literatürdeki yöntemlerle karşılaştırılabilir bir şekilde değerlendirilmesini sağlar.

1.5 Tezin Genel Bakışı

Bu tez aşağıda kısaca özetlenen altı bölümden oluşmaktadır:

Bölüm 1: Tezin girişı, kişı ve futbolcu yeniden tanımlama görevine, araştırmanın ana motivasyonlarına, sorunun ne olduğunu ve araştırma hedeflerinin yanı sıra katkılarına genel bir bakış sunmaktadır.

Bölüm 2: Farklı kişı yeniden tanımlama yaklaşımlarına, veri kümelerine ve değerlendirme metriklerine ilişkin bilgi sağlamaktadır ve derin öğrenme tekniklerini vurgulamaktadır.

Bölüm 3: Literatür taraması, kişı ve futbolcu yeniden tanımlama ile ilgili mevcut çalışmalar ve bu çalışmanın yer aldığı boşluğun belirginleştirilmesi sunulmaktadır.

Bölüm 4: Bu bölümde metodoloji ele alınarak, önerilen kişı Re-ID çerçevesi ve uygulamanın teknik detayları açıklanmaktadır.

Bölüm 5: Önerilen yöntemi değerlendirmek için yapılan deneylerin sonuçlarını sunarak sonuçları tartışmaktadır.

Bölüm 6: Son bölüm, sunulan çalışmayı özetleyerek tezdeki bulguları, katkıları ve gelecekteki mümkün çalışmaları tartışmaktadır.

Bu bölümde, kişinin yeniden tanımlama görevini destekleyen ilgili kavramları ve çalışmaları kapsamlı bir şekilde sunulmuştur. Derin öğrenme tekniklerine, özellikle kişi yeniden tanımlama için uygulanan CNN ve Transformer'lere genel bir bakış verilmiştir. Son olarak, bu bölümün sonunda bazı iyi bilinen kıyaslama veri kümeleri ve değerlendirme yöntemleri sunulmuştur.

2.1 Kişi Yeniden Tanımlanmasının Mevcut Durumuna Genel Bir Bakış

Gözetim sistemlerinin giderek artan popülaritesiyle birlikte, insanların tespit edilmesi, takip edilmesi ve son zamanlarda kişi yeniden tanımlama sorunları üzerine birçok araştırma gerçekleştirilmiştir. Son yıllarda, tek bir kamera içinde insan tespiti ve çoklu kişi takibi gibi zorlu problemler, önemli ölçüde doğruluk, verimlilik ve sağlamlık açısından gelişme göstermiştir [5], [6]. Bu nedenle, bu çalışmada insan tespiti ve kişi takibi göz ardı edilmiştir. Ana odak noktası, farklı zamanlarda ve farklı yerlerde çeşitli güvenlik kameraları tarafından kaydedilen aynı kişinin farklı görüntülerini bulma ve eşleştirme süreci olan kişi yeniden tanımlama üzerine odaklanmaktadır. Kişi yeniden tanımlama süreci genellikle iki temel aşamadan oluşur. İlk aşama, sağlam öznitelik tanımlayıcılarının veya temsillerinin çıkarılmasıdır, bu temsiller kişileri hem tanımlayabilen hem de ayırt edebilen bilgileri içerir. İkinci aşama ise, çıkarılan öznitelik tanımlayıcıları arasında uygun bir mesafe metriği kullanarak benzerlik eşleştirme prosedürünün geliştirilmesidir. Kişi yeniden tanımlama araştırmaları, bilinen kaynaklara göre 2005 yılında başlamıştır. Bu terim, Amsterdam Üniversitesi'nden W. Zajdel, Z. Zivkovic ve B.J.A. Krose'un çoklu kamera takibi üzerine yaptıkları çalışmalarla ortaya atılmıştır. Bu çalışmalar, bir kişinin görüş alanından çıkıp daha sonra tekrar girdiğinde onu yeniden tanımlamayı hedeflemiştir [7]. Ardından bir yıl sonra, Gheissari ve arkadaşları, ön plan algılaması için uzamsal-zamansal bir bölümlendirme algoritması uygulandıktan sonra insanların sadece görsel ipuçlarını kullandılar [8]. Bu çalışma, kişi yeniden tanımlamayı çoklu kamera takibinden ayırdığı için, bağımsız bir bilgisayarlı görüş görevinin başlangıcını göstermektedir [9]. Son yıllarda pek çok kişi yeniden tanımlama yöntemi önerilmiş ve performansları önemli ölçüde geliştirilmiştir. Bu yöntemler öznitelik temsili ve benzerlik eşleme sorunlarını ele almak için bilgisayarlı görüş

tekniklerinden yararlanarak geliştirilmiştir. Son zamanlarda derin öğrenme yöntemleri, kişi yeniden tanıma alanında kullanılmış ve elde edilen sonuçlar mevcut en iyi performansı sunmaktadır. Literatürdeki bazı derleme ve kapsamlı incelemelere [3], [9]–[12] kişi yeniden tanımlama yöntemlerini, geleneksel el yapımı ve derin öğrenme tabanlı olmak üzere iki kategoriye ayırmaktadır.

2.1.1 Geleneksel El Yapımı Kişi Re-ID Yöntemleri

Kişi Re-ID için önerilen geleneksel yaklaşımlar, el yapımı tasarlanmış düşük seviyeli özniteliklerin (örneğin renk [13], doku [14], uzamsal yapı [15] veya diğer öznitelik kombinasyonları) kullanımına dayanmaktadır. Bu öznitelikler kişinin görünümünü tanımlamak için kullanılırken, benzerlik skorlarını hesaplamak için uygun bir mesafe ölçüm fonksiyonu kullanılır. Bu nedenle, geleneksel kişi Re-ID yöntemleri öznitelik tabanlı ve ölçüt tabanlı olmak üzere iki kategoriye ayrılabilir.

Öznitelik tabanlı yöntemler: Öznitelik tabanlı yöntemler, el ile tasarlanmış özniteliklerin kullanımına odaklanarak görüntüleri tanımlama ve farklı insanları güvenilir bir şekilde ayırt etme yeteneği kazandırmayı hedefler. Bu öznitelikler, kişi Yeniden Tanımlama problemlerinde sıklıkla karşılaşılan poz, aydınlatma, bakış açısı değişikliği, kapanma gibi yaygın sorunlara karşı ayırt edici bir güç ve değişmezlik sağlamalıdır. Kişi yeniden tanımlama çalışmalarında, insanlar çeşitli düşük seviyeli öznitelikler kullanarak veya birleştirerek birçok farklı türde el yapımı öznitelik geliştirmişlerdir. Bu düşük seviyeli öznitelikler, renk [13], [16]–[19], doku [14], [18], [20] veya ilgi noktası dedektörleri [14], [18] gibi yöntemlerle elde edilebilir. El yapımı öznitelikler, bir kişinin tüm görüntüsünden veya vücudun farklı bölgelerinden sınırlayıcı kutuyu bölerek ve öznitelikleri tek tek alarak elde edilebilir. Renk histogramları [14], [16], [18], [21], yerel ikili desenler (local binary pattern) [21], [22], renk isimleri [17], [19], ölçeğe dayanlı öznitelik dönüşümleri (scale-invariant feature transforms) [14], [18], [23] ve ölçeğe dayanlı yerel üçlü desenler (scale-invariant local ternary patterns) [16], [24] genellikle kişi yeniden tanımlama için kullanılan el yapımı özniteliklerdir. RGB, HSV ve YCbCr renk uzaylarında 3 farklı renk histogramının ve Gabor [25] ile Schmid [26] doku filtrelerinin kullanıldığı 2 dokusal özelliğin bir kombinasyonu ile öznitelik tanımlayıcıları oluşturulmuştur [27]. Bu öznitelik gösterimi, birçok bilginin birleştirilmesini sağlar. Ayrıca, yerel maksimum oluşum öznitelikleri (local maximal occurrence features) olarak adlandırılan histogram öznitelikleri, bir dereceye kadar bakış açısı değişikliklerine karşı değişmezlik sağlayabilir ve birçok karşılaştırma veri setinde oldukça etkileyici

performans sergileyebilir [16], [24]. Son dönemde, el yapımı öznitelikleri kullanmak yerine orta düzey öznitelikler olarak bilinen belirgin insan vücut bölgelerinin kullanıldığı bazı el yapımı öznitelik öğrenme çalışmalarında yeni bir eğilim ortaya çıkmıştır. Belirgin bölgeler veya ayırt edici bölgeler, bir kişinin çevresindeki insanlardan ayrılmasına yardımcı olan bölgelerdir. Bu ayırt edici bölgeler, kişi Re-ID modellerinin performansını artırabilecek değerli bilgiler sağlayabilir. Zhao ve diğerleri [18] tarafından kişi Re-ID alanında yeni bir yaklaşım sunulmuştur. Bu araştırmanın ana odak noktası, genellikle görüntüleri girdi olarak alan ancak alt düzey özniteliklerini kullanmayan DL tabanlı modeldir. Derin öznitelikler düşük düzeyli özniteliklerden daha ayırt edici ve sağlam olduğu kanıtlanmıştır. Dolayısıyla, bahsi geçen öznitelikler bu çalışmada kapsamlı bir şekilde incelenmeyeceklerdir.

Metrik tabanlı yöntemler: Metrik tabanlı yöntemler, benzerliği veya ilişkiyi ölçebilen etkili bir mesafe metriği öğrenmeyi amaçlayarak geliştirilmiştir. İyi bir metriğin bulunması, el yapımı özniteliklerin sınıflandırma başarısındaki etkinliği üzerinde önemli bir etkiye sahiptir. Ayrıca, uygun bir mesafe metriği öğrenmek, ayırt edici bir tanımlayıcı bulmaktan daha zor bir süreçtir [9]. Kişi Re-ID alanında, bu tür bir mesafe ölçümü öğrenerek aynı kişinin görüntüleri arasındaki mesafeyi en aza indirgerken, farklı kişilerin görüntüleri arasındaki mesafeyi en üst düzeye çıkarmayı hedefler. Bu tür bir metrik kullanarak eşleştirme ve tanımlama daha kolay hale gelebilir ve önemli bir performans artışı elde edilebilir. Kişi Re-ID sorunlarını ele almak için birçok önemli mesafe öğrenme algoritması [28]–[32], önerilmiştir. Bu algoritmalar arasında en popüler olanı KISSME'dir [28]. KISSME, eşdeğerlik kısıtlamalarına dayalı olarak Mahalanobis ölçü öğrenimi yöntemini kullanmaktadır. Mahalanobis mesafesi genellikle iki görüntü arasındaki kareli mesafeyi sağlar. KISSME, büyük ölçekli verilerle başa çıkabilen ve karmaşık optimizasyon problemlerini gerektirmeyen bir mesafe öğrenme algoritmasıdır. [31] araştırmasında kişi yeniden tanımlaması için, yerel Fisher ayrım analizi önerilmiştir. Bu analiz sınıf içi değişkenliği en aza indirmeyi ve sınıf dışı değişkenliği en üst düzeye çıkarmayı hedeflemektedir. Başka bir çalışmada ise doğrusal olmayan durumlarla başa çıkmak için, filtre tabanlı boyut indirgeme tekniklerinin kullanılması önerilmiştir [32]. Ayrıca veri örneklerine uyum sağlayabilen karar sınırlarını öğrenmek için destek vektör makinesi (SVM) öğrenmesi de önerilmiştir [30]. Son zamanlarda yapılan araştırmalar, eşleştirme sürecinde kullanılan ayırmacı bir alt uzay öğrenmenin, belirli bir mesafe metriği öğrenmekten daha faydalı olduğunu ortaya koymuşlar. Bu yaklaşımlara göre, orijinal

öznitelik uzayındaki öznitelikler, daha düşük boyutlu bir alt uzaya yansıtılır. Genel de görüntü öznitelik tanımlayıcılarının boyutunu azaltmak için yaygın olarak temel bileşen analizi (PCA) kullanılmaktadır ve ardından PCA alt uzayında metrik öğrenme işlemi gerçekleştirilmektedir. Kişi yeniden tanımlama problemi, sıralama problemi olarak ele alındığında, RankSVM yöntemi kullanılarak veri örneklerine uyarlanabilen bir alt uzayın öğrenilmesi amaçlanır. RankSVM aslında SVM'in bir varyasyonudur. SVM'de, zayıf bir dizi RankSVM öğrenilir ve daha sonra toplu öğrenme yöntemiyle daha güçlü bir sıralama elde edilir. Bu sayede RankSVM, büyük ölçekli problemler için uygulanabilir hale gelir [33]. Çapraz görünüm ikinci dereceden ayırım analizi (XQDA) olarak adlandırılan bir başka önerilen yöntem, ayırmacı bir düşük boyutlu alt uzayı ve aynı zamanda türetilmiş alt uzayda bir QDA metriği öğrenmeye odaklanmaktadır. Bu yöntem, KISSME yaklaşımının bir uzantısı olarak geliştirilmiştir. Genellikle görüntüler farklı açılara sahip kameralardan alındığı için XQDA, kişi yeniden tanıma görevi için uygun bir yöntemdir. Bu nedenle, XQDA ile öğrenilen ayırmacı alt uzay ve QDA metriği, farklı görünüm arasındaki benzerlikleri hesaplamak için etkili bir araç sağlamaktadır [16]. Bir başka yaygın olarak bilinen metrik öğrenme algoritması, pozitif yarı kesin (PSD) lojistik metrik öğrenme tekniğidir [24]. Bu yöntem, örnekleri asimetric olarak ağırlıklandırmak için etkili bir strateji kullanmaktadır. Bu metrik öğrenme yöntemleri el yapımı özniteliklerle iyi sonuçlar elde edebilirler. Ancak, derin öğrenme tabanlı kişi yeniden tanıma yöntemleri daha doğru öznitelik temsilleri sağladıkları için bu metrik öğrenme yöntemlerine daha az bağımlıdır. Derin öğrenme yöntemleri, doğrudan derin özniteliklere Öklid mesafesi veya Kosinüs mesafesi gibi önemli mesafe ölçümlerini uygulayarak iyi sonuçlar elde etme eğilimindedir. Bu nedenle, bu araştırmada bu metrik öğrenme yöntemleri detaylı bir şekilde incelenmeyecektir.

2.1.2 Derin Öğrenmeye Dayalı Kişi Yeniden Tanımlama Yöntemleri

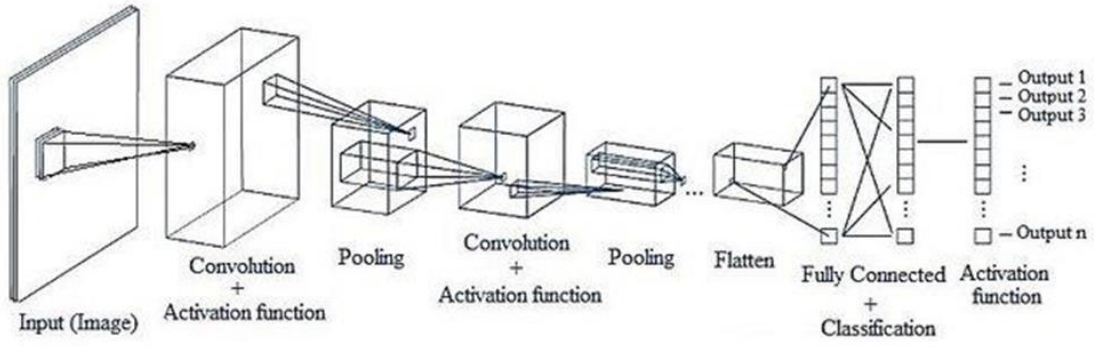
Geleneksel el yapımı tabanlı kişi yeniden tanımlama yöntemlerine ek olarak, derin öğrenme tabanlı yöntemler öznitelikle CNN'lerin yaygın olarak kullanıldığı kişi yeniden tanımlama görevinde önemli bir yer edinmiştir. Bu derin öğrenme modelleri, ayırmacı ve genelleme yeteneklerinden dolayı, geleneksel el yapımı öznitelik yöntemlerine kıyasla çok daha yüksek performans ve başarı sağlayabilirler. Derin öğrenme yöntemi, insan görüntülerindeki derin öznitelikleri tutarlı bir şekilde öğrenerek etkili bir benzerlik metriği oluşturabilir. Bu nedenle, son dönemdeki çalışmalar çoğunlukla el yapımı öznitelikler yerine derin öğrenme özniteliklerine odaklanmıştır [34]–[46]. Kişi yeniden

tanıma alanındaki derin öğrenme yöntemleri, hala global öznitelik çıkarma konusunda bazı zorluklarla karşılaşmaktadır. Ancak, Son yıllarda Transformer tabanlı yöntemlerin olgunlaşmasıyla birlikte, bu alanda bazı kişi yeniden tanıma yöntemleri de ortaya çıkmış ve tatmin edici sonuçlar elde etmiştir. Bu çalışma, temel olarak derin öğrenme tabanlı kişi yeniden tanıma yaklaşımlarına odaklanmaktadır. Bu nedenle, aşağıdaki bölümlerde CNN'ler ve Transformer'ler gibi derin öğrenme tekniklerinin kişi yeniden tanıma problemini ele almadaki genel bakışı sunulmaktadır. Ayrıca, kişi yeniden tanıma görevleri için derin öğrenme tekniklerini ilgilendiren bazı yeni çalışmaların sınıflandırılması yapılarak, avantajları ve dezavantajları eleştirel bir şekilde analiz edilecektir.

2.2 Derin Evrişimli Sinir Ağları

Makine öğreniminin dallarından biri olan DL, çok katmanlı doğrusal olmayan işlemleri kullanırken verilerin daha üst düzey temsillerini oluşturmayı amaçlamaktadır [47], [48]. Bu üst düzey temsiller, tanıma ve sınıflandırma için temel özniteliklerden çok daha faydalıdır. Günümüzde CNN'ler en popüler DL mimarisi haline gelmiştir. Doğal dil işleme, bilgisayarlı görü ve robotik gibi alanlardaki birçok görevde son teknoloji yöntemlerden daha iyi performans gösterebildiklerini ispatlamışlardır. Bu nedenle, son yıllarda Re-ID [35], [36], [38], [41], [42] için en uygun derin öznitelikleri öğrenmek için yaygın olarak kullanılmaya başlanmıştır. Bu bölümde CNN'lerin ana kavramları tanıtılmaktadır.

CNN [49], birden fazla katmanları olan ileri beslemeli bir sinir ağıdır. Sıklıkla bir havuzlama adımını içeren bir veya daha fazla Conv katmanı içermektedir. Ardından bir veya daha fazla tam bağlı (FC) katmanları takip etmektedir. CNN'in mimarisi, girdi olarak verilen görüntünün 2 boyutlu yapısından yararlanmak için tasarlanmıştır. Doğrudan eğitim veri kümesinden öğrenilebilen ve bir hiyerarşi içinde bağlantılı olan öznitelikler oluşturmaktadır. Ayrıca CNN modelleri daha kolay eğitilebilir ve aynı sayıda gizli katman ve nörona sahip FC ağlarına göre önemli ölçüde daha az parametreye sahiptir. Temel CNN modülleri, girdi katmanı, Conv katmanları, havuzlama katmanları, FC katmanı ve çıkış veya sınıflandırma katmanından oluşmaktadır. Genel derin CNN yapısı Şekil 2.1'de gösterilmiştir.



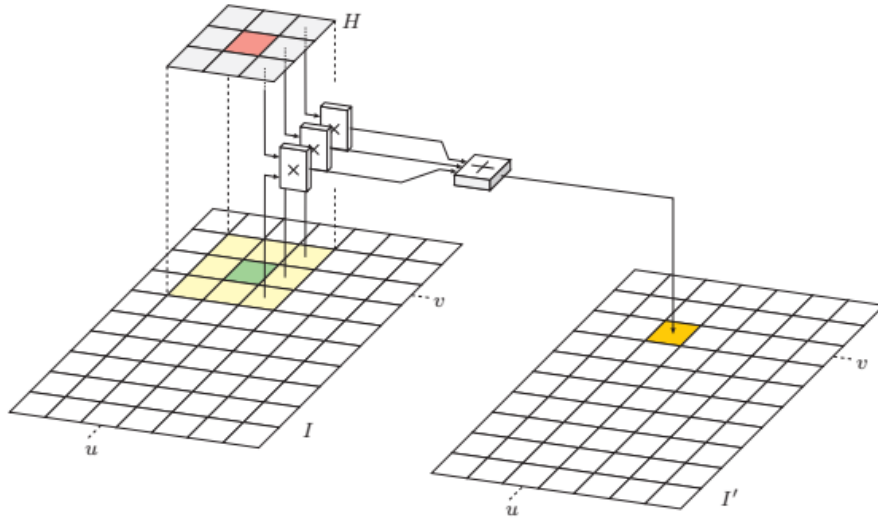
Şekil 2.1 Genel CNN yapısı [50].

2.2.1 Evrişimli Sinir Ağı Katmanları

Şekil 2.1'de gösterildiği gibi, CNN bir dizi katmandan oluşmaktadır. Bu bölümde en önemli katmanların özeti verilerek CNN'lerde oynadıkları rol, avantajları ve dezavantajları tartışılmaktadır.

- **Giriş katmanı:** CNN'de, ilk katman giriş görüntüsünü almaktadır ve belirli bir boyuta sahiptir. Bu katmanın boyutu, CNN'nin performansı için önemli bir faktördür. Giriş görüntüsünün boyutu artırıldığında, sistem ve eğitim süreci farklı koşullarda daha iyi bir başarı elde edebilir [51]. Ancak, giriş görüntüsünün boyutu küçültüldüğünde, eğitim sürecinde bir azalma olsa da aynı zamanda sistem başarısını da olumsuz etkileyebilir. Öte yandan, giriş görüntüsünün boyutu büyütüldüğünde, sistem başarısını artırabilir ancak eğitim sürecinde artışa yol açabilir. Bu nedenle, tasarlanan sisteme uygun bir boyutta görüntü seçmek önemlidir. Giriş katmanı, genişlik, yükseklik ve derinlik olmak üzere [genişlik $\times$ yükseklik $\times$ derinlik] boyutlu bir görüntüyü içeren bir piksel matrisidir.
- **Evrişim katmanları (Conv):** CNN'deki temel katman, görüntü özniteliklerinin çıkarılmasında kullanılan ana katmandır. Bu katman, genellikle Conv katmanı olarak adlandırılır. Conv katmanı, görüntülerin filtrelenmesi için bir dizi Conv çekirdeği kullanılarak öznitelik çıkarımı işlemini gerçekleştirir. Conv çekirdekleri, girdi görüntüsünün boyutlarını tarayarak bu Conv işlemlerini gerçekleştirir. Her bir Conv çekirdeği, farklı öznitelikleri çıkarmak için girdi verilerinden filtreleme yapmaktadır. Conv katmanında daha fazla Conv filtre kullanılması, girdi verilerinden daha fazla öznitelik çıkarılmasını sağlamaktadır. Conv katmanının temel tasarım hiperparametreleri, filtre boyutu, filtre sayısı ve çekirdek adımları gibi faktörleri

içerir. Bu filtreler farklı boyutlarda olabilir, örneğin 2×2 , 3×3 , 5×5 veya 7×7 gibi. "Sayı" ise, bir Conv filtresi kullanarak üst katmandan elde edilebilecek öznitelik haritalarının sayısıdır, "adım" ise filtre matrisinin girdi matrisi üzerinde kaydırılabileceği piksellerin sayısıdır. Bir adım kullanıldığında, filtreler her seferinde yalnızca bir piksel kayar. İki adım kullanıldığında, filtreler her seferinde iki piksel hareket eder. Bu adım değeri, Conv katmanının tasarımında önemli bir hiperparametredir ve özellikle örnekleme ve boyut azaltma gibi işlemlerde kullanılmaktadır. Bu nedenle bazen girdi matrisinin kenarları boyunca sıfırlarla dolgu yapmak en iyisidir. Böylece filtre, girdi matrisinin kenarındaki öğelere uygulanabilir. Sıfırlarla dolgu yapmanın faydalı bir özelliği, öznitelik haritasının boyutunu kontrolünü sağlamasıdır. Bu katmanın sonucunda elde edilen çıktı, bir aktivasyon haritası veya öznitelik haritası olarak adlandırılır. Bu, sigmoid, Tanh veya ReLU gibi aktivasyon fonksiyonları uygulanarak elde edilir. Bir Conv katmanının gösterimi Şekil 2.2'de sunulmuştur.



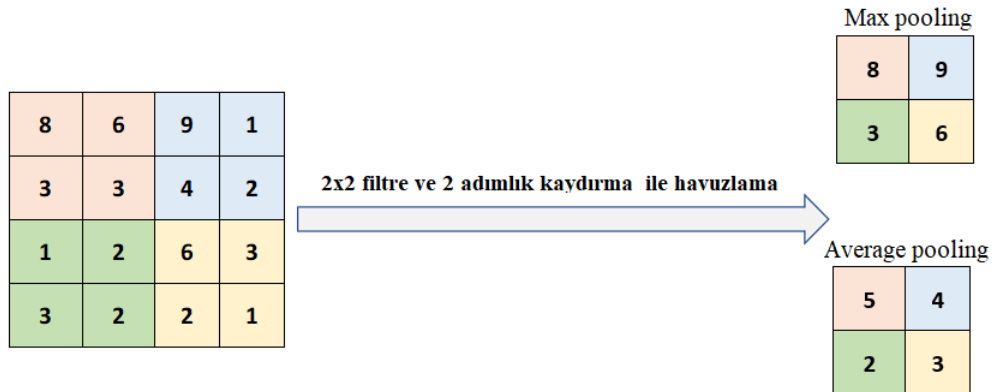
Şekil 2.2 Conv katmanının işlemlerini gösteren bir çizim[52].

$(n_h \times n_w)$ boyutunda bir görüntü üzerinde öznitelik çıkarımı için Conv filtresi boyutunu $(w \times w)$ olarak ve adım boyutunu k olarak ayarlayarak yapılmaktadır. Böylece, $\left(\frac{n_h - w + k}{k} \times \frac{n_w - w + k}{k}\right)$ boyutunda bir öznitelik haritası oluşturulabilir.

Conv çekirdeğinin boyutu, özniteliklerin çıkarılmasının kalitesini etkileyen bir faktördür. Genellikle daha küçük boyutlu Conv çekirdekleri daha yüksek kaliteli öznitelikler sağlar. Ancak, çekirdek boyutunun belirlenmesi, girdi görüntü boyutuna göre dikkatli bir şekilde yapılmalıdır.

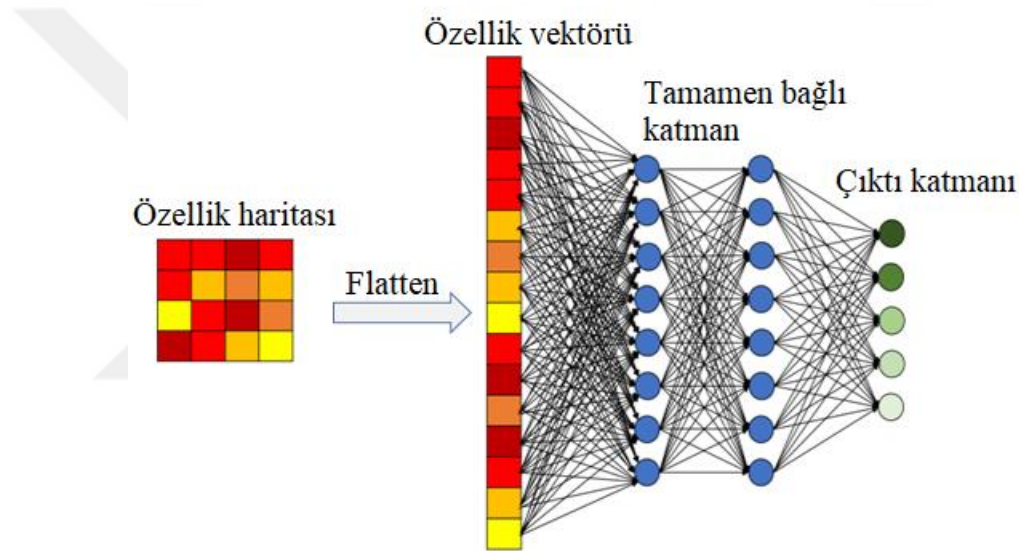
Bu nedenle, derin bir CNN'in eğitim amacı, bir görüntü içinde belirli görsel özniteliklerin varlığını algılayabilen bir dizi çekirdek öğrenmektir. Ağın girişindeki çekirdekler, düşük seviyeli özniteliklerin, örneğin çizgi ve kenar gibi temel yapıları algılamayı öğrenirken, ağın çıkışındaki çekirdekler ise daha yüksek seviyeli ve alana özgü görsel özniteliklerin varlığını algılamayı öğrenirler. Bu süreç, ağın katmanlarının derinleştikçe daha soyut ve karmaşık özniteliklere odaklanmasını sağlamaktadır.

- **Havuz katmanları (havuz):** Genellikle havuzlama katmanı, bir CNN'deki dönüştürme katmanından sonra gelmektedir. Temel amacı, girdi matrisinin boyutunu azaltarak önemli bilgileri korurken önemsiz bilgilerden kurtulmaktır. Havuzlama katmanı, Conv katmanı gibi belirli bir adım aralığında hareket edebilen tanımlanmış filtrelerle sahiptir. Her pikselin değerini bulmanın iki yaygın yöntemi vardır: ortalama havuzlama (average pooling) ve maksimum havuzlama (maxpooling) [50]. Maksimum havuzlama işleminde maksimum piksel değerleri seçilirken, ortalama havuzlama işleminde ortalama piksel değerleri seçilmektedir. Havuzlama katmanındaki iki önemli parametre, filtre boyutu ve adımdır. Genellikle havuzlama katmanları 2×2 gibi küçük filtrelerden oluşmaktadır. Bir ağda bu katmanların kullanılması bazı dezavantajlara sahiptir. Ortalama havuzlama kullanıldığında, düşük veya negatif piksel değerleri nedeniyle aktivasyon değerini minimumda tutabilir. Bu durumda, önemsiz veya düşük bilgi taşıyan piksellerin etkisi daha belirgin olabilir. Bununla birlikte, maksimum havuzlama kullanıldığında, düşük değerlere sahip pikseller göz ardı edildiği için önemli ve ayırt edici bilgilerin bu katmanda kaybolmasına yol açabilir [53]. Maksimum ve ortalama havuzlama örnekleri Şekil 2.3'te verilmiştir.



Şekil 2.3 2×2 filtre ve adım=2 ile maksimum ve ortalama havuzlama örnekleri.

- **Tamamen bağı katmanlar (FC):** CNN mimarisi içerisinde giriş boyutunun yeterince küçültülmesiyle birlikte, CNN'ler genellikle Conv katmanları ve havuzlama katmanlarından sonra bir veya daha fazla FC katmanına geçiş yapmaktadır. FC katmanları, her birimlerinin Conv veya havuzlama katmanlarındaki tüm düğümlere bağlı olduğu için "tam bağlantılı" olarak adlandırılır. CNN yapısına bağlı olarak, bu katmanların sayısı değişebilir. Toplam ağırlık sayısı, Conv katmanındaki düğüm sayısı ile bu FC katmanlarındaki düğüm sayısının çarpımına eşittir [54]. FC katmanlarının görevi, çıkarılan üst düzey öznitelikleri kullanarak giriş görüntüsünü farklı sınıflara ayırmaktır. Bunun dışında FC katmanı eklemek özniteliklerin lineer olmayan kombinasyonlarını öğrenmek için kullanımı kolay bir yöntemdir [55]. Şekil 2.4, iki FC katmanına sahip basit bir ağı nasıl kurulduğunu göstermektedir.



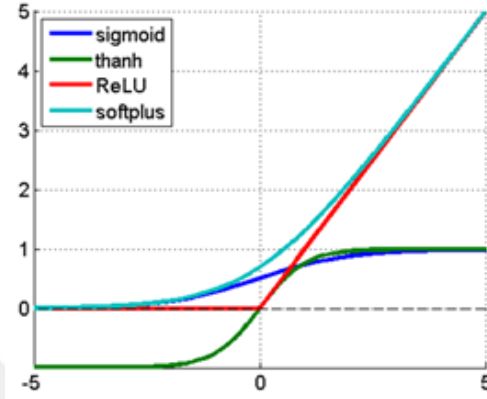
Şekil 2.4 İki FC katmanıyla basit ağı

- **Çıktı (sınıflandırma) katmanı:** Bir CNN'deki öğrenme süreci, tamamen bağı katmanından sonra gelen sınıflandırma katmanında gerçekleşir. Bu katman, çıktının sınıflar üzerinde bir olasılık dağılımı olarak yorumlanmasını sağlamaktadır [51]. Sınıflandırma için farklı kayıp fonksiyonları kullanılabilir. En yaygın olanı SoftMax fonksiyonudur, ancak Sigmoid veya Öklid gibi diğer fonksiyonlar da kullanılabilir. SoftMax, log olasılık dağılımı kullanan doğrusal bir sınıflandırıcıdır.

2.2.2 Aktivasyon fonksiyonları

Aktivasyon fonksiyonları, yapay sinir ağlarında kullanılan matematiksel işlevlerdir [56]. Bu fonksiyonlar, sinir ağlarının her bir katmanındaki çıktıların hesaplanmasında kullanılmaktadır. İşlevleri, girdi değerlerini dönüştürerek ve aktivasyon düzeyini

belirleyerek sinir hücrelerinin çıktılarını belirler. Çıkışı, 0 ile 1 arasındaki gibi istenilen bir aralığa eşlemek için tasarlanmıştır. Bu, ağı daha karmaşık eşlemeleri öğrenme yeteneğini artıran bir CNN'ye doğrusal olmayan öznelilikler eklemektedir. Genellikle NN'lerde en sık kullanılan aktivasyon fonksiyonları Sigmoid, Tanh, SoftPlus ve ReLU fonksiyonlarıdır. Şekil 2.5, NN'lerde en çok kullanılan aktivasyon fonksiyonlarını göstermektedir.



Şekil 2.5 Çoğunlukla NN'lerde kullanılan aktivasyon fonksiyonları.

İlk 2 aktivasyon fonksiyonu genellikle geleneksel geri yayımlı yapay sinir ağlarında kullanılmaktadır. Çünkü ReLU fonksiyonu, doymamış, doğrusal ve tek taraflı baskıya sahip olup seyrek aktivasyona sahiptir ve bu nedenle CNN'lerde daha fazla kullanılır [57]. Tüm Conv katmanları ve tam bağlantılı katmanların ilk 2 katmanı ReLU fonksiyonuyla aktive edilmektedir [58] bu da derin yapay sinir ağlarının son dönemdeki başarısının büyük bir kısmından sorumludur. Sinir düğümü aktivasyon fonksiyonunun $h^{(i)}$ olduğunu varsayarsak, ReLU fonksiyonu şu şekilde ifade edilebilir:

$$h^{(i)} = \max\left(\left(w^{(i)T}\right) x, 0\right) = \begin{cases} \left(w^{(i)T}\right) x, & \left(w^{(i)T}\right) x > 0 \\ 0, & \text{Otherwise} \end{cases} \quad (2.1)$$

Burada i , gizli katman düğümlerinin sayısı ve $w^{(i)}$ ağırlıklarıdır. ReLU'nun avantajları, hesaplama açısından daha verimli olması ve aynı zamanda kaybolan ve patlayan gradyan sorunlarının olumsuz etkilerini azaltmasıdır [59]. Ek olarak, ReLU aktivasyon fonksiyonlarına sahip derin CNN'ler, diğer aktivasyon fonksiyonlarını kullanan eşdeğer yapılandırmalara göre çok daha hızlı bir şekilde eğitilebilir [57]. Ancak, ReLU aktivasyon fonksiyonu "dying ReLU" olarak bilinen bir sorunla karşılaşır, burada ağırlıklar öyle güncellenir ki nöron hiçbir zaman aktive olmaz. Dolayısıyla, nöronun içinden geçen

gradyan sıfır kalır. Bu sorunu çözmek için "Leaky ReLU" gibi yöntemler denemiştir [60], burada ReLU'nun çıktısı $x < 0$ ise sıfır yerine çok küçük bir değere ayarlanır.

2.2.3 SoftMax Sınıflandırıcı ve Kayıp Fonksiyonu

Bir CNN görüntü sınıflandırma gibi bir görevi gerçekleştirmek için kullanıldığında, çıktı katmanına eklenmesi gereken bir işlev vardır. Bu işlev, mevcut değerleri alacak ve onları olasılık dağılımına dönüştürecektir. SoftMax [61], K gerçek değeri olan bir vektörü, toplamı = 1 olan K gerçek değere sahip bir vektöre dönüştürebilen bir işlevdir. Giriş değerlerinin negatif, pozitif, 0 veya 1'den büyük olması önemli değildir. Çünkü SoftMax, onları 0 ile 1 arasında değerlere dönüştürebilir. Böylece olasılıklar olarak yorumlanabilirler. SoftMax aktivasyonu, çoklu sınıf sınıflandırma yaparken kullanılabilir. Çünkü tek bir katmandaki tüm aktivasyonların toplamının = 1 olmasını sağlayabilir. SoftMax aktivasyon fonksiyonu aşağıdaki (2.2 denklemi) gibidir. Denklemdaki K aktivasyon sayısıdır ve z aktivasyon vektörüdür.

$$\delta(z)_j = \frac{e^{z_j}}{\sum_{K=1}^K e^{z_K}} \quad \text{For } j = 1, \dots, K \quad (2.2)$$

SoftMax regresyonunun amacı, çoklu sınıflandırma problemlerini çözmektir. Eğitim veri kümesi $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$ için etiket y'dir. $y^{(i)} \in \{1, 2, \dots, k\}$ k farklı değere sahip olabilir. Verilen bir x test girdisi için; hedef, hipotez fonksiyonu ile her bir j kategorisi için $p(y = j|x)$ olasılık değerini tahmin etmektir. Yani, her bir x kategorisinin olasılığını tahmin etmektir. Sonuç olarak, hipotez fonksiyonu, çıktı olarak toplamı 1 olan k boyutlu bir vektör verecektir, bu da tahminin olasılık değerini temsil etmektedir. Özetle, burada belirlenen hipotez fonksiyonu $h_\theta(x)$ aşağıda verildiği gibidir:

$$h_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1|x^{(i)}; \theta) \\ p(y^{(i)} = 2|x^{(i)}; \theta) \\ \dots \\ p(y^{(i)} = k|x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \dots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \quad (2.3)$$

Kayıp fonksiyonu, CNN'lerde önemli bir fonksiyondur. Bu fonksiyon, modelin tahmin ettiği değer ile gerçek değer arasındaki hatayı ölçmektedir. Daha basit bir ifadeyle, gradyanların hesaplanması kayıp fonksiyonu kullanılarak yapılmaktadır ve CNN'in ağırlıkları gradyanlar kullanılarak güncellenir. Kayıp fonksiyonunun kullanımı çözülmesi gereken göreve dayanarak eğitimin hangi hızda yakınsadığını doğrudan etkilemektedir.

Genellikle, CNN'ler çapraz-entropi kaybı fonksiyonunu kullanarak ağın doğruluğunu değerlendirir. Ağı rastgele değerlerle başlatıldığında, kayıp fonksiyonu yüksek olur. Ancak, ağı eğitirken amaç kayıp fonksiyonunu mümkün olduğunca düşürmektir. Düşük bir kayıp fonksiyonuna sahip olan bir CNN, eğitim setini daha yüksek bir doğrulukla sınıflandırır. Çoklu sınıf sınıflandırma problemleri için kullanılan çapraz entropi kaybının formülü aşağıdaki gibidir:

$$L_{CrossEnt.} = - \sum_{n=1}^N \log(p(n))q(n) \quad (2.4)$$

Burada $p(n) \in [0,1]$, CNN modelinin girdi görüntüsünün n 'inci sınıfa ait olduğunu tahmin etme olasılığını temsil ederken, $q(n)$, görüntü etiketlerinin dağılımını temsil etmektedir.

2.2.4 Derin Evrişimli Sinir Ağlarının Eğitimi

CNN'ler, diğer birçok makine öğrenimi problemi gibi, kullanılan eğitim verilerine bağlı olarak eğitilmektedir. Eğitim için optimize edici algoritmalar, ağırlıklar, biaslar ve diğer öğrenilebilir parametreler üzerindeki kaybı minimize etmek amacıyla kullanılır. Eğitim süreci yinelemeli olarak gerçekleştirilir. Dolayısıyla, her bir iterasyonda ağ parametreleri performansı iyileştirmek için güncellenir. Bir sinir ağı eğitirken, kullanıcı ona kullanılacak giriş örneğini sağlar. Öğrenme başladığında, parametreler küçük rastgele sayılarla başlatılır. Dolayısıyla ağ girişe nasıl tepki vermesi gerektiğini bilmemektedir. İleri geçiş, muhtemelen yanlış bir sonuç elde edecektir. Sinir ağı düzeltmek için, kayıp veya maliyet olarak bilinen hata ölçüsünü tanımlamak gereklidir. Kaybın hesaplanması, istenen çıktı üzerindeki kayıp fonksiyonunun değerlendirilmesi ve eğitim örneğinin tahmininin yapılmasıyla gerçekleştirilir. Ardından, öğrenme algoritması, kaybı azaltmak için tüm ağ parametrelerini güncellemeyi "öğrenmek" zorundadır. Sonuç olarak, kaybın parametrelere göre negatif gradyanının hesaplanması, zincir kuralını uygulayarak katman katman gerçekleştirilir ve girdiye doğru ilerler. Bu süreç geriye yayılma olarak bilinir ve her bir örnek için tekrarlanır. Daha sonra, elde edilen tüm negatif gradyanların ortalamasının bir kesri olan öğrenme oranı ağırlıklara eklenir ve onları günceller. Bu, tüm öğrenilebilir ağ parametrelerinin optimize edilmesini ve yerel minimumlara yakınsamayı sağlamaktadır [61]. NN'lerin eğitiminde birçok farklı algoritma kullanılmaktadır. En yaygın kullanılan algoritmalar arasında Stokastik gradyan inişi (SGD) [61] ve adaptif moment tahmini (Adam) [62] yer almaktadır. SGD, makine öğrenmesinde ve özellikle

derin öğrenmede en çok kullanılan optimizasyon algoritmasıdır. SGD aşağıdaki gibi ifade edilir:

$$\theta_{i+1} = \theta_i - \alpha \nabla_{\theta_i} L\theta_i \quad (2.5)$$

Burada θ_{i+1} yeni öğrenilen parametreleri, θ_i mevcut parametreleri, α öğrenme hızını (learning rate), $\nabla_{\theta_i} x_i$ 'ye göre gradyanı temsil etmektedir ve $L\theta_i$ kayıp fonksiyonudur.

Adam, adaptif öğrenme oranı optimizasyon algoritması olarak da yaygın olarak kullanılan bir yöntemdir. 'Adam' ismi, 'adaptif moment estimation' ifadesinden türetilmiştir. Bu isim, NN'deki her bir ağırlık için öğrenme hızını dönüştürmek için gradyanın 1'inci ve 2'nci moment tahminlerini kullanmasından dolayı verilmiştir. 1'inci moment gradyanın ortalamasıdır, 2'nci moment ise gradyanın merkezi olmayan varyansdır. Moment tahminini gerçekleştirmek için Adam, gradient üzerinde hesaplanan üstel olarak hareket eden ortalamalardan yararlanır ve aşağıdaki gibi hesaplanır:

$$m_{t+1} = \beta_1 m_t + (1 - \beta_1) g_t \quad (2.6)$$

$$v_{t+1} = \beta_2 v_t + (1 - \beta_2) g_t^2 \quad (2.7)$$

Burada, m ve v gradyanın ve karesi alınmış gradyanın hareketli ortalamalarıdır, g mevcut mini-batch üzerindeki gradyandır ve β 'ler gradyanın 1'inci ve 2'nci momentlerini tahmin etmek için kullanılan başlangıç azalma oranlarıdır. Varsayılan değerleri sırasıyla 0.9 ve 0.999'dur. Hareketli ortalama vektörleri ilk iterasyonda sıfırlarla başlatılır, bu da moment tahminlerinin sıfıra doğru eğilimli olmasına neden olmaktadır. Bu başlangıçtaki yanlılık kolaylıkla giderilebilir ve bu da 1'inci ve 2'nci momentler için yanlılık düzeltilmiş tahminlere yol açar. Bu tahminler, $(\widehat{m}_t$ ve $\widehat{v}_t)$ olarak gösterilir.

$$\widehat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (2.8)$$

$$\widehat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (2.9)$$

Ardından, bu yanlılık düzeltilmiş (bias-corrected) tahminler, Adam güncelleme kuralını kullanarak parametreleri güncellemek için kullanılır:

$$\theta_{t+1} = \theta_t - \alpha \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t} + \varepsilon} \quad (2.10)$$

Burada θ_{t+1} yeni öğrenilen parametreleri, θ_t mevcut parametreleri, α öğrenme hızını ve ε adım boyutunu temsil etmektedir. ε varsayılan değeri 10^{-8} 'dir.

CNN'lerde, doğru tahmin sonuçlarını elde etmek için eğitim sırasında optimize edilen ağırlık ve bias parametrelerinin yanı sıra, önceden belirlenen hiperparametreler adında bir dizi parametre bulunmaktadır. Hiperparametreler, ağıın mimarisini ve eğitim sürecini yönlendiren değerlerdir. Ağıın boyutu, gizli katmanların sayısı, her katmandaki birim sayısı gibi mimari hiperparametreler ağıın yapısal özniteliklerini belirlerken, öğrenme hızı, grup boyutu gibi eğitim hiperparametreleri ise ağıın nasıl öğrenileceğini kontrol etmektedir. Hiperparametreler arasında bazı bağımlılıklar vardır. Örneğin, belirli bir gizli katmandaki birim sayısı, gizli katmanların toplam sayısına bağılı olabilir [63]. Hiperparametrelerin doğru değerlerini bulmak genellikle deneme yanılma yöntemiyle yapılmaktadır. Veri kümesi üzerinde farklı hiperparametre değerleriyle tekrar tekrar eğitim yapılır ve her denemenin ardından ağıın performansı değerlendirilir. Bu şekilde en iyi hiperparametre değerleri belirlenmeye çalışılır [64].

2.2.5 Derin Evrişimli Sinir Ağlarında Aşırı Uyumla Başa Çıkma

Aşırı uyum (overfitting), CNN modelinin eğitim verilerine iyi uyum sağladığı, ancak yeni veya görülmemiş veriler üzerinde iyi bir genelleme yapamadığı durumlarda ortaya çıkmaktadır. Bu, modelin eğitim verilerine özgü desenleri öğrenmesi anlamına gelir, ancak diğer verilerde işe yaramazlar. Derin öğrenme modellerinde aşırı uyumun önlenmesi için en iyi yol daha fazla eğitim verisine sahip olmaktır. Ancak, maalesef gerçek hayatta bütçe, zaman veya teknik kısıtlamalar nedeniyle bu genellikle mümkün olmamaktadır.

Küçük veri setleri üzerinde eğitilen derin CNN modellerinde aşırı uydurmanın önlenmesi için çeşitli yaklaşımlar önerilmiştir. Rastgele kırpma (random cropping) [57], rastgele çevirme (random flipping) [65] ve rastgele silme (random erasing) [66] işlemleri, derin CNN modellerini eğitirken veri artırma yöntemleri olarak yaygın olarak kullanılmaktadır. Ayrıca, düzenleme (regularization) yöntemleri de aşırı uyumu önlemek için iyi bilinen bir yaklaşımdır. Kayıp fonksiyonuna fazladan bir terim ekleyerek karmaşık modelleri cezalandıran L1 veya L2 düzenlemesi [67], bir modeli daha kararlı hale getirmenin en yaygın yollarından biridir. Son yıllarda, CNN'leri eğitmek için yaygın olarak kullanılan

iki yöntem, batch normalization (BatchNorm) [68] ve dropout [69], olarak bilinir ve ikisi de aşırı uydurmayı önlemede yardımcı olduğu gösterilmiştir. Dropout, eğitim sürecinin bir parçası olarak, her bir gizli nöronun çıkışını olasılığa dayalı olarak rastgele kaldırmayı denemektedir. BatchNorm ise, her bir gizli nöronun çıkışını normalize etmek için minibatch'ın ortalamasını ve varyansını kullanarak içsel kovaryans kaymasını (ICS) azaltmayı hedeflemektedir. Kişi Re-ID veri setleri genellikle çok büyük olmadığından, yüksek doğrulukta bir kişi Re-ID modeli oluşturmak için aşırı uydurmayı engellemek için yöntemler bulmak gerekmektedir. Bu çalışmada, önerilen modelde aşırı uyumu önlemek için rastgele çevirme, rastgele silme ve rastgele kırma kullanılmaktadır.

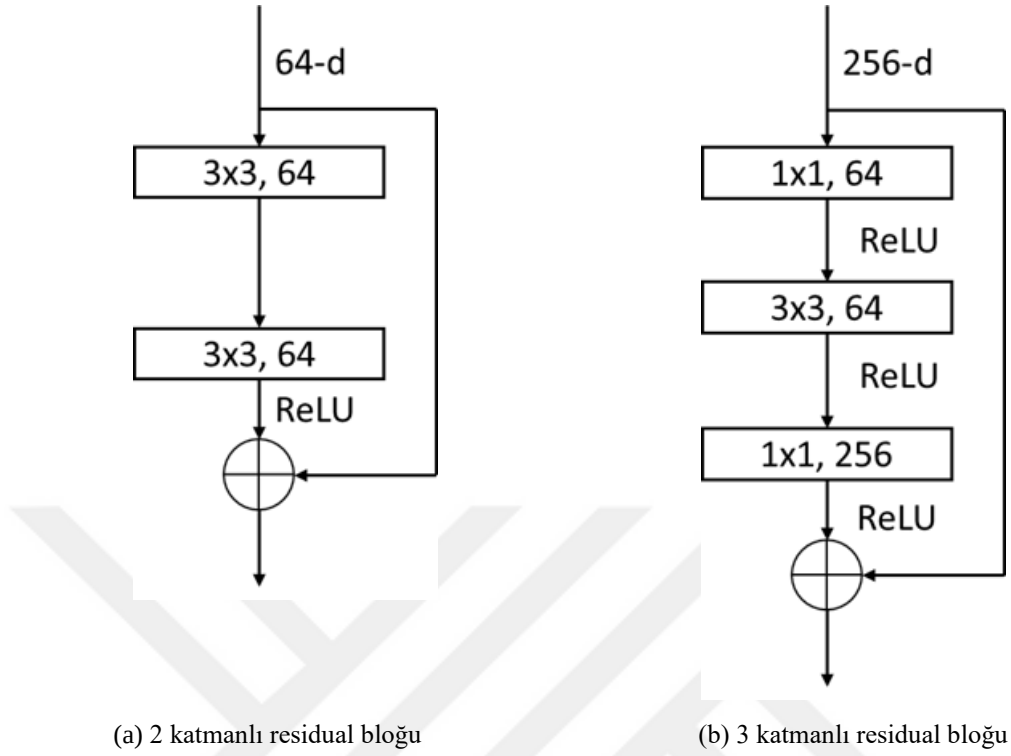
Ayrıca aşırı uyum ağın eğitim verileri üzerinde çok uzun süre eğitim vermesinden kaynaklanabilir [70]. Bunu önlemek için ağın performans artışı durduktan sonra ağın eğitimi durdurmasını sağlamak amacıyla erken durma (early stop) kullanılır. Bu nedenle, en iyi tahmin performansı veren ağırlıklar korunur.

2.2.6 Derin Evrişimli Sinir Ağı Mimarileri

Literatürde birçok derin evrişimli sinir ağı mimarileri modern bilgisayarlı görü uygulamalarını güçlendirmekte ve kişi Re-ID görevinde yaygın olarak kullanılmaktadır. Bu bölümde iyi bilinen CNN mimarileri olan ResNet [71] açıklanmaktadır. Her CNN mimarisi birbirinden çok farklı ağlardır ve her biri belirli problemler için özelleştirilmiştir.

- **Resnet:** ResNet, başlangıçta He ve diğerleri tarafından nesne tanıma için geliştirilen bir rezidüel CNN mimarisidir [71]. Bu derin sinir ağı, ImageNet [72] veri kümesinden 1,28 milyon görüntü üzerinde eğitilmiştir. ResNet, büyük ölçekli görsel tanıma yarışması 2015 (ILSVRC2015)'de iyi bir başarı elde etmiş ve ImageNet veri kümesindeki tespit, sınıflandırma ve yerelleştirme görevlerinde ve COCO veri kümesindeki tespit ve segmentasyon görevlerinde 1'inci sırayı almıştır [71]. ResNet'in temel motivasyonu, ağın derinliği arttıkça doğruluğun doygunlaşması (saturated) ve hızla düştüğü bir sorun olan "bozulma" (degradation) sorununu çözmektir. Ancak, modelin derinliği arttıkça ağın öğrenme yeteneği güçlenirken, derin modelin yüksek bir hata oranına ulaşmaması gerekmektedir. Bozunmanın nedeni, ağın optimize edilmesinin zor olmasıdır. Bu nedenle, bir artıklık (residual) yapısı önerilmiştir. ResNet bloğu, küçük ağlarda kullanılan ResNet 18, 34 gibi 2

katman derinlikte veya ResNet 50, 101, 152 gibi 3 katman derinlikte olabilir. Şekil 2.6, artıklık öğrenme yapı bloklarının yapısını göstermektedir.



Şekil 2.6 Tipik residual öğrenme yapı taşları. a) ResNet-18/34 yapı bloğu (sol) ve b) ResNet-50/101/152 yapı bloğu (sağ).

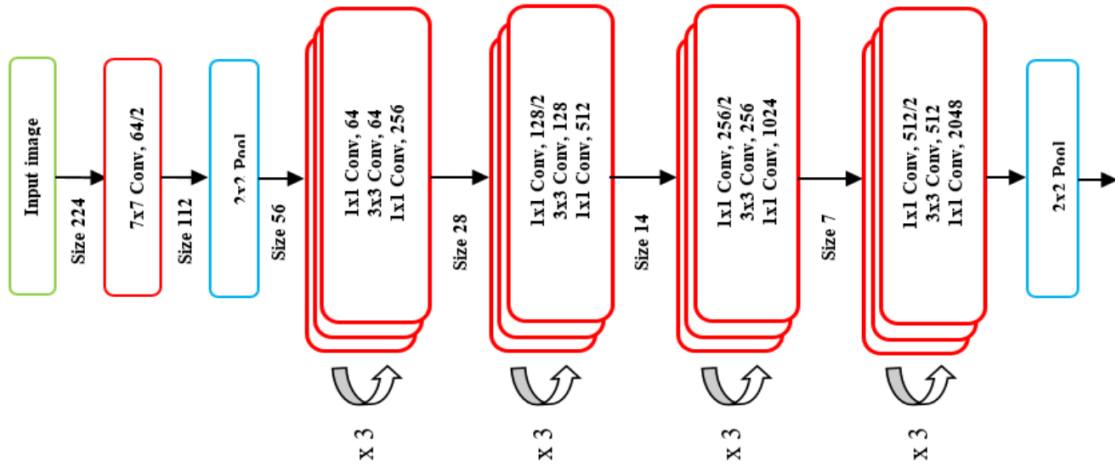
ResNet ağ mimarisi, 18, 34, 50, 101 ve 152 katmandan oluşan birçok farklı yapılandırmaya sahiptir [71]. İlk 2'si 2 katmanlı artıklık yapı bloğu kullanılarak oluşturulurken, son 3'ü 3 katmanlı artıklık yapı bloğu kullanılarak oluşturulur. Tablo 2.1 farklı ResNet ağ mimarilerinin detaylı açıklamasını göstermektedir. ResNet-50 ağ mimarisi Şekil 2.7'de verilmiştir.

ResNet-50'nin ayrıntılı bir açıklaması aşağıda verilmiştir. Bu ağ, $224 \times 224 \times 3$ RGB görüntüsü kullanan giriş katmanı ile başlamaktadır. ResNet-50 modeli, her biri bir Conv bloğu ve bir kimlik (identity) bloğu olan 5 aşamaya sahiptir. Şekil 2.8'de gösterildiği gibi, her Conv bloğu 3 Conv katmanından oluşmaktadır. Her Conv katmanı Conv, ReLU ve BatchNorm katmanlarına sahiptir. Kısa yol bağlantıları, kimlik eşleme işlemini gerçekleştirmek için kullanılır ve çıktıları yığılmış katmanların (stacked layers) çıktılarına eklenir. 3 Conv katmanı, 1×1 , 3×3 ve 1×1 boyutlarındadır. 1×1 katmanı boyutları azaltmalı ve ardından geri yüklemelidir. Ağın son katmanı, global ortalama havuzlama katmanıdır ve sonra SoftMax ile 1000

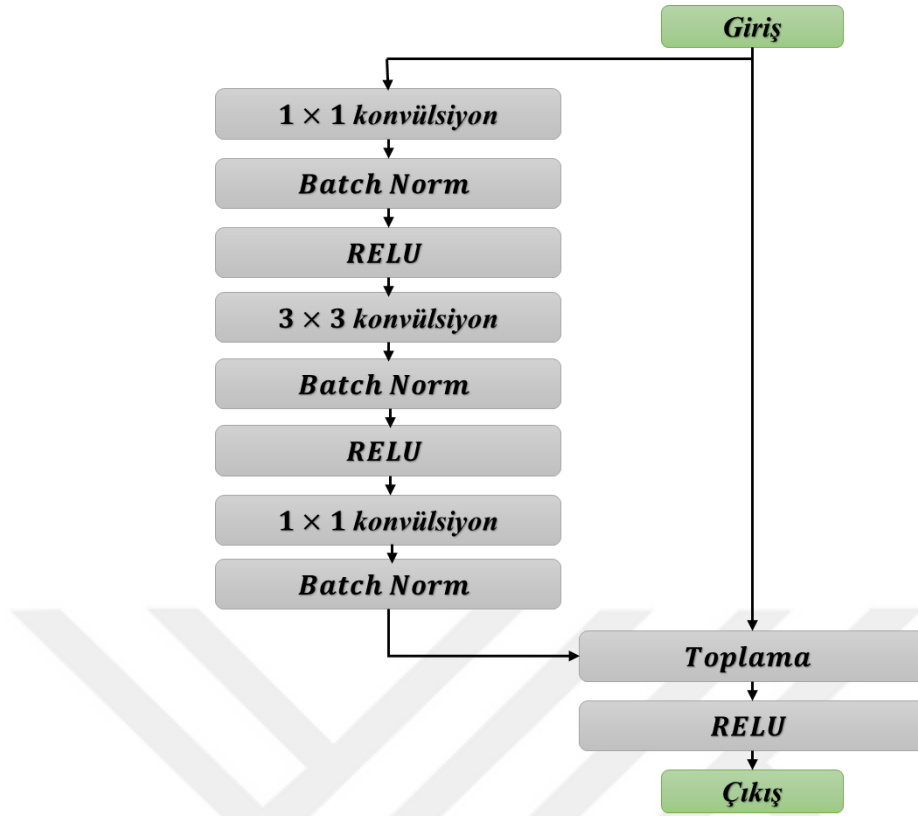
sınıflı tam bağlantılı bir katmandır. ResNet-50'nin eğitilebilir parametre sayısı 25 milyonun üzerindedir [71].

Tablo 2.1 ResNet Mimarisi [71].

Katman adı	Çıktı boyutu	18-katman	34-katman	50-katman	101-katman	152-katman
Conv1	112 × 112	7 × 7, 64, stride 2				
Conv2_x	56 × 56	3 × 3, max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Conv3_x	28 × 28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
Conv4_x	14 × 14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
Conv5_x	7 × 7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1 × 1	Average pool, 1000-d fc, softmax				
FLOPs		$1,8 \times 10^9$	$3,6 \times 10^9$	$3,8 \times 10^9$	$7,6 \times 10^9$	$11,3 \times 10^9$



Şekil 2.7 ResNet-50 Mimarisi. Her bir rezidüel bloktan önce, adım boyutu 2 olan bir örnekleme işlemi gerçekleştirilir. Çıkış katmanını dışındaki tüm katmanlar ReLU aktivasyonunu kullanır. Çıkış katmanını ise SoftMax kullanır [73].



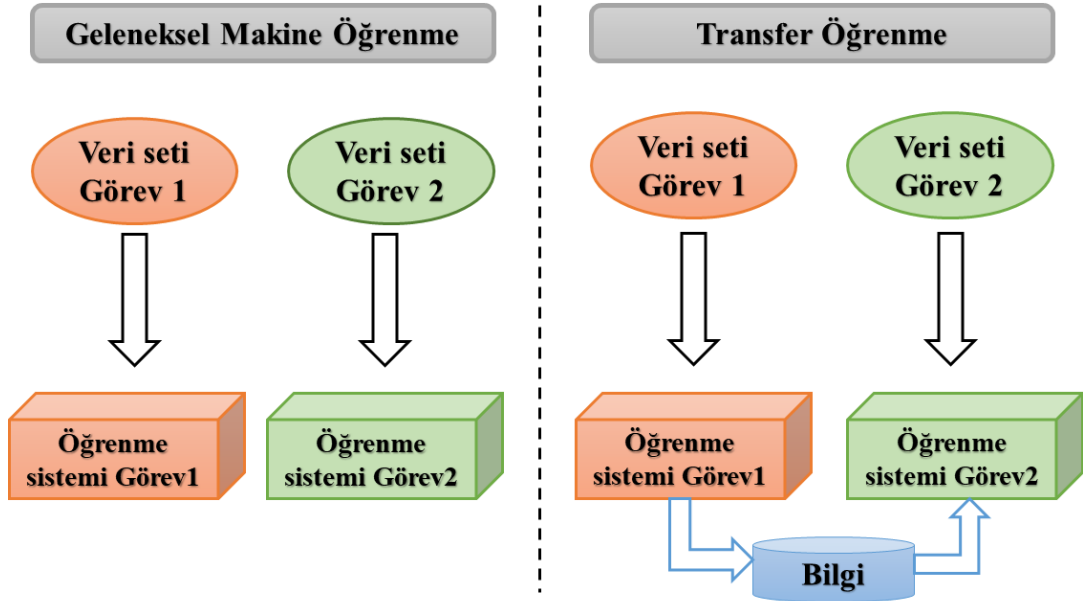
Şekil 2.8 ResNet-50'nin tipik bir rezidüel bloğunun görsel temsili.

2.3 Transfer Öğrenme ile Evrişimli Sinir Ağı Modelleri Geliştirme

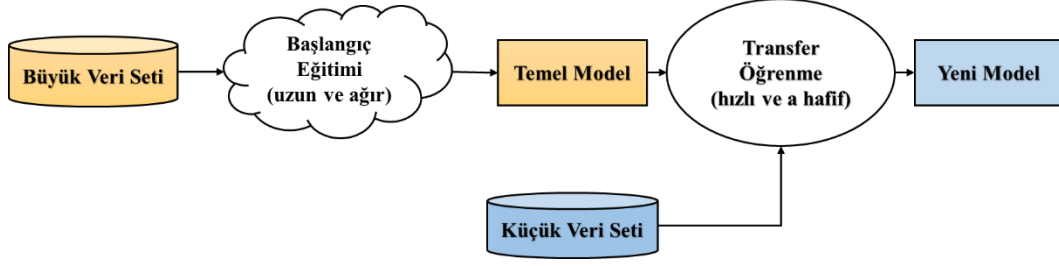
CNN modelini sıfırdan eğitmek önemli bir zaman ve hesaplama kaynağı gerektirebilir. Ayrıca CNN modelleri performanslarını artırmak için büyük miktarda etiketli veriye ihtiyaç duyulmaktadır. Yeterli veri eksikliği, derin bir CNN'in eğitiminde aşılması en zor sorunlardan biridir. Bu nedenle, bir ağı sıfırdan eğitmek için gereken kaynakları azaltmak ve yeterli veri eksikliği sorununu ele almak amacıyla CNN'lerin rastgele ağırlıklarla başlatmak yerine, benzer bir problem de önceden eğitilmiş ağırlıklarla başlatılması yaygın olarak kullanılmaktadır. Bu, insanların bir görevden edindikleri bilgiyi başka bir görevi öğrenmek için kullanabilmesine benzemektedir [74]. Bu öğrenme yaklaşımına transfer öğrenme denilmektedir. Transfer öğrenme, makine öğrenmesinin bir dalı olarak, belirli bir problemi çözerken farklı ancak ilişkili problemlerin çözümünden elde edilen bilgiyi kullanmayı hedeflemektedir [75]. Bu bölüm transfer öğrenme kavramını ve bilgisayarlı görü uygulamaları için CNN geliştirirken nasıl kullanılacağını açıklamaktadır.

2.3.1 Transfer Öğrenme Kavramı

Transfer öğrenme kavramı genel olarak bir modelin birinci problem üzerinde eğitildikten sonra ikinci bir ilişkili problemde uygulanma fikrine atıfta bulunmaktadır. Genellikle derin öğrenme modelleri, aynı alan uzayından elde edilen eğitim ve test veri kümeleri kullanılarak belirli bir problem için tasarlanırlar. Transfer öğrenme, farklı alanlardan veya dağılımlardan elde edilen eğitim ve test verilerini kullanabilen algoritmaların geliştirilmesi yoluyla bunu değiştirmeyi amaçlamaktadır. Gerçekte bu oldukça önemlidir çünkü birçok gerçek yaşam uygulamasında gerekli eğitim verisinin yeterli miktarda toplanması ve modellerin yeniden oluşturulması zordur. Bilginin başarılı bir şekilde transfer edilmesi durumunda, veri etiketleme maliyetini ortadan kaldırmak ve öğrenme kapasitesinin performansını iyileştirmek mümkündür [74]. Şekil 2.9, geleneksel öğrenme yaklaşımı ile transfer öğrenme yaklaşımı arasındaki farkı ve öğrenme süreçlerini göstermektedir [75]. Geleneksel öğrenme yaklaşımı her bir görevi sıfırdan öğrenmeyi hedeflerken, transfer öğrenme yaklaşımı önceden gerçekleştirilen görevlerden elde edilen bilginin bir kısmını mevcut hedef göreve aktarmayı amaçlamaktadır. Eğitim veri setinin az olması nedeniyle önceki görevlerin yerine getirilmesinden öğrenilen bilgi daha sonra eğitim verilerinin yerine kullanılır. Transfer öğreniminin genel akışını gösteren şematik diyagram Şekil 2.10'da gösterilmektedir.



Şekil 2.9 Geleneksel makine öğrenimi ve transfer öğrenimi yaklaşımları arasındaki fark.



Şekil 2.10 Transfer öğreniminin genel akışının şematik bir diyagramı.

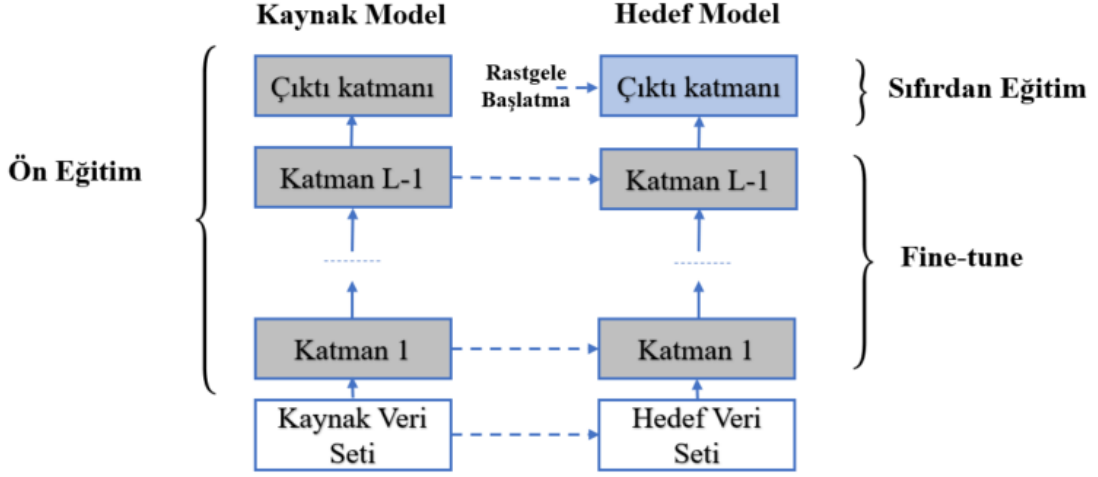
Kaynak ve hedef görevleri arasındaki fark ve eğitim verilerinin etiketli olup olmamasına bağlı olarak, transfer öğrenmede kullanılan çeşitli ayarlamalar bulunmaktadır: inductive, transductive ve gözetimsiz transfer öğrenme (unsupervised transfer learning). Inductive transfer öğrenmede, kaynak görev ve hedef görev farklıdır ve aynı alandan olup olmadıklarına önem verilmez. Transductive transfer öğrenmede ise hedef ve kaynak görevler aynıdır ancak etki alanları farklıdır. Son olarak gözetimsiz transfer öğrenmede kaynak ve hedef görevler farklıdır ancak ilişkili olabilirler [75]. Bu modellerin kişi Re-ID probleminde kullanılabilmesi için görüntü tanımadan (image recognition) elde edilen bilginin kişi Re-ID alanına aktarılması gerekmektedir. Burada hedef görev olan kişi Re-ID ve kaynak görev farklıdır ve aynı şekilde alanları da farklıdır. CNN'lerde transfer öğrenimi kullanılırken, sadece yeni eğitim verileriyle CNN'in yeniden eğitilmesi değil, aynı zamanda önceden eğitilmiş ağ ağırlıklarının da geri yayılım süreci devam ederken fine-tuning yapılması gerekmektedir.

2.3.2 Önceden Eğitilmiş Derin Modellerin İnce Ayar ile Transfer Öğrenme

İnce ayar (fine-tuning), transfer öğrenme stratejisinin bir parçasıdır ve özellikle kişi yeniden tanımlama gibi küçük veri setleriyle çalışan problemlerde başarılı bir şekilde kullanılan bir yöntemdir. Bu yöntem, CNN modelini yeniden eğitmenin yanı sıra, önceden eğitilmiş CNN modelinin ağırlıklarını da eğitim süreci devam ederken ince ayar yapmayı içermektedir. Önceden eğitilmiş CNN modelleri genellikle nesne tanıma görevleri için ImageNet veri kümesiyle eğitilmiş temel ağlardır [72]. ImageNet veri kümesi, sınırlayıcı kutularla birlikte 1000'den fazla farklı sınıfı temsil eden bir milyondan fazla görüntü içermektedir. Bu nedenle, kişi Re-ID gibi ilgili alanlar için yeterli miktarda bilgi içermektedir. İnce ayar işlemi, önceden eğitilmiş modelin genel özneliliklerini korurken, yeni veri kümesine daha iyi uyarlanmasını sağlamaktadır [76]. Bu sayede, model daha spesifik öznelilikleri öğrenir ve kişi Re-ID gibi özel bir görev için daha iyi performans sergiler. Temel olarak ince ayarlama işleminin kişi Re-ID problemleri için

uygun bir seçenek olması iki önemli parametreye dayanmaktadır. İlk olarak, eğitim veri setinin boyutu küçüktür ve ikinci olarak, orijinal eğitim veri setinden oldukça farklıdır [75]. Önceden eğitilmiş ağırlar, ağırlık başlangıcı için kullanılır. Atfedilen bir tahmin ağının ağırlıklarını ImageNet'in önceden eğitilmiş ağırlıklarıyla başlatarak, eğitim hızlı ilerlerken daha az zaman alır ve daha iyi performans göstermektedir. Ayrıca ağı aşırı uyum olasılığını da azaltabilmektedir. Önceden eğitilmiş ağırlıklarla bir ağı başlattıktan sonra, bir sonraki adım, ağı hedef alana uyacak şekilde yapılandırmaktır. Genellikle, hedef alanındaki sınıf sayısı ile aynı sayıda birim içeren bir final FC katmanı ile değiştirmeyi gerektirir [77].

CNN modelleri, farklı öznelikleri hiyerarşik olarak öğrenen katmanlı mimarilere sahiptir. Bu katmanlar genellikle bir FC katmanı olan son katmana bağlanarak denetimli öğrenme ile son çıktı elde etmek için kullanılır. Bu katmanlı yapı, ResNet-50, DenseNet-121 vb. gibi önceden eğitilmiş CNN modellerinin son katmanları kullanmadan, sabit bir öznelik çıkarıcısı olarak kullanılmasına olanak sağlamaktadır. İnce ayarlama işlemi için, önceden eğitilmiş modelin FC sınıflandırıcı katmanları çıkarılarak yeni veri üzerinde eğitim yapmak ve yeni görevi gerçekleştirmek için rastgele değerlere sahip yeni FC katmanları eklenir. Katmanların hepsine veya bazısına ince ayar yapmak mümkündür. Genel olarak aşırı uyumdan kaçınmak için alt katmanların aynı kalması ve üst katmanların yalnızca bazılarında ince ayar yapılması gerektiği kabul edilmiştir. Bunun nedeni alt katmanların daha genel öznelikleri, örneğin kenar ve şekil özneliklerini, temsil ederken üst katmanların ise problem alanıyla ilgili daha spesifik öznelikleri içermesidir. Üst katmanlar, yeni eklenen sınıflandırıcı katmanlarla birlikte ortak bir şekilde eğitilir ve önceden eğitilmiş CNN modelindeki üst düzey öznelik temsillerinin ince ayarlanmasına olanak sağlayarak yapılan göreve daha uygun hale getirmektedir. Şekil 2.11, ince ayarlama işlemi için CNN modelinin yapısını gösteren bir gösterim sunmaktadır. Transfer öğrenme yoluyla bir alandan diğerine bilgi aktarmanın tahminlerin doğruluğunu geliştirdiği gösterilmiştir [37], [78], [79].

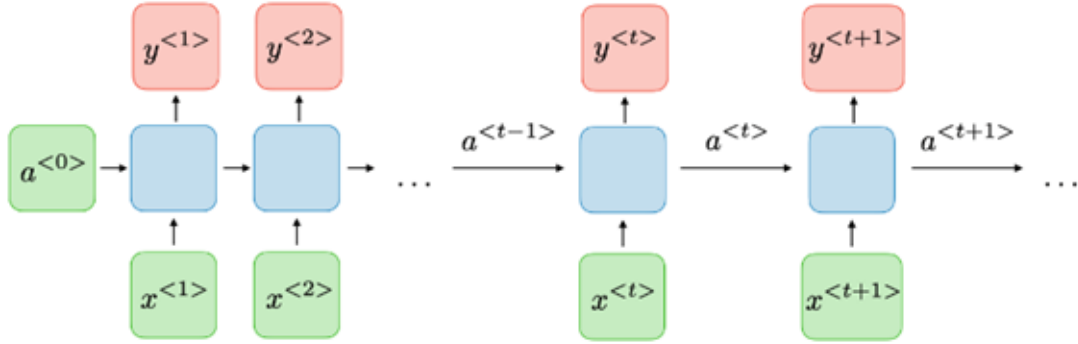


Şekil 2.11 İnce ayar prosedürü için CNN modelinin yapısı.

2.4 Transformer'ler

Transformer'ler, makine öğrenmesinde kullanılan derin öğrenme modelleridir ve dikkat mekanizmalarını kullanarak girdi verilerinin her bir bölümünün önemini farklı şekilde değerlendirir. Bu modeller, öz-dikkat katmanları adı verilen birçok katmandan oluşmaktadır. Dikkat mekanizmasına dayalı Transformer, ilk olarak 2017 yılında Vaswani ve diğerleri tarafından "Attention Is All You Need" başlıklı makalede tanıtılmıştır [80]. Transformer modelleri, oldukça az eğitim süresiyle yüksek performans elde edebilme yeteneklerini göstermiştir. Bu model ilk olarak Doğal Dil İşleme (NLP) alanında uygulanmıştır. Daha önce RNN, LSTM ve Gated Recurrent ağlar gibi modeller, sıralı veri üzerindeki ilgili bilgiyi bulma yetenekleri nedeniyle NLP uygulamalarında yaygın olarak kullanılmaktaydı. Ancak bu ağların NLP görevlerinde önemli dezavantajları bulunmaktadır. Öznitelikle sıralı işleme yapısı nedeniyle, metin verilerinden genel anlamı anlamak için yeterli olmamaktadırlar. Şekil 2.12'deki geleneksel RNN mimarisi görüntüsünde olduğu gibi, her bir zaman adımının çıktısı bir sonraki zaman adımına girdi olarak verilmelidir. Bu süreç, kısa süreli belleğe neden olur ve paralel eğitimi engeller. LSTM, bellek kapasitesini artırabilmesine rağmen yeterli olmayabilir ve hesaplama açısından hala maliyetlidir. Transformer mimarisi, kodlayıcı ve kod çözücü olmak üzere iki ana bölümden oluşmaktadır. Geleneksel RNN ve LSTM modelleri genellikle metin verilerini kelime kelime işleyerek gömme katmanlarında işlemektedir. Ancak Transformer mimarisi farklı bir yaklaşım sunmaktadır. Transformer'ler, tüm cümle gibi tüm girdileri bir kerede gömme olarak alır ve dikkat

mekanizmaları kullanarak işlem yapmaktadır. Dikkat bloğu, giriş cümlesindeki kelimeler arasındaki ilişkileri ve bağımlılıkları çıkarmak için kullanılan bir mekanizmadır. Transformer mimarisi, dikkat mekanizmaları ve gömmelerin yanı sıra LayerNorm, ileri beslemeli ağlar ve softmax fonksiyonları gibi diğer bileşenleri de kullanmaktadır.

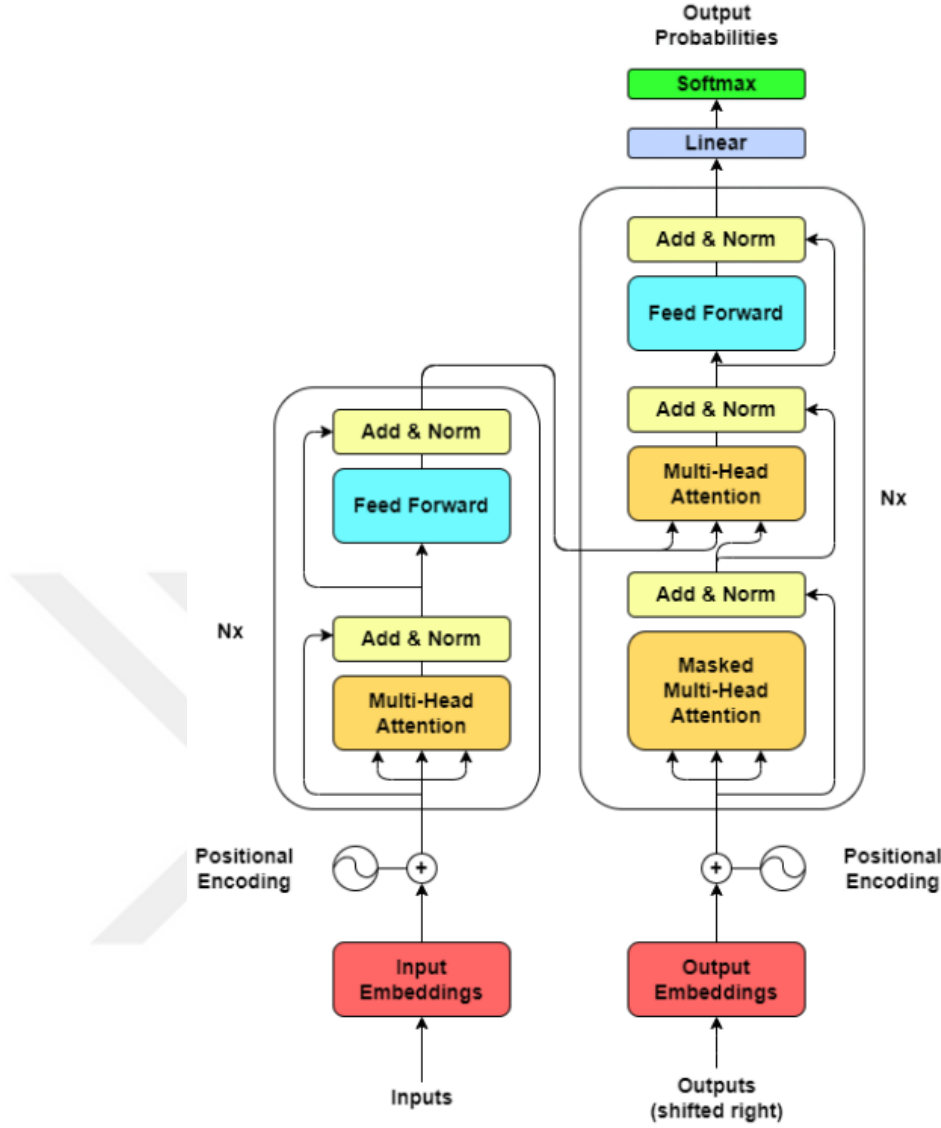


Şekil 2.12 RNN genel mimarisi[81].

2.4.1 İlk Transformer

İlk Transformer'ler oluşturduğunda, diğer modeller tarafından da yaygın olarak kullanmakta olan encoder-decoder yapı kullanmıştır [80]. Kodlayıcı (encoder), bir sembol gösterimler dizisini $(x_1, \dots, x_n)$ sürekli gösterimler dizisine $z = (z_1, \dots, z_n)$ dönüştürmektedir ve decoder daha sonra tekrardan semboller dizisi $(y_1, \dots, y_n)$ oluşturmak için kullanılmaktadır. Model her adımda otomatik regresif (regressive) olduğu için bir sonrakini oluştururken ek girdi olarak önceden oluşturulmuş sembolleri kullanmaktadır.

Şekil 2.13 modelin hem kodlayıcı hem de decoder için yığılmış öz-dikkat ve tamamen bağlantılı katmanlara sahip bir yapı kullandığını göstermektedir.



Şekil 2.13 Orijinal Transformer mimarisi [80]

Kodlayıcı sol tarafta, decoder ise sağ tarafta yer almaktadır. Hem Kodlayıcı hem de decoder, birbirlerinin üzerine yığılabilen modüllerden oluşmaktadır ve bu da şekilde N olarak gösterilmiştir. Modüller çoğunlukla Çok Başlı Dikkat (MSA) ve İleri Beslemeli katmanlardan oluşmaktadır. Diziler doğrudan kullanıldığı için girdiler ve çıktılar (hedef cümleler) önce n -boyutlu bir uzaya gömülmektedir. Girdi gömmesinden elde edilen çıktı, daha sonra konumsal kodlayıcıya giden bir vektördür. Kelimeler bir cümle içinde bulundukları yere bağlı olarak farklı anlamlara sahip olduklarından konumsal kodlayıcı, cümle içindeki kelimeler arasındaki mesafeler hakkında bilgi içeren bir vektör üretmektedir. Ramachandran ve arkadaşları 2019 yılında bu mesafeyi elde etmek için aşağıdaki işlevleri kullanmışlar [82]:

$$PE_{(pos,2i)} = \sin \left(\frac{pos}{10000^{\frac{2i}{d_{model}}}} \right) \quad (2.11)$$

$$PE_{(pos,2i+1)} = \cos \left(\frac{pos}{10000^{\frac{2i}{d_{model}}}} \right) \quad (2.12)$$

Sırasıyla denklem (2.11) ve (2.12), Sin konumsal kodlayıcı ve Cos konumsal kodlayıcıdır. d_{model} , model içindeki alt katmanların çıktı boyutu, pos , kelimenin cümle içindeki konumunu, i ise cümlenin boyutunu (veya uzunluğunu) göstermektedir. Bu aşamada kelime bir cümleye bağlam sağlayan ve konumsal bilgi içeren bir kelime vektörüne çevrilmiştir.

MSA katmanı, bir cümle içindeki kelimelerin önemini hesaplamakla ilgilenmektedir. Her kelime için bir cümle içindeki kelimeler arasındaki bağlamsal ilişkileri yakalayan bir dikkat vektörü çıkarmaktadır. Bu, i sayısında dikkat vektörü çıkartılmaktadır. Ardından diğer modüllerin anlamasını kolaylaştırmak için her bir dikkat vektörüne ileri besleme katmanı eklenmektedir. Katman ekleme ve normalleştirme adımı, katman normalleştirmeyi kullanmaktadır [83].

$$LayerNorm(x + Sublayer(x)) \quad (2.13)$$

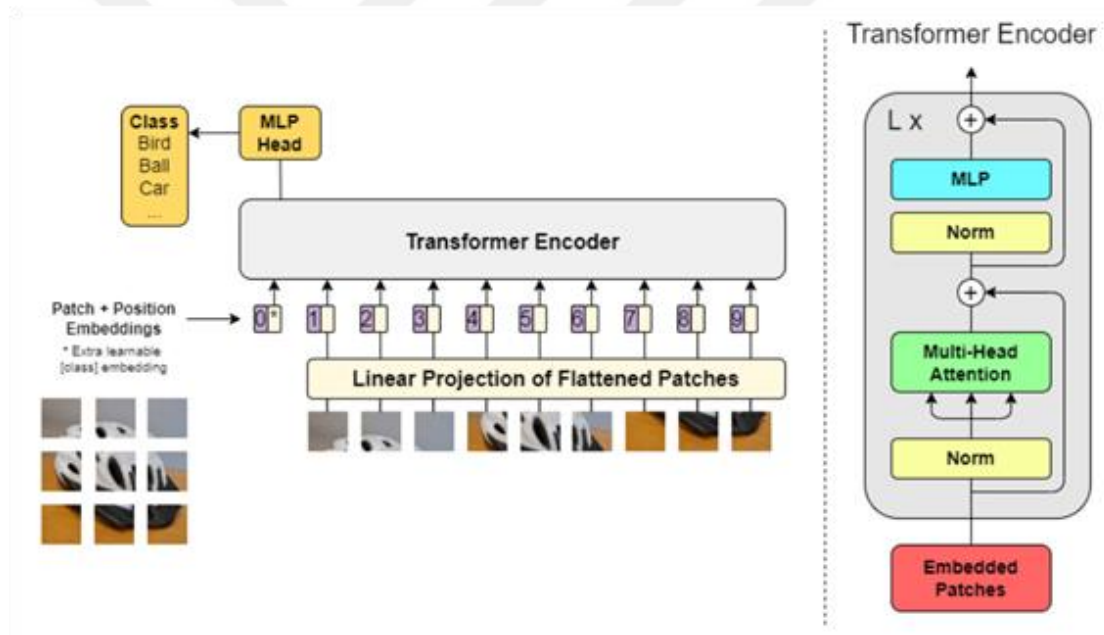
(2.13) denklemi, $LayerNorm$ kullanarak $Sublayer(x)$ olarak adlandırılan önceki katmanın çıktısına x (kelime vektörü) eklemek için kullanılır.

Encoder'ın iki benzer alt katmanı ve bir ek maskeli MSA katmanı bulunmaktadır. Maskeli MSA katmanının rolü, MSA bloğuna benzerdir. Aradaki farkı ağı bir dilin diğerine çevirmek için eğitirken, bazı hedef dil etiketlerin modele gizlenmiş (maskelenmiş) olmasıdır. Bu şekilde ağ çeviri için hemen tüm cevabı kullanamamaktadır. Dolayısıyla cümle içindeki kelimeler tekrarlanırken, cümlenin çoğu kullanılabilir hale gelir.

Maskeli MSA katmanından sonra işlem kodlayıcı ile benzerdir ancak hem orijinal cümle hem de yeni hedef cümle kullanılmaktadır. İleri besleme bloğundan sonra verinin boyutunu hedef dildeki kelime sayısına genişletmek için doğrusal katmana gönderilmektedir. Son olarak softmax katmanı veriyi olasılık dağılımına dönüştürür ve bu da insanlar tarafından okunabilir hale gelmektedir. Çıktı kelime, en yüksek olasılığa sahip olan kelimedir.

2.4.2 Vision Transformer

Dosovitskiy ve arkadaşları, 2020 yılında bilgisayarlı görü alanına Transformer'leri tanıtır ViT ismini verdiler [84]. NLP alanından standart bir transformer olarak mümkün olduğu kadarıyla bazı değişiklikler yaparak doğrudan görüntülere uygulamışlardır. "Orta ölçekli" veri kümeleri üzerinde eğitildiklerinde, transformer modellerinin iyi genelleme yapmadığı gözlenmiştir [84]. Bu nedenle, araştırmacılar modellerini başka veri kümeleri ve kendi veri kümeleri üzerinde önceden eğitmiştir. Diğer görüntü tanıma modellerine kıyasla çeşitli benchmark testlerinde daha iyi performans elde ederek sonuç olarak ImageNet'te %88,55 doğruluk elde etmişlerdir. Bir süre sonra [85] çalışmasında ViT modelini ImageNet veri setini kullanarak ön eğitim yapmadan test etmişler ve %85,2 doğruluk elde etmeyi başarmışlardır. Bu sonuçlar, ViT'nin bilgisayarlı görü alanında temel bir model olarak kabul edilmesini sağlamaktadır [84].



Şekil 2.14 ViT'nin görsel temsili [84]

Şekil 2.14, ViT'in görsel temsili sunmaktadır. Orijinal Transformer'e (Şekil 2.13) kıyasla, ViT'de kodlayıcıya girmeden önce bazı farklı adımlar görülebilmektedir. Standart transformer bir belirteç dizisi olarak tek boyutlu bir giriş alırken, görüntüleri işleyebilmek için onları yeniden şekillendirmek gerekmektedir. ViT'de [84] bir görüntüyü $x \in R^{H \times W \times C}$ olduğu halden düzleştirilmiş iki boyutlu yamalar dizisine $x_p \in R^{N \times (p^2 \cdot C)}$ dönüştürmektedir. Burada, (H, W) orijinal görüntünün yüksekliği ve genişliği,

C ise renk kanallarının sayısıdır. (P, P) her görüntü parçasının boyutlarıdır ve N , denklem (2.14)'de görüldüğü gibi, ortaya çıkan yama sayısını ifade etmektedir.

$$N = HW/P^2 \quad (2.14)$$

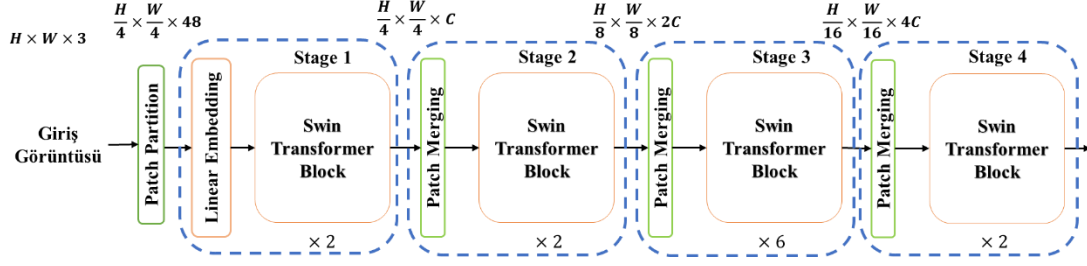
Transformer kodlayıcı bloğundaki katmanların sırası biraz değiştirilmiş ve MSA ve çok katmanlı perceptron (MLP) öncesinde katman normalizasyonu (LN) uygulanmıştır. Şekil 2.13'te "Ekle ve Norm" adımını bloklar içinde kullanılmıştır. Ancak Şekil 2.14'te gösterildiği gibi "Ekle" kısmı artı (+) işareti gibi görselleştirilmiş ve "Norm" LN ile ilişkilendirilmiştir. ViT'ki [84] yapılan bu işlem, [86], [87] çalışmalarından esinlenmiştir.

2.4.3 Swin Transformer

Swin Transformer, Ze ve arkadaşları tarafından önerilen bir sinir ağı mimarisidir [88] Görüntü tanıma görevlerinde etkileyici bir performans sergilemektedir. Bu mimari, ViT mimarisinin sınırlı ölçeklenebilirliğini ve yüksek hesaplama maliyetini artırma zorluğunu aşmak için bir çözüm olarak önerilmiştir. Swin Transformer, görüntüyü birden çok ölçekte işleyen bir hiyerarşik yapı sunmaktadır. Ayrıca, "kaydırılmış pencere" mekanizmasını kullanarak görüntüyü daha az sayıda belirteçle işlemesine izin vererek hesaplama maliyetini azaltmaktadır. Bu özellikler Swin Transformer'ın verimli ve ölçeklenebilir olmasını sağlamaktadır. Swin Transformer, görüntü sınıflandırma, nesne algılama ve semantik bölümlendirme gibi çeşitli bilgisayarlı görü görevlerinde önceki transformer tabanlı modellere kıyasla önemli gelişmeler göstermiştir. Görüntü tanıma görevlerinde son teknoloji modellerden biri haline gelmiştir ve başarılı transformer tabanlı mimarilere yönelik daha fazla araştırma yapılmasına ilham kaynağı olmuştur.

2.4.3.1 Swin Transformer modelin mimarisi

Swin Transformer [88], kaydırılan pencereleri kullanarak ViT'ye dayanan bir yapıya sahiptir. Swin Transformer'ın mimarisi, Şekil 2.15'de gösterildiği gibi ViT ile aynıdır. Bu iki modül, ViT'dekiyle yaklaşık olarak aynıdır. Ancak Çoklu-Kafa Dikkat mekanizması (MSA), sabit pencere (W-MSA) ve kaydırılmış pencere (SW-MSA) dikkat mekanizmalarıyla değiştirilmiştir. Bu dikkat mekanizmaları, belirli bir pencere ve kaydırılan bir pencere kullanarak MSA'yi gerçekleştirilmektedir. Böylece hesaplama verimliliğini artırır ve sonuçta sınıflandırma doğruluğunu iyileştirmektedir.



Şekil 2.15 Swin Transformer modeli.

Giriş olarak verilen $x \in R^{H \times W \times C}$ bir görüntüdür ve boyutları H , W ve C sırasıyla görüntünün yükseklik, genişlik ve kanal sayısını temsil etmektedir. Bu boyutlandırma, ViT çerçevesindeki yama boyutlarının bölünmesiyle uyumludur. Kayan pencere mekanizması kullanılarak, x görüntüsü kısmen piksel örtüşen yama bloklarına bölmek için kaydırılmaktadır. Giriş görüntüsü x , $H \times W$ çözünürlüğe sahip N sabit boyutlu yama bloğuna aşağıdaki denklemleri kullanarak bölünmektedir:

$$N = N_H * N_W = \left\lfloor \frac{H + S - P}{S} \right\rfloor * \left\lfloor \frac{W + S - P}{S} \right\rfloor \quad (2.15)$$

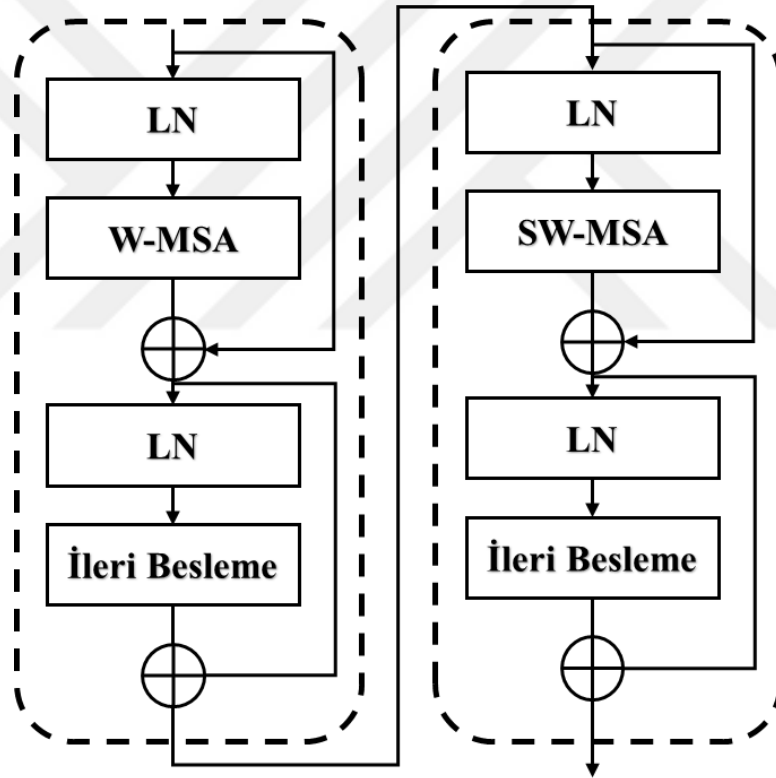
N_H ve N_W , x görüntüsünün yükseklik ve genişlik boyunca kaç yama bloğuna bölündüğünü temsil etmektedir. S , kaydırma adım uzunluğunu, P ise yamanın kenar uzunluğunu göstermektedir.

Her yama ayrıca bir belirteç olarak adlandırılır ve her belirtecin pikselleri RGB üç renkli kanallardır. Uygulamada bölünmüş yamaların boyutu 4×4 olarak belirlenmiştir. Bu nedenle her yamanın öznitelik boyutu $4 \times 4 \times 3 = 48$ 'dir. Bu orijinal değer özniteliklerine bir doğrusal gömme katmanı uygulanmaktadır ve onu belirli bir boyuta (C olarak gösterilir) yansıtmaktadır. Ardından her giriş yaması birkaç Swin transformer bloğundan geçirilmektedir ve dolayısıyla C her belirtecin sayısını göstermektedir. Bu yama belirteçlerine birkaç Swin transformer blokları uygulanmaktadır. Transformer blokları, belirteç sayısını ($H/4 \times W/4$) koruyarak doğrusal gömme ile birlikte "Aşama 1" olarak anılmaktadır.

Ağ daha derinleştikçe, hiyerarşik bir temsili üretmek için belirteç sayısı yama birleştirme katmanlarıyla azaltılır. İlk yama birleştirme katmanı, her biri 2×2 bitişik yamanın özniteliklerini birleştirilerek $4C$ boyutunda birleştirilmiş özniteliklere doğrusal bir katman uygulamaktadır. Bu, belirteç sayısını $2 \times 2 = 4$ kat azaltarak, çıkış boyutunu $2C$ olarak ayarlanmaktadır. Daha sonra Swin Transformer blokları öznitelik dönüşümü için

uygulanırken, çözünürlük $H/8 \times W/8$ olarak korunur. Bu ilk yama birleştirme ve öznitelik dönüşüm bloğu "Aşama 2" olarak adlandırılmaktadır.

Bu işlem, "Aşama 3" ve "Aşama 4" olarak adlandırılan iki kez tekrarlanır ve sırasıyla $H/16 \times W/16$ ve $H/32 \times W/32$ çözünürlüklerinde çıkış üretmektedir. Bu aşamalar, VGG [65] ve ResNet [71] gibi tipik evrişimli ağların öznitelik haritası çözünürlüklerine sahip olan bir hiyerarşik temsili birlikte oluşturulmaktadır. Swin Transformer Şekil 2.16'da gösterildiği gibi iki ardışık blok içermektedir. Bir Swin Transformer bloğu, bir GELU katmanına sahip 2 katmanlı bir MLP tarafından takip edilen kaydırma penceresi tabanlı bir MSA modülünden oluşmaktadır. Swin Transformer'de her MSA modülü ve her MLP modülünden önce LN katmanı uygulanmaktadır. Ayrıca her modülün çıkışı bir artıklık bağlantı (residual connection) ile önceki katmanın çıkışına eklenmektedir.



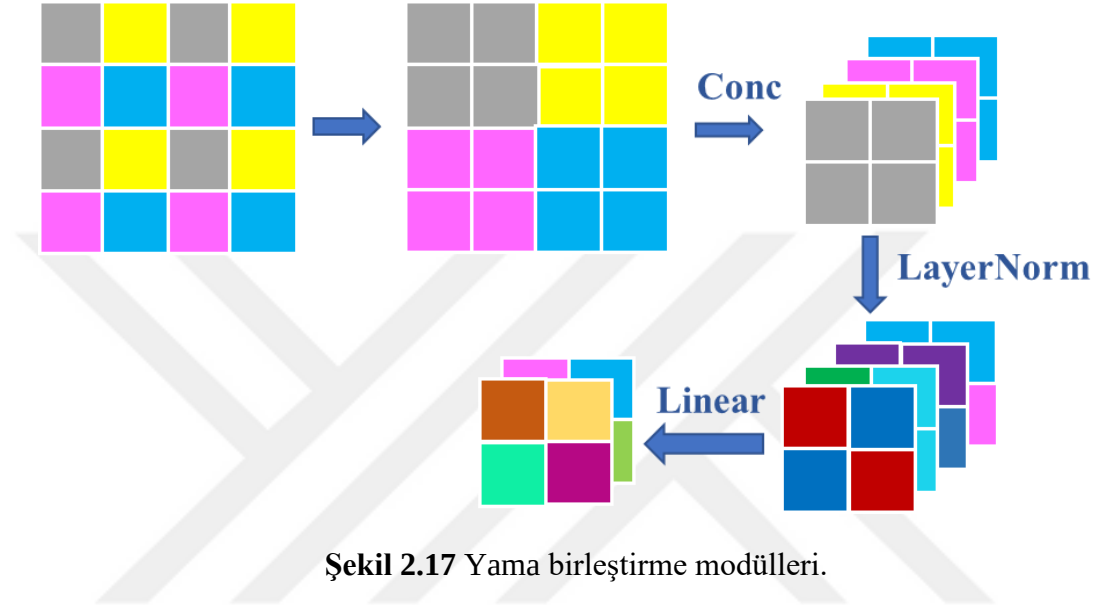
Şekil 2.16 Swin transformer blok modülleri.

2.4.3.2 Yama Birleştirme Modülleri

Yama Birleştirme modülü $H \times W$ boyutunda bir yamayı alır, öncelikle onu keserek kanal boyutunda tam olarak birleştirir, böylece $H/2 \times W/2 \times 4C$ öznitelik haritası oluşturmaktadır. Daha sonra düzenleme için normalleştirme işlemi gerçekleştirilir ve ardından doğrusal katmanlar aracılığıyla boyut azaltma yapılarak $H/2 \times W/2 \times 2C$

boyutunda öznitelik haritası oluşturulur. Bu şekilde öznitelik haritasının alt-örnekleme işlemi gerçekleştirilmiştir. Şekil 2.17'te gösterildiği gibi boyut orijinal boyutun yarısına indirgenirken, derinlik ise orijinal boyutun iki katına çıkarılmaktadır.

Her yama birleştirme işlemi gerçekleştirildiğinde, yan yana olan dört belirteç birleştirilir. Yama birleştirme, evrişimli sınır ağında havuzlama katmanını gibi görüntüyü alt-örnekleyerek alandan büyük bir değeri elde etmektedir.



Şekil 2.17 Yama birleştirme modülleri.

2.4.3.3 W-MSA modülü

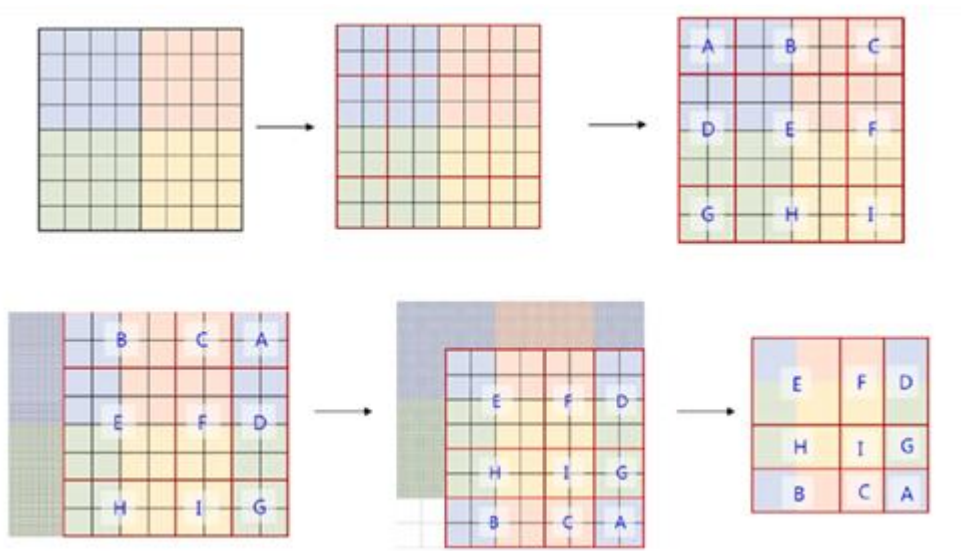
ViT ağındaki MSA, her pikselin diğer piksellerle iç çarpım yaparak tüm piksellerin bilgisini öz-dikkatli bir işlemle elde etmesini sağlayarak zengin global detaylar elde etmektedir. Ancak her pikselin diğer bütün pikseller ile bilgi alışverişinde bulunması gerekmektedir. Bu nedenle, ağın yürütülmesi için hesaplama yoğunluğu ve verimsizlik yaratmaktadır. Bu zorluğa istinaden Swin-T, MSA'yı birden çok sabit pencereye bölerek W-MSA'yı oluşturmaktadır. Bir penceredeki her piksel, yalnızca aynı penceredeki diğer piksellerle iç çarpım yaparak bilgi almaktadır. Dolayısıyla bu işlem hesaplama miktarını dramatik bir şekilde azaltır ve ağın verimliliğini artırmaktadır.

Giriş olarak bir görüntü alınarak bir tensöre (tensor) dönüştürülmektedir. Genellikle giriş görüntüler üç kanaldan (RGB) oluşmaktadır. Yama gömme işlemi yapıldıktan sonra derinlik 96'ya artarak orijinal görüntüdeki yama boyutu 4×4 olmaktadır. RGB kanallarına sahip olan görüntü birden fazla pencereye bölünmektedir. Bu pencereler oluşturulduktan sonra her bölümlendirilmiş pencere için W-MSA kullanılarak öz-dikkat hesaplaması yapılmaktadır. Bu pencereler arasında herhangi bir ilişki bulunmamaktadır.

Girişte ne kadar belirteç mevcut ise o kadar çıktı üretilmektedir. Öz-dikkat hesaplaması her pencere içinde gerçekleştirilmektedir.

2.4.3.4 SW-MSA modülü

W-MSA, pencereleri bölerek hesaplama yükünü azaltırken, pencereler arasında bilgi etkileşimi olmaması nedeniyle sınırlı bir "algılama alanı" olarak görülmektedir. Bu durum daha global olarak doğru bilgilere ulaşmayı engeller ve bu da ağın doğruluğunu etkilemektedir. Diğer pencereler arasında bilgi etkileşimi sağlamak için pencereyi kaydırabilir, pencereyi farklı pikseller içerecek şekilde farklı bir konuma taşıyabilir, ardından bir W-MSA hesaplama işlemi gerçekleştirilebilir. İki W-MSA hesaplama sonucu ve iki farklı pencerenin piksellerinde bulunan bilgiler birleştirilerek pencereler arasında ortak bilgi elde edilmektedir. Kaydırılmış pencerenin W-MSA'sı, SW-MSA modülünü oluşturmaktadır. Bu modülde pencere, W-MSA'nın sağ alt köşesine iki yama kadar kaydırılarak farklı boyutlarda dokuz blok oluşturmaktadır. Bu yapı şeması Şekil 2.18'te gösterilmiştir. Daha sonra dokuz blok dairesel kaydırmalar kullanılarak W-MSA ile aynı boyuta sahip dört bloğa dönüştürülmektedir. Maskeli MSA, bu dört birleştirilmiş blok için ilgili şablonları hesaplayarak bilgi elde etmektedir. Bilgi veri yamalarını orijinal konumlarına geri kaydırmak için dairesel değişimi tersine çevirmektedir. SW-MSA mekanizması, kaydırılmış pencerenin piksellerinin MSA'yı hesaplayarak farklı pencereler arasında piksellerin bilgi alışverişini gerçekleştirmektedir. Böylece ağın "algılama alanını" dolaylı olarak genişletir ve bilgi kullanım verimliliğini artırılmaktadır.



Şekil 2.18 Kaydırılan pencere bölmelerinde öz dikkat için kullanılan hesaplama yöntemi.

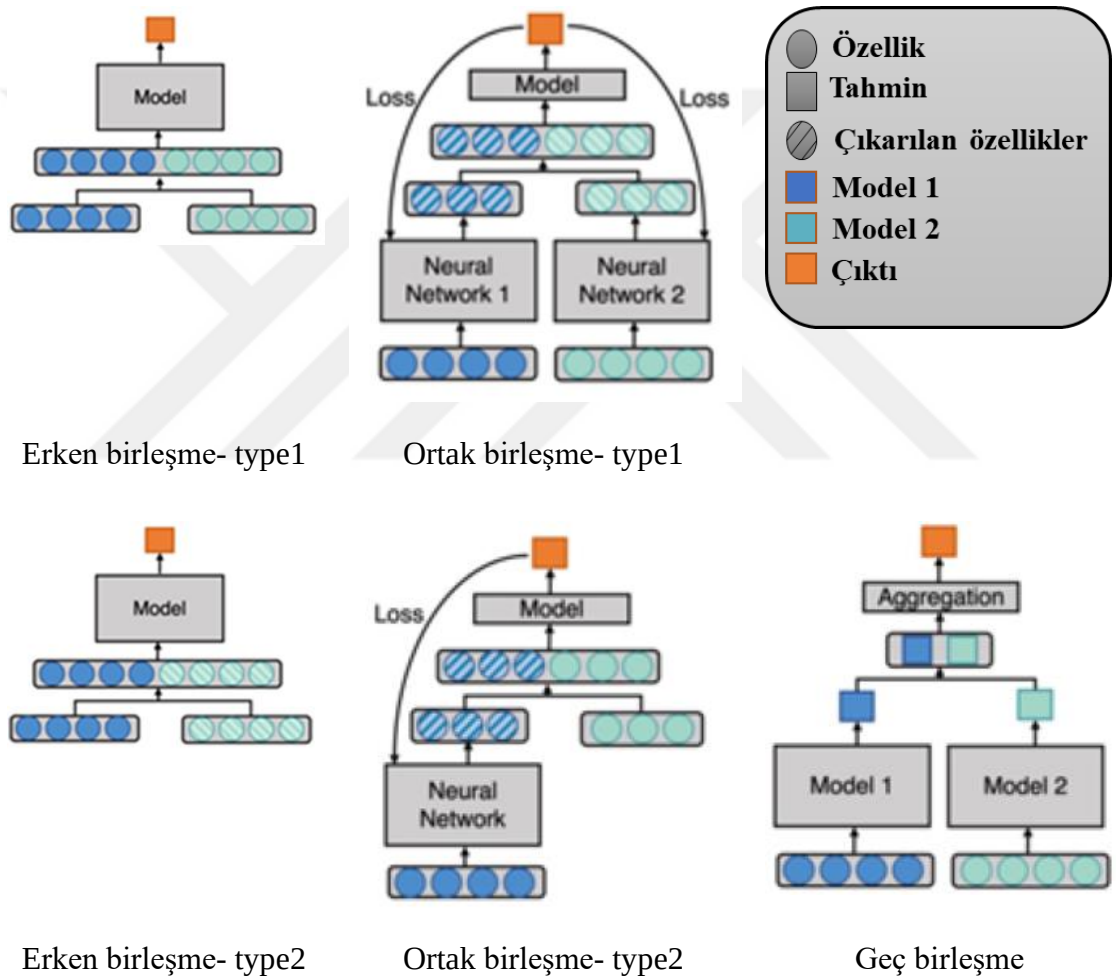
2.5 Kişinin Yeniden Tanımlanmasında Kayıp Metrikleri

Kişi yeniden tanımlanması için derin metrik öğrenme tasarımında kimlik kaybı (identity loss), doğrulama kaybı (verification loss) ve üçlü kayıp (triplet loss) gibi çeşitli popüler kayıp fonksiyonları yaygın olarak kullanılmaktadır. Kimlik kaybı, kişi yeniden tanımayı bir sınıflandırma problemine dönüştürerek elde edilmektedir. Re-ID sistemine bir sorgu görüntüsü verildiğinde, sistemin çıktısı ilgili kişinin kimliği veya adı olmaktadır. Kimlik kaybını elde etmek için genellikle çapraz entropi (cross entropy) [89] fonksiyonu kullanılmaktadır. Doğrulama kaybı, iki özne arasında optimal bir ikili çözüm aramaktadır. Bu amaçla yaygın olarak kontrastif kayıp (contrastive loss) [90] ve ikili doğrulama kaybı (binary verification loss) [39] gibi fonksiyonlar kullanılmaktadır. Kontrastif kayıp, gömme öznitelik uzayında bir ikili mesafe ve bir ikili etiketin (doğru/yanlış eşleşmesi için 1/0) doğrusal bir kombinasyonu ile gösterilmektedir. İkili doğrulama kaybı ise pozitif ve negatif görüntü çifti kümeleri arasındaki farkı göstermektedir. Üçlü kayıp [91] ise yeniden tanımayı, pozitif ve negatif çiftler arasındaki öznitelik mesafesini belirli bir eşikle kontrol etme ilkesi üzerine kurmaktadır. Daha spesifik olarak pozitif çift arasındaki mesafe, negatif çiftten belirlenen eşikten daha küçük olmalıdır. Gözlemlendiği kadarıyla bazı yöntemler yukarıda belirtilen üç tür kaybın kombinasyonunu kullanmaktadır. Ancak bilindiği kadarıyla, çok az yöntem yeniden kimlik belirleme için eğitim kaybı tasarımında odak kaybını kullanmamaktadır. Veri dengesizliği sorununu ele almak için odak kaybının kullanımı, eğitim sırasında zor örnekleri öğrenmeye odaklanan ağlar için etkili olduğu kanıtlanmıştır [92]. Derin metrik öğrenmede odak kaybını dahil etmek, kişinin kimliğinin yeniden belirlemenin genel performansını artırmak için faydalı olabilmektedir. Bu nedenle, bu araştırmada odak kaybını içeren kayıp fonksiyonların birleşimi kullanılmıştır.

2.6 Füzyon Terminolojisi ve Stratejileri: Birleştirme Yöntemleri ve Stratejileri

Veri birleştirme, derin öğrenme için farklı kaynaklardan veya yöntemlerden elde edilen birden çok veriyi tek bir temsil halinde birleştirme işlemidir. Bu işlem, verilerin ayrıntılarını artırarak derin öğrenme modelinin genel performansını iyileştir ve verilerin ayırt edici gücünü artırmayı amaçlamaktadır. Derin öğrenme bağlamında farklı seviyelerde öznitelik birleştirme yapılabilmektedir ve bunlar giriş veya erken seviye, orta seviye ve çıkış veya geç seviye olabilir. Bu yöntemler derin öğrenme modellerinin daha iyi öznitelik temsilleri oluşturmaya ve daha etkili sonuçlar elde etmesine yardımcı

olmaktadır. Erken füzyon, giriş düzeyinde farklı yöntemlerden gelen bilgileri birleştirerek öznitelik düzeyinde füzyon yapmaktadır. Örnek olarak görüntü ve metin verilerinin birleştirilmesi verilebilir. Orta füzyon ise ağı farklı katmanlarından veya aynı katmandaki farklı evrişim filtrelerinden çıkan öznitelikleri birleştirmektedir. Bu şekilde daha derin ve kapsamlı bir öznitelik temsili elde edilir. Geç füzyon ise çıktı düzeyinde kullanılır ve bağımsız olarak eğitilmiş birden fazla ağ veya modelin çıktılarını birleştirmektedir. Şekil 2.19, erken, orta ve geç füzyon stratejilerini göstermektedir ve her bir stratejiyi ayrıntılı olarak açıklamaktadır. Bu stratejiler, derin öğrenmede veri birleştirme işlemine yönelik farklı yaklaşımları temsil etmektedir.



Şekil 2.19 Farklı füzyon stratejileri için model mimarisi.

Erken birleştirme [93], birden fazla giriş yöntemin tek bir öznitelik vektöründe birleştirme sürecini ifade eden bir füzyon stratejisidir. Bu strateji genellikle öznitelik düzeyinde birleştirme olarak bilinmektedir. Farklı veri kaynaklarından gelen öznitelikleri birleştirmek için çeşitli yaklaşımlar kullanılır ve bunlar arasında toplama, çarpma ve

ağırlıklı ortalama gibi yöntemler bulunmaktadır [93]. Birleştirme, farklı kaynaklardan gelen öznitelikleri belirli bir boyuta sahip tek bir öznitelik vektörüne birleştirmeyi içermektedir. Toplama ve çarpma yöntemleri, öznitelikleri eleman bazında ekleyerek veya çarparak birleştirir. Ağırlıklı ortalama yöntemi ise farklı kaynaklardan gelen özniteliklere farklı ağırlıklar atayarak bunların ortalamasını alır ve tek bir öznitelik vektörü oluşturur. Erken birleştirme tip 1, orijinal özniteliklerin birleştirilmesini temsil ederken, erken birleştirme tip 2, başka bir sinir ağından öğrenilen temsilden çıkarılan özniteliklerin birleştirilmesini temsil etmektedir. Erken birleştirmeye ilgili sorunlardan biri, farklı veri kaynakları arasında zaman senkronizasyonunun belirlenmesidir. Farklı modalitelerden gelen verilerin eşzamanlı olması, doğru birleştirme yapabilmek için önemlidir.

Ortak füzyon [93], derin öğrenme modellerinin ara katmanlarından elde edilen öznitelik temsillerini diğer modalitelerden gelen özniteliklerle birleştirme stratejisidir. Bu füzyon stratejisi, erken füzyon ile karşılaştırıldığında temel bir farklılık gösterir. Kayıp, eğitim sırasında sinir ağlarına geri yayılır ve her eğitim iterasyonunda daha iyi öznitelik temsilleri oluşturulmasına olanak tanımaktadır. Ortak füzyon tip 1 olarak adlandırılan durumda kayıp, öznitelik çıkarma modelinden tahmin modeline doğrudan yayılır ve böylece tüm modalitelerden elde edilen öznitelik temsilleri dikkate alınmaktadır. Ancak Şekil 2.19 ortak füzyon tip 2’de tüm giriş öznitelikleri için ortak füzyon stratejisi gerekmeyebilir. Burada öznitelik çıkarma adımı ortak füzyon olarak tanımlanamamaktadır. Bu durumda ortak füzyon stratejisi, yalnızca tahmin modelinden öznitelik temsillerinin birleştirilmesini içermektedir. Ortak füzyon, sinir ağlarının özniteliklerini daha iyi bir şekilde birleştirmek ve genel performansı artırmak için kullanılan etkili yöntemlerden biri olmaktadır.

Geç füzyon (veya karar düzeyinde füzyon) [93], birden fazla modelin tahminlerinden yararlanarak nihai bir karar verme sürecini ifade etmektedir (Şekil 2.19 Geç Füzyon). Bu strateji genellikle farklı yöntemlerle eğitilen ayrı modellerin kullanılması ve tahminlerin birleştirilmesiyle gerçekleştirilir. Birleştirme için kullanılan toplama işlevi yerine örnek olarak ortalama alma, çoğunluk oylama, ağırlıklı oylama veya her modelin tahminlerine dayalı bir meta sınıflandırıcı kullanılabilir. Toplama işlevinin seçimi genellikle deneylere dayalıdır ve uygulamaya ve girdi yöntemlerine bağlı olarak değişiklik gösterebilmektedir. Bu nedenle, geç füzyon stratejisi birden fazla modelin tahminlerini birleştirerek daha güvenilir olarak doğru sonuçlar elde etmek için kullanılabilir.

Kişi yeniden tanımlama (Re-ID) görevleri için hangi füzyon yaklaşımının diğerlerinden daha iyi olduğunu kesin olarak belirtmek zordur çünkü her yaklaşımın kendi avantajları ve dezavantajları vardır. Öznitelik birleştirme yaklaşımının seçimi belirli veri kümesine, görev gereksinimlerine ve mevcut kaynaklara bağlı olabilmektedir. Geç füzyon stratejisi, modaliteler arasında net bir bilgi ayrımı olduğunda etkili olabilir ve uygulanması kolaydır. Ancak farklı modaliteler veya kaynaklar arasındaki karmaşık etkileşimleri yakalamak için ideal olmayabilir ve değişken uzunluklu girdilerin işlenmesi zorluk çıkarabilmektedir. Çok modlu verilerin erken birleştirilmesi, ilgili yöntemlerin tamamlayıcı doğasından tam olarak yararlanmayabilir ve çok büyük girdi vektörlerine yol açarak işlem gücüne neden olabilmektedir [93]. Ayrıca farklı veri kaynakları arasındaki zaman senkronizasyonunun belirlenmesi gibi sorunlarla karşılaşılabilir. Ara füzyon olarak adlandırılan alternatif bir yaklaşım, çok modlu verilerden öğrenilen temsillerin nasıl ve ne zaman birleştirileceği konusunda daha fazla esneklik sunmaktadır. Bu strateji, farklı özniteliklerin farklı aşamalarda birleştirilmesini sağlayarak daha iyi sonuçlar elde edebilmektedir [93]. Sonuç olarak hangi füzyon yaklaşımının en uygun olduğu spesifik duruma ve ihtiyaçlara bağlıdır. Bu nedenle, kişinin kimliğinin yeniden belirleme görevleri için en iyi füzyon yaklaşımını belirlemek için dikkatli bir analiz ve deneysel çalışmalar gerekmektedir.

2.7 Kişi Kimliğinin Yeniden Belirleme Benchmark Veri Setleri ve Değerlendirme Metrikleri

Kişi kimliğinin yeniden belirleme yöntemlerinin sağladığı performans değerlendirmek için tutarlı veri setleri oluşturmak, araştırmacıların sonuçlarını değerlendirmelerine ve karşılaştırmalarına yardımcı olmak, ayrıca olası ancak öngörülemeyen sınırlamaları belirlemek ve performansı geliştirmeye katkıda bulunmak için önemlidir. Güvenilir bir kişi Re-ID oranına ulaşmak için aydınlatma değişimi, kapanmalar, poz değişiklikleri, dağınık arka plan ve örtüşen bedenler gibi özniteliklere sahip mevcut kişi Re-ID veri setlerinin bulunması çok önemlidir [3]. Ayrıca önerilen kişi Re-ID yöntemlerinin performansını değerlendirmek için uygun performans analiz metriklerine sahip olmak da önemlidir. Aşağıdaki alt bölümler, kişi Re-ID görevini değerlendirmek için yaygın olarak kullanılan veri setlerini ve değerlendirme metriklerini özetlemektedir.

2.7.1 Kişi Re-ID Karşılaştırma veri kümeleri

Önerilen teknikler ve algoritmalar için parametrelerin veya yaklaşımların ön yargı içermediğini garanti etmek amacıyla çeşitli veri setleri kullanılarak doğrulama yapmak

son derece önemlidir. Bu nedenle, kişi yeniden tanımlama görevinin değerlendirilmesi için özel olarak hazırlanmış birçok genel veri seti bulunmaktadır. Bu veri setleri, görüntü sayısı, kimliklerin sayısı, kameraların sayısı ve görüntü türleri gibi öznitelikler bakımından birbirinden farklıdır. Derin modellerin eğitimi için veri toplama sürecinin zaman alıcı bir iş olduğu göz önüne alındığında, mevcut araştırma, halka açık veri kümelerinin değerlendirilmesini gerçekleştirmektedir. Tablo 2.2, popüler benchmark kişi Re-ID veri setlerinin özetini vermektedir.

Bu veri setleri birçok farklı halka açık alandan toplanmıştır. Tablo 2.2'de veri seti boyutunun önemli ölçüde arttığı görülebilmektedir. Yıllar önce veri setleri aslında oldukça küçüklerdi ancak son zamanlarda veri setleri daha geniş bir görüntü çeşidi içermeye başlamıştır. Dahası, veri setlerinin çoğunda kişi görüntüleri genellikle el yapımı sınırlayıcı kutular kullanılarak işaretlenmiştir. Günümüzde kişi tespit teknolojisindeki gelişmeler ve otomatik tespit ve takip algoritmaları, NN'ler ve deformasyona dayalı parça modeli (DPM) [94] gibi, sınırlayıcı kutuların otomatik olarak çizilmesi için sıklıkla kullanılmaktadır. Bu araştırmada Market-1501 [19] veri seti, oldukça büyük bir hacme sahip olmaları ve derin öğrenme modellerini değerlendirmek için kişi Re-ID görevi üzerinde çalışan araştırmacılar arasında popüler seçenek olması nedeniyle kullanılmıştır [3]. Ancak bu araştırmada kişi ReID'sinin bir alt dalı olarak bilinen futbolcu Re-ID'sini ele almak istendiği için SoccerNet-V3 Re-ID veri kümesi [95] de kullanılmıştır. Önerilen algoritmanın tutarlı karşılaştırmalı çalışmaları için bu iki veri seti kullanılmıştır. Aşağıdaki bölümlerde bu veri setlerinin öznitelikleri ve ayrıntıları sunulmaktadır.

Tablo 2.2 Kişi RE-ID'si için yaygın olarak kullanılan veri kümelerinin özeti.

Veriseti	Yayın yılı	Kamera sayısı	Kimlik sayısı	Görüntü sayısı	Dedektör	Görüntü boyutu	Değerlendirme
VIPeR [27]	2007	2	632	1264	El	128 x 48	CMC
ETHZ [96]	2007	1	148	8580	El	Çeşitli	CMC
PRID2011 [97]	2011	2	934	24541	El	128 x 64	CMC
CUHK03 [39]	2014	10	1467	13164	El/DPM	Çeşitli	CMC
i-LIDS [98]	2014	2	300	42495	El	Çeşitli	CMC
Market1501 [19]	2015	6	1501	32668	El/DPM	128 x 64	CMC/mAP
DukeMTMC - reID [99]	2017	8	1812	36411	El	Çeşitli	CMC/mAP
Airport [100]	2017	6	9651	39902	ACF	128 x 64	CMC/mAP
MSMT17 [101]	2018	15	4101	126441	Faster RCNN	Çeşitli	CMC/mAP

SoccerNet-ReID [95]	2022	4	243432	340.993	--	Çeşitli	CMC/mAP
------------------------	------	---	--------	---------	----	---------	---------

- **Market-1501 Veri Kümesi:** Market-1501 [19], günümüzde kullanılabilecek en büyük kişi Re-ID veri setlerinden biridir. 32.668 görüntüden oluşmaktadır ve 1501 kimlik içermektedir. Kişi görüntülerinin yakalanması, Tsinghua Üniversitesi kampüsünde bir süpermarketin önünde konumlandırılan 5 adet yüksek çözünürlüklü ve 1 adet düşük çözünürlüklü olmak üzere toplam 6 kamera kullanılarak gerçekleştirilmiştir. Veri seti, özellikle bisikletler ve çantalar tarafından oluşan birçok kapanma durumunu içermektedir. Görüntülerin birçoğu yüksek kalitede olmasına rağmen birçoğu da yüksek düzeyde bulanıklıkla karşılaşmaktadır. Görüntüleri tanımak ve kırpabilmek için DPM yöntemi [94] kullanılmıştır. Görüntüler 128×64 boyutuna ölçeklendirilmiştir ve duruş, aydınlatma ve arka plan gibi önemli ölçüde değişkenliklerle karşılaşmaktadır. Ayrıca, kırpma süreci sırasında bir kişinin vücudunun bazı bölümlerini görüntüden çıkarılmıştır. Market-1501 veri seti, eğitim, galeri ve sorgu olmak üzere 3 bölümden oluşmaktadır. Eğitim bölümü 751 kimliğin 12.936 görüntüsünden oluşurken, galeri bölümü 750 kimliğin 19.732 görüntüsünü içerir ve sorgu bölümü ise galerideki aynı 750 kimliğin 3368 görüntüsünden oluşmaktadır. Market-1501 veri setinden örnek görüntüler Şekil 2.20'de gösterilmiştir.



Şekil 2.20 Market-1501 veri kümesi görüntülerinin örnekleri. Sütunların her biri bir kimliği temsil etmektedir.

- **SoccerNet-ReID veri kümesi:** Futbol videoları, oyuncular, çizgiler ve diğer belirli nesneler gibi birçok öğeyi içeren zengin bir oyun alanı sunmaktadır. Bu videolar, bilgisayarlı görü için geniş bir veri kaynağı oluşturur ve futbol analizlerinde kullanılan otomatik yöntemler için önemli bir kaynaktır. Bu nedenle, bu futbol maçının

zenginliğini yakalamak ve detaylı otomatik analizlere imkân tanımak için SoccerNet-V3'ü yayınlanmıştır. SoccerNet-V3 veri seti [95], SoccerNet-V2'nin [102] bir genişletmesi olarak, farklı ve zorlu görevleri içeren bir veri kümesidir. Bu görevler arasında olay algılama, sınır tespiti ve kamera çekimi segmentasyonu, tekrar topraklama (replay grounding), kalibrasyon, çoklu görüntülü oyuncu kimliği yeniden belirme ve çok oyuncu takip etme bulunmaktadır. Futbolcuların kimliğinin yeniden belirlenmesi için SoccerNet Re-ID, SoccerNet-V3 veri setinin bir parçasıdır. SoccerNet-ReID'in yayın videoları, aynı eylemi farklı bakış açılarından yakalamayı sağlayan önemli eylemlerin tekrarlarını içermektedir. Kapsamlı yerel bilgilerle aynı anları gösteren canlı ve yeniden oynatılan eylem karelerine açıklamalar eklenmiştir. SoccerNet Re-ID veri seti, 6 büyük ligdeki 400 futbol maçından yayınlanan videoların görüntü çerçevelerinden çıkarılan 340.993 oyuncunun küçük resimlerinden oluşmaktadır. Veriler, eğitim, doğrulama, test ve meydan okuma (challenge) bölümlerine ayrılmıştır. Eğitim verisi toplamda 161,443 kimliğe ait 248234 örneğe sahiptir. SoccerNet-ReID veri seti doğrulama (valid) bölümünde, galeri ve sorgulama olmak üzere iki bölümden oluşmaktadır. Toplamda 29,534 kimliğe ait galeri 34,355 görüntüden ve sorgu bölümü ise galeride aynı 29,534 kimliğin ait 11,638 görüntüden oluşmaktadır. Test bölümü 30,059 kimliğe ait 11777 sorgu görüntüsüne ve 34989 galeri görüntüsüne sahiptir. Challenge bölümünde gerçek değer etiketleri bilinmemektedir. Bu bölümde 22,396 kimliğe ait 9021 sorgu örneği ve 26082 galeri örneği mevcuttur. SoccerNet-ReID veri setinden alınan görüntülerin örnekleri Şekil 2.21'de gösterilmektedir.

Futbolcuların kimlikleri eylemler arasında geçerli değildir ve belirli bir futbolcu tespit edildiği her eylem için farklı bir kimliği vardır. Bu nedenle, değerlendirme sürecinde sadece aynı eylem içindeki örnekler birbiriyle eşleştirilir. Yayınlanan videoda eylemin görüldüğü ilk kareye "aksiyon karesi" denilmektedir. Aynı eylemi gösteren (aynı zamanda) ve daha sonra yayınlanan diğer karelere "tekrar oynatma karesi" denilmektedir. Bu veri setinin amacı oyun sırasında aynı eylemi gösteren birden fazla kamera bakış açısıyla futbolcuları yeniden tanımlamaktır.



Şekil 2.21 SoccerNet-ReID veri kümesi görüntülerinin örnekleri. Satırların her biri bir kimliği temsil etmektedir.

2.7.2 Kişi Re-ID Değerlendirme Metrikleri

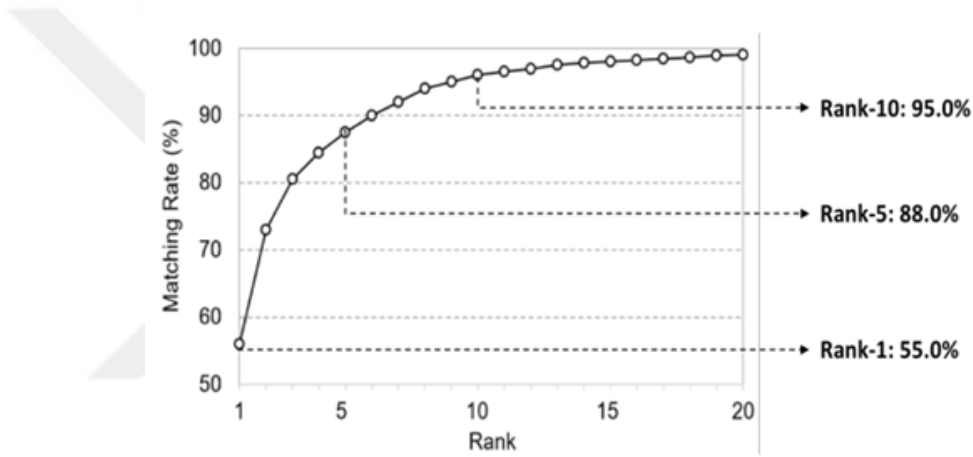
Günümüzde kişi kimliğinin yeniden tanımlama (Re-ID) değerlendirmesi için en yaygın olarak kullanılan 2 standart metrik, kümülatif eşleşme karakteristiği (CMC) eğrisi [103] ve ortalama hassasiyet (mAP) [104] metrikleridir. Bu nedenle, bu çalışmada önerilen modelin performansını değerlendirmek için bu iki metrik benimsenmiştir. Böylece literatürdeki önceki yaklaşımlarla iyi bir karşılaştırma yapılabilmiştir.

- **Kümülatif Eşleşme Karakteristiği (CMC):** CMC eğrisi, kişi Re-ID görevinde yaygın olarak kullanılan bir metriktir. Bu metrik, sorgu görüntüsü ile galeri görüntülerinin benzerliklerine veya uzaklıklarına dayanarak galeri görüntülerinin sıralamasını yapmayı ifade etmektedir. Her bir sıralama için eşleştirme doğrulukları hesaplanır ve her bir görüntü için sorguya göre bir sıralama elde edilir. Doğru bir eşleştirme sıralamasının belirli bir değerden küçük veya eşit olma olasılığı, galeri görüntülerinin sayısına bağlı olarak çizilmektedir [105]. CMC eğrisi, doğru şekilde yeniden tanımlanan sorguların kümülatif sayısını gösterir ve bu sayı, sıralama listesine dayanarak hesaplanmaktadır. Rank (sıra) 'de doğru şekilde yeniden tanımlanmış sorguların sayısının $q(i)$ olduğu bir durumda, o rank için CMC değeri aşağıda verildiği gibi tanımlanır:

$$CMC(i) = \sum_{r=1}^i q(r) \quad (2.16)$$

Burada, r sıra indeksini temsil etmektedir. CMC eğrisinde, ilk nokta ilk yeniden tanımlama oranı olarak bilinir ve Rank-1 olarak adlandırılmaktadır. Rank-1 için eşleştirme oranı, yalnızca ilk sıranın eşleştirme doğruluğudur ve Re-ID operatörleri

için önceliklidir. Yüksek bir Rank-1 eşleştirme oranı, kişi Re-ID modelinin gerçek dünya uygulamalarında verimliliğini önemli ölçüde artırmaktadır. CMC değerlendirme metriği, yalnızca Rank-1'i hesaplamakla kalmaz, aynı zamanda diğer en üst sıralarda doğru eşleştirilen sorguları da belirleyebilir. Örneğin Rank-1, -5 ve -10 gibi belirli sıralamaların oranları kullanılmaktadır. Şekil 2.22, CMC eğrisinin Rank-1, -5 ve -10'a dönüştürülmesini göstermektedir. CMC, her sorgu için yalnızca 1 doğru sonuç varsayımına dayanmaktadır. Bu nedenle, doğru bir değerlendirme yöntemi olarak kabul edilmektedir. Ancak galeri birden fazla doğru sonuç içeriyorsa, genel performansın değerlendirilmesinde mAP metriği kullanılabilir. mAP, sorguların doğru sonuçlara olan ortalama hassasiyetini hesaplayarak birden fazla doğru sonuç durumunda daha kapsamlı bir değerlendirme sağlamaktadır.



Şekil 2.22 CMC eğrisinin Rank 1, 5 ve 10'a dönüştürülmesi.

- **Ortalama Hassasiyet (mAP):** mAP, arama algoritmalarını karşılaştırmak için standart bir ölçüdür. Birden fazla doğru sonucun olduğu durumlarda galerideki tüm eşleşen örnekleri bulmak için kişi Re-ID algoritmalarının performansının değerlendirilmesine izin vermek amacıyla tanıtılmıştır [19]. Her bir sorgu görüntüsü için AP, hassasiyet-geri çağırma eğrisinin altındaki alan (the area under its precision-recall curve) olarak hesaplanmaktadır. Bir sorgu görüntüsü verildiğinde, AP aşağıdaki denklemde belirtildiği gibi tanımlanabilir:

$$AP = \frac{\sum_{k=1}^n P(k) \times G(k)}{N_{gt}} \quad (2.17)$$

Burada n , test görüntülerin sayısıdır, $P(k)$, sıralama listesinde kesim k 'teki kesinliği temsil etmektedir. $G(k)$ k 'inci eşleşme doğru ise 1'e eşittir; aksi halde 0'ya eşittir ve

N_{gt} doğru değerlerin sayısıdır. Dolayısıyla mAP, tüm sorguların AP'lerinin ortalamasıdır [19]. mAP, aşağıdaki denklemde belirtildiği gibi tüm sorgular üzerinde hesaplanabilmektedir:

$$mAP = \frac{\sum_{q=1}^Q AP(q)}{Q} \quad (2.18)$$

Burada Q, sorgu sayısını temsil etmektedir. Her bir sorgu görüntüsü için eşleşen birden fazla görüntü içeren veri kümeleri varsa, mAP'ler CMC eğrisiyle birlikte değerlendirilerek daha iyi bir performans değerlendirmesi yapılabilmektedir. Mevcut kişi yeniden tanımlama yöntemlerinin performansları genellikle geri alma değerlendirmesi (recall evaluation) için mAP ve geri alım hassasiyeti değerlendirmesi (retrieval precision evaluation) için CMC eğrisinin birleştirilerek incelenmektedir.

2.8 Bölüm Özeti

Bu bölümde, kişi Re-ID görevinin temel kavramlarının bir tanımı sunulmuştur. Bu tanımda kişi Re-ID görevine ilişkin araştırmaların bir özeti sunulmuş ve bu araştırmalar el yapımı yöntemlerden yeni derin öğrenme yöntemlerine kadar geniş bir alanda ele alınmıştır. Ayrıca kişi Re-ID için kullanılan derin öğrenme teknikleri, özellikle CNN'ler, transfer öğrenimi ve Transformer'ler gibi önemli teknikler detaylı bir şekilde tanıtılmıştır. Son olarak, kişi Re-ID yöntemlerinin değerlendirildiği halka açık veri kümeleri ve bu yöntemlerin değerlendirilmesi için kullanılan metriklerin önemi vurgulanmıştır.

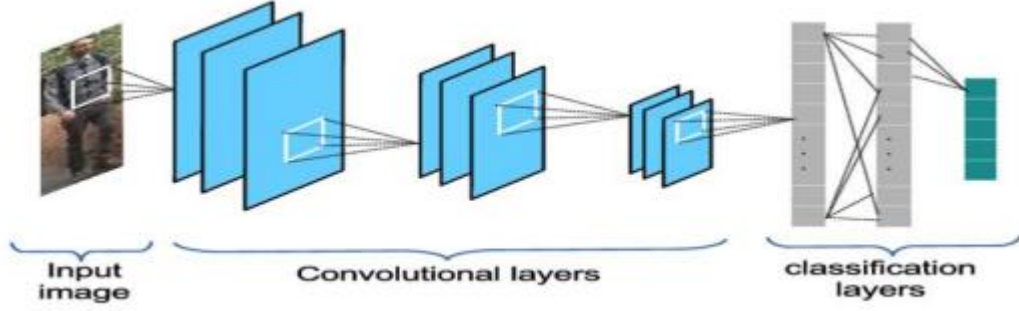
DERİN ÖĞRENMEYE DAYALI KİŞİ KİMLİĞİNİN YENİDEN BELİRLEME YÖNTEMLERİNİN TAKSONOMİSİ

Derin öğrenme tabanlı yöntemler, bilgisayarlı görü işlemlerinde başarılı sonuçlar elde etmesi nedeni ile kişi yeniden tanımlama konularında da sıkça kullanılmaktadır. Bu yöntemler, görüntülerdeki yayaların özniteliklerini adaptif bir şekilde öğrenme ve iyi sonuçlarla benzerlik metriklerini öğrenme yeteneğine sahiptirler [106]. Derin öğrenme gelişimiyle birlikte, kişi Re-ID performansını artırmak amacıyla mevcut derin öğrenme mimarilerinin değiştirilmesi veya yeni derin modellerin tasarlanması yoluyla farklı derin öğrenme tabanlı modeller önerilmiştir. Derin öğrenme model türüne bağlı olarak kişi Re-ID için tasarlanmış beş temel derin öğrenme tabanlı yöntem bulunmaktadır; tanımlama modeline dayalı yöntemler, doğrulama modeline dayalı yöntemler, mesafe metriğine dayalı yöntemler, vücut parçasına dayalı yöntemler ve Transformer modeli. Bu yöntemler yaygın olarak kullanılmakta ve kişi Re-ID topluluğunda geleceğin trendlerinin bir göstergesi olabilmektedir.

Bu bölümde kişinin yeniden tanımlama görevine yönelik ilgili literatürün kapsamlı bir incelemesi yapılmıştır. Ayrıca takım sporlarında oyuncuların yeniden kimliklendirmesi için yapılan çalışmalarda incelenmiştir. Bu inceleme, kişi Re-ID alanında gerçekleştirilen araştırmaların özetini içermektedir ve elde edilen sonuçlar değerlendirilmiştir. Yöntemler derin öğrenme tekniklerinin kullanıldığı farklı kategorilere ayrılarak analiz edilmiştir. Bu kategorizasyon, kullanılan derin öğrenme yönteminin türüne, yöntemlerin avantajlarına ve dezavantajlarına dayanarak yapılmıştır.

3.1 Tanımlama Modeline Dayalı Yöntemler

Burada kişinin yeniden tanımlama işlemi bir sınıflandırma sorunu olarak ele alan derin öğrenme tabanlı yöntemler incelenmektedir [3]. Bu yöntemlerde derin model bir kişinin görüntüsünü alır, ardından girdi kişi görüntülerine karşılık gelen sınıfın olasılığını hesaplayarak çıktı olarak ilgili etiketi vermektedir [1]. Şekil 3.1, tanımlama modelinin genel derin mimarisini görselleştirmektedir.



Şekil 3.1 Derin öğrenme tabanlı tanımlama modelinin temel mimarisi [1].

Literatürdeki birçok çalışma, kişi Re-ID görevini ele almak için kimlik tanımlama derin modellerini kullanmaktadır. Wu ve diğ. [41], kişi yeniden tanımlama için öznetelik çıkarma için birleşik bir öznetelik ağı (FFN) önermiştir. Bu model CNN özneteliklerini, renk histogramı ve doku gibi el yapımı öznetelikleri içeren öznetelikleri bir arada kullanmaktadır. Her iki öznetelik elde edildikten sonra bir buffer katmanına girip ardından bir FC katmanı kullanılmıştır. Önemli olan buffer katmanı, çok farklı olan bu iki öznetelik arasındaki boşluğu kapatmak için birleştirme işleminde kullanılmıştır. Ardından el yapımı özneteliklerin çeşitliliği nedeniyle, CNN öznetelikleri geri yayılım aşamasında kısıtlanır. Tüm ağın eğitimi, SoftMax kaybı kullanılarak gerçekleştirilir. Kişi Re-ID için geniş görünüm farklılıklarına sahip kişi örnekleri için doğrusal olarak ayrılabilir gizli temsiller üretmek amacıyla [107]'de bir hibrit derin mimari önerilmiştir. Bu mimari, kişi görüntülerini temsil edebilen doğrusal olmayan öznetelikleri üretmek için bir Fisher vektörü ile birleştirilmiş derin bir CNN modelinden oluşmaktadır. Ağ, bir LDA tabanlı amaç fonksiyonu optimize edilerek uçtan uca şekilde eğitilmektedir. Bu optimize edilmiş fonksiyon sayesinde, LDA tabanlı gradyanlar geri yayılarak Fisher vektörü kodlama parametrelerini güncellemektedir. Böylece görünüm farklılıklarına sahip kişi örnekleri için doğru temsiller oluşturulmaktadır. Transfer öğreniminden ilham alan Zheng ve diğerleri [9] ve [89], ImageNet [72] üzerinde önceden eğitilmiş derin CNN modellerinden doğrudan faydalanmışlardır. Bunları IDE olarak bilinen bir yaklaşım aracılığıyla fine-tuning yaparak geliştirdiler. IDE, kişi Re-ID eğitim sürecini çoklu sınıflandırma görevine benzetmektedir [108]. [89] çalışmasında elde edilen IDE sonuçları, bu yaklaşımın kişi Re-ID performansını artırmak için büyük bir potansiyele sahip olduğunu göstermektedir. Bu nedenle, günümüzde en çok kullanılan DL tabanlı yöntemler de genellikle önceden eğitilmiş modelleri temel olarak kullanmakta ve kişi Re-ID sistemlerinin performansını artırmak için yeni teknolojileri araştırmaktadır [109]. Xiao ve diğer araştırmacılar

tarafından yapılan bir çalışmada [42], çok çeşitli veri setlerinden genel derin öznelik temsilleri öğrenmek ve her eğitim veri seti için etkili nöronları keşfetmek amacıyla bir CNN modeli önerilmiştir. Bu çalışmada, birden fazla kişi Re-ID veri setinden veri ve etiketleri birleştirerek güçlü bir temel model oluşturulmuştur. Model, SoftMax kaybı ile eğitilmiştir. Ardından her veri seti için tüm örnekler üzerinde bir ileri geçiş gerçekleştirilmiştir. Bu geçişler sırasında her nöronun hedef fonksiyon üzerindeki etkisi ortalama olarak hesaplanmıştır. Son olarak, gereksiz nöronları kaldırmak için standart dropout işlemi yerine bir deterministic domain guided Dropout kullanılarak CNN modelini birkaç tekrar daha eğitilmiştirler. [110] çalışmasında derin temsil öğrenme sürecinin tekil vektör ayrışımı (SVD) ile optimize edilebileceği önerilmiştir. Öznelikle kısıtlama ve gevşeme iterasyonu (RRI) eğitim şemasını kullanarak, CNN'nin eğitiminde ortogonalite kısıtını iteratif olarak entegre etmeyi başarmışlardır. Bu da SVDNet olarak bilinen bir yöntem ortaya çıkarmıştır. SVDNet'in kişi Re-ID veri kümelerine ilişkin deneysel analizinden elde edilen sonuçlar, önemli bir performans artışıyla birlikte rekabetçi bir Re-ID doğruluğuna ulaşıldığını göstermektedir. Ardından Zheng ve diğerleri tarafından yayınlanan [111] çalışmada, yayaların hizalamasını öğrenen bir kimlik tanımlama görevi önerilmiştir. Bu çalışmada, kişi tanımlarını öğrenebilen ve aynı zamanda onları sınırlayıcı bir kutuyla eşleştirebilen bir yaya hizalama ağı (PAN) tasarlanmıştır.

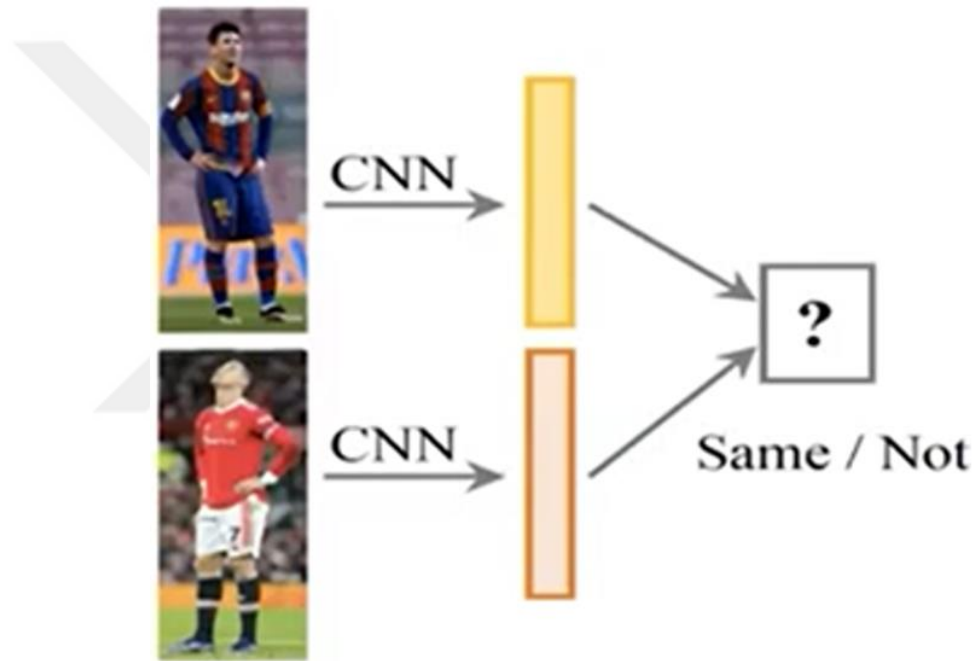
Gerçek yaşam senaryoları ile kişi Re-ID veri kümeleri arasındaki boşluğu doldurmak için Xiao ve ark. [112] tarafından kişi algılama ve Re-ID görevlerini uçtan uca bir şekilde ele alabilen bir derin ağ mimarisi önerilmiştir. Bu çalışmada, derin mimarinin eğitiminde kullanılan SoftMax kaybı için dengesiz ve seyrek etiketlere yönelik rastgele örnekleme yöntemini tanıtılmıştır. Ardından Xiao ve ekibi [113], ortak öğrenme sistemine yönelik ağın eğitiminde yaygın olarak kullanılan SoftMax kaybının yerine geçebilecek bir çevrimiçi örnek eşleştirme (OIM) kayıp fonksiyonu tasarlanmışlar. OIM, SoftMax kaybıyla karşılaştırıldığında parametrik olmayan bir yöntemdir. Bu sayede bir dairesel sıra kullanarak etiketsiz bir yayayı kullanma yeteneği elde edebilir. Ancak OIM kaybının parametrik olmaması nedeniyle aşırı uyuma eğilimi olduğu için bir dezavantaja sahiptir.

Genel olarak kimlik tanımlama modelinin girişi kolaylıkla gerçekleştirilebilir ve veri kümeleri içindeki etiket bilgilerini tam olarak kullanabilmektedir [9]. Bu nedenle, eğitim sürecinin verimliliğini artırmasına yardımcı olmaktadır. Ancak kimlik tanımlama modelinin eğitim amacı test yöntemiyle uyumlu olmadığı için sonuçların doğruluğunu

etkilemektedir. Ayrıca mevcut kişi Re-ID veri kümelerinde ölçekler çok küçük olduğu için kimlik tanımlama modelinin tamamen eğitilememesine neden olmaktadır.

3.2 Doğrulama Modeline Dayalı Yöntemler

Bu bölümde, kişi yeniden tanımlamasını ikili sınıflandırma problemleri olarak ele alan derin öğrenme tabanlı yöntemler tartışılmıştır [34], [39], [40]. Bu yöntemlerde derin model çift kişi görüntülerini giriş olarak kullanmaktadır. Ardından çift görüntülerin aynı kişiye ait olup olmadığını belirlemek için bir benzerlik değeri üretilmektedir [3]. Temel doğrulama derin model mimarisi Şekil Şekil 3.2'de gösterilmektedir. Literatürde birçok çalışma, kişi Re-ID işlemini ele almak için doğrulama derin modellerini kullanmışlardır.



Şekil 3.2 Derin öğrenme tabanlı doğrulama modelinin temel mimarisi.

Li ve diğerleri [39] tarafından, kişi Re-ID problemiyle başa çıkmak için doğrulama modelini kullanma konusunda ilk çalışmayı sunarak FPNN adlı bir ağ mimarisi önerilmiştir. FPNN, iki farklı giriş görüntüsünden öğrenilen iki öz nitelik vektörü karşılaştıran bir yama eşleştirme katmanını içermektedir ve ardından bir dizi yer değiştirme matrisi hesaplanmaktadır. Bu katman, en büyük aktivasyonlara sahip yer değiştirme matrislerini bir sonraki katmanlara aktarır. Sonraki katmanlar, Conv ve maksimum havuzlama katmanlarından oluşmaktadır. Ardından bir FC katmanı bulunmaktadır. Bu sayede giriş görüntülerinin her birine son öz nitelik vektörü elde edilmektedir. Son katman, iki giriş görüntüsünün aynı kimliği temsil edip etmediğini

belirlemek için SoftMax fonksiyonu kullanılmaktadır. Bu ağ, karmaşık geometrik ve fotometrik dönüşümlerin bir karışımını modellemek için geniş bir öğrenme kapasitesine sahiptir. FPNN ağı Ahmed ve arkadaşları [34] tarafından genişletilmiştir. [34] çalışmasında, çapraz giriş komşuluk farkı özniteliklerinin öğrenilmesi fikrini önerdiler. Bu, birinci giriş resminde belirli bir bölgeden elde edilen özniteliklerin, diğer giriş resmindeki yakın bölgelerden elde edilen özniteliklerle karşılaştırılmasıyla gerçekleştirilmektedir. Bu katmanın çıktısı, iki giriş resminin çıkarılan öznitelikleri arasındaki yaklaşık ilişkiyi sağlamaktadır. Ayrıca komşuluk farkı özniteliklerini daha küçük bir öznitelik açıklamasına özetleyen ek bir yenilikçi katman önerilmiştir. Bu katman, bir Conv ve maksimum havuzlama katmanından geçirilmiştir. Son olarak iki FC katman ve bir SoftMax fonksiyonu kullanılarak, iki giriş resminin aynı kimliği temsil edip etmediği belirlenmektedir. Daha sonra, [40] çalışmasında yukarıdaki bu iki çalışmaya dayalı olarak Wu ve diğerleri tarafından, kişi Re-ID için PersonNet modelini önerilmiştir. Model, yamalar arasındaki yerel ilişkiyi yakalamak için Köstinger ve diğerleri [28] tarafından önerilen yama eşleştirme katmanını kullanmıştır. Ayrıca daha küçük Conv filtreleri kullanarak mimarinin derinliğini artırmaya yardımcı olmuştur. Yukarıdaki doğrulama modellerinin dezavantajı ağ derinliklerinin oldukça sık olmasıdır, bu da görüntüleri ayırt edebilen derin öznitelikleri araştırırken pek fayda sağlamamaktadır. Bu doğrulama modellerinin bir diğer dezavantajı ise çift görüntülerin oluşturulması gerektiği ve yalnızca zayıf veri seti etiketlerinin kullanılabildiğidir [9]. Bu durum eğitim verimliliğini azaltmaktadır.

Aynı ağ mimarisi içerisinde hem kimlik kaybını hem de doğrulama kaybını birleştirmek mümkündür. Bu ağlar, bir kişi kimliğinin çoklu sınıf tanıma için uygun olan özniteliklerini öğrenirken aynı zamanda bir çift giriş görüntünün aynı kişiye ait olup olmadığını belirleyebilirler. Tek bir modelde kimlik tespiti ve doğrulamanın birleştirilmesi, kişi Re-ID için umut verici sonuçlara yol açmıştır. Chen ve ark. [114] tarafından yapılan bir çalışmada, iki kaybın bir kombinasyonunu kullanan bir ağ mimarisi önerilmiştir. Bu kayıplardan biri kimlik sınıflandırma kaybı, diğeri ise eşleşen bir kişi için öznitelik temsilini öğrenmeye çalışan doğrulama kaybıdır. Bir görüntü çiftinden elde edilen öznitelik haritalarına aynı dropout haritasının uygulandığından emin olmak için bir özel kayıp dropout birimi (loss-specific dropout unit) önerilmiştir. Kayıp özel dropout birimini ardından, elde edilen öznitelik haritası çifti bir doğrulama kaybı katmanına iletilir ve çiftin her bir öznitelik haritası ayrı ayrı bir kimlik kaybı katmanına iletilmektedir. Kısa

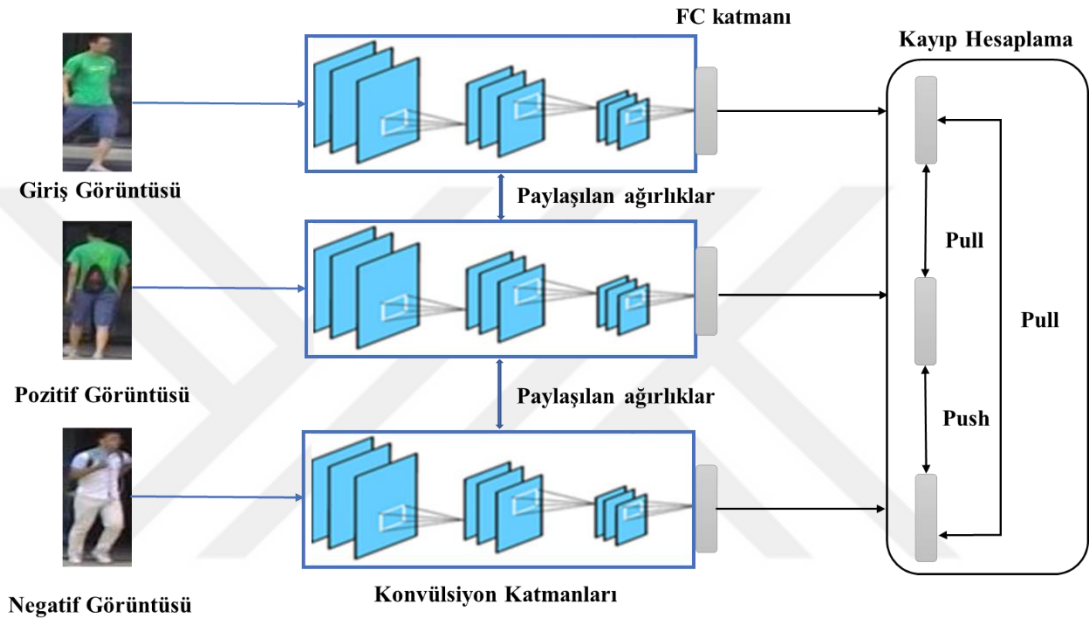
süre sonra, Zheng ve ark. [43], kişi Re-ID görevleri için daha ayırt edici yayaların özniteliklerini öğrenmek amacıyla doğrulama ve kimlik kayıplarını birleştiren bir doğrulama-kimlik tanımlama ağı önermiştir. Önerilen ağ yapısı bir giriş görüntü çifti verildiğinde, her bir görüntünün ayrı bir ağ dalından geçirilmesini sağlamaktadır. Bu iki dal, ağırlıkları paylaşan önceden eğitilmiş CNN modelleridir ve iki görüntü çiftinin kimlik etiketlerini eş zamanlı olarak tahmin etmektedir. Her bir dalda kimlik sınıflandırma kaybı katmanı kullanılarak giriş görüntüsünün kimliği SoftMax fonksiyonuyla tahmin edilmektedir. Her iki dalın nihai öznitelik haritaları, önerilen kare katmanı kullanılarak karşılaştırılmaktadır. Kare katmanı, iki öznitelik haritasını eleman düzeyinde çıkartarak ve kare işlemi uygulayarak benzerliklerini hesaplamaktadır. Son olarak çıktı öznitelik haritası doğrulama kaybı katmanına girdi olarak kullanılmakta ve SoftMax kaybı fonksiyonu kullanılarak iki giriş görüntüsünün aynı kimliği temsil edip etmediği tahmin edilmektedir.

Daha sonra Zheng ve arkadaşları [44], doğrulama kaybı ve kimlik kaybını birleştiren bir mimari önermişler. İlk olarak bir görüntü çiftinin her bir görüntüsü, öznitelik birleştirme ağı (FAN) olarak adlandırılan bir öznitelik çıkarma ağına yönlendirilmektedir. FAN, ResNet-50 [71] ağ mimarisine dayanarak farklı katmanlardan öznitelik haritaları çıkararak ve bunları eleman düzeyinde toplayarak birleştirilmiş bir öznitelik haritası oluşturmaktadır. Sonrasında doğrulama dalı için öznitelik haritaları, tekrarlayan karşılaştırmalı ağ (RCN) olarak adlandırılan bir ağa iletilmektedir. RCN, dikkat ve tekrarlayan ağlar kullanarak bir görüntü çiftini oluşturan görüntülerin görünümünü karşılaştırmaktadır. Dikkat bileşeni, öznitelik haritasının ayırt edici önemli bölgelerine ağırlıklar uygularken, tekrarlayan ağlar görüntü çiftinin ayırt edici önemli bölgelerini birleştirir. RCN çıktısı daha sonra doğrulama kaybı katmanına giriş olarak kullanarak iki giriş görüntüsünün aynı kişiye ait olup olmadığını hesaplanmaktadır. Kimliklendirme dalı için ise FAN'dan çıkarılan öznitelik haritaları, küresel ortalama havuzlama işlemi gerçekleştirilerek SoftMax kaybı ile donatılmış bir FC katmana yönlendirilmektedir. Bu katmanın çıktısı, görüntünün her bir sınıfa ait olma olasılığını tahmin etmektedir.

3.3 Mesafe-Metrik Tabanlı Modele Dayalı Yöntemler

Derin öğrenme tabanlı mesafe ölçütü modelleri, aynı kişiye ait görüntüler arasındaki mesafeleri minimize etmeyi ve farklı kişilere ait görüntüler arasındaki mesafeleri maksimize etmeyi hedeflemektedir. Bu amaçla yaygın olarak kullanılan yaklaşım üçlü tabanlı derin mimaridir. Üçlü kayıp yöntemleri, giriş görüntüsü, aynı kimliğe sahip pozitif

eşleşme görüntüsü ve farklı kimliğe sahip negatif eşleşme görüntüsü olmak üzere üç görüntüyü giriş olarak almaktadır. Ağ, çıktıları işleyerek giriş görüntüsünü pozitif eşleşme görüntüsüne yakınlaştırmak için aralarındaki mesafeyi çıktı öznetelik uzayında minimize etmektedir. Aynı zamanda, ağ giriş ve pozitif eşleşme görüntülerini negatif eşleşme görüntüsünden uzaklaştırarak çıktı öznetelik uzayında farklı kimlikler arasındaki mesafeyi maksimize etmektedir. Üçlü modelin temel derin mimarisi Şekil 3.3'te gösterilmiştir.



Şekil 3.3 Üçlü derin model mimarisi

Üçlü kayıp kullanarak bir ağı eğitilmesi, Re-ID alanında faydalıdır. Bu nedenle, birçok yöntem bu yaklaşımı benimsemiştir. Ding ve ark. [115], kişi Re-ID sorununu çözmek için üçlü modeli kullanan ilk araştırmacılarıdır. İlgili mesafe (relative distance) karşılaştırması kullanarak, ölçeklenebilir derin öznetelik öğrenme modeli önermişlerdir. Modelde CNN ağı, aynı üçlü set tarafından düzenlenen ilgili mesafe kısıtlamalarını karşılayabilen öznetelikleri üretebilmek için bir dizi üçlü ile eğitilmektedir. Üçlü sayısının üçüncü derece büyümesi sorunu çözmek için araştırmacılar ağı yinelenen ve verimli bir şekilde eğitebilen genişletilmiş bir ağ yayılım algoritması sunmuşlardır. Bu öğrenme algoritması, genel hesaplama yükünün çoğunlukla üçlü sayısından değil eğitim görüntülerinin sayısından etkileneceğini sağlayabilmektedir.

Ardından Cheng ve diğerleri [36], kişi Re-ID için geliştirilmiş bir üçlü yöntem önerdiler. Bu yöntemde, global görüntü öznetelikleri ve yerel vücut bölgesi görüntü öznetelikleri

birleştirilerek üçlü kayıp fonksiyonuna iletilmektedir. Ayrıca tipik üçlü kayıp eğitim stratejisi, aynı kişiye ait görüntülerden üretilen derin özniteliklerin ne kadar yakın olması gerektiğini belirleyememektedir. Bu nedenle, aynı kişiye ait olduğu görüntüler arasındaki iç sınıf mesafesi gereğinden büyük olmasına yol açmıştır. Bu sebeple bu çalışmada üçlü kayıp fonksiyonuna giriş görüntüsü ile pozitif giriş arasındaki mesafenin belirli bir eşik değerin altında olmasını sağlayan bir kısıtlama eklediler. Bu geliştirilmiş üçlü kayıp fonksiyonu, farklı kişi görüntülerini birbirinden daha da uzaklaştırırken aynı kişi görüntülerini yakınlaştırmaya yardımcı olmaktadır. Böylelikle, yeni öğrenilen derin öznitelik uzayı kullanılarak eşleştirme oranları %4'e kadar artırılabilir. Hermans ve diğerleri [116] tarafından önerilen bir çalışmada, minibatch içerisinde zorlu madencilik stratejisi (hard mining strategy) kullanılarak bir batch tabanlı üçlü kayıp yöntemi sunulmuştur. Bu yöntemde, rastgele olarak seçilen P kişi kimliği belirlenmiştir. Her bir P sınıfından K görüntü örnekleme yapılmıştır. Böylelikle minibatch içerisinde toplam $P \times K$ görüntü yer almaktadır. Son aşamada ise en zor negatif ve pozitif örnekler bir minibatch içerisinden seçilerek üçlü birimler oluşturulmuştur. Zhao ve diğerleri [117], ağırlıklı farklı katmanlarından alınan öznitelik haritalarının birleştirilmesiyle bir öznitelik tanımlayıcısı oluşturma yöntemi önermişlerdir. Bu yöntemde, sık katman bilgisi ile derin katman bilgisi birleştirilerek her bileşenin farklı görsel öznitelikleri yakalamasına olanak sağlanmaktadır. Dolayısıyla nihai öznitelik tanımlayıcısını oluşturulmaktadır. Bu çok katmanlı öznitelikler daha sonra üçlü kayıp eğitim stratejisinin bir parçası olarak kullanılmaktadır. Üçlü tabanlı derin modelin dezavantajı, yalnızca Re-ID veri kümesinden gelen zayıf etiket bilgisini kullanmasıdır [9]. Ayrıca modelin girdi olarak üçlü birimlerin üretilmesi ihtiyacı duymaktadır çünkü bunun yapılmaması eğitim sürecinin verimliliğini düşürecektir. Bazı yaklaşımlar, üçlü ve tanımlama kayıplarının [118]–[120] veya üçlü ve doğrulama kayıplarının [35], [121] birleştirilmesi yoluyla derin mimariyi optimize etmek için kullanılmıştır.

3.4 Pozisyon Yönlendirmeli ve Parça Tabanlı Modele Dayalı Yöntemler

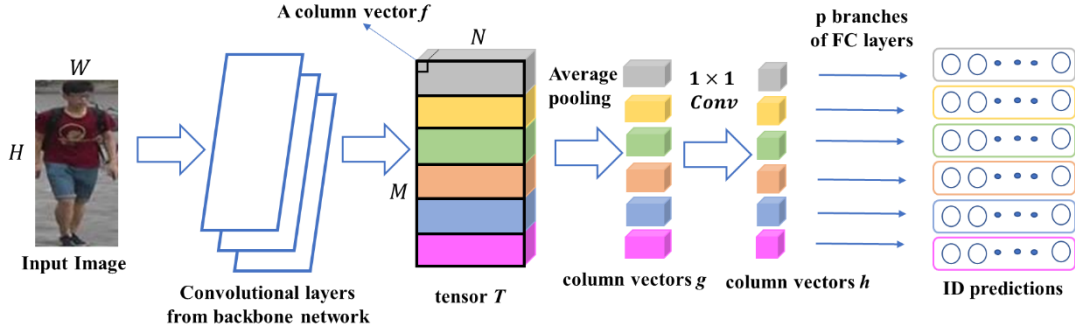
Parça tabanlı derin model, insan vücut parçalarından yerel öznitelik temsillerini öğrenmeyi hedefleyen bir yöntemdir. Bu model, global tam vücut temsilleri yerine yerel bilgiyi kullanarak yayaların tanınmasını amaçlamaktadır. Kişi görüntülerini açıklarken bölüm düzeyindeki özniteliklerin kullanılması, daha ayrıntılı bilgi sağlayarak kişi Re-ID için faydalı olabilmektedir [122]. Yerel öznitelik temsilleri genellikle parçaları veya

bölgeleri birleştirilmiş öznitelikler olarak öğrenmektedir ve bu da onları hizalama farklılıklarına karşı dirençli hale getirmektedir [123], [124]. Vücut parçaları, insan poz tahmini veya yaklaşık yatay bölünme gibi yöntemlerle elde edilebilmektedir [125].

Poz yer işaretleri, bir kişinin vücudunun nasıl konumlandığını göstermektedir. Bazı kişi Re-ID yöntemleri, poz yer işaretlerini kullanarak kişi Re-ID modellerine yardımcı olmaktadır [126]–[130]. Bu yöntemler, poz yer işaretlerine dayanarak kişi görüntülerini oluşturmak [126], [129] veya vücut kısımlarını hizalamak [128], [130] için kullanılmaktadır. Bu yöntemler, kişinin duruşuna veya oturma şekline bağlıdır. Ayrıca [131], [132] çalışmalarında, kapanma problemleriyle başa çıkmak için poz bilgisinden faydalanılmıştır. [132] çalışmasında ise, poz yer işaretlerini kullanarak vücut bölümlerinin konum bilgisini kodlayıp sorgu ve galeri görüntüleri arasında paylaşılan bölge parçalarını karşılaştırmaktadır.

Literatürde birçok çalışma, kişi Re-ID görevi için ayırt edici derin öznitelikler öğrenmek amacıyla bölüm tabanlı stratejiyi benimsemiştir. Cheng ve ark. [36], küresel Conv katmanını kullanarak küresel Conv özniteliklerini elde etmiştir. Ardından öznitelikleri eşit ve dört ayrı dal halinde bölmüşler, böylece bölüm tabanlı derin öznitelikleri elde etmişlerdir. Son adımda, küresel ve bölüm tabanlı öznitelik vektörlerini birleştirerek nihai derin öznitelikleri elde etmişlerdir. Benzer şekilde, [133] çalışmasında Li ve ark. de bölümlendirme stratejisini kullanmıştır. Ancak, tasarladıkları ağı optimize etmek için çoklu sınıflandırma kayıplarından faydalanmışlardır.

Sun ve diğerleri [124], Şekil 3.4'te gösterilen "parça tabanlı Conv temel modeli (PCB)" adlı bir derin model geliştirmiştir. Bu model, Conv katmanında eşit parçalara bölmeler yaparak parça düzeyindeki öznitelikleri öğrenebilmektedir. Girdi olarak tek bir görüntü kullanan bu model, 6 parça düzeyinde öznitelikten oluşan bir tanımlayıcı çıktı üretebilmektedir. Ayrıca, "iyileştirilmiş parça birleştirme (RPP)" yaklaşımını önermişlerdir. Bu yaklaşım homojen bölme içindeki her parçanın aykırı değerlerini en yakın oldukları parçalara yeniden atayarak daha gelişmiş bölge bölmeleri ve parça içi tutarlılık sağlamaktadır. Bu model, popüler bazı genel veri kümelerinde oldukça rekabetçi bir performans sergilemektedir.



Şekil 3.4 PCB derin modelinin yapısı [124].

Li ve diğerleri [38] çalışmalarında, yerel ve global vücut parçalarının özniteliklerini kullanmak için bağlamsal bilinçli çok ölçekli bir ağ tasarlamışlardır. Uygulama açısından önceden tanımlanmış bölüm parçaları yerine, gizli (latent) kişi parçalarını yerelleştirmek için mekânsal dönüşüm ağlarının kullanılmasını önermişlerdir. Son adımda ise tam vücut ile vücut parçalarını birleştirerek nihai temsili oluşturmuşlardır. [118], [134] çalışmalarında ise derin özniteliklerin ayırt edici yeteneklerini artırmak için dikkat mekanizmasını bir öznitelik öğrenme modülü olarak kullanmışlardır.

Li ve diğerleri [135], kişi Re-ID'de dikkat seçimi ve öznitelik temsillerinin ortak öğrenilmesi için yeni bir harmonik dikkat CNN (HACNN) modeli önermiştir. Yapılan kapsamlı karşılaştırmalı değerlendirmeler, HACNN modelinin birçok state-of-the-art yöntemle göre kişi Re-ID'de üstünlüğünü kanıtlamıştır. Yerel görsel ipuçları genellikle insan görsel alışkanlıklarına uygun ve global bilgiyle birlikte oldukça tamamlayıcıdır. Bu nedenle, global ve yerel özniteliklerin birleşimi kişi Re-ID için iyi bir seçimdir. Ancak parça tabanlı derin yöntemin bazı sınırlamaları vardır. Örneğin parça tabanlı dalların derin modele eklenmesi, modelin karmaşıklığını artırabilir ve eğitim verimliliğini azaltabilir. Ayrıca dikkat tabanlı derin modeller genellikle yalnızca bölge düzeyinde dikkate alınırken, piksel düzeyindeki belirginlik göz ardı edilmektedir. Bu nedenle, bu yöntemler küçük, etiketli eğitim veri setleri veya arka plan karmaşıklığına ve hizalamaya sahip gürültülü yaya görüntüleri ile başa çıkma konusunda etkisiz olabilir. Ayrıca ayırt edici temsiller için mekânsal bağlamsal veriler önemli bir unsurdur ancak mevcut yöntemlerin çoğu farklı parça tabanlı öznitelikler arasındaki mekânsal bağlam bilgisini dikkate almamaktadır.

3.5 Trensformer'ler Modeline Dayalı Yöntemler

Transformer modeli, sıralı verilerin işlenmesi için önerilen bir yöntem olup, başlangıçta doğal dil işleme (NLP) alanında kullanılmıştır [80]. Ancak daha sonra bilgisayarlı görü görevlerinde de etkinliği gösterilmiştir. Örneğin [136], [137] çalışmaları, Transformer'ın bilgisayarlı görü alanında nasıl uygulanabileceğini incelemiştir. Saf Transformer modelleri giderek daha popüler hale gelmektedir. IPT [138] örneğinde olduğu gibi büyük ölçekli ön eğitim Transformer'ı kullanarak yüksek performans elde edilebilmektedir. Bu da süper çözünürlük, gürültü giderme gibi çeşitli görüntü işleme görevlerinde son teknoloji çözümleri sunmaktadır. Son zamanlarda görüntü yamaları için doğrudan bir saf Transformer uygulayan ViT [84] önerilmiştir. Ancak ViT modeli önceden eğitmek için büyük bir veri kümesi gerektirmektedir. Bu eksikliği aşmak için Touvron ve diğerleri [85], DeiT adlı bir çerçeve sunarak, büyük ölçekli ön eğitim verilerine ihtiyaç duymadan ViT eğitimini hızlandırmayı amaçlamaktadır.

CNN tabanlı yöntemler, Re-ID çalışmalarında uzun yıllardır baskın bir rol oynamaktadır. Ancak saf transformer modelleri giderek daha popüler hale gelmektedir. He ve arkadaşları [139], kişi ve araç Re-ID görevlerinde son teknoloji performansı sağlayan TransReID adlı bir yöntem önererek, ViT'leri başarılı bir şekilde Re-ID görevlerine uygulayan ilk kişiler olmuşlardır. AAformer [140] ise parça temsillerini öğrenmek için ViT omurgasını kullanırken, ek öğrenilebilir "parça belirteçleri (part tokens)" vektörlerini kullanarak parça hizalamasını öz-dikkat mekanizmasıyla birleştirmişler. Bazı çalışmalar özniteliklerin veya CNN omurgalarından gelen bilgilerin toplanması için Transformer'ı kullanmaya çalışmışlardır. Örneğin [141]–[143] çalışmalarında, hiyerarşik öznitelikleri toplamak ve yerel öznitelikleri hizalamak için Transformer katmanları CNN omurgasına entegre edilmiştir. Video Re-ID için ise [144], [145] gibi çalışmalarda, görünüş, uzamsal ve zamansal öznitelikleri birleştirmek ve ayırt edici bir temsil öğrenmek için Transformer kullanılmıştır.

Transformer tabanlı yöntemler [139], [143], giderek daha fazla dikkat çekmiş ve kişi ReID'de büyük performans elde etmiştir. Kişi Re-ID alanında giderek daha fazla ilgi görmekte ve önemli bir performans sergilemektedir. Örneğin saf transformer tabanlı yöntem olan TransReID [139], mevcut CNN tabanlı yöntemlere kıyasla önemli bir performans artışı sağlamaktadır. Bununla birlikte ImageNet ve kişi ReID veri kümeleri arasında büyük bir alan boşluğu bulunmaktadır çünkü 1) ImageNet ve ReID veri kümelerinin görüntü içeriği çok farklıdır [146]; 2) ImageNet üzerinde denetimli ön

eğitim, zengin görsel bilgileri azaltan kategori düzeyinde denetime odaklanırken, kişi Re-ID'si ayrıntılı kimlik bilgilerini tercih etmektedir [147]. Bu nedenle, Transformer'lerin ön eğitim veri kümesine aşırı uyum sorununu önlemek için daha büyük ölçekli bir ImageNet-21K [72] veri kümesi üzerinde ön eğitime tabi tutulması gerekmektedir. Daha iyi transformer tabanlı Re-ID modelleri için ön eğitim ve ince ayar veri kümeleri arasındaki boşluğu kapatmak amacıyla TransReID-SSL [148] adlı çalışması, veri ve model yapısı perspektiflerinden soruna yaklaşmaktadır. Kişi Re-ID'nin mevcut ön eğitim paradigmalarının transformer tabanlı omurgalar için iyi performans gösteremediğini gözlemledikten sonra, LUPerson [146] üzerinde ViT'lerle [84] yapılan kendi kendine denetimli öğrenme (SSL) [149] yöntemlerini araştırarak DINO'nun [147] transformer tabanlı Re-ID için en uygun ön eğitim yöntemi olduğunu bulmuşlardır. Daha sonra ön eğitim ve Re-ID veri kümeleri arasındaki boşluğu daha da azaltmak için CFS'ye dayalı koşullu bir ön eğitim yöntemi önermişler ve bu yöntem performans düşüşü olmadan ön eğitimi hızlandığını ifade etmektedir. Son olarak performansı daha iyileştirmek için bir yama gömme olan IBN tabanlı evrişim çekirdeği önermişler. Bu çalışmanın dezavantajı olarak önceden eğitilmiş modeller yalnızca kişi Re-ID'si için uygun olmasıdır. Ancak araç Re-ID'si, insan ayrıştırması veya görüntü sınıflandırması gibi görevlerde iyi performans gösteremez.

Kişi Re-ID görevleri için ayrımcı öznitelik temsillerini çıkarmak önemlidir. Evrişimli sinir ağlarının güçlü öznitelik çıkarma yeteneklerine sahip olmasına rağmen, lokal olmayan ilişkileri modellemek hala zorluk oluşturmaktadır. Gelecek vaat eden strateji olarak, kendi dikkat mekanizması temelinde uzun bağımlılıkları modellemede başarılı olan görüntü transformer'leri benimsemektir. Ancak klasik görüntü transformer'ler, yerel ve ince ayrıntılı özniteliklere odaklanma konusunda eksikliklere sahiptirler. [150] çalışmasında, CNN ve Transformer'lerin avantajlarını birleştiren Swin Transformer'e [88] dayalı yeni ve güçlü bir kişi Re-ID modeli önermiş, uzun mesafe bağımlılıkları yakalama yeteneğini daha da artırmak için Uyarlamalı Bölünmüş Kendi Dikkat (ASA) modülünü tanıtmışlardır.

Swin transformer [88] , yerel ve genel öznitelikleri elde ettiği iyi sonuçlardan dolayı futbol oyuncuların Re-ID'si için mevcut çalışmada temel omurga olarak kullanılmıştır. Daha sonra Swin transformer'in yapısına yerel algı bloğu ekleyerek oyuncuların çok ayrıntıları elde edilmiş ve omurgaya entegre edilmiştir.

3.6 Takım Sporları Yayın Videolarında Oyuncuların Kimliklerinin Yeniden Belirlenmesi

Bilgisayarlı görüye dayanarak, görüntülerdeki kişilerin kimliklerinin yeniden belirlenmesi için literatürde farklı metotlar önerilmiştir. Bununla birlikte kişi Re-ID'sinin bir alt dalı olarak, oyuncu Re-ID'si araştırması spor alanı için önem taşımaktadır ancak şu ana kadar bu konuda yeterli çalışma yapılmamıştır. Oyuncu Re-ID, farklı kameralar tarafından yakalanan oyuncuların görüntülerinden oluşan bir galeriden belirli bir oyuncuyu tanımayı hedeflemektedir. Ancak geleneksel kişi Re-ID'ye kıyasla, oyuncu Re-ID çeşitli zorluklarla karşı karşıyadır. Örneğin oyuncular arasında yüksek düzeyde benzerlik, sınırlı veri kümeleri, değişken ve düşük çözünürlüklü görüntüler ve kapanma gibi zorluklarla karşılaşmaktadır. Bu zorlu görevi çözmek için [151] tarafından dikkate duyarlı çoklu ayrıntı ağı (AAMGN) olarak adlandırılan bir yöntem önerilmiştir. Bu yöntem, oyuncuların ayrımcı özniteliklerini yakalamak için farklı ayrıntılardan (granüleriteler) oluşan birden fazla dal kullanmaktadır. Üçlü kayıp ve çapraz entropi kaybı kriterleri kullanılarak, model oyuncunun farklı bölgelerini yerelleştirebilir ve her görüntü çifti arasında kapsamlı bir karşılaştırma yapabilmektedir.

[152] çalışması, takım sporlarının yayın videolarında oyuncuların Re-ID'si üzerine odaklanmaktadır. Bu çalışmada, basit ancak etkili bir hiyerarşik veri örnekleme prosedürü ve ağırlık merkezli kayıp fonksiyonu önerilmektedir. Veri örnekleme prosedürü olarak batch hard [116] yöntemini kullanılmıştır. Bu yöntem, eğitim ve test dağılımlarının benzerliğini iyileştirerek öznitelik vektörlerinin daha iyi tahmin edilen ağırlık merkezlerinin oluşturulmasına yardımcı olmuştur. Bu çalışma öznitelikle futbol maçlarına odaklansa da hiyerarşik örnekleme, liglere, ülkelere veya diğer spor özel durumlarına kolayca genişletilebilir. Sınırlı veri durumunda, basit bir öklid mesafesine dayalı centroid kayıp fonksiyonunun üçlü centroid kayıp fonksiyonundan önemli ölçüde daha iyi performans gösterdiği iddia etmektedirler.

3.7 Bölüm Özeti

Bu bölümde, kişinin kimliğinin yeniden belirleme görevinin temelini oluşturan ilgili literatürün kapsamlı bir değerlendirmesi yapılmıştır. Derin modelin türüne göre DL tabanlı kişi Re-ID yöntemlerinin bir taksonomisi sunulmuştur.

Derin öğrenme temelli kişi yeniden tanımlama yaklaşımlarının kullanımı genel olarak geliştirme aşamasındadır. Bugüne kadar sunulan derin öğrenme yöntemlerinin çoğu,

mevcut temel ağ modellerini iyileştirmeye odaklanmıştır. Ancak kişiye özgü görüntülerin öz nitelikleri dikkate alındığında, yeni bir derin ağ modelinin oluşturulması araştırmacıların ana hedefi haline gelmiştir. Bu yöntemler, sürekli olarak parametre ayarlamaları yapmanın zaman kaybı olduğunu ve ağ yapısı derinleştikçe aşırı uyum sorununun ortaya çıkabileceğini göz önünde bulundurarak daha iyi sonuçlar elde etmiştir. Ayrıca hesaplama maliyetlerini düşürme ve zamanı mümkün olduğunca etkin kullanma ihtiyacı da pratiklik açısından önemlidir. DL modelinin performansını etkileyen önemli bir faktör de kullanılan eğitim örneklerinin sayısıdır. Bu nedenle, karmaşık bir ortamda iyi genelleme yeteneği sağlamak için sadece standart büyük ölçekli bir veri kümesi oluşturmak önemlidir. Ancak günümüzde var olan yöntemler hem hızı hem de tanıma doğruluğunu aynı anda sağlayamamaktadır. Transformer tabanlı derin model, halihazırda var olan kişi kimliğinin yeniden tanımlama problemlerindeki sınırlamaları bir dereceye kadar ele alabildiği için daha güçlü özellik çıkarabilmek amacıyla kullanılabilecek bir teknik olarak kullanılmıştır. Ancak spor takımları oyuncu tanımlaması için çok fazla çalışma mevcut değildir. Oyunculardan kaynaklı zorluklardan dolayı güçlü öz nitelik çıkarma sadece transformer kullanarak yeterli değildir. Dolayısıyla bu çalışmada Swin transformer modelini geliştirerek daha güçlü ve ayırt edici öz nitelikler elde edilmiştir. RFES-ReID, mevcut araştırmada en azından bilindiği kadarıyla mevcut kişi Re-ID veri setlerine benzemeyen, düşük kaliteli futbol oyuncu görüntülerinin sağlam öz nitelik çıkarımında ilk kez kullanılmıştır.

Son zamanlarda spor profesyonellerinin ve taraftarların anlamsal bilgileri ayıklamaya yönelik artan ihtiyacı sayesinde otomatik spor video analizi büyük ilgi çekmektedir. Örneğin futbol, buz hokeyi, basketbol ve voleybol gibi takım sporlarında sıklıkla kullanılmaktadır. Spor analizi sonuçları, televizyon tartışma programları, antrenman programlarının değiştirilmesi, oyun istatistiklerinin oluşturulması ve bir takımın veya oyuncunun güçlü ve zayıf yönlerinin değerlendirilmesi gibi çeşitli şekillerde kullanılabilir. Spor video analizi, her karede topun ve oyuncuların tespit edilmesi, zaman içinde takip edilmesi ve etkileşimlerinin incelenmesini içermektedir. Oyuncu takibi, kapanma, oyuncuların karmaşık hareket kalıpları ve takımdaki benzer görünümeler nedeniyle zorlu bir süreç oluşturmaktadır. Oyuncuların takip edilme aşamasında, takip edilen oyuncular gözden kaybolabilir, yeni oyuncular takip edilebilir veya daha önce atanan takip kimlikleri değişebilir. Bu nedenle, oyuncu yeniden tanımlama otomatik spor analizinde önemli rol oynamaktadır [153]. Futbolcu yeniden tanımlama, kişi yeniden tanımlamanın bir alt kümesi olarak düşünülebilir. Ancak benzer giyimler, hızlı hareketler ve çok fazla kapanmadan dolayı normal kişi yeniden tanımlamada daha zor bir görevdir.

Derin öğrenme tekniklerinin ortaya çıkmadan önce kişiyi kimliğini yeniden belirleme üzerine yapılan ilk araştırmalar, manuel olarak görsel öznitelikleri daha iyi tasarlamaya ve benzerlik ölçümlerini geliştirmeye odaklanmıştır. Ancak son yıllarda derin öğrenme tekniklerinin gelişimiyle birlikte, kişi kimliği yeniden belirleme görevlerinde yaygın olarak kullanılmaktadır. Geleneksel yöntemlerden farklı olarak, derin öğrenme yöntemleri aynı anda daha iyi kişi görüntü özniteliklerini otomatik olarak çıkarabilmektedir. Dolayısıyla daha iyi benzerlik ölçümlerini öğrenebilmektedir. Nesne kimliği yeniden belirleme, belirli nesneleri sahneler ve kamera görünümeleri arasında ilişkilendirmeyi amaçlamaktadır. Görüntüler de sağlam ve ayırt edici özniteliklerin çıkarılmasında en fazla CNN tabanlı yaklaşımın kullanıldığı görülmektedir [124], [154], [155].

CNN ağlarına dayalı yöntemlerin başarısına rağmen, her seferinde farklı bir yerel komşuluk işleyerek, evrişim ve havuzlama tekniklerini kullandığından dolayı verilen görüntü üzerindeki bilgileri kaybetmeye olanak sağlamaktadır. CNN'ler kenarları,

şekilleri ve orijinal kişi özniteliklerini işlerken, bu öznitelikler arasındaki ilişkileri göz ardı etmektedir. Bu nedenle, farklı bakış açılarından yakalanan oyuncu görüntüleriyle karşılaşıldığında, modelin performansı genellikle tatmin edici olmamaktadır. Bununla birlikte dikkat mekanizmasının ortaya çıkması, evrişimli sinir ağlarındaki ayrıntı kaybı sorununu temel olarak çözmektedir.

Google, Transformer'dan esinlenerek bilgisayarlı görü alanında ViT adlı bir model önermiştir [84]. İlk olarak, sınıflandırma görevleri için iyi sonuçlar elde edebilen saf bir Transformer mimarisi doğrudan bir dizi görüntü bloğuna uygulanmıştır. 2020 yılında, Transformer modeli ilk kez bir görüntü sınıflandırma görevinde kullanılmış ve evrişimli sinir ağı modellerinden daha iyi sonuçlar elde etmiştir. Bu tarihten itibaren birçok araştırmacı, bilgisayarla görü alanında Transformer modelinin modelleme yeteneklerini kullanmaya başlamıştır.

Futbolcu yeniden tanıma alanında, CNN'ler genellikle kenarlar, şekiller ve futbolcu öznitelikleri gibi bilgileri içermekte ancak bu öznitelikler arasındaki ilişkileri dikkate almamaktadır. ViT'nin ortaya çıkışı bu sorunu başarılı bir şekilde çözmektedir. ViT, farklı pozlar, açılar ve belirsiz nesnelerle başa çıkma konusunda iyi performans göstermektedir. Ayrıca ViT ve DeiT [85], görüntü tanıma için öznitelik çıkarmada CNN temelli yaklaşımlardan daha pratik olduğunu kanıtlamıştır.

ViT modelindeki yama boyutu sabit ve tek ölçeklidir. Bu öznitelik doğal dil işleme alanı için uygundur ancak bilgisayarla görü alanında görüntünün yama boyutu değişken olabilmektedir. Örneğin bilgisayarlı görüde hedef tespiti ve piksel bazlı segmentasyon gibi aşağı yönlü görevler için genellikle yama boyutu ayarlanması gerekmektedir. Bu durum ViT için hesaplama sorunlarına yol açmaktadır ve sabit yama boyutları bu aşağı yönlü görevler için iyi bir çözüm sunmamaktadır. ViT'in diğer bir dezavantajı ise küçük boyutlu görüntüler kullanılmasıdır. ViT, küçük boyutlu görüntüleri yeniden boyutlandırma stratejisiyle kullanırken bilgi kaybı yaşanabilmektedir. Swin Transformer ise orijinal görüntü boyutunu, örneğin ImageNet için 224×224 , giriş olarak kabul etmektedir. Bu sayede Swin Transformer, bu sorunları önlemek için geliştirilmiştir. Swin Transformer farklı sürgülü pencere kullanarak ViT'i modifiye etmiştir. ViT pencere boyutu sabit olduğu için ve bu pencereler arasında ilişkiye yakalayamadığı için sadece genel öznitelikler elde etmektedir. Ancak Swin Transformer, kayan pencere mekanizması sayesinde farklı pencereler arasındaki bilgi alışverişini gerçekleştirir. Swin Transformer, genel ve yerel öznitelikleri yakalayabilmek için hiyerarşik bir yapı kullanmaktadır. Bu

model, bir görüntüyü daha küçük parçalara bölerek, her bir parçadaki yerel öznitelikleri yakalamak için öz-dikkat mekanizmasını kullanmaktadır. Farklı ölçeklerdeki parçalardaki öznitelikleri birleştirerek genel öznitelikleri elde etmektedir. Bu hiyerarşik yaklaşım, karmaşık uzamsal yapıya sahip görüntülerin etkili bir şekilde işlenmesini sağlamaktadır. Bu araştırmada, futbolcuların görüntülerinin yerel ve genel öznitelik haritalarını elde etmek için Swin Transformer omurga ağı olarak kullanılmıştır.

Swin Transformer'in sağladığı avantajlara rağmen futbolcu görüntülerinde küçük takım rozetleri veya forma modellerindeki farklılıklar gibi ayrıntılarını tam olarak yakalayamayabilir.

Bu sorunu ele almak için transformatörlerin ve evrişimli sinir ağlarının avantajlarını birleştiren bir yerel algı bloğu ile Swin Transformer'i geliştirmeyi tez kapsamında önermekteyiz. Reginal Feature Extraction Swin Transformer dayalı futbolcu kimliği yeniden belirleme (RFES-ReID) olarak adlandırdığımız bu geliştirilmiş mimari, ağın futbolcuların görünümündeki küçük ölçekli öznitelikleri tespit etme yeteneğini önemli ölçüde artırabilir. REFS, yerel öznitelik haritalarını birleştirerek futbolcuların görünümündeki ince ayrıntıları ve farklılıkları daha iyi yakalama yeteneği sunmaktadır. Bu öznitelik futbolcu Re-ID görevinde daha yüksek doğruluk sağlamaktadır. RFES'nin sağladığı bu avantajlar, futbolcu görüntülerindeki detayların önemini vurgulamakta ve yeniden tanıma sürecinde daha etkili bir performans sunmaktadır.

Günümüzde makine öğrenimi teknikleri sağlam ve doğru tahminler yapmak için yaygın bir şekilde kullanılmaktadır. Bununla birlikte öğrenmeye dayalı yaklaşımlar genellikle veri dengesizliği ve benzer görünümlü kişileri ayırt etme zorlukları ile karşılaşmaktadır. Re-ID görevinin performansı için genellikle yanlış pozitif ve yanlış negatif örnekler, kayıp fonksiyonunun tasarımında önemli bir rol oynamaktadır. Bu çalışmada, veri dengesizliği sorununu ele almak ve modelin zor örnekleri etkili bir şekilde öğrenmesini sağlamak amacıyla bir odak kaybı kullanılmıştır. Eğitim aşamasında ayırt etme gücünü ve sağlamlığı artırmak, daha iyi genelleme yapmak ve aşırı uyumu azaltmak için odak kaybı, üçlü kayıp ve çapraz entropi kaybı birleştirilmiştir.

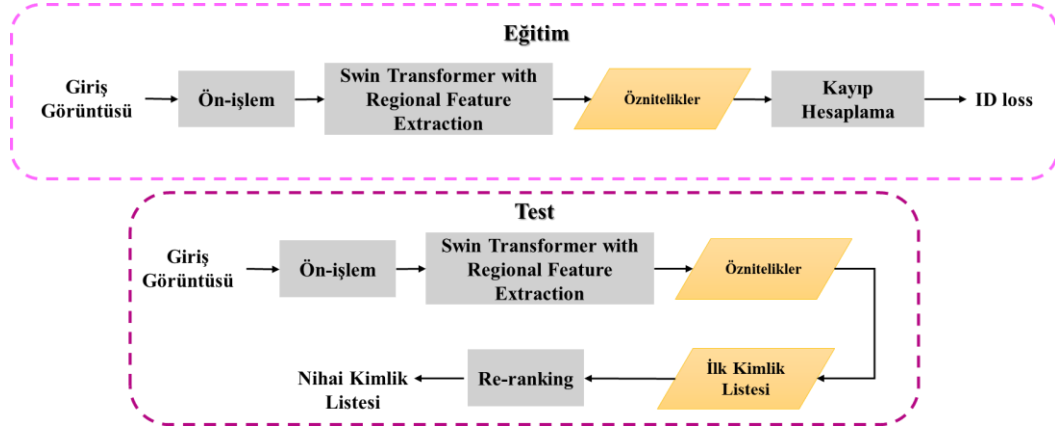
Son olarak, sonuç çıkarım (inference) adımında nihai tahminin doğruluğunu artırmak için yeniden sıralama optimizasyon yöntemi kullanılmıştır.

4.1 Önerilen Yöntem

Şekil 4.1, bu araştırmada önerilen RFES -ReID yönteminin futbolcu Re-ID'si için temel çerçevesini göstermektedir. RFES-ReID yöntemini kullanan futbolcu Re-ID prosedürü, aşağıdaki 4 adımda tanımlanmıştır:

1. Futbolcuların kimliklerinin yeniden tanımlanması için bir ağ modeli oluşturmak, CNN'lerin küresel bilgileri çıkarma konusundaki zayıf yeteneklerini aşmak için Swin Transformer'ını temel omurga ağı olarak seçilmiştir.
2. RFES-ReID, CNN'lerin ve transformerların avantajlarını birleştirerek, yerel algılamayı geliştirmek ve futbolcuların görünümündeki küçük ölçekli öznitelikleri daha iyi tespit etmek amacıyla tasarlanmıştır. Bu geliştirilmiş mimari, futbolcuların üniforma desenleri, rozetleri ve diğer küçük görsel ipuçları gibi ayrıntılarını daha etkili bir şekilde yakalayabilmektedir. RFES, yerel öznitelik haritalarını birleştirerek benzer görünümlü futbolcular arasındaki ince farklılıkları daha iyi ayırt edebilir ve bu da futbolcu yeniden tanımlamasının doğruluğunu artırmaktadır. Bu nedenle, tez kapsamında, RFES-ReID ağı önerilmektedir.
3. Futbolcu kimliği yeniden tanımlamanın tahmin doğruluğunu artırmak için eğitim sürecinde birleşik bir eğitim kaybı tasarımı önerilmektedir. Bu çalışmada, derin metrik öğrenme tasarımı için odak kaybı, üçlü kayıp ve çapraz entropi kaybını birleştirilmiştir. Bu yöntem, futbolcuların benzer ve farklı özniteliklerini ayırt edebilmek ve daha iyi eşleştirmeler elde edebilmek için eğitim sürecinde kullanılan farklı kayıp fonksiyonlarını bir araya getirmektedir.
4. Son aşamada, RFES-ReID modelinden yararlanarak sorgu ve galeri görüntülerinin ayırt edici öznitelikleri çıkarılmıştır. Galeri görüntülerinin her bir sorgu görüntüsüne göre sıralanması için Kosinüs mesafesi metriği kullanılarak aralarındaki benzerlik puanı hesaplanmıştır. Sıralama sonuçlarını optimize etmek için Re-Ranking (RR) yöntemi uygulanmıştır. Ayrıca bu aşamada Re-ID performansı hem niteliksel hem de kantitatif olarak değerlendirilmiştir.

İlerleyen bölümlerde, her aşama için tasarım ve uygulama süreçleri ayrıntılı olarak açıklanacaktır. Futbolcu Re-ID'si için önerilen yöntemin geliştirme sürecinde, Swin Transformer temel modeli kullanılarak ve eğitim stratejisi belirlenerek bir temel ağ modeli oluşturulacaktır. Ardından RFES-ReID modelini geliştirerek ayrıntıları ve farklılıkları yakalamak için eğitim stratejisini belirlemeye geçecektir. Son olarak birleştirilmiş kayıp fonksiyonunu kullanarak, RFES-ReID modeli eğitilecektir.



Şekil 4.1 RFES-ReID yönteminin genel blok diyagramı

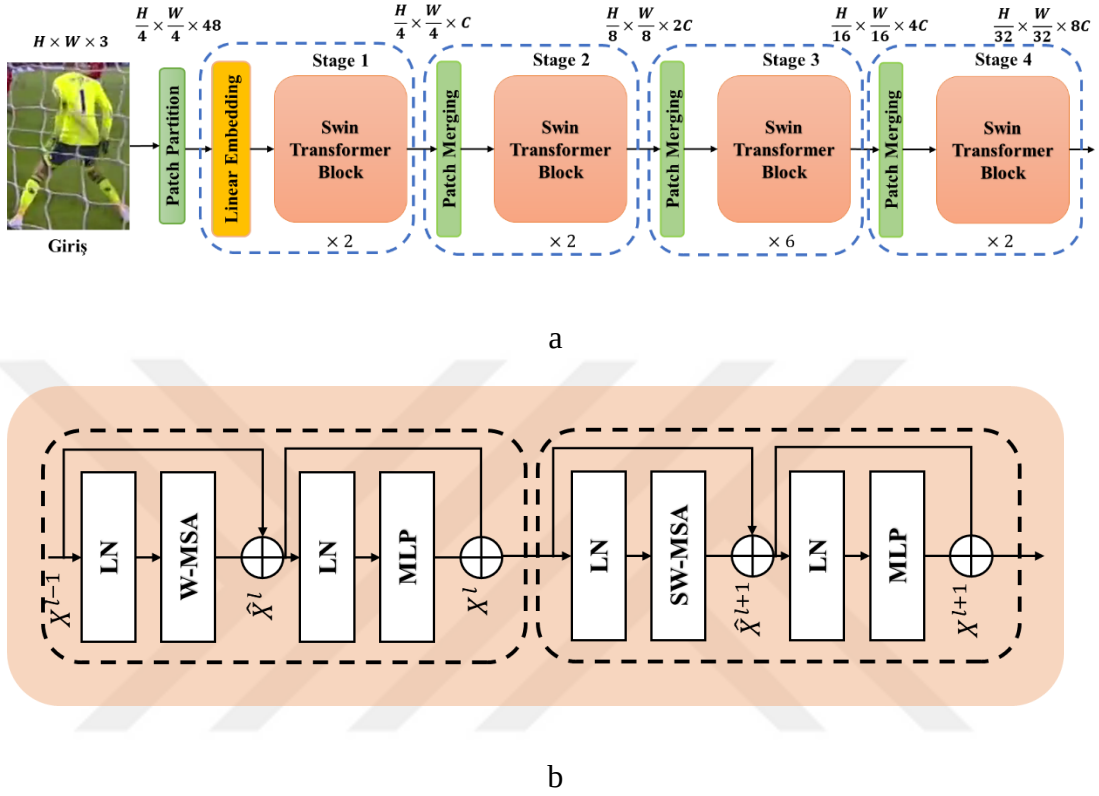
4.2 Futbolcu Kimliği Yeniden Tanımlama için Temel Model Oluşturma

Bu bölümde, futbolcu kimliğinin yeniden tanımlaması için derin öğrenme tabanlı temel modelin oluşturulmasıyla ilgili detaylar sunulacaktır. Swin Transformer temel modeli, mevcut araştırma yaklaşımının temelini oluşturmaktadır. Bu yaklaşımın üç katkısı vardır. İlk olarak Swin Transformer modelinin mimarisi değiştirilerek ve her aşama için yerel algı bloğu eklenerek nesne yeniden tanımlama sorunu için Transformer modellerine dayalı temel derin model mimarisi tasarlanmıştır. Bu değişiklik, futbolcunun görünümündeki küçük ölçekli özniteliklerin tespit doğruluğunu artırmaya yardımcı olmaktadır. Modelden elde edilen öznitelikler aynı zamanda iyi bir ayırt ediciliğe sahiptir. İkinci olarak odak kaybı, üçlü kayıp ve çapraz entropi kaybı gibi farklı kayıp fonksiyonlarını birleştirerek futbolcu yeniden tanımlamanın tahmin doğruluğunu artırmak için bir füzyon eğitim kaybı tasarlanmıştır. Son olarak bu önerilen temel modelin performansını iyileştirmek için RR [156] optimizasyon tekniği kullanılmıştır.

4.2.1 Temel Model Oluşturma

Temel modelin oluşturulmasında omurga olarak Swin Transformer kullanılmıştır. Bölüm 2.4.3'e göre, Swin Transformer [88] bir sinir ağı mimarisidir ve görüntü tanıma görevlerinde etkileyici performans sergilemektedir. Bu mimari, ViT mimarisinin ölçeklenebilirliğini artırmak için önerilmiştir çünkü ViT yüksek hesaplama maliyetine sahip olma eğilimindedir. Swin Transformer, görüntüyü birden çok ölçekte işleyen bir hiyerarşik yapı sunmaktadır. Ayrıca "kaydırılmış pencere" mekanizmasını kullanarak görüntüyü daha az sayıda belirteçle işleyerek hesaplama maliyetini azaltmaktadır. Bu

öznitelikler sayesinde Swin Transformer, görüntü sınıflandırma, nesne algılama, semantik bölümlendirme gibi çeşitli bilgisayarlı görü görevlerinde önemli gelişmeler kaydetmiştir. Görüntü tanıma görevlerindeki başarısı, transformatör tabanlı mimarilere yönelik daha fazla araştırmaya ilham vermiştir.



Şekil 4.2 a) Temel mimari, b) Swin Transformer bloğu.

Şekil 4.2'de, omurga olarak kullanılan Swin Transformer modeli gösterilmektedir. Genel olarak giriş görüntüsü örtüşmeyen yamalara bölünür ve ardından bir dizi Transformer bloğu tarafından işlenmektedir. Bu Transformer blokları MSA ve MLP katmanlarını içermektedir ve farklı ölçek ve konumlardaki yamaları hiyerarşik olarak işlemektedir. Öz-dikkat mekanizması, modelin küresel bağlamsal bilgileri yakalayabilmesini sağlamaktadır. MLP katmanları ise yerel öznitelik etkileşimlerini modellemek için kullanılmaktadır. Swin Transformer, isteğe bağlı boyuttaki giriş görüntülerini işleyebilir ve modelin son aşamasında kanal sayısını ayarlayarak istenen herhangi bir boyutta bir öznitelik vektörü çıkarabilecek şekilde yapılandırılabilir.

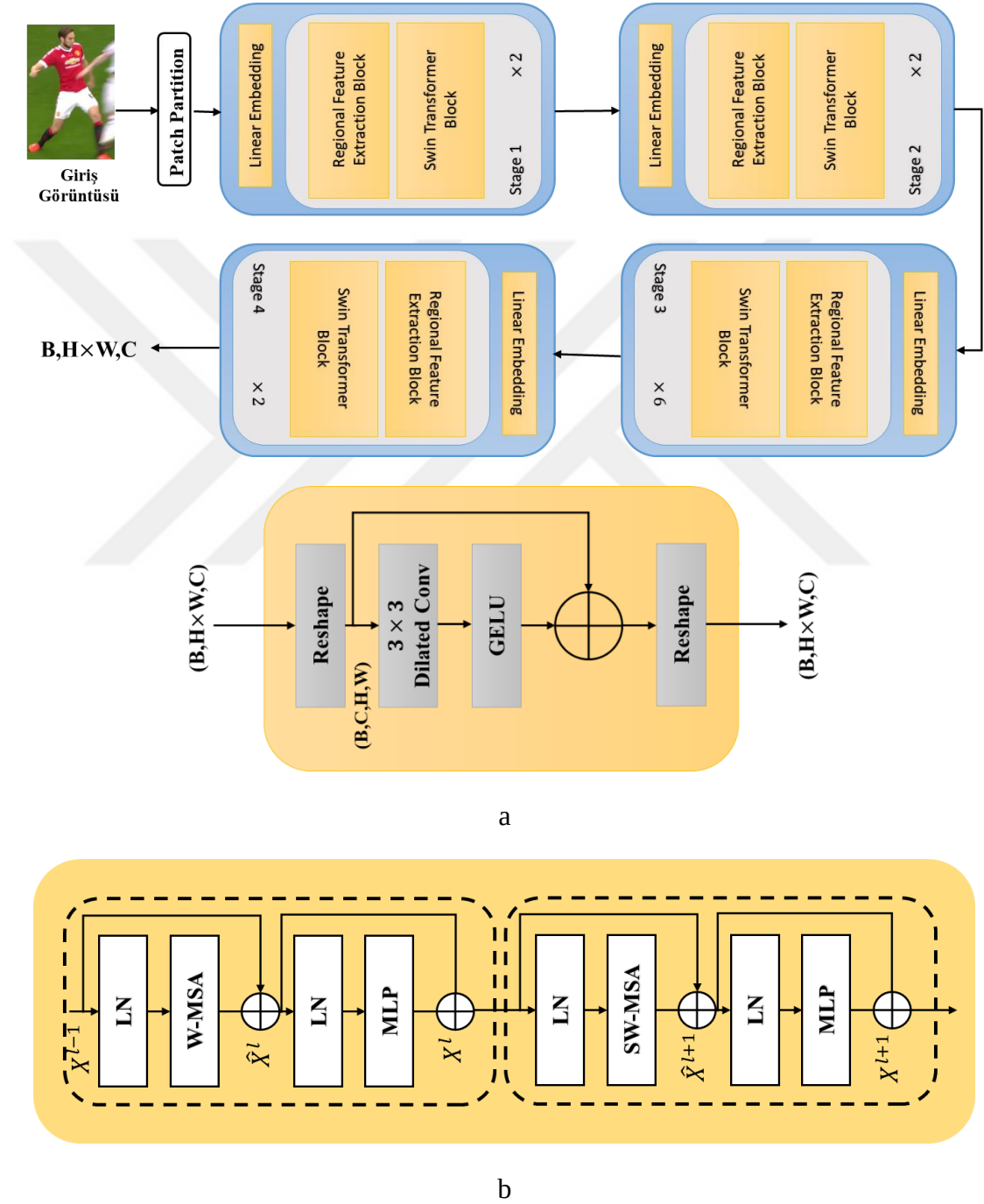
4.2.2 RFES-ReID Modeli

Bu bölümde, tasarlanan ağ yapısına odaklanılmaktadır. Önerilen RFES omurga ağının akış şeması Şekil 4.3'te gösterilmiştir. Swin transforer, modelin küçükten büyüğe Swin-

T, Swin-S, Swin-B ve Swin-L olmak üzere dört versiyonunu sağlamaktadır. Burada, Swin-T'e kullanılmıştır. Bu modelin hiper parametreleri aşağıdaki gibidir:

Swin-T: $C = 96$, katman numarası = $\{2, 2, 6, 2\}$

C birinci aşamadaki gizli katmanların kanal numarasıdır. Swin-T ve Swin-S'nin karmaşıklığının, sırasıyla ResNet-50 (DeiT-S) ve ResNet-101'inki ile aynıdır.



Şekil 4.3 Yerel algı Swin transformerin (RFES) mimarisi. (a) Yerel algı bloğunun ayrıntılı yapısı; (b) Swin transformer bloğunun ayrıntılı yapısı.

ViT'e benzer şekilde önce giriş olarak bir RGB görüntüsünü yama bölme katmanıya örtüşmeyen yamalara bölmektedir. Her yama bir "belirteç" olarak ele alınır ve özelliği piksel RGB değerlerinin bir birleşimi olarak ayarlanır. Swin Transformer, farklı sayıda belirteç üretmek için dört aşama içermektedir. $Y \times G$ boyutunda bir görüntü verildiğinde bir belirteç, 4×4 boyutunda bir RGB görüntü yamasının ham piksel birleştirme vektörüdür. C boyutuna sahip bir vektörde eşlemek için bu belirteç üzerinde doğrusal bir gömme kullanılır. Aşama 1, 2, 3 ve 4, sırasıyla $\frac{Y}{4} \times \frac{G}{4}$, $\frac{Y}{8} \times \frac{G}{8}$, $\frac{Y}{16} \times \frac{G}{16}$, ve $\frac{Y}{32} \times \frac{G}{32}$ belirteçler üretmektedir. Her aşama, bir yama birleştirme bloğundan (bir yama bölme katmanı ve bir doğrusal gömme katmanının bir kombinasyonu), yerel algılama bloğundan ve Swin transformer bloklarından oluşmaktadır.

4.2.2.1 Swin Transformer bloğu

Swin transformer bloğu, Swin transformer algoritmasının temel parçasıdır. Ayrıntılı yapı Şekil 4.3 b'de gösterilmiştir. Blok, sabit pencere çok başlı öz dikkat (W-MSA), kaydırılmış pencereler çok başlı öz dikkat (SW-MSA) ve çok katmanlı algılayıcıdan (MLP) oluşmaktadır. Ortaya bir LN katmanı eklemek, eğitimi daha kararlı hale getirmektedir ve her modülden sonra artık bağlantı kullanılmaktadır. Bu kısım aşağıdaki gibi ifade edilebilir:

$$\hat{X}^l = W - MSA(LN(X^{l-1})) + X^{l-1} \quad (4.1)$$

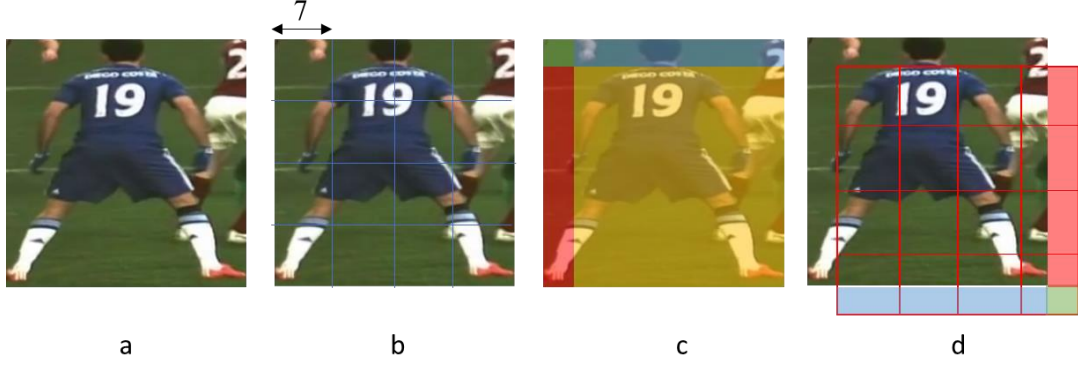
$$X^l = MLP(LN(\hat{X}^l)) + \hat{X}^l$$

$$\hat{X}^{l+1} = SW - MSA(LN(X^l)) + X^l$$

$$X^{l+1} = MLP(LN(\hat{X}^{l+1})) + \hat{X}^{l+1}$$

4.2.2.2 W-MSA ve SW-MSA

Geleneksel ViT'deki çok başlı öz-dikkat (MSA) [80] ile karşılaştırıldığında, Swin transformer bloğundaki W-MSA bir penceredeki hesaplama alanını bir birim olarak kontrol etmektedir. Burada pencere boyutu varsayım olarak 7'ye ayarlanmıştır. Bu, ağ hesaplamalarının miktarını azaltmaktadır ve Şekil 4.4'te gösterildiği gibi karmaşıklığı görüntü boyutunun doğrusal bir oranına indirgemektedir. MSA, pencereler arasında bağlantılardan yoksundur. SW-MSA'nın konumu, W-MSA katmanından sonradır. Bu nedenle, pencereler arası iletişimi gerçekleştirmek için W-MSA'dan sonra farklı bir pencere bölümlenme yöntemi sağlamak için SW-MSA gereklidir.



Şekil 4.4 Kaydırılan pencerelerin etki mekanizması. (a) Giriş görüntüsü; (b) Girdi görüntüsünün W-MSA aracılığıyla pencere segmentasyonu (pencere boyutu 7 olarak ayarlanmıştır); (c) Kaydırılan pencerelerin eylemi; (d) Kaydırılan pencereler SW-MSA aracılığıyla farklı bir pencere bölümlene yöntemi.

Girdi görüntüsünün W-MSA yoluyla pencere segmentasyonunun sonucu Şekil 4.4 b'de gösterilmektedir. Görüntünün her döngüsü pencere boyutunun yarısı kadar yukarı ve sola hareket ettirilir ve Şekil 4.4 c'deki mavi ve kırmızı alanlar, Şekil 4.4 d'de gösterildiği gibi sırasıyla görüntünün alt ve sağ taraflarına taşınır. Pencere, bu kaymalara dayalı olarak W-MSA'ya göre bölümlere ayrılır ancak SW-MSA, W-MSA'dan farklı bir pencere bölümlene prosedürü kullanmaktadır.

4.2.2.3 Yerel Algı Bloğu (Regional Feature Extraction Block - RFEB)

Bir transformer'deki konum kodlaması (position encoding), görüntünün yerel korelasyonunu ve yapısal bilgisini tespit etmekte başarısız olabilmektedir. Swin transformer'in hiyerarşik bir yapıda sıralı katmanlardan oluşan bir kaydırma penceresi şemasına sahip olmasına rağmen, çok çeşitli uzamsal bağlam bilgisi hala çok iyi kodlanmamıştır. Bu sorunu azaltmak için Swin transformer bloğunun önüne yerel algılama bloğu (RFEB) yerleştirilmiştir. Yerel algı bloğunun bileşimi Şekil 4.3 a'da gösterilmiştir.

Swin transformer'lerde veri akışı geleneksel CNN'lerdeki öznitelik haritaları yerine vektörlerden oluşmaktadır. LPB'de, öncelikle bir grup vektör özelliği bir uzamsal öznitelik haritasına yeniden şekillendirir.

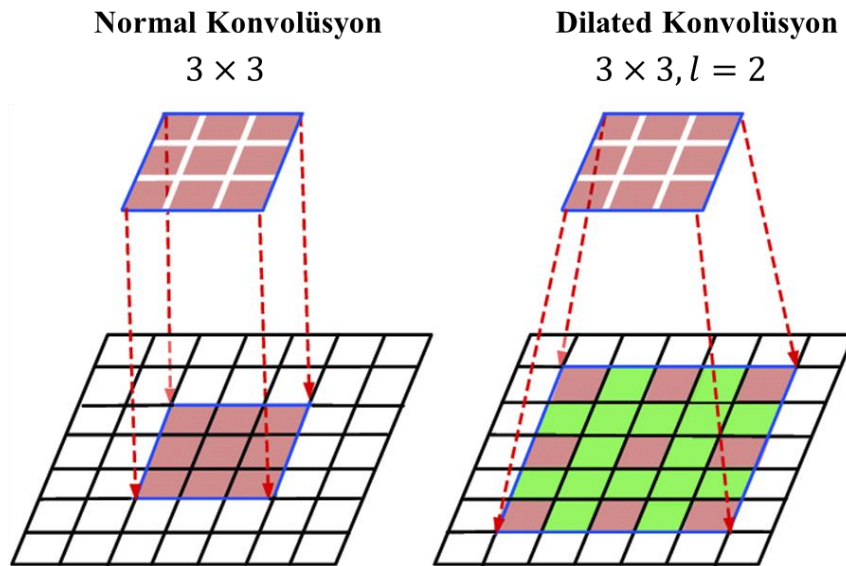
Örneğin bir $(B, H \times W, C)$ boyuttaki belirteç, öznitelik haritası (B, C, H, W) olarak yeniden şekillendirilir. Daha sonra 3×3 dilate evrişim katmanı (dilatasyon = 2) eklenmektedir. Genişletilmiş Evrişim [157] (Dilated Convolution) girdiye ardışık elemanları arasına delikler sokarak genişleten bir tekniktir. Daha basit bir ifadeyle,

evrişim ile aynıdır ancak girdinin daha geniş bir alanını kapsayacak şekilde piksel atlama içermektedir. Genişletilmiş evrişim, parametre sayısını artırmadan ağına daha geniş bir alıcı alana sahip olmasını sağlayan CNN'ler kullanılan bir tür evrişim işlemidir.

Düzenli bir evrişim işleminde, sabit boyutlu bir filtre giriş öznetelik haritası üzerinde kaymaktadır. Dolayısıyla filtredeki değerler, tek bir çıkış değeri üretmek için giriş öznetelik haritasındaki karşılık gelen değerlerle çarpılmaktadır. Bir nöronun çıktı öznetelik haritasındaki alıcı alanı, filtrenin "görebileceği" giriş öznetelik haritasındaki alandır. Alıcı alanın boyutu, filtrenin boyutuna ve evrişimin adımına göre belirlenmektedir.

Aksine genişletilmiş bir evrişim işleminde [157], filtre değerleri arasına boşluklar eklenerek filtre "genişletilir". Genişleme hızı, filtrenin alıcı alanını parametre sayısını artırmadan etkili bir şekilde artırmaktadır çünkü filtre hala aynı boyuttadır ancak değerler arasında boşluklar vardır. Bu, daha geniş bir alıcı alana ihtiyaç duyulan durumlarda yararlıdır. Genişletilmiş evrişim öznetelikleri birden çok ölçekte yakalayabilmektedir [158].

Ek bir parametre l (genişletme faktörü), girişin ne kadar genişletildiğini göstermektedir. Yani bu parametrenin değerine göre çekirdekte $(l-1)$ piksel atlanmaktadır. Şekil 4.5, normal ve genişletilmiş evrişim arasındaki farkı göstermektedir. Özünde, normal evrişim sadece 1-genişletilmiş bir evrişimdir.



Şekil 4.5 Normal Evrişim ve Genişletilmiş Evrişim

Geniřletilmiř evriřim, havuzlama olmadan kapsanan girdi grntsnn alanını geniřletmeye yardımcı olmaktadır. Ama her evriřim iřleminde elde edilen ıktıdan daha fazla bilgi elde etmektir. Bu yntem aynı hesaplama maliyetinde daha geniř bir grř alanı sunmaktadır. Dilation faktrnn (l) deęerini, farklı l deęerlerindeki her konvolsyonla elde edilen bilgi miktarını gzlemleyerek belirlenebilir.

Bu yntemi kullanarak, parametre sayısını artırmadan daha fazla bilgi elde edebilmektedir. řekil 4.5'de, saędaki grnt geniřlemiř konvolsyonu gstermektedir. $l = 2$ deęerini koruyarak, filtreyi giriře eřlerken 1 piksel atlamaktadır. Bylece her adımda daha fazla bilgi kapsayabilmektedir. Bu iřlem ařaęıdaki denklemi kullanarak gerekleřtirilmiřtir:

$$(F_{*l}k)(p) = \sum_{(s+lt=p)} F(s)k(t) \quad (4.2)$$

Burada, $F(s)$ Giriři, $k(t)$ uygulanan filtreyi, $*l$ l-geniřletilmiř evriřimi ve $(F_{*l}k)(p)$ ıktıyı temsil etmektedir. Normal evriřim yerine geniřletilmiř evriřim kullandığımızda daha geniř alıcı alanına sahip olarak hesaplama aısından daha verimli olmaktadır (aynı hesaplama maliyetinde daha geniř bir kapsam saęladığı iin). Ayrıca birleřtirme adımını atladığı iin daha az bellek tketmektedir ve ıktı grntsnn znrlęnde kayıp yařamamaktadır. Dolayısıyla bu evriřimin yapısı, verilerin sırasını korumaya yardımcı olmaktadır.

Geniřletilmiř evriřim katmanının ardından bir GELU aktivasyon fonksiyonu eklenmiřtir. GELU (Gaussian Error Linear Unit), bir aktivasyon fonksiyonu olarak kullanılan bir matematiksel iřlevdir. GELU, znitelikle derin sinir aęlarında daha hızlı ve daha iyi bir řekilde yakınsamayı saęlamak iin tasarlanmıřtır. GELU fonksiyonu, sigmoid fonksiyonu ve ReLU fonksiyonunun birleřimi olarak dřnlelebilmektedir. Standart Gauss kmlatif daęılım fonksiyonu olan $\Phi(x)$ ile tanımlanmaktadır. GELU, girişleri iřlemek iin ReLU'larda olduęu gibi iřaretleriyle deęil, yzdelik dilimlerine gre aęırlıklandırılır. Bu nedenle, GELU, daha yumuřak bir ReLU olarak dřnlebilir. Matematiksel olarak, GELU aktivasyon fonksiyonu řu řekilde formle edilmektedir:

$$\text{GELU}(x) = xP(X \leq x) = x\Phi(x), \quad (4.3)$$

where $\Phi(x) = P(X \leq x), X \sim N(0,1)$

$\Phi(x)$, standart normal daęılımın birikimli daęılım fonksiyonudur (CDF). Bu fonksiyonun seimi, sinir hcreti girişlerinin genellikle normal daęılımı takip etmesinden

kaynaklanmaktadır. Dolayısıyla, GELU, giriş değeri olan x 'e bağlı olarak $\Phi(x)$ fonksiyonunu kullanır. $\Phi(x)$, x 'in değerine bağlı olarak değişir ve x azaldıkça, bir sinir hücresini bırakma olasılığı (0 ile çarpma) artar. Bu nedenle, GELU'nun uyguladığı dönüşüm stokastik bir niteliğe sahiptir, ancak girişin değerine bağlıdır. Bu dağılımı seçme sebebi ise nöronun girişi, özellikle Toplu Normalizasyonun hemen ardından normal bir dağılım yapılmaktadır. Ancak herhangi bir aktivasyon fonksiyonunun çıktısı stokastik değil, deterministik olmalıdır. Böylece dönüşümünün beklenen değerini aşağıdaki gibi hesaplıyoruz:

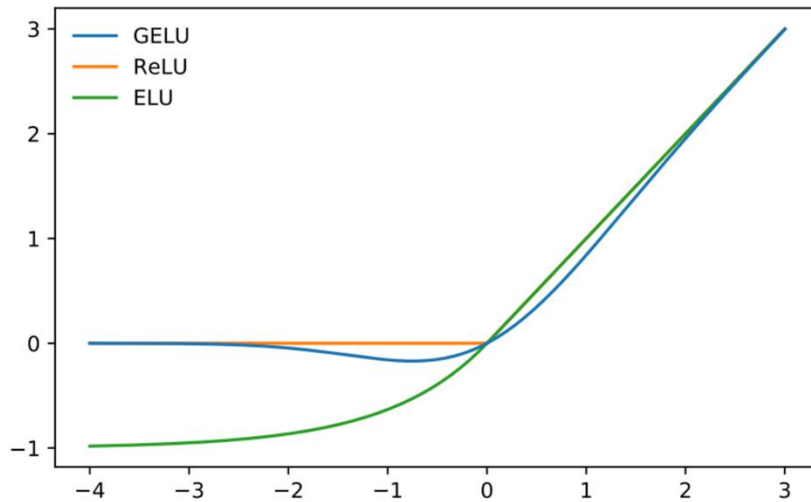
$$E[mx] = xE[m] \quad (4.4)$$

m bir Bernoulli rasgele değişkeni olduğundan, beklenen değeri $\Phi(x)$ 'dur.

$$\Rightarrow E[mx] = x\Phi(x) \quad (4.5)$$

$\Phi(x)$, Gauss dağılımının kümülatif bir dağılımı olduğundan ve genellikle hata fonksiyonuyla hesaplandığından, GELU aşağıdaki şekilde tanımlanmaktadır:

$$\begin{aligned} \text{GELU}(x) &= xP(X \leq x) = x\Phi(x) \\ &\approx 0.5x(1 + \tanh[\sqrt{2/\pi}(x + 0.044715x^3)]) \end{aligned} \quad (4.6)$$



Şekil 4.6 GELU ($\mu=0$, $\sigma=1$), ReLU ve ELU ($\alpha=1$) [159].

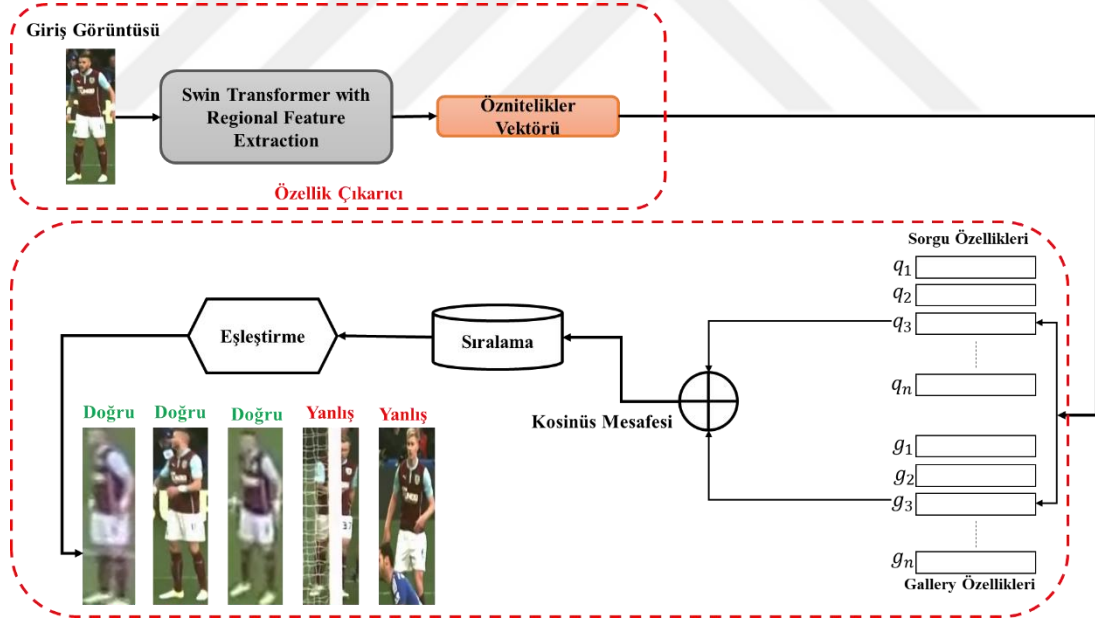
Şekil 4.6 görüldüğü gibi eğitim hızı, yakınsama oranı ve doğruluk açısından GELU diğer aktivasyon fonksiyonlarına göre daha iyi performans göstermektedir. Daha sonra bir artımlı bağlantı (residual connection), alansal veya uzamsal yerel özniteliklerin çıkarımını artırmak için ve aynı zamanda alıcı alanın yeterince geniş kalmasını sağlamak amacıyla

kullanılmaktadır. Son olarak öznitelik haritası (B, H, W, C) olarak yeniden şekillendirilir ve Swin transformer bloğuna gönderilmektedir.

4.3 Öznitelikleri Çıkarma ve Mesafe Metrik Öğrenme

Tüm sorgu ve galeri görsellerinin öznitelikleri çıkarıldıktan sonra bir dosyaya kaydedilerek sıralama (Ranking) işlemine hazırlanmaktadır. Sıralama (ranking) işlemi, bir sorgunun öznitelik vektörünün tüm galeri özniteliklerine göre benzerlik puanlarını belirlemeyi amaçlamaktadır. Daha sonra benzerlik puanlarının hesaplanması, mesafe metriği kullanılarak gerçekleştirilir ve daha küçük bir mesafe daha yüksek benzerlik puanına sahip olmaktadır. Öznitelik çıkarma ve mesafe metrik öğrenme süreçleri Şekil 4.7'te özetlenmiştir. Bu çalışmada kullanılan mesafe ölçüsü, aşağıdaki denklemde gösterildiği gibi hesaplanabilen X ve Y vektörleri arasındaki açısız kosinüs mesafesini sağlayabilen kosinüs mesafesidir:

$$\cos \theta = \frac{XY}{|X||Y|} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (4.7)$$



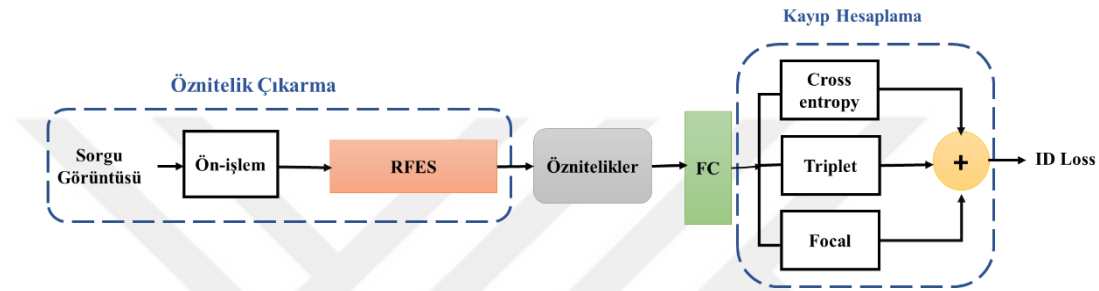
Şekil 4.7 Öznitelik çıkarma ve mesafe metriği öğrenme

4.4 Kayıp Hesaplaması

Önerilen kayıp hesaplaması, Şekil 4.8'te gösterilmektedir. Burada, oluşturulan öznitelikler, öznitelik üretici sürecinden sonra kimlik tahmini yapmak için tamamen bağlı (FC) bir katmanla beslenmektedir. Kaybı elde etmek için tahmin sonucu çapraz entropi,

üçlü kayıp ve odak kaybı dahil olmak üzere üç kayıp fonksiyonundan ayrı ayrı beslenmektedir. Daha sonra her kayıp dalından elde edilen sonuçlar birleştirilmektedir. Buradaki amaç her üç fonksiyonun avantajlarından yararlanarak tahmin doğruluğunu artırmaktır. Bunun için farklı kimliklere sahip benzer futbolcuları farklı bakış açılarından yakalanan aynı kimliğe sahip futbolculardan etkili bir şekilde ayırt edebilen eğitim aşamasında metrik öğrenme için üç kaybı da birleştiriyor. Birleşmiş kaybı L_{Fusion} aşağıdaki gibi tanımlanmaktadır:

$$L_{Fusion} = L_F + L_T + L_{CE} \quad (4.8)$$



Şekil 4.8 Birleşmiş kayıp hesaplaması.

Aşağıda her üç kayıp fonksiyonu ayrı ayrı açıklanmıştır:

- Çapraz entropi veya sınıflandırma kaybı, bir kişinin kimliğinin yeniden tanınmasındaki eğitim sürecinde kullanılan genel görüntü sınıflandırma göreviyle aynı prensiplere dayanmaktadır. Eğitim veri setindeki kişi kimlik etiketlerini daha iyi tahmin edebilmek adına FC katmanı için daha iyi parametreler öğrenmeye odaklanmaktadır. Test sürecinde kişi Re-ID eğitim ve test kümeleri arasında herhangi bir örtüşme olmamaktadır. Dolayısıyla tamamen genel öznitelik vektörüne bağlı olan mesafe matrisinin hesaplanmasını gerekmektedir. Sınıflandırma kaybı L_i aşağıdaki şekilde hesaplanmaktadır:

$$L_i = - \sum_{k=1}^{n_{id}} 1\{y_{n_i} = k\} \log \frac{e^{\hat{y}_{n_i}^i}}{\sum_{l=1}^{n_{id}} e^{\hat{y}_l^i}} \quad i = 1, \dots, N_c \quad (4.9)$$

Burada $\hat{y}_{n_i}^i$, n_i 'inci eğitim örneğinin tahmin edilen kişi sınıflandırma puanıdır. n_{id} , kişi kimliklerinin toplam kategori sayısıdır. Toplam çapraz-entropi kaybı L_{CE} aşağıdaki gibi hesaplanmaktadır:

$$L_{CE} = \sum_{i=1}^{N_c} L_i \quad (4.10)$$

- Üçlü kaybın (Triplet loss) amacı, bir öznitelik uzayında benzer örnekler arasındaki mesafeyi küçültmek ve farklı örnekler arasındaki mesafeyi büyütmektir. Yani, bu kayıp fonksiyonu, benzer örneklerin birbirine yakın olmasını ve farklı örneklerin birbirinden uzak olmasını sağlamayı hedeflemektedir. Bu amacı gerçekleştirmek için triplet loss, üç farklı örneklemden oluşan tripletler üzerinde çalışmaktadır: bir referans örnekleme (anchor), bir pozitif örnekleme (positive) ve bir negatif örnekleme (negative). Pozitif örnekleme, anchora benzerken negatif örnekleme anchora farklıdır. Üçlü kaybın L_T matematiksel gösterimi aşağıda gösterilmiştir:

$$L_T = \sum_i^N \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \tau \right] \quad (4.11)$$

Burada $f(x)$ x'in bir girişi, i i'inci girişi temsil etmektedir. a bir referans görüntüyü, p pozitif bir görüntüyü (aynı kimlik numarasına sahip görüntü) ve n negatif bir görüntüyü (aynı kimlik numarasına sahip görüntü), τ eşik değeri ifade etmektedir. Amaç, birinci terimi en aza indirerek ve ikinci terimi en üst düzeye çıkararak yukarıdaki denklemi en aza indirmektir.

- Focal loss, öznitelikle sınıf dengesizliği olan problemlerde kullanılan bir kayıp fonksiyonudur [160]. Odak kaybı sınıflandırma görevlerinde daha az temsil edilen sınıfların daha fazla vurgulanmasını sağlamaktadır. Dolayısıyla focal kaybı sınıf dengesizliği sorununu ele almak için çapraz entropi kaybı fonksiyonunun değiştirilmiş bir versiyonudur [161]. Odak kaybı L_F şu şekilde tanımlanmaktadır:

$$L_F = -(1 - L_T)^\gamma \times \log L_T \quad (4.12)$$

Burada L_T , tahmin edilen olasılığı temsil etmektedir ve γ , odaklanma parametresidir. Odak kaybı, çapraz entropi kaybının yeniden ağırlıklandırılması olarak görülebilir; burada her bir örneğin toplam kayba katkısı tahmin edilen olasılık ve γ değeri tarafından belirlenmektedir.

Odak kaybı denklemine bakacak olursak;

$(1 - L_T)^\gamma$, sınıflandırılması zor örneklerle daha fazla önem veren ağırlıklandırma terimidir. $\gamma = 0$ 'a ayarlandığında, ağırlıklandırma terimi tüm örnekler için 1'e eşittir ve bu da odak kaybının standart çapraz entropi kaybına düştüğü anlamına gelmektedir. γ daha yüksek bir değere ayarlandığında ağırlıklandırma terimi

sınıflandırması kolay örnekler için daha küçük, sınıflandırması zor örnekler için daha büyük olacaktır.

$\log L_T$, gerçek sınıf için tahmin edilen olasılığı cezalandıran çapraz entropi terimidir. Çapraz entropi kaybına benzer, ancak ağırlıklandırma terimiyle ölçeklenir.

4.5 Model optimization

Önerilen temel modelin performansını optimize etmek ve futbolcu Re-ID'sinde iyi sonuçlar elde etmek için son yıllarda önerilen birçok optimizasyon tekniği geliştirilmiştir. Bu araştırmada, önerilen temel modeli optimize etmek için RR [156] tekniği seçilmiştir. Bunun sebebi ise uygulanması ve temel modelle birleştirilmesi karmaşık değildir ve hiçbir ek parametrenin öğrenilmesi gerekmemektedir.

- **Re-Ranking:** Kişi Re-ID'si veri alışverişi (data retrieval) süreci olarak değerlendirildiğinde, sıralama işlemi doğrudan sonuçlarla ilişkilidir ve bu nedenle Re-ID'nin başarısı için temel bir adım olarak kabul edilmektedir. Bu nedenle, RR Re-ID'nin doğruluğunu artırmak için kritik bir adımdır. Dolayısıyla RR [156], kişinin Re-ID sisteminin performansını optimize etmek için bir optimizasyon tekniği olarak önerilmiştir. Re-ID'nin RR yapması için temel fikir, k-karşılıklı kodlama (k-reciprocal encoding) yöntemini [162] kullanmaktır.

[162]'a göre, $N(p, k)$, p arama görüntüsünün k-en yakın komşularının tipik bir örneği olduğunu varsayarsak, aşağıdaki denklem gibi tanımlanabilmektedir:

$$N(p, k) = g_1; g_2; \dots; g_k \quad (4.13)$$

k-karşılıklı en yakın komşular $R(p, k)$ aşağıdaki gibi tanımlanabilmektedir:

$$R(p, k) = \{g_i \mid (g_i \in N(p, k)) \wedge (p \in N(g_i, k))\} \quad (4.14)$$

k-karşılıklı en yakın komşuların sorgu p'ye k-en yakın komşulardan daha yakın ilişkili olduğu kanıtlanmıştır [156]. Ancak futbolcuların pozları, kapanmalar, aydınlatmalar ve kamera açıları gibi farklılıklar nedeniyle, k-en yakın komşular içerisindeki tüm pozitif görüntüleri dahil etmek mümkün olmayabilir. Dolayısıyla bu pozitif görüntülerin k-karşılıklı en yakın komşular içerisinde de dahil edilmemektedir. Bu nedenle, bu problemin üstesinden gelmek için $R(p, k)$ 'deki her adayın $\frac{1}{2}$ k-karşılıklı

en yakın komşularını aşağıdaki koşula dayanarak daha doğru bir $R^*(p, k)$ kümesine eklemektedir:

$$\begin{aligned} \mathcal{R}^*(p, k) &\leftarrow \mathcal{R}(p, k) \cup \mathcal{R}\left(q, \frac{1}{2}k\right) \\ \text{s.t. } &\left|\mathcal{R}(p, k) \cap \mathcal{R}\left(q, \frac{1}{2}k\right)\right| \geq \frac{2}{3} \left|\mathcal{R}\left(q, \frac{1}{2}k\right)\right| \\ &\forall q \in \mathcal{R}(p, k) \end{aligned} \quad (4.15)$$

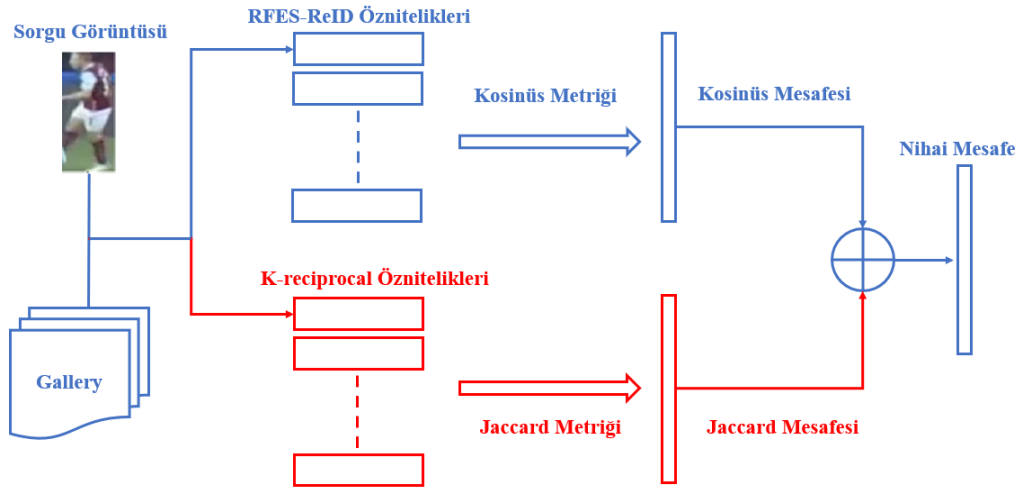
Daha sonra sorgu ve galeri kümelerindeki her görüntünün k-karşılıklı en yakın komşu kümesi bir öznitelik vektörü olarak oluşturulmaktadır. Ardından bu öznitelikler arasındaki mesafeyi hesaplamak için Jaccard uzaklığı kullanılmaktadır. Daha spesifik olarak, Jaccard mesafesi aşağıdaki gibi hesaplanmaktadır:

$$d_{Jaccard}(p, g_i) = 1 - \frac{|\mathcal{R}^*(p, k) \cap \mathcal{R}^*(g_i, k)|}{|\mathcal{R}^*(p, k) \cup \mathcal{R}^*(g_i, k)|} \quad (4.16)$$

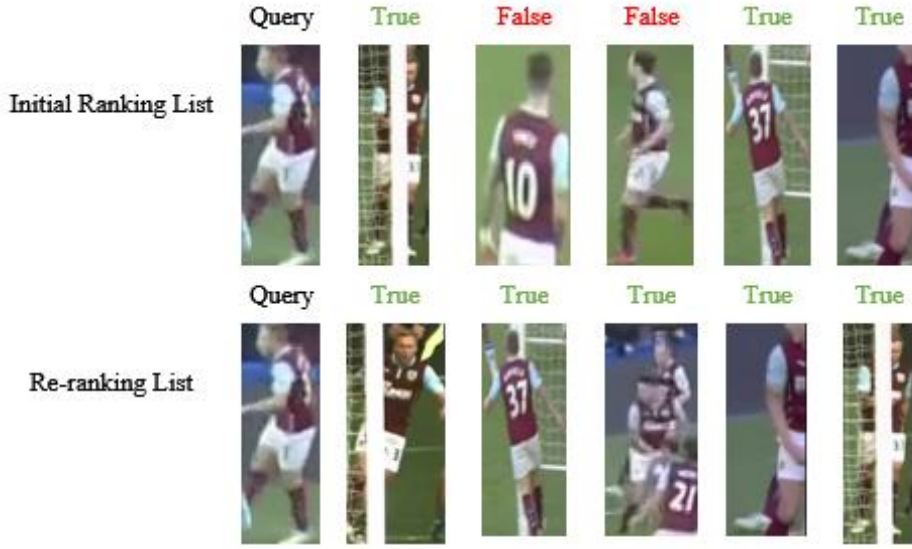
Burada, $|\cdot|$ ilgili kümedeki aday sayısını ifade etmektedir. Sıralama için Jaccard mesafesi ve orijinal mesafe (Kosinüs mesafesi) ağırlıklı bir şekilde birleştirilerek son mesafe hesaplanmaktadır. Bu çalışmada, her iki mesafenin ağırlıkları da 0.5 olarak belirlenmiştir.

$$D_{\text{Final}} = \frac{(D_{\text{Original}} + D_{\text{Jaccard}})}{2} \quad (4.17)$$

RR tekniği, kolay uygulanabilir olması ve öğrenme için ekstra parametrelere ihtiyaç duymaması nedeniyle bu çalışmada önerilen temel modelin optimize edilmesi için benimsenmiştir. Futbolcu Re-ID'si için RR çerçevesinin prosedürü Şekil 4.9'te gösterilmektedir.



Şekil 4.9 Futbolcu Re-ID'si için RR prosedürü.



Şekil 4.9 Devamı

4.6 Deneysel Çalışma

Bu bölüm, Futbolcu Re-ID'si için önerilen yöntemi değerlendirmek için kullanılan deneysel ayarları özetlemektedir. Deneylere geçmeden önce, deneylerde kullanılan futbolcu veri seti ve bu veri setinin hazırlanması için gerekli olan ön işleme süreçleri kısaca anlatılmaktadır. Ardından seçilen veri kümesi üzerinde önerilen modelin performansı açıklanmaktadır. Son olarak kayıp hesaplanması için önerilen modelin uygulanmasının ayrıntıları ve bunların eğitim yapılandırması tartışılmaktadır.

4.6.1 Benchmark Veri Setleri

Kişi Yeniden Kimliği için çeşitli veri kümeleri kamuoyuna açıktır. Bölüm 2.7.1'de, bu veri kümelerinin çoğu tanıtılmıştır. Önerilen yöntemi değerlendirmek için deneylerde SoccerNetV3 [95] veri seti kullanılmıştır. Tablo 4.1, bu veri setinin detaylarını özetlemektedir.

Tablo 4.1 Bu tez boyunca deneylerde kullanılan veri kümelerine ilişkin ayrıntılar.

Veri seti	Kimlik Sayısı		Görüntü Sayısı		
	Eğitim	Test	Eğitim	Query	Gallery
SoccerNetV3	161443	30059	248234	11777	34989

Futbolcu Re-ID modellerini doğru bir şekilde değerlendirmek için gerçek dünya senaryosunu yansıtan iyi bir veri kümesine ihtiyaç vardır. Bu veri kümesi, açı değişiklikleri, aydınlatmadaki farklılıklar, arka plandaki değişiklikler ve kamera öznitelikleri gibi gerçekçi video ayarlarını içermelidir. Ayrıca aynı futbolcu kimliğini farklı bakış açılarından ve yörüngelerden yakalamak için farklı kameralardan alınan görüntüleri içermelidir.

4.6.2 Veri Ön İşleme

RFES-ReID modelini öznitelik öğrenme amacıyla eğitmek için veri kümelerini kullanmadan önce, bazı ön işleme adımları uygulanması gerekmektedir. Bu adımlar, veri kümesinin eğitim süreci için daha uygun hale getirilmesi ve modellerin daha iyi performans göstermesine yardımcı olmak amacıyla gerçekleştirilmiştir. Önceden eğitilmiş bir ağdan faydalanmak için veri setinin giriş görüntüleri, ön eğitimde kullanılan veri seti ile tutarlı olacak şekilde normalize edilmektedir. Veri normalizasyonu, modelin eğitimi sırasında her bir giriş parametresinin (pikselin) benzer bir veri dağılımına sahip olmasını sağlamak için önemli bir adımdır. Bu da modelin daha hızlı bir şekilde yakınsamasını sağlamaktadır. Normalizasyon genellikle görüntülerin ortalaması her pikselden çıkarılarak başlamaktadır. Ardından bu değerler görüntülerin standart sapmasına bölünmektedir. Bu sayede pikseller arasında daha tutarlı bir ölçeklendirme elde edilebilmektedir. [99] çalışmasına göre, SoccerNet-V3 görüntüleri [0.485, 0.456, 0.406] ortalama ve [0.229, 0.224, 0.225] standart sapma değerleriyle normalize edilmelidir. Bu değerler [0, 255] aralığından [0, 1] aralığına ölçeklendirilmiştir. SoccerNet-V3 veri setindeki görüntülerin boyutları birbirine uymamaktadır. Tüm giriş görüntülerinin aynı boyuta ve en boy oranına sahip olmasını sağlamak için tüm giriş görüntüleri 244×224 boyutuna değiştirilmiştir.

Ayrıca görüntüler, eğitim sırasında 0,5 olasılıkla rastgele yatay olarak çevirmektedir. Bu yöntemin oldukça iyi sonuç verdiği gözlenmiştir. Rastgele Silme ise bir görüntüyü iyileştirmek için başka bir yöntemdir. Rastgele Silme, görüntüde rastgele bir dikdörtgen alanı gürültüyle değiştirir [66]. Rastgele silme veri artırma işlemi, giriş görüntüsünün yarı boyutunda bir alana 0,5 olasılıkla uygulanmaktadır. Son olarak da Rastgele kırpma işlemi 0,5 olasılıkla görüntüler üzerinde uygulanmıştır. Batch boyutu, kimlik başına 4 görüntü ile 32 olarak ayarlanmıştır. Modeli eğitmek için bir optimize edici olarak RR optimize edici kullanılmıştır. Momentum 0.9'a ve ağırlık düşüşü $5e-4$ 'e olarak ayarlanmıştır. Modelin eğitimi 120 dönem boyunca gerçekleştirilmiştir.

4.6.3 Performans Değerlendirme Metrikleri

Bu bölüm, seçili veri seti üzerindeki Futbolcu Yeniden Tanımlaması görevi için önerdiğimiz modelin performansını değerlendirmeye yönelik değerlendirme ölçütlerini açıklamaktadır.

Futbolcu Re-ID sürecinde temel ve önerilen modelin performansını değerlendirmek için [33]'da tanıtılan sıralamaya dayalı değerlendirme yönteminin aynısı benimsenmiştir. Bu yöntem, tüm galeri görüntülerinin öznitelik vektörlerini bir sorgu temsiliyle eşleştirir (kosinüs benzerliğine göre) ve ardından karşılık gelen benzerlik puanlarını azalan düzende sıralayarak gerçekleştirir. Re-ID performansını ölçmek için Bölüm 2.7.2'de açıklanan en sık kullanılan 2 değerlendirme ölçütü, yani CMC ve mAP kullanılmıştır.

CMC, sorgu görüntüsü ile galeri görüntülerinin benzerliklerine veya uzaklıklarına dayanarak galeri görüntülerinin sıralamasını yapmayı ifade etmektedir. CMC, her sorgu için yalnızca 1 doğru sonuç varsayımına dayanmaktadır. Bu nedenle, doğru bir değerlendirme yöntemi olarak kabul edilmektedir. Ancak galeri birden fazla doğru sonuç içeriyorsa, genel performansın değerlendirilmesinde mAP metriği kullanılır. mAP, sorguların doğru sonuçlara olan ortalama hassasiyetini hesaplayarak birden fazla doğru sonuç durumunda daha kapsamlı bir değerlendirme sağlamaktadır.

4.6.4 DeneySEL Kurulum

DeneySEL aşama, önerilen futbolcu Re-ID yöntemini eyleme geçirmeyi içermektedir. Hem temel modelin hem de önerilen modelinin uygulama ayrıntılarını ve tüm eğitim ve değerlendirme deneylerini içermektedir. Futbolcu Yeniden Tanımlamak için önerilen yöntemin ana deneyleri şunlardır:

- 1- SoccerNetV3 veri setini kullanarak önerilen modeli eğitmek.
- 2- Modelin tahmin doğruluğunu artmak için kayıp fonksiyonların birleşimini kullanmak.

Önerilen modelin uygulama ayrıntıları: Bölüm 4.2.2'de önerilen model, derin öğrenme için açık kaynaklı bir Python kitaplığı olan PyTorch çerçevesi kullanılarak uygulanmıştır. Deneyler, NVIDIA GeForce GTX 4080 Ti GPU'lu bir kişisel bilgisayarda gerçekleştirilmiştir. Kullanılan işletim sistemi 64 bit Windows'tur. Önerilen modelin eğitimi için mini-batch SGD ve birleştirilmiş kullanılmıştır. Momentum değeri 0,9 olarak ayarlanmıştır. Weight decay değeri $5e-4$ olarak belirlenmiştir. Modelin eğitimi 120 epoch süresince gerçekleştirilmiştir ve mini-batch boyutu 32 olarak ayarlanmıştır.

DENEYSEL SONUÇLAR VE TARTIŞMA

Bu bölümde, önerilen yöntemin kişi ve futbolcu Re-ID'si için performansının doğrulanması ve değerlendirilmesi amacıyla çeşitli sistemli deneyler gerçekleştirilmiştir. Tüm deneyler, Bölüm 4.6.1'te verilenlere göre uygulanmıştır. Deneyler, Market-1501 ve SoccerNetV3 veri setleri kullanılarak gerçekleştirilmiştir. Tablo 5.1'de iki farklı veri kümesi olan Market-1501 kişi Re-ID'si için ve SoccerNetV3'in futbolcu Re-ID'si için detayları yer almaktadır. Market-1501, 750 kişiye ait toplam 12396 eğitim görüntüsü, 751 kişi ait 3368 sorgu görüntüsü ve 19372 galeri görüntüsünü içermektedir. SoccerNetV3 ise 161443 farklı futbolcuya ait 248234 eğitim görüntüsü, test için ise 30059 farklı futbolcuya ait 11777 sorgu görüntüsü ve 34989 galeri görüntüsü içermektedir.

Tablo 5.1 Bu tez boyunca deneylerde kullanılan veri kümelerine ilişkin ayrıntılar.

Veri seti	Kimlik Sayısı		Görüntü Sayısı		
	Eğitim	Test	Eğitim	Query	Gallery
Market-1501	750	751	12396	3368	19372
SoccerNetV3	161443	30059	248234	11777	34989

Bu bölümde, temel ve RFES-ReID modellerinin eğitimi ve değerlendirilmesi için yapılan deneysel sonuçlar, bölüm 4.6.3'te açıklanan değerlendirme ölçütlerine göre sunulmuştur. İlk olarak veri kümeleri üzerinde temel modelin eğitim ve değerlendirmesinin sonuçları sunulmuştur. Ardından yeni RFES-ReID önerilen modelin eğitim ve değerlendirmesinin sonuçları, bölüm 4.4 ve 4.5'te bahsedilen kayıp ve optimizasyon yöntemleriyle sunulmuştur ve performans iyileştirmesine nasıl katkıda bulunduğunu göstermek için bir değerlendirmesi yapılmıştır. Ayrıca önerilen yöntemin üstünlüğünü ortaya koymak adına önerilen yöntem kullanılarak elde edilen bulgular, son teknoloji yöntemler ile karşılaştırılmıştır.

Tablo 5.2'de kişi ve futbolcu Re-ID'sinde önerilen yöntemi değerlendirmek için kullanılan deneysel ayarları özetlemektedir. Deneylere geçmeden önce, deneylerde kullanılan veri setleri ve bu veri setinin hazırlanması için gerekli olan ön işleme süreçleri kısaca anlatılmaktadır.

Tablo 5.2 Önerilen yöntemin hiper parametreleri.

Hiper parametreler	Ayarlar
Omurga	Swin-T
Optimizer	SGD
Ortalama	[0,485 0,456 0,406]
Standard Sapma	[0,229 0,224 0,225]
Feature dimension	1000
Training epoch	120
Batch size	32
Base learning rate	0,00035
Random erasing augmentation	0,5
Random horizontal flipping	0,5
Random cropping	0,5
Cross entropy loss	(epsilon, scale) = (0,1, 1)
Triplet loss	(margin, scale) = (0, 1)
Focal Tversky loss	Market1501: (α , β , γ) = (0,7, 0,3, 0,75)
	SoccerNet-V3: (α , β , γ) = (0,7, 0,3, 0,95)

5.1 Önerilen Yöntemin Eğitim Sonuçları

Bu bölüm, temel ve önerilen modelin eğitim sonuçlarını sunmaktadır. Eğitim sonuçları, kişi ve futbolcu Re-ID veri kümelerini kullanan hem eğitim hem de doğrulama verilerinin doğruluğunu ve kaybını içermektedir. Önerilen yaklaşımın etkinliğini ve genelliğini kanıtlamak amacıyla bir kişi ve futbolcu Re-ID'si için yeni temel model olan ve bir transformer türü olarak tanımlanan Swin mimarisi kullanılmıştır. Her ikisinin de sonuçları burada sunulmuştur. Önerilen temel modelin Re-ID performansını değerlendirmek için 2 veri setinden sorgu ve galeri görüntüleri kullanılmıştır. Sorgu ve galeri kişi görüntüleri temel modelden geçilerek öznitelikler elde edilmektedir. Daha sonra sıralanmış ve Kosinüs benzerliğine göre benzerlikleri hesaplanmıştır. Sonuçlar CMC eğrileri ve mAP kullanılarak sunulmuştur. Karşılaştırma için ilk olarak birleştirilmiş kayıp ve RR optimizasyon prosedürleri eklenmeden veri setlerini kullanarak temel modelin

değerlendirmesi yapılmıştır. Ardından birleşmiş kaybın ve RR'nin önerilen temel model üzerindeki etkinliği araştırılmıştır.

Kişi Re-ID'si için temel ve önerilen modelin eğitimi sırasında eğitim süreciyle ilgili mevcut yaklaşım, her kişinin kimliğinin ayrı bir sınıf olarak kabul edildiği bir sınıflandırma görevi olarak düşünülmektedir. Her kimliğin ilk görüntüsü bir doğrulama seti olarak kullanılıp geri kalan örnekler bir eğitim seti olarak kabul edilmiştir. Doğrulama setlerindeki sınıflandırma doğruluğu sonuçları aşağıdaki Tablo 5.3'de görülebilmektedir. Market-1501 veri kümesinde Swin'in doğruluk oranı %91,21 iken, RFES-ReID'nin doğruluk oranı %95,95'tir. SoccerNet-V3 veri kümesinde Swin'in doğruluk oranı %87,75 iken, RFES-ReID'nin doğruluk oranı %94,32'dir. Eğitim aşamasında sonuçlar, RFES-ReID'nin temel modelden daha iyi performans gösterdiğini göstermektedir.

Tablo 5.3 Doğrulama hassasiyeti

Omurga	Veri seti	
	Market-1501	SoccerNet-V3
Swin	91,21%	87,75%
RFES-ReID	95,95%	94,32%

5.2 Önerilen Yöntemin Değerlendirme Sonuçları

Kişi ve futbolcu yeniden tanımlama görevlerinde önerilen modelin etkinliğini göstermek için öncelikle temel modelin performansı değerlendirilmiştir. Bu değerlendirme, birleşik kayıp ve RR prosedürlerinin eklenmediği veri kümesi üzerinde gerçekleştirilmiştir. Daha sonra önerilen modele birleşik kayıp ve RR prosedürlerinin eklenmeden nasıl bir etkisi olduğu araştırılmıştır. Son olarak önerilen modele sırasıyla birleşik kayıp ve RR prosedürlerinin eklenmesinin etkinliği incelenmiştir.

Temel ve önerilen modelin birleştirilmiş kaybı ve RR'siz performansı, önerilen modelin birleştirilmiş kaybı olan ve RR'siz performansı ve önerilen modelin birleştirilmiş kaybı ve RR ile Market-1501 ve SoccerNet-V3 veri kümelerindeki performansı **Hata! Başvuru k aynağı bulunamadı.** de verilmiştir.

Tablo 5.4 Değerlendirme sonuçları

Yöntem			Market-1501		SoccorNet-v3	
	Fusion Loss	RR	R1	mAP	R1	mAP
Swin-T	-	-	94.15	83.00	75.34	80.16
+ RFEB	-	-	94.73	83.54	79.63	81.39
	✓	-	95.7	87.3	81.8	85.0
	✓	✓	96.2	89.1	84.1	86.7

Önerilen temel modelin etkinliğini ve genelliğini doğrulamak için kişi ve futbolcu ReID’inde temel model olarak Swin mimarisi ve RFES mimarileri uygulanmıştır. Yukarıda sunulan bulgular, önerilen temel modelin oldukça başarılı bir performans sergileyebildiğini göstermektedir.

Birleştirilmiş Kaybı olan ve RR'siz önerilen modelin performansı: Burada bölüm 4.4'te belirtilen Kayıp yöntemlerinin, yani Füzyon kaybının, önerilen temel modelin üzerindeki etkisine ilişkin sonuçlar sunulmaktadır. Market-1501 ve SoccerNet-V3 veri kümeleri için elde edilen Rank-1 ve mAP, Tablo **Hata! Başvuru kaynağı bulunamadı.**'te ve daha detaylı bir şekilde Tablo 5.5’de gösterilmektedir.

Tablo 5.5 Kayıp fonksiyonlarının etkinliği

Yöntem	Kayıp	Market-1501		SoccorNetV3	
		R1	mAP	R1	mAP
Swin+ RFEB	Cross-entropy	92,0	79,31	74,63	78,96
	Triplet	94,73	83,54	79,63	81,39
	Focal	94,89	86,78	80,06	84,25
	Cross-entropy + Triplet	94,86	84,23	79,87	82,54
	Cross-entropy + Focal	94,96	86,81	80,45	84,58
	Triplet + Focal	95,23	87,04	80,92	84,71
	Fusion	95,7	87,3	81,8	85,0

Füzyon Kaybı ekledikten sonra, Swin-T temelli önerilen model Market-1501 veri kümesinde Rank-1'de %0,5 ve mAP'de %3,61, SoccerNet-V3 veri kümesinde ise Rank-1'de %2,19 ve mAP'ta %3,63 iyileştirme sağlamıştır.

Temel modelin RR ile performansı: Burada bölüm 4.5'te belirtilen optimizasyon yönteminin, RR'nin, önerilen modelin verimliliği üzerindeki etkisine ilişkin sonuçlar sunulmaktadır. Market-1501 ve SoccerNet-V3 veri kümeleri için elde edilen Rank-1 ve mAP, Tablo **Hata! Başvuru kaynağı bulunamadı.**'un son satırında ve Tablo 5.6'da gösterilmektedir.

Tablo 5.6 Optimizasyon etkinliği

Method	optimization	Market-1501		SoccerNetV3	
		R1	mAP	R1	mAP
Swin+ RFEB	Without RR	95,7	87,3	81,8	85,0
	With RR	96,2	89,1	84,1	86,7

Tablo 5.6, RFES-ReID yöntemi için yapılan Re-Ranking etkinliğini değerlendirmektedir. RR varlığı veya yokluğuna bağlı olarak, Market-1501 ve SoccerNetV3 veri kümeleri için elde edilen R1 (en yüksek doğruluk oranı) ve mAP (ortalama doğruluk oranı) sonuçları görülmektedir. Optimizasyon yapılmadan önce Swin+RFEB yöntemi Market-1501 veri setinde %95,7 R1 ve %87,3 mAP elde etmiştir. SoccerNetV3 veri setinde ise %81,8 R1 ve %85,0 mAP değerleri elde edilmiştir. Optimizasyonun uygulandığı durumda, Market-1501 veri setinde %96,2 R1 ve %89,1 mAP, SoccerNetV3 veri setinde ise %84,10 R1 ve %86,7 mAP değerleri elde edilmiştir.

Sonuçlar, optimizasyonun Swin+RFEB yöntemi için hem Market-1501 hem de SoccerNetV3 veri kümelerinde performansı artırdığını göstermektedir. R1 ve mAP değerlerindeki iyileşmeler, önerilen yöntemin kişi ve futbolcu Re-ID görevlerinde etkili olduğunu ve doğruluk oranını artırdığını göstermektedir. Tablo 5.6, optimizasyonun yöntem performansına etkisini açıkça göstererek, önerilen yöntemin uygulanabilirliğini ve başarı potansiyelini vurgulamaktadır.

Kısaca, önerilen yöntemde Füzyon Kaybı ve RR optimizasyonu eklendikten sonra performans değerlendirme sonuçları, önerilen yöntemin kişi ve futbolcu ReID gerçekleştirilmesinde etkili olduğunu göstermiştir.

RR tekniğinin etkinliğinin görsel olarak sunulması için rastgele bir sorgu görüntüsü seçilmiş ve RR uygulanmadan önceki Rank-5 için galeri görüntüleri Şekil 5.1'de gösterilmiştir. Bu görüntüler, önerilen yöntemin başlangıçta elde ettiği iyi sonuçları temsil etmektedir. Ancak RR'nin uygulanmasıyla birlikte performansın daha da iyileştiği gözlemlenmiştir.

RR uygulanmadan önceki galeri görüntüleri incelendiğinde, bazı sorgu görüntülerinin doğru eşleşmeleri içeren galeri görüntülerine sahip olduğu görülmektedir. Bununla birlikte bazı sorgu görüntülerinin yanlış eşleşmeleri içeren galeri görüntülerine de sahip olduğu fark edilmektedir. RR'nin uygulanmasından sonra galeri görüntülerinin daha tutarlı ve doğru eşleşmeler içerdiği gözlemlenmektedir. Bu, RR'nin önerilen yöntemle eklenmesinin doğru eşleşmelerin daha iyi sıralanması ve hatalı eşleşmelerin azaltılması açısından etkili olduğunu göstermektedir. Sonuç olarak, görsel olarak sunulan Şekil 5.1, RR tekniğinin önerilen yöntemin performansını daha da geliştirdiğini ve doğru eşleşmelerin belirlenmesinde daha güvenilir bir yaklaşım sağladığını göstermektedir.



RR'dan önce



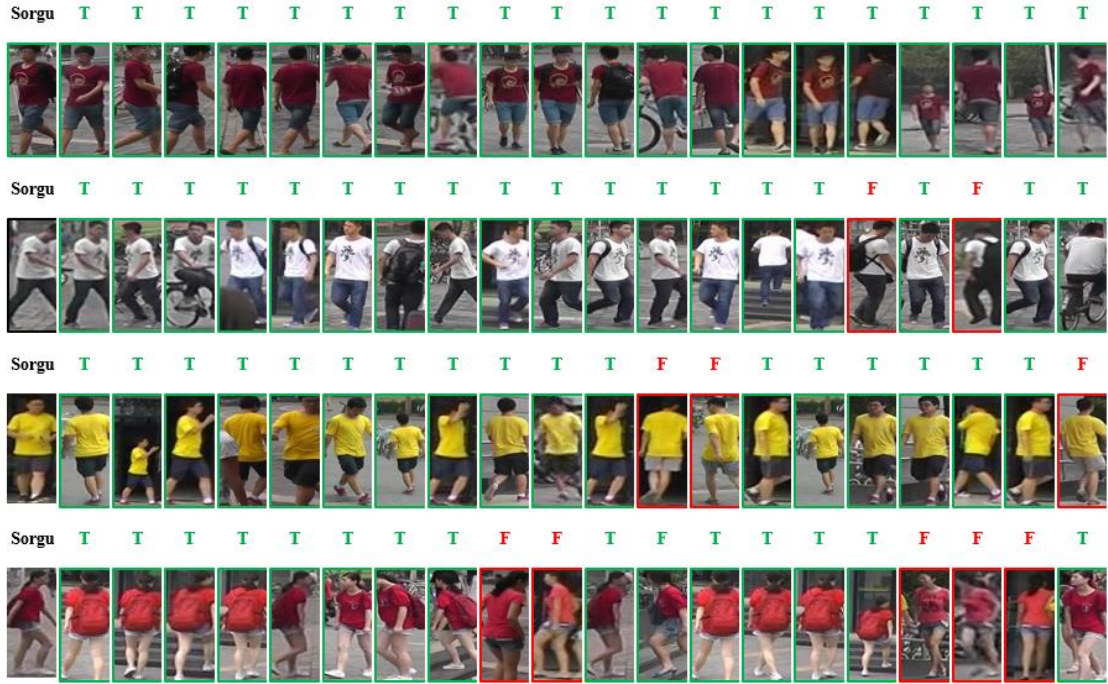
RR uygulandıktan sonra

Şekil 5.1 RR uygulamasından önce ve sonra rank-5 sonuçları örneği.

5.3 Kişi Yeniden Tanımlamanın Niteliksel Analizi

Önerilen yöntemin etkinliğini görsel olarak sunmak için Şekil 5.2, rastgele seçilen bazı sorgu görüntülerini ve bu görüntülerin Market-1501 galeri setinden alınan ilk 20 görüntüsünü, benzerlik puanlarına göre soldan sağa doğru sıralayarak göstermektedir. SoccerNet-V3 için aynı görsel sunumu yapamadık çünkü yeterli sayıda görüntü mevcut değildir. Önerilen yöntemin, ışıklandırma, pozlar veya arka plandaki farklılıklar gibi

etkenlere rağmen kişilerin doğru bir şekilde yeniden tanınmasında başarılı olduğu gözlemlenmiştir. Ancak benzer giysilere sahip kişiler arasında kafa karışıklığı yaşanabileceği unutulmamalıdır ve bu da kişi yeniden tanımlama görevindeki en büyük zorluklardan biri olarak kabul edilir.



Şekil 5.2 Önerilen yöntemin uygulanması yoluyla Market 1501 veri kümesine ait belirli sorgular için alınan ilk 20 sonuç. Sorgular en soldaki sütunda gösterilirken, galeriden alınan görseller benzerlik puanlarına göre soldan sağa doğru sıralanır. Her görüntünün üst kısmındaki T (yeşil) ve F (kırmızı) harfleri, doğru ve yanlış eşdeğerlerini göstermektedir.

5.4 Son Teknoloji Yöntemlerle Karşılaştırma

Sonuç olarak önerilen modelin Market-1501 ve SoccerNet-V3 veri setlerinde elde ettiği nihai performans, mevcut en son teknolojiye sahip DL yaklaşımlarıyla karşılaştırılmıştır. Karşılaştırma için yöntemler CNN tabanlı ve Transformer tabanlı olmak üzere omurgalarına göre 2 gruba ayrılmıştır. Elde edilen sonuçlar Tablo 5.7 ve Tablo 5.8'de sunulmuş olup, CMC Rank-1 doğruluğu ve mAP ölçütleri kullanılarak değerlendirilmiştir.

Tablo 5.7 Market-1501 veri kümesinde, en son derin öğrenme yaklaşımlarıyla performansın karşılaştırılması.

Yöntem	Omurga	Boyut	Rank-1	mAP
ResNet50	CNN	256×128	88,84	71,59
OSNet		256×128	94,8	84,9
PCB[124]		384×128	93,8	81,6
ABDNet[163]		384×128	95,6	88,3
SAN[164]		256×128	96,1	88,0
PGFA[132]		256×128	91,2	76,8
MGN[165]		384×128	95,7	86,9
RGA-SC[46]		256×128	96,1	88,4
Baseline	ViT-B/16	256×128	94,7	86,8
TransReID[139]		256×128	95,0	88,2
TransReID-SSL[166]		256×128	96,2	91,3
Baseline	DeiT-B/16	256×128	94,4	86,6
TransReID[139]		256×128	94,7	88,0
Baseline	Swin-T	224×224	94,1	86,0
RFES -ReID (Ours)		224×224	96,2	89,1

Yukarıdaki tablo, Market-1501 veri kümesinde en son derin öğrenme yaklaşımlarının performansını karşılaştırmaktadır. Yöntemler, kullanılan omurga türüne (CNN, Transformer veya Swin) ve elde edilen Rank-1 doğruluk oranı ile ortalama hassasiyet (mAP) değerlerine göre sıralanmıştır.

Tablodan görülebileceği gibi, farklı omurga türleri arasında performans farklılıkları bulunmaktadır. CNN tabanlı yöntemlerin (ResNet50, OSNet, PCB, ABDNet, SAN, PGFA ve MGN) çoğu yüksek doğruluk oranlarına sahipken, Transformer tabanlı yöntemler (Baseline ViT-B/16, TransReID, TransReID-SSL, Baseline DeiT-B/16 ve TransReID) genellikle daha yüksek doğruluk oranları ve mAP değerleri sergilemektedir. Swin tabanlı RFES-ReID (Ours) yöntemi ise diğer yöntemlerle rekabet edebilecek başarılı sonuçlar göstermektedir. Önerilen yöntem, mevcut çalışmaların birçoğunu aşarak %96,2 Rank-1 doğruluk ve %89,1 mAP elde etmiştir. Basline (Swin-T) ile

karşılaştırıldığında önerilen yöntem, Rank-1 doğruluğunda %2,1 ve mAP'de %3,1 oranında sonuçlarını geçmiştir. Bu sonuçlar, önerilen yöntemin Market-1501 veri kümesindeki etkinliğini ve performansını göstermektedir.

Tablo 5.8 SoccerNet-V3 veri kümesinde, en son derin öğrenme yaklaşımlarıyla performansın karşılaştırılması.

Yöntem	Omurga	Boyut	Rank-1	mAP
ResNet50	CNN	256×128	48,41	59,11
OSNet	CNN	256×128	69,20	76,40
Baseline	ViT-B/16	256×128	68,20	75,70
TransReID[139]	ViT-B/16	256×128	68,6	77,23
TransReID-SSL[148]	ViT/S-16	256×128	83,80	80,10
Inspur[167]	ViT/S-16	256×128	89,41	91,68
Sports Re-ID[152]	ViT/S-16	256×128	81,50	86,00
Baseline	DeiT-B/16	256×128	65,00	73,2
TransReID[139]	DeiT-B/16	256×128	65,4	74,6
Baseline	Swin-T	224×224	75,3	80,2
REFS-ReID (Ours)	Swin-T	224×224	84,1	86,7

Tablo 5.8, SoccerNet-V3 veri kümesindeki derin öğrenme tabanlı yaklaşımlarının performansını karşılaştırmaktadır. Omurga, modelin temel mimarisini ve yapılandırmasını belirleyen önemli bir unsurdur. Yöntemler, kullanılan omurga türüne (CNN, ViT, DeiT ve Swin) göre sıralanmıştır. RFES-ReID (Bizim önerilen yöntem), Swin-T omurgasını temel alarak SoccerNet-V3 veri kümesinde üstün bir performans sergilemiştir. Swin-T omurgası, boyutu 224x224 piksel olan görüntüler üzerinde çalışmaktadır. RFES-ReID, %84,1'lik bir Rank-1 doğruluk oranı ve %86,7'lik bir mAP değeri elde etmiştir. Bu sonuçlar, önerilen yöntemin futbolcu yeniden tanıma görevinde etkili olduğunu ve futbolcuları doğru bir şekilde tanımlayabildiğini göstermektedir. Tablodaki diğer yöntemler arasında ViT/S-16 ve DeiT-B/16 omurgaları dikkat çekmektedir. ViT/S-16 omurgasına dayalı Inspur yöntemi, %89,41'lik bir Rank-1 doğruluk oranı ve %91,68'lik bir mAP değeri ile yüksek performans göstermektedir. Diğer yandan DeiT-B/16 omurgasına dayalı yöntemler daha düşük performans sergilemektedir.

Bu çalışmanın kapsamında, önerilen kişi ve futbolcu Yeniden Tanımlama yönteminin performansı deneysel olarak analiz edilmiş ve tartışılmıştır. Performans değerlendirmesi için Market-1501 ve SoccerNet-V3 olmak üzere iki farklı genel veri seti kullanılmıştır. Elde edilen bulgular, Swin transformer'e eklenen yerel algı bloğunun kişi ve futbolcu yeniden tanımlama performansını artırdığını göstermiştir. Ayrıca geç füzyon tarafından üretilen kayıp, yeniden tanımlama sonuçlarını başarılı bir şekilde iyileştirmiştir. Bu nedenle, birleştirilmiş kayıplar ve RR optimizasyonu eklemek bu işlem için daha yüksek başarı elde edilmesini sağlamıştır.

Önerilen yöntem, SoccerNet-V3 veri setinde elde ettiği iyileştirmelerin Market-1501'deki kadar büyük olmasa da rekabetçi sonuçlar elde etmiştir. Önerilen temel model RGB görüntülerini giriş olarak kullandığından, hatalı sıralanan görüntülerin çoğu sorgu görüntüsüne çok benzeyen galeri görüntülerinden etkilenmiştir. Bu durum, insanların genellikle aynı veya benzer renkte kıyafetler giymesi ve futbolcuların aynı takımda olmaları nedeniyle ortaya çıkan görsel kafa karışıklığından kaynaklanmaktadır.

Sonuç olarak bu çalışmada önerilen yöntemin kişi ve futbolcu yeniden tanımlama sürecindeki performansı değerlendirilmiştir. Swin transformer ve yerel algı bloğunun entegrasyonu, rekabetçi sonuçlar elde edilmesini sağlamıştır.

6.1 Sonuçlar

Kişi Yeniden Tanımlama (Person Re-ID), günümüz gözetim sistemlerinde insan öznelerinin tanımlanması için en dikkat çekici ve en faydalı yaklaşımlardan biri olarak hızla kabul görmektedir. Bu çalışmada, kişi Re-ID'sinin bir alt alanı olarak futbol sporunun yayınlanan videolarında oyuncu yeniden tanımlama alanına odaklanmaktadır. Öznitelikle bir müsabakanın herhangi bir anında farklı kamera bakış açılarından elde edilen görüntülerde aynı oyuncuyu tanımlama hedeflenmektedir. Kişi yeniden tanımlama yaklaşımının karşılaştığı temel zorluklar arasında aydınlatma farklılıkları, farklı bakış açıları, kapanma, bireylerin poz değişiklikleri ve arka plan karmaşası bulunmaktadır. Bu çalışma, kişi yeniden tanımlama konusundaki geleneksel uygulamalardan farklılıklar içermektedir. Öncelikle aynı takımdan olan oyuncular benzer takım forması giydikleri için onları ayırt etmek daha zor hale gelmektedir. İkincisi, her bir kimlik için sınırlı sayıda örnek bulunması, bir yeniden tanımlama sistemi eğitimi daha zor hale getirmektedir. Ayrıca görüntülerin çözünürlükleri genellikle düşük olup ve yoğun örtüşmeler ve oyuncuların hızlı hareketleri gibi faktörlerden dolayı futbolcu yeniden tanımlama için daha fazla zorluk ortaya çıkmaktadır. Son dönemde derin öğrenme yöntemleri kişi Re-ID sistemlerinin performansını artırmada önemli bir rol oynamış ve yukarıda belirtilen zorlukların etkisini kısmen azaltmıştır. Bununla birlikte futbolcu Re-ID eğitimi için gerekli olan ayırt edici öznitelikler elde etme konusunda yeni bir zorluk ortaya çıkmıştır. Bu tez, futbolcu Re-ID için sağlam ve ayırt edici öznitelikler elde edebilen derin modellerin başarılı bir şekilde geliştirilmesine olanak tanıyan yeni bir yöntem önerme motivasyonu üzerine odaklanmaktadır. Geliştirilen yöntem, derin öğrenme tabanlı mevcut yöntemlerin sınırlamalarını aşmak ve futbolcu Re-ID performansını artırmak amacıyla tasarlanmıştır. Öznitelikle futbolcu özniteliklerini vurgulayabilen ve ayırt edici hale getirebilen bir derin modelin tasarlanmasıyla ilgilenmektedir. Bu sayede futbolcu Re-ID eğitimi için daha etkili bir temsil elde edilebilir ve tanıma performansı geliştirilebilir.

Bu tez RFES-ReID adlı bir yöntem geliştirerek futbolcu Re-ID sonuçlarının kalitesini artırmayı amaçlamaktadır. RFES-ReID yöntemi, yerel ve genel özniteliklerin çıkarılması

açısından literatürde önerilen diğer yaklaşımları geride bırakarak Swin-T adlı yeni bir transformer modelini kullanmaktadır. Ayrıca daha ayırt edici öznitelikler elde etmek için CNN'nin avantajlarından yararlanarak Swin Transformer modeline yerel algı bloğu entegre edilmiştir. Daha sonra modelin eğitimi için veri dengesizliği ve benzer görünümlü futbolcuları ayırt edilmesi zorluklarıyla başa çıkabilmek için birleştirilmiş bir kayıp önerilmiştir. Bu süreçte önerilen temel modelle birlikte RR (Re-Ranking) kullanılarak ek performans artışı sağlanmıştır.

Yeni geliştirilen yöntemin etkinliğini değerlendirmek için Market-1501 ve SoccerNet-V3 veri setleri kullanılmıştır. Bu yöntemin performansı, sistemli bir şekilde değerlendirme kriterlerine tabi tutularak değerlendirilmiştir. Deneysel sonuçlar önerilen yöntemin kişi ve futbolcu yeniden kimlik belirleme konusunda rekabetçi bir performans sergileyebileceğini göstermektedir. Bununla birlikte önerilen yöntem hala geliştirilebilir.

Aşağıda, bu tez tarafından yapılan katkılar bölüm 1.4'te vurgulandığı gibi sıralanmıştır.

1. **Nesne sınıflandırmasında kullanılmak üzere geliştirilmiş temel bir Swin transformer modelini futbolcu Re-ID'sinde kullanılmak üzere güçlü ve ayırt edici öznitelikler çıkarmak için uyarlayarak.** Önerilen temel model, bölüm 4.2.1'de ayrıntılı olarak sunulmaktadır. Bölüm 5.1'de sunulan deneysel analizlerde, kişi ve futbolcu Re-ID'si için bu temel modelin üstün sonuçlarını ve etkinliğini doğrulamıştır.
2. **Ardından futbolcu yeniden tanımlama problemi alanına daha uygun olması için RFES-ReID adlı yöntem olarak yerel algı blok yaklaşımını kullanarak Swin transformer modelini geliştirmek.** Bilindiği kadarıyla, halihazırda var olan futbolcu Re-ID için önerilen SoccerNet-v3 veri seti üzerinde çok az çalışma vardır. Mevcut olan çalışmalar arasında RFES-ReID modeli en ayırt edici öznitelikler üretmek. RFES-ReID modeli, bölüm 4.2.2'te ayrıntılı olarak sunulmuştur. Bölüm 5.2'de gösterildiği gibi yürütülen deneyler ve niteliksel ve niceliksel değerlendirmeler, RFES-ReID'in üstün sonuçlarını ve etkinliğini açıkça göstermiştir.
3. **Öğrenmeye dayalı yaklaşımlar genellikle veri dengesizliği ve bir kişiyi güçlü görünüm benzerliğine sahip diğerlerinden ayırt etme konusunda zorluklarla karşılaşır.** Nesne yeniden tanımlama performansını iyileştirmek için yanlış pozitifler ve yanlış negatifler, kayıp fonksiyonunun tasarımındaki ayrılmaz faktörlerin bir parçası olmalıdır. Dolayısıyla veri dengesizliği konusunu ele

alınarak ve bir birleştirilmiş kaybı dahil ederek modelin zor örneklerden etkili bir şekilde öğrenmesini kolaylaştırmak. Bu katkı, önerilen temel modeli eğitmek için önerilen kayıp fonksiyonunun kullanıldığı bölüm 4.4'te sağlanmıştır. Bölüm 5.2'de sunulan deneysel analizler kısmında ise RFES-ReID ve önerilen kaybın birleştirerek iyi genelleme yapmak ve aşırı uyumu azaltmak elde edilen sonuçların kişi ve futbolcu yeniden kimlik belirleme süreci için değerini göstermiştir.

- 4. Önemli bir performans artışı elde etmek ve fazla uydurmayı önlemek için yeniden sıralamayı (RR) bu yeni önerilen temel modelle birleştirmek.** Bu katkı, tahminin doğruluğunu artırmak için yeniden sıralama optimizasyon yöntemi kullanılarak bölüm 4.5'te sağlanmıştır. Bölüm 5.2'te, kişi yeniden kimlik belirleme işlemi için RFES-ReID üzerinde RR yönteminin etkinliği gösterilmiştir.

6.2 Öneriler

Gelecekteki çalışmalar için önerilen temel modelin performansını daha da iyileştirmek amacıyla daha fazla deneysel analiz yapılabilir. Bu analizlerde hesaplama maliyetini artırmadan, MSA'leri paralel gruplara ayırarak yatay ve dikey şeritlerde öz-dikkati gerçekleştiren ve her belirtecin dikkat alanını verimli bir şekilde genişletebilen CSwin gibi bir Transformer bloğu kullanılabilir. Gelecekteki başka bir potansiyel çalışma, görüntülerin çözünürlüklerini artırmaya yönelik çabaları içerebilir. Bu tür çalışmalar, görüntülerin çözünürlüğünü yükselterek daha yüksek başarı seviyelerine ulaşmayı hedefler. Bu, kişi ve futbolcu yeniden kimlik belirleme sürecinde daha net ve ayrıntılı görüntüler elde etmeyi sağlayabilir. Gelecekteki çalışmalardan bir diğeri, SoccerNet-V3 veri setindeki veri dengesizliği ve örnek sayısı eksikliğiyle başa çıkmak için GAN kullanarak yeni veriler üretmeyi içerebilir. Bu yaklaşım eksik verileri doldurarak ve daha çeşitli örnekler oluşturarak önerilen yöntemin performansını artırabilir. Yeni üretilen veriler kullanılarak eğitilen model, daha iyi tahminler yapabilme yeteneğine sahip olabilir. Ayrıca önerilen yöntem sadece futbolcu Re-ID veri seti üzerinde uygulanmıştır. Ancak bu yöntem diğer takım sporları için de geliştirilebilir. Farklı takım sporlarından veri setleri kullanılarak önerilen yöntemin genelleştirilebilmesi için yeni çalışmalar yapılabilir. Öte yandan, yeniden tanımlamanın temel amacı doğru kimlikleri doğru bir şekilde tahmin etmek ve ilişkilendirmektir. Bu nedenle, önerilen yöntemle elde edilen özniteliklerin sporcu takip mekanizmalarına entegre edilmesi, daha yüksek performans elde etmek için potansiyel sağlayabilir. Bu entegrasyon, spor etkinliklerindeki oyuncu hareketlerini daha doğru bir şekilde izlemeyi ve analiz etmeyi mümkün kılabilir. Böylece

hem gerek zamanlı takip sistemlerinde hem de video analizinde daha iyi sonular elde etmek iin nerilen yntemin sporcu takip mekanizmalarıyla entegre edilmesi nemli bir adım olabilir. Ayrıca mevcut araştırma futbolcu algılama adımlarını gz ardı ederek, kısa vadeli grnt tabanlı futbolcu yeniden kimliklendirme zerine odaklanmıştır. nerilen yntemin gerekleştirlmesi de bir sre gerektirebilir ve bu da gerek dnya uygulamalarının gerek zamanlı gereksinimlerini karşılamada zorluklar yaratabilir. Bu nedenle, gelecekteki araştırmalar ayrıca video tabanlı futbolcu yeniden kimliklendirme modellerini, gerek zamanlı futbolcu kimlięi yeniden tanılamak iin sistemleri ve karmaşıık ortamlarda futbolcu yeniden kimliklendirmelerin gerekleştirlmesini keşfetmeye ynelmelidir. Bu şekilde daha karmaşıık ve dinamik futbol sahalarında gerek zamanlı performans saęlanabilir ve futbolculardaki kimlik deęiřikliklerine daha etkin bir şekilde uyum saęlanabilir.

- [1] B. Lavi, M. F. Serj, and I. Ullah, “Survey on deep learning techniques for person re-identification task,” *arXiv preprint arXiv:1807.05284*, 2018.
- [2] “TRACAB Technologies - Tracab.” <https://tracab.com/products/tracab-technologies/> (accessed Feb. 20, 2023).
- [3] D. Wu *et al.*, “Deep learning-based methods for person re-identification: A comprehensive review,” *Neurocomputing*, vol. 337, pp. 354–371, 2019.
- [4] A. Bedagkar-Gala and S. K. Shah, “A survey of approaches and trends in person re-identification,” *Image Vis Comput*, vol. 32, no. 4, pp. 270–286, 2014.
- [5] A. Solichin, A. Harjoko, and A. E. Putra, “A survey of pedestrian detection in video,” *International Journal of Advanced Computer Science and Applications*, vol. 5, no. 10, 2014.
- [6] Y. Wu, J. Lim, and M.-H. Yang, “Online object tracking: A benchmark,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 2411–2418.
- [7] W. Zajdel, Z. Zivkovic, and B. J. A. Kröse, “Keeping track of humans: Have I seen this person before?,” *Proc IEEE Int Conf Robot Autom*, vol. 2005, pp. 2081–2086, 2005, doi: 10.1109/ROBOT.2005.1570420.
- [8] N. Gheissari, T. B. Sebastian, P. H. Tu, J. Rittscher, and R. Hartley, “Person reidentification using spatiotemporal appearance,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 1528–1535, 2006, doi: 10.1109/CVPR.2006.223.
- [9] L. Zheng, Y. Yang, and A. G. Hauptmann, “Person Re-identification: Past, Present and Future,” Oct. 2016, doi: 10.48550/arxiv.1610.02984.
- [10] K. Wang, H. Wang, M. Liu, X. Xing, and T. Han, “Survey on person re-identification based on deep learning,” *CAAI Trans Intell Technol*, vol. 3, no. 4, pp. 219–227, Dec. 2018, doi: 10.1049/TRIT.2018.1001.
- [11] M. O. Almasawa, L. A. Elrefaei, and K. Moria, “A Survey on Deep Learning-Based Person Re-Identification Systems,” *IEEE Access*, vol. 7, pp. 175228–175247, 2019, doi: 10.1109/ACCESS.2019.2957336.

- [12] H. Wang, H. Du, Y. Zhao, and J. Yan, "A Comprehensive Overview of Person Re-Identification Approaches," *IEEE Access*, vol. 8, pp. 45556–45583, 2020, doi: 10.1109/ACCESS.2020.2978344.
- [13] I. Kviatkovsky, A. Adam, and E. Rivlin, "Color invariants for person reidentification," *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 7, pp. 1622–1634, 2013, doi: 10.1109/TPAMI.2012.246.
- [14] R. Zhao, W. Ouyang, and X. Wang, "Person re-identification by salience matching," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 2528–2535.
- [15] L. Bazzani, M. Cristani, and V. Murino, "Symmetry-driven accumulation of local features for human characterization and re-identification," *Computer Vision and Image Understanding*, vol. 117, no. 2, pp. 130–144, Feb. 2013, doi: 10.1016/J.CVIU.2012.10.008.
- [16] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2197–2206.
- [17] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li, "Salient color names for person re-identification," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8689 LNCS, no. PART 1, pp. 536–551, 2014, doi: 10.1007/978-3-319-10590-1_35/COVER.
- [18] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3586–3593.
- [19] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1116–1124.
- [20] B. Ma, Y. Su, and F. Jurie, "Bicov: a novel image representation for person re-identification and face verification," in *British Machine Vision Conference*, 2012, pp. 11-pages.

- [21] F. Xiong, M. Gou, O. Camps, and M. Sznajder, "Person re-identification using kernel-based metric learning methods," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8695 LNCS, no. PART 7, pp. 1–16, 2014, doi: 10.1007/978-3-319-10584-0_1/COVER.
- [22] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans Pattern Anal Mach Intell*, vol. 24, no. 7, pp. 971–987, Jul. 2002, doi: 10.1109/TPAMI.2002.1017623.
- [23] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int J Comput Vis*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/B:VISI.0000029664.99615.94/METRICS.
- [24] S. Liao and S. Z. Li, "Efficient psd constrained asymmetric metric learning for person re-identification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 3685–3693.
- [25] I. Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biol Cybern*, vol. 61, no. 2, pp. 103–113, 1989.
- [26] C. Schmid, "Constructing models for content-based image retrieval," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 2001, pp. II–II.
- [27] D. Gray and H. Tao, "Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features.," *ECCV (1)*, vol. 2008, pp. 262–275, 2008.
- [28] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2288–2295, 2012, doi: 10.1109/CVPR.2012.6247939.
- [29] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7724 LNCS, no. PART 1, pp. 31–44, 2013, doi: 10.1007/978-3-642-37331-2_3/COVER.

- [30] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith, "Learning locally-adaptive decision functions for person verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3610–3617.
- [31] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, "Local fisher discriminant analysis for pedestrian re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3318–3325.
- [32] W. S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 3, pp. 653–668, 2013, doi: 10.1109/TPAMI.2012.138.
- [33] B. J. Prosser, W.-S. Zheng, S. Gong, T. Xiang, and Q. Mary, "Person re-identification by support vector ranking.," in *Bmvc*, 2010, p. 6.
- [34] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3908–3916.
- [35] W. Chen, X. Chen, J. Zhang, and K. Huang, "A Multi-Task Deep Network for Person Re-Identification," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, pp. 3988–3994, Feb. 2017, doi: 10.1609/AAAI.V31I1.11201.
- [36] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng, "Person re-identification by multi-channel parts-based cnn with improved triplet loss function," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1335–1344.
- [37] H. Chen *et al.*, "Deep Transfer Learning for Person Re-Identification," *2018 IEEE 4th International Conference on Multimedia Big Data, BigMM 2018*, Oct. 2018, doi: 10.1109/BIGMM.2018.8499067.
- [38] D. Li, X. Chen, Z. Zhang, and K. Huang, "Learning deep context-aware features over body and latent parts for person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 384–393.
- [39] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 152–159.

- [40] L. Wu, C. Shen, and A. van den Hengel, "Personnet: Person re-identification with deep convolutional neural networks," *arXiv preprint arXiv:1601.07255*, 2016.
- [41] S. Wu, Y. C. Chen, X. Li, A. C. Wu, J. J. You, and W. S. Zheng, "An enhanced deep feature representation for person re-identification," *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*, May 2016, doi: 10.1109/WACV.2016.7477681.
- [42] T. Xiao, H. Li, W. Ouyang, and X. Wang, "Learning deep feature representations with domain guided dropout for person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1249–1258.
- [43] Z. Zheng, L. Zheng, and Y. Yang, "A Discriminatively Learned CNN Embedding for Person Reidentification," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 14, no. 1, Dec. 2017, doi: 10.1145/3159171.
- [44] W. Zhong, L. Jiang, T. Zhang, J. Ji, and H. Xiong, "Combining multilevel feature extraction and multi-loss learning for person re-identification," *Neurocomputing*, vol. 334, pp. 68–78, Mar. 2019, doi: 10.1016/J.NEUCOM.2019.01.005.
- [45] X. Ni, L. Fang, and H. Huttunen, "Adaptive L2 regularization in person Re-identification," *Proceedings - International Conference on Pattern Recognition*, pp. 9601–9607, 2020, doi: 10.1109/ICPR48806.2021.9412481.
- [46] Z. Zhang, C. Lan, W. Zeng, X. Jin, and Z. Chen, "Relation-aware global attention for person re-identification," in *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, 2020, pp. 3186–3195.
- [47] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [48] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015, doi: 10.1016/J.NEUNET.2014.09.003.
- [49] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998, doi: 10.1109/5.726791.

- [50] M. Hacibeyoglu and M. H. Ibrahim, "Human Gender Prediction on Facial Mobil Images using Convolutional Neural Networks," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 6, no. 3, pp. 203–208, Sep. 2018, doi: 10.18201/ijisae.2018644778.
- [51] Ö. \Inik and E. Ülker, "Deep learning and deep learning models used in image analysis," *Gaziosmanpasa Journal of Scientific Research*, vol. 6, no. 3, pp. 85–104, 2017.
- [52] W. Burger and M. J. Burge, *Principles of digital image processing: fundamental techniques*. Springer Science & Business Media, 2010.
- [53] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," *IEEE Trans Pattern Anal Mach Intell*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015, doi: 10.1109/TPAMI.2015.2389824.
- [54] P. Sermanet, S. Chintala, and Y. LeCun, "Convolutional neural networks applied to house numbers digit classification," in *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*, IEEE, 2012, pp. 3288–3291.
- [55] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Comput*, vol. 29, no. 9, pp. 2352–2449, Sep. 2017, doi: 10.1162/NECO_A_00990.
- [56] A. HODGKIN and A. HUXLEY, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *Bull Math Biol*, vol. 52, no. 1–2, pp. 25–71, Jan. 1990, doi: 10.1016/S0092-8240(05)80004-7.
- [57] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
- [58] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.
- [59] D. Pedamonti, "Comparison of non-linear activation functions for deep neural networks on MNIST classification task," *arXiv preprint arXiv:1804.02763*, 2018.

- [60] B. Xu, N. Wang, T. Chen, and M. Li, “Empirical evaluation of rectified activations in convolutional network,” *arXiv preprint arXiv:1505.00853*, 2015.
- [61] I. Goodfellow, “Deep Learning-Ian Goodfellow, Yoshua Bengio, Aaron Courville-Google Books.” MIT Press. <https://books.google.com/books>, 2016.
- [62] D. P. Kingma and J. L. Ba, “Adam: A Method for Stochastic Optimization,” *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, Dec. 2014, doi: 10.48550/arxiv.1412.6980.
- [63] M. Claesen and B. De Moor, “Hyperparameter search in machine learning,” *arXiv preprint arXiv:1502.02127*, 2015.
- [64] J. N. Van Rijn and F. Hutter, “Hyperparameter importance across datasets,” *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2367–2376, Jul. 2018, doi: 10.1145/3219819.3220058.
- [65] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, Sep. 2014, doi: 10.48550/arxiv.1409.1556.
- [66] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, “Random Erasing Data Augmentation,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 13001–13008, Apr. 2020, doi: 10.1609/AAAI.V34I07.7000.
- [67] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer, 2009.
- [68] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *International conference on machine learning*, pmlr, 2015, pp. 448–456.
- [69] N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
- [70] Y. Gal and Z. Ghahramani, “A theoretically grounded application of dropout in recurrent neural networks,” *Adv Neural Inf Process Syst*, vol. 29, 2016.

- [71] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [72] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, Ieee, 2009, pp. 248–255.
- [73] S. Abdulatif, F. Aziz, K. Armanious, B. Kleiner, B. Yang, and U. Schneider, "Person identification and body mass index: A deep learning-based study on micro-Dopplers," in *2019 IEEE Radar Conference (RadarConf)*, IEEE, 2019, pp. 1–6.
- [74] L. Torrey and J. Shavlik, "Transfer learning. Handbook of research on machine learning applications," *IGI Global*, vol. 3, pp. 17–35, 2009.
- [75] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans Knowl Data Eng*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [76] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *Int J Comput Vis*, vol. 115, pp. 211–252, 2015.
- [77] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans Med Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [78] X. Qian *et al.*, "Pose-normalized image generation for person re-identification," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 650–667.
- [79] Q. Xiao, K. Cao, H. Chen, F. Peng, and C. Zhang, "Cross domain knowledge transfer for person re-identification," *arXiv preprint arXiv:1611.06026*, 2016.
- [80] A. Vaswani *et al.*, "Attention is All you Need," *Adv Neural Inf Process Syst*, vol. 30, 2017.
- [81] "CS 230 - Recurrent Neural Networks Cheatsheet." <https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks> (accessed Mar. 01, 2023).

- [82] P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens, “Studying Stand-Alone Self-Attention in Vision Models.” 2019. Accessed: Mar. 01, 2023. [Online]. Available: <https://research.google/pubs/pub48538/>
- [83] J. L. Ba, J. R. Kiros, and G. E. Hinton, “Layer normalization,” *arXiv preprint arXiv:1607.06450*, 2016.
- [84] A. Dosovitskiy *et al.*, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” Oct. 2020, doi: 10.48550/arxiv.2010.11929.
- [85] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, “Training data-efficient image transformers & distillation through attention,” in *International conference on machine learning*, PMLR, 2021, pp. 10347–10357.
- [86] Q. Wang *et al.*, “Learning Deep Transformer Models for Machine Translation,” *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pp. 1810–1822, Jun. 2019, doi: 10.48550/arxiv.1906.01787.
- [87] A. Baevski and M. Auli, “Adaptive Input Representations for Neural Language Modeling,” *7th International Conference on Learning Representations, ICLR 2019*, Sep. 2018, doi: 10.48550/arxiv.1809.10853.
- [88] Z. Liu *et al.*, “Swin transformer: Hierarchical vision transformer using shifted windows,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10012–10022.
- [89] L. Zheng, H. Zhang, S. Sun, M. Chandraker, Y. Yang, and Q. Tian, “Person re-identification in the wild,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1367–1376.
- [90] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, and J. Jiao, “Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 994–1003.
- [91] Y. Yuan, W. Chen, Y. Yang, and Z. Wang, “In defense of the triplet loss again: Learning robust person re-identification with fast approximated triplet loss and label distillation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 354–355.

- [92] N. Abraham and N. M. Khan, “A novel focal tversky loss function with improved attention u-net for lesion segmentation,” *Proceedings - International Symposium on Biomedical Imaging*, vol. 2019-April, pp. 683–687, Apr. 2019, doi: 10.1109/ISBI.2019.8759329.
- [93] D. Ramachandram and G. W. Taylor, “Deep multimodal learning: A survey on recent advances and trends,” *IEEE Signal Process Mag*, vol. 34, no. 6, pp. 96–108, Nov. 2017, doi: 10.1109/MSP.2017.2738401.
- [94] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part-based models,” *IEEE Trans Pattern Anal Mach Intell*, vol. 32, no. 9, pp. 1627–1645, 2010, doi: 10.1109/TPAMI.2009.167.
- [95] A. Cioppa, A. Delière, S. Giancola, B. Ghanem, and M. Van Droogenbroeck, “Scaling up SoccerNet with multi-view spatial localization and re-identification,” *Sci Data*, vol. 9, no. 1, p. 355, 2022.
- [96] W. R. Schwartz and L. S. Davis, “Learning discriminative appearance-based models using partial least squares,” *Proceedings of SIBGRAPI 2009 - 22nd Brazilian Symposium on Computer Graphics and Image Processing*, pp. 322–329, 2009, doi: 10.1109/SIBGRAPI.2009.42.
- [97] M. Hirzer, C. Beleznaï, P. M. Roth, and H. Bischof, “Person re-identification by descriptive and discriminative classification,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6688 LNCS, pp. 91–102, 2011, doi: 10.1007/978-3-642-21227-7_9/COVER.
- [98] T. Wang, S. Gong, X. Zhu, and S. Wang, “Person re-identification by video ranking,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8692 LNCS, no. PART 4, pp. 688–703, 2014, doi: 10.1007/978-3-319-10593-2_45/COVER.
- [99] Z. Zheng, L. Zheng, and Y. Yang, “Unlabeled samples generated by gan improve the person re-identification baseline in vitro,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 3754–3762.

- [100] S. Karanam, M. Gou, Z. Wu, A. Rates-Borras, O. Camps, and R. J. Radke, “A Systematic Evaluation and Benchmark for Person Re-Identification: Features, Metrics, and Datasets,” *IEEE Trans Pattern Anal Mach Intell*, vol. 41, no. 3, pp. 523–536, Mar. 2019, doi: 10.1109/TPAMI.2018.2807450.
- [101] L. Wei, S. Zhang, W. Gao, and Q. Tian, “Person transfer gan to bridge domain gap for person re-identification,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 79–88.
- [102] A. Deliege *et al.*, “Soccernet-v2: A dataset and benchmarks for holistic understanding of broadcast soccer videos,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4508–4519.
- [103] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, “Shape and appearance context modeling,” *Proceedings of the IEEE International Conference on Computer Vision*, 2007, doi: 10.1109/ICCV.2007.4409019.
- [104] X. Liu, W. Liu, T. Mei, and H. Ma, “A deep learning-based approach to progressive vehicle re-identification for urban surveillance,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9906 LNCS, pp. 869–884, 2016, doi: 10.1007/978-3-319-46475-6_53/FIGURES/8.
- [105] A. Bedagkar-Gala and S. K. Shah, “A survey of approaches and trends in person re-identification,” *Image Vis Comput*, vol. 32, no. 4, pp. 270–286, Apr. 2014, doi: 10.1016/J.IMAVIS.2014.02.001.
- [106] J. Yang, “Review of Image-Based Person Re-Identification in Deep Learning,” *Journal of New Media*, vol. 2, no. 4, p. 137, 2020.
- [107] L. Wu, Chunhua Shen, and A. van den Hengel, “Deep linear discriminant analysis on fisher networks: A hybrid architecture for person re-identification,” *Pattern Recognit*, vol. 65, pp. 238–250, May 2017, doi: 10.1016/J.PATCOG.2016.12.022.
- [108] S. H. Salem Hussin and R. Yildirim, “StyleGAN-LSRO Method for Person Re-Identification,” *IEEE Access*, vol. 9, pp. 13857–13869, 2021, doi: 10.1109/ACCESS.2021.3051723.
- [109] F. Xiong, Y. Xiao, Z. Cao, K. Gong, Z. Fang, and J. T. Zhou, “Good practices on building effective CNN baseline model for person re-identification,” in *Tenth*

- international conference on graphics and image processing (ICGIP 2018)*, SPIE, 2019, pp. 142–152.
- [110] Y. Sun, L. Zheng, W. Deng, and S. Wang, “Svdnet for pedestrian retrieval,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 3800–3808.
 - [111] Z. Zheng, L. Zheng, and Y. Yang, “Pedestrian alignment network for large-scale person re-identification,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 10, pp. 3037–3045, 2018.
 - [112] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, “End-to-end deep learning for person search,” *arXiv preprint arXiv:1604.01850*, vol. 2, no. 2, p. 4, 2016.
 - [113] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, “Joint detection and identification feature learning for person search,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3415–3424.
 - [114] H. Chen *et al.*, “Deep Transfer Learning for Person Re-identification,” *2018 IEEE 4th International Conference on Multimedia Big Data, BigMM 2018*, Nov. 2016, doi: 10.48550/arxiv.1611.05244.
 - [115] S. Ding, L. Lin, G. Wang, and H. Chao, “Deep feature learning with relative distance comparison for person re-identification,” *Pattern Recognit*, vol. 48, no. 10, pp. 2993–3003, Oct. 2015, doi: 10.1016/J.PATCOG.2015.04.005.
 - [116] A. Hermans, L. Beyer, and B. Leibe, “In defense of the triplet loss for person re-identification,” *arXiv preprint arXiv:1703.07737*, 2017.
 - [117] C. Zhao, K. Chen, Z. Wei, Y. Chen, D. Miao, and W. Wang, “Multilevel triplet deep learning model for person re-identification,” *Pattern Recognit Lett*, vol. 117, pp. 161–168, Jan. 2019, doi: 10.1016/J.PATREC.2018.04.029.
 - [118] H. Liu, J. Feng, M. Qi, J. Jiang, and S. Yan, “End-to-end comparative attention networks for person re-identification,” *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3492–3506, Jul. 2017, doi: 10.1109/TIP.2017.2700762.
 - [119] X. Bai, M. Yang, T. Huang, Z. Dou, R. Yu, and Y. Xu, “Deep-Person: Learning discriminative deep features for person Re-Identification,” *Pattern Recognit*, vol. 98, p. 107036, Feb. 2020, doi: 10.1016/J.PATCOG.2019.107036.

- [120] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, “Learning discriminative features with multiple granularities for person re-identification,” *MM 2018 - Proceedings of the 2018 ACM Multimedia Conference*, pp. 274–282, Oct. 2018, doi: 10.1145/3240508.3240552.
- [121] F. Wang, W. Zuo, L. Lin, D. Zhang, and L. Zhang, “Joint learning of single-image and cross-image representations for person re-identification,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1288–1296.
- [122] H. Yao, S. Zhang, R. Hong, Y. Zhang, C. Xu, and Q. Tian, “Deep Representation Learning with Part Loss for Person Re-Identification,” *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 2860–2871, Jun. 2019, doi: 10.1109/TIP.2019.2891888.
- [123] W. Yang, H. Huang, Z. Zhang, X. Chen, K. Huang, and S. Zhang, “Towards rich feature discovery with class activation maps augmentation for person re-identification,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 1389–1398.
- [124] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, “Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline),” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 480–496.
- [125] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. H. Hoi, “Deep Learning for Person Re-Identification: A Survey and Outlook,” *IEEE Trans Pattern Anal Mach Intell*, vol. 44, no. 6, pp. 2872–2893, Jun. 2022, doi: 10.1109/TPAMI.2021.3054775.
- [126] Y. Ge *et al.*, “FD-GAN: Pose-guided Feature Distilling GAN for Robust Person Re-identification,” *Adv Neural Inf Process Syst*, vol. 31, 2018, Accessed: Jan. 01, 2023. [Online]. Available: <https://github.com/yxgeee/FD-GAN>
- [127] S. Tang, M. Andriluka, B. Andres, and B. Schiele, “Multiple people tracking by lifted multicut and person re-identification,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3539–3548.

- [128] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, and Q. Tian, “Pose-driven deep convolutional model for person re-identification,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 3960–3969.
- [129] J. Liu, B. Ni, Y. Yan, P. Zhou, S. Cheng, and J. Hu, “Pose transferrable person re-identification,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4099–4108.
- [130] Y. Suh, J. Wang, S. Tang, T. Mei, and K. M. Lee, “Part-aligned bilinear representations for person re-identification,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 402–419.
- [131] S. Zhang, J. Yang, and B. Schiele, “Occluded pedestrian detection through guided attention in cnns,” in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2018, pp. 6995–7003.
- [132] J. Miao, Y. Wu, P. Liu, Y. Ding, and Y. Yang, “Pose-guided feature alignment for occluded person re-identification,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 542–551.
- [133] W. Li, X. Zhu, and S. Gong, “Person Re-Identification by Deep Joint Learning of Multi-Loss Classification,” *IJCAI International Joint Conference on Artificial Intelligence*, vol. 0, pp. 2194–2200, May 2017, doi: 10.48550/arxiv.1705.04724.
- [134] ‘Cheng ’ “Wang,” Q. Zhang, C. Huang, W. Liu, and X. Wang, “Manacs: A Multi-task Attentional Network with Curriculum Sampling for Person Re-identification,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, Sep. 2018.
- [135] W. Li, X. Zhu, and S. Gong, “Harmonious attention network for person re-identification,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2285–2294.
- [136] Y. Liu *et al.*, “A survey of visual transformers,” *IEEE Trans Neural Netw Learn Syst*, 2023.
- [137] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, “Transformers in Vision: A Survey,” *ACM Computing Surveys (CSUR)*, vol. 54, no. 10s, pp. 1–41, Sep. 2022, doi: 10.1145/3505244.

- [138] H. Chen *et al.*, “Pre-trained image processing transformer,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 12299–12310.
- [139] S. He, H. Luo, P. Wang, F. Wang, H. Li, and W. Jiang, “Transreid: Transformer-based object re-identification,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 15013–15022.
- [140] K. Zhu *et al.*, “Aaformer: Auto-aligned transformer for person re-identification,” *arXiv preprint arXiv:2104.00921*, 2021.
- [141] Y. Li, J. He, T. Zhang, X. Liu, Y. Zhang, and F. Wu, “Diverse Part Discovery: Occluded Person Re-Identification With Part-Aware Transformer,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [142] F. Shen, Y. Xie, J. Zhu, X. Zhu, and H. Zeng, “GiT: Graph Interactive Transformer for Vehicle Re-identification,” *IEEE Transactions on Image Processing*, pp. 1–1, Jul. 2021, doi: 10.48550/arxiv.2107.05475.
- [143] G. Zhang, P. Zhang, J. Qi, and H. Lu, “HAT: Hierarchical Aggregation Transformers for Person Re-identification,” *MM 2021 - Proceedings of the 29th ACM International Conference on Multimedia*, pp. 516–525, Oct. 2021, doi: 10.1145/3474085.3475202.
- [144] T. Zhang *et al.*, “Spatiotemporal Transformer for Video-based Person Re-identification,” Mar. 2021, doi: 10.48550/arxiv.2103.16469.
- [145] X. Liu, P. Zhang, C. Yu, H. Lu, X. Qian, and X. Yang, “A video is worth three views: Trigeminal transformers for video-based person re-identification,” *arXiv preprint arXiv:2104.01745*, 2021.
- [146] D. Fu *et al.*, “Unsupervised pre-training for person re-identification,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 14750–14759.
- [147] M. Caron *et al.*, “Emerging properties in self-supervised vision transformers,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9650–9660.

- [148] H. Luo *et al.*, “Self-supervised pre-training for transformer-based person re-identification,” *arXiv preprint arXiv:2111.12084*, 2021.
- [149] L. Jing and Y. Tian, “Self-Supervised Visual Feature Learning with Deep Neural Networks: A Survey,” *IEEE Trans Pattern Anal Mach Intell*, vol. 43, no. 11, pp. 4037–4058, Nov. 2021, doi: 10.1109/TPAMI.2020.2992393.
- [150] B. Zhang, Y. Liang, and M. Du, “Interlaced Perception for Person Re-identification Based on Swin Transformer,” in *2022 7th International Conference on Image, Vision and Computing (ICIVC)*, IEEE, Jul. 2022, pp. 24–30. doi: 10.1109/ICIVC55077.2022.9886403.
- [151] Q. An, K. Cui, R. Liu, C. Wang, M. Qi, and H. Ma, “Attention-Aware Multiple Granularities Network for Player Re-Identification,” *MMSports 2022 - Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports*, pp. 137–144, Oct. 2022, doi: 10.1145/3552437.3555695.
- [152] B. Comandur, “Sports Re-ID: Improving Re-Identification Of Players In Broadcast Videos Of Team Sports,” *arXiv preprint arXiv:2206.02373*, 2022.
- [153] S. Akan and S. Varlı, “Use of deep learning in soccer videos analysis: survey,” *Multimed Syst*, pp. 1–19, Dec. 2022, doi: 10.1007/S00530-022-01027-0/METRICS.
- [154] K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks,” *Int J Res Appl Sci Eng Technol*, vol. 10, no. 12, pp. 943–947, Nov. 2015, doi: 10.22214/ijraset.2022.47789.
- [155] Y. Sun *et al.*, “Circle loss: A unified perspective of pair similarity optimization,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 6398–6407.
- [156] Z. Zhong, L. Zheng, D. Cao, and S. Li, “Re-ranking person re-identification with k-reciprocal encoding,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1318–1327.
- [157] F. Yu and V. Koltun, “Multi-Scale Context Aggregation by Dilated Convolutions,” *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, Nov. 2015, Accessed: Apr. 28, 2023. [Online]. Available: <https://arxiv.org/abs/1511.07122v3>

- [158] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs,” *IEEE Trans Pattern Anal Mach Intell*, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: 10.1109/TPAMI.2017.2699184.
- [159] D. Hendrycks and K. Gimpel, “Gaussian error linear units (gelus),” *arXiv preprint arXiv:1606.08415*, 2016.
- [160] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.
- [161] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [162] D. Qin, S. Gammeter, L. Bossard, T. Quack, and L. Van Gool, “Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 777–784, 2011, doi: 10.1109/CVPR.2011.5995373.
- [163] T. Chen *et al.*, “Abd-net: Attentive but diverse person re-identification,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 8351–8361.
- [164] X. Jin, C. Lan, W. Zeng, G. Wei, and Z. Chen, “Semantics-Aligned Representation Learning for Person Re-Identification,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 11173–11180, Apr. 2020, doi: 10.1609/AAAI.V34I07.6775.
- [165] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, “Learning discriminative features with multiple granularities for person re-identification,” *MM 2018 - Proceedings of the 2018 ACM Multimedia Conference*, pp. 274–282, Oct. 2018, doi: 10.1145/3240508.3240552.
- [166] H. Luo *et al.*, “Self-Supervised Pre-Training for Transformer-Based Person Re-Identification,” Nov. 2021, doi: 10.48550/arxiv.2111.12084.

- [167] S. Giancola *et al.*, “SoccerNet 2022 Challenges Results,” *MMSports 2022 - Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports*, pp. 75–86, Oct. 2022, doi: 10.1145/3552437.3558545.



TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirileri

1. Akan, Sara, and Songül Varlı. "Reidentifying soccer players in broadcast videos using Body Feature Alignment Based on Pose." In Proceedings of the 2023 4th International Conference on Computing, Networks and Internet of Things, pp. 440-444. 2023.

Makaleler

1. Akan, Sara, and Songül Varlı. "Use of deep learning in soccer videos analysis: survey." Multimedia Systems 29, no. 3 (2023): 897-915.