

**ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

DOKTORA TEZİ

**HAREKETLİ NESNELERDE ALGILANMASI GÜÇ OLAN DEĞİŞİMLERİN
VIDEO BÜYÜTME YÖNTEMİYLE TESPİTİ VE DERİN ÖĞRENME İLE
ANALİZİ**

ABDULLAH ASIM YILMAZ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

**ANKARA
2020**

Her hakkı saklıdır

ÖZET

Doktora Tezi

HAREKETLİ NESNELERDE ALGILANMASI GÜÇ OLAN DEĞİŞİMLERİN VIDEO BÜYÜTME YÖNTEMİYLE TESPİTİ VE DERİN ÖĞRENME İLE ANALİZİ

Abdullah Asım YILMAZ

Ankara Üniversitesi

Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. İman ASKERBEYLİ

Hareketli objelerin tespiti ve analizi işlemleri, son yıllarda hem akademik hem de ticari dünyadaki potansiyeli nedeniyle görüntü işleme, bilgisayarla görme ve yapay zeka alanında dikkat çeken önemli bir çalışma alanıdır. Bu tez çalışması kapsamında; obje tespiti, video büyütme yaklaşımları ve derin öğrenme yöntemleri bir araya getirilerek obje tespiti ve analizi yaklaşımları alanında optimize bir model üretilmiştir. Bu çalışmadaki amaç video kareleri arasındaki farkların analizine dayalı olarak objelerin tespitini, izlenmesini sağlamak ve sonrasında obje hareketlerini izleyen video içerisinde analiz edilecek bölgeye; video büyütme yöntemi ile yaklaşım işlemi gerçekleştirilerek o bölge üzerinde derin öğrenme yöntemi ile video analiz işlemleri gerçekleştirmektir. Bu kapsamda çalışma içerisinde üretilen model, hibrit bir modele dayalı yeni bir derin öğrenme temelli mimari önermektedir. Mimari temel olarak transfer öğrenme metodolojisi ve genetik algoritmanın kullanılması ile önceden eğitilmiş ağlara dayalı olarak tasarlanmıştır ve üç ana aşamadan oluşmaktadır. Bunlar sırasıyla; obje tespiti ve tanımlama, video büyütme ve video analizi işlemleridir. Burada önerilen yöntemin ana katkıları; literatürde yaygın olarak kullanılan, önceden eğitilmiş dört ağ modelini optimize edilmiş bir şekilde birleştiren bir hibrit model önermesi ve öznel optimizasyonu işlemi için yeni bir sezgisel algoritma modellemesi geliştirilmesidir. Kapsamlı veri setleri dayanılarak gerçekleştirilen performans karşılaştırmaları, önerilen mimarinin literatürdeki en gelişmiş mimarilerden daha iyi sonuç verdiğini göstermiştir.

Eylül 2020, 123 sayfa

Anahtar Sözcükler: Objeler tespiti ve analizi, derin sinir ağları, derin öğrenme, transfer öğrenimi, genetik algoritma, makine öğrenmesi, yapay zeka

ABSTRACT

Ph. D. Thesis

DETECTION OF SUBTLE CHANGES ON MOVING OBJECTS WITH VIDEO MAGNIFICATION METHOD AND ANALYSIS WITH DEEP LEARNING

Abdullah Asım YILMAZ

Ankara University
Graduate School of Natural and Applied Sciences
Department of Computer Engineering

Supervisor: Prof. Dr. İman ASKERBEYLİ

Detection of moving objects and analysis operations are remarkable an important working field in image processing, computer vision and artificial intelligence due to its potential in both academic and commercial worlds in recent years. Within the scope of this thesis study, an optimized model has been produced in the field of object detection and analysis by combining object detection, video magnification approaches and deep learning methods. The aim of this study is to ensure the detection and tracking of the objects based on the analysis of the differences between the video frames, and then the process of approaching the region to be analyzed in the video following the object movements is carried out by using the video magnification method and performing video analysis operations on that region with the deep learning method. In this context, the model produced in the study proposes a new deep learning based architecture based on a hybrid model. The architecture is basically designed based on pre-trained networks using transfer learning methodology and genetic algorithm and consists of three main stages. These are object detection and recognition, video magnification and video analysis operations, respectively. The main contributions of the method proposed here are the development of a hybrid model proposition that combines four well-known pre-trained network models in an optimized way and the development of a new genetic algorithm modeling for the feature optimization process. Performance comparisons based on extensive data sets showed that the proposed architecture yields better results than most advanced architectures in the literature.

September 2020, 123 pages

Keywords: Object detection and analysis, deep neural networks, deep learning, transfer learning, genetic algorithm, machine learning, artificial intelligence

TEŞEKKÜR

Çalışmalarımı yönlendiren, araştırmalarımın her aşamasında bilgi, öneri ve yardımlarını esirgemeyen danışman hocam Prof. Dr. İman ASKERBEYLİ'ye (Ankara Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı),

Tez çalışmalarım süresince yapılan tez izleme komitesi toplantılarına katılarak fikir ve önerileriyle çalışmamın olgunlaşmasına yardımcı olan Prof. Dr. Nizami GASİLOV'a (Başkent Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı) ve Doç. Dr. Erkan Bostancı'ya (Ankara Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı),

Değerli yorum ve önerileriyle sağladıkları katkıdan dolayı tez jüri üyeleri; Prof. Dr. Fatih Vehbi ÇELEBİ (Yıldırım Beyazıt Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı) ve Doç. Dr. Mehmet Serdar GÜZEL'e (Ankara Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı),

Bugüne kadar gösterdikleri eşsiz sevgi, destek, fedakarlık ve hoşgörü nedeniyle değerli aileme,

Sağladığı koşulsuz ve sınırsız manevi desteği için Annem Serap YILMAZ'a en içten teşekkürlerimi sunarım.

Abdullah Asım YILMAZ

Ankara, Eylül 2020

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
SİMGELER VE KISALTMALAR DİZİNİ	vi
ŞEKİLLER DİZİNİ	ix
ÇİZELGELER DİZİNİ	xii
1. GİRİŞ	1
1.1 Tezin Kapsamı	2
1.2 Tezin Taslağı.....	4
2. LİTERATÜR İNCELEMESİ	5
2.1 Obje Algılama Yaklaşımları	5
2.2 Video Büyütme Yaklaşımları	10
2.3 Derin Öğrenme Yöntemleri.....	12
2.4 Diğer Yöntemler ve Yaklaşımlar	18
3. OBJE TESPİTİ VE ANALİZİ SİSTEMİNİN GELİŞTİRİLMESİ	29
3.1 Obje Tespiti ve Tanımlama İşlemleri.....	31
3.1.1 Arka plan bulma işlemleri.....	31
3.1.2 Muhtemel obje pozisyonlarının bulunması işlemleri.....	34
3.1.3 Görüntü düzeltme işlemleri.....	37
3.1.4 Tespit edilen konumların belirlenmesi işlemleri	39
3.1.5 Obje tanımlama işlemleri	40
3.2 Video Büyütme İşlemleri	43
3.2.1 Euler (piksel bazlı işleme) yaklaşımı	44
3.2.2 Langrangian (nokta izleme) yaklaşımı	47
3.3 Video Analizi İşlemleri	54
3.3.1 Durum çalışmaları	54
3.3.1.1 Duygu analizi işlemleri	55

3.3.1.2 Hareket ve davranış analizi işlemleri	57
3.3.2 Derin sinir ağı mimarisinin tasarlanması işlemleri	62
3.3.3 Derin sinir ağı mimarisinin eğitilmesi ve optimizasyonu işlemleri.....	72
4. DENEYSEL SONUÇLAR VE KARŞILAŞTIRMA	84
4.1 Kullanılan Veri Setleri ve Özellikleri	84
4.2 Elde Edilen Sonuçların Grafikselleştirilmesi ve Deneysel Karşılaştırılması.....	89
5. SONUÇ	103
KAYNAKLAR	106
ÖZGEÇMİŞ	122

SİMGELER VE KISALTMALAR DİZİNİ

$x^{(i)}$	Eğitim setinden rastgele örneklenmiş yerel yamaları
W_1	Görünür düğümleri gizli düğümlere eşleyen bir ağırlık matrisi
b_1	Gizli eğilim (bias) vektörü
s	Doğrusal olmayan sigmoid fonksiyonu
W_2	Görünür düğümü gizli düğümden yeniden yapılandıran ağırlık matrisi
Z	Düzenleme fonksiyonu (Regularization function)
i	Mini yığın içerisindeki ankorun indeksi
p_i	Ankorun tahmini olasılık değeri
P_i^*	Ankorun yer gerçeği etiketi (Ground-truth label)
t_i	Tahmini sınırlayıcı kutunun 4 parametrelili koordinatını temsil eden vektör
t_i^*	Pozitif ankor ile ilişkili yer gerçeği etiketi
L_{cls}	İki sınıf (nesne/ nesne olmayan) üzerindeki log kaybı değeri
L_{reg}	Gerileme kaybı değeri
X_i	Hedef bölgedeki normalleştirilmiş piksel konumları
y	Çalışılan resim karesindeki lokasyon bilgisi
q_u	Hedefin renk öznitelik olasılığı
$\widehat{p}_u(\widehat{y}_0)$	Taylor genişleme değerleri
k	Monotonik azalan kernel profili
w_i	Ağırlıklandırma veri değerleri
B_i	Tespit için tahmin edilen kutu sınırlaması değeri
$g_i(z)$	$2 * n + 3$ boyutlu bir vektör
ϵ_i	Gaussian gürültüsü değeri
Ψ_p	Yama (patch) değeri
$ \Psi_p $	Ψ_p yamasının alan değeri
$C(p)$	p noktası merkezli Ψ_p yamasının güven değeri
$C(q)$	Önceden atanmış Ψ_p yamasının güven değeri

$\perp$	Ortogonal (dikey) operatörü
n_p	p noktasında ön $\delta\Omega$ ortogonal olan birim vektörü değeri
I_{R^i}	Örnek girdi dokunun (texture) yama değeri
$I_{\partial R^i}$	Örnek girdi dokunun sınır bölgesi (boundary zone)
$\delta()$	Dirac'ın deltası (Dirac's delta)
β_i	Normalize benzerlik ölçek faktörü (Normalized similarity scale factor)
α	Büyültme faktörü değeri (Amplification factor)
λ_c	Uzaysal frekans limiti (Spatial frequency cutoff)

Kısaltmalar

BF	Kaba kuvvet algoritması (Brute force algoritması)
CLAHE	Kontrast sınırlı adaptif histogram eşitleme (Contrast limited adaptive histogram equalization)
CNN	Evrışimli sinir ağı (Convolutional neural network)
CK+	Genişletilmiş cohn-kanade veri seti (The extended cohn-kanade dataset)
DA	Derin otomatik kodlayıcı (Deep auto-encoder)
DPM	Deforme Edilebilir Parça Modeli (Deformable part model)
DVM	Destek vektör makinesi (Support vector machine)
EM	Beklenti Maksimizasyonu (Expectation maximization)
ESD	Enerji spektral yoğunluğu (Energy spectral density)
FCM	Bulanık C-Ortalamalar (Fuzzy c-means)
GPU	Grafik işlemci ünitesi (Graphics processing unit)
GWO	Gri Kurt Optimizasyonu (Grey wolf optimization)
HOG	Yönelimli degradelerin histogramı (Histogram of oriented gradients)

JAFFE	Japon kadın yüz ifadesi veri seti (The japanese female facial expression dataset)
KYM	Kırmızı, yeşil, mavi (Red green blue)
mAP	Ortalama hassasiyet (Mean average precision)
MS	Ortalama Kayma (Mean Shift)
MRF	Markov rastgele alanı (Markov random field)
NMS	Maksimum olmayan bastırma (Non-Maximum suppression)
RANSAC	Rastgele örnek konsensüsü (Random sample consensus)
RBF	Radyo temelli fonksiyon (Radio basis function)
RELU	Düzeltilmiş Doğrusal Birim (Rectified Linear Unit)
RMSE	Kök Ortalama Kare Hatası (Root mean square error)
RPN	Bölge önerme ağı (Region proposal network)
ROI	Bölge önerileri (Regions of interest)
R-CNN	Evrişimli sinir ağı özellikli bölgeler (Regions with convolutional neural network features)
SGD	Stokastik gradyan inişi (Stochastic gradient descent)
SIFT	Ölçekten bağımsız öznitelik dönüşümü (Scale invariant feature transform)
SSIM	Yapısal benzerlik endeksi (Structural similarity index)
SVM	Destek vektör makinesi (Support vector machine)
VATAS	Video analizi ve yörünge tabanlı video açıklama sistemi (Video analysis and trajectory based video annotation system)

ŞEKİLLER DİZİNİ

Şekil 1.1 Geliştirilen obje tespiti ve analizi sisteminin akış şeması.....	3
Şekil 2.1 Video nesne algılaması çalışması akış şeması	7
Şekil 2.2 HOG çalışması yöntemi için genel bakış (Dalal ve Triggs 2005)	8
Şekil 2.3 Boosted cascade çalışması için sonuç resimleri (Freund ve Schapire 1995)	9
Şekil 2.4 Video büyütme çalışması için sonuç resmi (Oh vd. 2018)	11
Şekil 2.5 El bileği kan damarlarını görüntüleme örneği (Wu vd. 2012)	12
Şekil 2.6 R-CNN yöntemi akış şeması (Girshick vd. 2014)	13
Şekil 2.7 Hızlı R-CNN nesne algılama şeması (Ren vd. 2015)	15
Şekil 2.8 Hızlı R-CNN için sonuç resimleri 1 (Ren vd. 2015)	16
Şekil 2.9 Hızlı R-CNN için sonuç resimleri 2	16
Şekil 2.10 YOLO çerçevesi resmi (Redmon vd. 2016)	18
Şekil 2.11 VATAS yöntemi genel bakış (Li vd. 2010)	19
Şekil 2.12 Hareket nesnesi algılama işlemi akış şeması (Li vd. 2010)	20
Şekil 2.13 Bireşimsel olarak oluşturulmuş eğitim görüntüleri örnekleri (Tang vd. 2014)	22
Şekil 2.14 İnsan algılaması çalışması için sonuç resimleri (Liang ve Thorpe 2000)	23
Şekil 2.15 Parametrik olmayan doku sentezi işlemi algoritması genel taslağı (Yılmaz vd. 2019) ...	24
Şekil 2.16 Parametrik olmayan doku sentezi yaklaşımı için sonuç resimleri (Yılmaz vd. 2019)	25
Şekil 2.17 Örnek tabanlı görüntü boyama algoritması için sonuç resimleri (Criminisi, Perez ve Toyama 2004)	27
Şekil 2.18 Yama tabanlı bir sentez algoritması için sonuç resimleri (Liang vd. 2001)	28
Şekil 3.1 Obje tespiti ve analizi sistemi modelinin genel mimarisi	30
Şekil 3.2 Girdi video karesi	32
Şekil 3.3 Arka plan görüntüsü	33
Şekil 3.4 Arka plan çıkarım algoritmaları için hareket maskeleri	33
Şekil 3.5 Arka plan algoritmaları için precision/recall eğrisi	34
Şekil 3.6 Video karesiyle arka plan görüntüsünün farkı	35
Şekil 3.7 Video karesinin ölçeklendirilmiş fark verisi	36
Şekil 3.8 Video karesinin segmentasyon işlemine uğramış fark verisi	36
Şekil 3.9 Fazlalıkları atılmış video karesi	37
Şekil 3.10 Morfolojik işlem uygulanmış video karesi	38
Şekil 3.11 Tespit edilen konumların belirlendiği video karesi	39
Şekil 3.12 Obje tanımlama işlemi sonrası elde edilen görüntü karesi 1	41
Şekil 3.13 Obje tanımlama işlemi sonrası elde edilen görüntü karesi 2	42
Şekil 3.14 Obje tanımlama işlemi sonrası elde edilen görüntü karesi 3	42
Şekil 3.15 Euler yaklaşımı için genel bakış (Wu vd. 2012)	43
Şekil 3.16 Langrangian yaklaşımı için genel bakış (Liu vd. 2005)	44
Şekil 3.17 Laplacian piramidi dönüşümü örneği (Burt ve Adelson 1983, Yılmaz vd. 2017)	45

Şekil 3.18 İnsan nabzını görüntüleme örneği (Wu vd. 2012)	47
Şekil 3.19 Lagrangian yaklaşımı örneği (Liu vd. 2005)	49
Şekil 3.20 Euler, Lagrangian ve diğer temel enterpolasyon yöntemlerinin karşılaştırması grafiği (SSIM)	50
Şekil 3.21 Euler ve Langragian yöntemlerinin karşılaştırılması grafiği (RMSE).....	51
Şekil 3.22 Video büyütme/yaklaşım işlemi sonrası elde edilen görüntü karesi 1	52
Şekil 3.23 Video büyütme/yaklaşım işlemi sonrası elde edilen görüntü karesi 2.....	52
Şekil 3.24 Video büyütme/yaklaşım işlemi sonrası elde edilen görüntü karesi 3.....	53
Şekil 3.25 Video büyütme/yaklaşım işlemi sonrası elde edilen görüntü karesi i 4.....	53
Şekil 3.26 Obje tespiti ve analizi sistemi modeli için duygu analizi işlemleri durum çalışmasının genel boru/akış hattı diyagramı.....	57
Şekil 3.27 Obje tespiti ve analizi sistemi modeli için hareket ve davranış analizi işlemleri durum çalışmasının genel boru/akış hattı diyagramı.....	62
Şekil 3.28 AlexNet ağının temel mimarisi (Krizhevsky vd. 2012, Yılmaz vd. 2020b)	63
Şekil 3.29 VGG-16 ağının temel mimarisi (Simonyan ve Zisserman 2014, Yılmaz vd. 2020b).....	64
Şekil 3.30 GoogleNet ağının temel mimarisi (Szegedy vd. 2015, Yılmaz vd. 2020b)	65
Şekil 3.31 Naif Inception Modülü (Szegedy vd. 2015)	66
Şekil 3.32 Resnet-152 ağının temel mimarisi (He vd. 2016, Yılmaz vd. 2020b)	67
Şekil 3.33 Residual blok diyagramı (He vd. 2016).....	68
Şekil 3.34 Duygu analizi işlemleri durum çalışması için çalışılan sistemin akış şeması gösterimi ..	70
Şekil 3.35 Hareket ve davranış analizi işlemleri durum çalışması için çalışılan sistemin akış şeması gösterimi	71
Şekil 3.36 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 1	79
Şekil 3.37 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 2	79
Şekil 3.38 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 3	80
Şekil 3.39 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 4	80
Şekil 3.40 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 5	81
Şekil 3.41 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 6	81
Şekil 3.42 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 7	82
Şekil 3.43 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 8	82
Şekil 3.44 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 9	83

Şekil 3.45 Eğitilmiş modelimiz ile her bir obje tespiti ve analizi sistemi modelinin genel mimarisinin aşamalarının entegre edilmiş gösterimi	83
Şekil 4.1 Duygu analizi işlemleri durum çalışması için kullanılan veri setleri içerisindeki her bir yüz duygu kategorisine ait örnek resimler (Lyons vd. 1997, Lucey vd. 2010)	85
Şekil 4.2 Hareket ve davranış analizi işlemleri durum çalışması içerisinde insan eylemlerini tanımlama işlemleri için kullanılan veri setlerine ait örnek resimler (Zappi vd. 2008, Anguita vd. 2012, Rodriguez 2010, Schuldt vd. 2004)	88
Şekil 4.3 Hareket ve davranış analizi işlemleri durum çalışması içerisinde el hareketi tanımlama işlemleri için kullanılan veri setlerine ait örnek resimler (Hsiao vd. 2014, Bao vd. 2017)	89
Şekil 4.4 AlexNet, VGG-16, GoogleNet, ResNet152 ve önerilen derin sinir ağı modellerinin JAFFE, CK+ ve Kişisel veri setleri üzerindeki elde edilen accuracy, sensitivity, specificity ve f-score metrik değerleri gösterimleri	95
Şekil 4.5 AlexNet, VGG-16, GoogleNet, ResNet152 ve önerilen derin sinir ağı modellerinin KTH, UCF Sports Action, Skoda ve UCI Smartphone veri setleri üzerindeki elde edilen doğruluk, hassasiyet, özgüllük ve f puanı metrik değerleri gösterimleri.....	96
Şekil 4.6 AlexNet, VGG-16, GoogleNet, ResNet152 ve önerilen derin sinir ağı modellerinin LaRED ve TinyHands veri setleri üzerindeki elde edilen doğruluk, hassasiyet, özgüllük ve f puanı metrik değerleri gösterimleri.....	96
Şekil 4.7 AlexNet, VGG-16, GoogleNet, ResNet152 ve önerilen ağ modellerinin kişisel veri seti üzerindeki çalışılan yedi yüz duygu kategorisi için karışıklık matrisleri gösterimleri.....	97
Şekil 4.8 AlexNet, VGG-16, GoogleNet, ResNet152 ve önerilen ağ modellerinin KTH veri seti üzerindeki altı hareket kategorisi için karışıklık matrisleri gösterimleri.....	98
Şekil 4.9 AlexNet, VGG-16, GoogleNet, ResNet152 ve önerilen ağ modellerinin TinyHands veri seti üzerindeki altı hareket kategorisi için karışıklık matrisleri gösterimleri	99

ÇİZELGELER DİZİNİ

Çizelge 3.1 Derin öğrenme yöntemleri, veri setleri ve öznitelik çıkarıcı birleşimleri için mAP değeri karşılaştırması	40
Çizelge 4.1 Duygu analizi işlemleri durum çalışması için kullanılan veri setleri içerisindeki yüz duygu sınıflarına ait örnek resim sayıları.....	85
Çizelge 4.2 Hareket ve davranış analizi işlemleri durum çalışması içerisinde el hareketi tanımlama işlemleri için kullanılan verisetlerinin karakteristik özellikleri	89
Çizelge 4.3 JAFFE, CK+ ve Kişisel veri setleri için hiperparametre yapılandırmaları gösterimleri	93
Çizelge 4.4 KTH, UCF Sports,Skoda ve UCI Smartphone verisetleri için hiperparametre yapılandırmaları gösterimleri.....	94
Çizelge 4.5 LaRED ve TinyHands verisetleri için hiperparametre yapılandırmaları gösterimleri	95
Çizelge 4.6 Önerilen ağ ve diğer state-of-the-art algoritmalarının sonuçları için CK + veri seti üzerinde elde edilen yüz duygusu sınıflandırma doğruluk değerleri tablosu.....	100
Çizelge 4.7 Önerilen ağ ve diğer state-of-the-art algoritmalarının sonuçları için sırasıyla KTH, UCF Sports Action, Skoda, UCI Smartphone veri setleri üzerindeki hareket/eylem sınıflandırma doğruluk değerleri tabloları	100
Çizelge 4.8 Önerilen ağ ve diğer state-of-the-art algoritmalarının sonuçları için LaRED veriseti üzerinde elde edilen el hareketi tanıma doğruluk değerleri tablosu.....	102
Çizelge 4.9 Önerilen modelimizin genetik algoritma kullanılarak ve kullanmadan elde edilen doğruluk sonuçları için t-testi sonuçları tablosu	102

1. GİRİŞ

Günümüzde birçok dalda yeni teknolojik gelişmeler günden güne artarak gerçekleşmektedir. Bu teknolojik gelişmelerin sonucu olarak, birçok sorunlar veya ihtiyaçlar ortaya çıkmıştır. Bu gibi sorunların veya ihtiyaçların tespit edilmesi ile yaşanabilecek bazı olumsuzluklar veya eksiklikler, çeşitli yöntemler vasıtasıyla en alt seviyeye düşürülebilir. Bundan sebeple; teknolojik ve bilimsel gelişmelere paralel düşme olması gerekirken, trafik kazaları, hırsızlık olayları, tespit edilemeyen şüpheli vakalar ve bunun gibi durumlar hala ülkemizde ve dünyada gündemin en ön sıralarında yer almaktadır. Bu nedenle görüntü işleme alanının üzerinde durduğu konulardan biri olan hareketli objelerin tespiti ve analizi işlemleri çok büyük önem arz etmekte ve üzerinde fazlaca çalışılan bir alan olmaktadır. Obje tespiti ve analizi yaklaşımlarında, art arda gelen görüntü karesi çerçeveleri kullanılarak hareket tabanlı çözümler ile obje tespiti ve teşhisi işlemleri gerçekleştirilmektedir. Bu yaklaşımlar, bir trafik kazasına karışan kişilerin olaydaki kusur durumlarının ortaya konulmasında, şüpheli araçların veya kayıp insanların tespitinde, çeşitli suçlara karışmış insanların ve araç sistemi yazılımlarında faydalanılmak üzere duygu ve hareket analizi işlemlerinde, çeşitli uzay araştırmalarında (gezegenler, uydular, gökyüzü olayları), askeri uygulamalarda (hedef tespiti, insansız hava araçları, gece görüş sistemleri), elverişli tarım alanlarının tespitinde, bitki hastalıklarının teşhisinde ve bunun gibi birçok alanda kullanılmaktadırlar.

Son zamanlarda obje tespiti ve analizi işlemleri içerisinde kullanılan diğer başlıca çalışma alanları video büyütme (resim teleskobu) ve derin öğrenme yaklaşımlarıdır. Bu yaklaşımlar ile çeşitli çalışmalar yapıp bu çalışmalarla birlikte yeni çözümler ve algoritmalar her geçen gün geliştirilerek meydana getirilmektedirler. Bu yaklaşımlar kısaca tanımlanırsa; video büyütme (resim teleskobu) yaklaşımlarında, çıplak gözle görülmesi ve açığa çıkarılması veya gösterilmesi zor olan zamansal değişimleri, elde edilen video görüntüleri veya resimlerin çeşitli ileri görüntü işleme algoritmaları ile iyileştirilerek ilgili hareket değişimlerinin büyütülerek ortaya çıkarılması işlemlerini oluşturmaktadır. İkinci olarak, derin öğrenme yöntemlerini tanımlayacak olursak; bir veya daha fazla gizli katman içeren

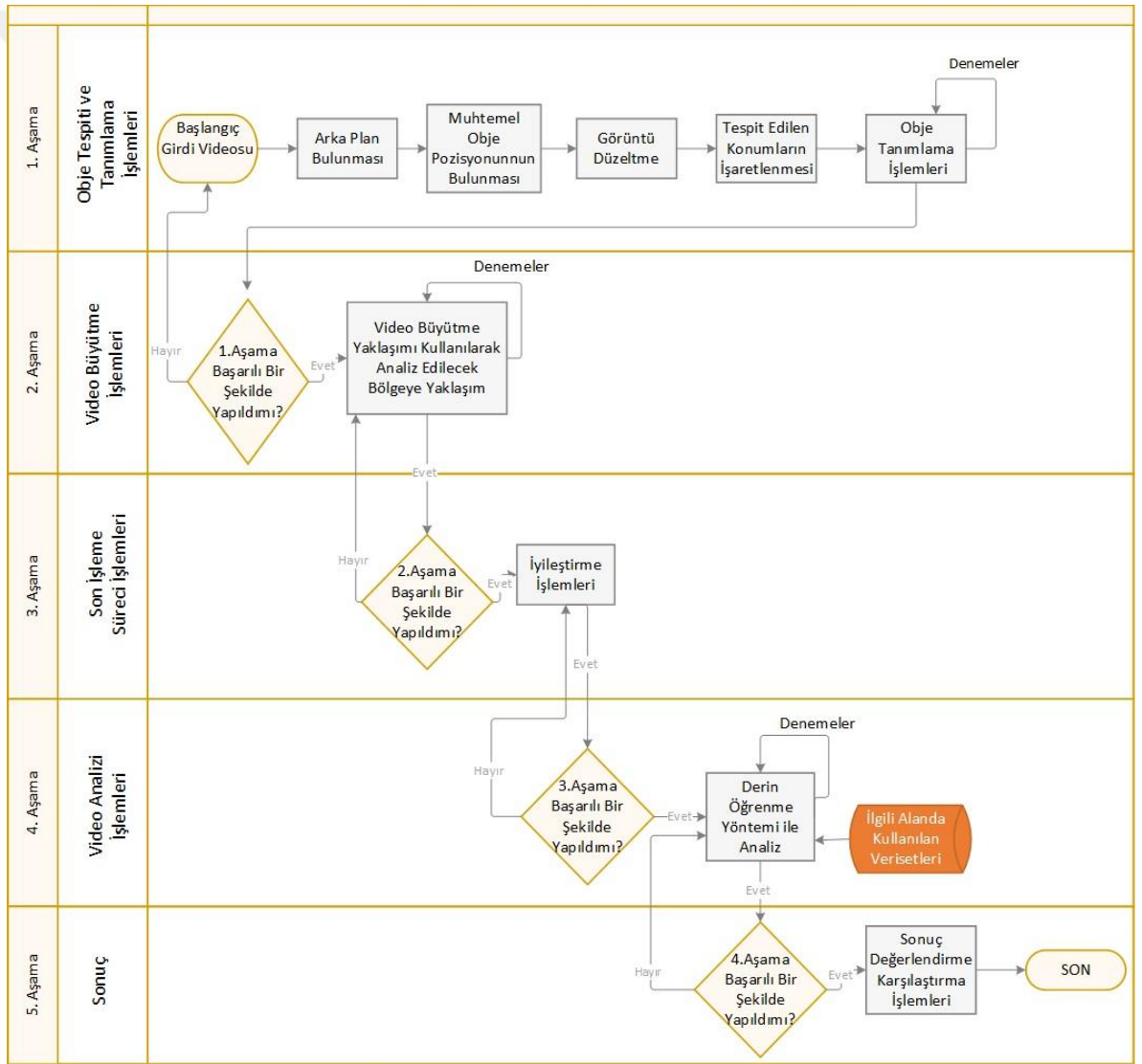
yapay sinir ağı ve benzeri makine öğrenme algoritmalarını kapsayan ve büyük miktarlarda etiketlenmiş eğitim verilerinden özellik saptama yapabilen yöntemlerdir.

1.1 Tezin Kapsamı

Tez çalışması kapsamında hipotezimiz video büyütme yaklaşımları ve derin öğrenme yöntemleri kullanılarak obje tespiti ve analizi işleminin yapılabilirliğinin değerlendirilmesidir. Bu çalışmadaki amaç video çerçeveleri arasındaki farkların analizine dayalı olarak objelerin tespitini, izlenmesini sağlamak ve sonrasında obje hareketlerini izleyen video içerisinde analiz edilecek bölgeye video büyütme yöntemi ile yaklaşım işlemi gerçekleştirilerek o bölge üzerinde derin öğrenme yöntemi ile video analiz işlemleri gerçekleştirmektir. Bu kapsamda çalışma içerisinde üretilen model, hibrit bir modele dayalı yeni bir derin öğrenme temelli mimari önermektedir. Mimari temel olarak transfer öğrenme metodolojisi ve genetik algoritmanın kullanılması ile önceden eğitilmiş ağlara dayalı olarak tasarlanmıştır. Gerçekleştirilen modelimizin literatüre katkıları listelenirse; ilk olarak, literatürde yaygın olarak kullanılan tanınmış önceden eğitilmiş dört ağ modelini optimize edilmiş bir şekilde birleştiren bir melez model geliştirilmiştir. İkinci olarak, çalışma içerisinde modele özgü yeni bir sezgisel algoritma modellemesi önerilen mimariye göre tasarlanmıştır ve üçüncü olarak, bu geliştirilen modelleme ile derin ağ modellerinde öznetelik seçimi/optimizasyonu gerçekleştirilmiştir.

Tez çalışmamızın içerisinde geliştirilen obje tespiti ve analizi sistemi modelinin metodolojisi beş ana aşamadan meydana gelmektedir. Bunlar sırasıyla Şekil 1.1'deki şemadan anlatılırsa; Birinci aşama içerisinde ilk olarak girdi videosu üzerinde arka plan hesaplanmıştır, sonrasında muhtemel obje pozisyonlarının bulunması işlemleri gerçekleştirilmiştir. Muhtemel obje pozisyonlarının bulunması işlemlerinde ilk adım olarak gri tonlara çevrilmiş her bir video karesinden elde edilen arka plan çıkarılmıştır. Sonrasında, ilk adımın sonucunda elde edilen video karesindeki değerler ileride daha optimize sonuçlar alınabilmesi için ölçeklendirilmiştir. Son adım olarak ise segmentasyon işlemleri gerçekleştirilmiştir. Devamında elde edilen sonuçlar üzerinde oluşan fazlalıkların

ve gürültülerin giderilmesi için gaussian filtresi ve eşikleme (thresholding) yöntemi uygulanmış ve obje görüntüleri üzerindeki boşlukların doldurulması içinde morfolojik işlemler gerçekleştirilerek görüntü düzeltme işlemleri yapılmıştır. Bunun sonucunda objelerin videodaki pozisyonları tespit edilmiş ve devamında tespit edilen konumlar sınırlayıcı kutu içerisine alınmıştır. Bu işlem sonrasında, derin öğrenme yöntemleri ile birlikte uygun hazır veri setleri kullanılarak algılanan konumlar üzerinde obje tanımlama işlemleri gerçekleştirilmiştir. İkinci aşamada, objelerin tespit edildiği video üzerinde analiz



Şekil 1.1 Geliştirilen obje tespiti ve analizi sisteminin akış şeması

edilecek bölgeye video büyütme yöntemleri ile yaklaşım işlemi gerçekleştirilmiştir. Bir sonraki aşamada, yaklaşım işlemi gerçekleştirilen analiz edilecek bölge üzerinde iyileştirme işlemleri yapılarak son işleme süreci işlemleri gerçekleştirilmiştir. Dördüncü aşamada, yaklaşılan bölge üzerinde geliştirilen derin öğrenme yöntemi mimarisi ile video analiz işlemleri yapılmıştır. Son aşama olarak, çalışma alanı içerisinde obje tespiti ve analizinde kullanılan diğer farklı geçmiş çalışmalarla da önerilen çalışma değerlendirilmiştir. Tez kapsamında gerçekleştirilen modelimiz ile sağlık hizmeti (health-care), sürüş güvenliği (driving-safety), akıllı ev teknolojileri hizmeti (smart home technology), akıllı gözetim ve güvenlik sistemi hizmetleri (intelligent surveillance and security system service) ve bunun gibi birçok uygulama alanında videolardaki hareketli objelerin algılanması ve analizi işlemlerinin yapılmasına olanak sağlanmaktadır.

1.2 Tezin Taslağı

Tezin 2. Bölümünde, tezimizin temel aldığı obje tespiti, video büyütme yaklaşımları ve derin öğrenme yöntemlerinin literatürdeki genel tanımlamaları yapılmış ve bunlarla ilgili literatürde yer alan geçmiş çalışmalardan kapsamlı özetler halinde bahsedilmiştir. 3. bölümde, obje tespiti ve analizi sistemi önerilmiş ve bu sistemin her bir geliştirilme aşaması bölümler halinde detaylı olarak açıklanmıştır. Ek olarak, geliştirilen obje tespiti ve analizi sistemi için gerçekleştirilen durum çalışmaları anlatılmıştır. 4. bölümde, geliştirilen obje tespiti ve analizi sistemi modeli için uygulanan uygulama detaylarından, deneyler sırasında kullanılan verisetleri ve özelliklerinden, gerçekleştirilen deneysel analiz karşılaştırmaları sonucu elde edilen deneysel sonuçlardan bahsedilmiş ve çalışılan modelin diğer güncel ve geleneksel mimarilere göre performans karşılaştırmaları yapılmıştır. 5.bölümde, tez çalışmasının genel sonuç özeti ve öneriler paylaşılmıştır.

2. LİTERATÜR İNCELEMESİ

Bu bölüm içerisinde tezimizin temel aldığı obje algılama, video büyütme yaklaşımları ve derin öğrenme yöntemleri ile ilgili literatürde yer alan geçmiş çalışmalardan ve tezimiz ile ilgili diğer farklı alanlardaki yöntemlerden veya yaklaşımlardan detaylı özetler halinde bahsedilmiş ve değerlendirmeler yapılmıştır. Bu kapsamda bu bölüm dört alt başlık altında incelenmiştir. Alt başlıklar içerisinde olan bölüm 2.1, bölüm 2.2 ve bölüm 2.3'de sırasıyla obje tespiti, video büyütme yaklaşımları ve derin öğrenme yöntemleri ile ilgili olarak öncelikle literatürdeki genel tanımlamaları yapılmış ve sonrasında literatürde var olan bu alanlardaki geçmiş çalışmalardan kapsamlı olarak bahsedilmiştir. Bölüm 2.4'de tez çalışma metodolojimiz ile ilgili diğer farklı alanlardaki yöntemler veya yaklaşımlar ayrıntılı olarak anlatılmıştır.

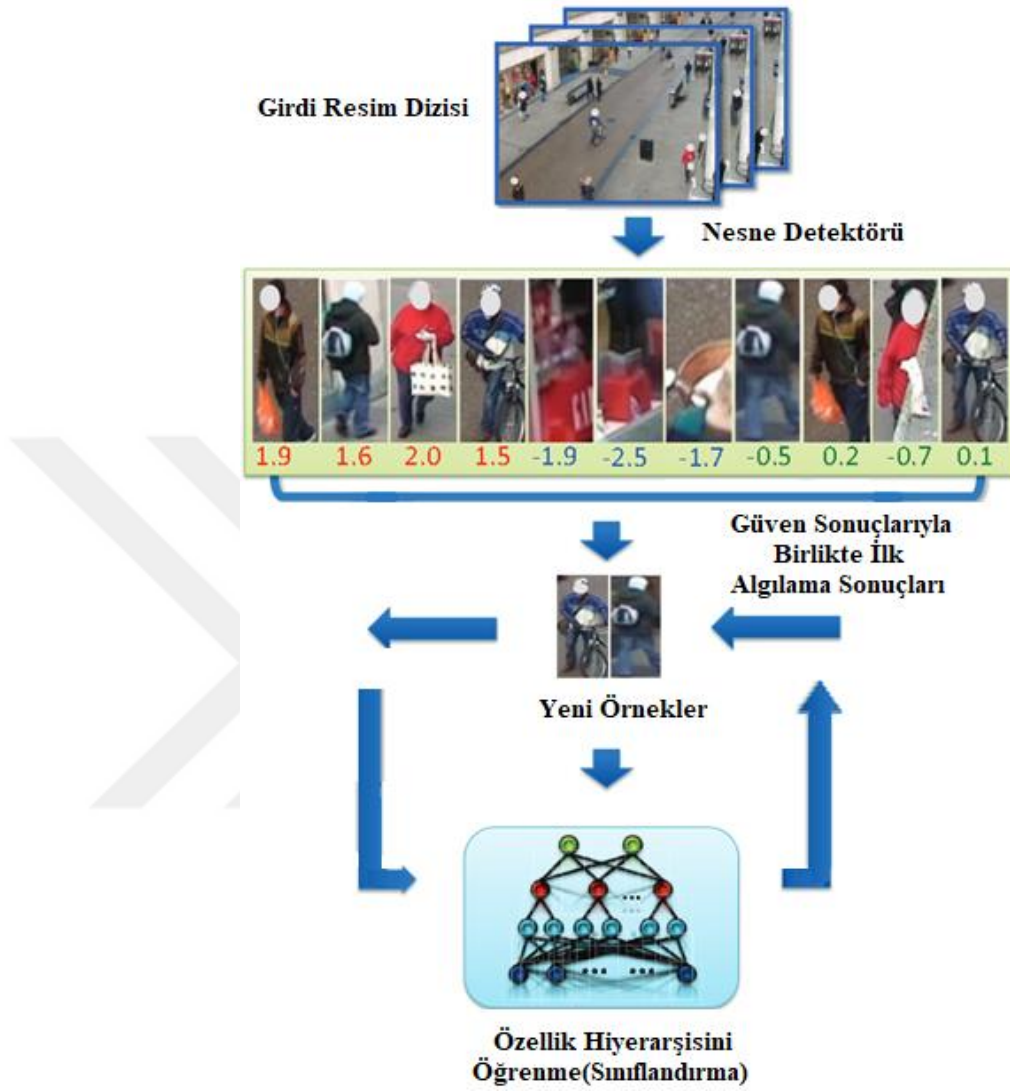
2.1 Obje Algılama Yaklaşımları

Obje algılama yaklaşımları içerisinde art arda gelen görüntü karesi çerçeveleri kullanılarak hareket tabanlı çözümler ile obje tespiti ve teşhisi işlemleri gerçekleştirilmektedir. Burada obje algılama işlemi, dijital görüntü ve videolar içerisindeki belirli sınıfların anlamsal örneklerini (örneğin insanlar, arabalar, hayvanlar, binalar, vb.) belirleme ve bulma işlemlerini ifade eder ve bilgisayarla görme ve görüntü işleme alanları ile ilgili bir problem olarak tanımlanır. Obje algılama yaklaşımları, makine öğrenimi tabanlı yaklaşımlar ve derin öğrenme tabanlı yaklaşımlar olarak iki grupta incelenmektedirler. Makine Öğrenimi yaklaşımlarında, öncelikle SIFT (Scale-Invariant Feature Transform) (Lowe 1999), HOG (Histogram of Oriented Gradients) (Dalal ve Triggs 2005) ve bunun gibi yöntemlerden birisi kullanarak özniteliklerin tanımlanmaktadır ve ardından destek vektör makinesi (DVM) (Cortes ve Vapnik 1995) gibi bir teknik kullanılması ile sınıflandırma yapılarak nesne algılama işlemleri gerçekleştirilmektedir. Diğer yandan, derin öğrenme yaklaşımlarında öznitelikler tanımlanmadan uçtan uca obje tespiti yapılmaktadır ve bu yaklaşımlar genellikle evrişimli sinir ağlarına (CNN) dayanmaktadır (Anonymous 2019a, Jiao vd. 2019).

Yang vd. (2013), videolardaki genel nesne detektörlerinin performansını artırmak için derin sinir ağı kullanılarak videolara özgü özniteliklerin öğrenilmesi yönünde bir video içerisinde nesne algılama ile ilgili bir yaklaşım önermiştir. Önerilen yaklaşımın temel dayandığı fikir, bir videonun farklı karelerinde görünen bir nesnenin, daha iyi dedektörler oluşturmak adına öğrenilebilen benzer öznitelikleri paylaşması gerektiğidir. Önerilen yöntem, izleme yöntemleriyle gerçekleştirilen denetlenen dedektör uyarlamaları (Andriluka vd. 2008, Gavrilu ve Munder 2007, Graner vd. 2008) ve algılama, Wang vd. 2012) aksine, ek açıklamalar gerektirmemektedir veya zamansal yazışmayı kullanmamaktadır. Çalışmada, genel bir detektörden gelen yüksek güven tespitleri işlemlerinin (Javed vd. 2005, Levin vd. 2003, Nair ve Clark 2004, Rosenberg vd. 2005, Roth vd. 2005, Roth vd. 2009 ile işleme başlanır. Devamında yeni videoya özgü öznitelikler iteratif bir şekilde tekrar edilerek öğrenilir. Bu işlemler sonucunda algılama puanları hassas bir şekilde elde edilmiş olur ve video içerisinde nesne algılama işlemi tamamlanır. Şekil 2.1'de çalışılan yöntemin akış şeması gösterilmiştir. Ek olarak çalışma içerisinde, ayırıcı ve kompakt özellikleri öğrenmek adına; derin ağları önceden eğitmek için kullanılan denetimsiz bir öğrenme mimarisi olan otomatik kodlayıcılara dayanan bir özellik öğrenme yöntemi kullanılmıştır. Burada özellik öğrenme yöntemi için aşağıdaki denklem 2.1'de gösterilen matematiksel formülasyon kullanılmıştır.

$$E_{AE} = \sum_1^N \|x^i - W_2 s(W_1 x^{(i)} + b_1) + b_2\|^2 + Z(W_1 x^{(i)} + b_1) \quad (2.1)$$

Bu matematiksel formülasyonun içerisindeki $x^{(i)}$ eğitim setinden rastgele örneklenmiş yerel yamaları, W_1 görünür düğümleri gizli düğümlere eşleyen bir ağırlık matrisini, b_1 gizli bir eğilim (bias) vektörünü, s doğrusal olmayan bir sigmoid fonksiyonunu, W_2 görünür düğümü gizli düğümden yeniden yapılandıran bir ağırlık matrisini, b_2 bir giriş eğilim (bias) vektörüdür. Z bir düzenleştirme fonksiyonunu göstermektedir. Çalışılan yöntemle, insan ve at algılama konusunda ayrıntılı deney sonuçları üzerinde önemli performans artışı elde edebileceği çalışma içerisinde gösterilmiştir (Yang vd. 2013).

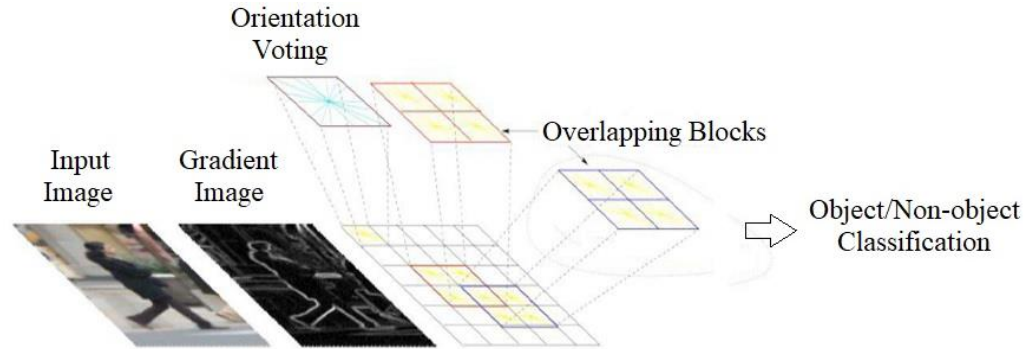


Şekil 2.1 Video nesne algılaması çalışması akış şeması

Dalal ve Triggs (2005) çalışmasında görüntü işlemede ve bilgisayarla görmede nesne tespiti için kullanılan bir özellik tanımlayıcısı yöntemi geliştirmişlerdir. Histogram of Gradient (HOG) olarak adlandırılan yöntem temel olarak bir görüntünün alt hücrelere ve bloklara bölünüp her bir alt bloktaki yön vektörlerinin açisal değerinin belirli açı aralıkları kullanılarak histogram ile hesaplanması esasına dayanmaktadır. Bu teknik bir resmin veya bir resim karesinin belli bir parçasındaki değişim oryantasyonlarını hesaplar (Freeman ve Roth 1995, Freeman vd. 1996). Bu metod SIFT (Lowe 1999) ve şekil içerik (shape context)

(Belongie vd. 2001) gibi bazı öznitelik tanımlama (feature descriptor) yöntemleriyle benzer yönler taşır. Fakat bu teknik düzgün aralıklı hücre ızgaraları arasında çalıştırıldığı için diğer öznitelik tanımlama yöntemlerinden ayrılır.

HOG yöntemi içerisinde yapılan işlemler Şekil 2.2 üzerinden sırasıyla anlatılırsa; ilk olarak renk normalizasyonu ve gamma doğrulaması işlemleri yapılır. Ardından ikinci olarak, konvolüsyon (convolution) metodu ile gradient değerleri hesaplanır. Sonraki aşamada ise her bir piksel ağırlıklı ortalamaya tutularak oryantasyon işlemleri gerçekleştirilir. Daha sonra dördüncü aşama olarak, çakışan blokların önlenmesi amacıyla normalizasyon işlemi gerçekleştirilir ve devamında, her bir video tespit karesindeki HOG bileşenleri toplanır. Son olarak, toplanan HOG bileşenleri destek vektör makinesine sınıflandırma işlemi için verilerek eğitilir ve bu işlem sonrasında videoda veya resim karesinde aranılan nesne türünün olup olmadığı belirlenir. Bu çalışma içerisinde nesne türü olarak yayalar kullanılmıştır ve ayrıca yazarlar ilerleyen çalışmalarında HOG yöntemi üzerindeki testlerini çeşitli hayvan ve araçları dahil edecek şekilde genişletmişlerdir. Burada elde edilen deneysel sonuçlar ile önerilen yaklaşımın geçerliliği bu çalışma içerisinde ispatlanmıştır.



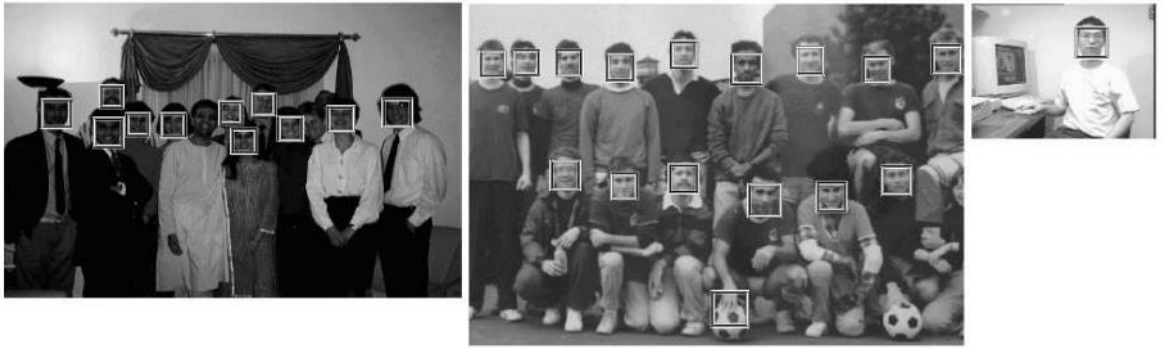
Şekil 2.2 HOG çalışması yöntemi için genel bakış (Dalal ve Triggs 2005)

Viola and Jones (2001), görsel nesne algılama işlemleri için görüntüleri çok hızlı bir şekilde işleyebilen ve yüksek tespit oranlarına ulaşabilen bir makine öğrenmesi yaklaşımı önermiştir. Bu çalışma üç önemli katkı ile ayırt edilmektedir. Birincisi, çalışma içerisinde kullanılan dedektörün kullandığı özelliklerin çok hızlı bir şekilde hesaplanmasını sağlayan,

özellik değerlendirmesine izin veren ve aşağıdaki denklemde matematiksel formülasyonu gösterilen “İntegral Görüntü” adı verilen yeni bir görüntü temsidir (Denklem 2.2).

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (2.2)$$

Denklem 2.2'de gösterilen matematiksel formülasyonun içerisindeki $ii(x, y)$ integral görüntüyü, $i(x', y')$ orijinal resmi göstermektedir. İkincisi, geniş bir kümeden az sayıda kritik görsel öznitelik seçen ve son derece etkili sınıflandırıcılar sağlayan ve AdaBoost algoritmasını temel alan bir öğrenme algoritmasıdır (Freund ve Schapire 1995). Üçüncüsü, fazla karmaşık sınıflandırıcıların bir “kaskad” yapısı içerisinde birleştirilmesi sağlayan “Attentional Cascade” adı verilen yöntemdir; bu yöntem, görüntünün arka plan bölgelerinin hızlıca atılmasına olanak verirken, nesne olma olasılığı konusunda gelecek vaat eden bölgelere daha fazla hesaplama olanağı sağlar ve çalışılan dedektörün hızını önemli ölçüde artırır. Kaskad, önceki yaklaşımlardan farklı olarak, atılan bölgelerin ilgilenilen nesneyi içermesinin olası olmadığı konusunda istatistiksel garantiler sağlayan, nesneye özel bir odaklanma mekanizması olarak görülebilir. Çalışma içerisinde gerçekleştirilen yöntem yüz algılama işlemleri üzerinde denenmiş ve en iyi önceki sistemlerle karşılaştırılabilir algılama oranları elde etmiştir. Çalışmanın sonuçları Şekil 2.3'de gösterilmiştir.



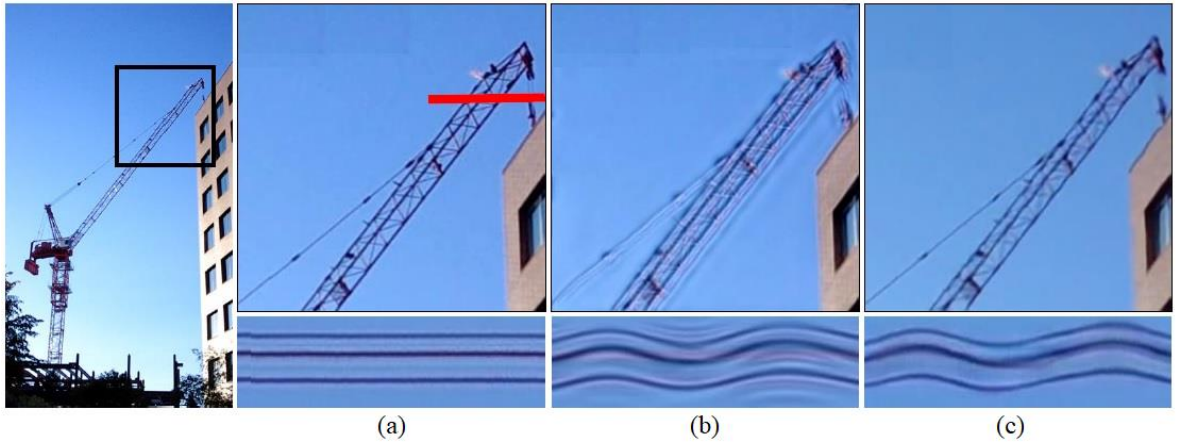
Şekil 2.3 Boosted cascade çalışması için sonuç resimleri (Freund ve Schapire 1995)

2.2 Video Büyütme Yaklaşımları

Video büyütme yaklaşımları, çıplak gözle görülmesi ve açığa çıkarılması veya gösterilmesi zor olan zamansal değişimleri, elde edilen video görüntüleri veya resimlerin çeşitli ileri görüntü işleme algoritmaları ile iyileştirilerek ilgili hareket değişimlerinin büyütülerek ortaya çıkarılması işlemini ihtiva etmektedir. Video büyütme yaklaşımları, temel enterpolasyon tabanlı yaklaşımlar, Euler tabanlı yaklaşımlar, Lagrange tabanlı yaklaşımlar olarak üç grupta incelenmektedirler. Burada temel enterpolasyon tabanlı yaklaşımlarda, görüntü kalitesi kaybı minimum olacak şekilde geometrik dönüşümler kullanılarak dijital resim karelerinin ölçeklendirilmesi işlemleri yapılırken, diğer Euler ve Lagrange tabanlı yaklaşımlarda, gelişmiş görüntü işleme formülasyonları ile videolar içerisinde bulunan insan gözünün tespit edemediği görünemeyen sinyallerdeki hareketler ortaya çıkarılmaktadır ve bu hareketler üzerinde büyütme işlemleri gerçekleştirilmektedir (Anonymous 2019b, Kaş ve Tanyıldızı 2017).

Oh vd. (2018) çalışmalarında video hareket büyütmesi işlemi için ilk öğrenme tabanlı yaklaşım önerilmiştir. Mevcut video hareket büyütme teknikleri, video karelerini genel olarak hareketi büyütmelerine izin veren temsiller olarak ayrıştırır (Wadhwa vd. 2013, Wadhwa vd. 2014, Wu vd. 2012, Zhang vd. 2017). Ayrışmaları genel olarak, optimal olmayan elle tasarlanmış filtreler (örnek olarak karmaşık yönlendirilebilir filtreler (complex steerable filters) (Freeman ve Adelson 1991)) dayanır. Bu yaklaşımda video hareket büyütme teknikleri içerisinde literatürde kullanılan el ile tasarlanmış filtreleri kullanmak yerine, filtreler doğrudan derin evrimsel sinir ağlarını kullanan örneklerden öğrenilmiştir ve bu öğrenilen filtreler çalışma içerisinde kullanılmıştır. Çalışmadaki amaç hareketi sadece çarpma işlemi ile manipüle etmek ve büyütülmüş çerçeveyi yeniden oluşturmak için kullanabilecek bir görüntü temsili çıkaran bir ağ tasarlamaktır. Bu nedenle çalışmadaki ağ üç bölümden oluşmaktadır. Bu bölümler kodlayıcı, manipülatör, kod çözücü'dür. İlk olarak, kodlayıcı, hareketi manipüle etmek için kullanabileceğimiz ve görüntü temsili (Jones ve Poggio 1998) tek bir kareden çıkarabilen bir mekansal ayrışma filtresi görevi görür. İkinci olarak, manipülatör, kodlayıcı tarafından elde edilen görüntü

temsili gösterimini alır ve hareketi büyütmek için farkı çarparak manipüle eder. Son olarak, kod çözücü, manipülatör tarafından değiştirilmiş görüntü temsilini, elde edilen hareket büyütmeli çerçevelere yeniden yapılandırır. Ek olarak, burada eğitimi izlenebilir hale getirmek için, küçük hareketleri iyi yakalayan sentetik bir veri kümesi tasarlanmıştır ve eğitim için iki çerçeveli girdi kullanılmıştır. Çalışma içerisinde öğrenilen filtrelerin, önceki yöntemlerden daha az zil artefaktları (ringing artifacts) ve daha iyi gürültü karakteristiği ile gerçek videolarda yüksek kaliteli sonuçlar elde ettiği gösterilerek önerilen yaklaşımın geçerliliği ispatlanmıştır. Çalışmanın sonucu oluşan sonuç resim karesi Şekil 2.4'de gösterilmiştir. Burada Oh vd. (2018) çalışmasındaki yöntem ile elde edilen büyütülmüş resim karesi (Şekil 2.4.c), (Wadhwa vd. 2013) çalışmasındaki yöntem ile elde edilen büyütülmüş resim karesine (Şekil 2.4.b) göre daha az zil artefaktı ve gürültü içerdiği gözlemlenmektedir.

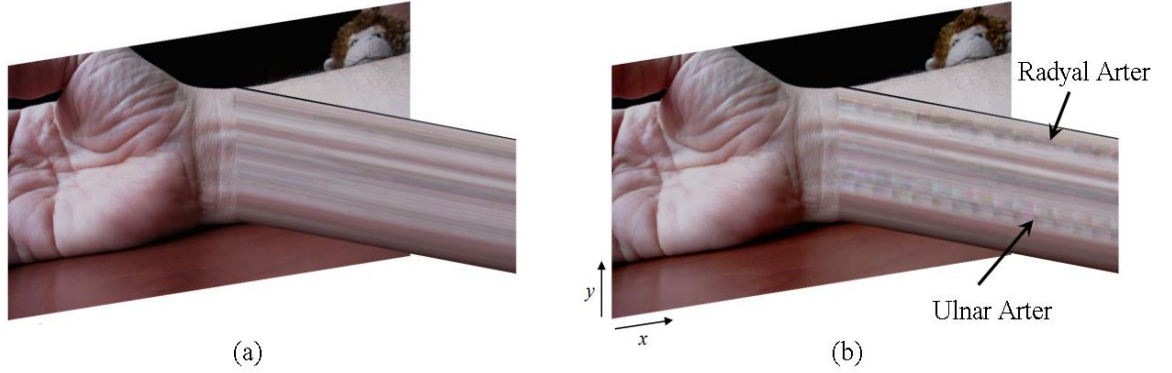


Şekil 2.4 Video büyütme çalışması için sonuç resmi (Oh vd. 2018)

(a) Girdi Resim Karesi (Kırmızı çizgi, x-t ve y-t dilim görünümünü çizmek için örneklenmiş bölgeleri göstermektedir), (b) Wadhwa vd. (2013) çalışmasındaki yöntem ile büyütülmüş resim karesi, (c) Oh vd. (2018) çalışmasındaki Yöntem ile Büyütülmüş Resim Karesi

Wu vd. (2012) çalışmalarında dünyanın en önemli yönleri arasında yer alan gözle görülmeyen değişimlerin, videoların uzaysal ve zamansal işlenmesinin birleşimi ile büyütülerek ortaya çıkarılması işlemi gerçekleştirilmektedir. Önerilen yaklaşım dört ana aşamadan meydana gelmektedir. Önerilen yaklaşımda ilk adım olarak girdi video dizilerini

farklı uzaysal frekans bantlarına ayırıştırma işlemi yapılmaktadır. İkinci adım olarak, ayırıştırılan bantlara aynı zamansal filtre uygulanmıştır. Devamındaki adımda, filtrelenmiş bantlar sonrasında belirlenen büyültme değerine göre büyütülmüştür. Son aşamada ise büyütülen bantlar orijinal sinyalin arkasına eklenmiştir ve sonunda daraltılarak çıkış videosunu meydana getirilmiştir. Çalışmanın sonucu oluşan sonuç resim karesi Şekil 2.5'de gösterilmiştir. Burada kan akışından kaynaklanan kan damarlarının hemen göze çarpmayan ince hareketleri girdi resim karesinde (Şekil 4.a) belli belirsiz görülürken, hareket büyütülme işlemi sonrası elde edilen sonuç resim karesinde (Şekil 4.b) bu hareketler önemli ölçüde daha belirgin görülebilir hale gelmişlerdir.



Şekil 2.5 El bileği kan damarlarını görüntüleme örneği (Wu vd. 2012)

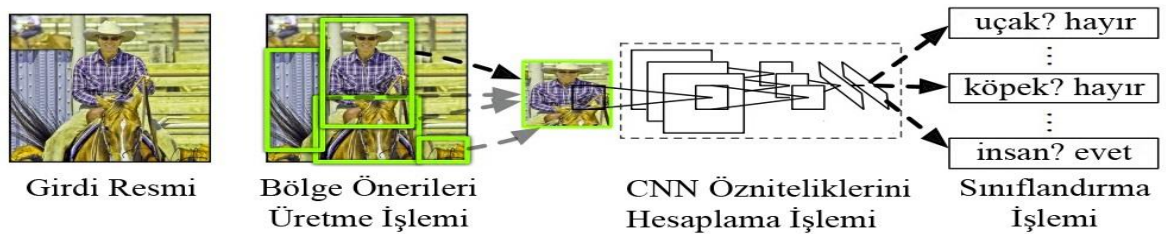
(a) Girdi resim karesi, (b) Büyütülmüş resim karesi

2.3 Derin Öğrenme Yöntemleri

Derin öğrenme yöntemleri, temeli yapay sinir ağlarına dayanan, birden çok doğrusal olmayan işlem birimi katmanını kullanarak ham girdiden üst düzey özneteliklerin aşamalı olarak çıkartılmasını ve dönüştürülmesini sağlayan bir makine öğrenme yöntemidir. Derin öğrenme algoritmaları gözetimli, yarı gözetimli veya gözetimsiz olarak gerçekleştirilmektedirler (Anonymous 2020a, Şeker vd. 2017, Bengio vd. 2013).

Girshick vd. (2014) çalışmalarında bir derin öğrenme yöntemi olarak kullanılan bir nesne tespit sistemi yöntemi önermiştir. R-CNN (Regions with Convolutional Neural Network Features) olarak adlandırılan bu çalışmada, Pascal-VOC veri seti (Everingham vd. 2010) 2012'den önceki en iyi sonuca göre mAP'ı (Mean Average Precision) % 30 dan fazla arttıran basit ve ölçeklenebilir bir tespit algoritması tasarlanmıştır ve % 53.3'lük bir mAP'a ulaşılmıştır. Önerilen yöntem iki temel anlayışı birleştirmektedir. Birincisi nesneleri lokalize etmek ve parçalamak için yüksek kapasiteli konvolüsyonel sinir ağlarının aşağıdan yukarıya bölge tekliflerine uygulanması işlemidir. İkincisi; etiketli eğitim verileri yetersiz olduğunda ise, alana özgü ince ayarlarla yardımcı bir görev için yapılan denetimli bir ön eğitim yapılmasının izlenmesidir. Bu durumda da önemli bir performans artışı sağlamaktadır. Bu çalışmada, diğer bölge önerisi methodlarının (Carreira ve Sminchisescu 2012, Alexe vd. 2012, Arbelaez vd. 2014, Cireşan vd. 2013, Endres ve Hoiem 2010) yerine seçmeli arama (Uijlings vd. 2013) yöntemi kullanılmıştır.

Şekil 2.6'de çalışılan yöntemin akış şeması gösterilmiştir. Buna göre, sistem temel olarak üç modül üzerinden çalışmaktadır. İlk modülde, alınan girdi verileri üzerinden kategoriden bağımsız bölge önerileri (ROI (Regions of Interest)) seçmeli arama algoritması kullanılarak üretilmektedir. İkinci modül içerisinde ilk olarak bölge önerisi alanı olarak belirlenen her bir veri kırılır. Devamında, her bir kırılmış bölge önerisi büyük bir konvansiyonel sinir ağına gönderilir ve bunun sonucunda sabit uzunlukta öznitelik vektörleri elde edilir. Burada kullanılan konvansiyonel sinir ağı, beş adet konvolüsyon katmanı ve iki adet tam bağlı katman'dan oluşmaktadır. Üçüncü ve son modülde, ikinci modülde elde edilen öznitelikler sınıflandırma işlemi için DVM'ne verilir ve sonrasında nesne algılama işlemi tamamlanır.

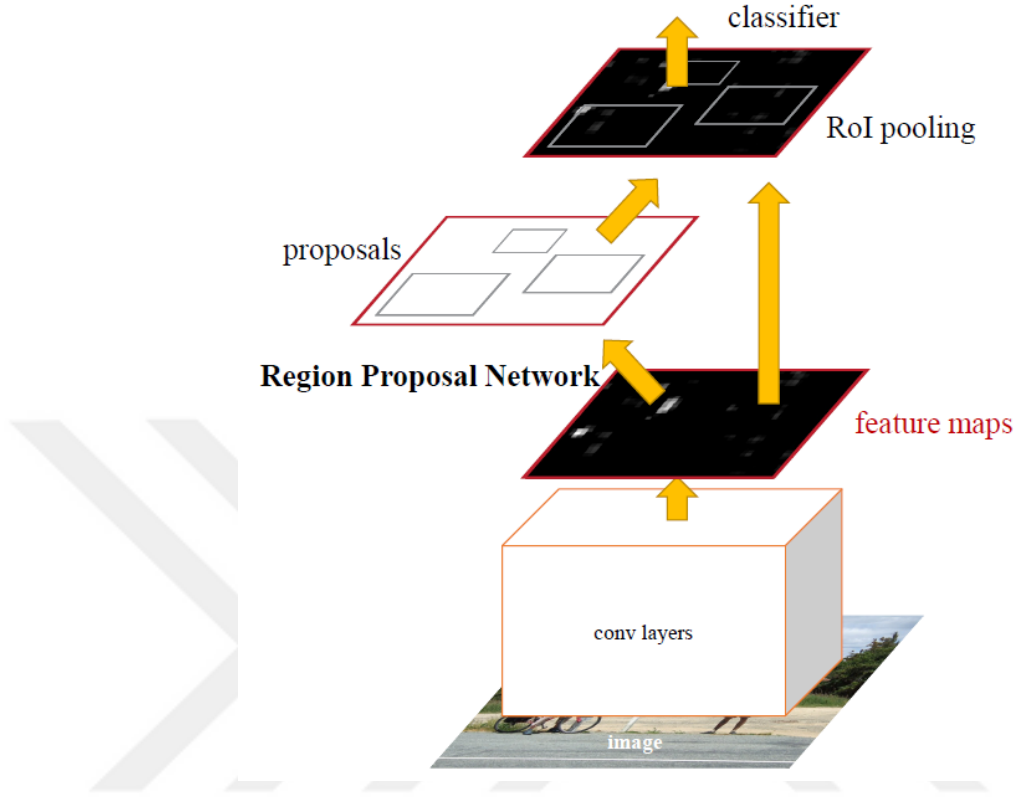


Şekil 2.6 R-CNN yöntemi akış şeması (Girshick vd. 2014)

Ren vd. (2015) çalışmasında Hızlı R-CNN (Faster R-CNN) olarak adlandırılan bir nesne tespit sistemi önermiştir. Hızlı R-CNN nesne tespit sistemi iki modülden oluşmaktadır. İlk modül, Region Proposal Network (RPN) olarak adlandırılan bölgeleri önermek için kullanılan derin tam konvansiyonel bir ağıdır ve ikinci modül, önerilen bölgeleri kullanan hızlı R-CNN dedektörüdür.

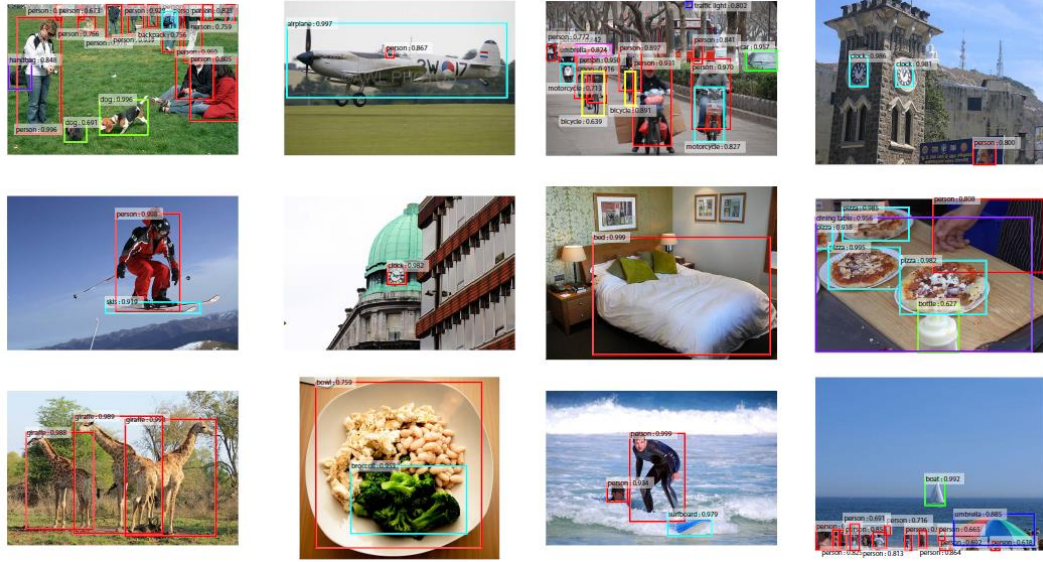
Sistemin tamamı, nesne algılama için oluşturulmuş tek bir birleşik ağdan meydana gelmektedir. Bu çalışma içerisinde RPN (Region Proposal Network), herhangi bir boyuta sahip bir görüntüyü girdi olarak almaktadır ve her biri nesnellik puanı olan bir dikdörtgen nesne teklifi seti üretmektedir. Bu işlem tam konvolüsyonel bir ağ ile modellenmiştir. Hesaplamalar bir Hızlı R-CNN nesne algılama ağı ile paylaşılacağı için her iki ağın da ortak bir konvolüsyon katmanları kümesi paylaştığı varsayılmıştır. Burada bölge önerileri üretmek için, son paylaşılan konvolüsyon katmanı tarafından konvolüsyon özellik haritası çıkışı üzerine küçük bir ağ taşınmaktadır. Bu küçük ağ girdi olarak konvolüsyonel öznitelik haritasının $n \times n$ 'lik ($n=3$ olarak kullanılmıştır) uzaysal penceresini almaktadır. Burada her kayan (sliding) penceresi düşük boyutlu bir öznitelige eşlenmektedir. Bu özellik, iki kardeş tam bağlı kutu-regresyon katmanı ve kutu-sınıflandırma katmanı katmanlarından beslenmektedir. Bu mini ağ, Şekil 2.7'da tek bir pozisyonda gösterilmiştir. Mini ağ kayan pencere (sliding window) biçiminde çalıştığı için, tam bağlantılı katmanlar tüm mekansal konumlarda paylaşılmaktadır. Bu mimari $n \times n$ konvolüsyon katmanı ve bunu takiben iki kardeş 1×1 konvolüsyon katmanı ile gerçekleştirilmektedir. Bu çalışmada RPN, geri yayılım ve SGD (Stochastic Gradient Descent) ile uçtan uca eğitilmektedir. Bu ağı eğitmek için görüntü merkezli örnekleme stratejisi uygulanmaktadır. Çalışmada, her bir mini yığın, birçok pozitif ve negatif örnek ankor'u içeren tek bir görüntüden meydana gelmektedir. Burada tüm ankorların kayıp fonksiyonlarını optimize etmek için aşağıdaki denklemde gösterilen matematiksel formülasyon uygulanmıştır (Denklem 2.3).

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i L_{reg}(t_i, t_i^*) \quad (2.3)$$

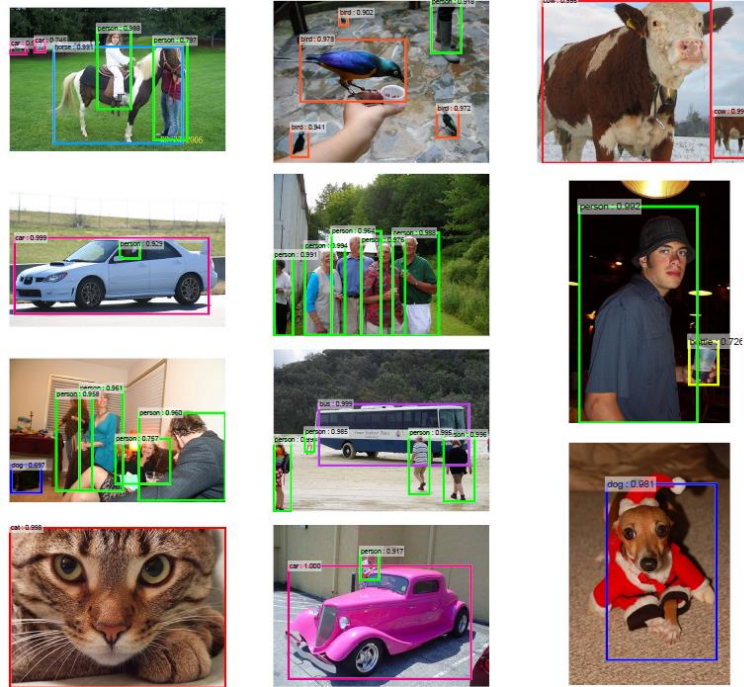


Şekil 2.7 Hızlı R-CNN nesne algılama şeması (Ren vd. 2015)

Denklem 2.3'de gösterilen matematiksel formülasyonun içerisindeki i mini yığın içerisindeki ankorun indeksini, p_i ankorun tahmini olasılığını, P_i^* ankorun durumuna göre yer gerçeği etiketini (ankor pozitif ise 1, ankor negative ise 0 değerini gösterir), t_i tahmini sınırlayıcı kutunun 4 parametrelili koordinatını temsil eden bir vektörü, t_i^* pozitif ankor ile ilişkili yer gerçeği etiketini, L_{cls} iki sınıf üzerinden log kaybını, L_{reg} gerileme kaybını göstermektedir.. Sonuç olarak, bu çalışma ile birlikte tam görüntü konvolüsyon özelliklerini algılama ağıyla paylaşan bir bölge önerme ağı (RPN) gerçekleştirilmektedir ve böylece neredeyse maliyetsiz bölge önerilerinde bulunulabilmektedir. Geliştirilen RPN ile birlikte Hızlı R-CNN'de, diğer nesne algılama ağlarının aksine harici bölge önerilerine ihtiyaç duyulmamaktadır. Ek olarak, RPN bölge öneri kalitesini artırmaktadır ve dolayısıyla nesne tespitinin genel doğruluğunu ve hızını da geliştirmektedir. Bu çalışmanın sonucunda elde edilen sonuç resim kareleri Şekil 2.8 ve Şekil 2.9'da gösterilmiştir.



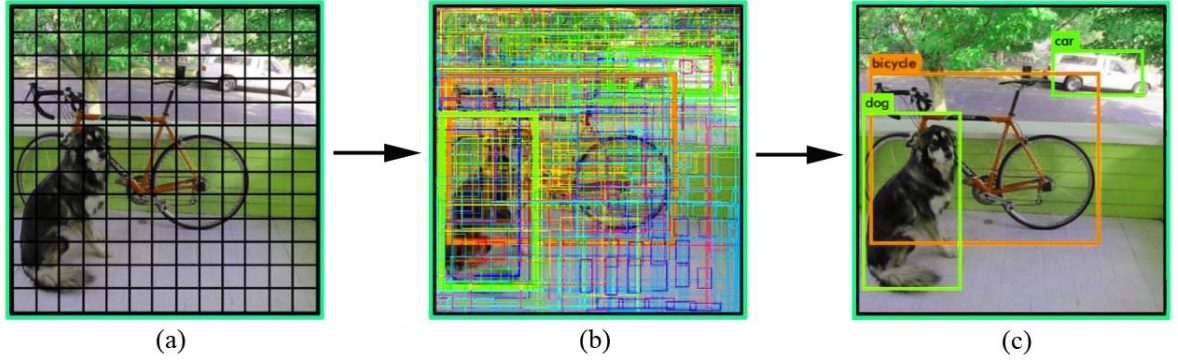
Şekil 2.8 Hızlı R-CNN için sonuç resimleri 1 (Ren vd. 2015)



Şekil 2.9 Hızlı R-CNN için sonuç resimleri 2

Redmon vd. (2016) nesnelerin sınıflandırılması ve yerlerinin tespit edilmesi için tek bir konvolüsyonel sinir ağı kullanan ve ismi YOLO olarak adlandırılan bir nesne tespit çerçevesi önermişlerdir. YOLO, nesne algılama işlemini bir regresyon problemi olarak değerlendirmektedir. Bu regresyon problemi içerisinde sınırlayıcı kutular, resim pikselleri ve sınıf olasılıkları hesaplanmaktadır.

YOLO'da ilk olarak verilen girdi veya video resim karesi $S \times S$ 'lik hücre ızgaralarına (örnek olarak 3×3 , 5×5 , 19×19) bir matris şeklinde bölünür. Burada her bir hücre ızgarası kendi içerisindeki alanda nesnenin olup olmadığını, eğerki var ise nesnenin merkez noktasının kendi matris hücresi ızgarası içinde olup olmadığını algılamaktan ve ek olarak, nesnenin merkez noktası da çalışılan matris hücresi ızgarası içindeyse nesnenin yüksekliğini, uzunluğunu ve hangi sınıftan olduğunu bulmaktan sorumludur. Buna göre algoritma içerisinde ikinci olarak, her matris hücresi ızgarası için sınırlayıcı kutu ve bu kutulara ait olasılık puanı tahmin edilir. Tahmin edilen olasılık puanı, yöntemin sınırlayıcı kutunun bir nesne içerdiğinden ne kadar emin olduğuna ve sınırlayıcı kutunun tahmin işlemini ne kadar doğru gerçekleştirdiğini göstermektedir. Burada elde edilen her bir sınırlayıcı kutu x , y , w , h , and 'confidence' değeri olmak üzere 5 tane değerden oluşmaktadır. Burada (x, y) matris hücresi ızgarasına göre hesaplanan sınırlayıcı kutunun merkez değerlerini, (w, h) tüm girdi resmine göre hesaplanan sınırlayıcı kutunun en ve boy değerlerini, 'confidence' sınırlayıcı kutunun gerçek sınırlayıcı kutuya oranı ile elde edilen olasılık puanı değerini göstermektedir. Son olarak, elde edilen fazla gereksiz sınırlayıcı kutuların NMS (Non-Maximum Suppression) algoritması ile atılması işlemi gerçekleştirilir. Bu işlem sayesinde her bir nesne için bir sınırlayıcı kutu kalmış olur ve nesnelerin sınıflandırılması ve yerlerinin tespit edilmesi işlemi gerçekleştirilmiş olur (Anonymous 2019c). YOLO yönteminin bahsedilen her bir aşaması sonucu elde edilen resim kareleri Şekil 2.10'de gösterilmiştir.



Şekil 2.10 YOLO çerçevesi resmi (Redmon vd. 2016)

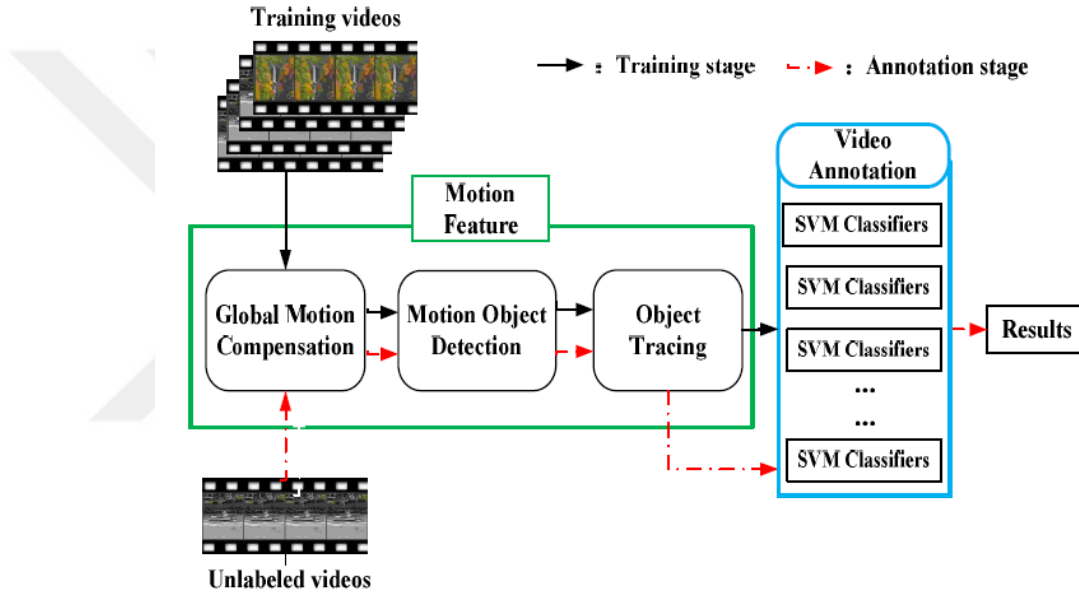
(a) Girdi Resminin $S \times S$ 'lik hücre ızgaralarına bölünmesi, (b) Her matris hücresi ızgarası için sınırlayıcı kutu ve bu kutulara ait olasılık puanlarının tahmin edilmesi, (c) Fazla gereksiz sınırlayıcı kutuların Non-Maximum Suppression algoritması ile atılması işlemi sonucu oluşan sonuç resim karesi

2.4 Diğer Yöntemler ve Yaklaşımlar

Bu alt bölüm içerisinde tez çalışma metodolojimiz ile ilgili diğer farklı alanlardaki (video etiketleme ve olay anlama yöntemleri, resim iyileştirme algoritmaları, insan algılaması ve izlemesi yöntemleri, vb.) yöntemlerden veya yaklaşımlardan detaylı özetler halinde bahsedilmiştir.

İlk olarak, video etiketleme ve olay anlama ile ilgili bir çalışmadan (Li vd. 2010) bahsedilecektir. Çalışmanın amacı, tanınmış video olaylarının etiketlenmesidir. Bu çalışmada, otomatik analizin amacına ulaşmak ve videolardaki olaylara etiketleme yapmak için VATAS olarak adlandırılan yeni bir video analizi ve yörünge tabanlı video açıklama sistemi geliştirilmiştir. Şekil 2.11'de genel bakış resmi gösterilen sistem küresel hareket tahmini, hareket nesnesi algılama, nesne izleme ve video etiketleme işlemlerini içeren dört ana modülden oluşmaktadır. İlk modülde, girdi videosunun girilmesinden sonra, hareket tahmini modülü videoyu kaydeder ve sonrasında karşılık gelen hizalanmış görüntüleri üretir. Burada sistemin esnekliğini arttırmak için SIFT temeline dayanan etkili bir özellik uyuma algoritması kullanılmıştır. Ardından, RANSAC (Random Sample Consensus) temelli sağlam parametre tahmini uygulanmaktadır. Devamında Şekil 2.12'de akış şeması

gösterilen ikinci modül içerisinde kayıtlı görüntüler alınıp onlardan hareketli nesneleri algılama işlemleri gerçekleştirilmiştir. Bu aşamada arka plan ve ön piksellerini belirlemek için üç kare fark ve morfolojik filtre yöntemleri kullanılmaktadır. Üçüncü modülde algılama nesneleri, Mean Shift tabanlı izleme yaklaşımını kullanarak video boyunca izlenmektedir. Burada nesne algılama işlemi için aşağıdaki denklem 2.4’de gösterilen matematiksel formülasyon kullanılmıştır.

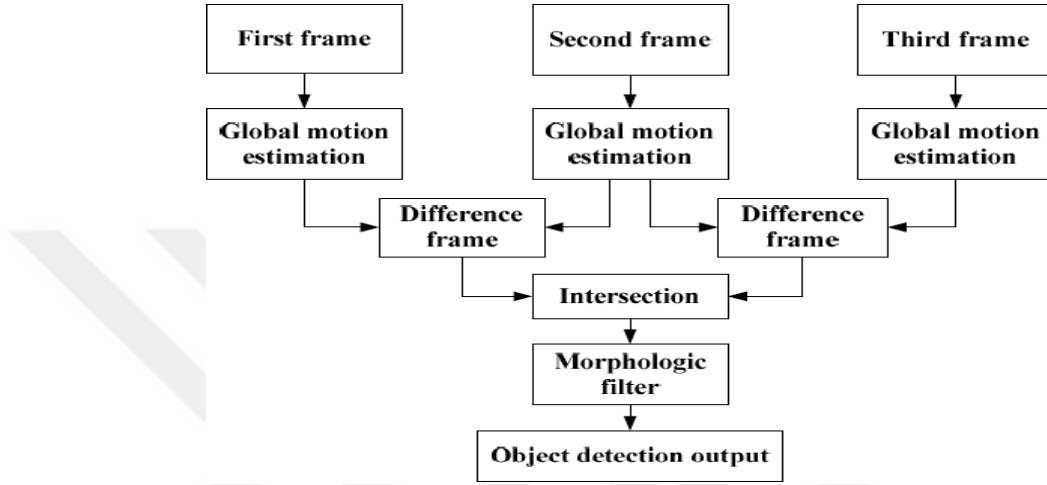


Şekil 2.11 VATAS yöntemi genel bakış (Li vd. 2010)

$$\frac{1}{2} \sum_{u=1}^m \sqrt{\widehat{p}_u(\widehat{y}_0) \widehat{q}_u} + \frac{c_h}{2} \sum_{i=1}^{n_h} w_i k\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \quad (2.4)$$

Denklem 2.4’de gösterilen formülasyon içerisindeki x_i hedef bölgedeki normalleştirilmiş piksel konumlarını, y çalışılan resim karesindeki lokasyonu, q_u hederin renk öznelik olasılığını, dışbükey ve k monotonik azalan kernel profilini, $\widehat{p}_u(\widehat{y}_0)$ taylor genişlemenin değerlerini, w_i ağırlıklandırma verilerini göstermektedir. Son olarak çalışmada, video etiketleme problemi çok sınıflı bir sınıflandırma sorunu olarak ele alınmıştır ve çok sınıflı sınıflandırıcı, tüm sınıf çiftleri sınıflandırıcılarına dayalı olarak oluşturulmuştur. Burada

sınıflandırma modeli olarak DVM kullanılarak işlemler gerçekleştirilmiştir. Deneysel sonuçlar ile önerilen yaklaşımın geçerliliği bu çalışma içerisinde ispatlanmıştır (Li vd. 2010).



Şekil 2.12 Hareket nesnesi algılama işlemi akış şeması (Li vd. 2010)

İkinci olarak, insan algılaması ve izlemesi ile ilgili bir çalışmadan (Tang vd. 2014) bahsedilecektir. Bu çalışmadaki amaç birden çok insanın bulunduğu kalabalık sokak sahnelerinde insanları tespit etmek ve izlemektir. Son zamanlardaki güncel metotlar (Andriyenko vd. 2012, Barinova vd. 2010, Breitenstein 2009, Dollár vd. 2009, Felzenszwalb vd. 2010, Huang 2008, Leibe vd. 2005, Wu ve Nevatia 2007) insan sayısının kısmen az olduğu sahnelerde iyi performans gösterirken; insan sayısının fazla, yoğun olduğu veya obje sayısının fazla olup iç içe geçtiği durumlarda yetersiz kalmaktadır. Bu sınırlama güncel insan detektörlerinin insan yoğunluğunun fazla olduğu durumlarda başarısız olduğundan dolayı oluşmaktadır. Çalışmalarda, insanların görüntülerde kesişmelerinden dolayı bu tarz çakışma durumlarının meydana geldiği gözlemlenmiş ve insan detektörünün çeşitli çakışma seviyelerine uygun hale gelmesi için çalışılmıştır. Bu çalışmada bölgesel çakışmaların üzerine gidilip onun üzerinde çalışmak yerine, tek tek insan çakışmaları üzerine gidilip bu model üzerindeki çakışmalarda daha iyi tespit sonuçları alınacağı düşünülmüş ve işlemler buna göre uygulanmıştır. Bu çalışmada çakışma

probleminin çözümünde üzerinde çalışılan insan detektörünün performans durumu çeşitli, kontrollü ve yoğun çakışmalar içeren insan veri setlerinde gözlemlenmiştir ve iki zorlu uygulama noktasında da tekil insan detektöründe daha iyi sonuçlar alınmıştır.

Bu çalışma ile çakışma problemleri üzerinde üç ana katkı sağlanmaktadır. İlk olarak, bu çalışmada insan çakışmalarının %50 veya daha üstü olduğu durumlarda (örnek olarak iki insanın çakışma durumu verilebilir) başarılı olarak insanı tahmin edebilen ve kutu ile sınırlayabilen yeni bir insan detektörü tasarlanmıştır. Ayrıca bunun yanı sıra, bu tasarlanan detektör için yeni bir eğitim metodu da tanımlanmıştır. Ek olarak, bu yaklaşım yoğun çakışmaların var olduğu görüntü karelerinde diğer tekil insan detektörleri (joint person detector) arasında daha iyi performans sağlamaktadır. İkinci olarak, bu çalışmada ortak kişi detektörü geliştirilmiştir. Bu detektörün geliştirilmesindeki amaç ise hem tekil insan tespitinde hemde çiftli insan çakışmalarındaki tespit işlemlerinde ortak bir şekilde sistemin eğitilmesi amaçlanmış ve gerçekleştirilmiştir. Bu ortak kişi detektörü zorlu ve gerçekçi verisetleri üzerinde son model performans sağlamaktadır. Son olarak, yukarıda bahsedilen ortak kişi detektörü izleme yaklaşımına entegre edilmiştir. Bu entegrede ise ortak kişi detektörü uygulamasının izleme yaklaşımına karşı çakışmış görüntüler üzerinde potansiyel başarısı incelenmiştir.

Bu çalışma DPM (Deformable Part Model) yaklaşımı (Felzenszwalb vd. 2010) üzerine inşa edilmiştir. Bu çalışmadaki çift-insan modelinin tek tek insan çakışmaları üzerine yoğunlaşp onun üzerinde çalışması bu modelin anahtar kavramıdır ve ayrıca bu kavrama ek olarak bu model iki kişinin varlığını tespit etmek yerine daha çok oluşan çakışmaları önlemek ve iyileştirmek için eğitilmiştir. Özellikle, çalışmadaki çift insan modeli aynı kök filtresine ait iki kişi arasında deforme parçaları paylaşmaktadır. Bir insan üzerindeki lokalizasyon işlemlerinden, benzer lokalizasyonların olduğu yoğun çakışmalar ve deforme parçalar içeren veri setlerinde daha iyi sonuçlar almak veya alınan sonuçların geliştirilebilmesini kolaylaştırmak için faydalanılmaktadır. Bu durumun örneği Şekil 2.13'de gösterilmiştir. Sonuç olarak, bu çalışma DPM modeli yaklaşımına dayanmaktadır. Bu çalışmada iki insanın varlığı tespit edilmekte ve her iki insanda önceden tahmin edilen

sınırlarla kutu ile sınırlandırılmaktadırlar. Burada kutu ile sınırlama işlemi doğrusal regresyon kullanılarak gerçekleştirilmiştir. Kutu ile sınırlanan her bir lokasyonun modellenmesi aşağıda denklem 2.5’de gösterilen matematiksel formülasyon ile yapılmıştır (Tang vd. 2014).

$$\mathbf{B}_i = \mathbf{g}_i(\mathbf{z})^T \boldsymbol{\alpha}_c + \epsilon_i \quad (2.5)$$

Bu formülasyonda $\mathbf{B}_i$ tespit için tahmin edilen kutu sınırlamasını ifade etmektedir. i ve c , DPM bileşenlerinin indeks değerlerini göstermektedirler. $\mathbf{g}_i(\mathbf{z})$ ise boyutsal bir vektörü ifade etmektedir. Bu vektörün değeri $2*n+3$ ’dür ve bu vektör kök filtrenin sol üst köşelerinden ve n parça filtrelerden oluşturulmuştur. ϵ_i ise gaussian gürültüsünü göstermektedir ve bu simge tahmin edilen ile gözlemlenen sınırlayıcı kutular arasındaki sapmaları tanımlamaktadır.



Şekil 2.13 Bireşimsel olarak oluşturulmuş eğitim görüntüleri örnekleri (Tang vd. 2014)

(a) 'dan (f)' ye kadar, çakışma seviyeleri kademeli olarak artmaktadır

Üçüncü olarak, insan algılaması ile ilgili bir çalışmadan (Liang ve Thorpe 2000) bahsedilecektir. Bu çalışma içerisinde kullanılan insan algılama yöntemi, ön plandaki var olan nesneleri, görüntünün veya video karesinin arka planından ayırmak için kullanılan bir modeldir. Çalışmada, nesne algılama ve insan tanıma işlemlerinde mesafe bilgilerini sağlamak için bir video oranlı stereo sistemi (video-rate stereo system) kullanılmıştır (Konolige 1997). Ek olarak, çalışma içerisinde birçok gerçek zamanlı akıllı araç sisteminde (Pomerleau 1993, Siyal vd. 1997, Wohler vd. 1998) başarıyla uygulanan sinir ağları, bu sistem içerisinde geri yayılım algoritması (Rumelhart vd. 1986) ile eğitilerek, yayaları diğer

nesnelerden ayırt etmek için kullanılmıştır. Bu model aydınlatma ve arka plan değişikliklerine karşı dayanıklılık sağlayarak diğer geleneksel modellere göre daha yüksek performans göstermektedir.

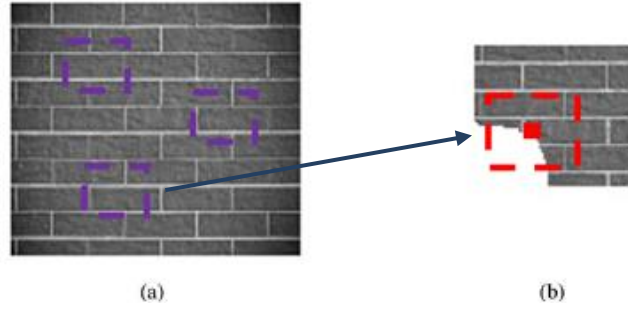
Çalışma içerisinde önerilen insan algılama yöntemi içerisinde, ilk olarak, birbirinden belirli mesafe ile ayırık olan iki video kamera ile derinlik bilgisi çıkarılarak ön plandaki nesneler ortaya çıkarılır. Ardından, yapay sinir ağı ile insan algılama sistemi eğitilir. Son olarak, eğitilen sistem ile nesneler arasından yaya olanlar diğerlerinden ayrıştırılma işlemi gerçekleştirilir. Bu çalışma sonucunda elde edilen örnek görüntü resimleri Şekil 2.14' de gösterilmiştir.



Şekil 2.14 İnsan algılaması çalışması için sonuç resimleri (Liang ve Thorpe 2000)

Dördüncü olarak, obje tespiti, video analizi, resim veya video büyütme, görüntü işleme ve bunun gibi birçok alanda son işleme süreci olarak kullanılan bir resim iyileştirme çalışmasından (Efros ve Leung 1999) bahsedilecektir. Bu çalışma içerisinde doku sentezi

işlemi için parametrik olmayan bir yöntem geliştirilmiştir. Parametrik olmayan doku sentezi (Non-parametric Sampling) yaklaşımı içerisindeki doku sentezi süreci, başlangıçtaki tohumdan dışarıya doğru piksel piksel yeni bir görüntü dokusu oluşturularak gerçekleştirilmektedir. Sentezleme işlemi bu çalışma içerisinde birkaç adımda yapılmaktadır ve bu gerçekleştirilen sentezleme işlemine ait algoritmanın genel taslağı Şekil 2.15’de gösterilmiştir. İlgili şekle göre ilk olarak, sentezlenecek olan pikselin komşularına benzeyen örnek görüntü üzerindeki tüm komşuları bulunur. Ardından rasgele olacak şekilde bir komşu seçilir ve devamında bu komşunun merkezi yeni sentezlenen piksel haline getirilerek sentezleme işlemleri gerçekleştirilmektedir.

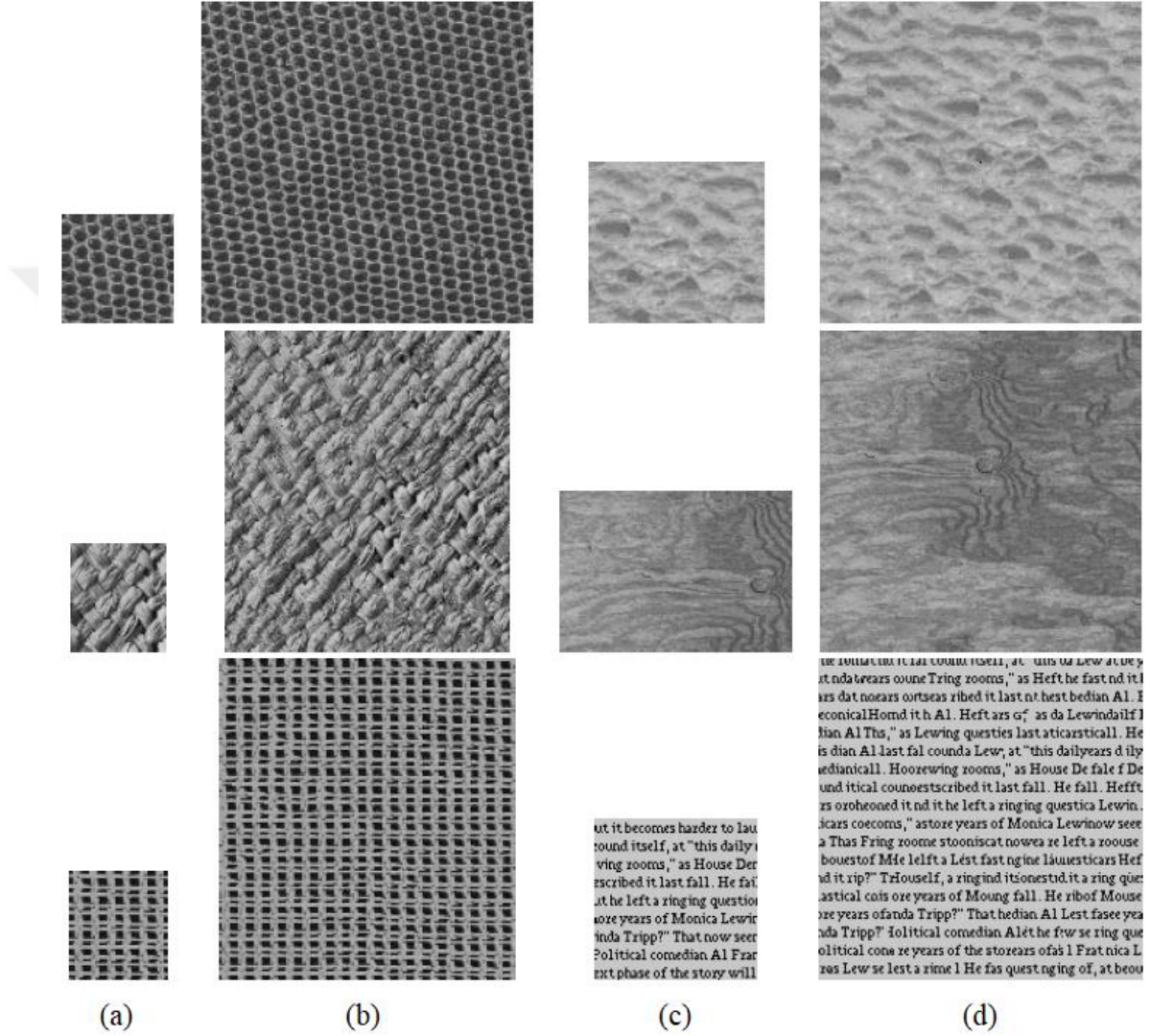


Şekil 2.15 Parametrik olmayan doku sentezi işlemi algoritması genel taslağı (Yılmaz vd. 2019)

(a) Örnek Doku Resmi (b) Hedeflenen Sentez Piksel

İlgili çalışmada MRF (Markov Random Field) modeli, doku modeli olarak kullanılmaktadır ve çalışma içerisinde MRF modeli dahilinde tüm komşuları sentezlenmiş olarak verilen bir pikselin koşullu dağılımı, örnek resim sorgulanarak ve benzer tüm komşuları bulunarak tahmin edilmektedir. Burada bir pikselin komşusu, bu pikselin etrafında kare bir pencere olarak modellenerek işlemler gerçekleştirilmektedir. Ek olarak, Rasgelelik (randomness) derecesi, algısal olarak sezgisel parametre tarafından kontrol edilmektedir ve yöntem içerisinde yerel yapı korunarak çeşitli yapay ve gerçek dünya dokuları üzerinde sentezleme işlemi için iyi sonuçlar üretmek amaçlanmaktadır (Efros ve Leung 1999). Bu çalışma sonucunda elde edilen sonuçlar Şekil 2.16’de gösterilmiştir. Bu resim içerisinde 2.16.a ve 2.16.c sütunları girdi olarak verilen resim karelerini, 2.16.b ve 2.16.d sütunları girdi resim

kareleri üzerinde parametrik olmayan doku sentezi işleminin uygulanması sonucu elde edilmiş sonuç resim karelerini göstermektedir.



Şekil 2.16 Parametrik olmayan doku sentezi yaklaşımı için sonuç resimleri (Yılmaz vd. 2019)

(a) ve (c) Girdi resim kareleri, (b) ve (d) Girdi resim karesi üzerinde gerçekleştirilen sentezleme işlemi sonrası elde edilen sonuç resim kareleri

Beşinci olarak, görüntülerden büyük nesneleri çıkarmak için tasarlanan yeni bir resim iyileştirme algoritmasından (Criminisi, Perez ve Toyama 2004) bahsedilecektir. Çalışmada,

büyük nesneleri çıkarma işlemi iki aşamadan oluşmaktadır. Bunlar nesnenin çıkarılması ve nesnenin çıkarılmasından sonra oluşan boşluk alanının doldurulması işlemleridir. Geçmiş çalışmalarda, alan doldurma işlemi iki algoritma sınıfı tarafından ele alınmıştır. Birincisi, örnek dokulardan geniş görüntü bölgelerinin doldurulması işlemi için "doku sentezi" algoritmaları'dır. Bu işlemlerde dokular üzerinde çalışmalar gerçekleştirilmiştir. İkincisi, küçük görüntü boşluklarını doldurması işlemi için "inpainting" teknikleridir. Burada çizgi ve nesne konturları gibi tek boyutlu desenler olarak düşünülebilecek doğrusal yapılar üzerinde çalışılmıştır. Örnek Tabanlı Görüntü Boyama (Exemplar-Based Image Inpainting) yaklaşımında, bu iki yaklaşımın avantajlarını birleştiren bir algoritma geliştirilmiştir.

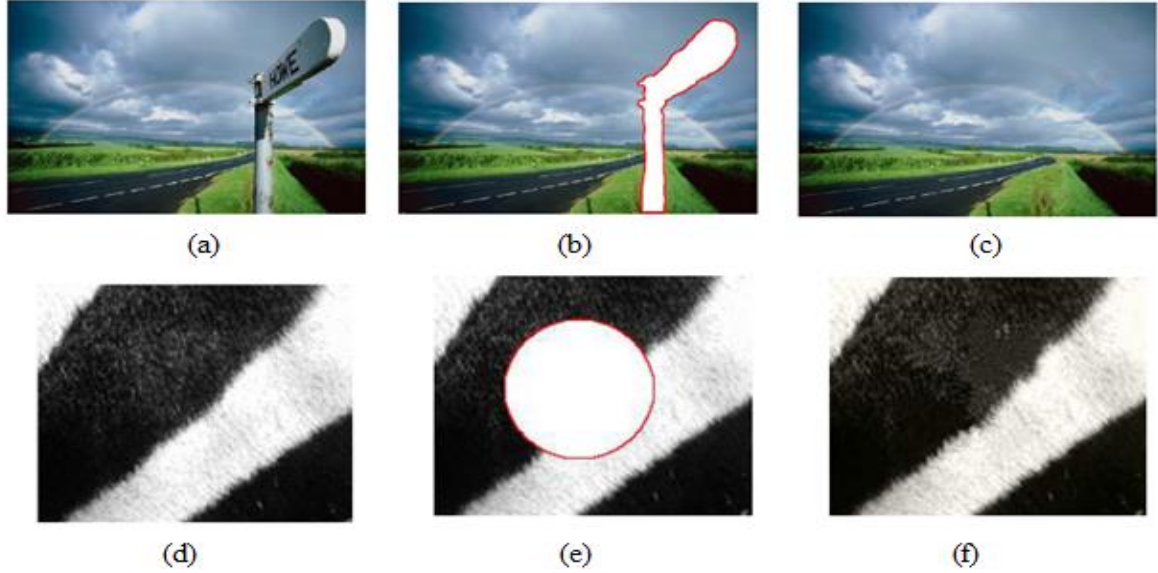
Çalışma içerisinde sentezleme görevi için en iyi olan en önce doldurulmalı mantığıyla yola çıkılarak algoritma boyunca sürekli güncellenmek üzere geçici öncelik değerlerine bağımlı bir yaklaşım izlenmektedir. Öncelikli hesaplamalarda, yüksek güvenceli piksellerle çevrelenmiş ve güçlü kenarların devamında olan yamalara karşı önyargılı (biased) olarak davranılmaktadır. Burada doldurulacak sınır üzerindeki her piksele bir öncelik değeri atanması yapılarak o piksel merkezli yamanın doldurulmasıyla, lineer yapıyı bozmadan, doldurma yapılması sağlanır. Öncelikli hesaplamaların formülasyonu denklem 2.6 ve 2.7'de gösterilmiştir.

$$P(p) = C(p)D(p) \quad (2.6)$$

$$C(p) = \frac{\sum q \psi_p \cap (I - \cap) C(q)}{|\psi_p|}, D(p) = \frac{|\nabla I_p^\perp \cdot n_p|}{\alpha} \quad (2.7)$$

Denklem 2.6'da gösterilen $C(p)$ terimi p noktasındaki pikselin çevresindeki güvenilir bilginin bir ölçüsünü vermektedir ve güven terimi olarak adlandırılmaktadır. Veri terimi olarak gösterilen $D(p)$ ise p noktasındaki parlaklık gradyanı ile o noktada sınıra dik olan normal vektörün çarpımıyla hesaplanan bir değer olup sınırın önüne doğru uzanan izotopun keskinliğinin bir ölçüsünü verir. Bu iki değer çarpımı sonucunda p merkezli yamaya verilen öncelik değeri $P(p)$ elde edilmektedir. Güven ve veri terimlerinin açık ifadeleri denklem 2.7'de verilmiştir. Bu çalışma sonucunda elde edilen sonuçlar Şekil 2.17'de

gösterilmiştir. Bu resim içerisinde ilk sütun içerisindeki Şekil 2.17.a ve Şekil 2.17.d’de girdi resmi kareleri, ikinci sütun içerisindeki Şekil 2.17.b ve Şekil 2.17.e’de girdi resim karesi üzerinde seçilen kırmızı bir sınırla işaretlenmiş hedef bölge gösterimi resim kareleri, üçüncü sütun içerisindeki Şekil 2.18.c ve Şekil 2.18.f’de örnek tabanlı görüntü boyama algoritması ile alan doldurma işlemi sonrası elde edilen resim kareleri gösterilmektedir.



Şekil 2.17 Örnek tabanlı görüntü boyama algoritması için sonuç resimleri (Criminisi, Perez ve Toyama 2004)

(a) ve (d) Girdi resmi kareleri, (b) ve (e) Girdi resim karesi üzerinde seçilen kırmızı bir sınırla işaretlenmiş hedef bölge gösterimi resim kareleri, (c) ve (f) Alan doldurma işlemi sonrası elde edilen resim kareleri

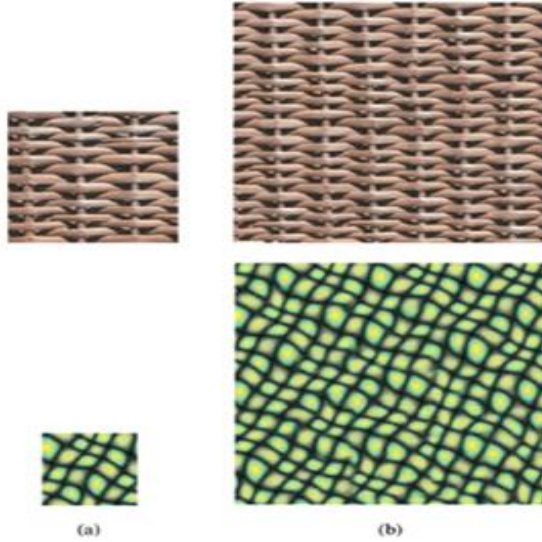
Son olarak, Liang vd. (2001) tarafından geliştirilen yama tabanlı bir sentez algoritmasından bahsedilecektir. Çalışmanın temel fikri, dokuları giriş dokusundan kopyalayarak dokuları sentezleme işlemini gerçekleştirmektir. Mevcut algoritmalarından en önemli farkı, en iyi eşleşen yamayı bulmak için aramayı hızlandırmak adına değiştirilmiş yaklaşık en yakın komşu tekniğini (modified approximate nearest neighbor) kullanması ve yüksek kaliteli doku sentezini gerçek zamanlı bir süreç haline getirmesidir. Bu geliştirilmiş arama hızıyla, algoritma gerçek zamanlı olarak çalışabilmektedir ve diğer görüntü tabanlı sentez algoritmaları ile benzer görüntü kalitesine ulaşabilmektedir. Yama tabanlı örnekleme

algoritması, düzenliden (regular) stokastik'e (stochastic) kadar çok çeşitli dokular için iyi çalışabilmektedir.

Bu çalışmada MRF doku modeli olarak kullanılmaktadır. MRF yönlendirilmemiş grafiklerin diline dayanan bir olasılık dağılımını kompakt bir şekilde temsil etmek ve görselleştirmek için kullanılan bir tekniktir. Bu yaklaşım içerisinde yerel koşullu MRF yoğunluk fonksiyonunun parametrik olmayan tahminine göre yamalar örneklenerek, yama sınırları boyunca uyumsuz öznitelikleri olması önlenmektedir. MRF'in matematiksel formülasyonu denklem 2.8'de gösterilmiştir.

$$P(I_R|I_{\partial R}) = \sum_i \beta_i \delta(I_R - I_{R^i}), \quad \sum_i \beta_i = 1 \quad (2.8)$$

Denklem 2.8 içerisindeki I_R , $I_{\partial R}$ doku yamalarını, α normalize benzerlik ölçek faktörünü ve δ Dirac'ın deltasını temsil etmektedir Burada sınır bölgeleri olarak bilinen $I_{\partial R}$ 'ye sahip olan tüm yamalar için bilinmeyen doku yaması I_R 'nin koşullu olasılık dağılımı tahmin edilerek işlemler gerçekleştirilmektedir. Bu çalışma sonucunda elde edilen sonuçlar Şekil 2.18'da gösterilmiştir.



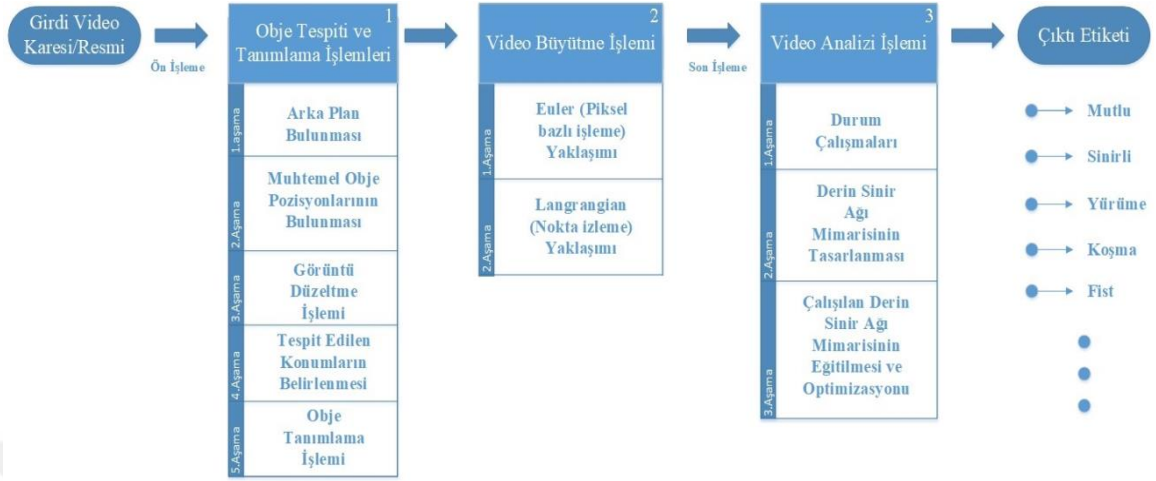
Şekil 2.18 Yama tabanlı bir sentez algoritması için sonuç resimleri (Liang vd. 2001)

(a) Giriş resmi kareleri (b) Sentezleme işlemi sonrası elde edilen resim kareleri

3. OBJE TESPİTİ VE ANALİZİ SİSTEMİNİN GELİŞTİRİLMESİ

Bu tezde obje tespiti, video büyütme yaklaşımları ve derin öğrenme yöntemleri bir araya getirilerek yeni bir obje tespiti ve analizi sistemi geliştirilmiştir. Mevcut obje tespiti ve analizi çalışmaları genellikle bu üç yaklaşım veya yöntemin tek bir tanesinin geliştirilmesi veya çiftinin bir araya getirilerek geliştirilmesi ile gerçekleştirilmişlerdir. Bu gerçekleştirilen yaklaşımlar veya yöntemler sınırlı uygulama alanları üzerinde kısıtlı başarı oranlarına ulaşabilmişlerdir. Bu yüzden çalışmada daha geniş bir uygulama alanı üzerinde daha yüksek başarı oranlarına ulaşabilmek için bu üç yöntem bir araya getirilmiştir. Bu çalışmadaki amaç video çerçeveleri arasındaki farkların analizine dayalı olarak objelerin tespitini, izlenmesini sağlamak ve sonrasında obje hareketlerini izleyen video içerisinde analiz edilecek bölgeye video büyütme yöntemi ile yaklaşım işlemi gerçekleştirilerek o bölge üzerinde geliştirilen derin öğrenme yöntemi mimarisi ile video analiz işlemleri gerçekleştirmektir. Bu kapsamda çalışma içerisinde üretilen modelde, hibrit bir modele dayalı yeni bir derin öğrenme temelli mimari geliştirilmiştir. Mimari temel olarak transfer öğrenme metodolojisi ve genetik algoritmanın kullanılması ile önceden eğitilmiş ağlara dayalı olarak tasarlanmıştır.

Tez kapsamında geliştirilen obje tespiti ve analizi sistemi modelinin mimarisi şekil 3.1’de verilmiştir. Burada numaralandırılmış her bir blok, sistem dahilinde uygulanacak işlemleri göstermektedir. Bundan sonraki açıklamalarda genel mimariye ait herhangi bir yöntemden bahsedilirken şekil 3.1’de hangi bloğa karşılık geldiği, ilgili blok numarası (Blok n) ile belirtilecektir.



Şekil 3.1 Obje tespiti ve analizi sistemi modelinin genel mimarisi

Geliştirilen obje tespiti ve analizi sisteminde ilk olarak girdi videosu/karesi üzerinde arka plan hesaplanmıştır, sonrasında muhtemel obje pozisyonlarının bulunması işlemleri gerçekleştirilmiştir (Blok 1). Muhtemel obje pozisyonlarının bulunması işlemleri içerisinde sırasıyla; hareket eden nesneleri algılamak adına her bir video karesinden arka plan çıkartım işlemi, değer kaybını en aza indirmek ve ileride daha optimize sonuçlar alınabilmesi için ölçeklendirilme işlemi ve segmentasyon işlemi gerçekleştirilmiştir (Blok 1). Devamında elde edilen sonuçlar üzerinde oluşan fazlalıkların ve gürültülerin giderilmesi için gaussian filtresi ve eşikleme (thresholding) yöntemi uygulanmış ve obje görüntüleri üzerindeki boşlukların doldurulması içinde morfolojik işlemler gerçekleştirilerek görüntü düzeltme işlemleri yapılmıştır (Blok 1). Bunun sonucunda objelerin videodaki pozisyonları tespit edilmiş ve devamında tespit edilen konumlar sınırlayıcı kutu içerisine alınmıştır (Blok 1). Bu işlem sonrasında, derin öğrenme yöntemleri ile birlikte uygun hazır veri setleri kullanılarak algılanan konumlar üzerinde obje tanımlama işlemleri gerçekleştirilmiştir (Blok 1). Sonrasında, objelerin tespit edildiği video üzerinde analiz edilecek bölgeye video büyütme yöntemleri ile yaklaşım işlemi gerçekleştirilmiştir (Blok 2). Akabinde, yaklaşılan bölge üzerinde geliştirilen derin öğrenme yöntemi mimarisi ile kapsamlı veri setleri kullanılarak video analiz işlemleri yapılmıştır (Blok 3). Bu bölümün alt başlıklarında bahsedilen mimarideki her bir adımın detayları verilmiştir.

3.1 Obje Tespiti ve Tanımlama İşlemleri

Şekil 3.1'de gösterilen obje tespiti ve analizi sistemi modelinin genel mimarisinin ilk aşaması olan obje tespiti ve tanımlama işlemleri dört ana aşamadan oluşmaktadır (Blok 1). Burada ilk aşama olarak giriş videosu üzerinde arka plan elde edilmiştir. Devamında, ikinci aşama olarak muhtemel obje pozisyonlarının bulunması işlemleri gerçekleştirilmiştir. Muhtemel obje pozisyonlarının bulunması içerisinde sırasıyla arka plan çıkarım, ölçeklendirilme ve segmentasyon işlemleri yapılmıştır. Üçüncü aşama olarak, ikinci aşamada elde edilen sonuçlar üzerinde oluşan fazlalıkların, gürültülerin giderilmesi ve obje görüntüleri üzerindeki boşlukların doldurulması için görüntü düzeltme işlemleri gerçekleştirilmiştir. Dördüncü olarak, objelerin videodaki tespit edilmiş konumları üzerinde tespit edilen konumların belirlenmesi işlemleri yapılmıştır. Son olarak, derin öğrenme yöntemleri ile birlikte uygun hazır veri setleri kullanılarak tespit edilen konumlar üzerinde obje tanımlama işlemleri gerçekleştirilmiştir. Bu paragraf içerisinde obje tespiti ve tanımlama işlemlerinin genel hatları belirtilmiştir. Bu kapsamda bu bölümün ilerleyen alt başlıklarında bahsedilen beş aşama ayrıntılı olarak anlatılmıştır. Ek olarak, bu aşamalar Şekil 3.2'de gösterilen bir videodan okunmuş bir video karesi üzerinden anlatılarak gerçekleştirilen işlemler gösterilmiştir.

3.1.1 Arka plan bulma işlemleri

Videodaki arka planın hesaplanması için ilk olarak, videodaki KYM (kırmızı, yeşil, mavi) renkte tanımlı olan kareler gri tonlara dönüştürülmüştür. Devamında gri tonlara çevrilmiş kareler toplanmış ve bu toplam videodaki kare sayısına bölünmüştür. Bu işlemlerin sonucunda videolarımızın arka plan resim kareleri gri tonda elde edilmiştir. Şekil 3.3'deki resimde videonun elde edilmiş arka planı görülmektedir. Bahsedilen bu yöntemin (Basit Metod (BM)) dışında ek olarak çeşitli arka plan çıkarım algoritmaları da çalışma içerisinde denenmiştir. Bu arka plan çıkarım algoritmaları; One Gaussian (1-G) (Wren vd. 1997), Minimum, Maximum and Maximum Inter-Frame Difference (MinMax) [Haritaoglu

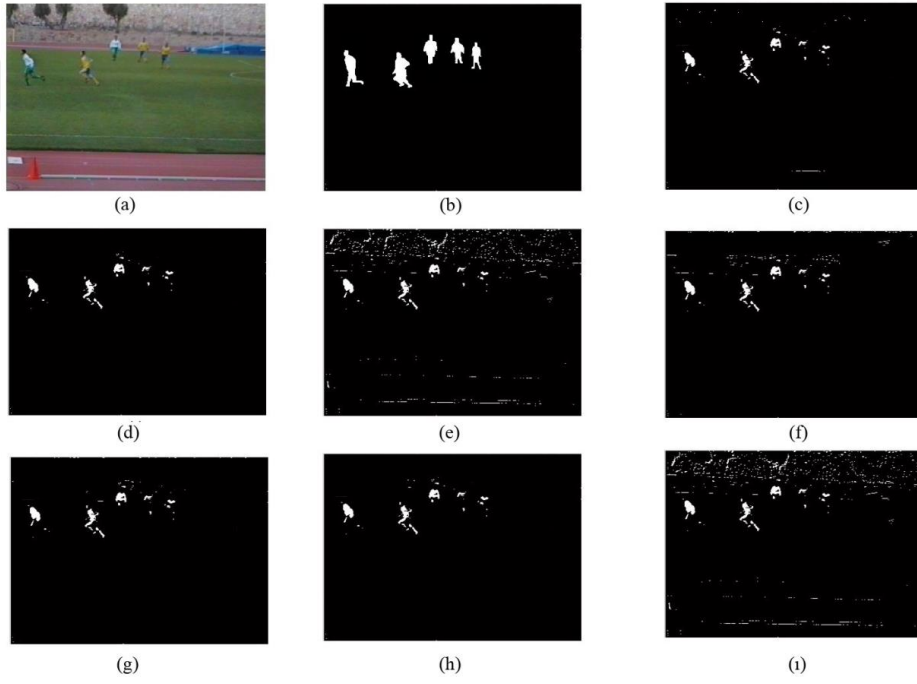


Şekil 3.2 Girdi video karesi

2000], Gaussian Mixture Model (GMM) (Stauffer vd. 1999), Kernel Density Estimation (KDE) (Elgammal vd. 2000), Codebook (CBRGB) (Kim vd. 2005), Eigen Backgrounds (Eigen) (Oliver vd. 2000)'dır. Bu arka plan çıkarma algoritmaları ve BM algoritmasının uygulanmaları sonucunda elde edilen hareket maskeleri görüntüleri Şekil 3.4'de ve performans değerlendirilmesi sonuçları Şekil 3.5'de precision/recall eğrisinde gösterilmiştir. Elde edilen hareket maskeleri görüntüleri ve precision/recall eğrisi sonuçlarına göre KDE, 1-G ve GMM algoritmaları diğer algoritmalara göre kısmen daha başarılı sonuçlar elde etmişlerdir. Fakat burada KDE, 1-G ve GMM algoritmaları yüksek performans değerleri göstermelerine rağmen daha yüksek bellek gereksinimi (memory requirement) ve hesaplama çabası (computational effort) gerektirmişlerdir. Bu nedenle, daha az bellek gereksinimi ve hesaplama çabası gerektiren BM algoritması, çalışmada büyük ölçekli veriler kullanılacağından tercih edilip kullanılmıştır.

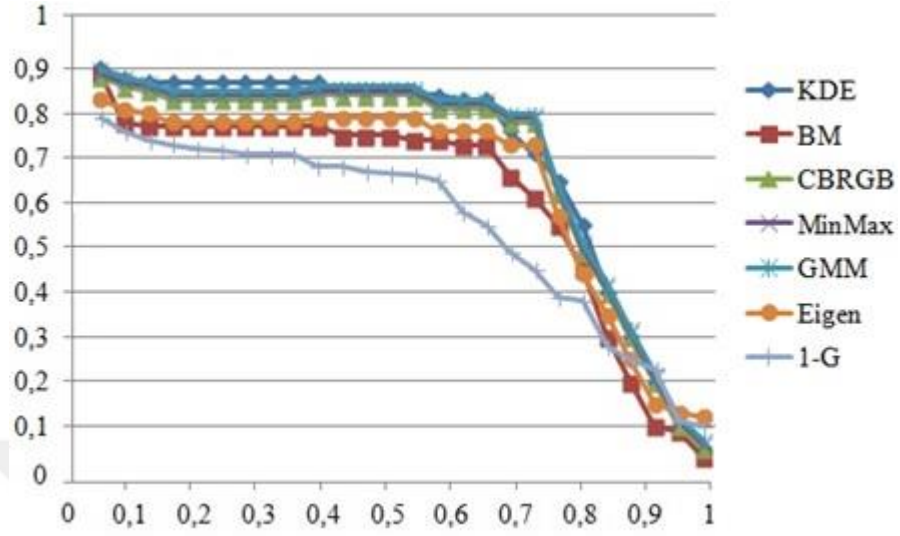


Şekil 3.3 Arka plan görüntüsü



Şekil 3.4 Arka plan çıkarma algoritmaları için hareket maskeleri

(a) Girdi resmi, (b) Ground truth, (c) BM, (d) KDE, (e) MinMax, (f) CBRGB, (g) GMM, (h) 1-G, (i) Eigen



Şekil 3.5 Arka plan algoritmaları için precision/recall eğrisi

3.1.2 Muhtemel obje pozisyonlarının bulunması işlemleri

Muhtemel obje pozisyonlarının bulunması işleminde, hareket eden nesneleri algılamak adına video içerisindeki her bir karenin renk değerleri renkli tonlardan gri tonlara çevirilmiş ve elde edilen arka plandan farkı alınmıştır. Bu işlemlerin sonucunda Şekil 3.6'daki resim elde edilmiştir. Burada elde edilen arka plan ve kare matrisleri değerleri 0 ile 1 arasındaki değerlerinden oluşmaktadırlar ve "0" siyahı, "1" beyazı, arada kalan diğer değerler gri renk tonlarını temsil eder. Yukarıdaki işlemler sonucunda ulaşılan matris içerisindeki değerler -1 ile +1 arasındadır ve bunlar genel olarak -0.5 ile +0.5 aralığında bulunmaktadırlar. 0 değerinin altındaki değerlerin bizim için bir anlamı yoktur ve bu nedenle bu değerler herhangi bir resim karesindeki 0 değeri gibi siyah olarak gözükeceklerdir. Bu durumdan dolayı ve daha az kayıp yaşanmasını sağlamak adına değerler 0 ile 1 arasında ölçeklendirilmiştir. Ölçeklendirme işlemi sonucunda elde edilen resim karesi Şekil 3.7'da gösterilmiştir. Ölçeklendirme işleminin sonrasında, segmentasyon işlemleri gerçekleştirilmiştir. Segmentasyon işlemleri içerisinde birden fazla segmentasyon metodu denenmiştir. Bu segmentasyon metodları; Fuzzy C-Means (FCM) (Cannon vd. 1986, Chi

vd. 1996), K-Means (Jain ve Dubes 1989), Expectation Maximization (EM) (Dempster vd. 1977), Mean Shift (MS) (Cheng 1995)'dir. Video görüntüsünü kullanılan videonun içinde makul sayıda obje olduğunu düşünmekteyiz. Buradaki makul sözcüğü ile demek istenilen anlam, video içerisindeki objelerin sayısının aşırı fazla sayıda olmaması ve anlık bir kareye bakıldığında arka planı temsil eden piksellerin objeleri temsil eden piksellerden daha fazla olduğunun varsayılmasıdır. Bu nedenle yukarıdaki işlemlere göre oluşan sonuca bahsedilen segmentasyon algoritmaları uygulanarak objeler gruplara bölünmüştür. Burada en iyi sonuçları EM, MS segmentasyon metodları elde etmiştir. Burada kısmen MS metodu daha iyi sonuç verdiği için çalışmamızda bu metod seçilerek kullanılmıştır. Sonuç olarak segmentasyon işlemi sonrasında elde edilen aşağıdaki Şekil 3.8'de gösterilen sonuca ulaşılmıştır.



Şekil 3.6 Video karesiyle arka plan görüntüsünün farkı



Şekil 3.7 Video karesinin ölçeklendirilmiş fark verisi



Şekil 3.8 Video karesinin segmentasyon işlemine uğramış fark verisi

3.1.3 Görüntü düzeltme işlemleri

Görüntü düzeltme işlemlerinde ilk olarak video görüntülerinde oluşan fazlalıkların veya gürültülerin giderilmesi gerekmektedir. Bu işlem için ilk işlem olarak bağlı bileşenlerin sayısını azaltmak adına “Gaussian” filtresi ile “smoothing” işlemi gerçekleştirilmiştir. Ardından “thresholding” işlemi uygulanmıştır. Bu işlem içerisinde seçilmiş belli bir boyun altında olan 1 ile işaretlenmiş kümeler kir olarak kabul edilmiş ve bu bölgeler kaldırılmıştır. Seçilen parametrenin belirlenmesi burada çok kritiktir çünkü eğer video görüntülerinde bu parametre değerini, objelerin temsil ettiği piksellerden daha büyük seçilirse objeyi veya objenin belli bir takım yerleri kaybedebilebilir veya tam tersi durumda gerektiğinden küçük bir değer seçilirse bazı gürültüler, fazlalıklar o anki karede kalabilir. İşlem sonrası gözüken kirler maksimize bir şekilde yok edilerek Şekil 3.9’daki ekran görüntüsü elde edilmiştir.



Şekil 3.9 Fazlalıkları atılmış video karesi

Bu düzeltmelerin ardından görüntü düzeltme işlemlerinin içerisinde ikinci olarak, objelerin gövde içeriklerinin belli bölgelerinde oluşan boşlukların giderilmesi işlemi gerçekleştirilmiştir. Bu boşlukların oluşmasının nedeni objenin boş olarak işaretlenmiş bölgesiyle hemen arkasındaki konumun renginin benzer olmasından kaynaklanmaktadır.

Bu tipte hataları gidermek için morfolojik işlemler kullanılarak objenin boş kalan bölgelerinin doldurulması sağlanmıştır. Morfolojik işlemler içerisinde temel morfolojik görüntü işleme teknikleri kullanılmıştır (Gonzalez ve Woods 1992, Boztoprak vd. 2007, Umbaugh 1999). Bunlar genişletme, aşındırma işlemleridir. Genişletme işlemi içerisinde resimde bulunan piksellerin filtre ile kesiştiği bölümler kadar büyütülme işlemi gerçekleştirilmiştir. Burada bu işlemi yapabilmek için filtre resim üzerinde piksel piksel dolaştırılmıştır. Bu işlem içerisinde eğer filtrenin orijini resim üzerinde "0" değerli bir piksel ile karşılaşılırsa herhangi bir değişiklik meydana getirilmemiştir. Fakat eğer değeri "1" olan bir piksel ile karşılaşılırsa filtre ile filtrenin altında kalan pikseller mantıksal "OR" işlemine tabi tutulmuştur. Aşındırma işleminde, genişletme işleminin tersi bir işlem gerçekleştirilmiştir. Burada genişletme işleminde de olduğu gibi aynı şekilde filtre resim üzerinde piksel piksel dolaştırılarak işlemler yapılmıştır. Filtrenin merkez pikselinde "1" değeri ile karşılaşılırsa filtre içerisindeki piksellerin durumuna bakılmıştır. Eğer filtre içerisindeki "1" olan piksellerden herhangi birinin altında resme ait "0" değeri varsa filtrenin diğer "1" değerlerinin altındakilerle beraber bu piksel "0" a dönüştürülmüştür. Aşağıda Şekil 3.10'daki resimde bu morfolojik yöntemle düzeltilmiş resim karesi görülmektedir.



Şekil 3.10 Morfolojik işlem uygulanmış video karesi

3.1.4 Tespit edilen konumların belirlenmesi işlemleri

Yukarıda anlatılan işlemler sonucu, her bir kare üzerindeki objelerin olduğu konumlar maksimize bir şekilde bulunmuştur. Tespit edilen konumların belirlenmesi işlemleri içerisinde, bu pozisyonlarla önceden okuduğumuz girdi videosunun her bir karesi sınırlayıcı kutu (bounding box) içerisine alınmıştır. Aşağıdaki Şekil 3.11'deki resimde bulunan yeni oluşturulan video görüntüsü karesi, ilk girdi göstergesi olarak gösterdiğimiz karenin yukarıda anlatılan süreçten geçerek işlenmiş halini göstermektedir.



Şekil 3.11 Tespit edilen konumların belirlendiği video karesi

3.1.5 Obje tanımlama işlemleri

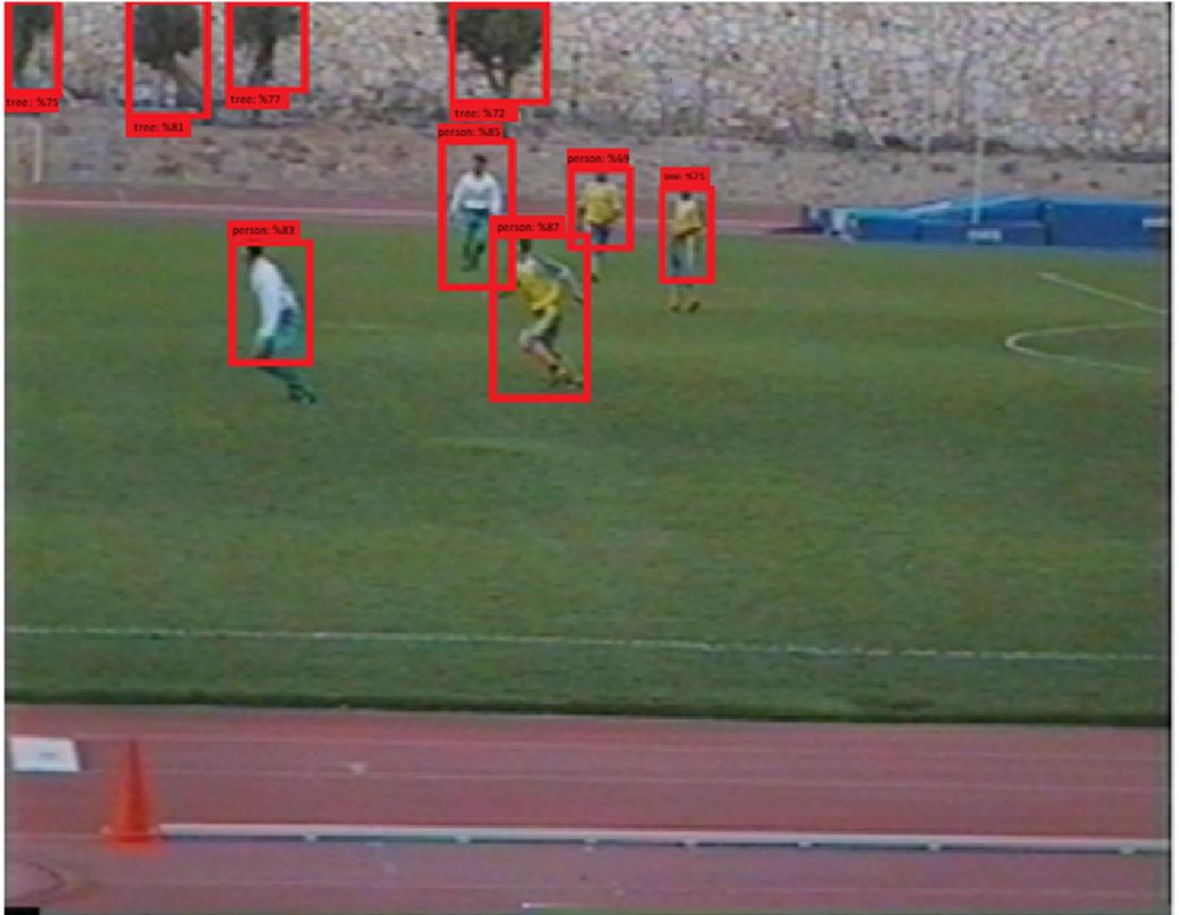
Obje tanımlama işlemleri içerisinde, derin öğrenme yöntemleri ile birlikte uygun hazır veri setleri kullanılarak obje tanımlama işlemleri yapılmıştır. Burada kullanılan derin öğrenme yöntemleri SSD (Single Shot Multibox Detector) (Liu vd. 2016), R-FCN (Region-based Fully Convolutional Networks) (Dai vd. 2016), Faster R-CNN (Faster Region-based Convolutional Network) (Ren vd. 2015), Mask R-CNN (Mask Region-based Convolutional Network) (He vd. 2018) ve kullanılan uygun hazır veri setleri Coco (Anonymous 2020b), Kitti veri setleri (Anonymous 2020c)'dir. Belirtilen derin öğrenme yöntemleri ve veri setleri ile birlikte öznitelik çıkarıcılar (MobileNet v1, Inception v2, Resnet 50, Resnet 101, Inception Resnet v2, Resnet 50 atrous, Resnet 101 atrous) kullanılarak Çizelge 3.1'deki sonuçlar elde edilmiştir. Çizelge 3.1'deki sonuçlara göre burada elde edilen en iyi sonuç en yüksek mAP değerine sahip olan Faster R-CNN derin öğrenme yöntemi, Kitti veri seti ve Resnet 101 öznitelik çıkarıcısı birleşimidir. Bu birleşim çalışmamız içerisinde en iyi sonucu verdiğinden, obje tanımlama işlemleri içerisinde kullanılmıştır. Sonuç olarak obje tanımlama işlemlerinin gerçekleştirilmesi sonucunda elde edilen ekran görüntüleri Şekil 3.12, Şekil 3.13 ve Şekil 3.14'de gösterilmiştir ve bu işleminde tamamlanması ile obje tespiti ve analizi sistemi modelinin genel mimarisinin ilk aşaması olan obje tespiti ve tanımlama işlemleri çalışmaları başarılı bir şekilde gerçekleştirilerek tamamlanmıştır.

Çizelge 3.1 Derin öğrenme yöntemleri, veri setleri ve öznitelik çıkarıcı birleşimleri için ortalama hassasiyet (mAP) değeri karşılaştırması

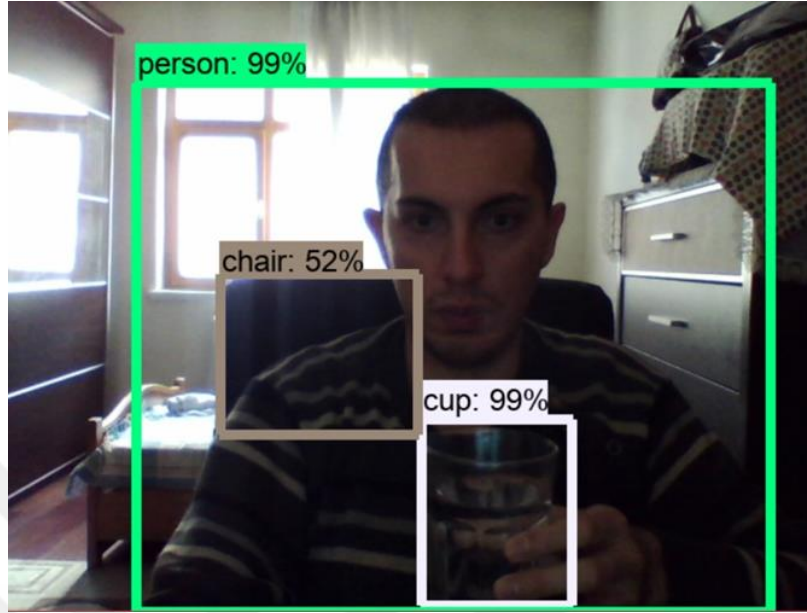
Derin Öğrenme Yöntemi	Veri seti	Öznitelik Çıkarıcı	mAP(%) (Precision/Recall)
SSD	Coco	MobileNet v1	51
SSD	Coco	Inception v2	54
R-FCN	Coco	Resnet 101	60
Mask R-CNN	Coco	Inception v2	55
Mask R-CNN	Coco	Resnet 50 atrous	59
Mask R-CNN	Coco	Resnet 101 atrous	63
Mask R-CNN	Coco	Inception Resnet v2	66

Çizelge 3.1 Derin öğrenme yöntemleri, veri setleri ve öznitelik çıkarıcı birleşimleri için ortalama hassasiyet (mAP) değeri karşılaştırması (devam)

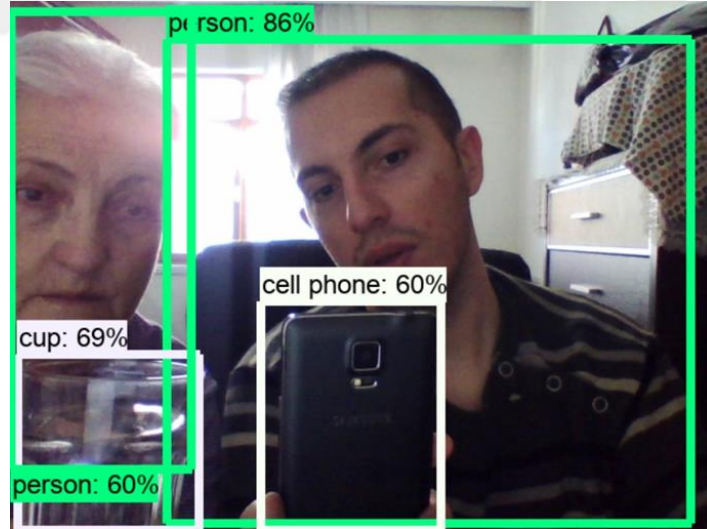
Faster R-CNN	Coco	Inception v2	58
Faster R-CNN	Coco	Resnet 50	60
Faster R-CNN	Kitti	Resnet 101	87
Faster R-CNN	Kitti	Inception Resnet v2	67



Şekil 3.12 Obje tanımlama işlemi sonrası elde edilen görüntü karesi 1



Şekil 3.13 Obje tanımlama işlemi sonrası elde edilen görüntü karesi 2

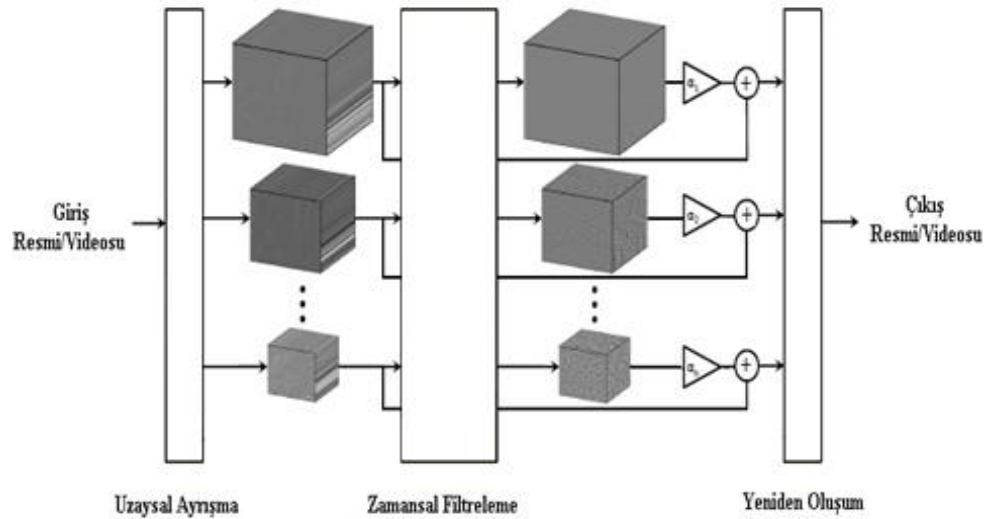


Şekil 3.14 Obje tanımlama işlemi sonrası elde edilen görüntü karesi 3

3.2 Video Büyütme İşlemleri

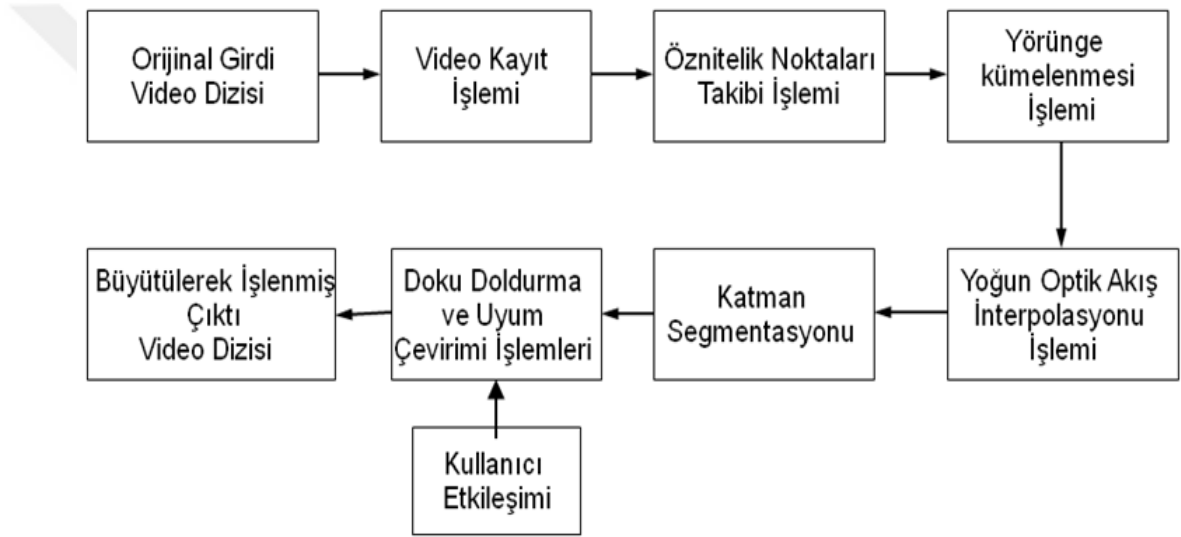
Bu bölüm içerisinde; obje tespiti ve analizi sistemi modelinin genel mimarisinin ikinci aşaması olan video büyütme işlemleri ile ilgili gerçekleştirilen çalışmalar anlatılmıştır (Blok 2). Çalışma içerisinde video büyütme işlemleri için iki yaklaşım üzerinde çalışılmıştır. Bu iki yaklaşım aşağıda "Euler (piksel bazlı işleme)" ve "Lagrange (nokta izleme)" olarak özetlenmiştir.

İlk olarak, Şekil 3.15'de genel bakış resmi gösterilen Euler yaklaşımı (Wu vd. 2012) dört ana aşamadan oluşmaktadır. Buradaki yaklaşımda ilk adım olarak girdi video dizileri farklı spatial frekans bantlarına ayrıştırılmaktadır. Devamında bir sonraki adımda, ayrıştırılma işlemi uygulanan bantlara aynı zamansal (temporal) filtre uygulanmıştır. Üçüncü adım olarak, filtreleme işlemi uygulanan spatial bantlar kullanıcı tarafından direkt olarak verilen veya otomatik olarak belirlenen büyütme faktörü değerine göre büyütülmüştür. Son olarak, büyütülen bantlar orijinal sinyalin arkasına eklenmiş ve devamında bu bantlar nihai çıkış videosunu oluşturmak adına daraltılmıştır.



Şekil 3.15 Euler yaklaşımı için genel bakış (Wu vd. 2012)

İkinci olarak, Lagrangian yaklaşımı (Liu vd. 2005) altı aşamadan oluşmaktadır. Aşamaları sırasıyla Şekil 3.16'da gösterilen yaklaşım içerisinde sırasıyla video kayıt, öznelik noktaları takibi, yörünge sınıflandırması, yoğun optik akış enterpolasyonu, katman segmentasyonu, büyütme, doku doldurma ve uyum çevirimi işlemleri gerçekleştirilmektedir. Yukarıda video büyütme işlemlerinin genel hatları belirtilmiştir. Bundan sonraki alt bölümlerde çalışılan Euler ve Lagrangian video büyütme yaklaşımlarının tek tek her aşamasının detayları sırasıyla anlatılacaktır.



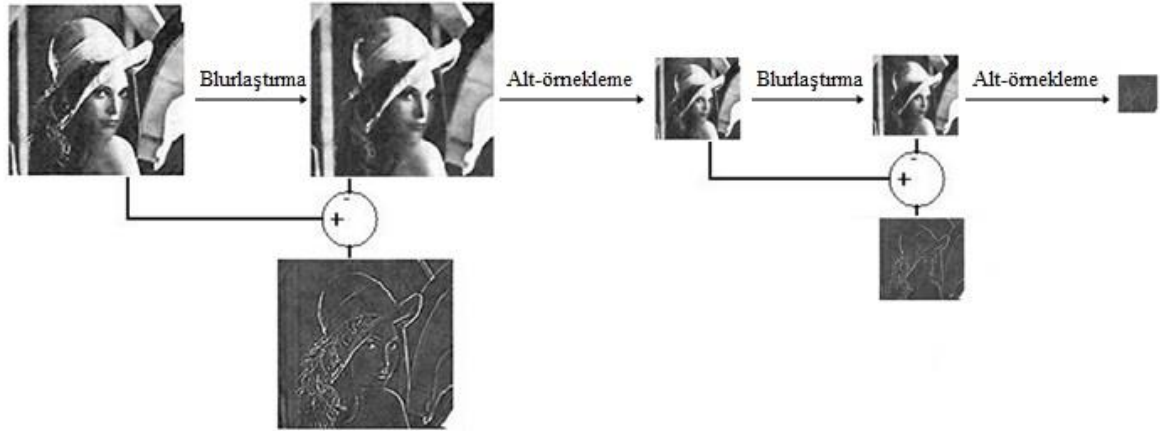
Şekil 3.16 Langrangian yaklaşımı için genel bakış (Liu vd. 2005)

3.2.1 Euler (piksel bazlı işleme) yaklaşımı

Euler yaklaşımında spatial ve temporal değişimler birleştirilmiştir. Yaklaşım içerisinde ilk adım olarak video veya resim dizileri farklı uzaysal frekans bantlarına ayrıştırılmıştır. İkinci olarak, artefaktları (artifacts) bastırmak için bu bantların amplifikasyonu azaltılmıştır. Uzaysal işlemenin (spatial processing) amacı temelde zamansal sinyal görüntü oranını çoklu piksel toplanması aracılığı ile artırmaktır. Ancak bu yaklaşım içerisinde uzamsal olarak sayısal verimlilik (computational efficiency) için video resim karelerine alçak geçirgen filtreleme (low-pass filter) ve alt örnekleme (downsampling) işlemleri yapılmıştır.

Bununla birlikte, Euler yaklaşımında genel durumda uzaysal işleme için Laplacian piramidi kullanılmıştır (Burt ve Adelson 1983).

Laplacian piramidi bir görüntü birleştirme yöntemidir. Görüntü birleştirme yöntemleri içerisinde bazı özel tekniklerle birden fazla girdi video veya resim karesinden çok daha elverişli ve kullanılabilir tek bir birleşik görüntü oluşturulabilmektedir ve bu sayede girdi videodaki gerekli bilgiler ortaya çıkarılabilmektedir. Bir görüntü karesinin Laplacian piramidi dönüşümü, her kademedede görüntü karesine bant geçiren filtre (bandpass filter) veya alçak geçiren filtre (lowpass filter) uygulanarak dönüşüm yapılan Gauss piramit yapısında ardışık görüntü kareleri arasındaki fark işlemi hesaplamasıyla gerçekleştirilen bir görüntü ayrıştırma yöntemidir (Sadjadi 2005, Olkkonen vd. 1996). Bu yöntemin örneği Şekil 3.17'de verilmiştir. Örnek içerisinde ilk olarak Gaussian dönüşümü gerçekleştirilmiş ve devamında ardışık Gauss piramidi düzeylerinde alçak geçiren görüntüler arasındaki farkın hesaplanması ile görüntülerin Laplacian piramidi dönüşümleri elde edilmiştir.



Şekil 3.17 Laplacian piramidi dönüşümü örneği (Burt ve Adelson 1983, Yılmaz vd. 2017)

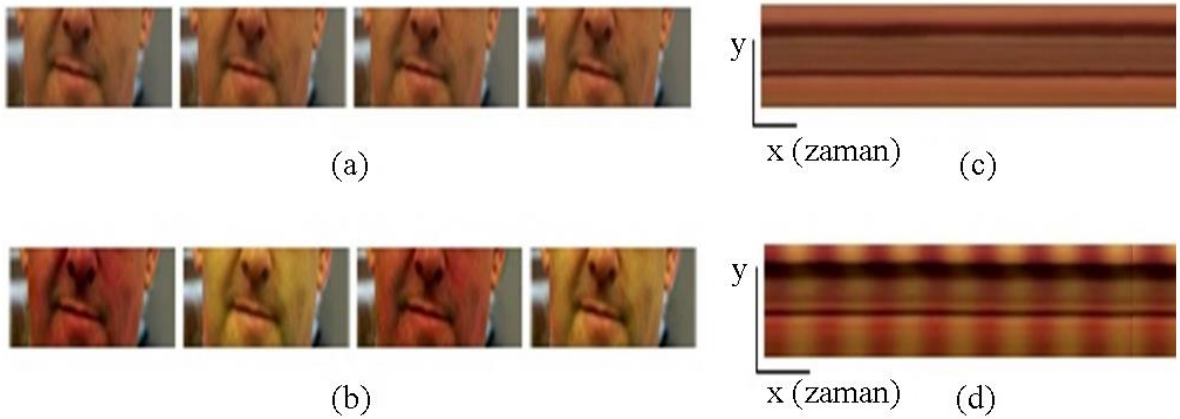
Devamında, her bir spatial bant için zamansal işleme (spatial processing) işlemi gerçekleştirilmiştir. Zamansal işleme işlemi içerisinde ilk olarak bir frekans bandında bir pikselin değerine karşılık gelen zaman serisi değerlendirilmiş ve sonrasında ilgili frekans bandını çekip çıkartmak için bant geçirgen filtresi uygulanmıştır. Yaklaşımda sonrasında elde edilen bant geçirgen sinyaller, büyütme (amplifikasyon) faktörü ile çarpılmıştır. Burada büyütme faktörü değeri otomatik olarak sonraki paragraflarda belirtilen hareket büyütülmesi işlemi esaslarına göre veya kullanıcı tarafından ayrıca manuel olarak ayarlanmaktadır. Son işlem olarak, orijinal sinyale büyütülmüş sinyal eklenmiş ve uzaysal piramit boyunca daraltılarak nihai sonuç çıktısı elde edilmiştir. Bu yaklaşımda zamansal işleme işlemi içerisinde hareket büyütülmesi işleminin üretimi, optik akış analizinde yaygın olan birinci dereceden Taylor Serisi açılımlarına dayanan analiz ile yapılmıştır. Burada Taylor serisine göre oluşturulan ve kullanılan hareket büyütme işlemi matematiksel formülasyonü denklem 3.1’de gösterilmiştir.

$$\mathbf{I}^*(\mathbf{x}, t) \approx \mathbf{f}(\mathbf{x} + \sum_k (\mathbf{1} + \alpha_k) \delta_k(t)) \quad (3.1)$$

Denklem 3.1 içerisinde $\mathbf{f}(\mathbf{x})$ t zamanındaki lokal resmi, $(\mathbf{1} + \alpha)$ büyütülme büyüklüğü değerini ve $\delta(t)$ uzaysal yerdeğişimini ifade etmektedir (Lucas ve Kanade 1981, Horn ve Schunck 1981). Şu ana kadarki varsayımlar ve işlemler, düz resimler ve yavaş değışimli (small motions) hareketler için uygulanmıştır. Ancak uygulanan bu işlemler hızlı değışen resim fonksiyonlarında uygulandığında başarılı olunamamıştır. Bu durumun nedeni; birinci dereceden Taylor serisi yaklaşımları içerisinde daha büyük büyültme faktörü ve hareket değeri alınması sonucu hatalı sonuçlar ortaya çıkması ve hareket büyütme hataları oluşmasıdır. Burada bu hataların oluşumunu engellemek adına büyültme faktörü değeri için kesilme (cutoff) noktası belirlenmiştir (Denklem 3.2) ve belirlenen bu kesilme noktası değeri limit noktası olarak işlemler gerçekleştirilmiş ve sonuç olarak, büyültme faktörü değeri kesilme noktasından fazla artırılmadığında hareket büyütme hatalarının oluşmadığı gözlemlenmiştir. Ek olarak, bazı videolar içerisinde bu limit değeri ihlal edilip çok ölçekli işlemlerde kullanıcı ayarlı parametrelerde kullanılmıştır.

$$(1 + \alpha)\delta(t) < \lambda / 8 \quad (3.2)$$

Sonuç olarak, Euler yaklaşımı ile video büyültme işlemi videolar üzerinde dört adımda gerçekleştirilmiştir. İlk olarak, zamansal bant geçirgen filtresi seçimi yapılmıştır. İkinci adım olarak, büyültme faktörü değeri seçilmiştir. Devamında, uzaysal frekans limiti (spatial frequency cutoff) seçilmiştir. Son olarak, büyültme faktörü değeri için azaltım (attenuation) oluşum değeri seçilmiştir. Video büyültme işleminin uygulanması öncesi ve sonrası oluşan video görüntü karelerinin ve dikey tarama doğrularının durumları Şekil 3.18’de gösterilmiştir (Yılmaz vd. 2017). Burada girdi video dizisinde değişim belirsiz iken, işlenerek büyütölmüş video dizisinin tarama doğrusunda değışim açık bir şekilde gözlemlenmektedir.



Şekil 3.18 İnsan nabzını görüntöleme örneđi (Wu vd. 2012)

(a) Orijinal video dizisinden alınan dört video karesi, (b) Orijinal video dizisinden alınan dört görüntü karesinin nabız sinyalleri ile büyütölmüş halleri, (c) İlgili video büyütme metodunun uygulanması öncesi periyodik renk değışimlerini gösteren dikey tarama doğrusu, (d) İlgili video büyütme metodunun uygulanması sonrası periyodik renk değışimleri gösteren dikey tarama doğrusu

3.2.2 Langrangian (nokta izleme) yaklaşımı

Lagrangian yaklaşımının hedefi videodaki küçük hareketleri tespit etmek ve sonrasında bu hareketleri büyütme işlemine sokarak göze çarpmayan veya güç algılanan değışimleri

ortaya çıkarmaktadır. Lagrangian yaklaşımı altı ana aşamadan oluşmaktadır. Bu aşamalar sırasıyla Şekil 3.16'da gösterilmiştir.

Lagrangian (nokta izleme) yaklaşımında ilk adım olarak video kayıt işlemi aşaması gerçekleştirilmiştir. Bu işlem içerisinde ilk adım olarak Harris köşe detektörü (corner detector) (Harris ve Stephens 1988) kullanılması ile referans çerçevesi üzerinde çeşitli farklı ölçeklerde köşeler tespit edilmiştir. Devamında, her bir çerçevenin kaydedilmesi yerine güvenilir öznitelik noktaları (feature point) bulunmuştur. Bu işlem içerisinde öznitelik noktaları izlemesi işlemi (feature point tracking) için BF (kaba kuvvet algoritması) kullanılmıştır. Son adım olarak, referans görüntü karesine göre her bir görüntü karesine global afin dönüşümü (global affine translation) yapılmış ve sonrasında yoğunluk normalizasyonu (intensity normalization) işlemi uygulanmıştır.

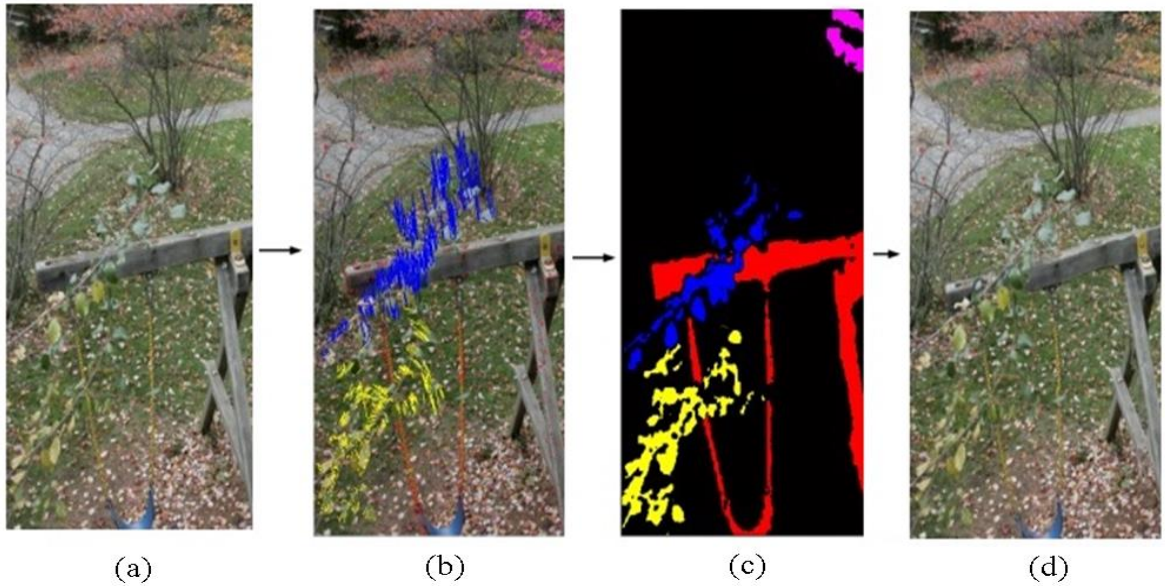
İkinci olarak öznitelik noktaları takibi işlemi yapılmıştır. Bu işlem uygulanırken kaba kuvvet arama tekniğinden (brute force search) faydalanılmıştır. Ayrıca, EM algoritması kullanılarak destek adaptif bölgesi öğrenilmiştir. Devamında, öğrenilen destek adaptif bölgesi üzerinden izleme yapılarak aykırı değerlerin tespiti ve kaldırılması işlemi enterpolasyon işlemi ile gerçekleştirilmiştir (Dempster vd. 1977).

Üçüncü olarak yörünge kümelenmesi işlemi gerçekleştirilmiştir. Yörünge kümelenmesi işlemindeki amaç alan içerisinde özellik kümelenmesini ve toplanmasını gerçekleştirmektir. Bu nedenle birbiriyle ilişkili hareketler bu işlem içerisinde tespit edilmiştir. Burada birbiriyle ilişkili hareketleri tespit edilmesi işlemi; yörüngeler arasında normalleştirilmiş korelasyonun, benzerliğin ölçülmesi için kullanılması ile gerçekleştirilmiştir. Devamında, tespit edilen birbiriyle ilişkili hareketler kümelenmiştir ve ilişkili bir dizi küme seti oluşturulmuştur.

Dördüncü aşama olan yoğun optik akış interpolasyonu işlemin içerisinde yerel ağırlıklı lineer regresyon kullanılarak yoğun optik akış alanı interpolasyon ile gerçekleştirilmiştir. Burada işlemler yoğun optik akış alanından seyrek özellik noktalarına doğru yapılmıştır.

Beşinci aşamada, katman segmentasyonu işlemi gerçekleştirilmiştir. Bu işlemde, her bir piksele hareket kümesi katmanı atanmıştır. Her bir piksele atanacak küme tahmini işleminde; her bir piksel için hareket, pozisyon, renk ve zamansal tutarlılıklar değerlendirilmesi ile küme tahmini yapılmıştır.

Son aşama olan büyütme, doku doldurma ve uyum çevirimi işlemleri aşamasında; ilk olarak, kullanıcı tarafından bir katman hareket büyütülmesi yapılması için seçilmiş ve büyütülme işlemi yapılmıştır. Devamında, değiştirilmiş hareket ile diğer her bir katman uyumlu hale getirilerek işlenmiştir. Son olarak, arka planda kaybolan veya eksik olan piksellerin kazanılması için Efros ve Leung (1999) tarafından önerilen doku sentezi yöntemi kullanılarak nihai sonuç çıkış videosu elde edilmiştir. Her bir aşamada elde edilmiş görüntü resim kareleri aşağıda gösterilmiştir (Şekil 3.19).

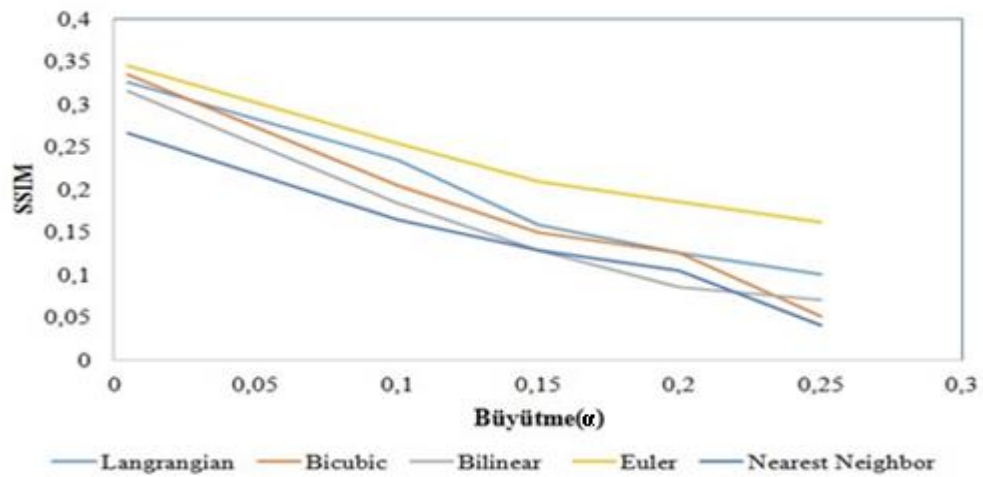


Şekil 3.19 Lagrangian yaklaşımı örneği (Liu vd. 2005)

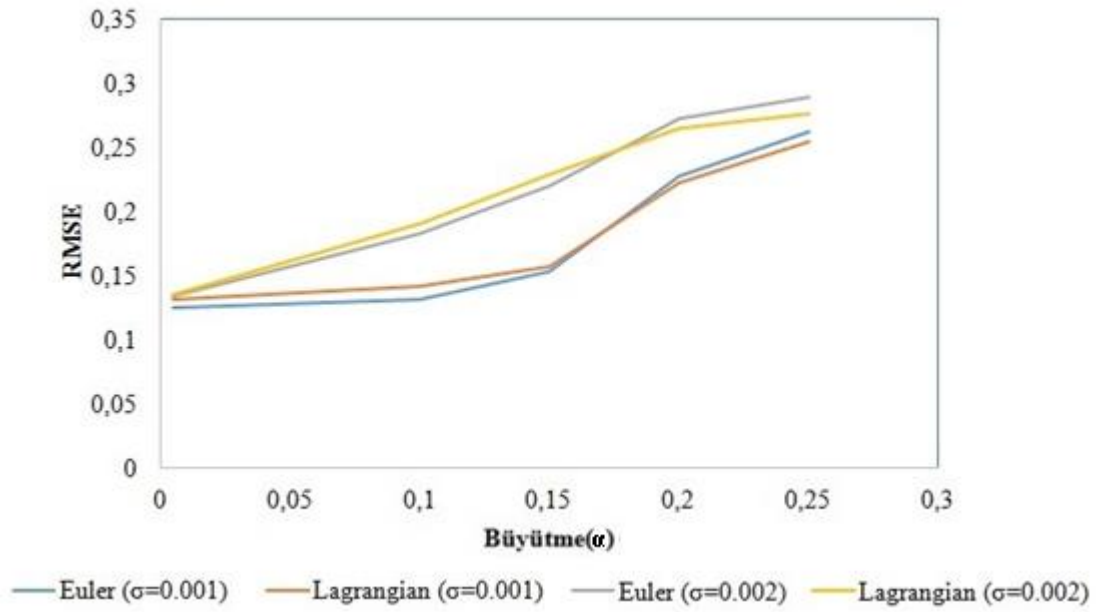
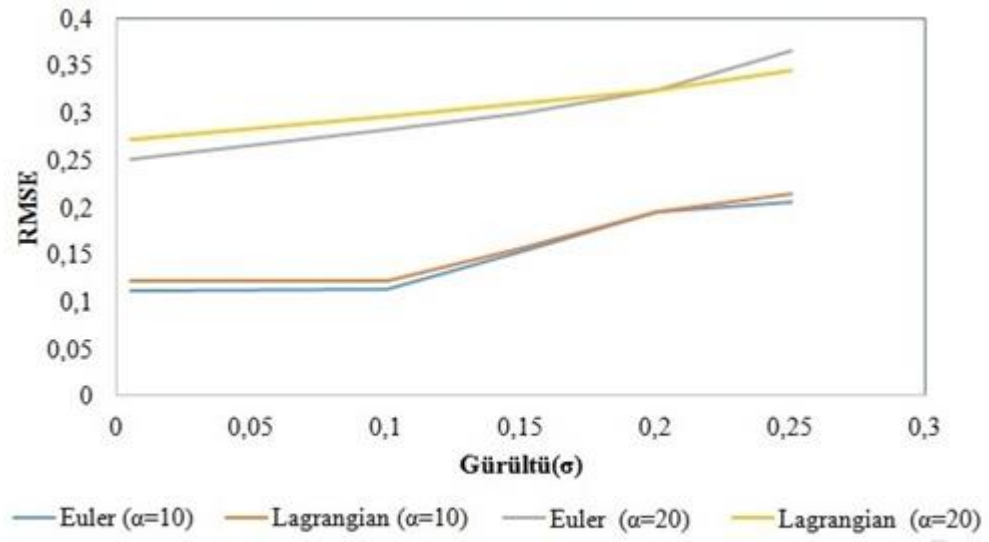
(a) Video kayıt işlemi aşamasında elde edilen görüntü karesi, (b) Özniteklemler takibi ve yörünge kümelenmesi aşamalarının uygulanması sonrasında elde edilen görüntü karesi, (c) Video kayıt işlemi aşamasında elde edilen görüntü karesi, (d) Yoğun optik akış interpolasyonu ve katman segmentasyonu aşamalarının uygulanması sonrasında elde edilen görüntü karesi

Sonuç olarak, iki video büyütme yaklaşımının aşamalarının gerçekleştirilmesi sonrasında işlenerek büyütülmüş nihai çıkış videoları veya resim dizileri başarılı bir şekilde elde edilmiştir. Euler ve Lagrangian yaklaşımlarının karşılaştırması yapılırsa, Euler yaklaşımının

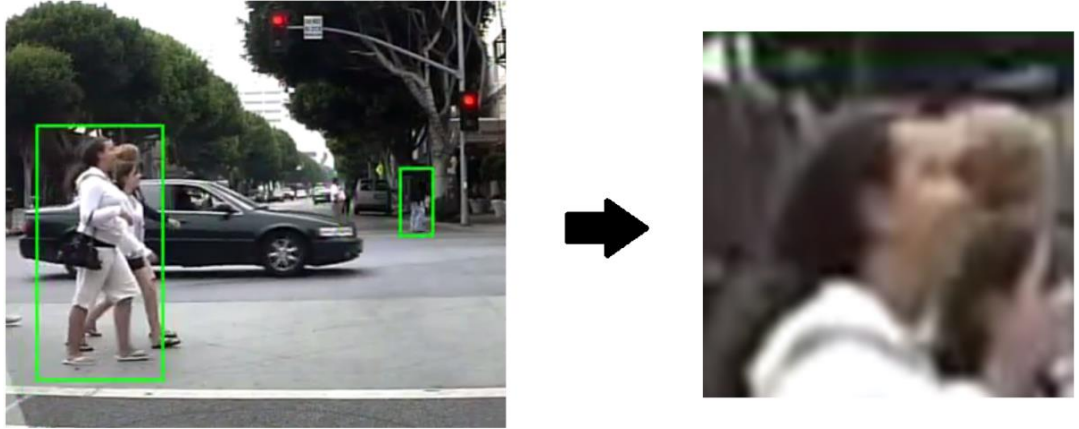
Lagrangian yaklaşımına göre iki ayrı avantajı bulunmaktadır. Bunlarda ilki, açık hareket kestirimi ve doku sentezi aşamalarının bulunmamasıdır. Bahsedilen bu durum Euler yaklaşımının kısmen daha hızlı olmasını sağlamaktadır. İkincisi, küçük amplifikasyonlar ve daha büyük gürültü seviyeleri için daha duyarlı olmasıdır. Bu durum Şekil 3.21'de gözlenmektedir. Şekilde görüldüğü üzere Euler (piksel bazlı işleme) yaklaşımının, küçük amplifikasyonlar ve daha büyük gürültü seviyeleri için daha az hata vererek daha iyi sonuçlar verdiği gözlemlenmiştir. İki yaklaşımın bahsedilen RMSE metrik değeri ile karşılaştırılmasına ek olarak, temel enterpolasyon teknikleriyle de (Bilinear (Bilineer), Bicubic (Bikubik), Nearest neighbor (En yakın komşu)) karşılaştırma işlemi yapılmıştır. Bu karşılaştırma işlemi SSIM metrik değeri kullanılması ile gerçekleştirilmiştir (Şekil 3.20). Burada da Euler yaklaşımı, elde edilen yüksek SSIM metrik değerleri ile diğer tüm tekniklere göre başarılı performans göstermiştir. Bahsedilen değerlendirmelere ve elde edilen sonuçlara göre Euler yaklaşımı, Lagrangian yaklaşımı ve diğer tekniklere göre çalışmamız içerisinde daha iyi sonuçlar verdiği için video büyütme işlemleri içerisinde kullanılmıştır. Sonuç olarak video büyütme işleminin gerçekleştirilmesi sonucunda elde edilen ekran görüntüleri Şekil 3.22, Şekil 3.23, Şekil 3.24, Şekil 3.25'de gösterilmiştir ve bu işleminde tamamlanması ile birlikte obje tespiti ve analizi sistemi modelinin ikinci aşaması olan video büyütme işlemleri çalışmaları başarılı bir şekilde gerçekleştirilerek tamamlanmıştır.



Şekil 3.20 Euler, Lagrangian ve diğer temel enterpolasyon yöntemlerinin karşılaştırması grafiği (SSIM)



Şekil 3.21 Euler ve Lagrangian yöntemlerinin karşılaştırılması grafiği (RMSE)



Şekil 3.22 Video büyütme/yaklaşım işlemi sonrası elde edilen görüntü karesi 1

(a) Girdi video karesi (Anonymous 2019d), (b) Video büyütme işleminin gerçekleştirilmesi sonucunda elde edilen büyütülmüş (yakınlaştırılmış) ekran görüntüsü karesi



Şekil 3.23 Video büyütme/yaklaşım işlemi sonrası elde edilen görüntü karesi 2

(a) Girdi video karesi (Anonymous 2019d), (b) Video büyütme işleminin gerçekleştirilmesi sonucunda elde edilen büyütülmüş (yakınlaştırılmış) ekran görüntüsü karesi



Şekil 3.24 Video büyütme/yaklaşım işlemi sonrası elde edilen görüntü karesi 3

(a) Girdi video karesi, (b) Video büyütme işleminin gerçekleştirilmesi sonucunda elde edilen büyütülmüş (yakınlaştırılmış) ekran görüntüsü karesi



Şekil 3.25 Video büyütme/yaklaşım işlemi sonrası elde edilen görüntü karesi 4

(a) Girdi video karesi (Anonymous 2015), (b) Video büyütme işleminin gerçekleştirilmesi sonucunda elde edilen büyütülmüş (yakınlaştırılmış) ekran görüntüsü karesi

3.3 Video Analizi İşlemleri

Bu bölümde; obje tespiti ve analizi sistemi modelinin genel mimarisinin üçüncü aşaması olan video analiz işlemleri ile ilgili yapılan çalışmalar anlatılmıştır (Blok 3). Bu aşamada ilk olarak obje tespiti ve analizi sistemi modelinin uygulanacağı durum çalışmaları ve modelleri belirlenmiştir. İkinci aşamada obje tespiti ve analizi sisteminin derin sinir ağı mimarisinin tasarlanması işlemi gerçekleştirilmiştir. Bu kapsamda çalışma içerisinde üretilen model, hibrit bir modele dayalı yeni bir derin öğrenme temelli mimari önermektedir. Mimari temel olarak transfer öğrenme metodolojisi ve genetik algoritmanın kullanılması ile önceden eğitilmiş ağlara dayalı olarak tasarlanmıştır. Son olarak, çalışılan derin sinir ağı mimarisinin bizim tarafımızdan geliştirilen genetik algoritma modellemesi ile optimizasyonu ve devamında eğitilmesi işlemleri gerçekleştirilmiştir. Bu paragraf içerisinde video analiz işlemleri işlemlerinin genel hatları belirtilmiştir. Bu kapsamda bu bölümün ilerleyen alt başlıklarında bahsedilen üç aşamanın detayları verilmiştir.

3.3.1 Durum çalışmaları

Obje tespiti ve analizi sistemi modelinin geliştirilmesinde ve değerlendirilmesinde kullanılmak üzere çeşitli durum çalışmaları yapılmıştır. Gerçekleştirilen durum çalışmaları duygu analizi işlemleri ve hareket ve davranış analizi işlemleri'dir. Duygu analizi işlemleri içerisinde insanların psikolojik durumlarının analiz edilmesi için yüz duygularını tanımlama işlemleri yapılmıştır. İkinci olarak, hareket ve davranış analizi işlemlerinde insanların eylemlerinin tanımlama ve el hareketi tanıma işlemleri gerçekleştirilmiştir. Gerçekleştirilen iki durum çalışması içerisinde alanına uygun kapsamlı ve popüler verisetlerinden (JAFPE, CK+, KTH, UCF Sports Action, Skoda, UCI Smartphone, LaRED, TinyHand) faydalanılmıştır. Bu verisetlerinin detayları kullanılan veri setleri ve özellikleri bölümünde (bölüm 4.1) ayrıntılı olarak verilmiştir. Yukarıda durum çalışmalarından genel olarak bahsedilmiştir. Bu kapsam içerisinde ilerleyen alt başlıklarda gerçekleştirilen durum çalışmalarının detaylarından bahsedilmiştir.

3.3.1.1 Duygu analizi işlemleri

İnsanlar arası iletişimde sözlü ifadeler kadar yüz duyguları da başkalarının niyetlerini anlamamıza yardımcı olan önemli faktörlerdir. İnsanlar genellikle diğer insanların yüz ifadeleri ve ses tonlarından yararlanarak şaşkınlık, sevinç ve üzüntü gibi duygusal durumları hakkında anlamlar çıkarırlar. Farklı araştırmalara göre, insanlar arası yüz yüze iletişimde sözel olmayan bileşenlerin sözlü olan bileşenlere göre çok daha fazla etkiye sahip olduğu gözlemlenmiştir (Mehrabian 1968, Essa ve Pentland 1997, Kaulard vd. 2012, Mandal ve Srivastava 2012). Bu nedenle, bilişim alanında da bilgisayar tarafından yüz ifadesinin tanınması işlemi, insan bilgisayar etkileşiminde önemli kilit bir rol oynamaktadır ve son yıllarda çok fazla araştırmanın ilgisini çekmiş, etkin bir araştırma konusu olmuştur. İnsan-bilgisayar etkileşimi (Dornaika ve Raducanu 2007, Bartneck ve Lyons 2007), sürüş güvenliği (Gao vd. 2014, Yılmaz vd. 2018), sağlık bakımı (Muhammed vd. 2017), sanal gerçeklik (Hickson vd.), artırılmış gerçeklik (Chen vd. 2015), görüntü rekonstrüksiyonu (image reconstruction) (Mutlu vd. 2014) ve bunun gibi uygulama alanları yüz duyguları tanıma işlemlerinin kullanıldığı uygulama alanlarına örnek verilebilir.

Yüz duygularını tanıma problemini genel olarak tanımlamak gerekirse şu şekilde tanımlanabilir: bir bilgisayar sistemine, görüntü veya videolarda var olan insan yüzlerinin optimize bir şekilde bulunması ile algılanan yüz bölgesinden yüz özelliklerinin çıkarılması ve sonrasında yüz özelliklerinin hareketi veya yüz özelliklerinin görünümündeki değişikliklerin analiz edilmesi sonucunda elde edilen bu bilgiyi gülümseme, kızma gibi bazı yüz ifadeleri olarak yorumlaması yeteneğini kazandırma işlemidir. Yüz duygularını tanıma yöntemleri, literatür içerisinde özniteliklerin el yapımı (conventional-based) (Guzel 2016) veya derin bir sinir ağının çıktısı (deep learning-based) (Gao vd. 2014) yoluyla üretilmesine bağlı olarak iki grupta incelenmektedirler.

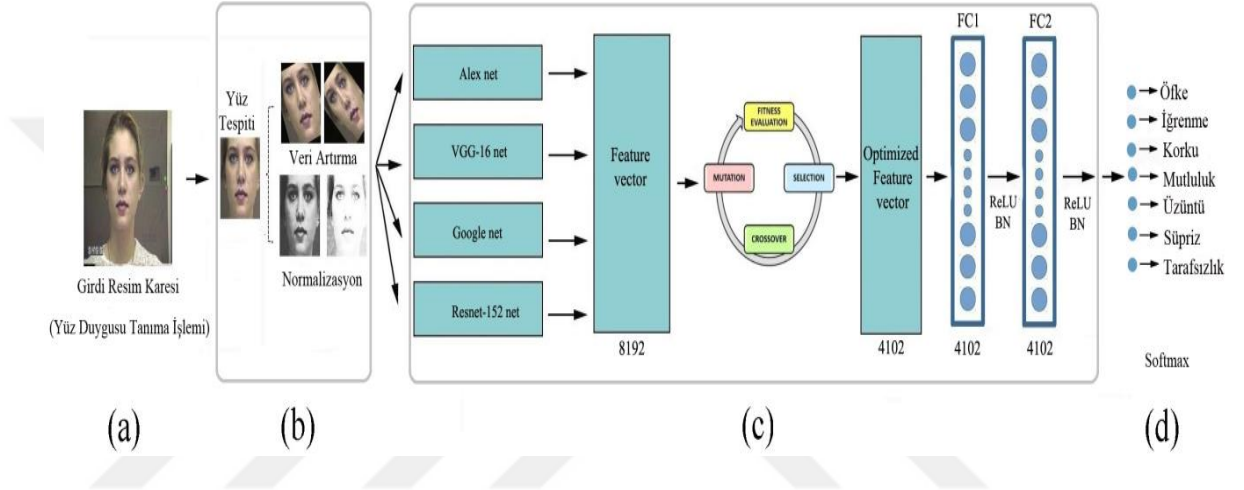
Duygu analizi işlemleri durum çalışmasında, gerçek zamanlı yüz duygusu tanımlama görevi için transfer öğrenimi (transfer learning) kullanılarak detayları derin sinir ağı mimarisinin tasarlanması bölümünde bahsedilen (bölüm 3.3.2) yeni bir derin sinir ağı

mimarisi geliştirilmiş ve bu geliştirilen derin sinir ağı mimarisi kullanılarak yüz duygusu tanımlama işlemleri gerçekleştirilmiştir. Ek olarak çalışma içerisinde, geliştirilen derin sinir ağı mimarimizin performansını değerlendirmek için deneysel analiz karşılaştırmaları yapılmıştır. Bunlar; seçilen diğer yüz duygusu tanımlama işlemlerinde kullanılan derin öğrenme mimarilerinin gerçekleştirilmesi ve diğer state-of-the-art algoritmalarının sonuçlarının alıntı yapılması işlemleridir. Bu durum çalışmasındaki amacımız videolardan elde edilen yüz görüntülerini analiz ederek yüz duygusunu tanımlama işlemlerini optimize bir şekilde gerçekleştirmektir. Burada yedi yüz duygu kategorisinde sınıflandırma yapılmıştır. Bu kategoriler öfke (angry), iğrenme (disgust), korku (fear), mutluluk (happy), üzüntü (sad), sürpriz (surprise) ve tarafsızlık (neutral)'tır.

Yüz duygusu tanıma sistemleri, öznitelik gösterimlerinde girdi verisi olarak kullandıkları resim kareleri (statik tabanlı) veya video karelerine (dinamik tabanlı) göre iki gruba ayrılabilirler. İlk olarak, statik tabanlı yüz duygusu tanıma sistemlerinde öznitelik gösterimi yalnızca geçerli tek görüntüdeki mekansal (spatial) bilgilerle kodlanır ve burada yalnızca el yapımı özniteliklerin çıkarılmasıyla elde edilen statik yüz öznitelikleri kullanılarak yüz duygusu tanıma işlemleri gerçekleştirilir. İkinci olarak dinamik tabanlı yüz duygusu tanıma sistemlerinde, yüz ifadeleri dizisindeki uzamsal-zamansal (spatio-temporal) özellikleri kullanılarak ifade dinamikleri elde edilir ve bu sistemlerde yüz fizik tabanlı modellerin kullanımı büyük oranda azaltılarak boru hattı içerisinde doğrudan girdi görüntülerinden uçtan uca öğrenme sağlanarak elde edilen yüz özniteliklerinin kullanılması ile yüz duygusu tanıma işlemleri yapılır (Ko 2018, Kim vd. 2017, Li ve Deng 2018, Walecki ve Rudovic 2017).

İlgili durum çalışmamız içerisindeki sistem, dinamik tabanlı yüz tanıma sistemleri kategorisine girmektedir ve çalışılan sisteminin metodolojisi dört ana adımdan oluşmaktadır. Bunlar sırasıyla Şekil 3.26'deki şemadan anlatılırsa; ilk olarak videolar üzerinden girdi resim kareleri alınır. İkinci olarak, alınan video kareleri üzerinden yüz algılama işlemi yapılır. Devamında, hızlı ve daha iyi eğitim için öncelikle gürültüyü azaltmak adına ön işleme işlemleri yapılır. Ek olarak bu aşama içerisinde veri arttırma (data

augmentation) ve normalizasyon (normalization) işlemleri de gerçekleştirilmektedir. Üçüncü aşama içerisinde öznitelik çıkarım işlemi (feature vector) ve derin sinir ağının eğitilmesi, optimizasyonu ve sınıflandırması işlemleri gerçekleştirilir. Son olarak, girdi video karesinin yüz duygusu tanımlama işlemi uygulanmış çıktı video karesi elde edilerek yüz duygusu tanımlama işlemi gerçekleştirilir.



Şekil 3.26 Obje tespiti ve analizi sistemi modeli için duygu analizi işlemleri durum çalışmasının genel boru/akış hattı diyagramı

(a) Giriş video çerçevesi (Lucey vd. 2010), (b) Yüz algılama ve ön işleme işlemleri, (c) Öznitelik çıkarım, öznitelik optimizasyonu, derin sinir ağları eğitimi ve sınıflandırma işlemleri, (d) Yüz duygusu tanımlama işleminin sonuç çıktısı

3.3.1.2 Hareket ve davranış analizi işlemleri

Hareket ve davranış analizi işlemleri durum çalışması içerisinde iki ayrı işlem için durum çalışmaları gerçekleştirilmiştir. Bu işlemler insan eylemlerini ve el hareketi tanıma işlemleri'dir. İnsan eylemlerini ve el hareketlerini tanıma işlemleri, son yıllarda çok sayıda yapılan akademik araştırmaları ve çeşitli uygulama alanlarındaki yüksek talebi ve sayısız uygulamaları nedeniyle bilgisayarla görme (computer vision) ve insan-bilgisayar etkileşimi (human-computer interaction) araştırmalarının önemli alanlarını oluşturmaktadırlar ve son zamanların popüler araştırma alanları olarak ortaya çıkmaktadırlar.

İnsan eylemlerini tanıma işlemleri, insanların günlük yaşamlarında, video veya sensör girdilerinden insan faaliyetleriyle ilgili derinlemesine bilgi edinme konusundaki yetkinliği için de önemli bir rol oynamaktadır (Wang vd. 2018). Sağlık bakımı (health-care) (Zhao vd. 2013), video izleme (video surveillance) (Qin vd. 2016), sürüş güvenliği (driving-safety) (Yılmaz vd. 2018), jest tanıma (gesture recognition) (Kim vd. 2016, Yılmaz vd. 2020a), akıllı ev teknolojisi (smart home technology) (Fatima vd. 2013) ve bunun gibi uygulama alanları insan eylemlerini tanıma işlemlerinin kullanıldığı uygulama alanlarına örnek verilebilir. İnsan eylemlerini tanıma problemini genel olarak tanımlamak gerekirse şu şekilde tanımlanabilir: bir eylem, insan vücudunun hareketlerinin bir dizisidir ve bu kapsamda bilgisayar görme açısından insan eyleminin tanınması, girdi değerlerinden (örnek olarak video, görüntü ve bunun gibi) elde edilen gözlemlerin önceden tanımlanmış modellerle eşleştirilmesi ve daha sonra eylem türüne göre bir etiket atanması (action label) işlemidir (Cheng vd. 2015). İnsan eylemlerini tanımanın amacı, video içerisinde devam eden eylemlerin otomatik olarak analiz edilmesinin sağlanmasıdır. İnsan eylemlerini tanıma işleminin amacını detaylandırmak gerekirse şu şekilde detaylandırılabilir: Bir videonun yalnızca bir insan aktivitesinin yürütülmesini içerecek şekilde olduğu basit durumlarda, insan eylemlerini tanıma sisteminin amacı videoyu aktivite kategorisine göre doğru şekilde sınıflandırmaktır. Bununla birlikte daha genel durumlar içeren insan eylemlerini tanıma sistemlerinde, bir videoda meydana gelen tüm aktivitelerin başlama ve bitiş zamanları tespit edilerek insan faaliyetlerinin sürekli olarak tanınmasının sağlanması gerekmektedir (Aggarwal vd. 2011).

El hareketi tanıma işlemleri, karmaşık ortamlarda (cluttered environment) insan-bilgisayar etkileşimi (human-computer interaction), işaret dili tanıma (sign language recognition), el eylem analizi (hand action analysis), sürücü el davranış izleme (driver hand behavior monitoring) ve sanal gerçeklik (virtual reality) gibi birçok bilgisayar görme uygulaması için çok önemli bir görevdir (Ren vd. 2011, Plouffe ve Cretu 2016, Liang vd. 2016, Nakjai ve Katanyukul 2019, Tao vd. 2018, Mohammed vd. 2019). El hareketi, özel anlamının avuç içi ve parmak konumu ile dil merkezimiz aracılığıyla oluşturulduğu yüksek pratik değeri olan bir beden dilidir ve insan diyalogu için doğal ve sezgisel bir iletişim modu sağlar. El

hareketleri genel olarak statik ve dinamik el hareketlerinden oluşur. Adından da anlaşılacağı gibi, statik el hareketleri elin bireysel şeklini ifade eder ve dinamik el hareketleri bir dizi el hareketinden oluşur. El hareketleri çeşitlilik ve belirsizliğe sahip olduğundan ve zamandan ve kültürel geçmişten etkilendiğinden, farklı insanların farklı el hareketi tanımlamaları vardır (Mitra ve Acharya 2007, Chen vd. 2013).

İnsan eylemlerini tanıma işlemleri, literatür içerisinde özniteliklerin geleneksel el yapımı (geleneksel el yapımı temelli) veya derin bir sinir ağının çıktısı (derin öğrenme temelli) yoluyla üretilmesine bağlı olarak iki grupta incelenmektedirler. Son yıllarda, derin öğrenme temelli insan eylemi tanıma yöntemleri (Xiao vd. 2016, Wu vd. 2017), geleneksel el yapımı temelli insan eylemi tanıma yöntemlerinin temelini aştı ve videolarda eylem tanıma konusunda kayda değer bir ilerleme ortaya koydu (Zhang vd. 2017). El hareketi tanıma işlemlerinde ise literatürde geleneksel el yapımı özelliklerin (hand-crafted features) kullanımı erken araştırmalara hakim olmuştur. Bu yaklaşımların çoğu, el hareketi tanıma için el ten rengi, dokusu ve görünüm özelliklerini kullanmışlardır (Dardas ve Georganas 2011, Chevtchenko vd. 2018, Matilainen vd. 2016, Mirehi vd. 2019). Bununla birlikte, bu yöntemlerin başarıları sadece iyi hazırlanmış bazı ortamlarda bulunmaktadır. El hareketi tanıma görevi, gerçek hayattaki uygulamalarda oklüzyon, bakış açısı, aydınlatma değişiklikleri ve bunun gibi problemlerden dolayı zorlayıcı olmaya devam etmektedir. Nesne tespiti ve tanınması için son zamanlarda ortaya çıkan derin öğrenme yaklaşımları ile cesaretlendirilen birçok araştırmacı (Girshick vd. 2014, Ren vd. 2017, He vd. 2016), belirtilen bu sorunlarla başa çıkmak için çok sayıda derin öğrenme temelli yaklaşım önermiştir (Mohammed vd. 2019, Bose ve Kumar 2019). İnsan eylemlerini ve el hareketi tanıma işlemleri için bahsedilen bu durumlarla birlikte, yeni bir derin öğrenme temelli insan eylemi veya el hareketi tanıma yöntemi modelini sıfırdan eğitmek, bazı durumlarda eğitim işlemleri için büyük miktarda veri, yüksek işlemsel kaynaklar ve saatler gerektirir. Bu durumda derin öğrenme temelli insan eylemi ve el hareketi tanıma yöntemi modellerini uygulamanın oldukça zorlayıcı olmasını sağlar. Bahsedilen bu zorluklarla mücadele etmek için önceden eğitilmiş bir derin öğrenme modelini yeni bir eğitim işlemi için kullanmak daha uygun olacaktır (Azizpour vd. 2015, Chatfield vd. 2014). Bu durumu ifade eden

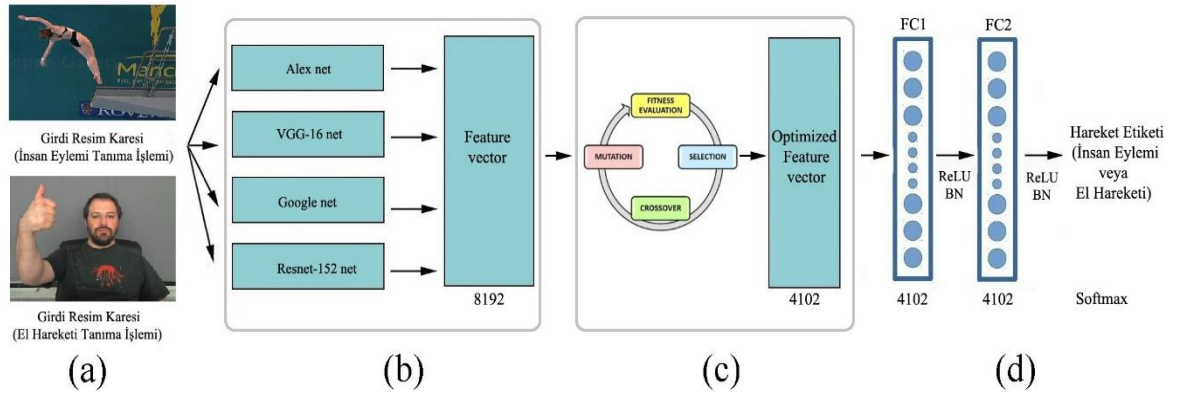
yapıya ise Transfer Öğrenme denilmektedir (Kaya vd. 2018). Hareket ve davranış analizi işlemleri durum çalışması içerisinde bahsedilen problemleri aşmak adına insan eylemlerini ve el hareketi tanıma görevleri için transfer öğrenimi kullanılarak detayları derin sinir ağı mimarisinin tasarlanması bölümünde bahsedilen (bölüm 3.3.2) yeni bir derin sinir ağı mimarisi tasarlanmış, uygulanmış ve sonrasında bu derin sinir ağı mimarisi kullanılarak insan eylemlerini ve el hareketi tanıma işlemleri gerçekleştirilmiştir. Ek olarak çalışma içerisinde, geliştirilen derin sinir ağı mimarimizin performansını değerlendirmek için deneysel analiz karşılaştırmaları yapılmıştır. Bunlar; seçilen diğer insan eylemi ve el hareketi tanımlama işlemlerinde kullanılan derin öğrenme mimarilerinin gerçekleştirilmesi ve diğer son teknoloji algoritmaların sonuçlarının alıntı yapılması işlemleridir. Buradaki amacımız girdi videolarından elde edilen insan eylemlerini ve el hareketlerini analiz ederek insan eylemlerini ve el hareketi tanıma işlemlerini optimize bir şekilde gerçekleştirmektir.

İnsan eylemlerini tanıma sistemleri, öznitelik gösterimlerinde giriş verisi olarak kullandıkları veri çeşidine göre iki gruba ayrılabilirler. Bunlar video tabanlı (video-based) ve sensor tabanlı (sensor-based) giriş veri değerleridir. Video tabanlı insan eylemi tanıma sistemleri (video-based human action recognition system) içerisinde, herhangi bir kameradaki insan hareketlerini içeren videolar analiz edilir. Diğer yandan sensör tabanlı insan eylemi tanıma sistemleri (sensor-based human action recognition system) içerisinde bir manyetometre (magnetometer), sıcaklık (temperature), ivmeölçer (accelerometer), jiroskop (gyroscope), bluetooth, akıllı telefon (smartphone) ve benzeri gibi sensörlerden gelen veriler kullanılarak işlemler gerçekleştirilmektedir (Wang vd. 2018, Aggarwal vd. 2011).

El hareketi tanıma sistemleri, genel olarak öznitelik gösterimlerinde girdi verisi olarak kullandıkları veri çeşidine göre iki grupta toplanabilirler. Bunlar statik ve dinamik el hareketleri girdi veri değerleridir. Burada statik el hareketleri kullanan sistemler, anlamlı bilgileri iletmek için yalnızca bir görüntü gerektirirken; dinamik el hareketleri kullanan sistemler bir hareketi gerçekleştirmek için bir dizi kare içerir. Burada amaç, çıplak elin belirli anlamları açıklamak için duruşlar yaptığı statik veya dinamik el hareketlerini

algılama ve tanıma problemini çözmektir. Ek olarak, el hareketlerini tanıma sistemleri birçok uygulamada önemli bir rol oynamaktadır. İşitme ve konuşma engelli kişiler için işaret dili tanıma (Nakjai ve Katanyukul 2019, Tao vd. 2018), sürücü dikkatinin dağılmasını azaltmak için sürücü el izleme ve el hareketi komutları (Le vd. 2016), insan ve makineler arasındaki arabirim için alternatif bir giriş yöntemi (Plouffe ve Cretu 2016, Dardas ve Georganas 2011), havada yazma etkileşimi (Huang vd. 2016), artırılmış ve görsel gerçeklik ortamlarında el nesnesi etkileşimi (Kang vd. 2017) ve bunun gibi diğer birçok uygulama alanları el hareketlerini tanıma sistemlerinin kullanıldığı uygulama alanlarına örnek verilebilir.

İlgili durum çalışmamız içerisindeki sistem, insan eylemi tanıma işlemi için video ve sensör verileri üzerinde gerçekleştirildiği için hem video tabanlı hemde sensör tabanlı insan eylemi tanıma sistemleri kategorisine ve el hareketi tanıma işlemi için statik el hareketi verileri üzerinde gerçekleştirildiği için statik el hareketleri kullanan sistemler kategorisine girmektedir ve çalışılan sisteminin metodolojisi dört ana adımdan oluşmaktadır. Bunlar sırasıyla Şekil 3.27'deki şemadan anlatılırsa; ilk olarak giriş verileri sisteme alınır. İkinci olarak, önceden eğitilmiş ağların kullanılması ile öznitelik çıkarım işlemleri (feature vector) gerçekleştirilir. Üçüncü aşama içerisinde, ikinci aşamada elde edilen öznitelikler genetik algoritmanın kullanılması ile optimize edilir (optimized feature vector). Son olarak, önerilen derin sinir ağının eğitilmesi ve sınıflandırması işlemleri yapılır ve insan eylemi ve el hareketi tanıma işlemleri eğitilmiş mimarinin kullanılması ile gerçekleştirilir.



Şekil 3.27 Obje tespiti ve analizi sistemi modeli için hareket ve davranış analizi işlemleri durum çalışmasının genel boru/akış hattı diyagramı

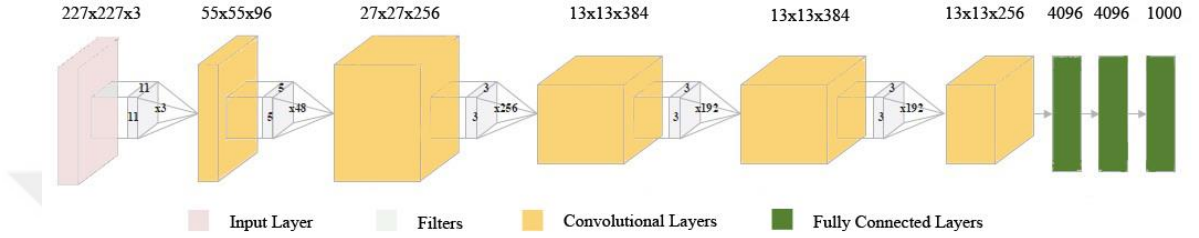
(a) Giriş video çerçevesi, (b) Öz nitelik çıkarma işlemleri, (c) Öz nitelik optimizasyonu işlemleri, (d) Derin sinir ağının eğitim, sınıflandırma işlemleri ve insan eylemlerini ve el hareketlerini tanımlama işleminin sonuç çıktısı

3.3.2 Derin sinir ağı mimarisinin tasarlanması işlemleri

Çalışılan obje tespiti ve analizi sistemi modelimizde, hibrit bir modele dayalı yeni bir derin öğrenme temelli mimari geliştirilmiştir. Geliştirilen mimari temel olarak transfer öğrenme metodolojisi ve genetik algoritmanın kullanılması ile önceden eğitilmiş ağlara dayalı olarak tasarlanmıştır. Burada, derin sinir ağı mimarisinin tasarımı için iki kritik ön işleme adımı yapılmıştır. Bu adımlardan ilki, bahsedilen durum çalışmalarının gerçekleştirilmesi sürecinde kullanılacak uygun bir derin öğrenme mimarisini tahmin etmektir. Ön deneyler, bir karma modülün daha iyi bir genel hassasiyet üretebileceğini ortaya koymaktadır. Bu nedenle, daha önce eğitilmiş mimarileri kullanarak bir hibrit bir modül üretilmiştir. Hibrit modül Alexnet (Krizhevsky vd. 2012), VGG-16 (Simonyan ve Zisserman 2014), GoogleNet (Szegedy vd. 2015) ve Resnet-152 (He vd. 2016) mimarilerini içermektedir. Bu derin sinir ağı mimarilerinin içerik detayları aşağıdaki paragraflarda açıklanmıştır.

AlexNet, ImageNet büyük ölçekli görsel tanıma mücadelesinde (ILSVRC-2012) tanıtılan öncü evrimsel sinir ağlarından biridir. Şekil 3.28'de mimarisi gösterilen AlexNet ağı, basit

bir mimariye sahiptir ve 8 katman derinliğindedir. Bu ağ, 5 adet konvolüsyon katmanından bazılarını takip eden 3 adet maxpool katmanı, 2 adet normalizasyon katmanı, 7 adet RELU (Rectified Linear Unit) katmanından oluşmaktadır. Ek olarak, mimaride 3 adet tam bağlı katmanı ve softmax katmanı vardır (Krizhevsky vd. 2012)



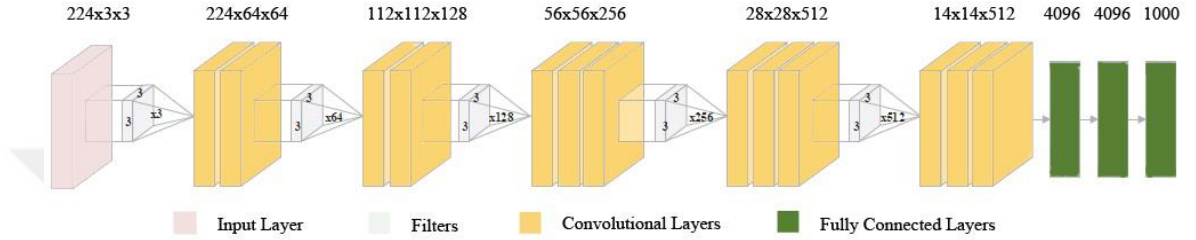
Şekil 3.28 AlexNet ağının temel mimarisi (Krizhevsky vd. 2012, Yılmaz vd. 2020b)

Yukarıdaki şekilde bulunan terimler içerisinde "Input Layer" girdi katmanını, "Filter" filtreyi, "Convolutional Layer" konvolüsyon katmanlarını ve "Fully Connected Layer" tam bağlantı katmanlarını temsil etmektedir.

VGG-16 ağı, ImageNet (Russakovsky vd. 2014) veritabanının kullanılması ile 1.3 milyondan fazla resim üzerinde eğitilmiş evrimsel sinir ağı mimarisidir. Bu ağ modelinin temel amacı, evrimsel ağ derinliğinin büyük ölçekli görüntü tanıma ayarlarındaki doğruluk üzerindeki etkisini araştırmayı amaçlamaktadır. Bu ağ modeli, tasarımı ve derinliği olan iki anahtar kelimeye sahiptir. Model, derinliği 16 katmana iterek önemli bir gelişme sağlar. Şekil 3.29'da gösterilen VGG-16 ağının mimarisi, baştan sona olmak üzere 3x3 evrimsel ve 2x2 ortaklama işlemleri içeren homojen bir yapı sunmaktadır. Bunlara ek olarak, ilgili mimari içerisinde üç adet tam bağlantı ve sınıflandırma işlemleri için bir adet softmax katmanı bulunmaktadır (Simonyan ve Zisserman 2014). Burada Softmax katmanı, bir dizi özelliğin K sınıflarından biri ile ilişkili olabileceği çoklu sınıf öğrenme sürecini sağlamak için kullanılır ve bu katman içerisinde kullanılan softmax fonksiyonunun matematiksel formülasyonu denklemde 3.3'de gösterilmiştir.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}, \quad j = 1, 2, \dots, K \quad (3.3)$$

Bu denklemde $\sigma(z)_j$ softmax sonucunu temsil ederken, z , çıkış katmanına girişlerin bir vektörünü temsil eder ve j , çıkış birimlerini indeksler. Çalışmada kullanılan mimari içerisinde girişten çıkışa doğru matrislerin yükseklik ve genişlik boyutları azalırken derinlik değerleri artmaktadır ve yaklaşık olarak 140 milyon parametre hesabı yapılmaktadır.

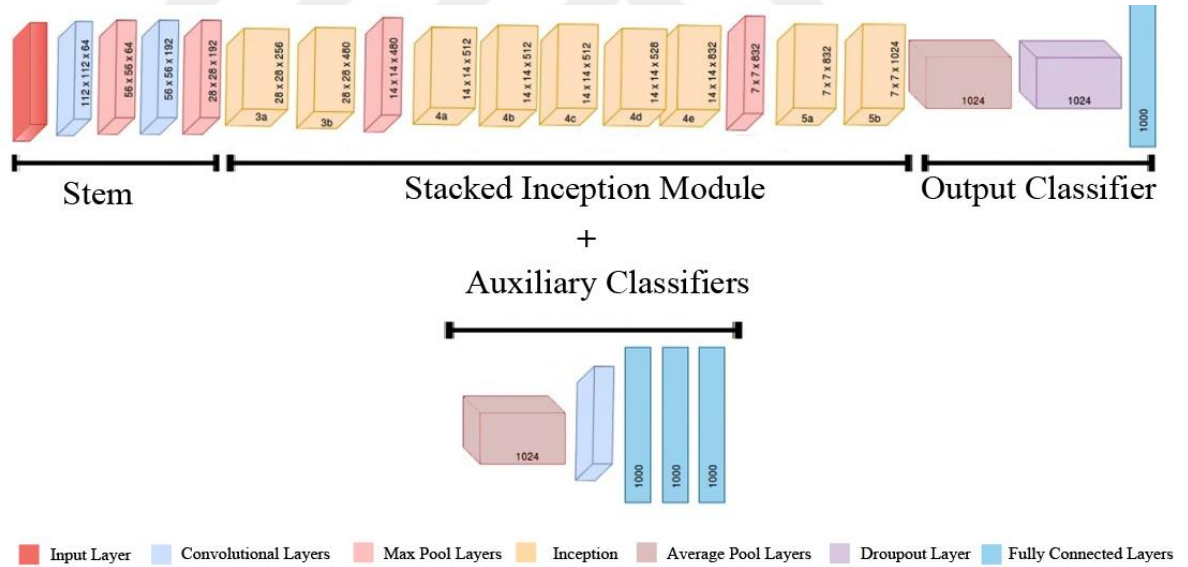


Şekil 3.29 VGG-16 ağının temel mimarisi (Simonyan ve Zisserman 2014, Yılmaz vd. 2020b)

Yukarıdaki şekilde bulunan terimler içerisinde "Input Layer" girdi katmanını, "Filter" filtreyi, "Convolutional Layer" konvolüsyon katmanlarını ve "Fully Connected Layer" tam bağlantı katmanlarını temsil etmektedir.

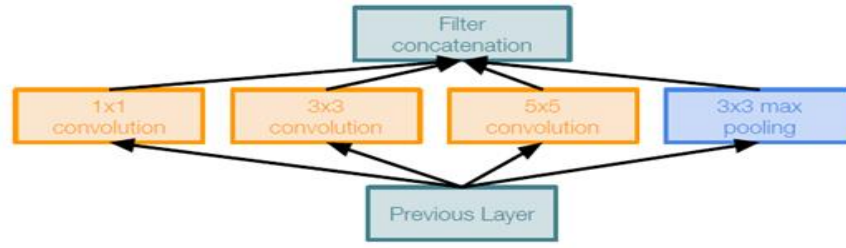
GoogleNet ağı, karmaşık bir mimariye sahip olmasına karşın evrişimli sinir ağı modelleri içerisinde genişlik kavramını kullanan ilk mimaridir. Model içindeki genişlik ögesi 'inception' adı verilen modüllerle elde edilmektedir. Inception modüllerindeki genişleme etkisi evrişim katmanlarındaki 1×1 , 3×3 , 5×5 filtrelerin ve 3×3 maksimum ortaklama işleminin paralel olarak gerçekleştirilmesiyle elde edilmektedir. Şekil 3.31.a'da gösterilen bu modüle "Naif Inception Modülü (Naive Inception Module)" denilmektedir. Ancak burada çıkış boyutunun büyüklüğü ve parametre sayısının paralel işlemler yüzünden çok büyümesi nedeniyle işlem karmaşıklığı artmaktadır. Bu sorunu çözmek için her Naif Inception evrişim katmanından önce 1×1 evrişim katmanı kullanılarak işlem yükü optimize edilmiş ve boyut azaltma işlemi gerçekleştirilmiştir. Şekil 3.31.b'de gösterilen bu modül (Boyut indirgemeli Inception modülü), Inception Modüllerinin birinci versiyonunu oluşturmaktadır. Ek olarak, bu modül ile işlem karmaşıklığına bulunan çözüm ile birlikte hız ve başarımlar elde edilmiştir.

GoogleNet 22 katmanlı bir mimariden oluşmaktadır ve ortaklama katmanları da dahil edilirse 27 katman derinliğe sahiptir. Şekil 3.30'da detayları gösterilen mimari stem (gövde), stacked inception module, auxiliary classifiers (yardımcı sınıflandırıcılar) ve output classifier (çıkış sınıflandırıcısı) olmak üzere dört ana bölümden oluşmaktadır. Bu mimari içerisinde evrişim katmanı, maksimum ortaklama, softmax katmanı, tam bağlantı katmanı, RELU layer, giriş katmanı ve çıkış katmanı gibi 144 katman ve 9 inception modülü bulunmaktadır. Mimari içerisinde evrişim katmanlarında 1x1, 3x3 ve 5x5 boyutunda filtreler, ortaklama işlemleri içerisinde 3x3 filtre, aktivasyon işlemleri için RELU aktivasyon fonksiyonu kullanılmaktadır (Szegedy vd. 2015). RELU fonksiyonu lineer olmayan bir aktivasyon fonksiyonudur ve bu fonksiyonun matematiksel formülasyonu denklemde 3.4'daki gibi tanımlanmaktadır. Bu denklem ile girdi negatif ise çıktı 0, pozitif ise çıktı kendi (ham) değeri olmaktadır.

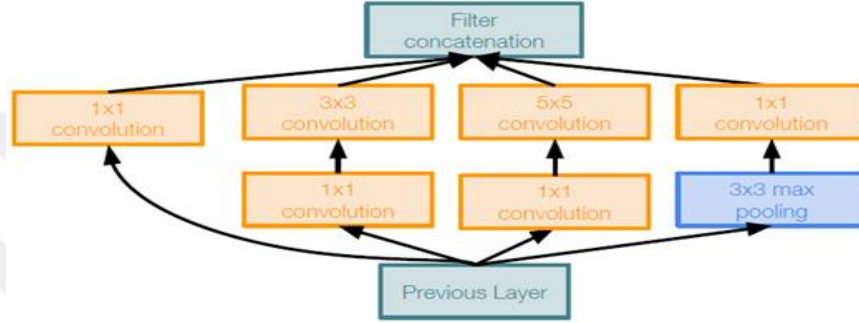


Şekil 3.30 GoogleNet ağının temel mimarisi (Szegedy vd. 2015, Yılmaz vd. 2020b)

Yukarıdaki şekilde bulunan terimler içerisinde "Input Layer" girdi katmanını, "Filter" filtreyi, "Convolutional Layer" konvolüsyon katmanlarını, "Max Pool Layer" maksimum havuz katmanlarını, "Average Pool Layer" ortalama havuz katmanlarını, "Dropout Layer" seyreltme katmanlarını ve "Fully Connected Layer" tam bağlantı katmanlarını temsil etmektedir.



(a)



(b)

Şekil 3.31 Naif Inception Modülü (Szegedy vd. 2015)

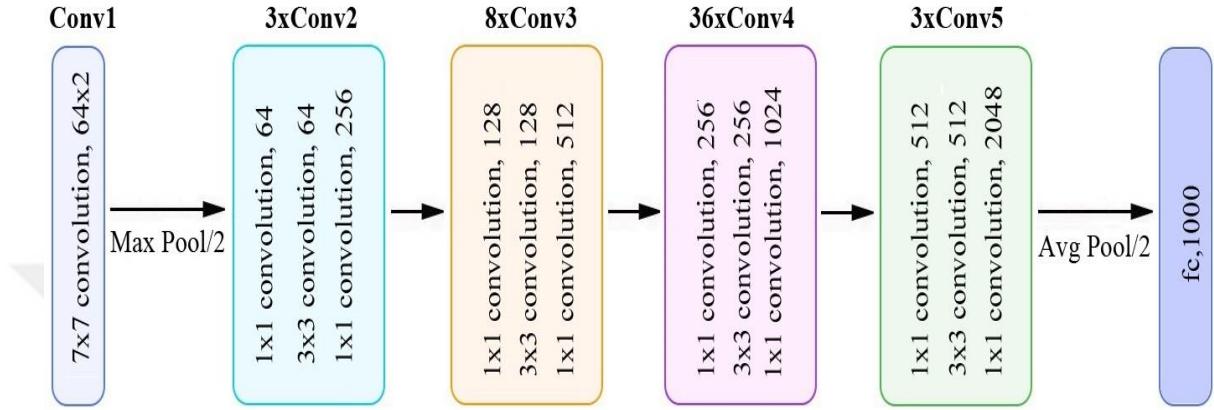
(a) Naif Inception modül versiyonu, (b) Boyut indirgemeli Inception modülü

$$f(x) = \max(0, x) \quad (3.4)$$

$$f(x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } x \geq 0 \end{cases}$$

ResNet-152 ağı modeli, eğitilmesi zor ağların eğitimini kolaylaştırmak için kalıcı bir öğrenme çerçevesi (residual learning framework) ortaya koymaktadır. Bu model içerisinde referans edilmemiş fonksiyonların öğrenilmesi yerine katman girişleri referans alınarak residual fonksiyonları öğrenilmesi üzerine bir derin sinir ağı mimarisi formülasyonu gerçekleştirilmiştir. Resnet152 ağının mimarisi, bahsedilen VGG-16 ağından temel alınarak oluşturulmuştur. Şekil 3.32’de temel mimarisi gösterilen bu ağ içerisinde evrişim katmanları 3x3 ve 1x1 filtreye sahiptir ve iki adet ortaklama işlemi gerçekleştirilmektedir. Burada altörnekleme işlemi 2 adıma (stride) sahip olan evrişimsel katmanlarla yapılmaktadır. Resnet-152 ağı ortalama havuz katmanı, 1000 yönlü tam bağlantı katmanı

ve softmax ile sona erer. Ek olarak, bu ağ içerisinde toplam ağırlıklı katman sayısı 152'dir (He vd. 2016).



Şekil 3.32 Resnet-152 ağının temel mimarisi (He vd. 2016, Yılmaz vd. 2020b)

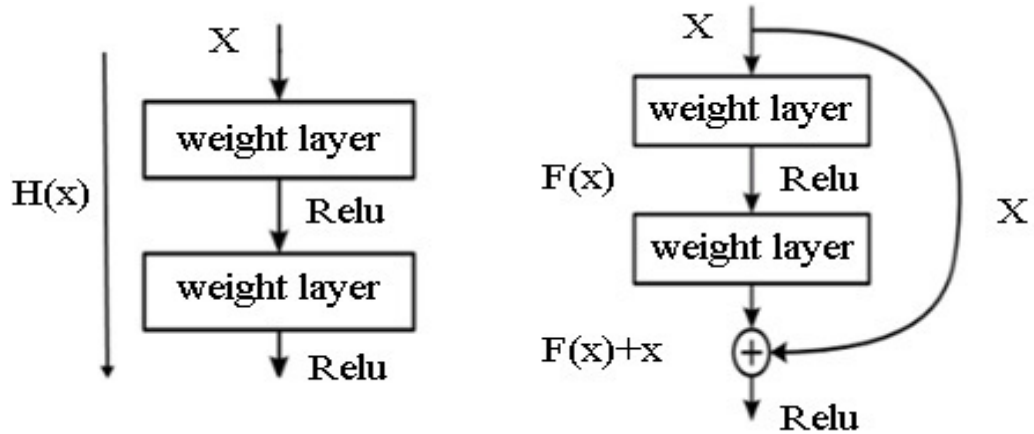
Derin sinir ağlarında ağ derinliği (katman sayısı) önemli bir parametredir. Derin sinir ağlarına teorik olarak bakıldığında katman sayısının artması ağın temsil kapasitesini (representational capacity) artırır ve bu durumunda daha başarılı ağ mimarileri oluşturması beklenir. Fakat pratikte bahsedilen bu durum geçerli değildir. Burada başarımın artmamasının sebebi "vanishing/exploiding gradients" ve optimizasyonun zorlaşması sorunlarıdır. Bu sorunları açıklamak gerekirse: "vanishing/exploiding gradients" sorunu, nöral ağların eğitilmesi sırasında güçlü aktivasyonları olmayan nöronların değerlerinin uzun katman zinciri arasında erimesi ve ağın eğitime katkı verememesi durumunu göstermektedir; optimizasyonun zorlaşması sorunu, derinlik arttıkça farklı yapılardan kaynaklı parametre oluşması sonucu ile parametre sayısının artması ve bu durumunda ağın eğitimini ve optimum çözümü bulma konusunu zorlaştırması durumunu ifade etmektedir. ResNet-152 ağının diğer derin sinir ağlarından farklı olarak en önemli katkılarından birisi ağın derinliğin artırırken yukarıdaki bahsedilen sorunlardan dolayı oluşabilecek negatif durumları engelleyebilmesidir. Kısacası ResNet ağı ile birlikte hem derinlik artırılabilir hem de daha hızlı ve başarılı şekilde ağlar eğitilebilmektedir. Yukarıda bahsedilen problemlerin çözümü için Resnet çalışması içerisinde oluşturulan residual blok

diyagram'ı şekil 3.33'de ve bunun formülasyonları denklem 3.5 ve denklem 3.6'da gösterilmiştir.

$$y = F(x, \{W_i\}) + x \quad (3.5)$$

$$y = F(x, \{W_i\}) + W_s x \quad (3.6)$$

Derin sinir ağı mimarilerinin genellikle girişten çıkışa kadar olan kısımları lineer olmayan bir $H(x)$ fonksiyonu ile haritalandırırken (Şekil 3.31.a), ResNet mimarisi içerisindeki şekil 3.31.b'de gösterilen yapıdaki bu yolda $H(x)$ yerine $F(x) = H(x) - x$ olarak tanımlanan başka bir lineer olmayan fonksiyon ile haritalandırılmaktadır. Buna ek olarak da girişten çıkışa bir kısayol bağlantıları yapılarak, x giriş değeri $F(x)$ fonksiyonuna aritmetik olarak eklenmektedir. Burada kısayol bağlantıları ile kimlik eşleşmesi (identity mapping) yapılmaktadır. Devamında, $F(x)+x$ fonksiyonu ReLU'dan geçirilmektedir. Dolayısıyla ikinci katmanın sonuna giriş değeri de eklenerek, geçmiş katmanlardaki değerlerin ileriki katmanlara daha güçlü bir şekilde iletilmesi sağlanmaktadır. Denklem 3.5 ve denklem 3.6 içerisinde x ve y , çalışılan katmanların giriş ve çıkış vektörlerini, $F(x, \{W_i\})$ fonksiyonu öğrenilecek residual haritalamayı temsil etmektedir (He vd. 2016).



Şekil 3.33 Residual blok diyagramı (He vd. 2016)

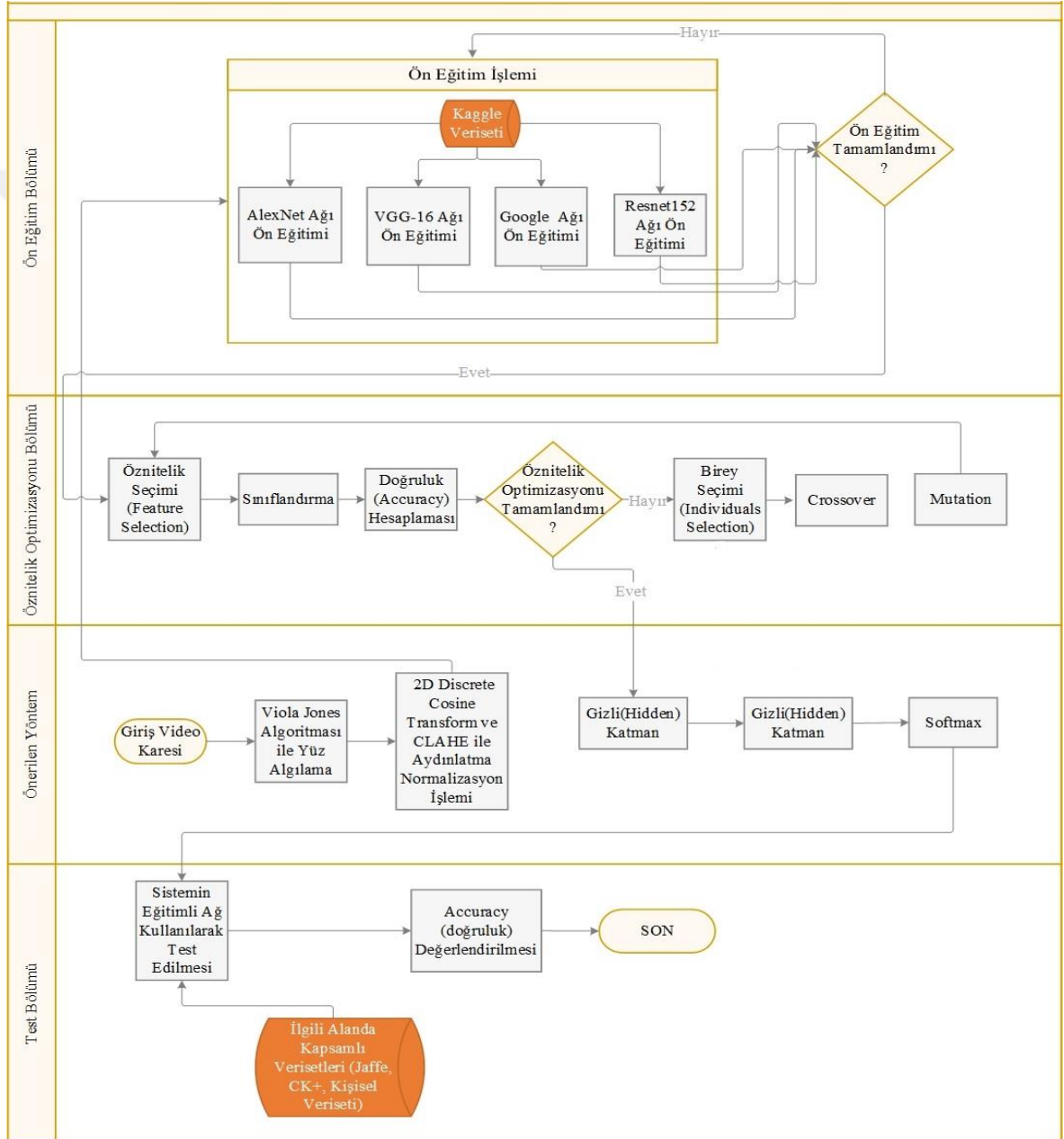
(a) Temel Derin Sinir Ağı Bloğu, (b) Residual Blok Bağlantılı Derin Sinir Ağı Bloğu

İkinci ön işlem adımı içerisinde önerilen derin sinir ağı mimarisini tasarlamak için seçilen derin öğrenme mimarilerine dayanan transfer öğrenme yaklaşımları araştırılmıştır. Çeşitli işlemlerde kullanmak üzere sınıflandırma yapılacak her uygulama veya her modelde, milyonlarca veri setinin elde edilmesi veya büyük bir veri setinin işlenmesi çeşitli faktörlerden dolayı mümkün olmamaktadır. Bu faktörlere model karmaşıklığı, veri setinin fazlaca büyüklüğü, donanım kaynakları kısıtları, zaman sınırlamaları ve bunun gibi örnekler verilebilir. Böyle durumlarda önceden eğitilmiş bir modelin (pre-trained model) kullanılması ile transfer öğrenme (transfer learning) yapılması çok daha uygun olabilir. Transfer öğrenme işleminde temel olarak iki yaklaşım kullanılır. Bunlardan ilki orijinal önceden eğitilmiş ağın korunması ile yeni eğitim veri setine dayanan ağırlıkların güncellenmesi işlemidir. İkincisi, önceden eğitilmiş ağı özellik çıkarıcı olarak kullanılması ve ardından genel bir sınıflandırıcı ile sınıflandırma işleminin yapılmasıdır.

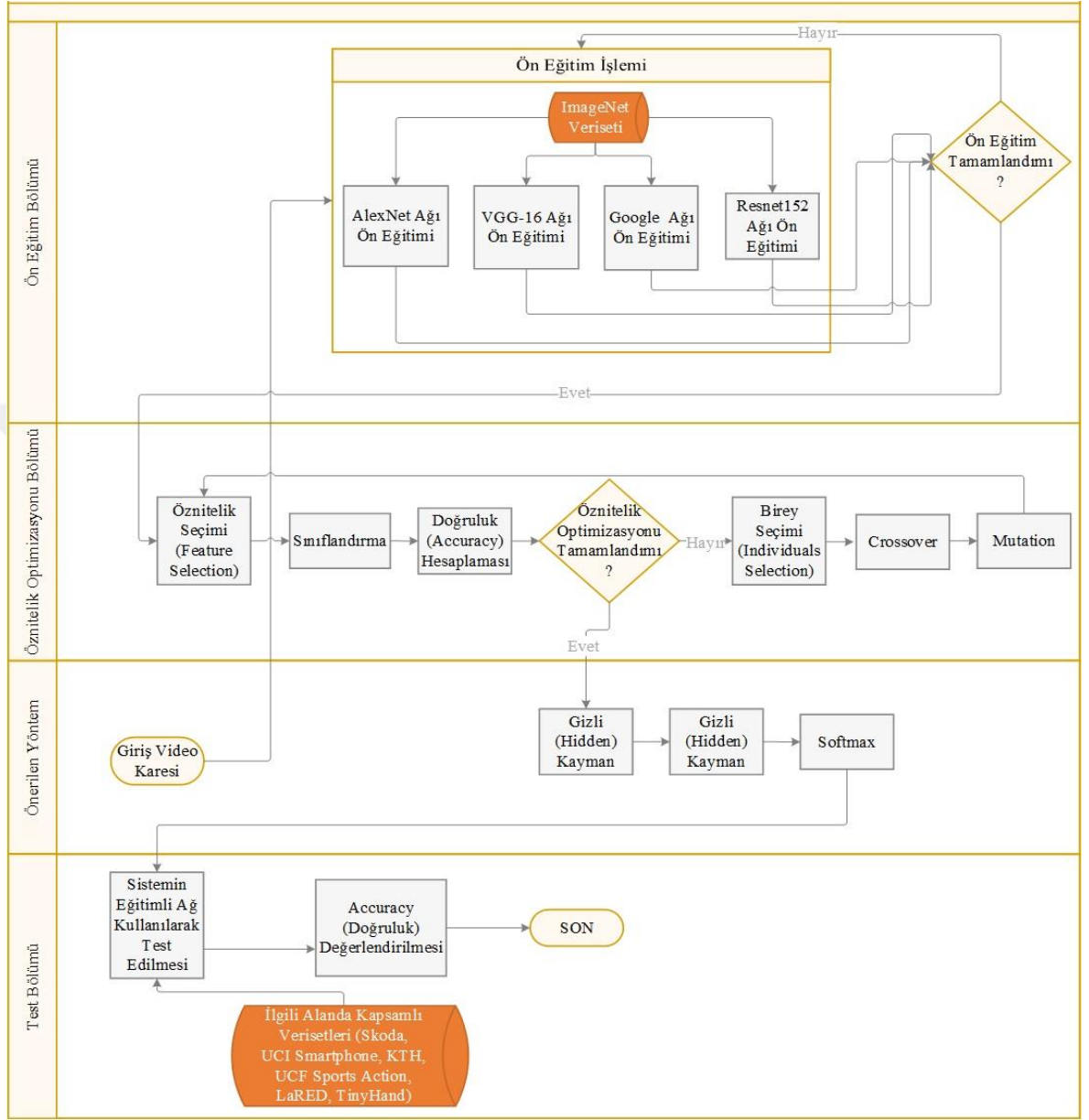
Sonuç olarak, daha önce belirtilen faktörler nedeniyle bahsedilen transfer öğrenme yaklaşımlarından ikincisi önerilen mimariye uyarlanmıştır. Önerilen mimariye, farklı mimarilerden elde edilen görüntülerin düşük ve yüksek özellikleri birleştirilmiştir. Şekil 3.34 ve Şekil 3.35'de bahsedilen durum çalışmaları (bölüm 3.3.1.1) için önerilen derin sinir ağı mimarisinin yapısı gösterilmiştir. Bu şekillerde ön eğitim bölümünde gösterilen önceden eğitilmiş ağlar, genel görüntü özellikleri için özellik çıkarıcıları olarak görev yaparlar. Ek olarak, bu şekiller içerisinde öznetelik optimizasyonu bölümünde gösterilen katmanlar genetik algoritma ile gerçekleştirilen öznetelik optimizasyonu sürecini gösterir ve son üç katman, öğrenme işlemi için tamamen bağlı katmanları ve sınıflandırma işlemi için softmax sınıflandırıcıyı temsil eder. Seçilen önceden eğitilmiş derin öğrenme mimarileri tarafından üretilen özelliklerin detayları aşağıdaki gibi özetlenmiştir:

- AlexNet Ağı: Her bir resim karesi üzerinde, Şekil 3.28'de gösterildiği gibi son bağlanan tamamen bağlı katmandan 2048 boyutlu bir öznetelik çıkarılır.
- VGG-16 Ağı: Her bir resim karesi üzerinde, Şekil 3.29'de gösterildiği gibi son bağlanan tamamen bağlı katmandan 2048 boyutlu bir öznetelik çıkarılır.

- GoogleNet Ağı: Her bir resim karesi üzerinde, Şekil 3.30'da gösterildiği gibi son bağlanan tamamen bağlı katmandan 2048 boyutlu bir öznitelik çıkarılır.
- Resnet152 Ağı: Her bir resim karesi üzerinde, Şekil 3.32'de gösterildiği gibi son bağlanan tamamen bağlı katmandan 2048 boyutlu bir öznitelik çıkarılır.



Şekil 3.34 Duygu analizi işlemleri durum çalışması için çalışılan sistemin akış şeması gösterimi



Şekil 3.35 Hareket ve davranış analizi işlemleri durum çalışması için çalışılan sistemin akış şeması gösterimi

Önerilen mimaride içerisinde ilk olarak, seçilen derin öğrenme mimarileri ilgili verisetleri ile eğitilerek ön eğitim işlemi gerçekleştirilmiştir. Burada ön eğitim işlemi içerisinde duygu analizi durum çalışması için 35887 adet görüntüden oluşan Kaggle veriseti (Anonymous 2019e) ve hareket davranış analizi için durum çalışması için yaklaşık 1.3 milyondan oluşan

ImageNet veriseti (Anonymous 2016) kullanılmıştır. İkinci olarak, dört derin öğrenme mimarisinden elde edilen öznitelikler 8192 boyutlu bir öznitelik vektörü oluşturmak için bir araya getirilmiştir. Üçüncü olarak, birleştirilen öznitelik vektörü detayları bölüm 3.3.3'de belirtilen genetik algoritmanın kullanılması ile optimize edilmiştir ve bu işlemin sonucunda optimize edilmiş 4102 boyutlu öznitelik vektörü elde edilmiştir. Dördüncü olarak, optimize edilen öznitelik vektörü sırasıyla öğrenme ve sınıflandırma işlemlerini gerçekleştirmek için üç tam bağlı katmandan geçirilmiştir. Burada üç gizli katmanın eklenmesiyle, ağıımızın öğrenme yeteneğinin artırılması hedeflenmiştir. Son olarak, yeni eğitilmiş modelimiz ile ilgili verisetleri üzerinde deneysel analiz karşılaştırmaları yapılmıştır.

3.3.3 Derin sinir ağı mimarisinin eğitilmesi ve optimizasyonu işlemleri

Çalışılan derin sinir ağı mimarisi'nin eğitilmesi ve optimizasyonu işlemleri içerisinde ilk olarak detayları derin sinir ağı mimarisinin tasarlanması bölümünde bahsedilen dört derin öğrenme mimarisinden elde edilen birleştirilmiş öznitelik vektörü, eğitim ve sınıflandırma işlemleri gerçekleştirilmeden önce önerilen mimarimizin daha verimli çalışabilmesi için genetik algoritma kullanılarak optimize edilmiştir. Burada amaç, daha az ama en ideal (optimal) özellikleri içeren bir öznitelik vektörü elde etmektir.

John Holland'ın temellerini geliştirdiği ve evrimsel hesaplama algoritmalarının ilki ve en bilineni olan genetik algoritmalar kısaca, en iyinin hayatta kalması ilkesine dayanarak çözüm uzayında en iyi çözümü bulmaya çalışan arama yöntemidir (Holland 1992a,b, Goldberg 1988). Charles Darwin'in “En güçlü ve en akıllı bireyler değil, değişime en iyi adapte olan bireyler yaşam mücadelesini sürdürür” mealindeki bir sözü üzerine inşa edildiği belirtilen Genetik Algoritmalar, optimizasyon problemlerini çözmeyi amaçlamaktadırlar. Genetik algoritmaların işleyişinde, ilk olarak her biri bir çözümü temsil eden rastgele oluşturulan bireylerden meydana gelen bir başlangıç popülasyonu ile çalışmaya başlanır. Devamında, her bir iterasyon sürecinde mevcut jenerasyondan seçilen bireylerin çaprazlanmasıyla veya mutasyona uğratılmasıyla yeni bir popülasyon elde edilir. Bu yeni jenerasyon sonrasında öncekin jenerasyonun yerini alır. Her yeni jenerasyon

içerisinde daha iyi genetik özelliklere sahip bireylerden oluşan bir popülasyonun oluşması umut edilir. Buradaki işlemler içerisinde, çaprazlama ve mutasyon işlemleri bir çaprazlama ve mutasyon olasılığına göre yapılır (Gülcü 2019). Genetik algoritmalarla optimizasyon problemlerinin çözümüne örnek verilirse; ilk olarak, Minaei-Bidgoli vd. (2003) çalışmasında bir veritabanı sisteminde daha önce saklanan veriler üzerinden elde edilen özellikleri kullanılması ile öğrencilerin son notlarını tahmin edecek şekilde sınıflandırmak için bir yöntem önermiştir. Bu yöntemde, bir grup sınıflandırıcıyı optimize etmek adına genetik algoritma kullanılmıştır. Buradaki amaç öznelik vektörlerinin optimize edilmiş bir şekilde ağırlıklandırılmasını sağlamaktır. Bu yaklaşım ham sınıflandırma üzerinde güçlü bir artış sağlamıştır. Handels vd. (1999) çalışmasında, dermatolojide deri tümörlerinin bilgisayar destekli analizine yeni bir yaklaşım sunmaktadırlar. Bu çalışmada, “nevositik (nevocytic)” “nevi” ve “malign (malignant)” melanomlarını otomatik olarak tanımak için yüksek çözünürlüklü cilt yüzeyi profilleri araştırılmıştır. Burada, özellik seçimi bir optimizasyon problemi olarak tanımlanmıştır ve tanıma sisteminin sınıflandırma performansını artırmak adına özellik seçimi için genetik ve açgözlü algoritmalar (greedy algorithm) kullanılmıştır. Burada çalışılan optimize edilmiş tanıma sisteminin kullanılması ile sınıflandırmada yaklaşık olarak % 98 doğruluk oranı elde edilmiştir. Bununla birlikte, optimizasyon problemlerini çözmek için genetik algoritmalarından farklı yöntemler vardır. Bunlara örnek olarak; Karim vd. (2018) çalışmasında verileri işlemek için temel olarak derin otomatik kodlayıcıları optimize etmeyi amaçlayan yeni bir derin otomatik kodlayıcı (Deep auto-encoder (DA)) tabanlı mimari önermişlerdir. Bu derin öğrenme mimarisi, parametreleri optimize etmek için Taguchi yöntemini DA tabanlı bir mimariye entegre eder. Sonuçlar, bu mimarinin birden fazla parametreyi optimize etmeyi başardığını göstermektedir. Bu durum, birkaç deneysel denemeden sonra daha ölçülebilir ve verimli verilerin elde edilmesini sağlamaktadır. Kumaran vd. (2018) çalışmasında kısıtlanmamış ortamlardan elde edilen videolardaki insan eylemlerini tanımak için hibrit bir CNN (Konvolüsyon Sinir Ağı) - GWO (Gri Kurt Optimizasyonu) yaklaşımını önermişlerdir. Bu yaklaşımın amacı, bir CNN sınıflandırıcısını “gradyan iniş (gradient descent)” ve “GWO” algoritmalarının yerel ve küresel keşif yetenekleri ile eğiterek sınıflandırma doğruluğunu arttırmaktır. Burada, optimal ağırlık başlangıç değerlerini bulmak için GWO algoritması

kullanılmıştır. Bu çalışmada önerilen yaklaşım, geleneksel yaklaşımlara kıyasla %99.9 tanıma doğruluk oranı ile daha iyi sonuçlar vermiştir. Karim vd. (2019) çalışmasında tıbbi veri işleme için DA ve enerji spektral yoğunluğu (Energy spectral density (ESD)) kullanılarak tasarlanan yeni bir çerçeve önermişlerdir. Bu çalışmanın yeniliği, ESD işlevini kullanarak seçilen özellikleri benzersiz bir derin seyrek otomatik kodlayıcılar modeline (deep sparse auto-encoders model) beslemektir. Bu durum, önerilen mimarinin optimize edilmesini sağlamıştır. Önerilen derin seyrek oto-kodlayıcılar mimarisi literatürde var olan çeşitli yöntemlerle karşılaştırıldığında, genel doğruluk oranının artırdığı ve derin otomatik kodlayıcı işlem süresini hızlandığı tespit edilmiştir.

Yapılan öznitelik optimizasyonu işleminin akış şeması, her bir durum çalışması için Şekil 3.33 ve Şekil 3.34'ün öznitelik optimizasyonu bölümlerinde gösterilmiştir. Gerçekleştirilen genetik algoritma ile öznitelik optimizasyonu işleminde ilk olarak, her biri 8192 elementin bir öznitelik vektörüne karşılık gelen 8192 genden oluşan belirli bir miktarda birey ile başlangıç popülasyonu oluşturulur. Ardından, elde edilen bireylerden rastgele öznitelik seçim işlemi gerçekleştirilmiştir. Bu aşamada her bir birey öznitelik vektörlerini elde etmek için kullanılmaktadır. Öznitelik seçim işlemi gerçekleştirildikten sonra öznitelik vektörü kullanılarak ileride detayları bahsedilen makine öğrenme modeli (SVM ve RBF) ile sınıflandırma işlemi yapılmıştır. Devamında, elde edilen sınıflandırma değerine göre doğruluk hesaplaması gerçekleştirilerek uygunluk değerlendirmesi yapılmıştır. Ardından bir sonraki aşama içerisinde bireyler seçilir. Burada seçim işlemi popülasyonun yüksek doğruluğa sahip bireylerinden ve çeşitliliği artırmak adına bazı diğer bireylerden rastgele yapılmıştır. Sonrasında, seçilen bireyler üzerinde çapraz geçiş (cross-over) ve mutasyon işlemleri (mutation) gerçekleştirilir. Bu işlem sonrasında mevcut iterasyon tamamlanır ve sonraki iterasyona geçilir. Sonlandırma kriteri sağlanana kadar iterasyonlar yukarıda bahsedildiği şekilde tekrar edilir.

Genetik algoritma ile öznitelik optimizasyonu işlemi konfigürasyonu, Cilla vd. (2009) çalışmasına göre yapılandırılmıştır. Popülasyon, öznitelik vektörünün büyüklüğüne karşılık gelen 8192 birey ile başlatılmıştır. Mutasyon oranı $\frac{1}{8192}$ olarak ayarlanmıştır. Her

iterasyondan sonra, genetik algoritma turnuva seçim (tournament selection) yöntemi ile popülasyonun % 2'si arasından en iyi bireyleri (best individuals) seçmektedir ve ayrıca diğer bireylerin %2'si arasından 0.02 olasılıkla rastgele bireyler seçilmektedir. Turnuva seçim yöntemi (Jebari ve Madiafi 2013), sıralamaya dayalı seçim yöntemlerinin bir çeşididir. Burada rastgele bir dizi k adet bireyler seçilir. Devamında, seçilen bu bireyler göreceli uygunluklarına göre sıralanırlar ve en uygun bireyler bir sonraki iterasyonda kullanılmak üzere seçilir. Tüm süreç, üzerinde çalışılan popülasyon için popülasyon içerisindeki birey sayısı (n) kadar kez tekrarlanır. Dolayısıyla, seçilecek her bireyin olasılığı ($p(i)$) denklem 3.7'deki formülasyon ile verilmiştir. Genetik algoritma'nın sonlandırma kriteri, modelin doğruluğu değişmeden kaldığında veya on yineleme geçtiğinde durması'dır.

$$p(i) = \begin{cases} \frac{c_{n-1}^{k-1}}{c_n^k}, & \text{if } i \in [1, n - k - 1] \\ 0 & \text{if } i \in [n - k, n] \end{cases} \quad (3.7)$$

Genetik algoritma içerisinde uygulanan makine öğrenme modeli'nde destek vektör makinesi (SVM) ve radio basis function (RBF) fonksiyonu kullanılmıştır (Hsu vd. 2003, Pham ve Nguyen 2018, Cao vd. 2012). Burada bahsedilen modeli optimize etmek için RBF çekirdeğinin γ ve C olan iki temel parametresi dikkate alınmıştır. Burada γ bir eğitim örneğinin etkisini tanımlar ve C yanlış sınıflandırma ile karar yüzeyinin sadeliği arasındaki dengeyi (trade-off) belirler. Ek olarak, SVM için en optimal γ ve C çiftini bulmak için loose grid-search arama tekniğinden (Hsu vd. 2003) faydalanılmıştır. Burada kullanılan önceden belirlenmiş değer aralıkları denklem 3.8'de gösterilmiştir.

$$\begin{aligned} C &= 2^a, \quad a \in \{-5, -3, -1, 1, \dots, 13, 15\} \\ \gamma &= 2^b, \quad b \in \{-15, -13, -11, -9, \dots, 1, 3\} \end{aligned} \quad (3.8)$$

Genetik algoritma ile öznetelik optimizasyonu işlemi gerçekleştirildikten sonra ikinci olarak bu bölüm içerisinde, detayları derin sinir ağı mimarisinin tasarlanması işlemleri bölümünde belirtilen çalışılan derin sinir ağı yapısı kullanılarak ağımız imagenet ve kaggle veri setleri ile eğitilmiş ve sonrasında yeni eğitilmiş modelimiz ile oluşturulan veri setimiz üzerinde

deneysel çalışmalar yapılmıştır. Eğitim ve test işlemleri sırasında kullanılan her bir verisetine ait detaylar ve gerçekleştirilen deneysel çalışmalar deneysel sonuçlar ve karşılaştırma bölümünde (bölüm 4) verilmiştir. Mevcut verilerin %70'i eğitim, %20'si test, %10'u onaylama verisi olarak seçilmiş ve kullanılmıştır. Ağ mimarisinin eğitim süreci daha önce tanımlanmış bir bilgisayarda, GPU desteği olmadan, duygu analizi işlemleri durum çalışması ve hareket davranış analizi işlemleri durum çalışması için sırasıyla yaklaşık 40 ve 90 saat (insan eylemlerini tanımlama işlemi için 60 saat, el hareketi tanımlama işlemleri için 30 saat) boyunca gerçekleştirilmiştir. Burada eğitimler duygu analizi işlemleri durum çalışması için 200 iterasyon ve hareket davranış analizi işlemleri durum çalışması için 400 iterasyonda (insan eylemlerini tanımlama işlemi için 300 iterasyon, el hareketi tanımlama işlemleri için 100 iterasyon) durdurulmuştur. Ek olarak bu bölümde, kaggle veri seti ile eğitim işlemi gerçekleştirilmeden önce, duygu analizi işlemleri durum çalışması işlemleri içerisinde eğitimin kalitesini geliştirmek adına bazı ön işlemler yapılmıştır. Bunlardan ilki, yüz algılama işleminin yapılmasıdır. Bu işlem içerisinde yüz sınırlayıcı kutular belirlenmiş ve belirlenen yüz sınırlayıcı kutular ile kırpma işlemi gerçekleştirilmiştir. Burada yüz sınırlayıcı kutuları bulmak için Viola Jones tarafından geliştirilen yüz algılama algoritması kullanılmıştır (Viola ve Jones 2004). Bu algoritma; integral görüntü, adaboost öğrenme algoritması ve haar-cascade sınıflandırıcısı olmak üzere üç temel bölümden oluşmaktadır. Burada integral görüntü dikdörtgen özelliklerinin çok hızlı bir şekilde hesaplamasını gerçekleştirmektedir ve bunun matematiksel formülasyonu denklem 3.9'de gösterilmiştir.

$$ii(x, y) = \sum_{\substack{x' \leq x \\ y' \leq y}} i(x', y') \quad (3.9)$$

Denklem 3.9 içerisinde $ii(x, y)$ integral görüntüyü, (x, y) imajın konumlandığı noktayı, $ii(x, y)$ girdi olarak verilen resmi temsil etmektedir. Adaboost algoritması ise hızlı bir şekilde hem en iyi haar-benzeri öz nitelikleri seçilmesini hemde öz niteliklerin sınıflandırmasını sağlamaktadır. Haar cascade sınıflandırıcı algoritması; dijital bir resim ya

da video karesi içerisinde bulunan belirli bir nesnenin tespit edilmesi amacıyla kullanılmaktadır. Haar-cascade sınıflandırıcı algoritmasının karar verme mekanizması denklem 3.10 ve 3.11’de gösterilmiştir.

$$C_m = \begin{cases} 1, & \sum_{i=0}^{l_m-1} F_{m,i} > \theta_m \\ 0, & \text{aksi halde} \end{cases} \quad (3.10)$$

$$F_{m,i} = \begin{cases} \alpha_{m,i}, & \text{if } f_{m,i} > t_{m,i} \\ \beta_{m,i}, & \text{aksi halde} \end{cases} \quad (3.11)$$

Burada, $f_{m,i}$ 2-boyutlu integrallerin ağırlıklı toplamıdır ve aynı zamanda i . öznetelik çıkarıcı için karar eşiğidir. $\alpha_{m,i}$ ve $\beta_{m,i}$, i . öznetelik çıkarıcı ile bağıntılı sabit değerlerdir. θ_m , sınıflandırıcı için karar eşiğidir. Bahsedilen yüz algılama algoritması ile haar dalgacık şablonlarına benzeyen 24x24 piksel boyutlarındaki dikdörtgen özellikli pencereler tüm resim boyunca kaydırılarak yüz görüntülerini algılama işlemi gerçekleştirilmiştir. İkinci olarak, ilk aşamada sonucunda elde edilen resim kareleri üzerinde ön işleme işlemleri gerçekleştirilmiştir. Burada kırılma işlemi yapılmış görüntüler ilk olarak gri tonlama formata dönüştürülmüş ve ardından bu görüntülere 2D Discrete Cosine Transform (Chen vd. 2006) ve adaptif histogram eşitleme tekniği olan CLAHE (Contrast Limited Adaptive Histogram Equalization) birlikte uygulanarak aydınlatma normalizasyon işlemi yapılmıştır (Thamizharasi ve Jayasudha 2016). CLAHE yöntemi (Zuiderveld 1994) ile gri tonlamalı görüntüler 8x8’lik bölgelere bölünmüş ve sonrasında her bir bölgeye histogram eşitlemesi uygulanmıştır. Ardından, elde edilen görüntüler 224 x 224 piksele yeniden boyutlandırılmış ve 3 kanala dönüştürülmüştür. Üçüncü olarak, ikinci aşamada elde edilen sonuç resim karelerinden veri artırma teknikleri kullanılarak sentetik görüntü verileri üretilmiştir. Buradaki amacımız elimizde veri sayısının çoğaltılması ile önerilen derin sinir ağımızın daha iyi eğitilmesinin sağlanmasıdır. Kullanılan veri artırma işlemleri döndürme (rotation), kaydırma (shifting), eğiltme (skew), ölçekleme (scaling), gürültü (noise), ve renk değişimi’dir (color jittering).

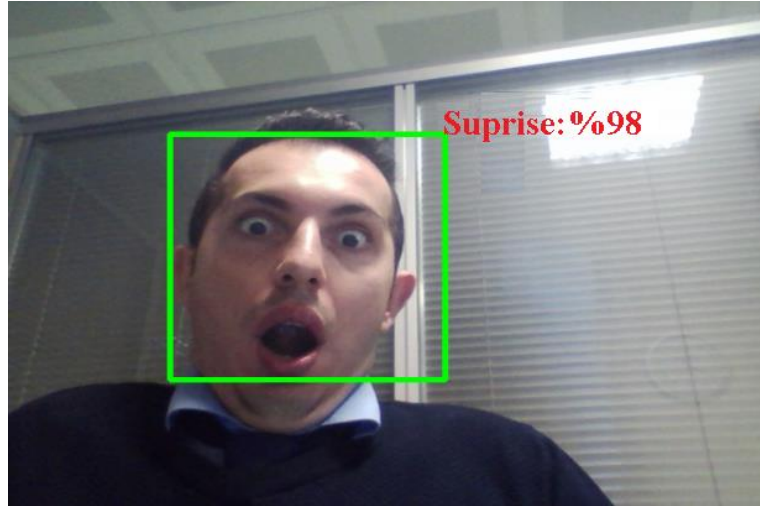
Sonuç olarak, obje tespiti ve analizi sistemi modelinin genel mimarisinin son aşaması olan video analizi işlemleri tamamlanarak obje tespiti ve analizi sistemi başarılı bir şekilde gerçekleştirilmiştir. Sistem içerisinde bahsedilen durum çalışmaları için transfer öğrenimi ve genetik algoritma kullanılarak yeni bir derin sinir ağı mimarisi geliştirilmiş ve bu geliştirilen derin sinir ağı mimarisi kullanılarak bahsedilen durum çalışmaları (duygu analizi işlemleri, hareket ve davranış analizi işlemleri) başarılı bir şekilde gerçekleştirilmiştir. Sistemde önceden eğitilmiş ağlara dayalı olarak geliştirilen hibrit modül için kullanılan derin öğrenme mimarileri Alexnet, VGG-16, GoogleNet ve Resnet-152'dir.

Geliştirilen obje tespiti ve analizi sistemi modelinin mimarisi üç ana aşamada gerçekleştirilmiştir. Bu aşamalar obje tespiti ve tanımlama işlemleri, video büyütme işlemleri ve video analizi işlemleri'dir. Bu aşamaların detayları yukarıdaki paragraf ve bölümlerde detaylıca bahsedilmiştir. Bu bahsedilen aşamalar gerçekleştirildiğinde çalışılan ağ mimarisi ile yeni eğitilmiş modelimiz elde edilmiştir. Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış resim kareleri Şekil 3.36, Şekil 3.37, Şekil 3.38, Şekil 3.39, Şekil 3.40 ve Şekil 3.41'de gösterilmiştir ve ek olarak, eğitilmiş modelimiz ile her bir obje tespiti ve analizi sistemi modelinin genel mimarisinin aşamalarının entegre edilmiş gösterimi Şekil 3.42'de gösterilmiştir. Şekil 3.42'deki resim karesinde her bir obje tespiti ve analizi sistemi modelinin genel mimarisi aşamasında elde edilen sonuçlar, sırasıyla aynı girdi video karesi üzerinden gösterilmiştir ve buna ek olarak ilgili resimde video büyütme ve video analizi işlemleri aşamasında Şekil 3.42.c ve Şekil 3.42.d'de gösterilen resim kareleri içerisinde üst alandaki resimler duygu analizi işlemleri durum çalışmasında elde edilen sonuç resim karelerini ve orta ve alt alandaki resimler hareket ve davranış analizi durum çalışmasında elde edilen sonuç resim karelerini göstermektedir (orta ve alt alandaki resimler sırasıyla insan eylemlerini ve el hareketi tanımlama işlemlerinde elde edilen sonuç resim karelerini göstermektedir). Ek olarak, geliştirilen derin sinir ağı mimarimizin performansını değerlendirmek için deneysel analiz karşılaştırmaları yapılmıştır. Bunların detayları deneysel sonuçlar ve karşılaştırma bölümünde (bölüm 4) anlatılmıştır. Burada çalışılan ağımızın, her bir durum çalışması için kapsamlı veri setleri üzerinde elde edilen

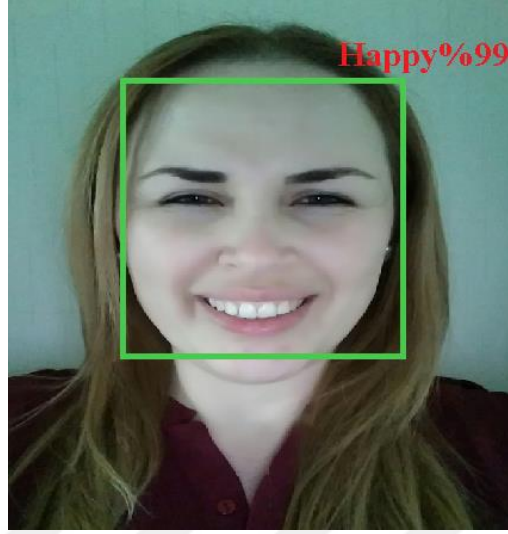
doğruluk oranları ile detayları bahsedilen bizim tarafımızdan gerçekleştirilen diğer dört derin sinir ağına ve diğer son teknoloji algoritmalara göre daha başarılı ve sağlam olduğu tespit edilmiştir.



Şekil 3.36 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 1



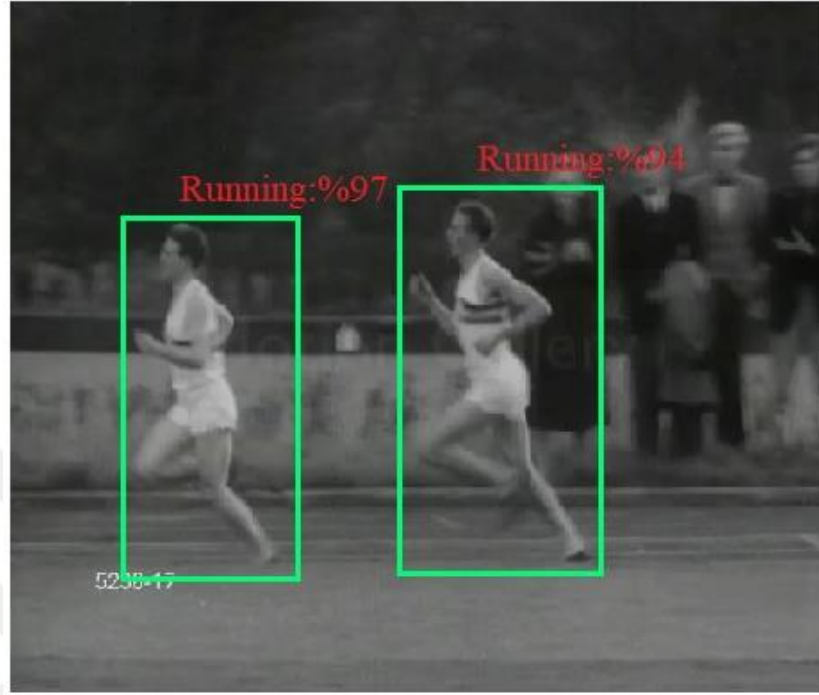
Şekil 3.37 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 2



Şekil 3.38 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 3



Şekil 3.39 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 4



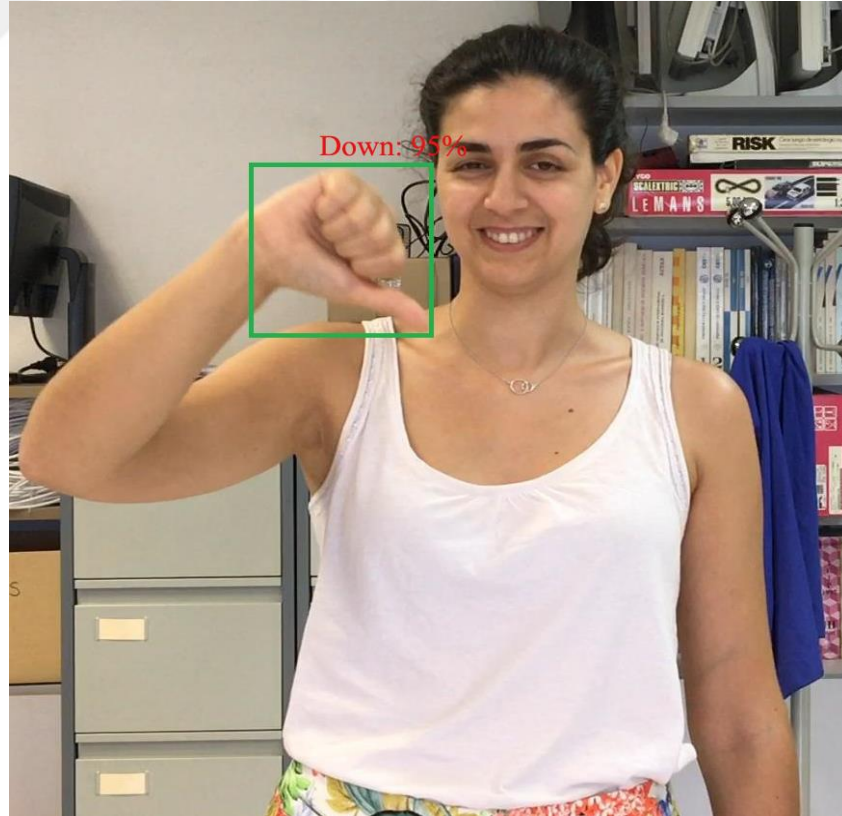
Şekil 3.40 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 5



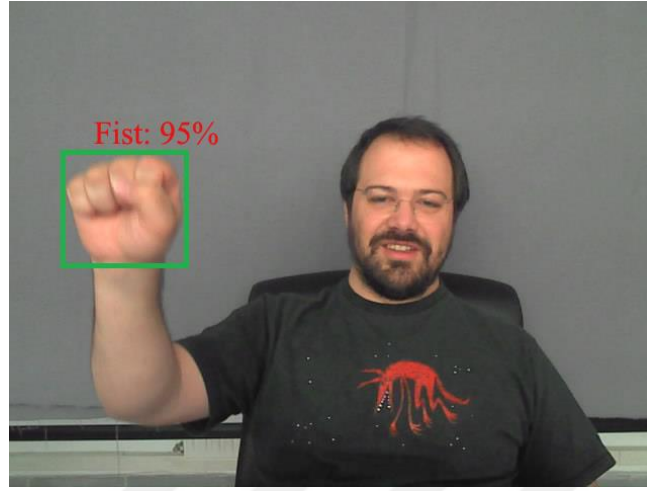
Şekil 3.41 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 6



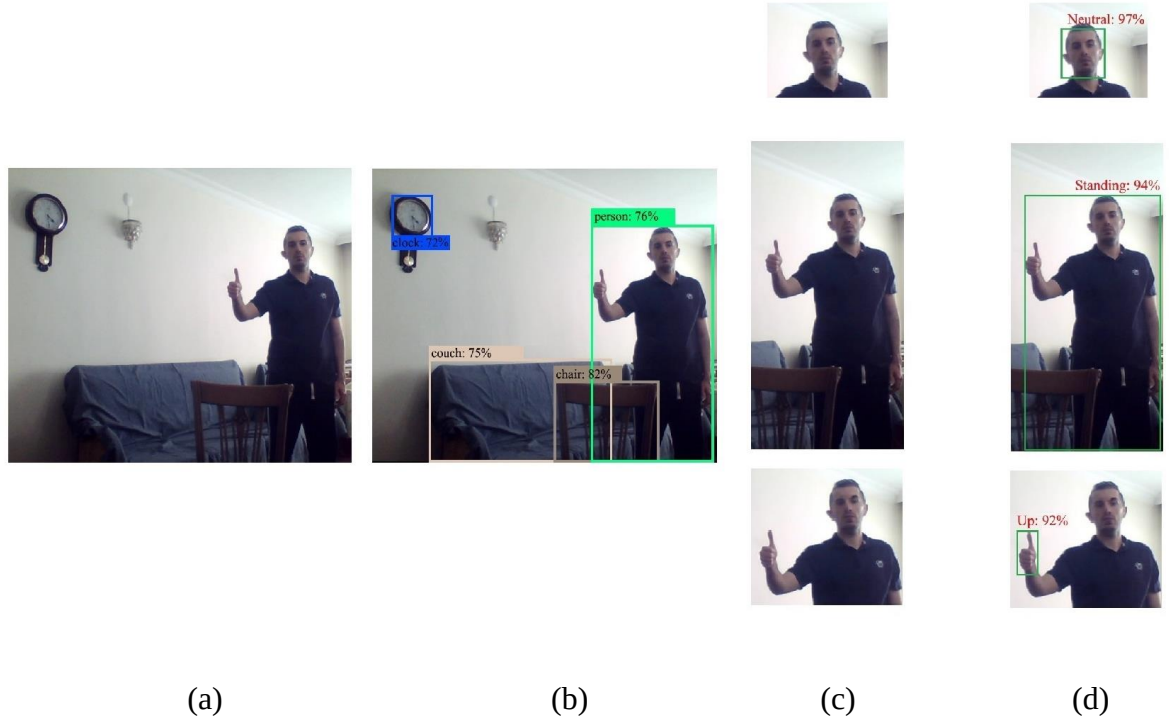
Şekil 3.42 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 7



Şekil 3.43 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 8



Şekil 3.44 Eğitilmiş modelimiz ile elde edilen bazı doğru sınıflandırılmış sonuç resim kareleri gösterimi 9



Şekil 3.45 Eğitilmiş modelimiz ile her bir obje tespiti ve analizi sistemi modelinin genel mimarisinin aşamalarının entegre edilmiş gösterimi

(a) Giriş resmi karesi, (b) Obje tespiti ve tanımlama işlemleri sonrası elde edilen görüntü karesi, (c) Video büyütme/yaklaşım işlemi sonrası elde edilen görüntü kareleri, (d) Video analizi işlemleri sonrası eğitilmiş modelimiz ile sınıflandırılmış görüntü kareleri

4. DENEYSEL SONUÇLAR VE KARŞILAŞTIRMA

Bu bölüm içerisinde obje tespiti ve analizi sistemi modelimiz için geliştirilen derin sinir ağı modelimizin değerlendirilmesi işlemleri gerçekleştirilmiştir. Bu kapsamda bu bölüm iki alt başlık altında incelenmiştir. İlk alt başlık olan bölüm içerisinde deneyler sırasında kullanılan verisetleri ve özellikleri anlatılmıştır. İkinci alt başlıkta, geliştirilen derin sinir ağı modelimiz için uygulanan uygulama detaylarından, gerçekleştirilen deneysel analiz karşılaştırmaları sonucu elde edilen deneysel sonuçlardan bahsedilmiş ve ek olarak, çalışılan derin sinir ağı mimarimizin diğer güncel ve geleneksel mimarilere göre performans karşılaştırılması yapılmıştır.

4.1 Kullanılan Veri Setleri ve Özellikleri

Obje tespiti ve analizi sistemi modelimiz içerisinde geliştirilen derin sinir ağı modelimizin değerlendirilmesinde, bahsedilen her bir durum çalışması için çeşitli ve kapsamlı veri setleri kullanılmıştır. İlk olarak, duygu analizi işlemleri durum çalışması için kullanılan verisetleri The Japanese Female Facial Expression (JAFFE) (Lyons vd. 1997), The Extended Cohn-Kanade (CK+) (Lucey vd. 2010) veri setleri ve kişisel fotoğraf koleksiyonundan oluşmaktadır. Burada CK+ veriseti 123 farklı öznenen toplanmış 593 farklı duygu ifade eden görüntü grubundan, JAFFE veri seti 10 farklı öznenen toplanmış 213 farklı duygu ifade eden görüntü grubundan ve kişisel veri setimiz veri artırma değişkenleri dahil olmak üzere 12 farklı öznenen toplanmış 458 farklı duygu ifade eden görüntü grubundan oluşmaktadır. Ek olarak, çalışma içerisinde duygu analizi işlemleri durum çalışması için yedi yüz duygu kategorisinde sınıflandırma gerçekleştirilmiştir. Bunlar öfke (angry), iğrenme (disgust), korku (fear), mutluluk (happy), üzüntü (sad), sürpriz (surprise) ve tarafsızlık (neutral)'tır. Bundan dolayı, çalışmada duygu analizi işlemleri durum çalışması içerisinde kullanılan veri setleri içerisindeki resimler bu yüz duygu kategorilerine göre düzenlenmişlerdir. İlgili durum çalışması için oluşturulan veri setinden her bir yüz duygu sınıfına (kategorisine) ait örnek resimler Şekil 4.1'de ve resimlerin sınıflara göre dağılımı Çizelge 4.1'de gösterilmiştir.



Şekil 4.1 Duygu analizi işlemleri durum çalışması için kullanılan veri setleri içerisindeki her bir yüz duygu kategorisine ait örnek resimler (Lyons vd. 1997, Lucey vd. 2010)

Çizelge 4.1 Duygu analizi işlemleri durum çalışması için kullanılan veri setleri içerisindeki yüz duygu sınıflarına ait örnek resim sayıları

Duygu Sınıfları	Örnek Resim Sayıları	Yüzdelik Oran(%)
Öfke	176	14(%)
İğrenme	189	15(%)
Korku	168	13(%)
Mutluluk	214	17(%)
Üzüntü	164	13(%)
Sürpriz	227	18(%)
Tarafsızlık	126	10(%)
Toplam	1264	100(%)

İkinci olarak, hareket ve davranış analizi işlemleri durum çalışmasında kullanılan veri setleri; insan eylemlerini tanımlama işlemleri için video ve sensör tabanlı olarak elde edilmiş veriler içeren halka açık dört veri setinden ve el hareketi tanımlama işlemleri için statik el hareketi görüntüleri içeren halka açık iki verisetinden oluşmaktadır. İlgili durum çalışması için her bir verisetine ait örnek resimler Şekil 4.2 ve Şekil 4.3'de gösterilirken, el hareketi tanımlama işlemleri için kullanılan verisetlerinin karakteristik özellikleri Çizelge 4.2'de özetlenmiştir. Ek olarak, bu altı veri setinin detayları aşağıda ayrı ayrı açıklanmıştır:

- **Skoda veriseti** (Zappi vd. 2008), montaj hattı çalışanlarının bir otomobil bakım senaryosundaki faaliyetlerini açıklar. Veri seti, fabrikada kalite kontrol noktalarından birinde 20 ivmeölçer giyen bir işçiyi 10 etkinlik gerçekleştirirken kaydeder. Faaliyetler arasında "her iki sol kapıyı kapat (close both left doors)", "gövde boşluklarını kontrol et (check trunk gaps)", "bagajı aç ve kapat (open and close trunk)", "motor kaputunu aç (open engine hood)" ve bunun gibi aktiviteler bulunur. Burada verilerin yakalanma hızı için kullanılan örnekleme sıklığı 96 Hertz'dir.
- **UCI Smartphone (Human Activity Recognition with Smartphone) veriseti** (Anguita vd. 2012), 30 öznenen toplanan ivmeölçer ve jiroskop sensörlerinden elde edilen verilerden oluşmaktadır. Burada her kişi bir akıllı telefon ile donatılmıştır ve altı etkinliği (yürümek (walking), yukarı yürümek (walking upstairs), aşağı yürümek (walking downstairs), oturmak (sitting), ayakta durmak ve uzanmak (laying)) gerçekleştirmeleri takip edilmiştir. Her kişiye takılan akıllı telefonlar içerisinde gömülü olarak bulunan ivmeölçer ve jiroskop sensörleri ile 3 eksenli doğrusal hızlanma ve 3 eksenli açısal hız içeren veriler 50 Hz sabit bir hızda yakalanmıştır. Ek olarak, veriler kaydedilen videolar kullanılarak manuel olarak burada etiketlenmiştir.
- **KTH veriseti** (Rodriguez 2010), 6 tür insan eyleminden (yürümek (walking), koşma (running), tempolu yavaş koşma (jogging), el sallama (hand waving), boks

(boxing) ve el ırpma (hand clapping)) oluřan kamuya aık verisetidir. Bu veritabanı, drt farklı senaryoda 25 zne tarafından birkaç kez gerekleřtirilerek elde edilmiřtir. İlgili senaryolar dıř mekan (outdoor), lekli farklılıklar ieren dıř mekan (outdoor with variation in scale), farklı giysilerle dıř mekan (outdoor with different clothes) ve aydınlatma farklılıkları olan dıř mekan'dır (outdoor with illumination variations). Burada veriler sabit bir kamera ve deėiřmeyen arkaplan ile 25 Hz kare hızında elde edilmiřtir.

- **UCF spor eylemi veriseti (UCF sports action dataset)** (Schuldt vd. 2004), ESPN ve BBC gibi televizyon kanallarında yayınlanan videolardan toplanan 10 tr spor eylemini iermektedir. Bu eylemler dalıř (diving), golf vuruřu (golf swing), tekme atma (kicking), kaldırma (lifting), at binme (riding horse), kořma (running), kaykayla kayma (skateboarding), salıncak bankı (swing-bench), salıncak tarafı (swing-side) ve yrmek (walking) 'dir. Veri seti, 720 x 480 znrlėe sahip toplam 150 diziye ierir. Burada veriler, 10 Hz hızında ve arka plandaki, aydınlatma kořullarındaki ve oklzyonlardaki farklılıkları gsterecek řekilde gerek spor ortamında kaydedilerek elde edilmiřtir.
- **LaRED veriseti** (Hsiao vd. 2014), yaklaşık 240.000 demet (tuple) grnt (renkli grnt, derinlik grnts ve el blgesinin maskesi) ieren byk bir RGB-D el hareketi veri setidir. Bu veri seti, 27 el hareketinin 3 farklı rotasyonda elde edilmesiyle oluřturulmuř 81 farklı sınıfla literatrde var olan en byk el hareketi veri setlerinden biridir. Veri seti kısa menzilli bir Intel derinlik kamerası kullanılarak toplanmıřtır ve burada 27 el hareket sınıfının (el hareketlerinin sınıflandırma isimleri: $G_1, G_2, G_3, \dots, G_{27}$ řeklindedir) byk bir kısmı amerikan iřaret dilinden, geri kalan kısmı elektronik cihazlarla etkileřim iin kullanılan bazı yaygın hareketlerden oluřturulmuřtur. Sınıflar 10 denek (beř erkek ve beř kadın) tarafından kaydedilmiřtir ve burada her denekten aynı el hareketini hafif hareketlerle tekrarlayarak sınıf bařına 300 hareket grnts elde edilmiřtir. alıřmamız ierisinde bu veri setinin sadece VGA znrlėnn renkli grntleri kullanılmıřtır ve veri setinin geri kalanı alıřmamıza dahil edilmemiřtir.

- **TinyHands veri setinin** (Bao vd. 2017), iki ayrı kurulum ortamında oluşturulması sağlanmıştır. Veri setinin yarısı basit bir arka planla (simple background), geri kalanı ise karmaşık bir arka planla (complex background) kaydedilerek elde edilmiştir. Burada karmaşık arka plan, oldukça karmaşık bir ortamla (cluttered environment) çeşitli aydınlatma koşullarına maruz bırakılmıştır. Veri setinde, hareketler görüntüdeki farklı konumlarda gerçekleştirilir ve eller tüm görüntünün yalnızca piksel olarak yaklaşık % 10'u kadar küçük bir bölgesini kaplar. Bu veri kümesini toplanması için 40 denek kullanılmıştır ve her deneğin yedi farklı el hareketi (fist (yumruk el işareti), L ('L' el işareti), ok (tamam el işareti), palm (avuç içi el işareti), pointer (işaretçi el işareti), down (aşağı el işareti), up (yukarı el işareti)) yapması gerçekleştirilerek her bir hareket sınıfından yaklaşık 1400 hareket görüntüsü elde edilmiştir.



(a) El sallama



(b) Salıncak tarafı



(c) Boks



(d) Salıncak bankı



(e) Golf vuruşu



(f) Motor kaputunu aç

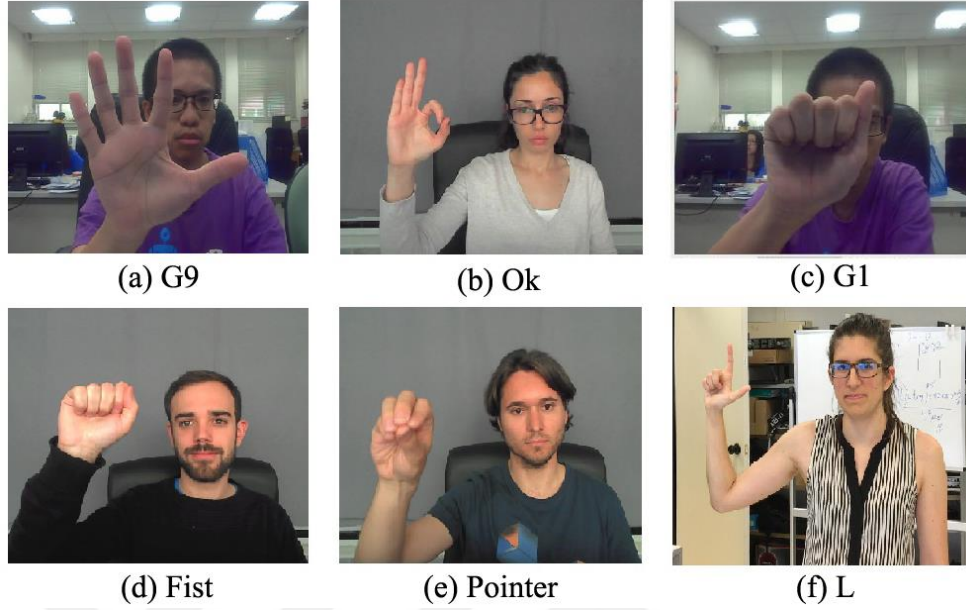


(g) Ata binme



(h) Yavaş Koşma

Şekil 4.2 Hareket ve davranış analizi işlemleri durum çalışması içerisinde insan eylemlerini tanımlama işlemleri için kullanılan veri setlerine ait örnek resimler (Zappi vd. 2008, Anguita vd. 2012, Rodriguez 2010, Schuldt vd. 2004)



Şekil 4.3 Hareket ve davranış analizi işlemleri durum çalışması içerisinde el hareketi tanımlama işlemleri için kullanılan veri setlerine ait örnek resimler (Hsiao vd. 2014, Bao vd. 2017)

(a) "G9" el işareti (b) "Ok" el işareti (c) "G1" el işareti (d) "Fist" el işareti (e) "Pointer" el işareti (f) "L" el işareti

Çizelge 4.2 Hareket ve davranış analizi işlemleri durum çalışması içerisinde el hareketi tanımlama işlemleri için kullanılan verisetlerinin karakteristik özellikleri

Veri Setleri	Özne/Denek Sayıları	El Hareketi Sınıfı Sayıları	Eğitim Seti Sayıları	Test Seti Sayıları
LaRED	10	81	~218,700	~24,300
TinyHands	40	7	~294,000	~98,700

4.2 Elde Edilen Sonuçların Grafikselleştirilmesi ve Deneysel Karşılaştırılması

Bu bölümde, obje tespiti ve analizi sistemi modelimiz için geliştirilen derin sinir ağı modelimiz değerlendirilmesi işlemleri için uygulanan uygulama detaylarından ve bahsedilen her bir durum çalışması için gerçekleştirilen deneysel analiz karşılaştırmaları sonucu elde edilen deneysel sonuçlardan bahsedilmiştir. Deneyler, bir Linux işletim

sisteminde 16 Gbayt RAM belleği ile 3.5 GHz'de çalışan bir Intel Core i7 makinesi ile gerçekleştirilmiştir. Önerilen mimariyi gerçekleştirmek için python programlama dili ve kütüphaneleri (Dlib, Numpy, OpenCV, Tensorflow, Scikit-learn ve bunun gibi) kullanılmıştır. Burada eğitim ve test işlemleri sırasıyla uygulanmıştır. Eğitim ve test işlemleri sırasında kullanılan her bir verisetine ait detaylar ve örnek resimler, kullanılan veri setleri ve özellikleri alt bölümünde (bölüm 4.1) verilmiştir. Mevcut verilerin %70'i eğitim, %20'si test, %10'u onaylama verisi olarak seçilmiş ve kullanılmıştır. Ağ mimarisinin eğitim süreci daha önce tanımlanmış bir bilgisayarda, GPU desteği olmadan, duygu analizi işlemleri durum çalışması ve hareket davranış analizi işlemleri durum çalışması için sırasıyla yaklaşık 40 ve 90 saat (insan eylemlerini tanımlama işlemi için 60 saat, el hareketi tanımlama işlemleri için 30 saat) boyunca gerçekleştirilmiştir. Burada eğitimler duygu analizi işlemleri durum çalışması için 200 iterasyon ve hareket davranış analizi işlemleri durum çalışması için 400 iterasyonda (insan eylemlerini tanımlama işlemi için 300 iterasyon, el hareketi tanımlama işlemleri için 100 iterasyon) durdurulmuştur. Değerlendirme işlemi için basit ama etkili değerlendirme metrikleri kullanılmıştır. Bu değerlendirme metrikleri doğruluk (accuracy), hassasiyet (sensitivity), özgüllük (specificity) ve f puanı (f-score) parametreleridir. Bu parametrelerin anlamlarından kısaca bahsedilirse; ilk olarak doğruluk değeri, değerlendirilen set ile doğru sonuçlar arasındaki oranın ölçülmesi ile modellerin doğruluk oranını vermektedir ve doğru sınıflandırılmış örnek sayısının, toplam örnek sayısına oranı ile hesaplanmaktadır. Sensitivity ve specificity değeri sırasıyla bir modelin pozitif ve negatif örnekler üzerindeki etkinliğini vermektedir. F-score parametresi, test edilen verilerin doğruluğunun ölçümünü göstermektedir ve ek olarak F-score, test edilen verilerin hassasiyet (precision) ve hatırlanma (recall) değerlerinin ağırlıklı harmonik ortalaması olarak tanımlanmaktadır. Çalışma içerisinde bu metrik değerleri aşağıda gösterilen (4), (5), (6) ve (7) numaralı denklemler kullanılarak hesaplanmıştır. Burada TP doğru pozitif (true positive), FP yanlış pozitif (false positive), TN doğru negatif (true negative) ve FN yanlış negatif (false negative) göstermektedir. Bahsedilen değerlendirme metrik değerleri, önerilen modelin performansını değerlendirmenin ilk adımıdır ve burada seçilen dört derin sinir ağı ile önerilen hibrit modelin karşılaştırılması ile karşılaştırma işlemleri gerçekleştirilmiştir.

$$\text{Doğruluk} = \frac{TP+TN}{TP+FP+FN+TN} \quad (4.1)$$

$$\text{Hassasiyet} = \frac{TP}{TP+FN} \quad (4.2)$$

$$\text{Özgüllük} = \frac{TN}{TN+FP} \quad (4.3)$$

$$\text{F puanı} = \frac{(2*TP)}{(2*TP+FP+FN)} \quad (4.4)$$

Çizelge 4.3, Çizelge 4.4 ve Çizelge 4.5'de her deneysel veri kümesi için tanımlanan konfigürasyon parametreleri gösterilirken; Şekil 4.4, Şekil 4.5 ve Şekil 4.6'de sırasıyla, önerilen ağ ve diğer AlexNet, VGG-16, GoogleNet ve ResNet152 derin sinir ağı modellerinin ayrı ayrı duygu analizi işlemleri durum çalışması (JAFfE, CK+, ve Personal veri setleri) ve hareket ve davranış analizi işlemleri durum çalışması içerisinde insan eylemi tanımlama (KTH, UCF Sports Action, Skoda ve UCI Smartphone veri setleri) ve el hareketini tanımlama işlemleri (LaRED ve TinyHands veri setleri) için kullanılan verisetleri üzerindeki elde edilen doğruluk (accuracy), hassasiyet (sensitivity), özgüllük (specificity) ve f puanı (f-score) metrik değerleri gösterilmiştir. Şekil 4.4, Şekil 4.5 ve Şekil 4.6'da gösterilen grafiklere göre, önerilen yöntemin çoğunlukla bahsedilen çağdaş mimarileri geride bıraktığı belirlenmiştir. Ek olarak, burada diğer dört derin sinir ağının performansları, bahsedilen durum çalışmaları için kullanılan toplam dokuz veri setine göre büyük farklılıklar gösterirken; çalışılan ağımızın performansı bu dokuz veri seti için benzer (yakın) performans sonuçları göstermiştir. Bahsedilen bu durumlar, çalışılan ağımızın diğer dört derin sinir ağından daha sağlam ve üstün performanslı olduğunu göstermektedir. Deneysel çalışmalarımız içerisinde ikinci olarak çalışılan ağ modelimizin başarımını daha iyi anlamak ve değerlendirmek için bahsedilen metriklerin yanı sıra karışıklık matrisleri ile birlikte her bir durum çalışması için çeşitli kategoriler incelenmiştir. Bu kategoriler duygu analizi işlemleri durum çalışması için duygu kategorileri, hareket ve davranış analizi işlemleri durum çalışması için hareket/eylem ve el hareketi kategorileri'dir. Önerilen ağ ve diğer derin sinir ağı modelleri için Şekil 4.7, Şekil 4.8 ve Şekil 4.9'de sırasıyla özel kişisel

veri seti, KTH ve TinyHand veri setleri üzerinde elde edilen karışıklık matrisleri gösterilmektedir. Burada Şekil 4.7'de kişisel veri seti üzerinde sırasıyla yukarıdan aşağıya olmak üzere gerçekleştirilen VGG-16, GoogleNet, ResNet152 ve önerilen ağ modellerinin çalışılan yedi yüz duygu kategorisi (öfke, iğrenme, korku, mutluluk, üzüntü, sürpriz ve tarafsızlık) için karışıklık matrisleri gösterilirken; Şekil 4.8 ve Şekil 4.9'de sırasıyla yukarıdan aşağıya olmak üzere AlexNet, VGG-16, GoogleNet, ResNet152 ve önerilen ağ modellerinin altı hareket (boks, el çırpma, el sallama, yürüme, tempolu yavaş koşma, koşma) ve yedi el hareketi kategorisi (fist (yumruk el işareti), 1 (bir el işareti), ok (tamam el işareti), palm (avuç içi el işareti), pointer (işaretçi el işareti), down (aşağı el işareti), up (yukarı el işareti)) için KTH ve TinyHands veri üzerinde oluşturulan karışıklık matrisleri sunulmaktadır. Karışıklık matrisleri ile birlikte her bir eylemin, el hareketinin ve yüz duygusunun doğruluk (accuracy) oranları sunulmuştur ve buna ek olarak x eksenini boyunca hedef sınıflarının ve y eksenini boyunca çıkış sınıflarının arasındaki uyum gösterilmiştir. Burada Şekil 4.7.e, Şekil 4.8. ve Şekil 4.9.e'de karışıklık matrisleri gösterilen önerilen yöntemimizin üzüntü, koşma, pointer dışındaki bütün duygu, hareket ve el hareketi sınıflandırma işlemlerinde diğer dört ağ modelinden daha iyi sonuç verdiği gözlemlenmiştir. Ek olarak, beş ağ modeli de duygu analizi işlemleri durum çalışması için mutluluk, şaşkınlık, tarafsızlık duygularını ve hareket ve davranış analizi işlemleri durum çalışması içerisinde insan eylemi tanımlama işlemleri için yürüme, el sallama hareketlerini ve el hareketi tanımlama işlemleri için down ve up el hareketlerini kolayca ayırt edebilmektedir. Şekil 4.7.d, Şekil 4.8.d ve Şekil 4.9.d'de gösterilen Resnet-152 ağ modeli üzüntü duygusunda, koşma hareketinde ve pointer el hareketinde diğer ağ modellerine kıyasla daha iyi sonuç vermiştir. Çalışmamız içerisinde önerilen modelimizin performansını değerlendirmek için üçüncü olarak yayınlarından doğrudan alıntı yapılan state-of-the-art (son teknoloji/ yeni fikir algoritmalar) sonuçlar ile karşılaştırma işlemi gerçekleştirilmiştir. Çizelge 4.7'de önerilen ağ modeli ve diğer state-of-the-art algoritmaları için sırasıyla KTH, UCF Sports Action, Skoda ve UCI Smartphone veri setleri üzerinde elde edilen doğruluk değerleri gösterilirken; Çizelge 4.6 ve Çizelge 4.8 'de sırasıyla önerilen ağ modeli ve diğer state-of-the-art algoritmaları için CK+ ve LaRED veri seti üzerinde elde edilen doğruluk değerleri sunulmuştur. Burada önerilen ağımızın

performansının daha yüksek doğruluk değerleri ile diğer state-of-the-art algoritmalarından çoğunlukla daha başarılı olduğu gözlemlenmiştir. Çalışmamız içerisinde değerlendirme işlemleri için son olarak, öznelik optimizasyonu süreci içerisinde kullanılan genetik algoritmanın etkinliğini kanıtlamak adına t-testi değerlendirmesi işlemi gerçekleştirilmiştir. Çok yaygın olarak kullanılan istatistiksel yöntemlerden biri olan t-testi, iki veri kümesi arasındaki farkın rastgele veya istatistiksel olarak anlamlı olup olmadığını değerlendirmek için kullanılan istatistiksel bir testtir. Çizelge 4.9'da, altı kapsamlı veriseti (JAFPE, KTH, UCF Sports, Skoda, LaRED ve TinyHands verisetleri) üzerinde önerilen hibrit modelimizin genetik algoritma kullanılarak ve kullanmadan elde edilen doğruluk sonuçları için t-testi sonuçları gösterilmektedir. Burada t Stat değerlerinin, hem tek uçlu (one-tailed) hem de iki uçlu (two-tailed) t kritik (t Critical) değerleri için iki doğruluk grubu arasındaki farkların istatistiksel olarak anlamlı olduğunu düşündürmektedir. Sıfıra çok yakın t olasılık değerleri ($P(T \leq t)$) değerleri, $P(T \leq t) \ll 0.05$ olduğu için değerlendirmenin güveninin %99'dan daha yüksek olduğunu gösterir. Bu nedenle, genetik algoritma kullanılarak geliştirilen önerilen yaklaşımımız dokuz veri seti üzerinde de yüz duygusu, insan eylemi ve el hareketi tanımlama işlemleri için elde edilen doğruluk değerlerinde önemli ölçüde artış olmasını sağlayarak ilgili işlemler için doğruluk değerlerinde performans artışı sağlamıştır.

Çizelge 4.3 JAFPE, CK+ ve Kişisel veri setleri için hiperparametre yapılandırmaları gösterimleri

Parametreler	JAFPE Veri seti	CK+ Veri seti	Kişisel Veri seti
Döngü (Epoch)	35	90	75
Regularizasyon	0.001	0.001	0.001
Veri Boyutu (Batch Size)	32	64	64
Öğrenme Oranı	0.01	0.01	0.01
0-500 iteration	0.005	0.005	0.005
500-1000 iteration	0.002	0.002	0.002
1000+ iteration			
Momentum	0.9	0.9	0.9

Çizelge 4.3 JAFFE, CK+ ve Kişisel veri setleri için hiperparametre yapılandırmaları gösterimleri (devam)

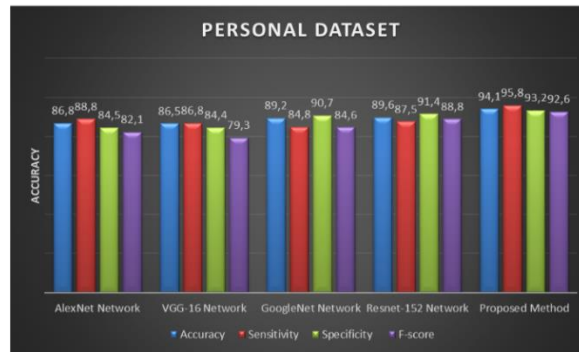
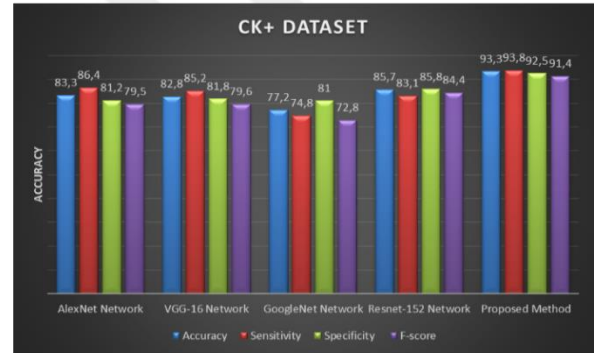
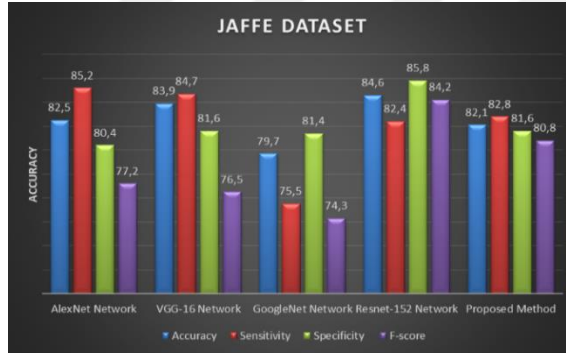
Seyreltme (Dropout)	0.5	0.5	0.5	0.5
Ağırlık İklendirme (Weight Initialization)	Xavier Initialization	Xavier Initialization	Xavier Initialization	Xavier Initialization
Optimizasyon Algoritması	SGD	SGD	SGD	SGD
Hata Fonksiyonu (Loss Function)	Categorical Cross Entropy	Categorical Cross Entropy	Categorical Cross Entropy	Categorical Cross Entropy

Çizelge 4.4 KTH, UCF Sports,Skoda ve UCI Smartphone verisetleri için hiperparametre yapılandırmaları gösterimleri

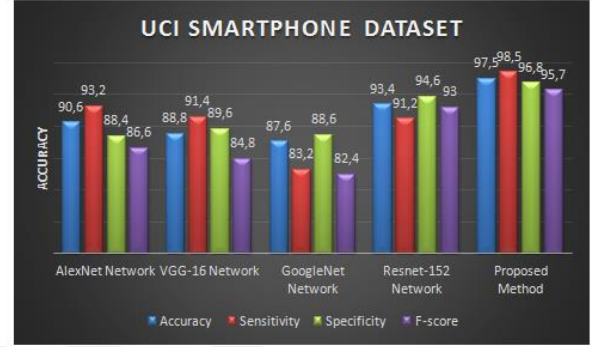
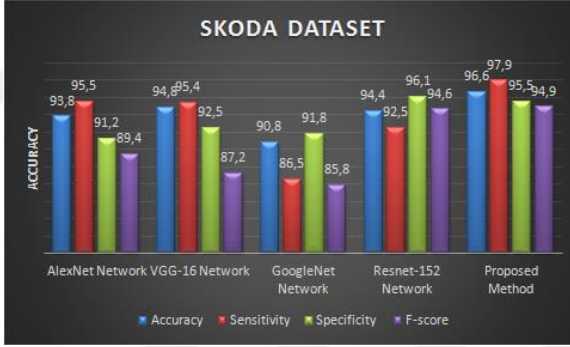
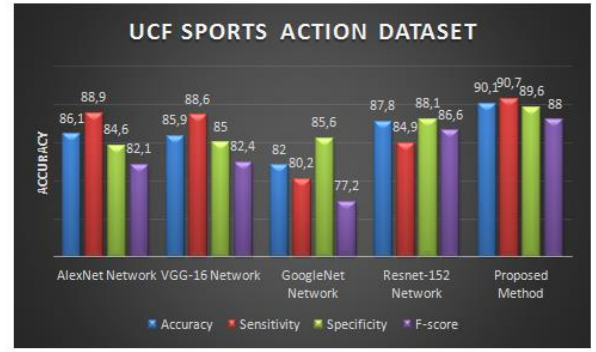
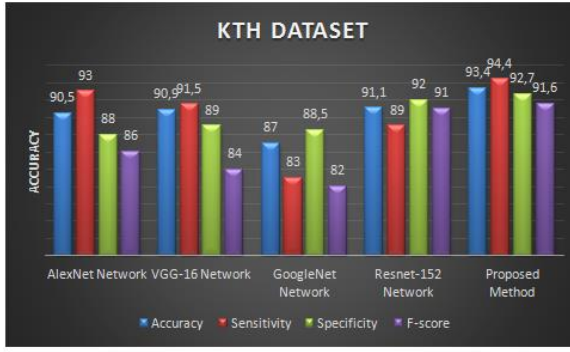
Parametreler	Skoda	UCI Smartphone	KTH	UCF Sports Action
Döngü (Epoch)	70	60	90	80
Regularizasyon	0.001	0.001	0.001	0.001
Veri Boyutu (Batch Size)	64	32	64	64
Öğrenme Oranı				
0-500 iteration	0.01	0.01	0.01	0.01
500-1000 iteration	0.005	0.005	0.005	0.005
1000+ iteration	0.002	0.002	0.002	0.002
Momentum	0.9	0.9	0.9	0.9
Seyreltme (Dropout)	0.5	0.5	0.5	0.5
Ağırlık İklendirme (Weight Initialization)	Xavier Initialization	Xavier Initialization	Xavier Initialization	Xavier Initialization
Optimizasyon Algoritması	SGD	SGD	SGD	SGD
Hata Fonksiyonu (Loss Function)	Categorical Cross Entropy	Categorical Cross Entropy	Categorical Cross Entropy	Categorical Cross Entropy

Çizelge 4.5 LaRED ve TinyHands verisetleri için hiperparametre yapılandırmaları gösterimleri

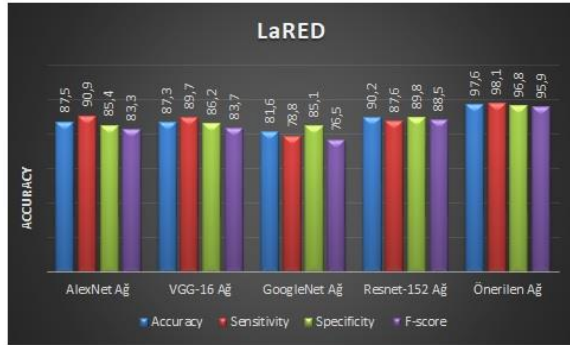
Parametreler	Skoda	UCI Smartphone
Döngü (Epoch)	90	80
Regularizasyon	0.001	0.001
Veri Boyutu (Batch Size)	64	64
Öğrenme Oranı	0.01	0.01
0-500 iteration	0.005	0.005
500-1000 iteration	0.002	0.002
1000+ iteration		
Momentum	0.9	0.9
Dropout	0.5	0.5
Seyreltme (Dropout)	Xavier Initialization	Xavier Initialization
Ağırlık İklendirme (Weight Initialization)	SGD	SGD
Optimizasyon Algoritması	Categorical Cross Entropy	Categorical Cross Entropy



Şekil 4.4 AlexNet, VGG-16, GoogleNet, ResNet152 ve önerilen derin sinir ağı modellerinin JAFFE, CK+ ve Kişisel veri setleri üzerindeki elde edilen doğruluk, hassasiyet, özgüllük, f puanı metrik değerleri gösterimleri



Şekil 4.5 AlexNet, VGG-16, GoogleNet, ResNet152 ve önerilen derin sinir ağı modellerinin KTH, UCF Sports Action, Skoda ve UCI Smartphone veri setleri üzerindeki elde edilen accuracy doğruluk, hassasiyet, özgüllük, f puanı metrik değerleri gösterimleri



Şekil 4.6 AlexNet, VGG-16, GoogleNet, ResNet152 ve önerilen derin sinir ağı modellerinin LaRED ve TinyHands veri setleri üzerindeki elde edilen doğruluk, hassasiyet, özgüllük, f puanı metrik değerleri gösterimleri

	Öfke	İğrenme	Korku	Muthuluk	Üzüntü	Sürpriz	Tarafsızlık
Öfke	82,8	5,5	2,5	2,3	3,4	2,0	1,5
İğrenme	6,9	76,8	3,9	4,2	4,1	1,6	2,5
Korku	3,8	1,9	82,8	3,1	6,2	1,1	1,1
Muthuluk	0,0	0,0	2,1	94,4	2,1	0,1	1,3
Üzüntü	3,5	2,2	2,9	1,9	82,2	1,8	5,5
Sürpriz	0,0	0,4	0,9	0,0	1,9	92,1	4,7
Tarafsızlık	0,1	0,1	1,3	1,1	0,5	0,4	96,5

(a)

	Öfke	İğrenme	Korku	Muthuluk	Üzüntü	Sürpriz	Tarafsızlık
Öfke	82,3	5,8	2,6	1,9	3,8	1,9	1,7
İğrenme	7,2	76,0	4,4	3,8	4,4	1,5	2,7
Korku	3,9	2,1	82,2	2,7	6,8	1,5	0,8
Muthuluk	0,0	0,0	2,4	94,1	1,9	0,0	1,6
Üzüntü	3,9	1,9	2,7	2,1	81,9	1,7	5,8
Sürpriz	0,0	0,3	1,2	0,0	2,1	92,0	4,4
Tarafsızlık	0,1	0,0	1,5	0,9	0,6	0,2	96,7

(b)

	Öfke	İğrenme	Korku	Muthuluk	Üzüntü	Sürpriz	Tarafsızlık
Öfke	86,8	4,4	1,8	1,7	2,6	0,8	1,9
İğrenme	5,8	81,7	4,3	2,6	1,8	0,9	2,9
Korku	5,7	2,2	84,6	1,1	3,3	1,2	1,9
Muthuluk	0,0	0,0	1,6	96,3	1,7	0,0	0,4
Üzüntü	3,3	0,9	2,9	2,2	84,9	1,5	4,3
Sürpriz	0,0	0,0	0,7	0,0	1,8	94,0	3,5
Tarafsızlık	0,0	0,0	1,7	0,3	1,6	0,6	95,8

(c)

	Öfke	İğrenme	Korku	Muthuluk	Üzüntü	Sürpriz	Tarafsızlık
Öfke	88,5	4,1	1,1	1,2	3,1	0,9	1,1
İğrenme	5,5	80,8	4,9	3,7	2,2	1,1	1,8
Korku	4,9	0,7	83,2	1,8	5,7	1,1	2,6
Muthuluk	0,0	0,0	2,1	95,2	2,7	0,0	0,0
Üzüntü	2,5	0,4	2,6	1,4	87,9	0,5	4,7
Sürpriz	0,0	0,0	0,4	0,0	2,3	94,4	2,9
Tarafsızlık	0,2	0,0	1,3	0,1	1,2	0,1	97,1

(d)

	Öfke	İğrenme	Korku	Muthuluk	Üzüntü	Sürpriz	Tarafsızlık
Öfke	94,4	3,4	0,4	0,6	0,8	0,1	0,3
İğrenme	2,9	89,8	0,2	0,9	3,9	0,9	1,4
Korku	1,1	1,3	88,7	3,9	2,5	1,5	1,0
Muthuluk	0,0	0,0	0,0	98,9	1,1	0,0	0,0
Üzüntü	3,1	0,1	1,7	0,3	87,6	1,3	5,9
Sürpriz	0,0	0,0	0,4	0,0	0,0	99,4	0,2
Tarafsızlık	0,1	0,0	0,0	0,0	0,0	0,1	99,8

(e)

Şekil 4.7 AlexNet, VGG-16, GoogleNet, ResNet152 ve önerilen ağ modellerinin kişisel veri seti üzerindeki çalışılan yedi yüz duygu kategorisi için karışıklık matrisleri gösterimleri

(a) AlexNet Ağı, (b) VGG-16 Ağı, (c) GoogleNet Ağı, (d) ResNet152 Ağı, (e) Önerilen Ağ

	Boks	El Çırpma.	El Sallama	Yavaş Koş.	Koşma	Yürüme
Boks	87,8	5,8	2,6	1,9	3,8	1,9
El Çırpma.	7,2	85,7	4,4	3,8	4,4	1,5
El Sallama	3,9	2,1	93,3	2,7	6,8	1,5
Yavaş Koş.	0,0	0,0	2,4	90,6	1,9	0,0
Koşma	3,9	1,9	2,7	2,1	92,8	1,7
Yürüme	0,0	0,3	1,2	0,0	2,1	92,5

(a)

	Boks	El Çırpma.	El Sallama	Yavaş Koş.	Koşma	Yürüme
Boks	84,9	4,4	1,8	1,7	2,6	0,8
El Çırpma.	5,8	85,2	4,3	2,6	1,8	0,9
El Sallama	5,7	2,2	89,8	1,1	3,3	1,2
Yavaş Koş.	0,0	0,0	1,6	88,3	1,7	0,0
Koşma	3,3	0,9	2,9	2,2	85,1	1,5
Yürüme	0,0	0,0	0,7	0,0	1,8	88,5

(b)

	Boks	El Çırpma.	El Sallama	Yavaş Koş.	Koşma	Yürüme
Boks	88,2	5,8	2,6	1,9	3,8	1,9
El Çırpma.	7,2	85,9	4,4	3,8	4,4	1,5
El Sallama	3,9	2,1	93,6	2,7	6,8	1,5
Yavaş Koş.	0,0	0,0	2,4	91,6	1,9	0,0
Koşma	3,9	1,9	2,7	2,1	93,0	1,7
Yürüme	0,0	0,3	1,2	0,0	2,1	92,8

(c)

	Boks	El Çırpma.	El Sallama	Yavaş Koş.	Koşma	Yürüme
Boks	88,4	4,1	1,1	1,2	3,1	0,9
El Çırpma.	5,5	86,3	4,9	3,7	2,2	1,1
El Sallama	4,9	0,7	93,7	1,8	5,7	1,1
Yavaş Koş.	0,0	0,0	2,1	91,9	2,7	0,0
Koşma	2,5	0,4	2,6	1,4	93,4	0,5
Yürüme	0,0	0,0	0,4	0,0	2,3	93,1

(d)

	Boks	El Çırpma.	El Sallama	Yavaş Koş.	Koşma	Yürüme
Boks	92,7	3,2	0,7	0,4	1,4	0,1
El Çırpma.	3,2	91,8	0,1	1,2	3,7	1,0
El Sallama	1,7	1,9	94,8	3,8	2,9	1,3
Yavaş Koş.	0,0	0,0	0,0	92,4	1,4	0,0
Koşma	2,6	0,1	1,9	0,2	93,2	1,0
Yürüme	0,0	0,0	0,8	0,0	0,0	95,3

(e)

Şekil 4.8 AlexNet, VGG-16, GoogleNet, ResNet152 ve önerilen ağ modellerinin KTH veri seti üzerindeki altı hareket kategorisi için karışıklık matrisleri gösterimleri

(a) AlexNet Ağı, (b) VGG-16 Ağı, (c) GoogleNet Ağı, (d) ResNet152 Ağı, (e) Önerilen Ağ

	Fist	L	Ok	Palm	Pointer	Down	Up
Fist	87,1	1,3	1,7	2,9	3,7	2,1	1,2
L	0,4	86,5	0,7	2,8	4,7	3,5	1,4
Ok	1,1	0,8	87,3	2,7	5,6	2,2	0,3
Palm	0,0	0,5	1,1	94,7	2,3	0,1	1,3
Pointer	1,2	0,0	0,2	0,6	97,2	0,2	0,6
Down	0,5	0,4	0,9	1,7	0,8	95,5	0,2
Up	0,3	0,2	0,6	0,1	0,4	0,5	97,9

(a)

	Fist	L	Ok	Palm	Pointer	Down	Up
Fist	86,6	1,9	1,8	2,6	3,5	2,4	1,2
L	1,2	86,3	2,4	3,2	2,6	1,4	2,9
Ok	0,8	3,5	85,4	1,4	4,2	3,2	1,5
Palm	0,3	0,1	1,1	94,5	1,6	0,2	2,2
Pointer	0,5	0,9	0,1	0,3	97,0	0,7	0,5
Down	1,3	0,3	0,7	0,1	2,1	95,2	0,3
Up	0,2	0,5	0,1	0,7	0,5	0,3	97,7

(b)

	Fist	L	Ok	Palm	Pointer	Down	Up
Fist	89,1	1,2	1,5	2,7	1,9	0,7	2,9
L	2,6	88,9	3,8	1,9	1,6	0,9	0,3
Ok	1,5	2,9	90,8	0,3	2,4	1,7	0,4
Palm	0,0	0,2	2,2	95,1	1,5	0,4	0,6
Pointer	0,7	0,5	0,0	0,0	98,5	0,0	0,3
Down	0,0	0,1	0,8	0,0	1,5	97,4	0,2
Up	0,1	0,0	0,5	0,3	0,0	0,2	98,9

(c)

	Fist	L	Ok	Palm	Pointer	Down	Up
Fist	89,5	1,6	0,4	1,2	2,1	3,1	2,1
L	2,7	89,4	1,3	0,8	1,9	2,4	1,5
Ok	2,5	0,3	91,2	1,4	2,6	1,7	0,3
Palm	0,0	0,0	2,1	95,2	2,7	0,0	0,0
Pointer	0,2	0,0	0,1	0,2	99,3	0,2	0,0
Down	0,4	0,2	0,7	0,2	0,0	97,7	0,8
Up	0,2	0,2	0,1	0,2	0,0	0,1	99,2

(d)

	Fist	L	Ok	Palm	Pointer	Down	Up
Fist	99,8	0,0	0,0	0,1	0,0	0,1	0,0
L	0,0	98,1	0,3	0,1	0,7	0,5	0,3
Ok	0,0	0,0	99,9	0,0	0,0	0,0	0,1
Palm	0,2	0,1	0,0	98,5	0,2	0,9	0,1
Pointer	0,4	0,1	0,0	0,3	98,8	0,1	0,3
Down	0,1	0,0	0,1	0,1	0,0	99,6	0,1
Up	0,0	0,2	0,0	0,0	0,0	0,1	99,7

(e)

Şekil 4.9 AlexNet, VGG-16, GoogleNet, ResNet152 ve önerilen ağ modellerinin TinyHands veri seti üzerindeki altı hareket kategorisi için karışıklık matrisleri gösterimleri

(a) AlexNet Ağı, (b) VGG-16 Ağı, (c) GoogleNet Ağı, (d) ResNet152 Ağı, (e) Önerilen Ağ

Çizelge 4.6 Önerilen ağ ve diğer state-of-the-art algoritmalarının sonuçları için CK + veri seti üzerinde elde edilen yüz duygusu sınıflandırma doğruluk değerleri tablosu

Model/Yöntem	Doğruluk(%)
3DCNN (Liu vd. 2014a)	85,9
ITBN (Wang vd. 2013a)	86,3
STM (Liu vd. 2014b)	91,13
LOMo (Sikka vd. 2016)	92,0
AUDN (Liu vd. 2013)	92,05
3DCNN-DAP (Liu vd. 2014a)	92,4
Önerilen Ağ	93,4

Çizelge 4.7 Önerilen ağ ve diğer state-of-the-art algoritmalarının sonuçları için sırasıyla (a) KTH, (b) UCF Sports Action, (c) Skoda, (d) UCI Smartphone veri setleri üzerindeki hareket/eylem sınıflandırma doğruluk değerleri tabloları

Model/Yöntem	Doğruluk(%)
Ahad vd. (2016)	86,7
Chaaaraoui vd. (2013)	89,86
Ji vd. (2013)	90,2
Charalampous and Gasteratos (2016)	91,99
Önerilen Ağ	93,4

(a)

Model/Yöntem	Doğruluk(%)
Atmosukarto vd. (2015)	82,6
Wang vd. (2009)	85,6

Çizelge 4.7 Önerilen ağ ve diğer state-of-the-art algoritmalarının sonuçları için sırasıyla (a) KTH, (b) UCF Sports Action, (c) Skoda, (d) UCI Smartphone veri setleri üzerinde elde edilen insan eylemlerini tanıma doğruluk değerleri tabloları (devam)

Le vd. (2011)	86,5
Kovashka vd. (2010)	87.27
Wang vd. (2013b)	88.0
Wang vd. (2011)	88.2
Önerilen Ağ	90,1

(b)

Model/Yöntem	Doğruluk(%)
Zeng vd. (2014)	86,10
Alsheikh vd. (2016)	89,30
Ordóñez and Roggen (2016)	95,80
Önerilen Ağ	96,6

(c)

Model/Yöntem	Doğruluk(%)
Ronao and Cho (2015a)	90,0
Ronao and Cho (2015b)	94,79
Ronao and Cho (2016)	94,61
Jiang and Yin (2015)	95,18
Önerilen Ağ	97,5

(d)

Çizelge 4.8 Önerilen ağ ve diğer state-of-the-art algoritmalarının sonuçları için LaRED veri seti üzerinde elde edilen el hareketi tanıma doğruluk değerleri tablosu

Model/Yöntem	Doğruluk(%)
Sanchez-Riera vd. (2016) - Deep belief networks	66,13
Sanchez-Riera vd. (2016) - Restricted boltzman machines	72,95
Hsiao vd. (2014) - Baseline network	74,55
Sanchez-Riera vd. (2016) - Stacked autoencoders	81,09
Sanchez-Riera vd. (2016) -Convolutional Neural Network	88,72
Mohammed et al. (2019) - Lightweight MobileNet	97,25
Önerilen Ağ	97,6

Çizelge 4.9 Önerilen modelimizin genetik algoritma kullanılarak ve kullanmadan elde edilen doğruluk sonuçları için t-testi sonuçları tablosu

LaRED		TinyHand	
t Stat	20,85391	t Stat	30,09557
P(T<=t) tek-uçlu	3,05E-12	P(T<=t) tek-uçlu	1,99E-14
t Kritik tek-uçlu	1,76131	t Kritik tek-uçlu	1,76131
P(T<=t) iki-uçlu	6,1E-12	P(T<=t) iki-uçlu	3,99E-14
t Kritik iki-uçlu	2,144787	t Kritik iki-uçlu	2,144787
KTH		UCF SPORTS ACTION	
t Stat	20,9965	t Stat	29,3854
P(T<=t) tek-uçlu	2,78E-12	P(T<=t) tek-uçlu	2,78E-14
t Kritik tek-uçlu	1,7613	t Kritik tek-uçlu	1,7613
P(T<=t) iki-uçlu	5,56E-12	P(T<=t) iki-uçlu	5,55E-14
t Kritik iki-uçlu	2,14479	t Kritik iki-uçlu	2,14479
JAFPE		SKODA	
t Stat	8,964933	t Stat	20,9371
P(T<=t) tek-uçlu	1,77E-07	P(T<=t) tek-uçlu	2,89E-12
t Kritik tek-uçlu	1,7613	t Kritik tek-uçlu	1,7613
P(T<=t) iki-uçlu	3,55E-07	P(T<=t) iki-uçlu	5,78E-12
t Kritik iki-uçlu	2,14479	t Kritik iki-uçlu	2,14479

5. SONUÇ

Tez kapsamında obje tespiti, video büyütme yaklaşımları ve derin öğrenme yöntemleri bir araya getirilerek yeni bir obje tespiti ve analizi sistemi geliştirilmiştir. Mevcut obje tespiti ve analizi çalışmaları genellikle bu üç yaklaşım veya yöntemin tek bir tanesinin veya çiftinin bir araya getirilerek geliştirilmesi ile gerçekleştirilmişlerdir. Bu gerçekleştirilen yaklaşımlar veya yöntemler sınırlı uygulama alanları üzerinde kısıtlı başarı oranlarına ulaşabilmişlerdir. Bundan dolayı çalışmada daha kapsamlı bir uygulama alanı üzerinde daha yüksek başarı oranlarına ulaşabilmek adına ilgili üç yöntem bir araya getirilmiştir.

Bu çalışmadaki amaç video çerçeveleri arasındaki farkların analizine dayalı olarak objelerin tespitini, izlenmesini sağlamak ve sonrasında obje hareketlerini izleyen video içerisinde analiz edilecek bölgeye video büyütme yöntemi ile yaklaşım işlemi gerçekleştirilerek o bölge üzerinde geliştirilen derin öğrenme yöntemi mimarisi ile video analiz işlemleri gerçekleştirmektir. Bu kapsamda çalışma içerisinde üretilen modelde, hibrit bir modele dayalı yeni bir derin öğrenme temelli mimari geliştirilmiştir. Mimari temel olarak transfer öğrenme metodolojisi ve genetik algoritmanın kullanılması ile önceden eğitilmiş ağlara dayalı olarak tasarlanmıştır.

Geliştirilen obje tespiti ve analizi sistemi modelinin mimarisi üç ana aşamadan meydana gelmektedir. Bu üç aşama kısaca sırasıyla anlatılırsa; birinci aşama içerisinde ilk olarak giriş videosu üzerinde arka plan hesaplanmış ve devamında ikinci olarak muhtemel obje pozisyonlarının bulunması işlemleri yapılmıştır. Muhtemel obje pozisyonlarının bulunması işlemleri içerisinde sırasıyla; hareket eden nesneleri algılamak adına her bir video karesinden arka plan çıkarım işlemi, değer kaybını en aza indirmek ve ileride daha optimize sonuçlar alınabilmesi için ölçeklendirilme işlemi ve segmentasyon işlemleri gerçekleştirilmiştir. Üçüncü olarak, elde edilen sonuçlar üzerinde görüntü düzeltme işlemleri yapılmıştır. Bu işlemler içerisinde ilk olarak önceki adımda oluşan fazlalıkların ve gürültülerin giderilmesi için gaussian filtresi ve eşikleme (thresholding) yöntemi uygulanmış ve devamında obje görüntüleri üzerindeki boşlukların doldurulması içinde

morfolojik işlemler gerçekleştirilmiştir. Sonuç olarak objelerin videodaki pozisyonları tespit edilmiş ve devamında tespit edilen konumlar sınırlayıcı kutu içerisine alınmıştır. Bu işlem sonrasında son adım olarak derin öğrenme yöntemleri ile birlikte uygun hazır veri setleri kullanılarak algılanan konumlar üzerinde obje tanımlama işlemleri gerçekleştirilerek birinci aşamamız tamamlanmıştır. Bir sonraki aşama olan ikinci aşama içerisinde objelerin tespit edildiği video üzerinde analiz edilecek bölgeye video büyütme yöntemleri ile yaklaşım işlemi yapılmıştır. Devamında son aşamada, yaklaşılan bölge üzerinde geliştirilen derin öğrenme yöntemi mimarisi ile kapsamlı veri setleri kullanılarak video analiz işlemleri gerçekleştirilmiştir.

Önerilen yöntemin ana katkıları, tanınmış önceden eğitilmiş dört ağ modelini optimize edilmiş bir şekilde birleştiren bir hibrit model önermesi ve öznitelik optimizasyonu işlemi için yeni bir sezgisel algoritma modellemesi geliştirilmesidir. Önerilen mimarinin performansı, kapsamlı ve popüler veri setlerinin faydanılması ile değerlendirilmiştir. Burada önerilen hibrit modelin performansı diğer bireysel modellerle karşılaştırıldığında, önerilen model tüm senaryolar için tüm popüler bireysel modellerden neredeyse daha iyi performans göstermiştir. Bu deneylerin yanı sıra, çeşitli kapsamlı akademik makalelerde yayınlanan son teknoloji çalışmalarla da karşılaştırma işlemleri yapılmıştır. Burada çalışılan ağımızın; sırasıyla CK+, KTH, UCF Sports Action, Skoda, UCI Smartphone ve LaRED veri setleri üzerinde elde edilen %93,3, %93,4, %90,1, %96,6, %97,5 ve %97,6 doğruluk oranları ile seçilen diğer dört derin sinir ağına ve diğer state-of-the-art algoritmalarına göre daha başarılı ve sağlam olduğu tespit edilmiştir. Elde edilen doğruluk oranları Çizelge 5.1'de gösterilmiştir. Sonuç olarak, elde edilen sonuçlar önerilen yöntemin, en gelişmiş yöntemlere göre üstünlüğünü ortaya koymakta ve onaylamaktadır.

Önerilen hibrit yaklaşımımızın yazarları bilgisayarla görme, insan-bilgisayar etkileşimi ve bunun gibi alanlar içerisinde daha karmaşık görevlerin üstesinden gelmek için önceden eğitilmiş modellerden hibrit modeller oluşturmaya teşvik edeceğine inanıyoruz. Gelecekteki çalışmalarda, önerilen obje tespiti ve analizi sistemi modelinin kapsamı, bahsedilen durum çalışmalarının dışında farklı durum çalışmaları yapılarak genişletilebilir.

Burada farklı araştırma ve uygulama alanlarında elde edebilecek verisetleri kullanılması ile çalışılan sistem eğitilebilir ve performans durumu gözlemlenerek sistemin geliştirilmesi için çeşitli katkılar gerçekleştirilebilir. Ek olarak, parçacık sürü optimizasyonu (particle swarm optimization), karın kolonisi Optimizasyonu (ant colony optimization), diferansiyel gelişim algoritması (differential evaluation algorithm) ve bunun gibi farklı sezgisel algoritmalar önerilen yöntemimiz içerisinde kullanılarak daha başarılı sonuçlar alınabilir. Son olarak, çalışılan sistem farklı deneysel kurumlar ile gerçekleştirilerek daha az veri ile daha kısa sürede daha verimli sonuçlar alınması sağlanabilir. Bu deneysel kurulumlara, grafik işlemci ünitelerinin (GPU) kullanılması örnek verilebilir. Burada GPU'nun çok çekirdekli yapısından faydalananarak daha hızlı bir şekilde eğitim ve optimizasyon işlemleri gerçekleştirilebilir ve bu sayede çalışma performansı önemli derecede artırılarak daha etkili bir şekilde işlemler gerçekleştirilebilir.

KAYNAKLAR

- Aggarwal, J. K. ve Ryoo, M. S. 2011. Human activity analysis: A review. *ACM Computing Surveys*, 43(3), 1–43. <https://doi.org/10.1145/1922649.1922653>
- Ahad, M. R., Islam, M.N. ve Jahan, I. 2016. Action recognition based on binary patterns of action-history and histogram of oriented gradient, *Journal on Multimodal User Interfaces*, 10, 335–344. <https://doi.org/10.1007/s12193-016-0229-4>
- Alexe, B., Deselaers, T. ve Ferrari, V. 2012. Measuring the objectness of image windows, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34, 11, 2189-2202. <https://doi.org/10.1109/TPAMI.2012.28>
- Alsheikh, M.A., Selim, A., Niyato, D., Doyle, L., Lin, S. ve Tan, H.-P. Deep activity recognition models with triaxial accelerometers, *The Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence Artificial Intelligence Applied to Assistive Technologies and Smart Environments*, 8-13. arXiv:1511.04664
- Andriluka, M., Roth, S. ve Schiele, B. 2008. People-tracking-by-detection and people detection-by-tracking, *Computer Vision and Pattern Recognition*, 1-8. <https://doi.org/10.1109/CVPR.2008.4587583>
- Andriyenko, A., Schindler, K. ve Roth, S. 2012. Discrete-continuous optimization for multitarget tracking. 2012 *IEEE Conference on Computer Vision and Pattern Recognition*, 1926-1933. <https://doi.org/10.1109/CVPR.2012.6247893>
- Anguita, D., Ghio, A., Oneto, L., Parra, X. ve Reyes-Ortiz, L.J. 2012. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine, *International Workshop on Ambient Assisted Living*, Springer, 216–223. https://doi.org/10.1007/978-3-642-35395-6_30
- Anonymous. 2015. Web Sitesi: [http://www.vision.caltech.edu/Image_Datasets/Caltech Pedestrians/](http://www.vision.caltech.edu/Image_Datasets/Caltech_Pedestrians/), Eriřim Tarihi: 01.07.2020.
- Anonymous. 2016. Web Sitesi: <http://www.imagenet.org/>, Eriřim Tarihi: 01.07.2020.
- Anonymous. 2019a. Web Sitesi: https://en.wikipedia.org/wiki/object_detection, Eriřim Tarihi: 01.07.2020.
- Anonymous. 2019b. Web Sitesi: https://en.wikipedia.org/wiki/image_scaling#bilinear_and_bicubic_algorithms, Eriřim Tarihi: 01.07.2020.
- Anonymous. 2019c. Web Sitesi: <https://medium.com/deep-learning-turkiye/yolo-algoritmasını-anlamak-290f2152808f>, Eriřim Tarihi: 01.07.2020.
- Anonymous. 2019d. Web Sitesi: <http://people.csail.mit.edu/nwadhwa/phase-video/>, Eriřim Tarihi: 01.07.2020.
- Anonymous. 2019e. Web Sitesi: <https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data>, Eriřim Tarihi: 01.07.2020.

- Anonymous. 2020a. Web Sitesi: https://en.wikipedia.org/wiki/deep_learning, Erişim Tarihi: 01.07.2020.
- Anonymous. 2020b. Web Sitesi: <http://cocodataset.org/>, Erişim Tarihi: 01.07.2020.
- Anonymous. 2020c. Web Sitesi: <http://www.cvlibs.net/datasets/kitti/>, Erişim Tarihi: 01.07.2020.
- Arbelaez, P., Pont-Tuset, J., Barron, J., Marques, F. ve Malik, J. 2014. Multiscale combinatorial grouping, IEEE Conference on Computer Vision and Pattern Recognition, 328-335. <https://doi.org/10.1109/CVPR.2014.49>
- Atmosukarto, I., Ahuja, N. ve Ghanem, B. 2015. Action recognition using discriminative structured trajectory groups, IEEE Winter Conference on Applications of Computer Vision, 899-906. <https://doi.org/10.1109/WACV.2015.124>
- Azizpour, H., Razavian, S.A., Sullivan, J., Maki, A. ve Carlsson, S. 2015. From generic to specific deep representations for visual recognition, IEEE Conference on Computer Vision and Pattern Recognition Workshops, 36-45. <https://doi.org/10.1109/CVPRW.2015.7301270>
- Bao, P., Maqueda, A.I., Del-Blanco, C.R. ve Garcíá, N. 2017. Tiny hand gesture recognition without localization via a deep convolutional network, IEEE Transactions on Consumer Electronics, 63, 251–257. <https://doi.org/10.1109/TCE.2017.014971>
- Barinova, O., Lempitsky, V. ve Kohli, P. 2012. On detection of multiple object instances using hough transform, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1773-1784. <https://doi.org/10.1109/TPAMI.2012.79>
- Bartneck, C., Lyons, M.J. 2007. HCI and the face: Towards an art of the soluble, Proceedings of the 12th International Conference on Human-Computer Interaction: Interaction Design and Usability, 20–29. https://doi.org/10.1007/978-3-540-73105-4_3
- Belongie, S., Malik, J. ve Puzicha, J. Matching shapes. 2001. The 8th IEEE International Conference on Computer Vision, 454–461.
- Benezeth, Y., Jodoin, P.-M., Emile, B., Laurent, H. ve Rosenberger, C. 2010. Comparative study of background subtraction algorithms, Journal of Electronic Imaging. <https://doi.org/10.1117/1.3456695>
- Bengio, Y., Courville, A. ve Vincent, P. 2013. Representation Learning: A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35 (8), 1798–1828. arXiv:1206.5538. <https://doi.org/10.1109/tpami.2013.50>
- Bose, S. R. ve Kumar, V. S. 2019. Hand gesture recognition using faster R-CNN inception v2 model, Advances in Robotics, 1–6. <https://doi.org/10.1145/3352593.3352613>

- Boztoprak, H., Çağlar, M.F. ve Merdan, M. 2007. Alternatif Morfolojik Bir Yöntemle Plaka Yerini Saptama, XII. Elektrik, Elektronik, Bilgisayar, Biyomedikal Mühendisliği Ulusal Kongresi.
- Breitenstein, M.D., Reichlin, F., Leibe, B., Koller-Meier, E. ve Gool, L. V. 2009. Robust tracking-by-detection using a detector confidence particle filter, IEEE 12th International Conference on Computer Vision, 1515-1522. <https://doi.org/10.1109/ICCV.2009.5459278>
- Burt, P. ve Adelson, E. 1983. The laplacian pyramid as a compact image code, IEEE Transactions on Communications, 31, 532–540.
- Cannon, R.L., Dave, J.V. ve Bezdek, J.C. 1986. Efficient implementation of the fuzzy cmeans clustering algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2, 248–255. <https://doi.org/10.1109/TPAMI.1986.4767778>
- Carreira, J. ve Sminchisescu, C. 2012. CPMC: Au-tomatic object segmentation using constrained parametric min-cuts, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34, 1312–1328. <https://doi.org/10.1109/TPAMI.2011.231>
- Cao, H., Nguyen, N.M., Phua, C., Krishnaswamy, S. ve Li, X. 2012. An integrated framework for human activity classification, ACM International Joint Conference on Pervasive and Ubiquitous Computing. <https://doi.org/10.1145/2370216.2370268>
- Chaaraoui, A.A., Climent-Pérez, P. Ve Flórez-Revuelta, F. 2013. Silhouette-based human action recognition using sequences of key poses, Pattern Recognition Letters, 34(15), 1799-1807. <https://doi.org/10.1016/j.patrec.2013.01.021>
- Charalampous, K. ve Gasteratos, A. 2016. On-line deep learning method for action recognition, Pattern Analysis and Applications, 19(2), 337-354. <https://doi.org/10.1007/s10044-014-0404-8>
- Chatfield, K., Simonyan, K., Vedaldi, A. ve Zisserman, A. 2014. Return of the devil in the details: Delving deep into convolutional nets. arXiv:1405.3531
- Chen, W., Er, M. J. ve Wu, S. 2006. Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 36(2), 458–46. <https://doi.org/10.1109/TSMCB.2005.857353>
- Chen, L., Wang, F., Deng, H. ve Ji, K. 2013. A survey on hand gesture recognition, Int. Conference on Computer Sciences and Applications, 313-316. <https://doi.org/10.1109/CSA.2013.79>
- Chen, C.H., Lee, I.J. ve Lin, L.Y. 2015. Augmented reality-based self facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders, Research in Developmental Disabilities, 36, 396–403. <https://doi.org/10.1016/j.ridd.2014.10.015>

- Cheng, Y. Mean shift, mode seeking, and clustering. 1995. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(8), 790-799. <https://doi.org/10.1109/34.400568>
- Cheng, G., Wan, Y., Saudagar A.N., Namuduri, K. ve Buckles, B. P. 2015. Advances in Human Action Recognition. arXiv:1501.05964
- Chevtchenko, S.F., Vale, R.F., Macari, V. ve Cordeiro, F.R. 2018. A convolutional neural network with feature fusion for real-time hand posture recognition, Applied Soft Computing, 73, 748–766. <https://doi.org/10.1016/j.asoc.2018.09.010>
- Chi, Z., Yan, H. ve Pham, T. 1996. Fuzzy algorithms: with application to image processing and pattern recognition, Advances in Fuzzy Systems Applications and Theory. <https://doi.org/10.1142/3132>
- Cilla, R., Patricio, M.A., Garcia, J., Berlanga, A. ve Molina, J. M. 2009. Recognizing human activities from sensors using hidden markov models constructed by feature selection techniques, Algorithms, 2(1), 282–300. <https://doi.org/10.3390/a2010282>
- Cireşan, D., Giusti, A., Gambardella, L. ve Schmidhuber, J. 2013. Mitosis detection in breast cancer histology images with deep neural networks, Medical Image Computing and Computer-Assisted Intervention, 411-418. https://doi.org/10.1007/978-3-642-40763-5_51
- Cortes, C. ve Vapnik, V. 1995. Support-vector networks. Machine Learning, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>.
- Criminisi, A., Perez, P. ve Toyama, K. 2004. Region filling and object removal by exemplar-based image inpainting, IEEE Transactions on Image Processing, 13, 1200-1212. <https://doi.org/10.1109/TIP.2004.833105>
- Dalal, N., Triggs, B. 2005. Histograms of Oriented Gradients for Human Detection, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1, 886-893. <https://doi.org/10.1109/CVPR.2005.177>
- Dardas, N.H. ve Georganas, N.D. 2011. Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques, IEEE Transactions on Instrumentation and Measurement, 60, 3592–3607. <https://doi.org/10.1109/TIM.2011.2161140>
- Dollár, P., Wojek, C., Schiele, B. ve Perona, P. 2009. Pedestrian Detection: A Benchmark.”, 2009 IEEE Conference on Computer Vision and Pattern Recognition, 304-311. <https://doi.org/10.1109/CVPR.2009.5206631>
- Dempster, A., Laird, N. ve Rubin, D. 1977 .Maximum likelihood estimation from incomplete data via the EM algorithm, Journal of the Royal Statistical Society, 39(7), 1-38.
- Dai, J., Li, Y, He, K.ve Sun, J. 2016. R-FCN: Object detection via region-based fully convolutional networks, Proceedings of the 30th International Conference on Neural Information Processing Systems, 379–387. arXiv:1605.06409

- Dornaika, F. ve Raducanu, B. 2007. Efficient facial expression recognition for human robot interaction, International Work-Conference on Artificial Neural Networks, 700–708. https://doi.org/10.1007/978-3-540-73007-1_84
- Efros, A. A. ve Leung, T. K. 1999. Texture synthesis by nonparametric sampling, IEEE International Conference on Computer Vision, 1033–1038. <https://doi.org/10.1109/ICCV.1999.790383>
- Elgammal, A., Harwood, D., and Davis, L. 2000. Non-parametric model for background subtraction, European Conference on Computer Vision, 751–767. https://doi.org/10.1007/3540-45053-X_48
- Endres, I. ve Hoiem, D. 2010. Category independent object proposals, European Conference on Computer Vision, 575–588. https://doi.org/10.1007/978-3-642-15555-0_42
- Essa, I.A., Pentland, A. P. 1997. Coding, analysis, interpretation and recognition of facial expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), 757–763. <https://doi.org/10.1109/34.598232>
- Everingham, M., Gool, L. V., Williams, C. K. I., Winn, J. ve Zisserman, A. 2010. The PASCAL Visual Object Classes (VOC) Challenge. International Journal of Computer Vision, 88, 303–338. <https://doi.org/10.1007/s11263-009-0275-4>
- Fatima, I., Fahim, M., Lee, Y.-K ve Lee, S. 2013. A genetic algorithm-based classifier ensemble optimization for activity recognition in smart homes. KSII Transactions on Internet and Information Systems, 7, 2853–2873. <https://doi.org/10.3837/tiis.2013.11.018>
- Felzenszwalb P. F., Girshick, R. B., McAllester, D. ve Ramanan, D. 2010. Object detection with discriminatively trained part-based models, IEEE Transactions on Pattern Analysis and Machine Intelligence, 32, 9, 1627–1645. <https://doi.org/10.1109/TPAMI.2009.167>
- Freeman, W.T., Adelson, E.H. 1991. The design and use of steerable filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(9), 891–906. <https://doi.org/10.1109/34.93808>
- Freeman, W. T. ve Roth, M. 1995. Orientation histograms for hand gesture recognition, Intl. Workshop on Automatic Face and Gesture- Recognition, IEEE Computer Society, 296–301.
- Freeman, W. T., Tanaka, K., Ohta, J. ve Kyuma, K. Computer vision for computer games, 2nd International Conference on Automatic Face and Gesture Recognition, Killington, 100–105.
- Freund, Y. ve Schapire, R. E. 1991. A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and System Sciences, 55, 119–139. <https://doi.org/10.1006/jcss.1997.1504>

- Gao, H., Yüce, A. ve Thiran, J. 2014. Detecting emotional stress from facial expressions for driving safety, 2014 IEEE International Conference on Image Processing, 5961-5965. <https://doi.org/10.1109/ICIP.2014.7026203>
- Gavrila, D. M. ve Munder, S. 2007. Multi-cue pedestrian detection and tracking from a moving vehicle, International Journal of Computer Vision, 73, 41–59. <https://doi.org/10.1007/s11263-006-9038-7>
- Girshick, R., Donahue, J., Darrell, T. ve Malik, J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation, IEEE Conference on Computer Vision and Pattern Recognition, Computer Vision and Pattern Recognition (CVPR). <https://doi.org/10.1109/CVPR.2014.81>
- Goldberg, D. E. ve Holland, J. H. 1988. Genetic algorithms and machine learning, Machine learning, 3, 95-99. <https://doi.org/10.1023/A:1022602019183>
- Gonzalez, R.C. ve Woods, R.E. 1992. Digital Image Processing, Prentice Hall.
- Graner, H., Leistner, C. ve Bischof H. 2008. Semi-supervised on-line boosting for robust tracking, European Conference on Computer Vision, 234-247. https://doi.org/10.1007/978-3-540-88682-2_19
- Guzel, M. S. 2016. Performance evaluation for feature exctractors on street view images. The Imaging Science Journal, 64(1), 26-33. <https://doi.org/10.1080/13682199.2015.1109783>
- Gülcü, A. ve Kuş, Z. 2019. Konvolüsyonel Sinir Ağlarında Hiper-Parametre Optimizasyonu Yöntemlerinin İncelenmesi, Gazi University Science Journal: PART:C ‘Design and Technology’, 7(2), 503-522, <https://doi.org/10.29109/gujsc.514483>.
- Handels, H., Ross, T., Kreuschb, J., Wolffb, H.H. ve Pöpl, S.J. 1999. Feature selection for optimized skin tumor recognition using genetic algorithms, Artificial Intelligence in Medicine, 283–297. [https://doi.org/10.1016/s0933-3657\(99\)00005-6](https://doi.org/10.1016/s0933-3657(99)00005-6)
- Haritaoglu, I., Harwood, D. ve Davis, L.S. 2000. W/sup 4/: real-time surveillance of people and their activities, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), 809-830.
- Harris, C., Stephens, M. 1988. A combined corner and edge detector, Proceedings of 4th Alvey Vision Conference, 147-151.
- He, K., Zhang, X., Ren, S., ve Sun, J. 2016. Deep residual learning for image recognition, IEEE Conference on Computer Vision and Pattern Recognition, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- He, K., Gkioxari, G., Dollár P. ve Girshick, R. 2017. Mask R-CNN, IEEE International Conference on Computer Vision, 2980-2988. <https://doi.org/10.1109/ICCV.2017.322>
- Hickson, S., Dufour, N., Sud, A., Kwatra, V. ve Essa, I.A. 2017. Eyemotion: Classifying facial expressions in VR using eye-tracking cameras, 2019 IEEE Winter Conference

- on Applications of Computer Vision, 1626-1635.
<https://doi.org/10.1109/WACV.2019.00178>
- Holland, J.H. 1992a. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control and artificial intelligence, MIT press.
- Holland, J. H. 1992b. Genetic algorithms, Scientific American, 267(1), 66-73.
- Horn, B. K. P., Schunck, B. G.1981. Determining optical flow, Artificial Intelligence, 17, 185–203. [https://doi.org/10.1016/0004-3702\(81\)90024-2](https://doi.org/10.1016/0004-3702(81)90024-2)
- Hsu, C.-W., Chang, C.-C. ve Lin C.-J. 2003. A practical guide to support vector classification.
- Hsiao, Y.S., Sanchez-Riera, J., Lim, T., Hua , K.L. ve Cheng, W.H. 2014. LaRED: A large RGB-D extensible hand gesture dataset, ACM Multimedia Systems Conference, 53–58. <http://dx.doi.org/10.1145/2557642.2563669>
- Huang, C., Wu, B. ve Nevatia, R. 2008. Robust object tracking by hierarchical association of detection responses, 10th European Conference on Computer Vision, 788-801. https://doi.org/10.1007/978-3-540-88688-4_58
- Huang, Y., Liu, X., Zhang , X. ve Jin, L. 2016. A pointing gesture based egocentric interaction system: Dataset, Approach and Application, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 370–377. <https://doi.org/> <https://doi.org/10.1109/CVPRW.2016.53>
- Jain, A. K. ve Dubes, R. C. 1988. Algorithms for Clustering Data, Prentice-Hall. <https://doi.org/10.2307/1268876>
- Javed, O., Ali, S. ve Shah, M. 2005. Online detection and classification of moving objects using progressively improving detectors, Computer Vision and Pattern Recognition, <https://doi.org/10.1109/CVPR.2005.259>
- Jebari, K. ve Madiafi, M. 2013. Selection methods for genetic algorithms. International Journal of Emerging Sciences, 3(4), 333–344.
- Ji, S., Xu, W., Yang, M. ve Yu, K. 3D convolutional neural networks for human action recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(1), 221- 231. <https://doi.org/10.1109/TPAMI.2012.59>
- Jiang, W. ve Yin, Z. 2015. Human activity recognition using wearable sensors by deep convolutional neural networks, Proceedings of the ACM International Conference on Multimedia, 1307–131. <https://doi.org/10.1145/2733373.2806333>
- Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z. ve Qu, R. 2019. A Survey of Deep Learning Based Object Detection. In IEEE Access, 7, 128837-128868. <https://doi.org/10.1109/ACCESS.2019.2939201>.
- Jones, M. J. ve Poggio, T. 1998. Multidimensional morphable models: A framework for representing and matching object classes. International Journal of Computer Vision, 29(2), 107–131. <https://doi.org/10.1023/A:1008074226832>

- Kang, B., Tan, K.H., Jiang, N., Tai, H.S., Tretter, D. ve Nguyen, T. 2017. Hand segmentation for hand-object interaction from depth map, IEEE Global Conference on Signal and Information Processing, 259–263. <https://doi.org/10.1109/GlobalSIP.2017.8308644>
- Karim, A.M., Güzel, M.S., Tolun, M.R., Kaya, H. ve Çelebi, F. V. 2018. New Generalized Deep Learning Framework Combining Sparse Auto-encoder and Taguchi Method for Novel Data Classification and Processing, Mathematical Problems in Engineering, <https://doi.org/10.1155/2018/3145947>
- Karim, A.M., Güzel, M. S., Tolun, M. R., Kaya, H., ve Çelebi, F. V. 2019. A new framework using deep auto-encoder and energy spectral density for medical waveform data classification and processing, Biocybernetics and Biomedical Engineering, 39(1), 148-159. <https://doi.org/10.1016/j.bbe.2018.11.004>
- Kaş, Ü. ve Tanyıldızı, E. 2017. Euler Renk ve Hareket Büyütme Yöntemlerinin Performans Analizi. Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 17 (2), 506-515. <https://dergipark.org.tr/tr/pub/akufemubid/issue/43399/524719>
- Kaulard, K., Cunningham, D. W., Bühlhoff, H. H. ve Wallraven, C. 2012. The MPI facial expression database—A validated database of emotional and conversational facial expressions, PLoS ONE. <https://doi.org/10.1371/journal.pone.0032321>
- Kaya, A., Keçeli, A.S. ve Can, A.B. 2018. Akciğer nodül özelliklerinin tahmininde çeşitli sınıflama stratejilerinin incelenmesi, Gazi Üniversitesi Mimar Mühendislik Fakültesi Dergisi, 34(2), 709-725. <https://doi.org/10.17341/gazimmfd.416530>
- Kim, K., Chalidabhongse, T.H., Harwood, D. ve Davis, L. 2005. Real-time foreground background segmentation using codebook model, Real-Time Imaging, 11(3), 172–185. <https://doi.org/10.1016/j.rti.2004.12.004>
- Kim, Y. ve Toomajian, B. 2016. Hand Gesture Recognition using Micro-Doppler Signatures with Convolutional Neural Network, IEEE Access, 4, 7125-7130. <https://doi.org/10.1109/ACCESS.2016.2617282>
- Kim, D.H., Baddar, W., Jang, J. ve Ro, Y.M. 2017. Multi-objective based Spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition, IEEE Transactions on Affective Computing, 10(2), 223 - 236. <https://doi.org/10.1109/TAFFC.2017.2695999>
- Ko, B.C. 2018. A Brief Review of Facial Emotion Recognition Based on Visual Information, Sensors, 18(2). <https://doi.org/10.3390/s18020401>
- Konolige, K. 1997. Small vision systems: Hardware and implementation. Proc. of the Intl. Symp. of Robotics Research (ISRR), 111-116. https://doi.org/10.1007/978-1-4471-1580-9_19
- Kovashka, A. ve Grauman, K. Learning a hierarchy of discriminative space-time neighborhood features for human action recognition, Computer Vision and Pattern Recognition, 2046-2053. <https://doi.org/10.1109/CVPR.2010.5539881>

- Krizhevsky, A., Sutskever, I. ve Hinton, G.E. 2012. Imagenet classification with deep convolutional neural networks, Neural Information Processing Systems, 1097–1105. <https://doi.org/10.1145/3065386>
- Kumaran, N., Vadivel, A., Kumar, S.S. 2018. Recognition of human actions using CNN-GWO: A novel modeling of CNN for enhancement of classification performance, Multimedia Tools and Applications, <https://doi.org/23115-23147.10.1007/s11042-017-5591-z>
- Le, Q.V., Zou, W.Y., Yeung, S.Y. ve Ng, A.Y. 2011. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis, Computer Vision and Pattern Recognition, 3361-3368. <https://doi.org/10.1109/CVPR.2011.5995496>
- Le, T.H.N., Zhu, C., Zheng, Y., Luu, K. ve Savvides, M. 2016. Robust hand detection in vehicles, Int. Conference on Pattern Recognition, 573–578. <https://doi.org/10.1109/ICPR.2016.7899695>
- Leibe, B., Seemann, E. ve Schiele, B. 2005. Pedestrian detection in crowded scenes, 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1, 878-885. <https://doi.org/10.1109/CVPR.2005.272>
- Levin, A., Viola, P. A. ve Freund, Y. 2003. Unsupervised improvement of visual detectors using co-training, International Conference on Computer Vision. <https://doi.org/10.1109/ICCV.2003.1238406>
- Li, Y., Zhang, Y., Lu, J., Lim, R. ve Wang, J. 2010. Video Analysis and Trajectory Based Video Annotation System. Asia-Pacific Conference on Wearable Computing Systems, APWCS, 1, 307-310. <https://doi.org/10.1109/APWCS.2010.84>
- Li, S. ve Deng, W. 2018. Deep Facial Expression Recognition: A Survey, IEEE Transactions on Affective Computing, <https://doi.org/10.1109/TAFFC.2020.2981446>
- Liang, Z. ve Thorpe, C. 2000. Stereo- and neural network-based pedestrian detection. Intelligent Transportation System, 1(3), 148-154. <https://doi.org/10.1109/6979.892151>
- Liang, L., Liu, C., Xu, Y., Guo, B. ve Shum, H. Y. 2001. Real-time texture synthesis by patch-based sampling, ACM Transactions on Graphics, 127–150. <https://doi.org/10.1145/501786.501787>
- Liang, C., Song, Y. ve Zhang, Y. 2016. Hand gesture recognition using view projection from point cloud, IEEE International Conference on Image Processing, 4413–4417. <https://doi.org/10.1109/ICIP.2016.7533194>
- Liu, C., Torralba, A., Freeman, W. T., Durand, F. ve Adelson, E. H. 2005. Motion magnification, ACM Transactions on Graphics (SIGGRAPH), 519–526. <https://doi.org/10.1145/1073204.1073223>

- Liu, M., Li, S., Shan, S. ve ChenX. 2013. AU-aware deep networks for facial expression recognition, IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, 1-6. <https://doi.org/10.1109/FG.2013.6553734>.
- Liu, M., Li S.Z., Shan, S., Wang, R. ve Chen, X. 2014a. Deeply Learning Deformable Facial Action Parts Model for Dynamic Expression Analysis, Asian Conference on Computer Vision (ACCV), 143-157. https://doi.org/10.1007/978-3-319-16817-3_10
- Liu, M., Shan, S., Wang, R. ve Chen, X. 2014b. Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition, Computer Vision and Pattern Recognition, 1749-1756. <https://doi.org/10.1109/CVPR.2014.226>
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. ve Berg, A. C. 2016. SSD: Single shot multibox detector, European Conference on Computer Vision. https://doi.org/10.1007/978-3-319-46448-0_2
- Lowe, D.G. 1999. Object recognition from local scale-invariant features. Proceedings of the Seventh IEEE International Conference on Computer Vision, 2, 1150-1157.
- Lowe, D. G. 2004. International Journal of Computer Vision, International Journal of Computer Vision, 60(2), 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- Lucas, B. D. ve Kanade, T. 1981. An iterative image registration technique with an application to stereo vision, Proceedings of the 7th International Joint Conference on Artificial Intelligence, 674–679.
- Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z. ve Matthews, I. 2010. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 94-101. <https://doi.org/10.1109/CVPRW.2010.5543262>
- Lyons, M. J., Kamachi, M., Lyons, J. ve Gyoba, J. 1997. Japanese female facial expressions (JAFPE), Database of digital images. <http://doi.org/10.5281/zenodo.3451524>
- Mandal, F. B. ve Srivastava, S. 2012. Real Time Facial Expression Recognition using a Novel Method, The International Journal of Multimedia & Its Applications.
- Matilainen, M., Sangi, P., Holappa, J. ve Silvén, O. 2016. OUHANDS database for hand detection and pose recognition, Int. Conference on Image Processing Theory, Tools and Applications, 1–5. <https://doi.org/10.1109/IPTA.2016.7821025>
- Mehrabian, A. 1968. Communication without words, Psychology Today 2, 53–56.
- Minaei-Bidgoli, B. ve Punch, W. 2003. Using genetic algorithms for data mining optimization in an educational web-based system, Genetic and Evolutionary Computation Conference, 2252–2263. https://doi.org/10.1007/3-540-45110-2_119

- Mirehi, N., Tahmasbi, M. ve Targhi, A.T. 2019. Hand gesture recognition using topological features, *Multimedia Tools and Applications*, 78, 13361–13386. <https://doi.org/10.1007/s11042-019-7269-1>
- Mitra, S. ve Acharya, T. 2007. Gesture recognition: A survey, *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 311-324. <https://doi.org/10.1109/TSMCC.2007.893280>
- Muhammad, G., Alsulaiman, M., Amin, S. U., Ghoneim, A. ve Alhamid, M. F. 2017. A Facial-Expression Monitoring System for Improved Healthcare in Smart Cities, *IEEE Access*, 5, 10871-10881. doi: <https://doi.org/10.1109/ACCESS.2017.2712788>
- Mohammed, A.A.Q., Lv, J. ve Islam, M.S. 2019. A Deep learning-based end-to-end composite system for hand detection and gesture recognition, *Sensors*, 19(23). <https://doi.org/10.3390/s19235282>
- Mutlu, B., Haciomeroglu, M., Guzel, M. S., Dikmen, M. ve Sever, H. 2014. Silhouette extraction from street view images, *International Journal of Robotics Research*, 11/114. <https://doi.org/10.5772/58740>
- Nair, V. ve Clark, J. J. 2004. An unsupervised, online learning framework for moving object detection, *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *Computer Vision and Pattern Recognition*, <https://doi.org/10.1109/CVPR.2004.1315181>
- Nakjai, P. ve Katanyukul, T. 2019. Hand sign recognition for thai finger spelling: An application of convolution neural network, *Journal of Signal Processing Systems*, 91, 131–146. <https://doi.org/10.1007/s11265-018-1375-6>
- Oh, T., Jaroensri, R., Kim, C., Elgharib, M. A., Durand, F., Freeman W. T., ve W. Matusik. 2018. Learning-based video motion magnification, *Computer Vision – ECCV 2018*. arXiv:1804.02684
- Oliver, N.M., Rosario, B. ve Pentland, A.P. 2000. A bayesian computer vision system for modeling human interactions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 831-843. <https://doi.org/10.1109/34.868684>
- Olkkonen, H. ve Pesola, P. 1996. Gaussian Pyramid Wavelet Transform for Multiresolution Analysis of Images, *Graphical Models and Image Processing*, 58, 394- 398. <https://doi.org/10.1006/gmip.1996.0032>
- Ordóñez, F.J. ve Roggen, D. 2016. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition, *Sensors*, 16(1),i 115. <https://doi.org/10.3390/s16010115>
- Ordóñez, F.J. ve Roggen, D. 2016. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition, *Sensors*, 16(1),i 115. <https://doi.org/10.3390/s16010115>

- Pham, H.-A. ve Nguyen, T.D.T. 2018. An Improved Human Activity Recognition by Using Genetic Algorithm to Optimize Feature Vector, International Conference on Knowledge and Systems Engineering, <https://doi.org/10.1109/KSE.2018.8573335>
- Plouffe, G. ve Cretu, A.M. 2016. Static and dynamic hand gesture recognition in depth data using dynamic time warping, IEEE Transactions on Instrumentation and Measurement, 65(2), 305–316. <https://doi.org/10.1109/TIM.2015.2498560>.
- Pomerleau D.A. 1993. Knowledge-based training of artificial neural networks for autonomous robot driving, In: Robot Learning. Connell, J. ve Mahadevan, S. (eds), 19-43. Norwell.
- Qin, J., Liu, L., Zhang, Z., Wang Y. ve L. Shao. 2016. Compressive Sequential Learning for Action Similarity Labeling, IEEE Transactions on Image Processing, 25(2), 756-769, <https://doi.org/10.1109/TIP.2015.2508600>
- Redmon, J., Divvala, S., Girshick, R. ve Farhadi, A. 2016. You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779-788, <https://doi.org/10.1109/CVPR.2016.91>
- Ren, Z., Yuan, J. ve Zhang, Z. 2011. Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera, ACM Multimedia Conference and Co-Located Workshops, 1093–1096. <https://doi.org/10.1145/2072298.2071946>
- Ren, S., He, K., Girshick, R. ve Sun, J. 2015. Faster R-CNN, "Towards real-time object detection with region proposal networks", IEEE Transactions on Pattern Analysis and Machine Intelligence, 39, 1137-1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- Ren, S., He, K., Girshick, R. ve Sun, J. 2017. Faster R-CNN: Towards real-time object detection with region proposal networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39, 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- Rodriguez, M. 2010. Spatio-temporal maximum average correlation height templates in action recognition and video summarization. Doktora Tezi, University of Central Florida, College of Engineering and Computer Science, Electrical Engineering and Computer Science, CFE0003313, Orlando.
- Ronao, C.A. ve Cho, S.B. 2015a. Evaluation of deep convolutional neural network architectures for human activity recognition with smartphone sensors, Proceedings of the KIISE Korea Computer Congress, 858–860. Corpus ID: 111714508
- Ronao, C.A. ve Cho, S.B. 2015b. Deep convolutional neural networks for human activity recognition with smartphone sensors, International Conference on Neural Information Processing, 46–53. https://doi.org/10.1007/978-3-319-26561-2_6
- Ronao, C.A. ve Cho, S.B. 2016. Human activity recognition with smartphone sensors using deep learning neural networks, Expert System with Applications, 59, 235–244. <https://doi.org/10.1016/j.eswa.2016.04.032>

- Rosenberg, C., Hebert, M. ve Schneiderman, H. 2005. Semi-supervised self-training of object detection models, 2005 Seventh IEEE Workshops on Applications of Computer Vision, 1, 29-36. <https://doi.org/10.1109/ACVMOT.2005.107>
- Roth, P. M., Grabner, H., Skoaj, D. ve Bischof, H. 2005. Online conservative learning for person detection, 2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 223-230. <https://doi.org/10.1109/VSPETS.2005.1570919>
- Roth, P. M., Sternig, S., Grabner, H. ve Bischof, H. 2009. Classifier grids for robust adaptive object detection, 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2727-2734. <https://doi.org/10.1109/CVPR.2009.5206616>
- Rumelhart, D. E., Hinton, G. E. ve Williams, R. J. 1986. Learning internal representations by error propagation, In: Parallel Distribution Processing: Explorations in the Microstructure of Cognition. Rumelhart, D. E. ve McClelland, J. L. (eds), MIT Press, 318–362, Cambridge.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg A.C. ve Li, F.F. 2015. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision, 115, 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- Sadjadi, F. 2005. Comparative Image Fusion Analaysais, Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. <https://doi.org/10.1109/CVPR.2005.436>
- Sanchez-Riera, J., Hua, K.L., Hsiao, Y.S., Lim, T., Hidayati, S.C., Cheng, W.H. 2016. A comparative study of data fusion for RGB-D based visual recognition, Pattern Recognition Letters, 73, 1–6. <https://doi.org/10.1016/j.patrec.2015.12.006>
- Schuldt, C., Laptev, I. ve Caputo, B. 2004. Recognizing human actions: A Local SVM approach, International Conference on Pattern Recognition, 3, 32-36. <https://doi.org/10.1109/ICPR.2004.1334462>
- Sikka, K., Sharma, G. ve Bartlett, M. 2016. LOMo: Latent ordinal model for facial analysis in videos, IEEE Conference on Computer Vision and Pattern Recognition, 5580–5589. <https://doi.org/10.1109/CVPR.2016.602>
- Simonyan, K. ve Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556.
- Siyal, M. Y., Fathy, M. ve Dorry, F. 1997. Neural-vision based approach for real-time road traffic applications, Electronics Letters, 33, 969–970. <https://doi.org/10.1049/el:19970646>
- Stauffer, C. ve Grimson, W.E.L. 1999. Adaptive background mixture models for real time tracking, International Conference on Computer Vision and Pattern Recognition, 2, 246-252. <https://doi.org/10.1109/CVPR.1999.784637>

- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. ve Rabinovich, A. Going deeper with convolutions, IEEE conference on computer vision and pattern recognition, 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>
- Şeker, A., Diri, B. ve Balık, H. 2017. Derin Öğrenme Yöntemleri ve Uygulamaları Hakkında Bir İnceleme. Gazi Mühendislik Bilimleri Dergisi, 3(3), 47-64. <https://dergipark.org.tr/tr/download/article-file/394923>
- Tang, S., Andriluka, M. ve Schiele, B. 2014. Detection and tracking of occluded people, International Journal of Computer Vision, 58–69. <https://doi.org/10.1007/s11263-013-0664-6>
- Tao, W., Leu, M.C. ve Yin, Z. 2018. American sign language alphabet recognition using convolutional neural networks with multiview augmentation and inference fusion, Engineering Applications of Artificial Intelligence, 76, 202–213. <https://doi.org/10.1016/j.engappai.2018.09.006>
- Thamizharasi, A., Jayasudha, Js. 2016. An Illumination Invariant Face Recognition Using 2D Discrete Cosine Transform and Clahe, International Journal of Computer Science and Information Technology, 8, 45-53. [10.5121/ijcsit.2016.8304](https://doi.org/10.5121/ijcsit.2016.8304)
- Uijlings, J. R. R., Sande, K. E. A., Gevers, T. ve Smeulders, A. W. M. 2013. Selective Search for Object Recognition, International Journal of Computer Vision, 104, 154–171. <https://doi.org/10.1007/s11263-013-0620-5>
- Umbaugh, S.E. 1999. Computer Vision and Image Processing, Prentice Hall.
- Viola, P. ve Jones, M. 2001. Rapid object detection using a boosted cascade of simple features, Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1, I-511-I-518. <https://doi.org/10.1109/CVPR.2001.990517>
- Viola, P. A. ve Jones, M. J. 2004. Robust Real-Time Face Detection, International Journal of Computer Vision, 57(2), 137–154. <https://doi.org/10.1109/ICCV.2001.937709>
- Wadhwa, N., Rubinstein, M., Durand, F. ve Freeman, W.T. 2013. Phase based video motion processing. ACM Transactions on Graphics(SIGGRAPH), 32(4), 80:1-80:10. <https://doi.org/10.1145/2461912.2461966>
- Wadhwa, N., Rubinstein, M., Durand, F. ve Freeman, W.T. 2014. Riesz pyramids for fast phase-based video magnification. Conference: 2014 IEEE International Conference on Computational Photography. <https://doi.org/10.1109/ICCPHOT.2014.6831820>
- Walecki, R. ve Rudovic, O. 2017 Deep structured learning for facial expression intensity estimation. Image and Vision Computing, 259, 143–154
- Wang, H., Ullah, M.M., Klaser, A., Laptev, I. ve Schmid, C. 2009. Evaluation of local spatio-temporal features for action recognition, British Machine Vision Conference, BMVA Press. <https://doi.org/10.5244/C.23.124>

- Wang, H., Klaser, A., Schmid, C. ve Liu, C.L. 2011. Action recognition by dense trajectories, *Computer Vision and Pattern Recognition*, 3169–3176. <https://doi.org/10.1109/CVPR.2011.5995407>
- Wang, X., Hua, G. ve Han, T. X. 2012. Detection by detections: Non-parametric detector adaptation for a video, 2012 IEEE Conference on Computer Vision and Pattern Recognition, 350-357. <https://doi.org/10.1109/CVPR.2012.6247695>
- Wang, Z., Wang, S. ve Ji, Q. 2013a. Capturing Complex Spatio-temporal Relations among Facial Muscles for Facial Expression Recognition, *IEEE Conference on Computer Vision and Pattern Recognition*, 3422-3429. <https://doi.org/10.1109/CVPR.2013.439>
- Wang, H., Klaser, A., Schmid, C. ve Liu, C.L. 2013b. Dense trajectories and motion boundary descriptors for action recognition, *International Journal of Computer Vision*, 103(1), 60-79. <https://doi.org/10.1007/s11263-012-0594-8>
- Wang, J., Chen, Y., Hao, S., Peng, Xi. ve Hu, L. 2018. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119, 3-11. <https://doi.org/10.1016/j.patrec.2018.02.010>
- Wohler, C., Aulanf, J. K., Portner, T. ve Franke, U. 1998. A time delay neural network algorithm for real-time pedestrian recognition. *International Conference on Intelligent Vehicles*, Almanya.
- Wren, C., Azarbayejani, A., Darrell, T. ve Pentland, A. 1997. Pfinder: Real-time tracking of the human body, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 780-785. <https://doi.org/10.1109/34.598236>
- Wu, B. ve Nevatia, R. 2007. Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors, *International Journal of Computer Vision*, 247–266. <https://doi.org/10.1007/s11263-006-0027-7>
- Wu, H-Y., Rubinstein, M., Shih E., Guttag, Jo., Durand, F. ve Freeman, W. 2012. Eulerian Video Magnification for Revealing Subtle Changes in the World. *ACM Transactions on Graphics* 31, 4, 1-8. <https://doi.org/10.1145/2185520.2185561>
- Wu, D., Sharma, N. ve Blumenstein, M. 2017. Recent advances in videobased human action recognition using deep learning: A review. *International Joint Conference on Neural Networks (IJCNN)*, 2865-2872. <https://doi.org/10.1109/IJCNN.2017.7966210>
- Xiao, X., Xu ve D., Wan, W. 2016. Overview: Video recognition from handcrafted method to deep learning method. *International Conference on Audio, Language and Image Processing (ICALIP)*, 646-651. <https://doi.org/10.1109/ICALIP.2016.7846652>
- Yang, Y., Shu, G. ve Shah, M. 2013. Semi-supervised Learning of Feature Hierarchies for Object Detection in a Video, 2013 IEEE Conference on Computer Vision and Pattern Recognition, 1650-1657. <https://doi.org/10.1109/CVPR.2013.216>

- Yang, Y., Shu, G. ve Shah, M. 2013. Semi-supervised Learning of Feature Hierarchies for Object Detection in a Video, 2013 IEEE Conference on Computer Vision and Pattern Recognition, 1650-1657. <https://doi.org/10.1109/CVPR.2013.216>.
- Yılmaz, A. A., Guzel, M. S. ve Askerbeyli, İ. 2017. Algılanması güç olan değişimlerin ortaya çıkarılması için resim teleskobu çalışmalarının analiz çalışmaları, Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 19 (57) ,723-732. <https://doi.org/10.21205/deufmd.2017195765>
- Yılmaz, A. A., Guzel, M. S., Askerbeyli, İ. and Bostanci, E. 2018. A Vehicle Detection Approach using Deep Learning Methodologies, II. International Conference on Theoretical and Applied Computer Science and Engineering, 64-71. arXiv:1804.00429
- Yılmaz, A. A., Guzel, M. S., Askerzade, İ. and Bostanci, E. 2019. Comparison of Region Filling Algorithms Using Texture Synthesis Methodologies, 2019 6th International Conference on Electrical and Electronics Engineering, 340-343. <https://doi.org/10.1109/ICEEE2019.2019.00072>
- Yılmaz, A.A., Guzel, M. S., Askerbeyli, I. ve Bostancı, E. 2020a. A Hybrid Facial Emotion Recognition Framework Using Deep Learning Methodologies, In: Human-Computer Interaction, Özseven, T. (eds), Nova Science Publishers.
- Yılmaz, A. A., Guzel, M. S., Askerzade, İ. and Bostanci, E. 2020b. A Novel Action Recognition Framework Based on Deep-Learning and Genetic Algorithms, IEEE Access Dergisi, 8, 100631-100644, <https://doi.org/10.1109/ACCESS.2020.2997962>
- Zappi, P., Lombriser, C., Stiefmeier, T., Farella, E., Roggen, D., Benini, L. ve Tröster, G. 2008. Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection, Wireless Sensor Networks, Springer, 17–33. https://doi.org/10.1007/978-3-540-77690-1_2
- Zeng, M., Nguyen, L.T., Yu, B., Mengshoel, O.J., Zhu, J., Wu, P. ve Zhang, J. 2014. Convolutional neural networks for human activity recognition using mobile sensors, International Conference on Mobile Computing, Applications and Services, 197–205. <https://doi.org/10.1109/10.4108/icst.mobibase.2014.257786>
- Zhang, Y., Pintea, S.L. ve Van Gemert, J.C. 2017. Video Acceleration Magnification, Computer Vision and Pattern Recognition. arXiv:1704.04186
- Zhang, Z., Ma, X., Song, R., Rong, X., Tian, X., Tian, G. ve Li, Y. 2017. Deep learning based human action recognition: A survey, Chinese Automation Congress (CAC), 3780-3785. <https://doi.org/10.1109/CAC.2017.8243438>
- Zhao, K., Du, J., Li, C., Zhang, C., Liu, H. ve Xu, C. 2013. Healthy: A Diary System Based on Activity Recognition Using Smartphone, IEEE International Conference on Mobile Ad-Hoc and Sensor Systems, 290-294, <https://doi.org/10.1109/MASS.2013.14>
- Zuiderveld, K. 1994. Contrast Limited Adaptive Histogram Equalization, Graphics Gems IV, 474-485.

ÖZGEÇMİŞ

Adı Soyadı : Abdullah Asım YILMAZ

Doğum Yeri : Altındağ/Ankara

Doğum Tarihi : 20.05.1988

Medeni Hali : Bekar

Yabancı Dili : İngilizce

Eğitim Durumu (Kurum ve Yıl)

Lise : MEV Özel Ankara Fen Lisesi (2005)

Lisans : Başkent Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği
(2010)

Yüksek Lisans : TOBB Ekonomi ve Teknoloji Üniversitesi Fen bilimleri Enstitüsü
Bilgisayar Mühendisliği Ana Bilim Dalı (2012)

Doktora : Ankara Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği
Anabilim Dalı (2020)

Çalıştığı Kurum/Kurumlar ve Yıl

T.C. Gıda,Tarım ve Hayvancılık Bakanlığı / 2011 – Devam Ediyor

Yayınlar (SCI/SCI-Expanded)

Yılmaz, A. A., Guzel, M. S., Askerzade, İ. and Bostanci, E. 2020. "A Novel Action Recognition Framework Based on Deep-Learning and Genetic Algorithms", IEEE Access Dergisi, Cilt 8, s. 100631-100644, <https://doi.org/10.1109/ACCESS.2020.2997962>

Kitap Bölümleri

Yılmaz, A. A., Guzel, M. S., Askerbeyli, İ. and Bostanci, E. 2020. “A Hybrid Facial Emotion Recognition Framework using Deep Learning Methodologies”, Human-Computer Interaction, NOVA Publishers.

Hakemli Dergiler

Yılmaz, A. A., Guzel, M. S. ve Askerbeyli, İ. 2017. " Algılanması Güç Olan Değişimlerin Ortaya Çıkarılması İçin Resim Teleskobu Çalışmalarının Analiz Çalışmaları", Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, Cilt. 19 (57) , s. 723-732. <https://doi.org/10.21205/deufmd.2017195765>

Uluslararası Bildiri ve Sempozyumlar

Yılmaz, A. A., Guzel, M. S., Askerzade, İ. and Bostanci, E. 2019. "Comparison of Region Filling Algorithms Using Texture Synthesis Methodologies", 2019 6th International Conference on Electrical and Electronics Engineering (ICEEE), Istanbul, Turkey, pp. 340-343, IEEE. <https://doi.org/10.1109/ICEEE2019.2019.00072>

U., Sayar, R., Yilmaz, B., Bostanci, E., Güzel, M. S. and **Yılmaz, A. A.** 2019. "Desktop Artillery Simulation Using Augmented Reality", 2019 6th International Conference on Electrical and Electronics Engineering (ICEEE), Istanbul, Turkey, pp. 79-85, IEEE. <https://doi.org/10.1109/ICEEE2019.2019.00023>

Yılmaz, A. A., Guzel, M. S., Askerbeyli, İ. and Bostanci, E. 2018. "A Vehicle Detection Approach Using Deep Learning Methodologies", II. International Conference on Theoretical and Applied Computer Science and Engineering, pp. 64-71, arXiv:1804.00429

Yılmaz, A. A., Guzel, M. S., Askerbeyli, İ. 2017. " Pedesterian Detection Studies Using Picture Telescope Approach", International Conference on Theoretical and Applied Computer Science and Engineering, Ankara, Turkey