

**KARADENİZ TEKNİK ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**

**BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**DOĞAL ORTAMLARDA YÜRÜYÜŞTEN KİŞİ TANIMA**

**DOKTORA TEZİ**

**Büşranur YAPRAK**

**MAYIS 2025**

**TRABZON**



**KARADENİZ TEKNİK ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**

**BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**DOĞAL ORTAMLARDA YÜRÜYÜŞTEN KİŞİ TANIMA**

**Büşranur YAPRAK**

**Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünde**  
**"DOKTOR (BİLGİSAYAR MÜHENDİSLİĞİ)"**  
**Unvanı Verilmesi İçin Kabul Edilen Tezdir.**

**Tezin Enstitüye Verildiği Tarih : 05 / 05 / 2025**

**Tezin Savunma Tarihi : 30 / 05 / 2025**

**Tez Danışmanı : Doç. Dr. Eyüp GEDİKLİ**

**Trabzon 2025**

## ÖNSÖZ

“Doğal Ortamlarda Yürüyüşten Kişi Tanıma” isimli bu tez Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, Doktora Programı’nda hazırlanmıştır.

Tez çalışmam boyunca hiçbir desteğini esirgemeyen ve bilim insanı olma yolundaki ilk adımlarımı atarken kıymetli bilgileriyle ve tecrübesiyle yolumu aydınlatan danışman hocam Sayın Doç. Dr. Eyüp GEDİKLİ’ ye teşekkürü bir borç bilirim. Ayrıca, yapıcı eleştirileri ve önerileri ile tezime büyük katkıları bulunan saygıdeğer hocalarım Sayın Dr. Öğr. Üyesi Ramazan Özgür DOĞAN’ a ve Sayın Dr. Öğr. Üyesi Özkan BİNGÖL’ e, eğitim öğretim hayatımda katkısı olan tüm hocalarıma ve tez çalışması süresinde de hiçbir yardımdan kaçınmayan arkadaşlarım Arş. Gör. Mehmet Ali KÜÇÜK’ e ve Öğr. Gör. Dilek KAYA’ ya teşekkür ederim.

Son olarak, doğduğum günden beri elimi hiç bırakmayan, hayatım boyunca aldığım tüm kararlarda bilgi ve tecrübelerinden faydalandığım ve her kararında arkamda olan, maddi ve manevi desteklerini hiçbir zaman esirgemeyen ve bugünlere gelmemde en büyük paya sahip olan annem Hulkiye KÜÇÜKUĞURLU ve babam Lokman KÜÇÜKUĞURLU’ ya ve eşim Yıldırım YAPRAK ‘a teşekkür eder, minnettarlığımı sunarım.

Bu tez çalışmasının bundan sonraki çalışmalara katkı sağlamasını temenni ederim.

Büşranur YAPRAK

Trabzon 2025

## TEZ ETİK BEYANNAMESİ

Doktora Tezi olarak sunduđum “Dođal Ortamlarda Yürüyüşten Kiři Tanıma” başlıklı bu çalışmayı baştan sona kadar danışmanım Doç. Dr. Eyüp GEDİKLİ’ nin sorumluluđunda tamamladıđımı, verileri/örnekleri kendim topladıđımı, deneyleri/analizleri ilgili laboratuvarlarda yaptıđımı/yaptırdıđımı, başka kaynaklardan aldıđım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiđimi, çalışma sürecinde bilimsel araştırma ve etik kurallara uygun olarak davrandıđımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiđimi beyan ederim. 30/05/2025

Büşranur YAPRAK

## İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ .....	III
TEZ ETİK BEYANNAMESİ.....	IV
İÇİNDEKİLER.....	V
ÖZET .....	VII
SUMMARY .....	VIII
ŞEKİLLER DİZİNİ.....	IX
TABLolar DİZİNİ .....	XI
SEMBOLLER VE KISALTMALAR DİZİNİ .....	XIII
1. GENEL BİLGİLER.....	1
1.1. Model Tabanlı Yaklaşımlar .....	2
1.2. Görünüm Tabanlı Yaklaşımlar .....	5
1.2.1. Silüet Çıkarma İşlemi ve RGB Çerçevesi .....	6
1.2.2. Silüet Tabanlı Yürüyüş Örüntüleri .....	14
1.2.3. Optik Akış ile Yürüyüş Örüntüsü .....	25
1.3. Makine Öğrenme İle Yürüyüş Tanıma .....	28
1.4. Derin Öğrenme Modelleri ve Transfer Öğrenme .....	31
1.4.1. VGG16 .....	32
1.4.2. ResNet50 .....	33
1.4.3. MobileNet .....	34
1.4.4. EfficientNet.....	37
1.4.5. ConvNeXt .....	38
1.4.6. Topluluk Öğrenme Mimarisi .....	41
1.4.7. Transfer Öğrenme.....	45
1.5. Literatür Veri Setleri ve Performans Karşılaştırması .....	45
2. YAPILAN ÇALIŞMALAR.....	49
2.1. Kullanılan Veri Setleri .....	50
2.2. Çoklu Modalite Ağ Mimarisi.....	52
2.3. Topluluk Öğrenmeli Parça-Tabanlı Ağ Mimarisi .....	56
2.3.1. Parça Eleme İşlemi .....	61
2.4. Benzerlik Hesaplaması ve Değerlendirme Kriteri.....	64

2.4.1. Kosinüs Benzerliği .....	64
2.4.2. Rank-1 Doğruluk Oranı.....	66
3. BULGULAR VE TARTIŞMA.....	67
3.1. Çoklu Modalite Ağ Mimarisi Başarı Değerlendirmesi .....	67
3.1.1. CASIA-B Veri Seti Performans Değerlendirmesi.....	68
3.1.2. Outdoor-Gait Veri Seti Performans Değerlendirmesi .....	76
3.1.3. MobileNet Hesaplama Maliyeti Analizi .....	78
3.2. Topluluk Öğrenmeli Parça-Tabanlı Ağ Mimarisi Başarı Değerlendirmesi.....	79
3.2.1. Parça Bazında Başarı Değerlendirmeleri .....	79
3.2.2. Parça Eleme İşleminin Tanımaya Etkisi .....	81
3.2.3. Topluluk Öğrenmeli Parça-Tabanlı Ağ Mimarisinin Performans Değerlendirmesi.....	82
4. SONUÇLAR.....	91
5. ÖNERİLER.....	93
6. KAYNAKLAR.....	94
ÖZGEÇMİŞ	

Doktora Tezi

## ÖZET

### DOĞAL ORTAMLARDA YÜRÜYÜŞTEN KİŞİ TANIMA

Büşranur YAPRAK

Karadeniz Teknik Üniversitesi  
Fen Bilimleri Enstitüsü  
Bilgisayar Mühendisliği Anabilim Dalı  
Danışman: Doç. Dr. Eyüp GEDİKLİ  
2025, 106 Sayfa

Yürüyüş, fiziksel temas veya açık rıza gerektirmeksizin kimlik tespiti yapılmasına olanak tanıyan davranış temelli bir biyometrik özelliktir. Bu avantajları sayesinde, özellikle yüz veya parmak izi gibi fiziksel biyometrik özelliklerin elde edilemediği güvenlik odaklı gözetim sistemlerinde kullanılabilir. Yürüyüşten kişi tanıma, benzer varyasyon koşullarında yüksek doğruluk sunarken, farklı varyasyonlar ile karşılaşıldığında hâlâ iyileştirilmeye açık bir araştırma problemi. Yürüyüş, farklı vücut parçalarının eşgüdümünü içeren davranışsal bir hareket olduğundan, her bir vücut bölgesinin tanıma üzerindeki katkısı farklılık göstermektedir. Buradan hareketle çalışmada, yenilikçi bir katkı olarak yerel vücut bölgeleri üzerinde çalışan bir topluluk öğrenme yaklaşımı önerilmektedir. Yatay vücut bölgelerinin her biri için ayrı bir derin öğrenme modeli eğitilmiş ve parçaları işleyen mimariye topluluk öğrenme uygulanmıştır. Parça tabanlı öznelikleri bütünleştirmek ve en ayırt edici olanları sentezlemek amacıyla, her bir vücut bölgesi için farklı bir derin öğrenme modeli daha eğitilmektedir. Ayrıca, görünüme dayalı varyasyonların tanıma üzerindeki etkisini azaltmak amacıyla, varyasyon içeren ve içermeyen görüntüler arasındaki mutlak farklara dayalı bir parça eleme yöntemi önerilmiştir. Önerilen yöntem, CASIA-B ve Outdoor-Gait gibi yaygın yürüyüş veri setleri üzerinde test edilmiş ve tanıma doğruluğu ile uygulanabilirlik açısından önemli iyileştirmeler sağlamıştır.

**Anahtar Kelimeler:** Parça-tabanlı yürüyüş tanıma, Topluluk öğrenme, Biyometri, Derin öğrenme, Uzaktan kimliklendirme

PhD. Thesis

## SUMMARY

### PERSON IDENTIFICATION FROM GAIT IN THE WILD

Büşranur YAPRAK

Karadeniz Technical University  
The Graduate School of Natural and Applied Sciences  
Computer Engineering Graduate Program  
Supervisor: Assoc. Prof. Eyüp GEDİKLİ  
2025, 106 Pages

Gait is a behavior-based biometric feature that enables identity recognition without requiring physical contact or explicit consent. Thanks to these advantages, it can be utilized in security-oriented surveillance systems, particularly when physical biometric features such as the face or fingerprints are unavailable. While gait-based person identification achieves high accuracy under similar conditions, it remains an open research problem when faced with varying scenarios. As gait is a behavioral motion involving the coordination of different body parts, the contribution of each region to recognition performance varies. Motivated by this, the study proposes an ensemble learning approach operating on localized body regions as a novel contribution. For each horizontal body region, an individual deep learning model is trained, and ensemble learning is applied to the models responsible for processing these body parts. To integrate part-based features and synthesize the most discriminative ones, an additional deep learning model is trained for each body region. Additionally, to mitigate the impact of appearance-based variations on recognition, a part elimination method based on absolute differences between images with and without variations is proposed. The proposed method is evaluated on widely used gait datasets such as CASIA-B and Outdoor-Gait, demonstrating significant improvements in recognition accuracy and practical applicability.

**Key Words:** Part-based gait recognition, Ensemble learning, Biometrics, Deep learning, Remote identification

## ŞEKİLLER DİZİNİ

	<u>Sayfa No</u>
Şekil 1. Model tabanlı yaklaşımlar için üretilmiş bazı insan vücut model yapıları (Sun vd., 2018) .....	3
Şekil 2. Görünüm tabanlı yaklaşımlar için bazı yürüyüş desenleri (Sun vd., 2018) .....	5
Şekil 3. Silüet elde etme işlemi .....	7
Şekil 4. 11 farklı açıda RGB yürüyüş çerçeveleri (Yu vd., 2006) .....	13
Şekil 5. İki farklı kişiye ait silüetler ve elde edilen GEI (Han & Bhanu, 2005) .....	16
Şekil 6. Optik akış hesaplama süreci .....	25
Şekil 7. Bir CNN mimarisi genel yapısı .....	32
Şekil 8. VGG16 ağı genel mimari yapısı.....	33
Şekil 9. Resnet50 mimarisi ve atlama bağlantısı yapısı.....	33
Şekil 10. Standart evrişim filtreleri (Howard vd., 2017) .....	35
Şekil 11. Derinlik bazlı evrişim filtreleri (Howard vd., 2017) .....	35
Şekil 12. Noktasal bazlı evrişim filtreleri (Howard vd., 2017) .....	35
Şekil 13. (a) Standart ve derinlik bazlı ayrılabilir evrişimli katman, (b) MobileNet yapısı .....	36
Şekil 14. EfficientNet ölçeklendirmesi (a) temel ağ yapısı, (b)-(d) sadece tek boyut ölçeklendirme, (e) EfficientNet bileşik ölçeklendirme (Tan & Le, 2019).....	37
Şekil 15. Temel EfficientNet mimari yapısı (Tan & Le, 2019).....	38
Şekil 16. Swin Transformer, ResNet ve ConvNeXt için blok tasarımları (Liu vd., 2022).....	41
Şekil 17. Farklı veri örneklemeleri üzerinde eğitilen ve üç temel sınıflandırıcıdan oluşan bir topluluk öğrenme mimarisi (Kazmaier & Van Vuuren, 2022) .....	42
Şekil 18. CASIA-B veri seti 11 farklı açıda GEI örnekleri .....	50
Şekil 19. CASIA-C veri seti GEI örnekleri .....	51
Şekil 20. Outdoor-Gait veri seti GEI örnekleri .....	51
Şekil 21. Çoklu modalite mimarileri (a) GEI ve SH, (b) OF, (c) HConL, (d) Leg ve Head.....	54
Şekil 22. Hizalanmış RGB görüntüsü, ikili maskesi, maskelenmiş RGB görüntüsü.....	56
Şekil 23. Topluluk Öğrenmeli Parça-Tabanlı Ağ Mimarisi .....	59
Şekil 24. Taşıma varyasyonu için parça eleme işlemi, soldan sağa sırasıyla varyasyonlu görüntü, varyasyonsuz görüntü, mutlak fark görüntüsü, parça elenmiş görüntü .....	62

Şekil 25. Giyim varyasyonu için parça eleme işlemi, soldan sağa sırasıyla varyasyonlu görüntü, varyasyonsuz görüntü, mutlak fark görüntüsü, parça elenmiş görüntü .....	62
Şekil 26. Görünüme dayalı varyasyonlar için parça eleme işlemli tanıma süreci örneği.....	63
Şekil 27. CASIA-B, NM varyasyonu, kombinasyonların performans karşılaştırması .....	70
Şekil 28. Her kombinasyonun NM, BG, CL varyasyonları altında ortalama <i>rank-1</i> karşılaştırması (a) $CNN_{GEI} + CNNSH$ , (b) $CNN_{GEI} + CNN_{OF}$ , (c) $CNN_H + CNN_L$ , (d) $CNN_{GEI} + CNN_{HConL}$ .....	72
Şekil 29. CASIA-B veri seti tüm varyasyonlar hazırlanan parça tabanlı öğrenme mimarilerinin karşılaştırılması (a) NM, (b) BG, (c) CL .....	84
Şekil 30. CASIA-B veri seti, görünüm tabanlı yöntemlerin <i>rank-1</i> oranları karşılaştırması .....	85
Şekil 31. CASIA-B veri seti, model tabanlı yöntemlerin <i>rank-1</i> oranları karşılaştırması .....	87
Şekil 32. CASIA-B veri seti önerilen yöntemlerin mevcut yöntemler ile özdeş açı altında tanıma doğruluğu karşılaştırması .....	88

## TABLolar DİZİNİ

	<u>Sayfa No</u>
Tablo 1. Literatürde en önde gelen veri setleri içerik karşılaştırma tablosu (Zhu vd., 2021) .....	46
Tablo 2. CASIA-B veri seti önde gelen çalışmalar performans karşılaştırması.....	47
Tablo 3. Topluluk öğrenmeli parça-tabanlı yürüyüş tanıma algoritması.....	60
Tablo 4. CASIA-B eğitim ve test küme bölümlenmesi .....	65
Tablo 5. CNN <sub>GEI</sub> + CNN <sub>SH</sub> , CNN <sub>GEI</sub> + CNN <sub>OF</sub> , CNN <sub>H</sub> + CNN <sub>L</sub> ve CNN <sub>GEI</sub> + CNN <sub>HConL</sub> kombinasyonlarının, CASIA-B veri seti NM varyasyonu, <i>rank-1</i> doğruluğu üzerinden karşılaştırması (özdeş açılar hariç) .....	68
Tablo 6. CNN <sub>GEI</sub> + CNN <sub>SH</sub> , CNN <sub>GEI</sub> + CNN <sub>OF</sub> , CNN <sub>H</sub> + CNN <sub>L</sub> ve CNN <sub>GEI</sub> + CNN <sub>HConL</sub> , CASIA-B, BG ve CL varyasyonu, <i>rank-1</i> (%) karşılaştırması (özdeş açılar hariç) .....	70
Tablo 7. CASIA-B veri seti NM varyasyonu önerilen bazı çoklu modalite mimarilerinin GaitNet (Song vd., 2019) ile çapraz görünüm tanıma oranlarının karşılaştırması.....	73
Tablo 8. Önerilen bazı çok modlu ağların, CASIA-B veri seti NM, BG, CL varyasyonları, literatürde önde gelen yöntemler ile ortalama <i>rank-1</i> (%) üzerinden (özdeş açılar hariç) karşılaştırması.....	74
Tablo 9. Çok modlu ağların, CASIA-B, NM varyasyonu MGan ile karşılaştırılması.....	75
Tablo 10. CASIA-B veri seti RGB ve maskelenmiş RGB (MRGB) bazlı tekli ve çoklu modalite mimarilerinin <i>rank-1</i> doğruluk oranı karşılaştırması (özdeş açılar hariç) .....	75
Tablo 11. Çoklu modalite ağının CASIA-B veri seti NM varyasyonu HGRBOL (Khan vd., 2023) yöntemi ile performans karşılaştırması (özdeş açılar hariç).....	76
Tablo 12. Outdoor-Gait veri seti üzerinde önerilen çok modlu ağların tüm varyasyonlar altında ortalama <i>rank-1</i> doğruluğu (%) üzerinden karşılaştırması .....	77
Tablo 13. Outdoor-Gait veri seti, ortalama <i>rank-1</i> doğruluğu (%) üzerinden bazı çok modlu ağlar ile görünüm tabanlı GaitNet (Song vd., 2019) yönteminin karşılaştırması .....	77
Tablo 14. Outdoor-Gait veri seti, ortalama <i>rank-1</i> doğruluğu (%) üzerinden bazı çok modlu ağlar ile model tabanlı 3DHuman (Meng vd., 2023) yönteminin karşılaştırması .....	78
Tablo 15. Ağların maliyet analizi. Boyut, nihai öznitelik vektörünün boyutudur .....	79
Tablo 16. CASIA-B veri seti tüm varyasyonlar parça ve bütün bazlı mimarilerin ortalama <i>rank-1</i> (%) oranları karşılaştırması.....	80

Tablo 17. CASIA-B veri seti tüm varyasyonlar parça elemeli (5part+elm) ve elemesiz 5 parçalı öğrenme (5part) <i>rank-1</i> (%) oranları.....	81
Tablo 18. CASIA-B tüm varyasyonlar, farklı parça tabanlı mimarilerin <i>rank-1</i> oranı (%) üzerinden karşılaştırılması (özdeş açılar hariç) .....	82
Tablo 19. CASIA-B veri seti, tüm varyasyonlar, görünüm ve model tabanlı yöntemler ile ortalama <i>rank-1</i> (%) üzerinden karşılaştırma (özdeş açılar hariç).....	85
Tablo 20. CASIA-B veri seti tüm varyasyonlar, önerilen yöntemlerin özdeş açı altında tanıma doğruluğu (%) performans karşılaştırması .....	88
Tablo 21. Önerilen yöntemin CASIA-C veri seti, mevcut yöntemlerle tanıma doğruluğu karşılaştırması.....	89
Tablo 22. Önerilen yöntemin Outdoor-Gait veri seti üzerinden mevcut yöntemlerle karşılaştırması .....	90



## SEMBOLLER VE KISALTMALAR DİZİNİ

AÖM	: Aşırı Öğrenme Makinesi
BBA	: Bağımsız Bileşenler Analizi
BGRU	: İki Yönlü Kapılı Tekrarlayan Birimler (Bidirectional Gated Recurrent Units)
BN	: Batch Normalizasyon
CCR	: Doğru Sınıflandırma Oranı (Correct Classification Rate)
CGI	: Krono Yürüyüş Görüntüsü (Chrono-Gait Image)
CNN	: Evrişimli Sinir Ağı (Convolutional Neural Network)
DOEH	: Doğrusal Olmayan Eşlenik Haritalama
DVM	: Destek Vektör Makineleri
EER	: Eşit Hata Oranı (Equal Error Rate)
EYK	: En Yakın Komşu
FCN	: Tam Konvolüsyonel Ağ (Fully Convolutional Network)
FConv	: Odaklı Konvolüsyon (Focal Convolution)
FDEI	: Çerçeve Farkı Enerji Görüntüsü (Frame Difference Energy Image)
ff-GEI	: Çerçeve Çerçeve GEI (frame-by-frame GEI)
FLOPs	: Saniye Başına Kayan Nokta İşlemleri (Floating Point Operations Per Second)
FPFE	: Çerçeve düzeyinde Parça Özellik Çıkarıcı (Frame-level Part Feature Extractor)
GAN	: Çekişmeli Üretici Ağ (Generative Adversarial Network)
GBDVM	: Genetik Bulanık Destek Vektör Makineleri
GCEM	: Yürüyüş Evrişim Enerji Haritaları (Gait Convolutional Energy Maps)
GCN	: Grafik Evrişimli Ağı (Graph Convolutional Network)
GDM	: Görüş Dönüşüm Modeli
GenI	: Yürüyüş Entropisi Görüntüsü (Gait Entropy Image)
GEI	: Yürüyüş Enerjisi Görüntüsü (Gait Energy Image)
GFI	: Yürüyüş Akışı Görüntüsü (Gait Flow Image)
GLFE	: Küresel ve Yerel Özellik Çıkarıcı (Global and Local Feature Extractor)
GTAA	: Genel Tensor Ayrımsal Analiz
HBF	: İnsan Vücut Odaklama (Human Body Focusing)
HMM	: Gizli Markov Modeli (Hidden Markov Model)
HPM	: Yatay Piramit Haritalama (Horizontal Pyramid Mapping)

IPC	: Düzensiz Tempo Dönüştürücü (Irregular Pace Converter)
KAA	: Kanonik Ayırıcı Analiz
KKA	: Kanonik Korelasyon Analizi
LDA	: Lineer Ayırıcı Analiz (Linear Discriminant Analysis)
LGSD	: Yerel Grafiksel İskelet Tanımlayıcı (Local Graphical Skeleton Descriptor)
LN	: Layer Normalizasyon
LRConv	: Yerel İlişki Evrişim (Local Relation Convolutional)
LSTM	: Uzun-Kısa Vadeli Bellek (Long Short-Term Memory)
LTA	: Yerel Zamansal Toplama (Local Temporal Aggregation)
MBCConv	: Ters Çevirilmiş Darboğaz (Inverted Bottleneck Convolution)
MCM	: Mikro-hareket Yakalama Modülü (Micro-motion Capture Module)
MGP	: Çok Katmanlı Küresel Boru Hattı (Multilayer Global Pipeline)
MKAA	: Marjinal Kanonik Korelasyon Analizi
MLP	: Çok Katmanlı Algılayıcı (Multilayer Perceptron)
MSFE	: Çok Ölçekli Özellik Çıkarıcı (Multi-scale Feature Extractor)
MSI	: Hareket Silüet Görüntüsü (Motion Silhouettes Image)
MvDA	: Çoklu-Açı Ayırıcı Analizi (Multi-view Discriminative Analysis)
PCA	: Temel Bileşen Analizi (Principal Component Analysis)
PEI	: Periyot Enerjisi Görüntüsü (Period Energy Image)
PHM	: Parça Tabanlı Yatay Eşleme (Part-based Horizontal Mapping)
PLD	: Işıklı Noktaların Görünümü (Point-Light Displays)
PSN	: Çiftler Arası Benzerlik Ağı (Pairwise Similarity Network)
PTP	: Zamansal Özellik Piramit Toplayıcı (Temporal Feature Pyramid Aggregator)
ReLU	: Düzeltmeli Doğrusal Ünite (Rectified Linear Unit)
RFAM	: Kalan Çerçeve Dikkat Mekanizması (Residual Frame Attention Mechanism)
RGB	: Renkli Görüntü
SED	: Bölme Çıkarma Cihazı (Slice Extraction Device)
SGD	: Stokastik Gradyan Azalma (Stochastic Gradient Descent)
SP	: Dizi Ortaklama (Set Pooling)
SRKAA	: Spektral Regresyon Kernel Ayrımsal Analiz
SVB	: Şekil Varyasyonu Tabanlı Friz (Shape Variation-based Frieze Pattern)
TDA	: Tekil Değer Ayrışımı
TM	: Toplayıcı Model

TP : Zamansal Ortaklama (Temporal Pooling)  
TTGS : Üçlü Yürüyüş Silüetleri (Triple Gait Silhouettes)  
ViT : Görüntü Dönüştürücü (Vision Transformer)



## 1. GENEL BİLGİLER

Bireyleri tanımlamak için kullanılabilen biyolojik ölçümler veya fiziksel özelliklere biyometri denir. Biyometri, genetik ve moleküler düzeydeki özellikleri kullanan biyolojik biyometri; vücudun fiziksel özelliklerini kullanan morfolojik biyometri; kişiye özgü kalıplara dayanan davranışsal biyometri olarak üç grupta incelenebilir. Parola ve anahtar gibi çalınması mümkün olmayan biyometri kişinin her zaman yanındadır ve kaybolmaz. Bu yönleri ile güvenlik amaçlı kullanımlarda gittikçe talep görmektedir. Biyometrik sistemler kişi tanıma ve kimlik doğrulama gibi farklı modlarda çalışır. İmza, güvenli alanlara giriş gibi işlemlerde kimlik doğrulama yapılırken, halka açık alanlarda kimlik belirleme/kişi tanıma yapılır. Kimlik doğrulamada, kişinin rızası gerekirken, kişi tanımada kişiden bilgi veya izin alınmasına gerek olmayabilir. Bu açıdan kişi tanıma daha çok istenmeyen durumların belirlenmesi için kullanılabilir. Kart gösterme, parmak izi cihazını kullanma, sağlık sistemine el okutma kişinin rızası ile kimliklerinin ispatlanmasını sağlarken bir hırsızın eşya üzerindeki parmak izi, DNA'sının belirlenebileceği kan veya bir saç teli delil niteliğindeki kişi tanıma için kullanılır. Günümüzde, biyolojik ve morfolojik biyometrilerin daha çok kimlik doğrulama, davranışsal biyometrilerin ise kişi tanıma için kullanıldığını görmekteyiz. Kişi tanımada rıza şartı olmadığından, yakalanma riskine karşılık, kişi davranışlarını değiştirme çalışması içinde olabilir. Sesini filtreden geçirmek, makyaj yapmak, kılık değiştirmek, farklı şekilde yazmaya çalışmak gibi davranışlar bunlara örnek olabilir. Biyometrik sistemlerin amacı, davranış sürecinde yapılan deterministik hareketi yakalayıp hedef kümesini küçültmek ve kişiye özgü davranışı belirlemektir. Özellikle kalabalık ortamlarda, caddeler, hava sahalarında yüzden, sestən, yürüyüş davranışlarından kişi tanıma uzak mesafelerde çözüm üretmektedir. Yine de bazı durumlarda insan yüzünü saklayabilir veya ses ayırıştırması zor olabilir. Kişi kılık değiştirse veya aksak yürüse de yürüyüşten herhangi bir açıdan biyometrik özelliğini saklamayı başarabilir. Yürüyüş tarzında bir kişinin davranışsal ve fiziksel özellikleri bir arada bulunur. Bu nedenle, yürüyüş tanınmanın temel amacı; insan vücudunun şekli hakkındaki uzamsal (spatial) ve zamansal (temporal) karakteristiklerden yola çıkarak, benzersiz ve değişmez özellikleri çıkarabilmek ve bu özellikleri kullanarak kimliklendirme işlemini başarıya ulaştırabilmektir. Yürüyüş biyometrisi yüz, parmak izi, iris, ses gibi fiziksel biyometrelerle karşılaştırıldığında; (I) herhangi bir kullanıcı iş birliğine (onayına) veya bir tarayıcıya gerek duymadan temassız

olarak elde edilebilmesi (II) verilerin düşük çözünürlükte veya uzak mesafeden alınmış olsa bile tanıma işleminde sonuç üretebilmesi ve (III) dinamik bir davranışı temsil ettiğinden kopyalanma ihtimalinin düşük olması sebepleriyle avantajlı bir tanıma işlemi sunar.

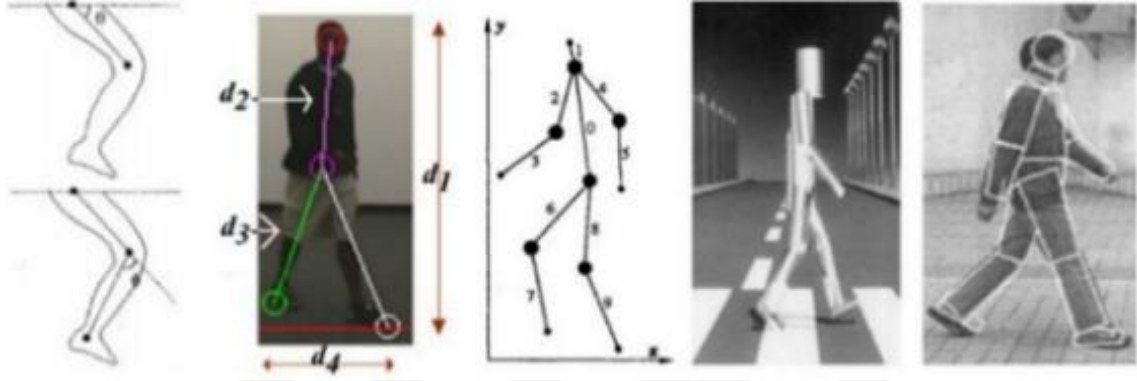
Yürüyüş şeklinin incelenmesi ilk olarak sağlık alanında gündeme gelmiştir. Murray ve diğerlerinin (1964) çalışmasında yürüyüşün farklı düzlemlerdeki kinematik bileşenlerini zamanla ilişkilendirerek ölçmeye ve yaş ve boy gibi faktörlerle ilişkilerini inceleyerek yürüyüş bozuklukları tespit edilmeye çalışılmıştır. Başka bir çalışmada Johansson (1975) tarafından ışıklı noktaların görünümü (Point-Light Displays, PLD) adıyla bilinen yöntem geliştirilmiştir. İnsan vücudunun belli noktalarına (örneğin eklemelere) ışıklar yerleştirilerek, karanlık bir ortamda sadece bu ışıklı noktaların hareketi izlenmiştir. Böylece, yalnızca bu noktaları izleyerek bile insanlar, yürüyen bir insan figürünü tanıyabilmiştir. Biyometrik olarak ilk çalışmalara örnek Niyogi ve Adelson' un (1994;1994) yürüyüş silüetini standart yöntemlerle incelemeleri gösterilebilir. Bu çalışmalarda silüet üzerine eğriler yerleştirilerek etiketleme yapılmış ve bu etiketler aracılığıyla yürüyüş deseni elde edilmeye çalışılmıştır.

Günümüzde, video gözetimi için kameraların yaygın olarak kullanılması yürüyüş tanımayı sosyal güvenlik, suç önleme veya adli kimliklendirme gibi 'gerçek dünya' uygulamaları için yararlı bir araç haline getirmektedir. Avantajlarının yanı sıra yürüyüş biyometrisi; çapraz açı problemi, giyim tarzı veya nesne taşıma, yürüyüş hızı değişimi gibi görünüm tabanlı varyasyonlarından oldukça etkilenir ve bu varyasyonlar altında tanıma başarısı ciddi oranda düşer. Bu varyasyonlar altında alınmış yürüyüş görüntülerindeki tanıma başarılarını artırmaya yönelik literatürde şimdiye kadar model tabanlı yaklaşımlar ve görünüm tabanlı yaklaşımlar olmak üzere iki farklı yaklaşım türü incelenmiştir. Bu çalışmada, model tabanlı yaklaşımların hesaplama maliyetinin yüksek olması sebebiyle, parçalı bölgelerin ayrı ayrı değerlendirilmesi ve farklı yürüyüş örüntülerinin bir arada kullanılmasına dayalı, özellikle tanıma oranının ciddi oranda düştüğü durumlarda başarıyı artırmaya yönelik bir görünüm tabanlı yaklaşım önerilmiştir. Sonraki bölümlerde literatürde sunulan model ve görünüm tabanlı yaklaşım türleri detaylandırılmıştır.

### **1.1. Model Tabanlı Yaklaşımlar**

Model tabanlı yaklaşımlar yürüyüş biyometrisinin iç yapısına yönelen yaklaşımlardır. Bu yaklaşımlarda, 2B (2-Boyutlu) veya 3B (3-Boyutlu) görüntülerden bireylerin vücut yapılarını ve farklı vücut bölümleri arasındaki yerel hareket desenlerini modelleyerek

yürüyüş tanıma gerçekleştirilir. Bu yaklaşımlara dayalı yapılan çalışmalarda genellikle kol, bacak gibi vücut bölümleri ve bunlar arasındaki oluşan hareket, bir çubuk modeli ile simüle edilir (Nixon & Carter, 2006) veya vücut farklı sayılarda bölümlere ayrılarak ve bu bölümler arasında oluşan eklem bölgelerindeki açıların değişiminden yola çıkarak bir tanıma işlemi gerçekleştirilir (Wang vd., 2004). Şekil 1’ de model tabanlı yaklaşımlar için insan vücut iskelet yapısını ve eklemler arası bağlantıyı farklı şekillerle temsil ederek üretilmiş bazı örnek model yapıları gösterilmiştir.



Şekil 1. Model tabanlı yaklaşımlar için üretilmiş bazı insan vücut model yapıları (Sun vd., 2018)

Derin öğrenme alanında meydana gelen son zamanlardaki ilerlemelerle birlikte model tabanlı yürüyüş tanıma yaklaşımlarından bir çalışmada (Feng vd., 2016), CASIA-B (Yu vd., 2006) veri setindeki renkli görüntülerden (RGB) kişinin sağ-sol ayak, sağ-sol bacak, sağ-sol el vb. 12 farklı uzvunun sıcaklık haritasını çıkararak poz tahmini yapabilen bir Evrişimli Sinir Ağı (Convolutional Neural Network, CNN) mimarisi ile her bir yürüyüş çerçevesi için ayrı ayrı özellik çıkarımı yapmışlardır. Zamansal özelliklerin çıkarımı için elde edilen ardışık sıcaklık haritaları bir Uzun-Kısa Vadeli Bellek (Long Short-Term Memory, LSTM) mimarisine verilmiştir. Yapılan deneyler sonucunda daha önce çapraz açılı problemi için önerilmiş yöntemlerden, ortalama başarı yüzdesi parametresi üzerinden daha yüksek performans elde edildiği görülmüştür. Bu alanda yapılmış diğer bir derin öğrenme temelli çalışmada (Liao vd., 2020), içerisinde dört farklı özelliği barındıran ‘PoseGait’ adında yeni bir özellik tanımlanmıştır. Bu özelliklerden ilki ‘fpoz’ renkli görüntülerin 3B poz görüntülerini, ikincisi ‘façı’ eklemler arası değişen açı bilgilerini, üçüncüsü ‘fuzuv’ iki eklem arasındaki uzunluğu, son olarak ‘fmotion’ eklemlerin hareket bilgisini barındırır. Önerilen yeni özelliğin başarımı, hazırlanan CNN mimarisi ile CASIA-B veri setinin üç

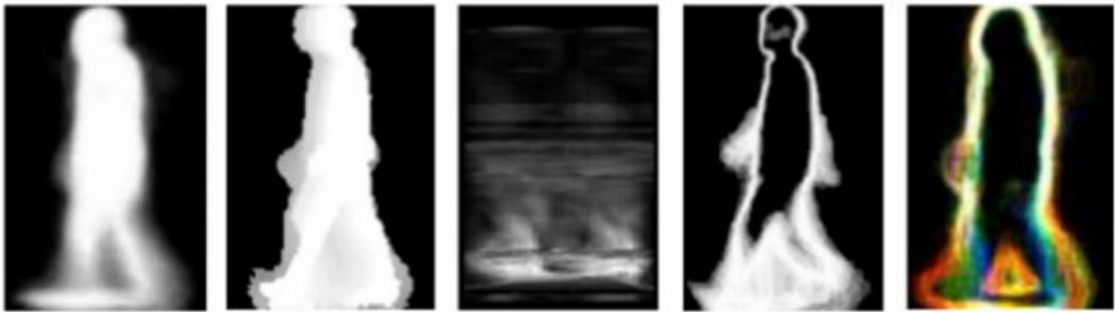
farklı yürüyüş varyasyonu olan normal yürüyüş (NM), çanta ile yürüyüş (BG) ve ceket ile yürüyüş (CL) varyasyonları üzerinde değerlendirilmiştir. Elde edilen sonuçlarda önerilen yeni özellik, özdeş ve çapraz açı değerlerinde karşılaştırılan diğer yöntemlere göre özellikle CL varyasyonunda daha yüksek performansa ulaşmıştır. Diğer yandan Teepe ve diğerleri (2021) yürüyüş özelliklerinin daha net ve daha etkili çıkarılması amacıyla iskelet pozlarını bir Grafik Evrişimli Ağ (Graph Convolutional Network, GCN) ile birleştiren, güçlü uzamsal ve zamansal modelleme sağlayan ‘GaitGraph’ı önermişlerdir. Bir GCN ağı, birçok ardışık ve zamansal anlamda klasik 2B Evrişim ağlarından ve dar boğazlı yedek bağlantılı bloklardan meydana getirilmiştir. CASIA-B veri seti üzerinde yapılan doğruluk oranları karşılaştırmasında PoseGait yöntemine göre tüm varyasyonlarda — NM, BG, CL — daha ileri seviyede performansa ulaşılmıştır. Önerilen diğer bir çalışmada (Xu vd., 2021), yürüyüş iskeletlerinin sınıf içi ve sınıflar arası yerel grafiksel desenlerini tanımlayabilmek amacıyla bir Yerel Grafiksel İskelet Tanımlayıcı (Local Graphical Skeleton Descriptor, LGSD) modülü tanımlanmıştır. Buna ek olarak, aynı veya farklı kimliklere ait yürüyüş özellikleri eşleştirilerek, doğru eşleşen çiftlerin benzerliğini maksimize ve yanlış eşleşen çiftlerin benzerliğini minimize eden Çiftler Arası Benzerlik Ağ (Pairwise Similarity Network, PSN) geliştirilmiştir. Önerilen yöntem CASIA-B ve CASIA-C (Tan vd., 2006) veri setleri üzerinden literatürdeki önde gelen çalışmalar ile karşılaştırılmış, elde edilen sonuçlarda çapraz açı doğruluklarında yüksek performansa erişilemese de özdeş açı doğruluklarında özellikle PoseGait çalışmasına göre daha başarılı performansa erişilmiştir. Başka bir çalışmada ise (Li & Zhao, 2022), zamansal yürüyüş özelliklerini çıkarabilmek amacıyla Zamansal Özellik Piramit Toplayıcı (Temporal feature Pyramid Aggregator, PTP) adında, paralel birçok zamansal evrişim operatörünü bir araya getirebilen bir modül hazırlanmıştır. Bu modül ile uzamsal yürüyüş özelliklerini çıkarabilen bir GCN ağı kullanılarak yürüyüş verisinden uzamsal-zamansal özellikler elde edilmiştir. Çalışmada ek olarak, düzenli iskelet çerçevelerini düzensiz çerçeveler haline dönüştürebilen ve farklı yürüyüş döngülerine sahip veriler üretebilen bir Düzensiz Tempo Dönüştürücü (Irregular Pace Converter, IPC) modülü de hazırlanmıştır. ‘CycleGait’ olarak adlandırılan mimarinin, CASIA-B veri seti üzerinde yapılan değerlendirmelerinde, tüm varyasyonlarda PoseGait, GaitGraph ve PSN model tabanlı metodlarından daha yüksek tanıma oranlarına ulaşılmıştır.

Model tabanlı yaklaşımlara dayalı yöntemler, literatürde oldukça fazla sayıda önerilmiştir. Bu yöntemler; çapraz açı, ceket, ayakkabı giyme ve çanta, klasör taşıma gibi görünüm varyasyonlarına karşı dayanıklıdır. Ancak insan vücudunu iyi bir şekilde

modelleyebilmek için yüksek çözünürlüklü kamera ile iyi yakalanmış görüntülere ihtiyaç duyarlar. Model çıkarma aşamalarının hesaplama maliyetleri de oldukça yüksektir. Ayrıca vücut yapısının çıkarılmasının oldukça zor olduğu düşük çözünürlüklü görüntülerde hata oranı yüksek olacağından dış mekân uygulamaları için de uygun çözümler üretmezler. Bu nedenle görünüm tabanlı yaklaşımlar literatürde oldukça geniş yer tutmaktadır. Bir sonraki bölümde görünüm tabanlı yaklaşımlar detaylandırılmıştır.

## 1.2. Görünüm Tabanlı Yaklaşımlar

Görünüm tabanlı yaklaşımlar videodan ayırt edici özellikleri doğrudan çıkarır, modeli açıkça oluşturmaları gerekmez ve model tabanlı yaklaşımlara göre daha düşük çözünürlük tanıma için yeterli olduğundan maliyetleri düşüktür. Görünüm tabanlı yaklaşımlara dayalı yürüyüş özniteliklerinin çıkarılması için literatürde birçok yöntem önerilmiştir ve bu yöntemler genellikle yürüyüş periyoduna dayalıdır. Örneğin, bir Yürüyüş Enerjisi Görüntüsü (Gait Energy Image, GEI) (Han & Bhanu, 2005), bir yürüyüş periyodu dizilerinin hizalanmış silüetlerinin tek bir görüntüye ortalaması alınarak oluşturulur. Benzer şekilde, Hareket Silüet Görüntüsü (Motion Silhouettes Image, MSI) (Lam & Lee, 2006), Şekil Varyasyonu Tabanlı Friz (Shape Variation-based Frieze Pattern, SVB) (Lee vd., 2007), Çerçeve Farkı Enerji Görüntüsü (Frame Difference Energy Image, FDEI) (Chen vd., 2009), Krono Yürüyüş Görüntüsü (Chrono-Gait Image, CGI) (Wang vd., 2010), Yürüyüş Entropisi Görüntüsü (Gait Entropy Image, GEnI) (Bashir vd., 2010), Yürüyüş Akışı Görüntüsü (Gait Flow Image, GFI) (Lam vd., 2011) ve Periyot Enerjisi Görüntüsü (Period Energy Image, PEI) (He vd., 2018) gibi desenler yürüyüş döngüsünü, yürüyüşü temsil etmek amacıyla tek bir görüntüde birleştirmişlerdir. Şekil 2' de görünüm tabanlı yaklaşımlara dayalı üretilmiş bazı örnek yürüyüş desenleri sunulmuştur.



Şekil 2. Görünüm tabanlı yaklaşımlar için bazı yürüyüş desenleri (Sun vd., 2018)

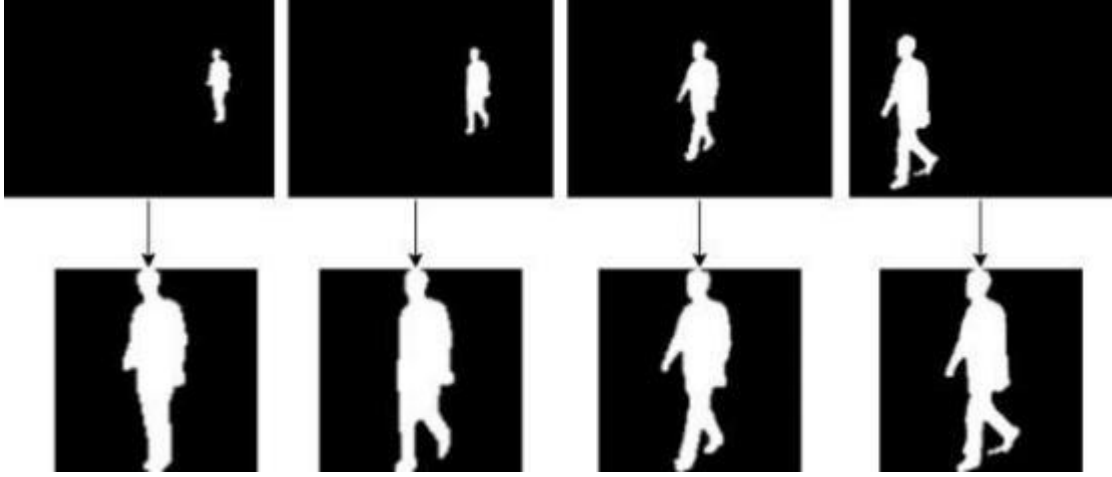
Şekil 2' de soldan sağa ilk yürüyüş deseni görüntüsü GEI' dir. İkinci görüntü MSI, her pikselin hareketinin zamansal geçmişine dayanırken, üçüncü görüntü SVB silüetlerin yatay ve düşey projeksiyonlarına dayanır. Dördüncü görüntü deseni GENI silüetleri Shannon entropisine göre kodlarken, beşinci desen CGI kodlama işlemi için bir renk eşleme işlevi kullanmıştır.

Bu tez kapsamında görünüm tabanlı yaklaşımlara yönelik çalışmalar yapıldığından sonraki bölümlerde yürüyüş verilerinden silüet elde etme işlemi ile literatürde en sık rastlanan ve tez kapsamında kullanılan görünüm tabanlı farklı yürüyüş desenleri detaylandırılmıştır.

### 1.2.1. Silüet Çıkarma İşlemi ve RGB Çerçevesi

Geçmişten günümüze görünüm tabanlı yaklaşımların tamamına yakını, yürüyüşü yürüyüş video dizilerinden çıkarılan silüetlere dayanarak temsil eder. Bunun nedeni, yürüyüşün davranışsal bir biyometri olması ve kişinin statik görünümünü yakalayan yüz gibi fiziksel biyometrilere farklı olarak, insan hareketinin dinamik yönünü yakalamasıdır. Yürüyüş dizilerinden silüetler çıkarıldığında, insanın fiziksel görünümüne ait birçok özellik görüntü temsilinden kaldırılmış olur. Ancak bir silüet hâlâ insan vücudunun şekli ve duruşu hakkında bilgi içerir.

Bir yürüyüş silüeti, bir yürüyüş çerçevesi görüntüsünün küçük bir kısmını kaplar. Bunun sebebi, görüş açıları farkından kaynaklanan arka plan ve kamera mesafesi faktörünün değişimidir. Bu nedenle, mesafe ve görüş açısı değişimlerinin etkilerini azaltmak için iki temel işlem uygulanır; ölçek normalizasyonu ve merkezi hizalama. Bu işlem için Takemura ve diğerleri (2018) tarafından önerilen yöntem bu çalışma kapsamında kullanılmıştır. Öncelikle, her bir yürüyüş çerçevesi için silüet alanlarının üst, alt ve yatay merkez noktaları elde edilmiştir. İkinci olarak, bu konumlara bir hareketli ortalama filtresi uygulanmıştır. Üçüncü adımda, ortalama konumlara göre silüet görüntülerinin boyu  $m$  piksele sabitlenecek şekilde boyut normalizasyonu gerçekleştirilmiş ve her bölgenin en boy oranı korunmuştur. Son olarak, ortalama yatay medyan görüntüsünün yatay merkezine karşılık geldiği  $n \times m$  piksel boyutlarında bir görüntü üretilmiştir. Burada  $n$  ve  $m$  istenen silüet boyutlarının en ve boy ölçüleridir. Şekil 3' te birinci satırda ikili yürüyüş çerçevesi, ikinci satırda ise ölçek normalizasyonu ve merkezi hizalama işlemleri sonucu elde edilen ikili silüet görüntülerine örnekler sunulmuştur.



Şekil 3. Silüet elde etme işlemi

Silüete dayalı yürüyüş tanıma işlemi özellikle son teknoloji yöntemlerde en sık kullanılan girdi verisi olmuştur. Bunların en başında Chao ve diğerleri (2019) tarafından geliştirilmiş ‘GaitSet’ mimarisi gelmektedir. Bu mimari temelde üç ana yapıdan oluşmaktadır. İlk yapı, bir CNN mimarisinden oluşmakta ve birbirinden bağımsız her bir silüet görüntüsünden ‘çerçeve seviyesi’ bazında özellikler çıkarmaktan sorumludur. Dizi Ortaklama (Set Pooling, SP) adı verilen ikinci yapı ise çerçeve seviyesi bazındaki özellikleri tek bir ‘dizi seviyesi’ özelliğinde birleştirmekten sorumludur. Böylece uzamsal ve zamansal özellikler başarılı bir şekilde elde edilmiş olur. Son yapı ise Yatay Piramit Haritalama (Horizontal Pyramid Mapping, HPM) olarak adlandırılmış ve dizi seviyesi özellikleri daha ayrılabilir bir uzaya haritalamaktan sorumlu tutulmuştur. Çalışmada ayrıca, yapısında CNN mimarisi bulunduran Çok Katmanlı Küresel Boru Hattı (Multilayer Global Pipeline, MGP) adlı farklı bir modül ile farklı seviyelerdeki ara katmanlardan çıkarılan dizi seviyesindeki bilgilerin kaybı önlenmiş, ‘Batch All Triplet Fonksiyonu (Hermans vd., 2017)’ kayıp fonksiyonu olarak kullanılmıştır. CASIA-B veri seti üzerinde ilk 24 kişinin eğitim, kalan 100 kişinin test olarak kullanıldığı küçük örneklem (ST); ilk 62 kişinin eğitim, kalan 62 kişinin test olarak kullanıldığı orta örneklem (MT) ve ilk 74 kişinin eğitim, kalan 50 kişinin test olarak kullanıldığı geniş örneklem (LT) isimlendirmeleri altında üç farklı deney yapılmıştır. Test verilerindeki galeri ve sorgu görüntüleri arasındaki benzerlik Öklid mesafesi ile hesaplanmış olup *rank-1* doğruluk oranları elde edilmiştir. CASIA-B veri seti üzerinde yapılan deneylerin sonuçları DeepCNN (Wu vd., 2016), çalışması ile karşılaştırılmış ve ST, MT, LT deneylerinin hepsinde DeepCNN’ e göre oldukça yüksek başarılar elde edilmiştir. Bu başarılarla göre LT deneyi için çapraz açısı ve NM varyasyonunda

tüm açılarda ortalama %95.0, BG varyasyonunda %87.2, CL varyasyonunda %70.4 başarı elde edilmiştir. OU-MVLP (Takemura vd., 2018) üzerinde yapılan çapraz açı deneyleri sonuçları GeiNet (Shiraga vd., 2016) ile karşılaştırılmış, tüm açılarda ortalama GeiNet'te %35.8 başarı alınırken GaitSet' te %87.1 başarı elde edilmiştir. Sonuç olarak bu mimari ile yürüyüş tanımadaki çapraz açı ve farklı yürüyüş varyasyonlardaki düşük doğruluk oranları ciddi oranda yükseltilmiştir.

Fan ve diğerleri (2020) tarafından önerilen performansı oldukça yüksek diğer bir çalışmada, insan bedeninin bölümlerinden özellik çıkarmaya dayalı 'GaitPart' adlı bir yöntem sunulmuştur. Yöntemde amaç, insan bedeninin her bir bölümünün kendine ait uzamsal-zamansal bir ifadesini elde edebilmektir. Önerilen yöntem temel olarak iki ana bileşenden oluşur; Çerçeve düzeyinde Parça Özellik Çıkarıcı (Frame-level Part Feature Extractor, FPFE) ve Mikro-hareket Yakalama Modülü (Micro-motion Capture Module, MCM). GaitPart mimarisi ağ girdisi olarak silüet görüntülerini alır. FPFE yapısı silüet görüntülerini yatay parçalara ayırır ve yapısındaki birbiri ardına sıralanmış CNN yapıları ile, her parça için ayrı ayrı, parça bazında daha faydalı uzamsal özellikleri çıkarmaktan sorumludur. MCM yapıları ise FPFE' den elde edilen parça bazındaki özellik satırlarını, satır bazında bir araya getirerek her bir parça için uzamsal-zamansal özellikleri elde eder. Ayrıca barındırdığı Zamansal Ortaklama (Temporal Pooling, TP) yapısı sayesinde, bir araya getirilen özelliklerden bir maksimize fonksiyonu yardımıyla silüet çerçevelerinin içerisindeki en ayrılabilir mikro hareket özelliklerini seçer. Son aşamada, birçok tam bağlantı katmanı kullanılarak özellik vektörleri daha ayrılabilir bir uzaya düşürülür. CASIA-B veri setinde yapılan deneylerde ilk 74 kişi eğitim, son 50 kişi test aşaması için kullanılmıştır. Çalışmada test setindeki galeri ve sorgu görüntüleri arasındaki benzerlik Öklid mesafesi ile hesaplanmıştır. DeepCNN ve GaitSet ile yapılan karşılaştırmalarda tüm varyasyonlarda bu çalışmalardan daha başarılı sonuçlar alındığı görülmüştür. OU-MVLP veri setinde yapılan deneylerde ise diğer yöntemlere göre yine performansı en yüksek yöntem olduğu sonucuna ulaşılmıştır.

Sadece özdeş açı üzerinde tanıma performansını artırmaya yönelik yapılan çalışmalardan Thapar ve diğerleri (2018) tarafından önerilen 'VGR-Net' adlı metotta, zamansal ve uzamsal özellikleri aynı anda elde edebilmek amacıyla; birinci mimari silüet görüntülerinin açısının tespiti için, ikinci mimari ise yürüyüş tespiti için olmak üzere iki farklı 3BCNN mimarisi kullanılmıştır. Bu iki farklı mimaride de video sınıflandırma için daha önce 'Sports-1M (Karpathy vd., 2014)' veri seti üzerinde eğitilmiş hazır bir 3BCNN

mimarisi kullanılmıştır. CASIA-B veri seti için transfer öğrenme yapılan bu mimarilerden birincisinde ağın son 5 katmanı çıkarılıp yerine 2 bağlantı katmanı ve bunları takip eden birer bırakma katmanı; ikincisinde ise 7 bağlantı katmanı ve bunları takip eden birer bırakma katmanı eklenmiştir. Mimarilerin ilkinde, veri seti toplam 11 farklı açı içerdiğinden sınıflandırıcı son katman 11 nöron; ikincisinde ise veri seti toplam 124 farklı kişi içerdiğinden 124 nöron olarak belirlenmiştir. İkinci mimarinin eğitimi 11 farklı açıya özel olarak 11 defa farklı açı bazında gerçekleştirilmiştir. Çalışmada orijinal videolar yerine, videoların 16' şarlı silüet görüntülerinden oluşturulan klipsler kullanılmıştır. Açısı tespit edilen klipsler, daha sonra yürüyüş tanıma işlemi için ilgili ikinci mimariye girdi olarak verilmiştir. Çalışmada ayrıca, deney aşamasında stereo görüntüler ve kısmi örtüşen klipsler de test amaçlı kullanılmış, elde edilen Doğru Sınıflandırma Oranı (Correct Classification Rate, CCR) yüzdeleri, Wolf ve diğerleri (2016) tarafından önerilen 3BCNN tabanlı çalışmanın CCR parametreleri ile karşılaştırılmıştır. Elde edilen deney sonuçlarında birçok özdeş açı değerinde daha yüksek performans elde edildiği görülmüştür.

Çapraz açı problemine yönelik daha ayırıcı yürüyüş özelliklerini elde etmek amacıyla yapılmış 'ACL' adı verilen diğer bir çalışmada (Zhang vd., 2019), yeni bir kayıp fonksiyonu önerilmiştir. Bu çalışmada amaç yürüyüş tanımında sıkça kullanılan Merkez Kayıp Fonksiyonu (Wen vd., 2016)' ndan yola çıkarak sadece ayrılabilir özellikleri elde edebilen değil, gerçek anlamda ayırt edebilen bir kayıp fonksiyonu elde edebilmektir. Bu amaçla Merkez Kayıp Fonksiyonu'ndan farklı olarak, aynı kişiye ait tüm açılara tek bir merkez noktası atamak yerine, her bir açı için de alt bir merkez nokta atanır. Yani atanan toplam merkez sayısı kişi sayısı  $\times$  açı sayısı şeklinde hesaplanır. Her bir merkez ve seçilen 'anchor' arasındaki mesafe açısının maksimum değeri kayıp olarak seçilir. Burada anchor tanımlı bir kimliğe ait olan, sistemin referans noktası olarak kullandığı örnek özelliktir. Alt merkez noktalarının kullanılması ve açılarla bağdaştırılmalarından yola çıkılarak bu fonksiyon 'Açı Merkez Kayıp Fonksiyonu' olarak adlandırılmıştır. Çalışmada ayrıca, belirlenmiş sayıdaki ardışık silüet görüntüleri, literatürde önerilmiş bir öğrenme ağı kullanılarak yatay 4 ayrı parçaya bölünmüş, böylece en işlevsel 4 parça elde edilmiştir. Elde edilen parçalar, başta CNN ardından LSTM katmanlarından oluşan mimariye ayrı ayrı verilerek uzamsal ve zamansal özellikler kaybedilmeden bir arada elde edilmeye çalışılmıştır. Çalışmanın test aşamalarında önerilen yeni kayıp fonksiyonu ile yürüyüş tanımında sıkça kullanılan literatürdeki diğer kayıp fonksiyonları arasında ve silüet görüntülerini manuel olarak 2, 3, 4, ve 5 parçaya ayırma ile öğrenimli parçalara ayırma yöntemleri arasında karşılaştırmalar

yapılmıştır. Önerilen Açık Merkez Kayıp Fonksiyonu'nun CASIA-B veri setindeki doğruluk değerleri standart Merkez Kayıp Fonksiyonu'nun üzerine çıkmıştır. Yine 4 parçalı öğrenimli bölümlenimin performansı en yüksek bölümlenme şekli olduğu ispatlanmıştır. Çalışmadaki sonuçlar, CASIA-B, OU-ISIR (Iwama vd., 2012) ve OU-MVLP veri setleri üzerinde DeepCNN ve GaitSet yöntemleri başta olmak üzere literatürdeki diğer çalışmalar ile karşılaştırılmış, GaitSet'e yakın ancak ortalama başarı parametresi bazında daha yüksek sonuçlar elde edilmiştir. Çalışmanın görünüme dayalı yürüyüş varyasyonları üzerindeki başarısı ölçülmemiştir.

Çekişmeli Üretici Ağ (Generative Adversarial Network, GAN) (Goodfellow vd., 2020) mimarisi kullanılarak literatürde önerilen yöntemlerin, çapraz açı ve giyim varyasyonlarındaki performansını artırmaya yönelik yapılmış diğer bir çalışmada (Chen vd., 2021), 11 farklı açı değerinde görüntüler içeren CASIA-B ve OU-MVLP veri setlerinin her ikisinden de yararlanarak üretilmiş sahte silüetler ile yeni bir eğitim veri seti hazırlanmıştır. Bu veri seti için hazırlanan GAN tabanlı mimaride, hepsi farklı amaçları gerçekleştirmek için toplam 4 farklı kayıp fonksiyonundan faydalanılmıştır. Bunlardan ilki 'Ayrıcı Kayıp Fonksiyonu' GAN mimarisindeki ayırıcının, üreticinin ürettiği yürüyüş silüetinin gerçek veya sahte olma olasılıklarını hesaplamada kullanılmıştır. Burada üretici ise ayırıcıyı hataya düşürmeye çalışır. Dolayısıyla üretici bu kayıp fonksiyonunu minimize etmeye çalışırken, ayırıcı maksimize etmeye çalışır. İkincisi 'Açık Sınıflandırma Kayıp Fonksiyonu' ağın hedeflenen açıda silüet üretmesini sağlayan fonksiyondur. Üçüncüsü 'Döngü Tutarlılık Kayıp Fonksiyonu' ile sadece gerçeğe yakın ve hedeflenen açıda sorunsuz bir sahte silüet üretimi değil, üretim işleminde meydana gelebilecek diğer varyasyonlardan sahte silüetin yürüyüş bilgilerini kaybetmemesi hedeflenir. Sonuncu 'kimliklendirme kayıp fonksiyonu' ise üretilen sahte yürüyüş silüeti ile orijinal görüntü çiftinin aynı kişiye ait olup olmadığının belirlenmesi için kullanılmıştır. Son olarak bu 4 kayıp fonksiyonu birleştirilerek minimize edilmesi amaçlanan tek bir kayıp fonksiyonu elde edilmiştir. Çalışmada ayrıca orijinal ve üretilen sahte silüetler arasındaki dağılım farkını azaltmak ve bu iki silüeti birbirine yaklaştırmak için bir 'etki alanı hizalama' ağı da tasarlanmıştır. CASIA-B ve OU-MVLP veri setleri üzerinde yapılan deneylerde veri setindeki her bir açı değerine özel, hazırlanan GAN mimarisi belirli bir açı üzerinden alınmış yürüyüş silüeti üzerinden kalan tüm diğer açılara göre sahte silüetler üretmesi için ayrı ayrı eğitilmiştir. Daha sonra elde edilen sahte silüetler her iki orijinal eğitim veri setlerine eklenip GaitSet mimarisi eğitilmiş ve test görüntüleri üzerinde bu eğitilmiş ağ denenmiştir. Elde edilen sonuçlar, GaitSet' in orijinal

sonuçları ile literatürdeki diğer bazı GAN tabanlı yöntemlerin başarıları ile karşılaştırılmış, çapraz açı ve giyim varyasyonlarının tümünde oldukça iyi sonuçlar elde edilmiştir. Çalışmada ayrıca CASIA-B orijinal eğitim veri setine her iki veri setinden de üretilen sahte silüetler eklenerek toplamda 22 farklı açıdan oluşan yeni bir veri seti oluşturulmuş, bu veri seti üzerinden bir eğitim daha gerçekleştirilmiştir. Eğitim sonucunda OU-MVLP veri seti üzerinde yapılan deneylerde de başarılı sonuçlar elde edildiği görülmüştür.

GaitSet mimarisini genişleterek kullanan ‘mmGaitSet’ adlı diğer bir çalışmada (Zhao vd., 2022), hesaplanan sıcaklık haritası görüntüleri ile silüet görüntülerinden çıkarılan özelliklerin bir arada kullanılmasıyla çanta taşıma, BG ve ceket giyme, CL varyasyonlarına karşı daha dayanıklı özellikler elde edebilmek amaçlanmıştır. Bu çalışmada, model tabanlı yöntemlerde sıklıkla kullanılan poz çıkarma işlemi gerçekleştirilmiş ancak çıkarma işlemi literatürde daha önce önerilmiş bir evrişim tabanlı ağ yapısıyla sağlanmıştır. Bu ağ yapısı RGB yürüyüş çerçeve görüntülerini ağ girdisi olarak alır ve her biri farklı vücut bölgelerini gösteren, bu bölgelere denk gelen alanları daha parlak piksellerle işaretleyen 16 farklı görüntü üretir. Bu görüntülerin birleştirilmesiyle vücuttaki sağ-sol ayak, sağ-sol bacak, sağ-sol el vb. 16 farklı vücut bölgelerini işaretleyen gri seviyeli sıcaklık haritası görüntüleri elde edilir. Elde edilen sıcaklık haritaları ve silüet görüntüleri birbirine paralel ve genişletilmiş GaitSet tabanlı iki ayrı mimariye iki ayrı koldan verilerek özellik çıkarımı sağlanmıştır. İki ayrı koldan elde edilen çıktılar daha sonra tam bağlantı katmanı ile birleştirilmiştir. Kurulan ağ yapısı sayesinde düşük seviyeli yapısal özellikler ile yüksek seviyeli mantıksal özellikler birleştirilmiş, daha ayırıcı ve genişletilmiş bir yürüyüş temsili elde edilmiştir. Çalışmada kayıp fonksiyonu olarak ‘Üçlü Kayıp Fonksiyonu’ kullanılmış, test aşamasında galeri ve sorgu görüntüleri arasındaki benzerlik Öklid mesafesi ile hesaplanmıştır. Başta (Zhang vd., 2019), GaitSet ve GaitPart çalışmaları olmak üzere literatürdeki farklı birçok çalışma ile CASIA-B veri seti üzerinde çapraz açı ve giyim varyasyonları bazında yapılan performans değerlendirmesinde, bazı çapraz açı değerleri arasındaki doğruluk oranlarında daha yüksek, geriye kalan doğruluk oranlarında ise bu çalışmalara yakın sonuçlar elde edildiği görülmüştür.

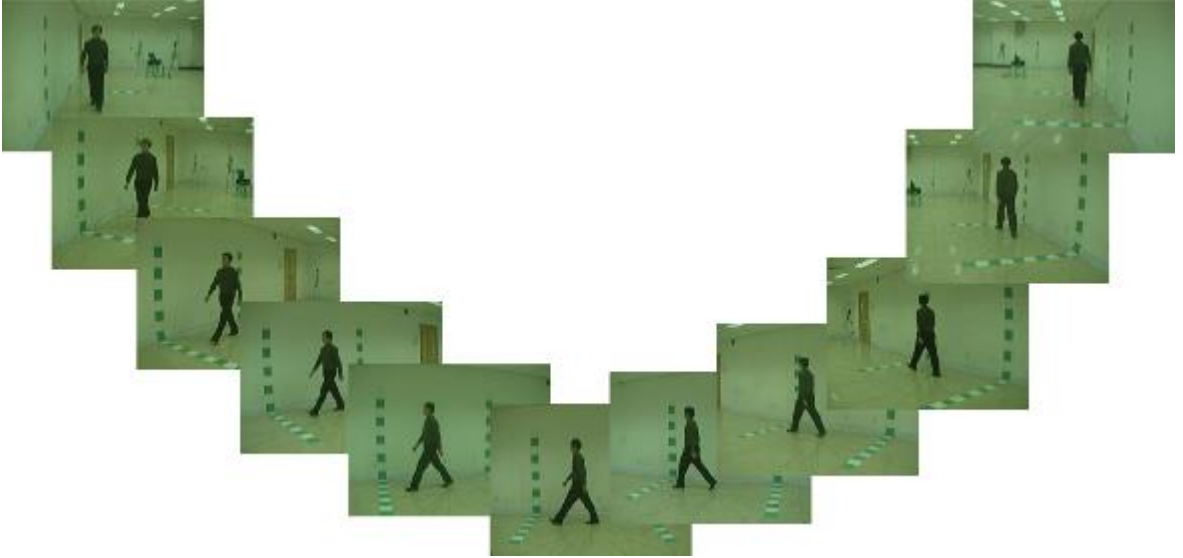
Han ve diğerleri (2022) tarafından yapılan diğer bir çalışmada çapraz açı varyasyonundaki performans düşüklüğüne çözüm olarak yürüyüş özelliklerini ayrılabilir yapmak yerine tamamen ayırt edici hale getirebilen, yeni bir kayıp fonksiyonu hesaplama yöntemi önerilmiştir. Çalışmada ana ağ yapısı olarak GaitSet mimarisi kullanılmış ve literatürde yürüyüş tanımada sıkça kullanılan tüm kayıp fonksiyonları karşılaştırılarak,

bunların arasından ‘Angular Softmax Kayıp Fonksiyonu ( $L_A$ ) (Liu vd., 2017)’ ile ‘Batch-All Triplet Kayıp Fonksiyonu ( $L_{BA}$ ) (Hermans vd., 2017)’ önerilen yöntem için seçilmiştir.  $L_A$  fonksiyonu ayrılabilir özellikleri kosinüs uzayında öğrenirken,  $L_{BA}$  fonksiyonu farklı kişilerin özellikleri arasındaki mesafeyi artırıp, aynı kişiler arasındaki mesafeyi azaltarak Öklid uzayını kullanır. Çalışmada ayrıca, fonksiyonların optimize uzayları farklı olduğundan bu durumun meydana getirebileceği etkiyi azaltmak ve eğitim işlemini uygulanabilir hale getirebilmek amacıyla, ağ mimarisindeki tam bağlantılı katmanlardan önce mimariye bir ‘Batch-Normalizasyon’ katmanı eklenmiştir. Eğitim aşamasında önce özellikler,  $L_{BA}$  kayıp fonksiyonunu hesaplamak için, sonra sınıflandırma vektörleri,  $L_A$  kayıp fonksiyonunu hesaplamak için, kullanılmıştır. Son olarak toplam kayıp  $L = L_{BA} + \alpha L_A$  denklemi ile elde edilmiştir. Çalışmada test aşamasındaki galeri ve sorgu görüntüleri arasındaki benzerlik,  $k$ -NN sınıflandırıcısı ile hesaplanmıştır. Yapılan deneylerde  $L_{BA}$  ve  $L_A$  kayıp fonksiyonlarının mimaride ayrı ayrı kullanılması ile bir arada kullanılması durumlarının performansları ele alınmış, Batch-Normalizasyon katmanının mimariye eklenmiş ve eklenmemiş hallerinin performansları karşılaştırılmıştır. Sonuç olarak, kayıp fonksiyonlarının bir arada olduğu ve mimarinin Batch-Normalizasyon katmanının eklenmiş versiyonunun en yüksek performansa ulaştığı görülmüştür. Çalışmanın performansı ayrıca literatürdeki (Wang & Yan, 2020), MGan (He vd., 2018), DeepCNN ve GaitSet çalışmalarının, CASIA-B veri setindeki çapraz açı ve giyim varyasyonları üzerindeki ortalama sonuçları ile karşılaştırılmış, daha iyi performans elde edildiği görülmüştür. Literatürdeki diğer bazı çalışmalar ile TUM-GAIT (Hofmann vd., 2014) veri seti üzerinde yapılan deneylerde de daha başarılı performansa ulaşılmıştır.

Uzamsal ve zamansal özellikleri aynı anda çıkarabilmeyi amaçlayan ‘GaitSlice’ adı verilen diğer bir çalışmada (Li vd., 2022) uzamsal özellikler için Bölme Çıkarma Cihazı (Slice Extraction Device, SED), zamansal özellikler için Kalan Çerçeve Dikkat Mekanizması (Residual Frame Attention Mechanism, RFAM) ile adlandırılmış iki modül hazırlanmıştır. İnsan vücudundaki birbirine bağlı bölümlerin yürüme esnasındaki etkileşen yapılarından yola çıkarak SED mimarisi, yürüyüş çerçevesindeki insan bedenini yatay olarak parçalara ayırıp bu parçalardan çıkardığı özellikleri baştan aşağı birbirleriyle ilişkilendirir. Birbirine paralel RFAM yapıları ise zamansal bilgiyi yakalarken elde edilen her bir yatay bölme için kilit çerçeve görüntüsünü elde tutarak ön plana çıkarır. Hazırlanan mimaride, ilk olarak ardışık evrişim katmanları ile uzamsal özellikler çıkarılır, elde edilen özellikler global maksimum ve minimum ortaklama işlemleri ile daha üst boyutta yatay

çizgiler halinde haritalanır. SED yapıları ile vücut parçaları birbiriyle ilişkilendirilir, paralel RFAM yapıları barındırdıkları önemli bilgilere göre çerçeveleri yeniden ağırlıklandırarak daha iyi yatay vücut bölmelerini elde tutar. Elde edilen yeniden ağırlıklandırılmış yatay bölmeler bir araya getirilerek bölme bazında uzamsal-zamansal özellikler hesaplanır. Son olarak bu özellikleri yüksek boyutlu vektör haline getirebilmek için bir tam bağlantı katmanı kullanılır. Ağ mimarisindeki modüllerde GaitPart' ın barındırdığı yapılardan faydalanılmıştır. Literatürdeki PoseGait, DeepCNN, GaitSet ve GaitPart, çalışmaları ile CASIA-B veri seti üzerinde yapılan çapraz açı ve giyim varyasyonları üzerindeki performans karşılaştırmalarında, tüm ortalama değerlerinde önerilen yöntem en performanslı sonuca ulaşmıştır. Bu durum OU-MVLP veri seti için de geçerli olmuştur.

GaitSlice mimarisine benzer olarak son zamanlarda yürüyüş tanıma alanında yapılan çalışmalarda silüetlerin uzamsal ve zamansal özniteliklerine aynı anda odaklanıldığı görülmektedir (Lin vd., 2021; Sepas-Moghaddam & Etemad, 2020; Sepas-Moghaddam vd.,2021; Ma vd., 2023; Chen vd., 2023; Wei vd., 2024; Pan vd., 2024). Bu çalışmalarda temel amaç, küresel ve yerel özniteliklerin daha verimli çıkarımını sağlamak veya elde edilen yerel öznitelikler arasındaki global ilişkileri daha anlamlı biçimde modelleyebilmektir. Bahsi geçen çalışmalar ilerleyen bölümlerde detaylandırılmıştır.



Şekil 4. 11 farklı açıda RGB yürüyüş çerçeveleri (Yu vd., 2006)

Yürüyüş tanımda RGB çerçeveleri genellikle kullanılacak olan silüet, silüet tabanlı yürüyüş örüntüleri, optik akış veya derinlik haritası gibi girdi verilerinin hesaplanmasının

ilk aşamasında kullanılır. Çerçevelerin direkt olarak öğrenme algoritmasına verilmesi, kıyafet ve arka plan faktörünün tanımaya dahil olabilmesi ve öğrenmeyi olumlu veya olumsuz etkileyebilme olasılığı sebebiyle literatürde nadirdir. Bu nedenle ön işlem basamaklarının ham girdi verisi olarak ele alınır. CASIA-B veri setinden 11 farklı açıda alınmış bazı örnek orijinal RGB yürüyüş çerçeveleri Şekil 4' te gösterilmiştir.

RGB çerçevelerinin direkt olarak öğrenme algoritmasına verildiği çalışmalardan birine örnek olarak Khan ve diğerleri (2023) tarafından önerilen bir çalışmada birden fazla girdi verisinin özniteliklerinin birleştirilmesi üzerine bir mimari hazırlanmıştır. Çalışmada ilk girdi verisi olarak kontrastı iyileştirilmiş RGB çerçeveleri, ikinci girdi verisi olarak ise optik akış görüntüleri kullanılmıştır. Bu iki farklı girdi verisinden transfer öğrenme ile ince ayarlanmış EfficientNetB0 (Tan & Le, 2019) ağı kullanılarak en faydalı öznitelikler çıkarılmış ve ağ parametreleri Bayes optimizasyonu kullanılarak optimize edilmiştir. Böylece ağların veriye en uygun parametreler ile eğitimi tamamlanmıştır. İki farklı girdi verisinden elde edilen en faydalı özellikler birleştirme işlemine tabi tutularak tanıma gerçekleştirilmiştir. Çalışmada CASIA-B ve CASIA-C veri setleri deneyler için kullanılmış bu veri setlerindeki tüm kişiler eğitim aşamasına dahil edilmiştir. Elde edilen deney sonuçlarında tüm varyasyonlarda oldukça yüksek tanıma oranlarına erişilse de bu tanıma oranlarına erişmede en büyük sebep RGB çerçevelerinin kullanımı ve eğitime veri setindeki tüm kişilerin dahil edilmesi gösterilebilir. Bu alanda yapılmış diğer bir çalışmada (Zhang vd., 2019) ise 'Ayrıştırılmış Temsil Öğrenimi' adlı oto kodlayıcı tabanlı bir ağ önerilmiştir. Oto kodlayıcı yapısı RGB çerçevelerinden elde edilen RGB yürüyüş görüntüsünü direkt olarak görünüm ve poz tabanlı özelliklere ayırır. Bu yapı, aynı kişiye ait, farklı kıyafetler ve kamera açılarıyla çekilmiş iki yürüyüş videosundaki iki özelliğin farkından, etiket verisi olmadan, görünüm özelliklerini ayırmayı ve çıkarmayı amaçlar. Çalışmada CASIA-B veri seti üzerinde yapılan deneylerde literatürde daha önce önerilmiş karşılaştırılan yöntemlere göre daha başarılı performansa erişildiği görülmüştür.

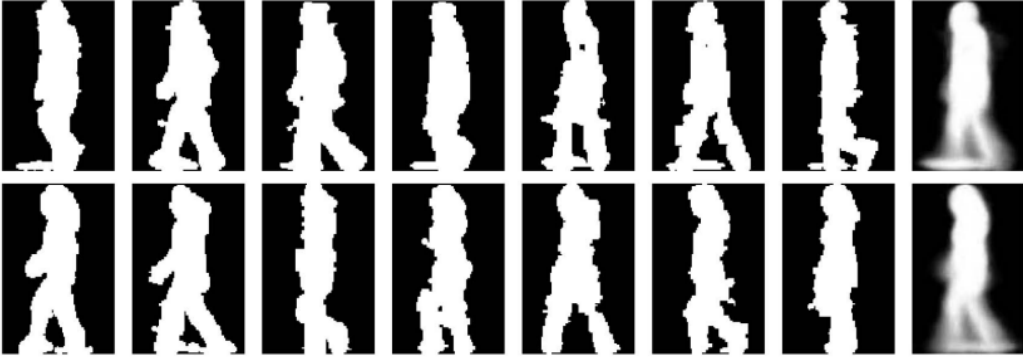
### **1.2.2. Silüet Tabanlı Yürüyüş Örüntüleri**

Yürüyüş tanıma alanında görünüm tabanlı yaklaşımlara yönelik yapılan ilk çalışmalarda, zamansal bilgiyi korumak amacıyla önerilen stratejiler iki kategoriye ayrılabilir. İlk kategoriye ait yöntemlerde yürüyüş verilerindeki zamansal bilgi eğitim aşamasında korunur. Örneğin, Sundaresan ve diğerleri (2003) bu amaçla Gizli Markov

Modeli (Hidden Markov Model, HMM) tabanlı bir yaklaşım sunarken, Sarkar ve diğerleri (2005) galeri ve sorgu dizileri arasındaki dizi çiftlerinin korelasyon bilgisini kullanmışlardır. Wang ve diğerleri (2003) ise yürüyüş çerçevelerinin istatistiksel uzamsal-zamansal özelliklerini çıkarmak için Temel Bileşen Analizi (Principal Component Analysis, PCA) uygulamıştır. Ancak bu yöntemlerin büyük ölçekli veri seti gerektirmeleri ve yüksek hesaplama karmaşıklıkları sebebiyle ikinci strateji olarak yürüyüş çerçeveleri dizisini tek bir şablona dönüştürme işlemi kullanılmıştır. Bu amaçla ilk olarak Liu ve Sarkar (2004) insan yürüyüşünü tüm silüetlerin ortalamasıyla temsil etmeyi önermiştir. Bu çalışmadan ilham alan Han ve Bhanu (2005), Yürüyüş Enerji Görüntüsü (Gait Energy Image, GEI) desenini sunmuştur. GEI insan hareketinin uzamsal ve zamansal özelliklerini bir arada ele alan etkili ve verimli bir yürüyüş örüntüsüdür. Bashir ve diğerleri (2010) ise GEI'lerdeki entropiye dayalı değişmez yürüyüş alt uzaylarını Yürüyüş Entropisi Görüntüsü (Gait Entropy Image, GEnI) adı verdikleri yeni bir yürüyüş örüntüsüyle araştırmıştır. Başka bir çalışmada (Wang vd., 2010), yürüyüş çerçeveleri arasındaki zamansal bilgi renk haritalama yoluyla kodlanmıştır. Her bir yürüyüş görüntüsündeki dış kenarlar çıkarıldıktan sonra, aynı yürüyüş dizisindeki her bir yürüyüş dış kenar görüntüsünü kodlamak için bir renk haritalama fonksiyonu kullanılmıştır. Bu görüntüler tek bir Krono Yürüyüş Görüntüsü (Chrono-Gait Image, CGI) adı verilen yürüyüş deseninde birleştirilmiştir. Yürüyüş Akışı Görüntüsü (Gait Flow Image, GFI), her döngüdeki ikili silüetlerden optik akış alanı belirlenerek elde edilmiştir. Literatürde, bahsi geçen yürüyüş çerçevelerini tek bir görüntüde birleştiren yöntemlere benzer olarak silüetlerin her bir pikselinin hareketinin zamansal geçmişine bağlı üretilen Hareket Silüet Görüntüsü (Motion Silhouettes Image, MSI), silüetlerin yatay ve dikey projeksiyonlarına dayanan Şekil Varyasyonu Tabanlı Friz (Shape Variation-based Frieze Pattern, SVB) deseni, ardışık çerçevelerin farklarının pozitif parçalarının GEI ile toplanmasıyla elde edilen Çerçeve Farkı Enerji Görüntüsü (Frame Difference Energy Image, FDEI), GEI ve CGI desenlerinin avantajlı yanlarından faydalanarak üretilen Periyot Enerjisi Görüntüsü (Period Energy Image, PEI) gibi birçok çalışma yapılmıştır. Ancak GEI, bu çalışmalar arasında literatürde en sık kullanılan yürüyüş deseni olarak ön plana çıkmıştır. Diğer temsillere kıyasla GEI, yürüyüş silüetlerinin genel formunu ve yürüyüş döngüsü boyunca zamansal değişimlerini yansıtırken kontur kenarının yerel bozulmalarına ve küçük ölçekli değişimlere de nispeten dayanıklıdır (Wang & Yan, 2020). Ancak görüş açısı farkından ve görünüme dayalı varyasyonlardan etkilenir. Bir GEI, yürüyüş çerçevelerinin düzgün hizalanmış silüetlerin ortalaması alınarak,

$$G(x, y) = \frac{1}{N} \sum_{t=1}^N B_t(x, y) \quad (1)$$

biçiminde hesaplanmaktadır ve düşük bir hesaplama maliyetine sahiptir. Burada  $N$ , bir yürüyüş döngüsündeki toplam çerçeve sayısı,  $t$ , döngüdeki çerçevenin numarası,  $x$  ve  $y$ , 2B görüntünün koordinatlarıdır. Şekil 5’ te iki farklı kişiye ait yürüyüş döngülerinin silüetleri ve her satırın en sağında bu silüet görüntülerinden elde edilen GEI’ ler gösterilmiştir.



Şekil 5. İki farklı kişiye ait silüetler ve elde edilen GEI (Han & Bhanu, 2005)

GEI’ nin kullanıldığı yürüyüş tanıma metotlarına örnek olarak GENI deseninin önerildiği çalışmada, değişmez özellikleri tespit edebilen, otomatik yürüyüş özellik seçici yeni bir yöntem önerilmiştir. Bu yöntemde GEI’ lerdeki her bir piksel konumunda ‘Shannon Entropisi’ ölçülerek bir maske görüntüsü elde edilmiştir. Böylece GEI’ lerdeki dinamik ve statik bölgelerinin ayrımı sağlanmıştır. Her bir galeri ve sorgu görüntüsüne özel hesaplanan GENI maskeleri ile GEI’ lerdeki daha ayırt edici özellikler ön plana çıkarılmış, çanta taşıma ve ceket giyme varyasyonlarındaki önemsiz kısımların yürüyüş tanıma aşamasına dahil edilmesi önlenmeye çalışılmıştır. Çalışmada ayrıca elde edilen faydalı özellikler üzerinden galeri ve sorgu görüntüleri arasındaki benzerliği hesaplayabilmek için hesaplama maliyeti daha az olan yeni bir Adaptif Bileşen ve Ayırıcı Analiz yöntemi önerilmiştir. CASIA-B ve Soton (Shutler vd., 2004) veri setleri üzerinde yapılan deneyler sonucunda tatmin edici sonuçlar elde edildiği görülmüştür. Önerilen diğer bir çalışmada (Choudhury & Tjahjadi, 2015) çanta taşıma ve ceket giyme varyasyonlarına karşı dayanıklılık sağlamak amacı ile GEI’ lerin sadece ayak bölgesi entropisini baz alarak, sorgu görüntülerinin eşleşen galeri görüntü açısını iki fazda bulan bir yürüyüş tanıma yöntemi önerilmiştir. Eşleşen görüntü açısını bulmak amacıyla birinci fazda 2B PCA ve Öklid mesafesi kullanılırken ikinci fazda

çok boyutlu şekil analizi kullanılmıştır. USFHumanID (Sarkar vd., 2005) veri seti üzerinde yapılan deneylerde, karşılaştırılan yöntemlerin bazılarında oldukça yüksek performans elde edilmişken, kalan diğer yöntemlerle yakın sonuçlar elde edildiği görülmüştür. Benzer bir diğer bir çalışmada ise (Ghebleh & Ebrahimi Moghaddam, 2018) galeri ve sorgu GEI' leri arasındaki benzerlik hesabı satır satır hesaplanmış, ayırt edici olmayacak aykırı satırlar, bu benzerlik üzerinden adaptif olarak tespit edilmeye çalışılmıştır. Aykırı satırın belirlenmesi için her bir satırdan gelen Öklid mesafelerinin ortalama değeri eşik değeri olarak belirlenmiş ve bu eşik değerin üzerinde kalan satırlar yürüyüş tanıma işlemine dahil edilmemiştir. Bu yöntem ile çanta taşıma ve ceket giyme varyasyonlarının yürüyüş tespiti performansı üzerindeki etkisi düşürülmeye çalışılmıştır. CASIA-B ve OU-ISIR (Iwama vd., 2012) veri setleri üzerinde hesaplanan *rank-1,5* ve 10 parametrelerinde karşılaştırılan yöntemlerden daha yüksek performansa ulaşılmıştır.

Yapay sinir ağlarındaki hızlı gelişmelerle birlikte model tabanlı yaklaşımlara benzer olarak görünüm tabanlı yaklaşımlarda da yürüyüş tanıma problemleri derin öğrenme yöntemleri ile çözülmeye çalışılmış ve başarılı sonuçlar elde edilmiştir. Bu alanda derin öğrenme metotları kullanılarak yapılan ilk çalışma Shiraga ve diğerleri (2016) tarafından 3BCNN mimarisi kullanılarak GEI girdi verisi üzerinden yapılmış ve önerilen ağ mimarisine 'GeiNet' adı verilmiştir. Çalışmada sırasıyla 3B evrişim, ortaklama ve normalizasyon katmanlarından oluşan iki ardışık blok ve bu blokları takip eden iki tam bağlantı katmanı ile bir mimari oluşturulmuştur. OU-ISIR veri setinde bulunan dört farklı açı üzerinden özdeş ve çapraz açılar için yapılan performans değerlendirmesinde literatürde daha önce derin öğrenme kullanmayan farklı metotların başarımlarına göre oldukça yüksek doğruluk değerlerine ulaşıldığı görülmüştür. Derin öğrenme ile yapılmış ilk çalışmalardan bir diğerinde ise (Zhang vd., 2016) aynı kişiye ait GEI' ler arası benzerlik mesafesini minimum, farklı kişilere ait GEI' ler arası benzerlik mesafesini maksimum yapabilen bir 'Siam ağ' tasarlanmıştır. Bu ağ yapısında birbiri ile tamamen aynı parametreleri paylaşan iki CNN ağ mimarisi birbirine paralel konumdadır. Benzer (aynı kişiye ait) veya benzer olmayan (farklı kişilere ait) GEI çiftleri sırasıyla '1' ve '0' ile etiketlenerek ağa verilmiş ve çıktılar, 'zıt kayıp fonksiyonu' nu hesaplamak için bir 'zıt kayıp katmanı' ile bir araya getirilmiştir. OU-ISIR veri setindeki dört farklı açı değerinde yapılan özdeş ve çapraz açı deneylerinde tatmin edici sonuçlara erişilmiştir.

Yu ve diğerleri (2017) tarafından önerilen bir çalışmada çapraz açı, çanta taşıma ve ceket giyme varyasyonlarının başarımlarını performansları üzerindeki etkisi, tek bir ağ mimarisi

kullanarak azaltılmaya çalışılmıştır. Geliştirilen ağ mimarisinde yedi adet ardışık Oto kodlayıcı (Hinton & Salakhutdinov, 2006) kullanılmıştır. Bunlardan ilki, GEI özelliği CL varyasyonunda ise, NM varyasyonuna; ikincisi, BG varyasyonunda ise NM varyasyonuna dönüştürmekten sorumludur. Diğer beş Oto kodlayıcı ise GEI hangi açıda olursa olsun, her bir katmanda ardışık farklı açılardan birine sırasıyla dönüştürmek koşuluyla, son katmanda GEI özelliğini yan görünüm açısına ( $90^\circ$ ) dönüştürür. Ayrıca ağın bazı katmanlarında daha iyi özellikleri de kaybetmemek için PCA kullanılmıştır. Elde edilen ayırt edici özellikler en yakın komşu,  $k$ -NN, yöntemi kullanılarak sınıflandırılmıştır. CASIA-B veri seti kullanılarak özdeş ve çapraz açı üzerinde yapılan deneylerde en düşük başarı çapraz açıda CL varyasyonunda elde edilmiş, NM varyasyonunda ise karşılaştırılan yöntemlere yakın performans elde edildiği sonucuna varılmıştır.

Alotaibi ve Mahmood (2017) tarafından önerilen başka bir çalışmada dört evrişim ile her bir evrişim katmanını takip eden dört ortaklama katmanı ve bir tam bağlantı katmanından oluşan, yapısında standart CNN mimarisindeki katmanlardan farklı sayılarda parametre içeren kendilerine özgü bir CNN mimarisi geliştirmişlerdir. Bu mimariyi GEI özelliği üzerinden eğiterek CASIA-B veri seti üzerinde üç farklı deney yapmışlardır. Yapılan ilk deneyde, CASIA-B veri setindeki 124 kişinin sadece  $90^\circ$  deki sırasıyla NM, BG ve CL varyasyonlarındaki GEI görüntüleri ile ağın eğitimleri tamamlanmış daha sonra ağı benzer varyasyonlar altında test edebilmek için, test görüntüleri de sırasıyla  $90^\circ$  deki NM, BG ve CL varyasyonlarındaki GEI görüntülerinden oluşturulmuştur. Gerçekleştirilen ikinci deneyde veri setindeki 124 kişinin tüm varyasyonlarından eğitim setine çok az sayıda örnekler konulmuş, test veri seti de yine bu varyasyonları barındıracak şekilde oluşturulmuştur. Çalışmada son deney çapraz-açı problemine yönelik gerçekleştirilmiş ve veri setindeki ilk 24 kişi ağın eğitim aşaması için, son 100 kişi test aşaması için kullanılmıştır. Literatürde daha önce önerilmiş, karşılaştırılan çalışmalardan birçok çapraz-açı aralığında daha yüksek doğruluk değerine ulaşıldığı görülmüştür.

Wu ve diğerleri (2017) tarafından önerilen çalışmada literatürde önerilmiş çalışmaların çapraz açı ve yürüyüş varyasyonları üzerindeki performans düşüşlerine yönelik olarak 'DeepCNN' adı altında başarılı bir yöntem sunulmuştur. Bu çalışmada oluşturulan ağ mimarilerinde, aynı kişiye ait GEI'lerden oluşturulan çiftlere '1', farklı kişilere ait GEI'lerden oluşturulan çiftlere '0' etiketi verilmiştir. Ağlar girdilerden yola çıkarak, yerel ve global özellikleri öğrenmek için çiftleri birbirleri ile karşılaştırır, otomatik olarak en ayırt edici yürüyüş özelliklerini çıkarır ve bu çiftler için benzerlik seviyesini tespit eder.

Çalışmada farklı sayılarda evrişim, normalizasyon, ortaklama ve tam bağlantı katmanları içeren tek bir CNN mimarisi ile siam ağlara benzer, birbirine paralel iki CNN mimarisi içeren iki farklı mimari olmak üzere toplam üç farklı CNN mimarisi kurulmuştur. CASIA-B veri seti için ilk 50 kişi ağların eğitimi, sonraki 24 kişi validasyon, son 50 kişi test için kullanılmıştır. Yapılan deneylerde CASIA-B veri setindeki NM, BG ve CL varyasyonlarında şimdiye kadar önerilmiş tüm çalışmalardan daha iyi ve oldukça yüksek doğruluk oranlarına ulaşılmıştır. Çalışmada ayrıca OU-ISIR ile USFHumanID veri setleri üzerinde yapılan deneyler sonucunda da literatürde önerilen çalışmalara göre çok daha başarılı performansa ulaşıldığı görülmüştür.

Derin öğrenme metotlarından herhangi birini kullanmayan ancak elde ettiği sonuçları DeepCNN ile karşıların bir diğer yöntem Ben ve diğerleri (2019) tarafından önerilmiştir. Yapılan çalışmada, iki farklı değişkenler kümesi arasındaki en büyük ilişkiyi bulmaya çalışan Kanonik Korelasyon Analizi (KKA) istatistiksel yönteminden esinlenilerek Eşleştirilmiş Yama Hizalaması adı verdikleri bir yöntem önermişlerdir. Yöntemde amaç, oluşturulan çapraz açılı GEI çiftlerini iz düşürüldükleri ortak alt uzayda aynı sınıfa ait komşuluklar arası mesafeyi minimize ederken, farklı sınıflara ait komşuluklar arası mesafeyi maksimize edebilmektir. Bu amaçla ilk olarak sınıf içi ve sınıflar arası komşulukları barındıran bir GEI çiftler kümesi oluşturulmuştur. Daha sonra her bir çift için sınıf içi bütünlüğü ve sınıflar arası ayrılabilirliği dengeleyebilecek bir amaç fonksiyonu tasarlanmıştır. Daha sonra tüm çiftler kolay çözümlenebilir tek bir amaç fonksiyonu elde edebilmek için bir araya getirilmiştir. Çalışmada CASIA-B, OU-ISIR ve USFHumanID veri setleri olmak üzere toplam üç veri seti üzerinde deneyler yapılmış ve DeepCNN ile performans karşılaştırmasına gidilmiştir. Elde edilen sonuçlarda, CASIA-B ve OU-ISIR veri setlerinde tüm açılar arası başarı oranlarında DeepCNN den daha yüksek, USFHumanID veri setinde ise daha düşük performans elde edildiği sonucuna varılmıştır. Çalışmada görünüm tabanlı varyasyonlar için bir deney sunulmamıştır.

GAN kullanılarak yapılan ilk başarılı çalışmalardan biri olan ve kurulan mimariye GaitGAN adı verilen yöntem Yu ve diğerleri (2017) tarafından önerilmiştir. Bu çalışmada amaç, herhangi bir açılı veya giyim koşulundaki GEI' den, aynı GEI' nin 90° (yan görünüm)'deki normal giyim varyasyonunun GEI' sini elde edebilmektir. Çalışmada standart GAN yapılarından farklı olarak tek bir ayırıcı mimari yerine kişiye ait biyometrik yürüyüş özelliklerinin de elde tutulmasını sağlayacak ikinci bir ayırıcı kurulan ağ yapısına eklenmiştir. Mimarinin üretici (piksel bazında dönüştürücü) ağı yapısında, kaynak

görüntüsünü anlamak için bir kodlayıcı ve hedef görüntüsünü üretebilmek için bir çözücü barındırır. Çözücünün ürettiği görüntüyü sınırlayabilmek için tanımlanan ilk ayırıcı ağ yapısı, kaynak ve hedef görüntülerini çift şeklinde alarak bu görüntülerin aynı kişiye ait olup olmadığına dair ilişkiyi sayısal bir olasılık değeri üreterek belirler. Böylece kişiye ait belirleyici mantıksal özellikler korunmaya çalışılır. İkinci ayırıcı ise standart GAN mimarisi yapısındaki ayırıcılar ile özdeş olup gelen görüntünün sahte mi yoksa orijinal görüntü mü olduğunu belirlemekten sorumludur. Bu iki ayırıcının ürettiği ortak sonuca göre görüntünün etiketi verilir. CASIA-B veri setinde yer alan ilk 62 kişi eğitim, son 62 kişi test veri seti olarak kullanılmasıyla yapılan deneylerde, başta Yu ve diğerleri (2017) tarafından önerilen çalışma olmak üzere, literatürdeki diğer çalışmalarla yapılan özdeş ve çapraz açı bazındaki karşılaştırmalarda tüm açı ve giyim varyasyonlarında benzer performans sonuçları elde edildiği görülmüştür.

He ve diğerleri (2018) tarafından derin öğrenme metotlarından GAN mimarisi kullanılarak ‘MGan’ adı verilen diğer bir yöntemde genel akış şu şekilde sıralanmaktadır: (1) sorgu görüntüsü  $p$ ’ den bir oto kodlayıcı ile açı bazında özellikler kodlanır, (2) bir açı sınıflandırıcı ile hem sorgu görüntüsü  $p$ ’nin hem de galeri görüntüsü  $g$ ’nin açıları tahmin edilir, (3) bir açı dönüştürme katmanı ile açı bazında kodlanmış sorgu görüntüsü  $p$ , galeri görüntüsü  $g$ ’ nin açısına dönüştürülür. Ayrıca tüm galeri görüntülerinin de açı bazında kodlanmış özellikleri hesaplanır, (4) sorgu görüntüsünün hangi galeri görüntüsü ile eşleştiği en yakın komşu sınıflandırıcısı ile belirlenir. Ağ mimarisi temelde 5 ana yapıdan oluşur; kodlayıcı, açı sınıflayıcı, açı dönüştürme katmanı, üretici ve ayırıcı. Çalışmada ayrıca GEI ve CGI özelliklerinin mantığından yola çıkarak zamansal bilgiyi daha az kayıpla elde edebilmek amacıyla PEI yürüyüş deseni tanımlanmıştır. Bu desen birçok kanaldan oluşmakta ve her bir kanalın hesaplanmasında tüm yürüyüş döngüsündeki belirlenmiş genlik değerine sahip çerçevelerin ortalama görüntüsü bulunmaktadır. Üreticinin amacı gerçek PEI’ den ayrılamayacak kadar birbirine yakın bir PEI üretmektir. Açı dönüştürme katmanından elde edilen çıktıların üreticiye verilir. Ayırıcının girdileri ise aynı sayıda kanala sahip üretilmiş yürüyüş görüntüsü ile gerçek yürüyüş görüntüleridir. Ayırıcı ayrıca kişiye özgü yürüyüş karakteristiklerinin korunup korunmadığını da hesaplamasına dahil eder. CASIA-B, OU-ISIR ve USFHumanID veri setleri üzerinde yapılan deneylerde PEI, GEI ve CGI özellikleri bazında denenen, önerilen GAN mimarisi çapraz açı doğruluk oranlarında OU-ISIR veri seti hariç diğer veri setlerinde en yüksek performansı göstermiştir. Ancak bu veri setindeki başarıyı DeepCNN’ den daha yüksektir. CASIA-B veri seti üzerindeki NM

varyasyonunda ise DeepCNN' in daha yüksek performansa ulaştığı görülmektedir. Yöntem ayrıca BG ve CL varyasyonları üzerinde de denenmiş, bazı açılarda tatmin eden sonuçlara ulaşılmıştır.

Özdeş açıda ve çapraz açılarılar arası farkın az olduğu açı varyasyonlarında %100 doğruluk oranına ulaşılmış bir çalışmada (Song vd., 2019) yürüyüş görüntüsü segmentasyonu ile yürüyüş tanıma işlemi tek bir ağda bir arada gerçekleştirilmiş, geliştirilen ağ mimarisine 'GaitNet' adı verilmiştir. Ağ temel olarak üç ana aşamadan oluşmuştur: Bir tam bağlantılı ağ mimarisi ile yürüyüş görüntüsü segment edilmiştir, bir zamansal birleştirme ünitesi ile elde edilen yürüyüş görüntüleri birleştirilmiştir, bu özellikler yürüyüş tanıma ağına verilerek hem özellik çıkarım hem de tanıma ağı ortaklaşa eğitilmiştir. Çalışmada iki ağ yapısının ortaklaşa gerçekleştirilen eğitimi ile ağ, yürüyüş tanıma daha uygun silüetleri çıkarması için zorlanmıştır. Böylece literatürdeki yöntemlerin sıklıkla özellik çıkarımı ile tanıma işlemlerini ayrı ayrı yapmalarından kaynaklanan bilgi kaybı engellenmeye çalışılmıştır. Ağın silüet çıkarma aşaması için kullandığı mantıksal segmentasyon ağı 'Tam Konvüsyonel Ağ (Fully Convolutional Network, FCN) (Long vd., 2015)' daha önce literatürde görüntü segmentasyonu için önerilmiş çok kanallı tam bağlantı katmanlarından oluşan bir ağ yapısıdır. Bu ağ, çalışmada yürüyüş görüntüleri için 'ince ayar' işlemine tabi tutulmuştur. Her bir görüntüden mantıksal segment edilmiş silüetler, sonrasında GEI çıkarma aşamalarına benzer şekilde bir ortalama alma işlemi ile birleştirilmişlerdir. Ağın yürüyüş tanıma aşamasında 4 evrişim ve 2 tam bağlantı katmanından oluşan ve MSCAN adı verilen bir yapı kullanılmış ancak geleneksel CNN mimarilerinden farklı olarak, buradaki evrişim katmanlarında evrişim işlemi farklı ölçeklerde aynı anda uygulanarak, 'çok ölçekli bağlam' da uygulanmıştır. Ağ mimarisindeki her bir alt ağda farklı kayıp fonksiyonları kullanılmış, sonrasında ortaklaşa eğitim işlemi için bir toplam kayıp fonksiyonu da hesaplanmıştır. Ağ CASIA-B veri setinde başarı ölçümü yapmanın yanı sıra, ağın gerçek dünyaya benzer koşullarda da başarısını test edebilmek amacıyla SZU RGB-D (Yu vd., 2013) ile gündelik yaşantıya daha yakın görüntülerden oluşan 'Outdoor-Gait' adındaki kendi veri setleri üzerinde de başarı değerlendirilmiştir yapılmıştır. CASIA-B üzerindeki deneylerde önerilen yöntem, GeiNet' e göre oldukça yüksek, DeepCNN ile benzer doğruluk oranları elde etmiştir. Önerilen yöntem ayrıca 180° hariç, geriye kalan özdeş açıların tümünde %100 doğruluk oranına ulaşırken, bu orana birbirleri arasındaki açı farkının az olduğu çapraz açı varyasyonlarında da ulaşmıştır. SZU RGB-D ile Outdoor-Gait veri setleri üzerinde yapılan başarı değerlendirilmesinde de

önerilen yöntem literatürdeki diğer bazı başarılı yöntemlere göre en performanslı yöntem olmuş, böylece gerçek ortam şartlarına daha uygun bir yöntem olduğu ispatlanmıştır.

Takemura ve diğerleri (2019) tarafından yapılan bir diğer çalışmada DeepCNN deki birbirine paralel CNN yapılarına benzer dört adet derin ağ tasarlanmıştır. Tasarlanan ağlardan ilk ikisinde, görüntüler arası açı farkı fazla ise bu farktan kaynaklanan özne içi uzamsal farklılıklar, özneler arası uzamsal farklılıklara göre daha ayırt edici olacağı savunularak, görüntü çiftinin her bir elemanı ayrı CNN kollarına verilmiştir. Ağların son ikisinde bu durumun tam tersi olacak şekilde, görüntüler arası açı farkı az ise, özneler arası uzamsal farktan yararlanma ön planda tutularak, özneler arası uzamsal farkın, açılar arası yetersiz uzamsal farktan kaynaklanan özne içi farktan daha etkili olacağı savunulmuştur. Bu sebeple görüntü çiftleri aynı anda aynı CNN koluna verilmiştir. Tasarlanan mimarilerin her bir CNN kolunda sırasıyla 2' şer adet ardışık 1 evrişim, 1 normalizasyon ve 1 ortaklama katmanlarından oluşan bloklar ile kanal sayısı kadar 1 adet ortaklama katmanı ve bu ortaklama katmanlarını birleştiren 'L2' normalizasyon katmanları kullanılmıştır. Tasarlanan ilk ağda, birbirine paralel iki CNN kolu bulunur ve ağın kayıp fonksiyonu 'zıt kayıp' fonksiyonudur. İkinci ağda, birbirine paralel üç CNN kolu bulunur ve ağın kayıp fonksiyonu 'Üçlü Sıralama Kayıp Fonksiyonu' dur. Üçüncü ağda, tek CNN kolu, dördüncü ağda ise iki CNN kolu bulunmakta ve bu ağların kayıp fonksiyonları da sırasıyla 'zıt' ve 'üçlü sıralama' fonksiyonlarıdır. OU-ISIR veri seti üzerinde yapılan deneylerde, sınıflar arası açı farkı fazla ise ilk iki mimaride, az ise son iki mimaride daha iyi sonuçlar elde edilerek savunulan görüş desteklenmiştir. Tasarlanan ağların kombinasyonu kullanılarak başta DeepCNN olmak üzere literatürdeki diğer çalışmaların mimarileri ile *rank-1* doğruluk oranı ve Eşit Hata Oranı (Equal Error Rate, EER) hata parametreleri üzerinde yapılan karşılaştırmalarda çok daha başarılı sonuçlar elde edildiği görülmüştür.

DeepCNN ile karşılaştırılmış diğer bir çalışmada (Wang vd., 2020) çok kanallı bir CNN mimarisi tasarlanmış ve bu mimariye girdi olarak yeni bir özellik tanımlanmıştır. Çalışma temel olarak, Üçlü Yürüyüş Silüetleri (Trituple Gait Silhouettes, TTGS) adı verilen yeni bir özelliğin her biri için bir silüet görüntüsü kayarak (son iki silüet görüntüsü ortak olacak şekilde), ardışık üç silüet çerçevesinin hesaplanması ve bu çerçevelerin ayrı ayrı, hazırlanan üç paralel kollu CNN mimarisine verilmesinden oluşmaktadır. Çalışmada GEI özellikleri, uzamsal ve zamansal bilgileri tek bir görüntüde birleştirmelerinden meydana getirebilecekleri kayıplar ile eğitim verisi sayısını azaltmaları yönlerinden TTGS ile birçok karşılaştırmaya tabi tutulmuştur. Önerilen çok kanallı CNN mimarisi, birbirine paralel üç

kolun her birinde 3 evrişim katmanı ve bu katmanlara bağılı 3 ortaklama katmanından oluşan bloklar ile sonrasında bu kolların bir ‘AND’ operatörü ile bir araya getirilerek tekrar 2 evrişim, 2 ortaklama ve 1 tam bağılantı katmanından oluşan kola birleştirilmesinden oluşmaktadır. Önerilen yöntemin performansı başta DeepCNN olmak üzere birçok farklı çalışma ile CASIA-A, CASIA-B ve OU-ISIR veri setleri üzerinde karşılaştırılmıştır. Elde edilen sonuçlarda, DeepCNN ile bazı *rank-1* değerlerinde daha üstün olmak üzere yakın performans elde edildiğı görülmüştür.

Evrişimli-LSTM yapısı kullanan diğeri bir çalışmada (Wang & Yan, 2020) ağı girdi olarak, GEI’ ler üzerinde iyileştirme yapılarak elde edilmiş çerçeve çerçeve GEI (frame-by-frame GEI, ff-GEI) adındaki özellikler kullanılmıştır. Ff-GEI’ lerin elde edilmesinde, GEI çıkarma aşamalarından segmentasyonda, silüet bölgesi çıkarılırken morfolojik operatörler kullanılmıştır. Normalizasyonda silüet bölgelerinin boyutunu değiştirmek yerine, görüntü boyutu değiştirilmiştir. Kombinasyon aşamasında ise yürüyüş periyodunun tamamı yerine her bir yarısına özel bir ortalama işlemi alma fonksiyonu tanımlanmıştır. Böylece, yürüyüş görüntülerindeki bilgiyi koruma ve eğitim veri sayısını artırma adına iyileştirme yapılmıştır. Hazırlana CNN mimarisinde, 3 evrişim katmanı ve her birini takip eden 3 ortaklama katmanı ile 1 tam bağılantı katmanından oluşan yapıya ek bunları takip eden 3 LSTM katmanı kullanılmıştır. Her bir ff-GEI bu yapılara ayrı ayrı verilmiştir. Çalışma içerisinde DeepCNN’in de olduğu literatürdeki diğeri çalışmalarla CCR parametresi üzerinden CASIA-B ve OU-ISIR veri setleri üzerinde karşılaştırma yapmış tatmin edici sonuçlara ulaşıldığı görülmüştür.

Elharrouss ve diğeri (2021) tarafından yapılan bir çalışmada GEI’ ler elde edilmeden önce her bir videodaki hareketi algılamak ve kişiyi segmente edebilmek için bir arka plan çıkarma algoritması önerilmiştir. Önerilen algoritmada, ilk olarak videodaki her bir görüntü çerçevesi bloklara ayrılmış, sonrasında ardışık gelen çerçevelerin bu blokları arasındaki benzerlik değerleri hesaplanarak belirlenen sayıdaki çerçeveler boyunca değişmeyen en benzer bloklar üzerinden arka plan elde edilmiştir. Önerilen algoritma, literatürdeki diğeri arka plan üreten yöntemlerle karşılaştırılmış ve daha iyi performans elde edildiğı görülmüştür. Oluşturulan arka plandan, adaptif belirlenen bir eşik değeri üzerinden ayrılan silüet görüntüleri sonrasında GEI hesaplaması için kullanılmıştır. Çalışmada elde edilen GEI’ ler ilk olarak açılarının belirlenmesi için bir CNN mimarisine verilmiştir. Açısı kesinleşen GEI’ ler daha sonra yürüyüş tanıma işlemi için her bir açığa özel ayrı ayrı eğitilen başka bir CNN mimarisine girdi olarak verilmiştir. Literatürdeki başta model tabanlı

PoseGait ve görünüm tabanlı DeepCNN çalışmaları olmak üzere, farklı yöntemlerle CASIA-B, OU-ISIR ve OU-MVLP (Takemura vd., 2018) veri setleri üzerinde sadece özdeş açı üzerinde yapılan karşılaştırmalarda bazı açılarda daha iyi performansa ulaşıldığı görülmüştür.

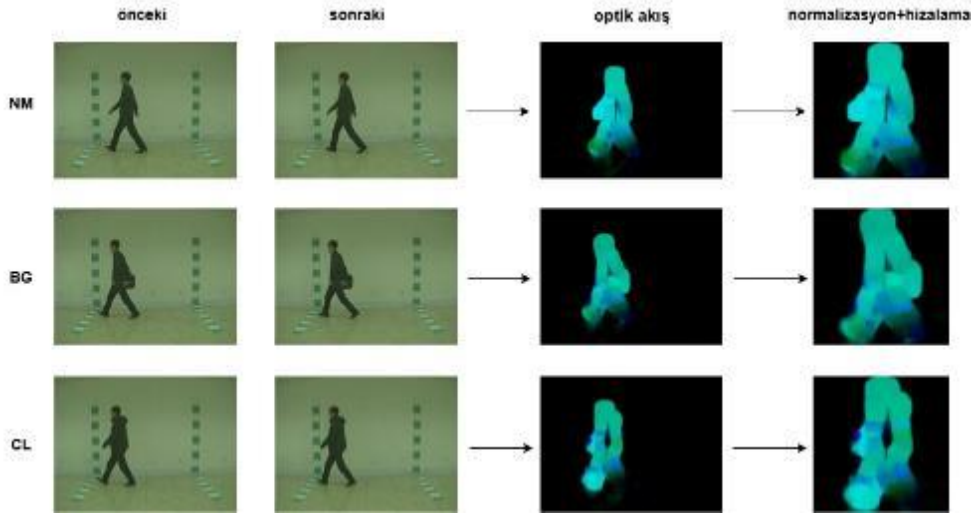
Wang ve Yan (2021) tarafından önerilen diğer bir yöntemde oluşturulan GEI çiftlerinden tamamen lokal olmayan özelliklerin çıkarımı hedeflenmiştir. Çalışmada birbirine paralel iki CNN mimarisi ve bu mimarilerden elde edilen çıktıları birleştirmek için üçüncü bir CNN mimarisi kullanılmıştır. Çalışmanın ilk aşamasında kullanılan veri setinden rastgele seçilen bir kişinin GEI görüntüsü ile bu kişiye ait başka bir GEI görüntüsü kullanılarak pozitif çiftler, farklı kişiye ait bir GEI görüntüsü kullanılarak negatif çiftler oluşturulmuştur. Sonrasında bir pozitif çift '0' etiketi ile bir negatif çift ise '1' etiketiyle olacak şekilde birbirine paralel ağ yapısına ardı sıra verilerek eğitim işlemi gerçekleştirilmiştir. Bu ağ yapısından, tek bir GEI görüntüsü ile iki GEI görüntüsü arasındaki lokal olmayan özellikleri aynı andan elde etme amaçlanmıştır. Kullanılan her bir CNN mimarisi ögesinde DeepCNN mimarilerinden biri olan 'LBNet' yapısı kullanılmıştır. İki CNN mimarisinden gelen çıktılar daha sonra üçüncü CNN mimarisinde %50' lik oranlarla birleştirilmiştir. Son CNN mimarisinden elde edilen özellik vektörü zayıf dinamik bölge, mikro dinamik bölge ve güçlü dinamik bölge olmak üzere vücut bölümlerini temsil edecek şekilde üç parçaya ayrılmıştır. Elde edilen bu son üç parça, üç ayrı 'Softmax' sınıflandırıcısına verilerek ayrı ayrı sınıflandırıldıktan sonra tüm sınıflandırma sonuçlarının bir araya getirilmesinden tek sınıflandırma sonucu elde edilmiştir. CASIA-B ve OU-ISIR veri setleri üzerinde yapılan deneyler sonucunda çapraz açı bazında literatürde daha önce önerilmiş diğer bazı çalışmalara göre tatmin edici sonuçlar elde edilmiştir.

Uzamsal-zamansal özellikleri aynı anda yakalayabilmek ve sadece özdeş açıda tanıma performansını artırmak amacıyla Gul ve diğerleri (2021) tarafından yapılan başka bir çalışmada GEI görüntüleri kurulan 3BCNN mimarisine ağ girdisi olarak verilmiştir. Kurulan ağ mimarisinde 2' şer 3B evrişim katmanını takip eden 1 ortaklama katmanı bloğundan ardışık iki adet ve bu blokları takiben 1 'Batch Normalizasyon' katmanı ile 2 adet tam bağlantı katmanı kullanılmıştır. Çalışmada kurulan ağ yapısının CASIA-B ve OULP veri setleri üzerinde en yüksek başarıyı gösterebilmesi için, bu veri setlerinin her birinin üzerinde Bayes Optimizasyonu uygulanarak ağda kullanılacak optimize fonksiyonu, aktivasyon fonksiyonu, öğrenme oranı değeri ve bırakma katmanı oranı parametreleri belirlenmiştir. Önerilen yöntem literatürde daha önce özdeş açı probleminde 3BCNN mimarisi kullanan

diğer çalışmalar ve CNN mimarisi kullanan Siam Ağ çalışması ile karşılaştırılmıştır. Deneylede, bazı açı değerlerinde bu çalışmalar ile yakın sonuçlar elde edilirken, bazı açı değerlerinde daha başarılı performans elde edilmiştir.

### 1.2.3. Optik Akış ile Yürüyüş Örüntüsü

Optik akış, bir görüntüdeki piksellerin zaman içindeki hareketini modelleyen bir tekniktir. Optik akış görüntüsü ise video çerçeveleri arasındaki hareketi yansıtır. Bu özelliği yürüyüş tanıma sürecini kolaylaştırır. Ancak GEI ve silüetlere benzer olarak yürüyüş tanımda insan vücudu için küresel bir yaklaşım sunar. Bu nedenle görüş açısı farkından ve görünüme dayalı varyasyonlardan etkilenmesi muhtemeldir. Yürüyüşe ait bir optik akış görüntüsü, iki ardışık RGB yürüyüş çerçevesi arasında hesaplanarak elde edilir. Bu tez kapsamında optik akış hesabı için ‘Farneback (Farneback, 2003)’ yöntemi kullanılmıştır. Hesaplanan optik akış çerçeveleri, silüet çerçevelerindeki gibi Takemura ve diğerleri (2019) yöntemine göre ölçek normalizasyonu ve hizalama işlemine tabi tutulmuştur. Ancak silüetlerden farklı olarak optik akışın baş ve ayak kısmının başlangıç ve bitiş nokta pikselleri hesabına 4 piksel ekleme yapılmıştır. Yürüyüşe ait bir optik akış görüntüsü hesaplama süreci Şekil 6’ da sunulmuştur.



Şekil 6. Optik akış hesaplama süreci

Farneback yöntemi, görüntüdeki piksellerin zaman içindeki hareketini her piksel için tahmin eder. Bu nedenle “dense” yani yoğun optik akış yöntemidir. Farneback’ in temel fikri, görüntüdeki küçük bölgeleri ikinci dereceden bir polinom ile modelleyerek ve bu

polinomun iki görüntü arasında nasıl kaydığını analiz ederek, piksellerin hareketini bulabilmektir. Yani, her görüntü parçası bir fonksiyonla yaklaşık olarak şöyle ifade edilir,

$$f(x) \approx x^T A x + b^T x + c \quad (2)$$

burada,  $x = [x, y]^T$ : piksel koordinatları,  $A$ :  $2 \times 2$  simetrik matris,  $b$ :  $2 \times 1$  vektör,  $c$ : sabit terimdir. Yöntemin hesaplama adımları sırasıyla aşağıdaki gibi verilebilir:

- Her bir küçük pencere (örneğin  $5 \times 5$ ) içindeki yoğunluk değerleri, yukarıdaki polinom ile yaklaşık olarak modellenir. Bu işleme “polynomial expansion” denir. Yani her pencerenin etrafındaki parlaklık değişimi, bir polinom ile ifade edilir. Bu, o bölgenin yüzey şekline benzer bir bilgi verir.
- İkinci görüntü (örneğin bir sonraki kare) aynı şekilde modellendiğinde, birinci ve ikinci görüntüdeki polinomlar arasındaki fark, hareketin (optik akışın) ne olduğunu verir.

$$f_2(x) = f_1(x - d) \quad (3)$$

bu denklem, ikinci görüntüdeki değerler birinci görüntüdeki aynı yapının  $d$  kadar kaymış hali demektir.

- Matematiksel olarak iki polinom arasındaki kayma  $d$  şu formülle hesaplanır;

$$d = -\frac{1}{2} A^{-1} b \quad (4)$$

burada,  $A$ : ilk görüntüdeki polinomun ikinci dereceden kısmı,  $b$ : İki polinom arasındaki birinci dereceden terim farkıdır ve bu formül doğrudan hareketi verir.

- Hesaplanan optik akışın gürültü içerebilmesi sebebiyle Farneback, hareket vektörlerini Gaussian pencereleri ile ağırlıklı olarak ortalar. Böylece daha düzgün bir akış haritası elde edilir.
- Son olarak büyük hareketlerin tahmini için görüntü çözünürlüğü düşürülerek katmanlı (piramit) bir yaklaşım uygulanır (Farneback, 2003).

Optik akış özelliği kullanılarak yapılan ilk derin öğrenme tabanlı bir çalışma Wolf ve diğerleri (2016) tarafından sunulmuştur. Çalışmada yeni bir özellik önerilmiş, bu özelliğin

ilk kanalı, video çerçevelerinin gri seviyeli görüntüsü, ikinci ve üçüncü kanalı ise  $x$  ve  $y$  yönünde hesaplanan optik akışlar kullanılarak hesaplanmıştır. Ayrıca veri setlerindeki video görüntülerinin çerçeveleri 16'şarlı klipler haline getirilmiştir. Bu özellikler ile içerisinde yedi adet 3B konvolüsyon katmanı barındıran bir 3BCNN mimarisi eğitilmiştir. Eğitim ve test aşamaları sadece özdeş açılar için gerçekleştirilmiş ve kullanılan veri setlerinden en bilinen CASIA-B veri setinde en yüksek  $90^\circ$  açısında %99.9 doğruluk oranı elde edilmiştir. Benzer bir çalışma Castro ve diğerleri (2017) tarafından yapılmıştır. Çalışmada  $x$  ve  $y$  yönündeki optik akışlar farklı sayılarda 2BCNN katmanları, 3BCNN katmanları veya artık bloklar içeren 4 farklı CNN modeline verilmiştir. TUM-GAIT veri seti üzerinde yapılan deneylerde artık blok içeren CNN mimarilerinde daha yüksek performansa erişildiği görülmüştür.

Jimenez ve diğerlerine (2021) ait bir çalışmada, yürüyüş tanımda video çerçevelerinden tek yararlı özellik çıkarmak yerine birçok özellik çıkarıp bu özellikleri farklı kombinasyonlar halinde kullanmak amaçlanmıştır. Çalışmada, video çerçevelerinden kullanılan veri setinin izin verdiği ölçüde; optik akış, gri seviye, derinlik ve silüet olmak üzere 4 farklı özellik çıkarılarak ayrı kollarla farklı CNN mimarilerinde işlenmiştir. Kurulan ağ yapısının her bir kolunun çıkışında bulunan 'kapı' yapıları sayesinde ilgili özelliğin veri setinden elde edilememesi durumunda, ilgili koldan gelmeyecek özellikler ağa dahil edilmeyerek hataya düşme engellenmiştir. Elde edilen tüm özellikler eşit boyutlandırılıp bir 'birleştirme' fonksiyonu ile bir araya getirildikten sonra ağın ilerleyen aşamalarındaki tam bağlantı katmanlarına iletilmişlerdir. Kurulan ağın her bir özellik için ayrı oluşturulan CNN kollarında başta GaitSet olmak üzere, literatürde önerilmiş farklı başarılı CNN mimarileri kullanılmıştır. Ağın eğitiminde, giyim ve açı varyasyonlarından etkilenmeden, aynı kişilere ait özellikler arası mesafeyi minimum, farklı kişilere ait özellikler arası mesafeyi maksimum yapabilmesi için üçlü kayıp fonksiyonu kullanılmıştır. Ayrıca test aşamasında galeri ve sorgu görüntüleri arası benzerlik  $k$ -NN sınıflandırıcısı ile hesaplanmıştır. TUM-GAIT ve CASIA-B veri setleri üzerinde yapılan deneylerde ağdaki kapı yapıları sayesinde her bir veri setinden farklı özellikler tekli veya 2'li, 3'lü kombinasyonlar şeklinde kullanılarak en iyi ağ girdi şekilleri araştırılmıştır. Yapılan deneylerde, TUM-GAIT ve CASIA-B veri setinin özdeş açı değerlendirmelerinde, literatürdeki önerilmiş diğer başarılı çalışmalardan daha iyi performans elde edilmişken, CASIA-B veri setinin çapraz açı NM, BG ve CL varyasyonlarında bazı çalışmaların performansının altında sonuçlar elde edilmiştir.

### 1.3. Makine Öğrenme İle Yürüyüş Tanıma

Yapay zekâ alanındaki hızlı gelişmeler, derin öğrenme modellerinin çeşitli uygulama alanlarında üstün performans sergilemesine olanak sağlamıştır. Bu doğrultuda, yürüyüş tanıma alanında da son yıllarda derin öğrenmeye dayalı yöntemlerin kullanımı artmış ve önemli ilerlemeler kaydedilmiştir. Bununla birlikte, derin öğrenme yöntemlerinin yaygınlık kazanmasından önce, yürüyüş tanıma problemleri ağırlıklı olarak makine öğrenmesi yaklaşımlarıyla ele alınmıştır. Makine öğrenmesine dayalı bir yürüyüş tanıma sistemi ise temelde iki ana aşamadan oluşur: Öznitelik temsili ve sınıflandırma. Herhangi bir makine öğrenmesine dayalı tanıma sisteminin başarımı, büyük ölçüde uygun ve etkili bir öznitelik temsil uzayının bulunmasına bağlıdır. Bu alanda ileri düzey öznitelik temsil teknikleri arasında boyut indirgeme yöntemlerinden PCA ve Lineer Ayırıcı Analiz (Linear Discriminant Analysis, LDA) ile bu iki yöntemin birleşiminin makine öğrenmesine dayalı yürüyüş tanıma problemini ele alma konusundaki etkinlikleri öne çıkmış ve yaygın olarak kullanılmışlardır. PCA, orijinal veri ile projeksiyonu alınmış veri arasındaki hataların en aza indirilmesi şeklinde formüle edilen, denetimsiz bir doğrusal boyut indirgeme tekniğidir. LDA ise veriyi farklı sınıfların birbirinden iyi bir şekilde ayrılabilirdiği daha düşük boyutlu bir alt uzaya yansıtmayı amaçlayan denetimli bir tekniktir. Başka bir ifadeyle LDA, sınıf içi varyasyonları en aza indirirken sınıflar arası varyasyonları en üst düzeye çıkarmayı hedefler. Sınıflandırma aşamasında ise En Yakın Komşu (EYK) ve Destek Vektör Makineleri (DVM) yöntemlerinin yaygın ve başarılı bir şekilde kullanıldığı görülmektedir. EYK, eğitim verisinin dağılımı hakkında önceden herhangi bir varsayımda bulunmaz ve yeni bir örnek, kendisine en yakın mesafeye sahip eğitim örneğinin sınıf etiketine atanır. DVM ise bir düzlem üzerindeki verileri en iyi şekilde ayıracak bir karar sınırı bulmayı amaçlayan ve bu sınırın, sınıflara ait örneklerden maksimum uzaklıkta (maksimum marj) olmasını hedefleyen denetimli bir öğrenme yöntemidir.

Makine öğrenmeye dayalı yürüyüş tanıma yöntemlerinin model tabanlı yaklaşımlar kategorisi temelde bir model uydurmaya dayanır. Bu kapsamda, Lee ve Grimson (2002), siluet bölgelerine uydurulan elipslerin parametrelerini analiz ederek yürüyüşün görünüşünden ve dinamik değişimlerinden yararlanmışlardır. Çalışmada sınıflandırma aşaması için DVM kullanılmıştır. Dockstader ve diğerleri (2003), çeşitli eklem açılarını çıkarmak için bir 3B model tanımlamış ve bu modelleri EYK yöntemi ile sınıflandırmışlardır. EYK ile sınıflandırma yapan diğer bir çalışmada (Wang vd., 2004)

insan vücudu, eklemlerle birbirine bağlanmış on dört sabit parçadan oluşan bir yapı olarak modellenmiştir. Zhang ve diğerleri (2007), eklem pozisyon izleklerinin çıkarılmasına olanak tanıyan beş bağlantılı iki ayaklı bir insan hareket modeli önermiş, sınıflandırma aşaması için HMM kullanmışlardır. Ariyanto ve Nixon (2012), kukla ve kütle-yay modeline dayalı yeni bir üç boyutlu model yaklaşımı tanıtmış, sınıflandırma aşaması için EYK kullanmışlardır. Tafazzoli ve Safabakhsh (2010), eklemlerle vücut parçalarının hareketlerini modellemek için aktif kontur modelleri ile Hough dönüşümünü birleştiren bir yöntem sunmuşlardır.

Makine öğrenmeye dayalı model tabanlı yaklaşımlarda başarılı sonuçlara ulaşılsa da modellerin karmaşıklığı ve bileşenlerinin video akışından çıkarılması oldukça maliyetli hesaplamalar gerektirmiştir. Bu nedenle görünüm tabanlı yaklaşımların daha geniş bir kullanım alanına sahip olmasına sebep olmuşlardır.

Makine öğrenmeye dayalı görünüm tabanlı yaklaşımlara yönelik yapılan ilk araştırmalarda daha çok silüetlere ve konturlara dayalı öznitelikler üzerine yoğunlaşmıştır. Kale ve diğerleri (2004), ikili silüetlerin dış kontur genişliğinden ve bütün ikili silüetlerden faydalanırken, Liu ve Sarkar (2006) manuel olarak çıkarılmış silüetler ile HMM eğitmişlerdir. Lam ve diğerleri (2007), hem dinamik hem de statik bilgileri çıkarmak amacıyla Statik Silüet Şablonları temsili sunarken, Choudhury ve Tjahjadi (2012) silüet konturlarından öznitelikleri Procrustes şekil analiziyle çıkarmıştır. Diğer taraftan, bazı çalışmalar, kontur ve silüetlerden çıkarılan öznitelikler için denetimli ve denetimsiz öğrenme tekniklerine dayalı olarak etkili ve uygun temsil uzayları bulmayı amaçlamıştır. Lu ve Zhang (2007), silüetlerden ayırt edici öznitelikleri Fourier ve Dalgacık (Wavelet) tanımlayıcıları kullanarak çıkarmış, tanıma aşamasında Bağımsız Bileşenler Analizi (BBA) ve Genetik Bulanık Destek Vektör Makineleri (GBDVM) sınıflandırıcılarını kullanmışlardır. Lu ve diğerleri (2008) ise çıkarılan silüetlere Çok Doğrusal Temel Bileşenler Analizi (ÇDTBA) ve LDA uygulamıştır.

Kontur ve silüet tabanlı özniteliklerin iyi sonuçlar vermesine rağmen, Han ve Bhanu (2005) tarafından önerilen GEI temsili, daha ayırt edici bir yöntem olarak kabul görmüştür. Önerildiği çalışmada tanıma süreci için başlangıçta Kanonik Ayırıcı Analiz (KAA) ile birlikte kullanılmıştır. Bu yöntem PCA ve ardından LDA uygulanarak elde edilen özniteliklerin EYK algoritması ile sınıflandırılması ile gerçekleştirilmiştir. Bu aşamadan sonra literatürde çok sayıda çalışma, GEI tabanlı sınıflandırma için uygun ve etkili bir öznitelik temsil uzayı bulmaya çalışmıştır. Tao ve diğerleri (2007), GEI şablonlarından en ayırt edici öznitelikleri çıkarmak için Gabor filtrelerini kullanmış, boyut azaltma için PCA

tekniki yerine Genel Tensor Ayrımsal Analizi (GTAA) uygulayarak EYK ile sınıflandırma işlemini gerçekleştirmiştir. Lishani ve diğerleri (2014), GEI'lerden Haralick özneliklerini çıkarmış ve bunları doğrusal olmayan bir DVM sınıflandırıcısına girdi olarak vermiştir. DVM kullanan başka bir çalışmada Lishani ve diğerleri (2016), GEI'lerden bir Gabor filtre bankası aracılığıyla öznelikler çıkarmış ve bu öznelikleri Spektral Regresyon Kernel Ayrımsal Analizi (SRKAA) ile birleştirmiştir. Benzer bir çalışma Wang ve diğerleri (2017) tarafından GEI Gabor Dalgacık özneliklerini 2BPCA kullanarak yeni bir alt uzaya yansıtılmasıyla sunulmuştur. Ben ve diğerleri ise (2016), düşük çözünürlüklü GEI tabanlı tanıma için yeni bir Doğrusal Olmayan Eşlenik Haritalama (DOEH) yöntemi önermişlerdir. Bu çalışmalara ek olarak, literatürde görünüm tabanlı varyasyonlarda ciddi oranda düşen tanıma performansını artırmak için bu varyasyonlar altındaki GEI'nin en ayırt edici yürüyüş vücut bölgelerini seçerek (Iwashita vd., 2013; Nandy vd., 2016; Lishani vd., 2017) veya bu işlemi makine öğrenmesine dayanarak (Rokanujjaman vd., 2015; Rida vd., 2015; 2016; Alotaibi & Mahmood, 2017; Isaac vd., 2017; Ghebleh & Ebrahimi Moghaddam, 2018) tanıma gerçekleştiren birçok yöntem sunulmuştur.

GEI üzerine yapılmış bahsi geçen başarılı yürüyüş tanıma yöntemlerine dayanarak literatürde GEI'ye benzer MSI, SVB, FDEI, CGI, GEnI, GFI, PEI gibi birçok temsil ve çalışma (Ekinci & Gedikli, 2005; Ekinci vd., 2007) yürüyüş çerçeveleri dizisini tek bir şablona dönüştürme işlemi için farklı teknikler kullanmıştır.

Makine öğrenmesine dayalı görünüm tabanlı yaklaşımlar çerçevesinde, farklı görüş açıları varyasyonuna karşı dayanıklı öznelik temsilleri elde etmeye yönelik çeşitli yöntemler geliştirilmiştir. Bu yaklaşımlardan biri, farklı görüş açılarından çıkarılan öznelikleri ortak bir görüş düzlemine dönüştürmeyi amaçlayan Görüş Dönüşüm Modelleri'dir (GDM). Bu bağlamda literatürde farklı dönüşüm modelleri önerilmiş olup, özellikle Tekil Değer Ayrışımı (TDA) (Makihara vd., 2006; Kusakunniran vd., 2009; Zheng vd., 2011; Muramatsu vd., 2011; Muramatsu vd., 2015) temelli yöntemler öne çıkan yaklaşımlar arasında yer almıştır. Yaklaşımlardan bir diğeri ise orijinal yürüyüş özneliklerini bir alt uzaya (genellikle daha düşük boyutlu) dönüştürerek görüş açıları farkına dayanıklı öznelikler elde etmeyi amaçlamıştır. Bu bağlamda, Bashir ve diğerleri (2010) KKA'dan faydalanırken Liu ve diğerleri (2011), KKA ve PCA ile ortak bir alt uzay öğrenme yöntemi önermişlerdir. Önerdikleri diğeri çalışmada (2013), alt uzaylardan daha verimli biçimde ayırt edici bilgilerin elde edilmesini sağlamak için Marjinal Kanonik Korelasyon Analizi (MKAA) uygulamışlardır. Al Mansur ve diğerleri (2014; 2015) ise

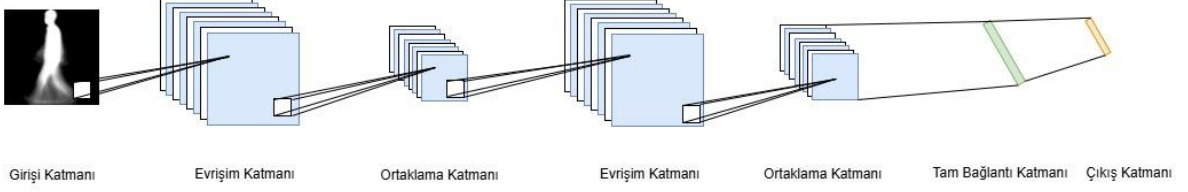
Çoklu-Açı Ayırıcı Analizi (Multi-view Discriminative Analysis, MvDA) ile bu analizin tensör versiyonunu kullanmışlardır.

Yapay zekâ alanındaki gelişmeler ile makine öğrenmesine dayalı yöntemlerin özellik çıkarımı, temsil öğrenimi ve sınıflandırma süreçlerini kapsayan uçtan uca (end-to-end) tanıma sistemleri geliştirmeyi amaçlayan Derin Öğrenme temelli çalışmaların sayısında belirgin bir artış gözlenmiştir. Bu yaklaşımlar, ham verileri doğrudan giriş olarak kullanan ve geleneksel Yapay Sinir Ağları'nda yaygın olan iki katmandan daha derin ağ yapıları oluşturan yöntemlerle ilerlemektedir. Bu yaklaşımlarda düşük seviyeli her katman, verinin belirli özelliklerini ilkel düzeyde kodlamakta ve bu ilkel yapılar, ardışık olarak gelen yüksek seviyeli katmanlar tarafından aşamalı şekilde birleştirilerek giriş verisinin daha temsil gücü yüksek ve ayırt edici gösterimlerine ulaşmaktadır. Bu durum, performansın önemli ölçüde artmasını sağlamaktadır. Ancak, derin öğrenme modellerin etkin şekilde çalışabilmesi için özellikle eğitim aşamasında geniş veri setlerine ihtiyaç duyulmaktadır. Ayrıca, bu modellerden yüksek performans elde edilebilmesi için karmaşık yapılar ve üst düzey hesaplama gücü gerekmektedir (Rida vd., 2019).

#### **1.4. Derin Öğrenme Modelleri ve Transfer Öğrenme**

Son zamanlarda yapay zekâ alanındaki hızlı gelişmeler derin öğrenme modelleri sayesinde gerçekleşmektedir. Grafik kartlarının hesaplama gücü ile yüksek hesaplama gerektiren derin öğrenme ağları kabul edilebilir sürede başarılı sonuçlar üretmiştir. Bu bölümde genel derin ağ modelleri sunulmuştur.

2012 ImageNet Görüntü Sınıflandırma Yarışması'nda AlexNet (Krizhevsky vd., 2012) ağının üstün bir performans göstererek yarışmada birinci olmasıyla birlikte derin öğrenme teknolojisine olan ilgi artış göstermiştir. Geleneksel makine öğrenme yöntemleriyle karşılaştırıldığında, derin öğrenme modelleri manuel öznitelik çıkarma işlemi yerine verilerden direkt olarak en faydalı öznitelikleri otomatik olarak çıkarmıştır. En yaygın kullanılan derin öğrenme modellerinden biri olan CNN' ler özellikle görüntüler üzerinde çalışmak ve farklı sayıda parametreyle uzaysal zamansal özellikler çıkararak dikkate değer bir başarı elde etmek amacıyla tasarlanmışlardır. Böylece, bugüne kadar görüntü işleme alanında mükemmel performans sergilemişlerdir. Genel bir CNN mimarisi Şekil 7' de gösterilmiştir.



Şekil 7. Bir CNN mimarisi genel yapısı

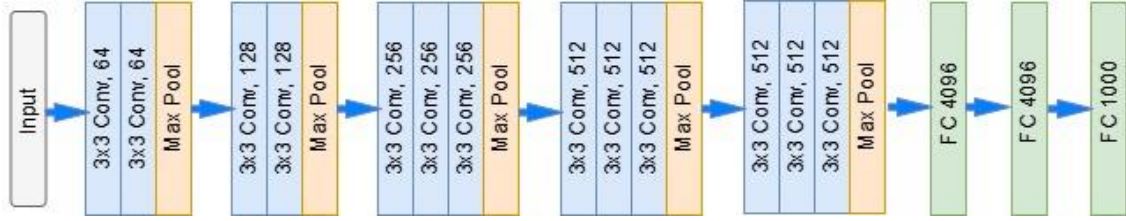
Bir CNN tipik olarak giriş katmanları, evrişim katmanları, ortaklama katmanları ve tam bağlantı katmanlarından oluşur. 2B veya 3B görüntüler giriş olarak verildiğinde, evrişim katmanları görüntülerden özellikler çıkarırken, ortaklama katmanları boyutu, veriyi ve parametre sayısını azaltır. Ardışık bir dizi evrişim ve ortaklama işlemleri ile CNN'ler aynı anda hem düşük seviyeli hem de yüksek seviyeli özellikleri öğrenir. Tam bağlantılı katmanlarında ise ağ, sınıflandırma işlemini kolaylaştıran en ayırt edici özellikleri elde eder.

Literatürde şimdiye kadar birçok farklı veri setleri üzerinde başarısı ispatlanmış derin öğrenme modelleri önerilmiştir. Sonraki bölümlerde bu tez kapsamında kullanılan CNN mimarileri — VGG16, ResNet50, MobileNet, EfficientNet ve ConvNeXt — ile transfer öğrenme işlemi detaylandırılmıştır.

#### 1.4.1. VGG16

VGG, Simonyan ve Zissermann (2014) tarafından önerilen derin bir evrişimli sinir ağıdır. VGG adı, Oxford Üniversitesi Görsel Geometri Grubu'nun kısaltmasıdır. Bir VGG modeli büyük ölçekli görüntülerle işlem yapabilmek için küçük ( $3 \times 3$ ) evrişimli filtre boyutlu katmanların derinliğini araştırır. VGG16 ağı ise ağırlıkları olan 16 katmandan oluşur.

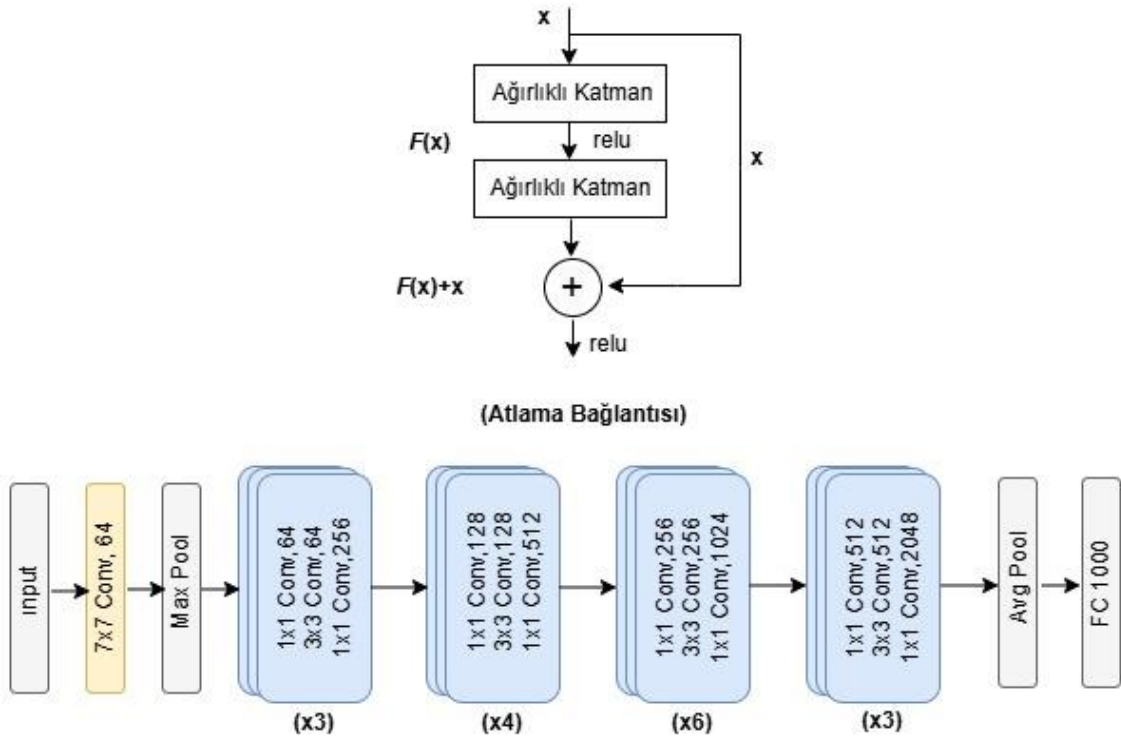
VGG16 ağı, 13 evrişimli katman, 5 maksimum ortaklama katmanı ve 3 tam bağlantı katmanından oluşur. Evrişimli ve tam bağlantı katmanlarından dolayı ayarlanabilir parametrelere sahip katman sayısı 16'dır. İlk bloktaki filtre sayısı 64'tür. Daha sonra bu sayı 512'ye ulaşana kadar sonraki bloklarda iki katına çıkar. Giriş tensör boyutu 3 RGB kanallı  $224 \times 224$  boyutudur. Evrişim ve maksimum ortaklama katmanları tüm mimari boyunca ardışık bir şekilde düzenlenmiştir. Evrişimli katman bloklarını 3 tam bağlantı katmanı takip eder. İlk iki tam bağlantı katmanının nöron sayıları 4096'dır. Çıktı katmanı, Imagenet veri setinin (Russakovsky vd., 2015) sınıf sayısına karşılık gelen 1000 nöronudur. Şekil 8' de VGG16 ağının genel yapısı sunulmuştur.



Şekil 8. VGG16 ağı genel mimari yapısı

#### 1.4.2. ResNet50

ResNet (He vd., 2016) mimarisi derin ağların eğitimi sırasında meydana gelen ve sık karşılaşılan bir problem olan ‘kaybolan gradyan’ problemine çözüm üretmek amacıyla geliştirilmiştir. Kaybolan gradyan problemi, geleneksel düz derin öğrenme mimarilerinin eğitimlerinde geri yayılım esnasında, küçük gradyan değerlerin sürekli olarak daha küçük gradyanlar ile çarpılmasının sonucunda ortaya çıkan yeni gradyan değerlerinin tamamen küçülüp ağırlıklarda neredeyse hiç değişiklik yapamayacak hale gelmesiyle ortaya çıkar. ResNet mimarisi bu problemi ‘atlama bağlantıları’ veya diğer adıyla ‘kimlik eşleme’ yapısı ile çözmeye çalışmıştır. Atlama bağlantısı ve ResNet50 mimarisi Şekil 9’ da gösterilmiştir.



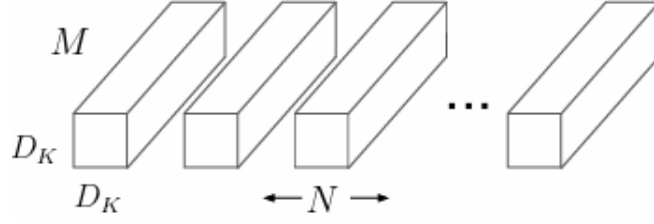
Şekil 9. Resnet50 mimarisi ve atlama bağlantısı yapısı

Atlama bağlantılarında amaç girdi ve çıktı arasındaki ilişkiyi doğrudan öğrenmek yerine artık fonksiyonları öğrenebilmektir. ResNet50, ResNet mimarileri içerisinde en bilinen ve en sık kullanılan mimaridir. ResNet50 mimarisi  $3 \times 224 \times 224$  boyutunda görüntüler alır ve 50 katmandan oluşur. Yapısında 16 atlama bloğu vardır ve bu bloklardan sırasıyla ilk üçlü: 64, 64, 256; sonraki dördü: 128, 128, 512; sonraki altı: 256, 256, 1024; son üçlü: 512, 512, 2048 filtre sayılı ve tamamı sırasıyla:  $1 \times 1$ ,  $3 \times 3$ ,  $1 \times 1$  filtre boyutlarına sahip evrişim katmanlarından oluşmaktadır. Mimari ayrıca ortaklama katmanları, tam bağlantı katmanı ve sınıflandırma için bir 'softmax' çıktı katmanı içerir. Yapısındaki atlama bağlantıları sayesinde iki Düzeltmeli Doğrusal Ünite (Rectified Linear Unit, ReLU) aktivasyon birimi arasında bulunan evrişim blokları atlanır ve ilk ReLU'nun çıkışı ( $x$ ) doğrudan ikinci ReLU'ya ( $F(x)+x$ ) beslenir. Bu yapı çok derin ağlarda avantaj sağlar.

### 1.4.3. MobileNet

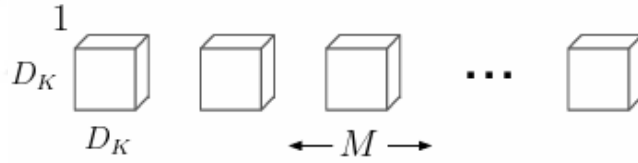
MobileNet'ler (Howard vd., 2017), sınırlı kaynakları göz önünde bulundurarak doğruluğu etkili bir şekilde en üst düzeye çıkarmak için tasarlanmış küçük, düşük gecikme süreli ve düşük güç tüketimli modellerdir. Düşük model boyutu sayesinde mobil cihazlar gibi sınırlı donanım kaynaklarına sahip ortamlarda derin öğrenme tabanlı uygulamalarda kullanılmaya elverişlidirler. Standart CNN mimarilerinden en büyük farkları standart evrişimli katmanlar yerine derinlik bazlı ayrılabilir evrişimli katmanlar kullanmalarıdır. Derinlik bazlı ayrılabilir evrişimler ile geleneksel evrişimler arasındaki temel fark, tek  $3 \times 3$  evrişim katmanı yerine  $3 \times 3$  derinlik bazlı ve  $1 \times 1$  noktasal bazlı evrişim olmak üzere iki ayrı katman kullanmalarıdır. Bu yapı, hem parametre sayısını azaltır hem de eğitim süresini kısaltarak pratik uygulamalarda verimliliği artırır. Temel olarak ilk katman giriş kanallarını filtrelemek için kullanılırken ikinci katman bunları birleştirerek yeni bir özellik oluşturmak için kullanılır.

MobileNet mimarisinde derinlik bazlı evrişim, her giriş kanalına yalnızca bir filtre uygular. Ardından noktasal bazlı evrişim, derinlik bazlı evrişimden elde edilen çıktıları birleştirmek için  $1 \times 1$  evrişim uygular. Standart bir evrişim işlemi ise filtreleme ve birleştirme işlemlerini tek bir adımda gerçekleştirir. Derinlik bazlı ayrılabilir evrişimin bu iki işlemi iki ayrı katmana ayırma yani faktörizasyon işlemi, hesaplama maliyetini ve model boyutunu önemli ölçüde azaltma etkisi sağlar. Şekil 10' da standart evrişim filtreleri gösterilmiştir.



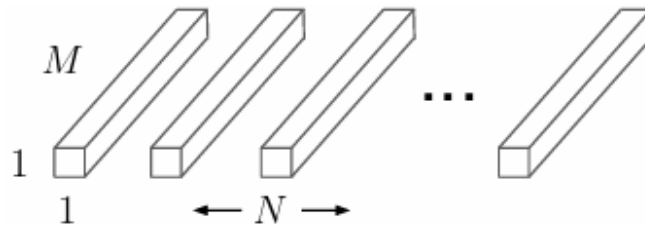
Şekil 10. Standart evrişim filtreleri (Howard vd., 2017)

$D_F \times D_F \times M$  boyutlu bir  $F$  özellik haritasını girdi olarak alan bir standart evrişim katmanının hesaplama maliyeti  $D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F$  şeklinde hesaplanır. Burada  $D_F$ ,  $F$  özellik haritasının boyutları,  $D_K$  evrişim filtre boyutu,  $M$  ve  $N$  sırasıyla girdi ve çıktı kanal boyutlarıdır. Standart bir evrişim için hesaplama maliyeti, giriş ve çıkış kanallarının sayısına ve giriş özellik haritasının ve evrişim filtresinin uzaysal boyutlarına çarpımsal olarak bağlıdır. Şekil 11’ de ise derinlik bazlı evrişim filtreleri sunulmuştur. Bir derinlik bazlı evrişim katmanının hesaplama maliyeti  $D_K \cdot D_K \cdot M \cdot D_F \cdot D_F$  şeklinde hesaplanır. Filtrelerin kanal boyutu  $M$ , 1 (bir)’ e eşittir.



Şekil 11. Derinlik bazlı evrişim filtreleri (Howard vd., 2017)

Derinlik bazlı evrişim yalnızca giriş kanalını filtrelemek için kullanıldığından, bunları yeni özellikler üretmek için birleştiremez. Bu nedenle,  $1 \times 1$  evrişim kullanarak derinlik bazlı evrişim katmanının çıktısının doğrusal bir kombinasyonunu hesaplayan noktasal bazlı evrişim katmanı adı verilen ek bir katman gereklidir. Noktasal bazlı evrişim filtresi Şekil 12’ de sunulmuştur.

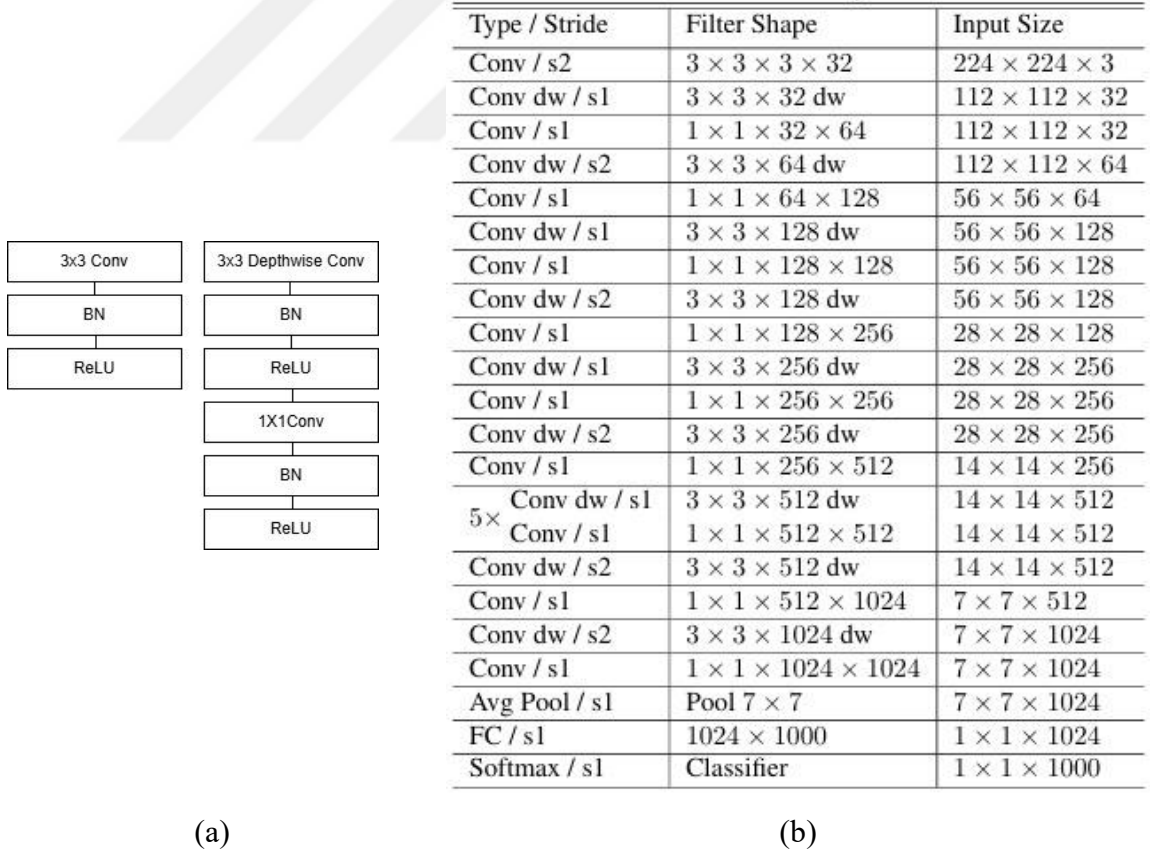


Şekil 12. Noktasal bazlı evrişim filtreleri (Howard vd., 2017)

Noktasal bazlı evrişim katmanının hesaplama maliyeti  $M \cdot N \cdot D_F \cdot D_F$  şeklinde hesaplanır. Sonuç olarak derinlik bazlı ayrılabilir evrişimlerin toplam hesaplama maliyeti  $D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F$  şeklinde hesaplanacaktır. Bunu standart evrişimin hesaplama maliyetiyle karşılaştırdığımızda, aşağıdaki şekilde ifade edilebilecek bir hesaplama azaltımı elde ederiz:

$$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_K^2} \quad (5)$$

Buradan, standart evrişimlere kıyasla yalnızca çok az bir doğruluk kaybıyla 8 ile 9 kat daha az hesaplama gerektiren  $3 \times 3$  derinlemesine ayrılabilir evrişimler kullanıldığı, ya da bir başka deyişle standart evrişimin, derinlik bazlı evrişime göre 8 ile 9 kat daha fazla çarpma sayısına sahip olduğu sonucuna varmaktayız (Howard vd., 2017). MobileNet yapısının ayrıntılı tanımı Şekil 13' te sunulmuştur.

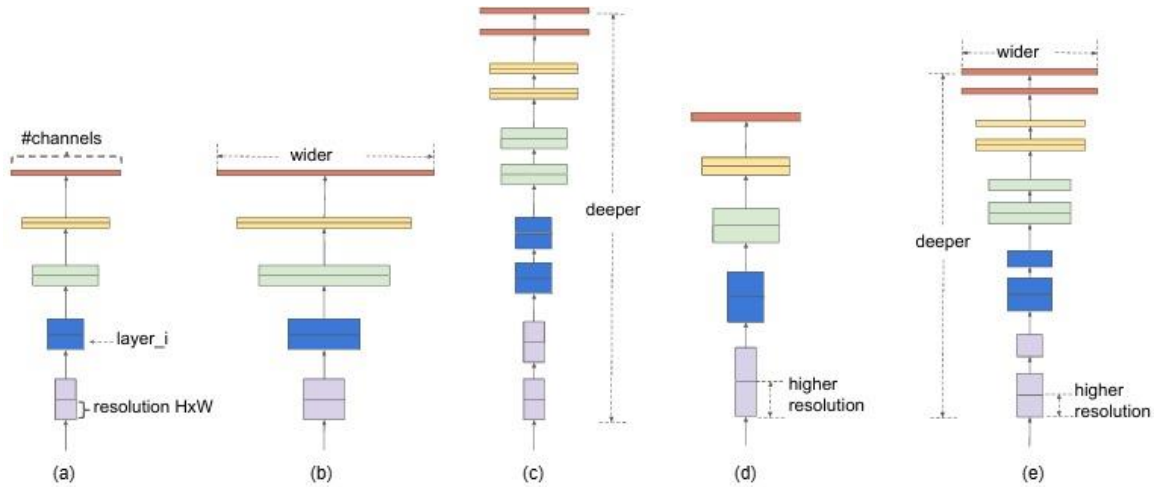


Şekil 13. (a) Standart ve derinlik bazlı ayrılabilir evrişimli katman, (b) MobileNet yapısı

Şekil 13 incelendiğinde MobileNet mimarisi, ilk katman dışında Şekil 13 (a)'da sağda gösterilen derinlik bazlı ayrılabilir evrişimlerden oluşacak şekilde tasarlanırken, ilk katman solda gösterilen standart tam evrişim olarak uygulanmıştır. Tüm katmanlar, istisnasız şekilde toplu (batch) normalizasyon işlemi ve doğrusal olmayan ReLU (Düzeltilmeli Doğrusal Ünite) aktivasyon fonksiyonu ile takip edilir. Yalnızca son tam bağlantı katmanı bu sürece dahil değildir. Bu katmanda herhangi bir aktivasyon fonksiyonu bulunmaz ve çıktısı doğrudan sınıflandırma işlemi için kullanılan softmax katmanına aktarılır. Son aşamada, tam bağlantılı katmandan önce ortalama ortaklama uygulanarak uzamsal boyut 1 (bir)'e düşürülür. Derinlik bazlı ve noktasal bazlı evrişimleri ayrı katmanlar olarak sayıldığında MobileNet yapısı toplam 28 katmandan oluşmaktadır. Bu çalışma kapsamında MobileNet ailesinin ilk versiyonu olan MobileNetV1 kullanılmıştır.

#### 1.4.4. EfficientNet

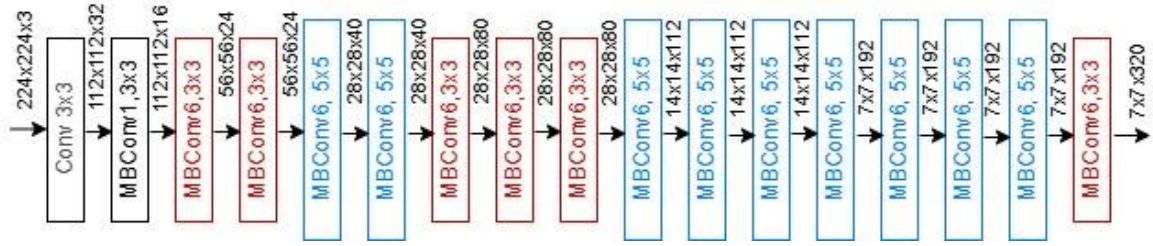
EfficientNet (Tan & Le, 2019)'in temel yeniliği, doğruluk ile hesaplama kaynakları arasındaki dengeyi optimize etmek amacıyla ağır derinlik (depth), genişlik (width) ve çözünürlük (resolution) boyutlarını eş zamanlı olarak ölçeklendirebilen bileşik bir ölçeklendirme yöntemi sunmasıdır. EfficientNet, ağ genişliği/ derinliği/ çözünürlüğü boyutlarını dengelemenin kritik bir öneme sahip olduğunu ve böyle bir dengenin bu boyutların her birini sabit bir oranda ölçekleyerek elde edilebileceğini savunur. Şekil 14' te EfficientNet ölçekleme yöntemi ile geleneksel yöntemler arasındaki fark gösterilmektedir.



Şekil 14. EfficientNet ölçeklendirmesi (a) temel ağ yapısı, (b)-(d) sadece tek boyut ölçeklendirme, (e) EfficientNet bileşik ölçeklendirme (Tan & Le, 2019)

Şekil 14 incelendiğinde (b)-(d) aralığındaki geleneksel CNN mimarileri sıklıkla genişlik, derinlik ve görüntü boyutlarından yalnızca birini ölçeklendirir veya seçilen herhangi iki ya da üç boyutu keyfi ve manuel ölçeklendirme yaparken, (e)' de EfficientNet bileşik ölçeklendirme yöntemi sabit bir oranda üç boyutu da eşit şekilde ölçekler. Bu üç boyutu eşit olarak ölçeklemek için derinlik:  $d = \alpha^\phi$ , genişlik:  $w = \beta^\phi$ , çözünürlük:  $r = \gamma^\phi$  ile temsil edildiğinde buradaki  $\alpha, \beta, \gamma$  sabitleri küçük ölçekli Grid arama ile belirlenir ve  $\alpha \geq 1, \beta \geq 1, \gamma \geq 1$  şeklindedir. Sezgisel olarak,  $\phi$  kullanıcı tarafından belirlenen ve modelin ölçeklendirilmesi için ne kadar ek kaynak ayrılacağını kontrol eden bir katsayıdır;  $\alpha, \beta, \gamma$  ise bu ek kaynakların sırasıyla ağırlıklı derinliği, genişliği ve çözünürlüğüne nasıl atanacağını belirler. Burada,  $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$  kısıtlaması getirilmiştir. Böylece herhangi bir yeni  $\phi$  değeri için toplam saniye başına kayan nokta işlemleri (Floating Point Operations Per Second, FLOPs) yaklaşık olarak 2 (iki) kat artmaktadır (Tan & Le, 2019).

EfficientNet' in ana yapı taşı, ilk olarak MobileNetV2 (Sandler vd., 2018) tarafından tanıtılan Ters Çevirilmiş Darboğaz (Inverted Bottleneck Convolution, MBConv) yapılarıdır. Bu yapılar, orijinal ResNet bloklarında bulunan kanalları önce sıkıştıran, ardından kanalları genişleten katmanların tersi olarak, önce kanalları genişleten, sonra bunları sıkıştıran katmanlardan oluşur. Böylece daha az kanala sahip katmanlar atlanarak bağlanır. Şekil 15' te temel EfficientNet (EfficientNetB0) yapısı sunulmuştur (Tan & Le, 2019).



Şekil 15. Temel EfficientNet mimari yapısı (Tan & Le, 2019)

Bu çalışma kapsamında EfficientNet ailesinin en az parametre sayısına sahip versiyonu EfficientNetB0 kullanılmıştır.

#### 1.4.5. ConvNeXt

Doğal dil işleme alanında büyük başarılar sağlayan dönüştürücü (transformer) mimarileri, Görüntü Dönüştürücü (Vision Transformer, ViT) (Dosovitskiy vd., 2020) ile

birlikte görsel tanıma problemlerine de uygulanmaya başlanmıştır. ViT, görüntüleri sabit boyutlu parçalara bölerek bu parçaları tıpkı kelime vektörleri gibi işlemekte ve saf dikkat (self-attention) mekanizması kullanmaktadır. Ancak ViT, yüksek hesaplama maliyeti ve lokal detayları yakalamada zayıflık gibi sınırlamalara sahiptir. Bu sorunlara çözüm olarak geliştirilen Swin Transformer (Liu vd., 2021), dikkat hesaplamasını lokal pencereler içinde yaparak hem hesaplama verimliliğini artırmış hem de hiyerarşik mimarisiyle farklı çözünürlük seviyelerinde temsil öğrenimini mümkün kılmıştır. Bu gelişmelerin ardından, CNN mimarileri yeniden değerlendirilmiş ve transformer tasarımlarından ilham alınarak ConvNeXt (Liu vd., 2022) mimarisi önerilmiştir. ConvNeXt, klasik ResNet yapısını temel alırken, Swin Transformer'dan alınan normalizasyon düzenlemeleri, kernel boyutları ve katman düzenlemeleri gibi modern tasarım bileşenleriyle güncellenmiş, böylece saf CNN yapısıyla yüksek performans elde edilmiştir.

Klasik ResNet50 mimarisini hiyerarşik yapıya sahip Swin Transformer yapısına göre modernize ederek tasarlanan ConvNeXt mimarisinde herhangi bir saf dikkat mekanizmasına dayalı yapı kullanılmamıştır. Ancak eğitim tekniklerinde, tasarımda, filtre boyutlarında veya kullanılan aktivasyon fonksiyonu çeşidi ve sayısında gibi performansı etkileyecek yapı ve parametrelerde Swin Transformer mimarisinden faydalanılmıştır. İlk olarak ResNet50 mimarisi transformerları eğitmek için kullanılan eğitim tekniklerine benzer olarak eğitilmiştir. Bu amaçla, ResNet modelleri için özgün olarak kullanılan 90 epoch'luk eğitim süresi, 300 epoch'a çıkarılmıştır. AdamW optimizasyon algoritmasıyla birlikte Mixup, Cutmix, RandAugment, Random Erasing gibi veri artırma teknikleri ve Stochastic Depth ile Label Smoothing gibi regülerizasyon yöntemleri kullanılmıştır. Bu gelişmiş eğitim stratejisi tek başına, ResNet50 modelinin doğruluk performansını % 76,1'den % 78,8'e yükseltmiştir (Liu vd., 2022).

Orijinal ResNet tasarımı, nesne algılama görevleri için "res4" aşamasına odaklanarak aşama hesaplama oranını (stage compute ratio) deneysel belirlemiştir. Swin Transformer ise benzer bir prensibe dayanarak belirlemiş ancak farklı olarak dört aşama arasında hesaplama yükünü sırasıyla 1:1:3:1 oranında dağıtmıştır. Böylece Swin Transformer makro tasarım yapısına yönelik ilk olarak ResNet50'nin blok numaraları Swin Transformer'ın oranıyla eşleşecek şekilde ayarlanmış ve ResNet50'nin doğruluk performansı %78,8' den %79,4'e çıkarılmıştır. İkinci olarak ResNet50'in 2 (iki) adım (stride) boyutlu  $7 \times 7$  konvolüsyon ve bunu takip eden maksimum havuzlama işlemine sahip ilk katmanı, Swin Transformer'ın daha basit tasarımlı, dört (dört) adım boyutlu  $4 \times 4$  konvolüsyon katmanına sahip, görüntüyü

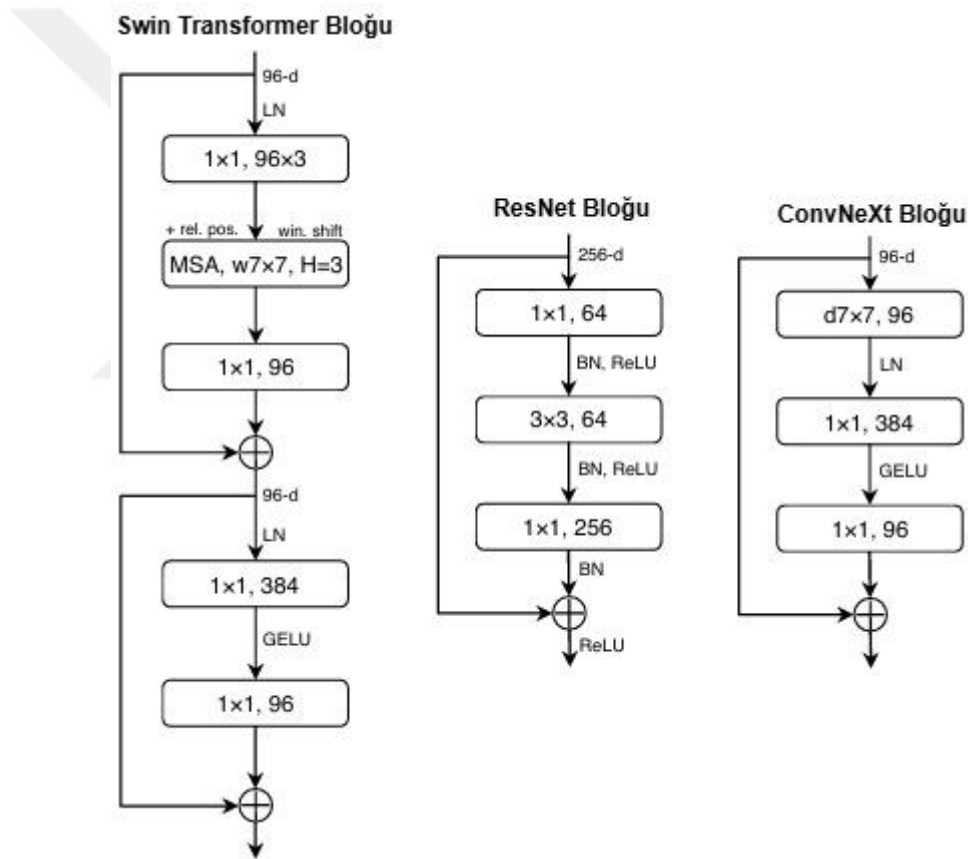
önce küçük parçalara (patch'lere) bölen "patchify" katmanı ile değiştirilmiştir. Bu değişiklik ile başarı oranı %79,4'ten %79,5'e ulaşmıştır. Makro tasarım yapısına yönelik son olarak MobileNet tarafından popülerlik kazanan derinlik bazlı evrişim katmanları ResNet-50 mimarisine kullanılmıştır. FLOP sayısının azalması doğruluğu geçici olarak azaltsa da, ağın genişliği Swin Transformer'a paralel olarak 64'ten 96'ya çıkarılarak ağ performansı %80,5'e yükseltilmiştir. Ayrıca transformer blok tasarımı, giriş boyutundan dört kat daha geniş gizli bir boyuta sahip Çok Katmanlı Algılayıcı (Multilayer Perceptron, MLP) bloğu içerir ve bu bloklar yapısında daha önceki bölümde EfficientNet'te detaylandırılan ters darboğaz yapılarını barındırır. ResNet50 mimarisinde ters darboğaz yapılarının kullanılması ile ağ performansı %80,5'ten %80,6'ya yükseltilmiştir (Liu vd., 2022).

ViT ve Swin Transformer gibi modellerde  $7 \times 7$ 'lik filtreler kullanılabilirken, klasik CNN modellerinde genellikle daha küçük  $3 \times 3$  filtreler tercih edilmektedir. Büyük filtre kullanımının ResNet50 mimarisindeki etkisini görebilmek amacıyla amaçlanan ConvNeXt mimarisinde derinlik bazlı evrişim katmanının konumu yukarı taşınmıştır. Konumu yukarı taşımak ResNet50 mimarisinin performansını geçici olarak %79,9'a düşürürken.  $7 \times 7$  boyutlu büyük filtre kullanımı performansı tekrar %80,6'ya yükseltmiştir (Liu vd., 2022).

Ağın mikro tasarım yapısına yönelik ViT mimarilerinde GELU aktivasyon fonksiyonu, klasik CNN tabanlı mimarilerde ise ReLU kullanılmasından yola çıkılarak amaçlanan ConvNeXt mimarisinde ReLU yerine GELU aktivasyon fonksiyonu tercih edilmiştir. Ancak, bu değişim ağı performansında bir değişiklik yaratmamıştır. Transformer ve ResNet blokları arasındaki diğer bir fark ise Transformer'ların daha az aktivasyon fonksiyonuna sahip olmasıdır. Bu strateji ResNet bloklarına uygulandığında ağın performansı %0,7 artarak %81,3'e ulaşmış ve bu başarı ile Swin Transformer başarısına oldukça yaklaşmıştır.

Transformer blokları genellikle daha az normalizasyon katmanına sahiptir. ResNet bloklarındaki iki Batch Normalizasyon (BN) katmanı kaldırılarak ve  $1 \times 1$  evrişim katmanlarından önce yalnızca bir BN katmanı bırakılarak ağın başarı performansı %81,4'e çıkarılmıştır. Bu başarı oranı Swin Transformer başarısını geride bırakmıştır. BN, yakınsamayı artırmak ve aşırı uyumu azaltmak için klasik CNN'lerde yaygın olarak kullanılır ancak karmaşıklıkları model performansına zarar verebilmektedir. Layer Normalizasyon (LN) ise Transformer bloklarında iyi performans sergilemekte ve daha basit yapıya sahiptir. ResNet yapısında BN ile LN doğrudan değiştirildiğinde modelin performansı bir miktar iyileşerek %81,5 doğruluk oranı elde edilmiştir. Son olarak, ResNet

mimarisinde, her aşamanın başında  $3 \times 3$  ve 2 adım boyutlu evrişim filtreleri kullanılarak uzaysal indirgeme (spatial downsampling) yapılırken, Swin Transformer'da her aşama arasına ayrı bir uzaysal indirgeme katmanı eklenir. Uygulanan yeni yaklaşımda ise  $2 \times 2$  ve 2 adım boyutlu evrişim filtreleri kullanılarak uzaysal indirgeme sağlanmıştır. Bu durum dengesiz bir eğitime yol açsa da normalizasyon katmanları eklenerek kararlılık sağlanmış ve ağ performansı %82,0' a ulaşarak Swin Transformer'ın %81,3'lük başarısını geride bırakmıştır (Liu vd., 2022). Şekil 16' da bir Swin Transformer, bir ResNet ve tasarlanan bir ConvNeXt bloğu gösterilmiştir. Bu tez kapsamında temel ConvNeXt mimarisi olan ConvNeXt-base kullanılmıştır.



Şekil 16. Swin Transformer, ResNet ve ConvNeXt için blok tasarımları (Liu vd., 2022)

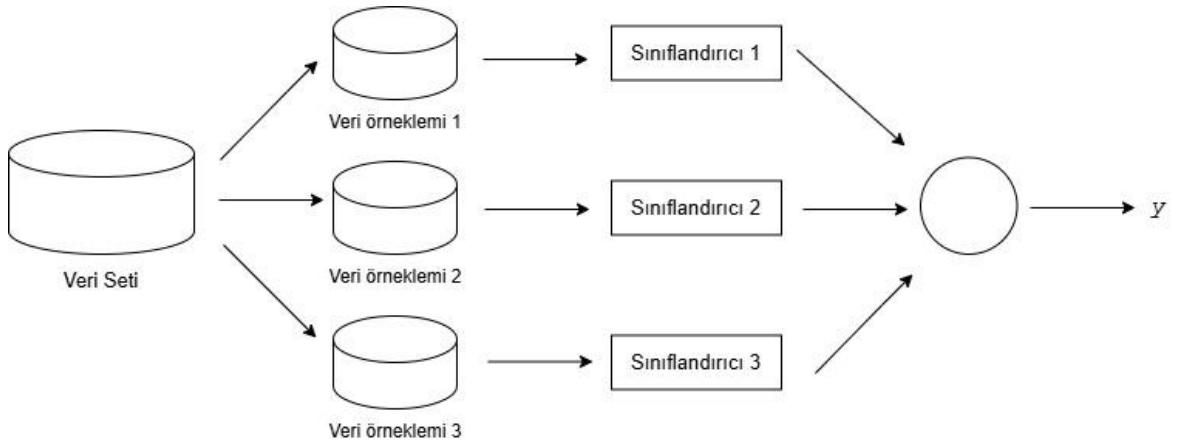
#### 1.4.6. Topluluk Öğrenme Mimarisi

Topluluk öğrenme, birden fazla modelin bir araya getirilmesiyle tahmin doğruluğunu artırmayı amaçlayan bir yaklaşımdır. Bu yöntemin temelinde, farklı modellerin güçlü

yönlerinden yararlanarak daha iyi sonuçlar elde etmek yatmaktadır. Bu kapsamda, veriyi farklı açılardan değerlendirebilmek için birbirinden farklı modeller ya da aynı modelin çeşitli varyasyonları kullanılmaktadır. Bireysel modellerin ürettiği tahminler; oylama, ortalama alma veya daha gelişmiş birleştirme teknikleriyle bir araya getirilir (Mohammed & Kora, 2023). Geleneksel topluluk öğrenmesi, çoğunlukla klasik makine öğrenmesi modellerinin entegrasyonu ile gerçekleştirilmekte ve çeşitli alanlarda uygulanmaktadır. Sınıflandırma performansını artırma ve genellenebilirlik kapasitesi sayesinde, topluluk öğrenmesi derin öğrenmeye dayalı sınıflandırma problemlerinde de oldukça yaygın kullanılmaktadır. Herhangi bir topluluk öğrenme sisteminin genel çerçevesi,  $k$  adet temel sınıflandırıcıyı,  $c_1, c_2, \dots, c_k$ , tek bir çıktıyı tahmin edecek şekilde birleştirmek için  $G$  toplama fonksiyonunu kullanmaktır. Boyutu  $n$  olan bir  $D$  veri kümesi ile  $m$  boyutlu özellikler verildiğinde,  $D = \{(x_i, y_i), 1 \leq i \leq n, x_i \in R^m\}$ , bu topluluk öğrenmesi ile tek çıktı  $y_i$ 'nin tahmini aşağıdaki gibidir (Mohammed & Kora, 2023):

$$y_i = \phi(x_i) = G(c_1, c_2, \dots, c_k) \quad (6)$$

Veri örnekleme yönteminin seçimi ve temel sınıflandırıcıların veri örnekleme kararlarında çeşitliliğin sağlanması topluluk öğrenmesinin performansını etkileyen en önemli faktörlerden biridir (Mohammed & Kora, 2023). Farklı veri örneklemeleri üzerinde eğitilen üç temel sınıflandırıcıdan oluşan bir topluluk öğrenme mimarisi örneği Şekil 17' de sunulmuştur.



Şekil 17. Farklı veri örneklemeleri üzerinde eğitilen ve üç temel sınıflandırıcıdan oluşan bir topluluk öğrenme mimarisi (Kazmaier & Van Vuuren, 2022)

Şekil 17, üç temel sınıflandırıcıya sahip genel bir topluluk öğrenme mimarisini göstermektedir. Bu mimaride, çeşitliliği sağlamak için orijinal veri kümesinden üç ayrı eğitim veri kümesi örneklenir. Her eğitim veri kümesi ayrı bir sınıflandırıcıyı eğitmek için kullanılır. Sınıflandırıcı çıktıları daha sonra bir araya (oylama, ortalama vb. yöntemler ile) getirilerek topluluk modelleme çıktısı üretilir. Topluluk öğrenmesinin tahmin performansını artırmadaki başarısının başlıca nedenleri aşağıda sıralanmıştır (Sagi & Rokach, 2018):

- Yalnızca küçük bir veri setinin mevcut olduğu durumlarda, bir model eğitim verisine mükemmel şekilde uyum sağlayan örüntüler öğrenebilir. Ancak bu durum modelin görülmemiş veriler üzerindeki genelleme yeteneğini olumsuz etkileyebilir. Birden fazla modelin tahminlerinin ortalamasının alınması, hatalı bir tahmin seçilme riskini azaltarak genel tahmin performansını artırır.
- Yerel arama algoritmalarına dayanan tekil modeller, yerel minimumlara takılma riski taşır. Ancak, birden fazla öğrencinin bir araya getirilmesiyle oluşturulan topluluk öğrenmesi, yerel minimum elde etme riskini azaltır.
- Doğru bir tahmin, tek bir modelin arama uzayının dışında kalabilir. Ancak, birden fazla modelin bir araya getirilmesiyle arama uzayı genişletilebilir veya çeşitlendirilebilir; bu da veriye daha uygun bir arama alanı oluşturularak doğru tahmine ulaşma olasılığını artırır.

Topluluk öğrenme teknikleri, özellikle aktivite tanıma, kişiyi yeniden kimliklendirme ve hareket tanıma gibi alanlarda, topluluk öğrenmenin model çeşitliliğini sağlama yoluyla tahmin doğruluğunu artırma özelliği sayesinde önemli başarılar elde etmiştir. Örneğin, aktivite tanıma alanında Gupta ve Semwal (2020), çoklu aktiviteye dayalı yürüyüş sınıflandırması için farklı sınıflandırma algoritmalarını, Destek Vektör Makinesi (DVM), En Yakın Komşu (KNN), Aşırı Öğrenme Makinesi (AÖM) ve PCA, bir arada kullanan bir topluluk öğrenme yaklaşımı önermiştir. Bu yaklaşım, farklı sınıflandırıcıların bir araya getirilmesinin aşırı öğrenmenin (overfitting) önüne geçerek hiper parametrelere olan bağımlılığı azalttığı ve ortalamada daha istikrarlı bir performans sağladığı düşüncesine dayanmaktadır. Başka bir çalışmada (Semwal vd., 2021), önerilen CNN tabanlı topluluk öğrenme hibrit modelinin, diğer hibrit modellerin her birinden daha yüksek performans gösterdiği kanıtlanmıştır. Kişiyi yeniden kimliklendirme alanında, Ye ve diğerleri (2020), farklı sınıflandırıcılar arasında bilgi aktarımını kolaylaştırarak ve daha az hiper parametre kullanarak iki kollu bir CNN' in performansını artıran işbirlikçi bir topluluk öğrenme stratejisi önermiştir. Liang ve diğerleri (2019) tarafından önerilen bir 3B hareket tanıma

çalışmasında, üç aşamalı CNN mimarisinin bir aşamasında, ağın genelleme yeteneğini artırabilmek amacıyla çok görevli ve topluluk öğrenme temelli bir ağ yapısı kullanılmıştır. Xu ve diğerleri (2018), model tabanlı hareket tanıma için atlama bağlantılı yapılar içeren dört adet 1B CNN alt ağı ile çeşitli ve tamamlayıcı özelliklerin çıkarıldığı bir topluluk sinir ağı önermiştir. Bu dört alt ağ, tek bir topluluk öğrenme ağına entegre edilerek yüksek performans elde edilmiştir. Topluluk öğrenme, aktivite tanıma, kişiyi yeniden kimliklendirme ve hareket tanıma gibi yürüyüş tanımaya yakın alanlarda başarılı sonuçlar vermiş olsa da yürüyüş tanımadaki uygulamaları görece sınırlı ve yeterince araştırılmamış durumdadır. Örneğin, Wang ve Yan (2020) tarafından önerilen bir çalışmada, iki yürüyüş bölgesi arasındaki mesafeye dayalı yeni bir yürüyüş özelliği çıkarılmış, ardından aynı bireyin farklı görüş açılarına ait yürüyüş görüntülerine göre ayrı ayrı eğitilen bir dizi açığa bağımlı HMM kullanılmış ve son olarak bu öğrenciler topluluk yöntemiyle birleştirilerek sınıflandırma yapılmıştır. Bir diğer çalışmada (Xiuhui & Yan, 2021), GEI'lerden rastgele örnekleme yöntemiyle bir dizi örneklem kümesi oluşturulmuş, ardından farklı hiper parametrelere sahip birbirinden farklı birincil CNN'ler eğitilmiş ve ikincil bir sınıflandırıcı ile bu modeller topluluk öğrenme yöntemiyle birleştirilmiştir. Ancak bu çalışmaların bazı yöntemsel ve deneysel sınırlılıkları bulunmaktadır. Özellikle, test aşamasında galeri ve sorgu görüntü setlerinin ayrılmamış olması, bu çalışmaların sonuçlarının literatürdeki son teknoloji yöntemlerle karşılaştırılmasını zorlaştırmakta ve topluluk öğrenmenin etkinliğinin ortaya konulmasını güçleştirmektedir.

Bu çalışma kapsamında topluluk öğrenme mimarisi, derin öğrenme modeli tarafından işlenen insan vücudunun farklı bölgelerinden en ayırt edici özelliklerin çıkarılabilmesi ve nihai tanıma başarı oranının artırılması amacıyla kullanılmıştır. Bu amaçla tez kapsamında hazırlanan mimaride veri örneklemede çeşitliliği sağlamak amacıyla iki temel yöntem kullanılmıştır;  $k$ -katlı çapraz doğrulama ( $k$ -fold cross-validation) ve önyükleme toplulaştırması (bootstrap aggregating veya bagging).  $k$ -katlı çapraz doğrulama yaklaşımında, veri seti eşit büyüklükte  $k$  alt kümeye bölünmekte ve iteratif olarak her yinelemede  $k-1$  alt küme üzerinde bir sınıflayıcı eğitilmektedir. Topluluk öğrenme mimarisinde  $k$ , veri örneklem sayısına denk gelmektedir. Örneğin üç temel sınıflandırıcının kullanıldığı bir topluluk öğrenme mimarisinde üç adet veri örnekleme kullanılacağından  $k$  sayısı üçe eşit olacaktır. Öte yandan, önyükleme toplulaştırması (bagging), yer değiştirmeli rastgele örnekleme yöntemini kullanarak eğitim veri örneklemelerini elde eder. Bu yaklaşım, bir örneğin birden fazla eğitim alt kümesinde yer almasına olanak tanır. Sonuç olarak, bir

veri setinden üretilen veri örneklemi  $k$ -katlı çapraz doğrulama yönteminde birbirleri ile ortak verilere sahip olamazken önyükleme toplulaştırması yönteminde aynı veriler birden fazla örnekleme görülebilir.

#### 1.4.7. Transfer Öğrenme

Transfer öğrenme, bir eğitimden öğrenilen bilginin farklı bir eğitimde yeniden kullanılmasıdır. Daha kapsamlı olarak bu yaklaşım, ağın daha önce farklı bir kaynak veri seti üzerinde önceden eğitilmiş bir model ile başlatılmasını ve ardından hedeflenen veri seti üzerinde ince ayar (fine-tuning) yapılmasını içermektedir. Bir CNN sıfırdan eğitildiğinde, modelin parametreleri genellikle rastgele olarak başlatılır. Ancak, eğitim veri setinin küçük olması durumunda bu rastgele başlatma, verideki desenlerin öğrenilmesini zorlaştırabilir ve bu durum modelin performansını olumsuz yönde etkileyebilir. Transfer öğrenme tekniklerinden yararlanılarak eğitim süresi kısaltılabilir, az öğrenme (underfitting) veya aşırı öğrenme (overfitting) problemleri etkili bir şekilde azaltılabilir ve modelin genelleme yeteneği artırılabilir (Liu vd., 2024).

ImageNet veri seti, bilgisayarla görme alanında kullanılan, manuel olarak etiketlenmiş 1000 sınıfa ait on milyondan fazla doğal görüntü içeren büyük ölçekli bir görsel veri setidir (Russakovsky vd., 2015). Transfer öğrenmede, ImageNet veri seti kaynak veri seti olarak alınır. Ağ modeli, öncelikle bu veri kümesi üzerinde eğitilmiş olup öğrenilen model parametreleri, çalışmaya özgü veri kümesi üzerinde eğitim yapılırken ağın başlatılmasında kullanılmaktadır. Eğitim oldukça büyük bir veri seti üzerinde gerçekleştirildiğinden daha verimli ağ kullanımı ve yüksek performanslı eğitim sağlanmaktadır (Liu vd., 2024). Doğal görüntüler ile yürüyüş görüntüleri arasındaki farklılıklar göz önünde bulundurularak, tez çalışmasında yalnızca kaynak veri setine ait model parametreleri ağın başlangıç parametreleri olarak kullanılmıştır. Ayrıca, modellerin sonuna 4096 boyutuna sahip yeni bir tam bağlantılı katman eklenmiş ve sınıflandırıcı katmandaki nöron sayısı yürüyüş veri setinde kullanılan eğitim seti kişi sayısı ile değiştirilmiştir. Böylece ağdaki tüm katmanların parametreleri yeniden eğitilerek yürüyüşe ait veri setleri üzerinde ince ayar yapılmıştır.

#### 1.5. Literatür Veri Setleri ve Performans Karşılaştırması

Yürüyüş tanımaya yönelik hazırlanmış çoğu yürüyüş veri seti, genellikle laboratuvar ortamı veya sabit dış mekânlar gibi nispeten sınırlı ve sabit ortamlarda elde edilmiştir. İlk

çalışmalarda kullanılmış, laboratuvar ortamında çekilmiş CMU MoBo (Gross & Shi, 2001), düzgün silüet verilerine sahip SOTON ve dış ortamda çekilmiş USFHumanID gibi veri setleri hem örnek sayısının azlığı hem de yeterli görünüm tabanlı varyasyonları sağlayamadığı için son teknoloji yöntemlerde kullanılmamaktadır. Özellikle CASIA-B tanımada başarıyı ciddi oranda düşüren giyim varyasyonlarını barındırdığı için günümüzdeki yürüyüş tanıma araştırmalarında en yaygın kullanılan veri setidir. Bunu OU-MVLP veri seti izler. CASIA-B, içerisinde 124 kişi ve 13.640 yürüyüş dizisi barındırmakta olup 2006 yılında oluşturulmuştur. OU-MVLP ise 10.307 kişiliktir ve 288.596 yürüyüş dizisi içermektedir. Bu da onu, kişi sayısı bakımından büyük bir yürüyüş veri kümesi yapmaktadır.

Tablo 1’ de literatürde şimdiye kadar en önde gelen, CMU MoBo, CASIA-A (Wang vd., 2003), SOTON, USFHumanID, CASIA-B, CASIA-C, OU-ISIRSpeed (Tsuji vd., 2010), OU-ISIRCloth (Hossain vd., 2010), OU-ISIRMV (Makihara vd., 2010), OU-LP (ISIR), TUM-GAIT (Hofmann vd., 2014), OU-LP Age (Xu vd., 2017), OU-MVLP, OU-LPBag (Uddin vd., 2018), OU-MVLPPose (An vd., 2020), GREW (Zhu vd., 2021), Gait3D (Zheng vd., 2022) yürüyüş tanıma veri setleri, barındırdıkları kişi (Id), dizi ve açı sayısı, veri tipleri, oluşturuldukları deney ortamı ve barındırdıkları çeşitli varyasyonlar üzerinden detaylandırılmıştır.

Tablo 1. Literatürde en önde gelen veri setleri içerik karşılaştırma tablosu (Zhu vd., 2021)

Veri Seti	Yayın Yeri	Id	Dizi	Açı	Veri Tipi	Ortam	Varyasyon
CMU MoBo	TR2001	25	600	6	RGB,Silüet	Lab.	A,ÇT,H,Y
CASIA-A	TPAMI2003	20	240	3	RGB	Lab.	A
SOTON	ASSC2004	115	2,128	2	RGB,Silüet	Lab.	A
USFHumanID	TPAMI2005	122	1,870	2	RGB	Lab.	A,ÇT,Y,AY
CASIA-B	ICPR2006	124	13,640	11	RGB,Silüet	Lab.	A,ÇT,G
CASIA-C	ICPR2006	153	1,530	1	Kızılötesi,Silüet	Lab.	ÇT,H
OU-ISIRSpeed	CVPR2010	34	612	1	Silüet	Lab.	H
OU-ISIRCloth	PR2010	68	2,764	1	Silüet	Lab.	G
OU-ISIRMV	ACCV2010	168	4,200	25	Silüet	Lab.	A
OU-LP	TIFS2012	4,007	7,842	2	Silüet	Lab.	A
TUM-GAIT	JVCIR2014	305	3,370	1	RGB,Derinlik,Ses	Lab.	ÇT, AY
OU-LP Age	CVA2017	63,846	63,846	1	Silüet	Lab.	Yaş
OU-MVLP	CVA2018	10,307	288,596	14	Silüet	Lab.	A
OU-LPBag	CVA2018	62,528	187,584	1	Silüet	Lab.	ÇT
OU-MVLPPose	TBIOM2020	10,307	288,596	14	2B İskelet	Lab.	A
GREW	ICCV2021	26,345	128,671	882	Silüet,Akış,2/3B İskelet	Dış Ort.	A,AP,ÇT,G, Y, ÖR, İŞ
Gait3D	CVPR2022	4,000	25,309	39	Silüet,2/3B İskelet,3B Mesh&SMPL	Dış Ort.	A,H

Tablo 1’ de varyasyonlar sütunundaki kısaltmalar sırasıyla; A: aç, AP: arkaplan, ÇT: çanta taşıma, G: giyim, Y: yüzey, H: hız, AY: ayakkabı, ÖR: örtülme ve İŞ: ışıklandırma şeklindedir. Tablo 1’ de yer alan veri kümelerinin çoğu, kontrollü ortamlarda oluşturulmuş ve önceden tanımlanmış bakış açıları arasında yürüyüş tanıma amacıyla tasarlanmıştır. Ancak, gerçek dünya senaryolarında yürüyüş tanıma; çeşitli görüş açıları, örtülmeler, farklı taşıma ve giyim biçimleri, karmaşık ve dinamik arka plan gürültüleri, aydınlatma koşulları, yürüme tarzı ve zemin etkisi gibi tamamen kontrolsüz zorluklardan oluşmaktadır. Bu gerçeklikten yola çıkılarak son yıllarda tamamen doğal koşullar altında, gerçek dünya senaryolarına uygun veri setleri üretilmeye başlamıştır. Ancak, bu durum beraberinde artan kişi, dizi, aç ve varyasyon çeşitliliğinin sayısına bağlı olarak hesaplama maliyeti problemini beraberinde getirmektedir. Bu veri setlerini işleyebilmek ve detaylı analizler yapabilmek için oldukça güçlü makinelere ihtiyaç duyulmaktadır. Kullanımı en yaygın olan CASIA-B veri seti için şimdiye kadar literatürde sunulmuş en önde gelen çalışmalar arasında bir performans analizi Tablo 2’ de gerçekleştirilmiştir.

Tablo 2. CASIA-B veri seti önde gelen çalışmalar performans karşılaştırması

Galeri NM#1-4	Sorgu	0-180											ort
		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	
NM #5-6	GEI+PCA	16.8	24.6	19.9	21.4	26.7	24.6	27.3	24.7	19.5	20.9	21.9	22.6
	GeiNet	45.8	57.6	67.1	66.9	56.3	48.3	58.3	68.4	69.4	59.0	46.5	58.5
	DeepCNN	82.6	90.3	96.1	94.3	90.1	87.4	89.9	94.0	94.7	91.3	78.5	89.9
	GaitSet	90.8	97.9	99.4	96.9	93.6	91.7	95.0	97.8	98.9	96.8	85.8	95.0
	GaitPart	94.1	98.6	99.3	98.5	94.0	92.3	95.9	98.4	99.2	97.8	90.4	96.2
	ACL	92.0	98.5	100.0	98.9	95.7	91.5	94.5	97.7	98.4	96.7	91.9	96.0
	GaitGL	96.0	98.3	99.0	97.0	96.9	95.4	97.0	98.9	99.3	98.8	94.0	97.4
	GaitBase	93.9	98.8	99.6	98.1	94.0	91.6	94.9	98.4	99.3	98.5	91.8	<b>97.6</b>
	PoseGait	55.3	69.6	73.9	75.0	68.0	68.2	71.1	72.9	76.1	70.4	55.4	68.7
BG #1-2	GaitGraph	85.3	88.5	91.0	92.5	87.2	86.5	88.4	89.2	87.9	85.9	81.9	87.7
	DeepCNN	64.2	80.6	82.7	76.9	64.8	63.1	68.0	76.9	82.2	75.4	61.3	72.4
	GaitSet	83.8	91.2	91.8	88.8	83.3	81.0	84.1	90.0	92.2	94.4	79.0	87.2
	GaitPart	89.1	94.8	96.7	95.1	88.3	84.9	89.0	93.5	96.1	93.8	85.8	91.5
	GaitGL	92.6	96.6	96.8	95.5	93.5	89.3	92.2	96.5	98.2	96.9	91.5	<b>94.5</b>
	GaitBase	91.9	95.5	96.8	94.7	90.9	88.9	91.7	94.9	96.2	95.5	86.3	94.0
	PoseGait	35.3	47.2	52.4	46.9	45.5	43.9	46.1	48.1	49.4	43.6	31.1	44.5
CL #1-2	GaitGraph	75.8	76.7	75.9	76.1	71.4	73.9	78.0	74.7	75.4	75.4	69.2	74.8
	DeepCNN	37.7	57.2	66.6	61.1	55.2	54.6	55.2	59.1	58.9	48.8	39.4	54.0
	GaitSet	61.4	75.4	80.7	77.3	72.1	70.1	71.5	73.5	73.5	68.4	50.0	70.4
	GaitPart	70.7	85.5	86.9	83.3	77.1	72.5	76.9	82.2	83.8	80.2	66.5	78.7
	GaitGL	76.6	90.0	90.3	87.1	84.5	79.0	84.1	87.0	87.3	84.4	69.5	<b>83.8</b>
	GaitBase	60.2	77.6	82.8	78.7	74.8	72.2	76.1	78.2	76.8	72.0	56.9	77.4
	PoseGait	24.3	29.7	41.3	38.8	38.2	38.5	41.6	44.9	42.2	33.4	22.5	36.0
GaitGraph	69.6	66.1	68.8	67.2	64.5	62.0	69.5	65.6	65.7	66.1	64.3	66.3	

Tablo 2’de GEI+ PCA (Wang vd., 2003) makine öğrenmesine dayanan bir yöntem iken, GeiNet ilk derin öğrenme tabanlı çalışma olarak kabul edilebilir. Ardından, literatürden görünüm tabanlı çalışmalardan DeepCNN, GaitSet, GaitPart, ACL, GaitGL (Lin vd., 2021) ve GaitBase (Fan vd., 2023) ile model tabanlı çalışmalardan PoseGait ve GaitGraph’ ın başarıları sunulmuştur. Tablo 2 incelendiğinde makine öğrenmesine dayanan GEI+ PCA yönteminin CASIA-B veri seti için oldukça düşük bir tanıma performansına ulaştığı görülmektedir. Derin öğrenmeye dayanan GeiNet’ e geçiş ile birlikte tanıma oranında %100’ ün üzerinde bir artış gözlenmektedir. Bu aşamadan sonra derin öğrenmeye dayanan görünüm tabanlı yürüyüş tanıma yöntemlerinin tanıma oranlarında kısa bir süre içinde ciddi bir iyileşme sağlandığı görülmektedir. Derin öğrenmeye dayanan model tabanlı yöntemlerde ise görünüm tabanlı yöntemlere göre bir performans düşüklüğü görülse de kendi kategorisi içerisinde tanıma oranında ciddi bir artış sağlanmıştır. Bu durum en iyi görünüme dayalı BG ve CL varyasyonlarında gözlemlenmiştir.

## 2. YAPILAN ÇALIŞMALAR

Yürüyüş tanıma problemlerinde genel amaç insan vücudunun şekli hakkındaki, uzaysal ve zamansal değişen karakteristiklerden yola çıkarak, benzersiz ve değişmez özellikleri çıkarabilmek ve bu özellikleri kullanarak kimliklendirme işlemini başarıya ulaştırabilmek olmuştur. Literatürde şimdiye kadar bu amaçla birçok farklı yürüyüş deseni üretilmiş ve farklı tanıma yöntemleri önerilmiştir. Tez kapsamında ilk olarak literatürde en sık kullanılan GEI, silüetler, optik akışlar, RGB çerçeveleri ve farklı vücut bölümleri desenlerinin önceki bölümlerde detaylandırılan VGG16, ResNet50, MobileNetV1, EfficientNetB0 ve ConvNext-base mimarileri üzerinden yürüyüş tanıma performansları literatürdeki kullanım sıklığına göre değerlendirilmiş, sonrasında bu desenlerin birbirleri ile olan kombinasyonlarının hazırlanan ‘Çoklu Modalite Ağ Mimarileri’ üzerinden tanıma başarısını ne derece artırdıkları araştırılmıştır. Buradan elde edilen sonuçlar üzerinden tanıma başarısında en etkili girdi verisine ve ağ mimarisine yönelik ikinci bir ‘Parça-Tabanlı Ağ Mimarisi’ hazırlanmış bu mimaride ağın genel tanıma performansını artırmak amacıyla topluluk öğrenme metotları kullanılmıştır. Çalışmada, topluluk öğrenme mimarisini oluşturan CNN’lerin çıktılarının tek ve daha sağlam bir yürüyüş öznitelik temsiline dönüştürebilmek amacıyla yeni bir model eğitilmiş ve bu modele ‘Toplayıcı Model (TM)’ adı verilmiştir. Klasik topluluk öğrenme yöntemlerinin çoğunluk oylaması veya ağırlıklı ortalama gibi birleştirme yaklaşımlarına kıyasla, bu yaklaşım birleştirilmiş öznitelik uzayını doğrudan iyileştirerek daha sağlam kararlar alınmasına olanak sağlamıştır. Hazırlanan ikinci mimaride ayrıca, yürüyüş tanıma problemlerinde başarıyı ciddi oranda düşüren ceket giyme ve çanta taşıma gibi görünüme dayalı varyasyonlar için bir parça eleme işlemi önerilmiştir. Bu işlem için varyasyonlu ve varyasyonsuz görüntüler arasında bir fark görüntüsü hesaplandıktan sonra değişimden en çok etkilenen parça, önceden tanımlanmış bir eşik değeri kullanılarak görüntüden çıkarılmış ve tanıma işlemine dahil edilmemiştir. Özetle bu çalışmada şu katkılar sağlanmıştır:

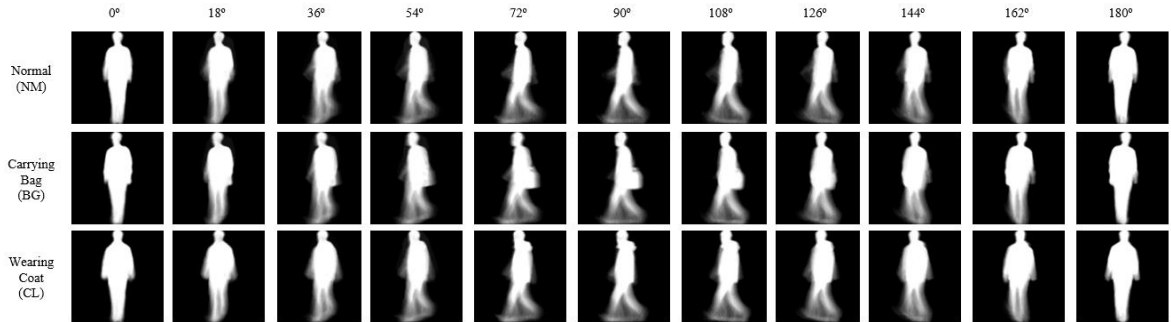
- Çoklu modalite ağ mimarileri üzerinden en yürüyüş tanımada en etkili yürüyüş deseni kombinasyonu tespit edilmiştir.
- Yeni bir Topluluk Öğrenmeli Parça-Tabanlı Ağ mimarisi önerilmiştir.
- Topluluk öğrenme çıktılarından en dayanıklı öznitelikleri üretebilen bir Toplayıcı Model (TM) hazırlanmıştır.

- Görünüme dayalı varyasyonların tanımadaki etkisini azaltabilmek amacıyla bir Parça-eleme işlemi önerilmiştir.
- Hazırlanan ağ mimarilerinin hem laboratuvar hem de dış mekân olmak üzere üç farklı veri seti üzerinde performans analizi gerçekleştirilmiştir.

## 2.1. Kullanılan Veri Setleri

Tez kapsamında hazırlanan ağların performans değerlendirmelerini yapmak amacıyla yürüyüş tanımadada en sık kullanılan veri setlerinin başında gelen, bünyesinde çapraz açı varyasyonu ile görünüm tabanlı varyasyonları barındıran CASIA-B (Yu vd., 2006) ile diğer veri setleri CASIA-C (Tan vd., 2006) ve Outdoor-gait (Song vd., 2019) kullanılmıştır.

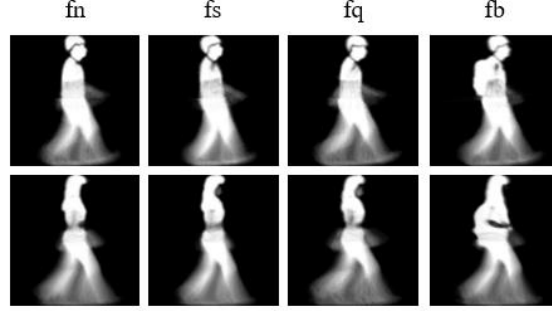
CASIA-B veri seti laboratuvar ortamındaki her biri 10 video grubuna sahip 124 farklı bireyden oluşmaktadır. Bu gruplardan 6 çerçeve dizisi yürüyüş varyasyonlarından basit giyimli normal yürüyüş (NM), 2 çerçeve dizisi çantalı yürüyüş (BG) ve 2 çerçeve dizisi ceketli veya montlu yürüyüş (CL) durumlarını temsil etmektedir. Her grup, sabit 18 derecelik aralıklarla alınmış 11 farklı açıdan ( $0^\circ$ ,  $18^\circ$ , ...,  $162^\circ$ ,  $180^\circ$ ) elde edilen 11 yürüyüş sekansı içermektedir. Sonuç olarak veri setindeki 124 kişinin her biri için 11 farklı açı ve 10 video grubu üzerinden 110 çerçeve dizisi bulunmaktadır. Şekil 18’ de CASIA-B veri setinden üretilmiş bazı GEI örnekleri sunulmuştur.



Şekil 18. CASIA-B veri seti 11 farklı açıda GEI örnekleri

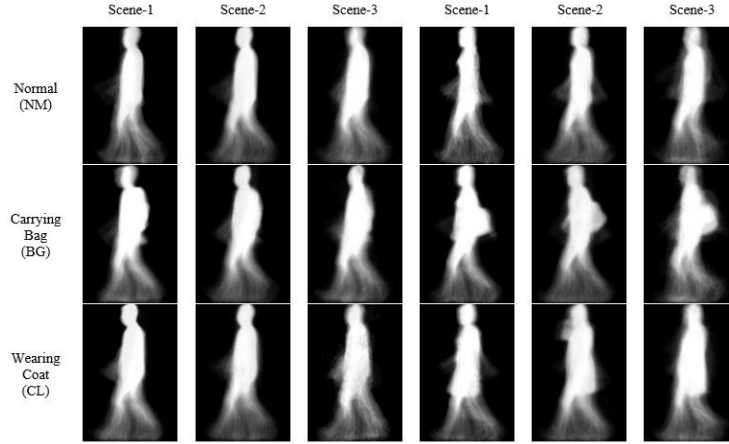
CASIA-C veri seti, termal (kızılötesi) kamera kullanılarak doğal ortamdan toplanan görüntülerden oluşmaktadır. Bu veri setinde, dört farklı koşul, normal yürüyüş (fn), yavaş yürüyüş (fs), hızlı yürüyüş (fq) ve çanta ile yürüyüş (fb), altında yürüyen 153 birey yer almaktadır. Her birey için toplam 10 yürüyüş sekansı bulunmaktadır; bunların dördü normal

(fn), ikisi yavaş (fs), ikisi hızlı (fq) ve ikisi çanta ile yürüyüş (fb) durumlarına aittir. Şekil 19’ da CASIA-C veri setinden üretilmiş bazı GEI örnekleri gösterilmiştir.



Şekil 19. CASIA-C veri seti GEI örnekleri

Outdoor-Gait, karmaşık dış mekân arka planlarına sahip kapsamlı bir dış ortam veri setidir. Bu veri kümesinde, üç farklı giyim koşulunda (NM: normal, CL: ceket ile, BG: çanta ile) ve üç farklı sahnede (SAHNE-1: sade arka plan, SAHNE-2: durağan ve karmaşık arka plan, SAHNE-3: hareketli nesnelere içeren dinamik ve karmaşık arka plan) yer alan toplam 138 birey bulunmaktadır. Şekil 20’de Outdoor-Gait veri setinden üretilmiş bazı GEI örnekleri verilmiştir.



Şekil 20. Outdoor-Gait veri seti GEI örnekleri

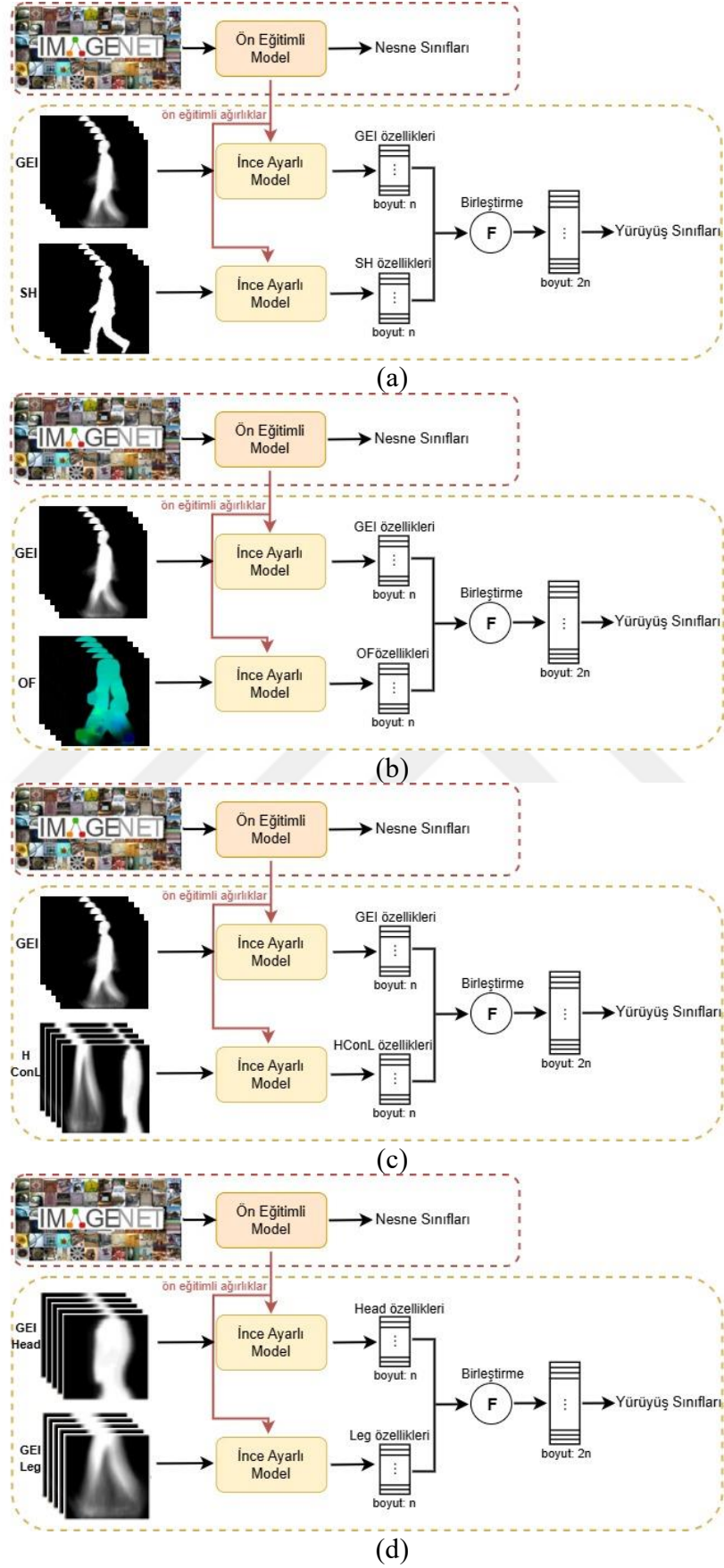
Deneylerde; CASIA-B veri seti için ilk 74 birey eğitim, geriye kalan 50 birey test; CASIA-C veri seti için ilk 24 ve 62 birey eğitim, kalan 53 birey test; Outdoor-Gait veri seti için ilk 69 birey eğitim kalan bireyler test veri seti olarak kullanılmıştır.

## 2.2. Çoklu Modalite Ağ Mimarisi

Bu bölümde, yürüyüş tanıma çözümünde en bilinen yürüyüş temsillerinin etkinliği, derin öğrenme tabanlı yaklaşımlar doğrultusunda araştırılmıştır. Bu amaçla, silüetler, optik akışlar ve hazırlanan bir GEI baş-bacak bölgelerinin birleştirilmiş görüntüsü ile GEI yürüyüş görüntüsünün kendisini içeren farklı modalitelerin kombinasyonu ile kapsamlı bir performans değerlendirmesi gerçekleştirilmiştir. Bu değerlendirme, adil bir karşılaştırma sağlamak amacıyla ince ayarlanmış başta EfficientNetB0, MobileNetV1 ve ConvNeXt-base gibi farklı çoklu modaliteleri işleyebilen CNN mimarileri üzerinden yapılmıştır. Bu ağ yapıları, birçok girdi modalitesinden gelen bilgileri bir araya getirerek daha zengin ve daha kompakt bir yürüyüş gösterimi oluşturmayı amaçlamaktadır. Hazırlanan modeller, GEI, silüet, optik akış ve hazırlanan GEI baş-bacak bölgelerinin birleştirilmiş görüntüsü üzerinde ayrı ayrı eğitilmiş ardından, GEI' den çıkarılan öznitelikler ile diğer modalitelerden çıkarılan özniteliklerle ikili olarak birleştirilerek en etkili yürüyüş kombinasyonu belirlenmeye çalışılmıştır. GEI, literatürde sık kullanımı ve yüksek verimliliği nedeniyle birincil modalite olarak seçilmiştir. GEI ile kombinleyecek diğer iki modalite; silüetler ve optik akışlar literatürde GEI' den sonra en sık kullanılan ve başarısı ispatlanan yürüyüş desenleri olmuştur. Hazırlanan GEI baş-bacak bölgelerinin birleştirilmiş görüntüsü ise insan vücudunun yatay bölgelere ayrılarak, özellikle en etkili yatay vücut parçasını tespit etmeye yönelik yapılmış önceki çalışmalardan yola çıkılarak hazırlanmıştır. Rida ve diğerleri (2015) tarafından yapılan bir çalışmada, insan vücudu dört bölgeye ayrılarak hareket temelli 'Grup Lasso' yöntemi ile en ayırt edici yatay vücut bölgeleri seçilmiştir. Rokanujjaman ve diğerlerinin (2012) çalışmasında ise insan vücudu beş bölgeye ayrılmış, tanıma sürecinde baş, bel ve bacak bölgelerinin olumlu ya da olumsuz etkilerine bağlı her bölgenin katkı oranı araştırılmıştır. Choudhury ve Tjahjadi (2015), GEI 'nin bacak bölgesinin daha fazla ayırt edici bilgi içerdiği, çanta taşıma ve ceket giyme varyasyonlarından en az etkilendiği bölge olduğu sonucuna varılmışlardır. Bashir ve diğerleri (2010), GEI' de bulunan dinamik yürüyüş bilgisi ile statik şekil bilgisini ayırt edebilmek amacıyla GEI desenlerine bir özellik seçim maskesi uygulanmış, sonuç görüntülerinde en dinamik bölgelerin kol ve bacak olduğu sonucuna varmışlardır. Başka bir çalışmada (Ghebleh & Ebrahimi Moghaddam, 2018), yürüyüş tanıma üzerinde giyim varyasyonunun etkisini azaltmak amacıyla bir adaptif aykırı bölge tespit yöntemi önerilmiştir. Bu yaklaşım ile silüetlerdeki kıyafet farklılığından etkilenen bölgelerin tespit edilip elenmesi hedeflenmiştir. Önceki çalışmalar doğrultusunda,

giyim ve taşıma varyasyonlarındaki değişkenlikler göz önünde bulundurulduğunda, hazırlanacak yeni modalite için yalnızca iki vücut bölgesi seçilmiştir; baş ve bacak bölgeleri. Bu bölgelerin seçimi, bacak bölgesinin yüksek ayırt ediciliği ile baş bölgesinin tamamen örtülüp gizlenme olasılığının düşük olması dikkate alınarak yapılmıştır. Sonrasında, GEI' ye ait baş ve bacak bölgeleri tek bir görüntüde birleştirilerek, görünüm tabanlı varyasyonlara karşı daha dayanıklı bir modalite elde edilmeye çalışılmıştır. Ayrıca, baş ve bacak bölgelerinin ayrı bir kombinasyonu da çoklu modalite ağları üzerinden değerlendirilmiştir.

Şekil 21, önerilen yürüyüş tanıma çerçevelerini göstermektedir. Şekilde, çoklu modaliteye sahip ağ yapılarının iki ayrı koluna iki farklı modalitenin verildiği gösterilmektedir. Bu modaliteler; GEI ve silüetler (SH), GEI ve optik akışlar (OF), GEI ve baş-bacak bölgelerinin birleştirilmiş görüntüsü (HConL) ile yalnızca bacak (Leg) ve baş (Head) bölgeleri olup, sırasıyla Şekil 21 (a), (b), (c) ve (d)'de sunulmuştur. Şekil'de hazırlanan her bir mimaride dalları oluşturan yapılar, aynı özelliklere sahip iki adet ince ayarlanmış CNN mimarisinden oluşmaktadır. Bu mimariler, tek bir modaliteye dayanmak yerine, birkaç modalitenin birleşiminden yararlanmaktadır ve verilen modalitelerden öznetelik çıkarmaktadır. Ardından, farklı modalitelerden elde edilen öznetelikler, birleştirme (concatenate) işlemi ile bir araya getirilmiştir. Bu işlem sırasında, boyutu  $n$  olan iki öznetelik vektörü dikey olarak birleştirilmekte ve sonuç olarak, boyutu  $2n$  olan bir nihai öznetelik vektörü elde edilmektedir. Ağ dallarında başta EfficientNet-B0, MobileNet-V1 ve ConvNeXt-base mimarileri olmak üzere VGG16 ile ResNet50 mimarileri de ince ayar uygulanmış CNN yapıları olarak kullanılmıştır. EfficientNet, MobileNet ve ConvNeXt, ImageNet veri kümesi üzerinde test edilmiş ve mevcut en gelişmiş CNN modelleriyle karşılaştırıldığında yüksek doğruluk sonuçları elde etmişlerdir. Özellikle ConvNeXt mimarisi daha önceki bölümlerde detaylandırıldığı gibi son teknoloji transformer mimarilerinden daha yüksek bir doğruluk oranına ulaşmıştır. EfficientNet-B0, RGB video kareleri kullanılarak yürüyüş tanıma alanında etkileyici bir performans sergilemiş olsa da (Khan vd., 2023) silüet verisi ile olan değerlendirmelerde hâlen bir boşluk bulunmaktadır. Ayrıca, MobileNet (Liu vd., 2024) ve ConvNeXt (Fu vd., 2023) mimarilerinin hareket tanıma üzerindeki başarıları literatürde tartışılmış olsa da yürüyüş tanıma performansları henüz değerlendirilmemiştir. Bu nedenle, kapsamlı bir performans analizi gerçekleştirmek amacıyla EfficientNet-B0, MobileNet-V1 ve ConvNeXt-base olmak üzere üç farklı CNN tabanlı ağ modeli karşılaştırmalı analiz için seçilmiştir.



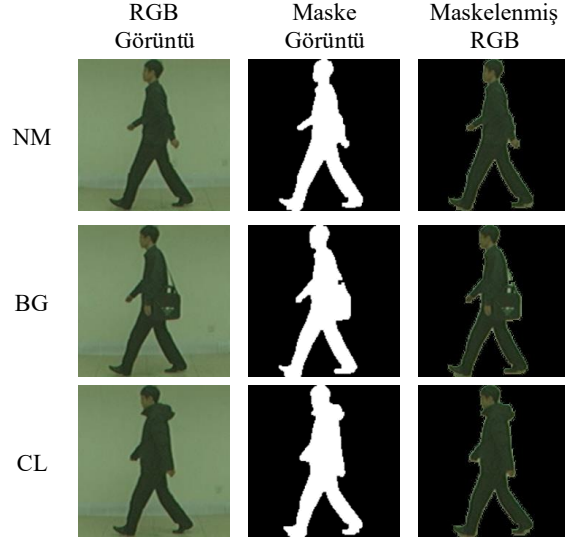
Şekil 21. Çoklu modalite mimarileri (a) GEI ve SH, (b) OF, (c) HConL, (d) Leg ve Head

Modalitelerin tanıma performansı üzerindeki etkileri değerlendirebilmek amacıyla, çoklu modalite yaklaşımına dayalı olarak tüm farklı modalite kombinasyonları derin CNN mimarilerine verilmiştir. Tüm CNN ağlarının iki ayrı evrişimli dalı, iki farklı modaliteden öznitelik çıkarmak için kullanılmış, ardından bu öznitelikler birleştirilerek tek bir öznitelik vektörü elde edilmiştir. Bu yaklaşım, ağın iki ayrı evrişimli dalının eşzamanlı olarak eğitilmesine olanak tanımakta ve farklı modalite kombinasyonlarının çapraz açılı ve görünüm tabanlı varyasyonlar altındaki başarısını değerlendirmeye olanak tanımaktadır. Ayrıca, birleştirme işlemi daha zengin ve daha kompakt bir yürüyüş temsili sağlamaktadır. Bununla birlikte, farklı CNN mimarilerinin kullanılması, bu ağların yürüyüş tanıma performanslarının eşzamanlı olarak karşılaştırılmasına da olanak sağlar.

Transfer öğrenme süreci için, ağ öncelikle ImageNet veri kümesi üzerinde eğitilmiş olan ön eğitilmiş modelin ağırlıklarıyla başlatılmaktadır. Daha sonra, ağ yürüyüş tanıma veri setleri üzerinde ince ayar işlemine tabi tutulmaktadır. Transfer öğrenme, ağın daha verimli kullanılmasını ve yüksek performanslı bir eğitimin gerçekleştirilmesini sağlamaktadır. Öyle ki eğitim, oldukça büyük bir veri kümesi olan ImageNet üzerinde gerçekleştirilmiştir.

Çoklu modalite ağ mimarilerinde performansı değerlendirilecek son girdi verisi olarak RGB yürüyüş çerçeveleri seçilmiştir. RGB görüntülerin doğrudan model girişinde kullanılması, kıyafet ve arka plan gibi öğelerin öğrenme sürecine müdahil olmasına yol açarak tanıma performansını etkileyebileceğinden, bu yaklaşım literatürde sık rastlanan bir yöntem değildir. Bu sebeple RGB yürüyüş çerçeveleri silüetler ve optik akışlarda olduğu gibi bir hizalama işlemine tabi tutulduktan sonra arka planı elimine etmeye yönelik bir ön işlemin ardından GEI ile olan kombinasyonun başarısı diğer modalitelerden ayrı olarak değerlendirilmiştir.

Şekil 22' de CASIA-B veri setinden alınmış, arka planı eleme sürecine tabi tutulmuş farklı varyasyonlar altında örnek bir RGB yürüyüş görüntüsü, ikili maskesi ve sonuç görüntüsü sunulmuştur. Şekil' de, hizalanmış RGB görüntüler üzerine doğrudan vücut maskelerinin (hizalanmış ikili silüetlerin) uygulanmasıyla maskelenmiş yürüyüş görüntüleri elde edilmiştir. Bu eleme işlemiyle arka planın ve aydınlatmanın öğrenmeye etkisi azaltılmaya çalışılmıştır. Ardından sonuç maskelenmiş RGB yürüyüş görüntüleri, GEI ile diğer modalitelerde olduğu gibi çoklu modalite ağ mimarileri üzerinden kombinleyerek tanıma başarısına etkisi değerlendirilmiştir.



Şekil 22. Hizalanmış RGB görüntüsü, ikili maskesi, maskelenmiş RGB görüntüsü

### 2.3. Topluluk Öğrenmeli Parça-Tabanlı Ağ Mimarisi

Bu bölümde, bir önceki bölümde detaylandırılan çoklu modalite ağ mimarilerinden farklı olarak global bir tanıma yaklaşımı yerine, yerel vücut bölgeleri üzerinde topluluk öğrenmesinden yararlanarak geliştirilmiş bir parça tabanlı yürüyüş tanıma yaklaşımı önerilmektedir. Mimaride ilk olarak GEI beş yatay parçaya bölünmekte ve her bir parçayı işleyen CNN mimarisine topluluk öğrenmesi uygulanmaktadır. Her bir vücut parçası için, topluluk öğrenmesiyle elde edilen parça tabanlı öznitelikleri birleştirmek ve en ayırt edici olanları sentezlemek amacıyla ayrı bir Toplayıcı Model (TM) eğitilmektedir. Ayrıca, görünüme dayalı değişimlerin etkisini azaltmak amacıyla, varyasyon içeren ve içermeyen görüntüler arasındaki mutlak farkları analiz eden bir parça eleme süreci geliştirilmiştir. Mimaride elde edilen en ayırt edici özniteliklerin bir araya getirilmesi, sağlam ve etkili bir tanıma performansı sağlamaktadır.

Yürüyüş tanımda sıklıkla görünüş değişimlerine karşı hassas olan vücut parçası detaylarını göz ardı edebilen küresel (global) verilere dayalı yürüyüş temsillerine odaklanılmıştır. Ancak son zamanlarda literatürde önerilen birçok çalışmada yürüyüş özelliği temsiline dayanıklılığını artırmak amacıyla küresel öznitelikler yerine yerel (lokal) öznitelikler tercih edilmiştir. Örneğin, Fan ve diğerleri (2020) tarafından önerilen GaitPart modelinde odaklanmış evrişimli katman kullanılarak parça düzeyindeki uzaysal öznitelikleri öğrenebilen bir FPFE ile yerel mikro hareket özelliklerini modelleyen bir MCM kullanılmıştır. Zhang ve diğerleri (2019), silüetlerin dört yatay parçasından ayırt edici

öznitelikleri çıkarmak üzere zamansal dikkat modeli önermiş, her bir yerel vücut bölgesi için ayrı ayrı hibrit CNN-LSTM mimarisi eğitmişlerdir. Lin ve diğerleri (2021), yürüyüş çerçevelerinden hem küresel hem de yerel bilgileri kullanarak ayırt edici temsiller elde etmek amacıyla 3B CNN tabanlı ‘GaitGL’ adlı bir yapı sunmuştur. Küresel ve Yerel Özellik Çıkarıcı (Global and Local Feature Extractor, GLFE) modülü ile küresel ve yerel özellikleri prensipli bir şekilde birleştirmişlerdir. Ayrıca mekânsal bilgiyi daha iyi korumak için Yerel Zamansal Toplama (Local Temporal Aggregation, LTA) adlı yeni bir işlem geliştirmişlerdir. Sepas-Moghaddam ve Etemad (2020), çerçeve düzey evrişimli özelliklerden Yürüyüş Evrişim Enerji Haritaları (Gait Convolutional Energy Maps, GCEM) çıkaran bir model önermiştir. Bu model, GCEM’in bölünmüş bölümlerinden elde edilen farklı kısmi temsiller arasındaki ilişkileri öğrenmek için dikkat yinelemeli bir model kullanmıştır. Başka bir çalışmada (Sepas-Moghaddam vd.,2021), daha derin parça bütün ilişkilerini öğrenmek üzere bir kapsül ağı hazırlanmıştır. Bu ağ, çok ölçekli kısmi vücut temsillerini elde ettikten sonra, ileri ve geri yönlerde İki Yönlü Kapılı Tekrarlayan Birimler (Bidirectional Gated Recurrent Units, BGRU) kullanarak desenler arasındaki korelasyonları ve birlikte görülme durumlarını öğrenmektedir. Li ve diğerleri (2022), üstten alta dilimlenmiş ve birbiriyle ilişkili bölme özniteliklerini oluşturmak için SED ve zaman içindeki önemli çerçeveleri vurgulamak için RFAM modülü kullanan GaitSlice modelini sunmuştur. Hazırlanan mimari ile insan yürüyüşünü uzamsal-zamansal bölme özelliklerine dayalı olarak analiz etmişlerdir. Ma ve diğerleri (2023), yürüyüş özelliklerinin bir dizi yerel hareket deseninden oluşan çoklu anahtar bölgelerdeki küresel hareket desenlerini içerdiğini öne süren yeni bir çerçeve önermiştir. Komşu pikseller arasındaki özellikler arasında dinamik bir dikkat mekanizması geliştirerek hem anahtar bölgelere uyarlanabilir biçimde odaklanmayı hem de daha etkileyici yerel hareket desenleri üretmeyi başarmışlardır. Chen ve diğerleri (2023), hem küresel hem de yerel düzeyde vücut hareket tanımlayıcılarını çıkarmak için bütünsel ve kısmi zamansal toplama stratejisi kullanan GaitAMR adlı bir çerçeve önermiştir. Görüş açısı domaininde, GaitAMR en uygun görüş açısından seçilen özellikleri uzamsal-zamansal alan için tamamlayıcı bilgi olarak kullanmaktadır. Zaman domaininde ise hem küresel hem de yerel düzeyde hareket yörüngesi boyunca zengin ve benzersiz zamansal ilişkileri yakalamaya odaklanmaktadır. Wei ve diğerleri (2024), Çok Ölçekli Özellik Çıkarıcı (Multi-scale Feature Extractor, MSFE) ve Parça Tabanlı Yatay Eşleme (Part-based Horizontal Mapping, PHM) olmak üzere iki ana modülden oluşan GMSN modelini sunmuştur. MSFE, farklı ölçeklerde yerel ve küresel bilgileri yakalayıp daha kapsamlı bir yürüyüş temsili sağlayan paralel çok

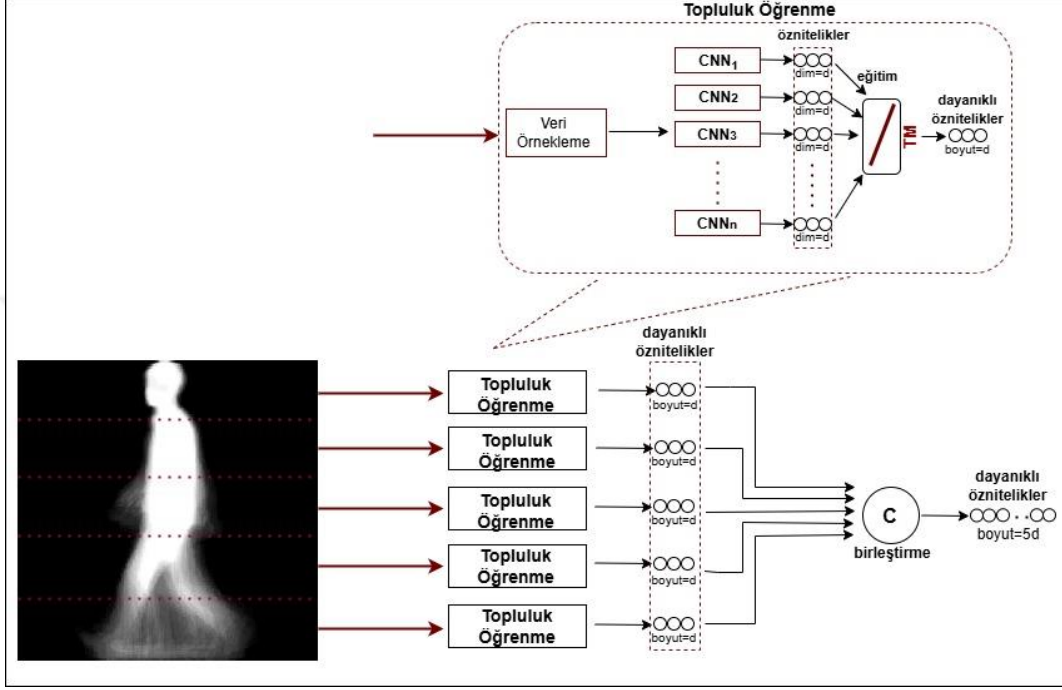
ölçekli CNN'ler kullanırken, PHM ise net vücut hatlarını ve hareket desenlerini vurgulayarak kritik vücut bölgelerinin öğrenilmesini geliştirmektedir. Başka bir çalışmada Pan ve diğerleri (2024), kenar detay bilgilerini koruyarak ve bağlamsal korelasyonlardan yararlanarak yerel detaylı yürüyüş özellikleri çıkaran Yerel İlişki Evrişim (Local Relation Convolutional, LRConv) katmanlarını ve kıyafet kaynaklı örtüşmelerden daha az etkilenen yürüyüş parçalarına odaklanan İnsan Vücut Odaklama (Human Body Focusing, HBF) modülünü önermiştir. Xiong ve diğerleri (2024), silüet ve iskelet modalitelerinin tamamlayıcı bilgilerini çıkarmak için çok modelli birleşik öğrenmeye dayalı bir ağ önermiştir. İnsan vücudu parçalarını ve eklemler arası hareket bağlantılarını grafik olarak soyutlamışlardır.

Yukarıda detaylandırılan yerel özniteliklere odaklanan çalışmaların temel hedefinin, küresel ve yerel özellikleri bir arada daha etkili bir şekilde çıkararak veya çıkarılan yerel özellikler arasındaki ilişkileri anlamlı biçimde modelleyerek yürüyüş tanıma başarımını en üst düzeye çıkarmak olduğu gözlemlenmektedir. Bu tez kapsamında hazırlanan ağ mimarisinde ise yerel bölgelerden en etkili ve sağlam öznitelikleri yeni bir şekilde çıkarmak amacıyla topluluk öğrenme yöntemi GEI örüntüsü üzerinde denenmiştir. İlk olarak, tanıma için yatay GEI vücut parçaları kullanılmakta, ardından her yerel GEI vücut bölgesinden sorumlu CNN'lere topluluk öğrenme uygulanarak ağın genel başarımı artırılmaktadır. Buna ek olarak, görünüş tabanlı varyasyonlardan etkilenen bölgeleri ortadan kaldırmak amacıyla bir parça-eleme işlemi önerilmiştir.

Çalışma kapsamında girdi verisi olarak GEI' nin seçilmesinde en büyük etken derin öğrenme mimarileriyle çalışırken, her bir silüetin ayrı ayrı işlenmesinin yüksek hesaplama maliyeti ve bellek kullanımı gerektirmesi ve sonuç olarak eğitim sürecinin ciddi ölçüde zorlaşması olmuştur. GEI' nin kullanımı, bu yükü azaltarak yürüyüş döngüsünü tek bir görüntüde özetleme avantajı sağlamıştır. Ayrıca, literatürde GEI' nin birçok veri kümesinde yüksek doğruluk oranlarına ulaştığı ve farklı bireylerin ayırt edilmesinde etkili bir temsil sunduğu gösterilmiştir. Bu nedenle hem hesaplama verimliliği sağlamak hem de tanıma performansından ödün vermemek amacıyla GEI temsili tercih edilmiştir.

İnsan vücudunun farklı bölümleri, yürüyüş sırasında kendine özgü morfolojik özellikler ve hareket desenleri sergilerler. Bu nedenle, her bir vücut bölgesinin kendine özgü bir temsil ile ele alınması gerekmektedir (Fan vd., 2020). Tez kapsamında önerilen mimaride GEI yerel bölgelere odaklanmak amacıyla yatay olarak beş parçaya ayrılmıştır. Topluluk öğrenmesi, farklı öznitelik temsillerinden yararlanarak bireysel modellerin sınırlılıklarını azaltırken genelleme yetisini artırır. Bu bağlamda, vücut bölgesi temsilleri, topluluk

öğrenme aracılığıyla daha sağlam ve ayırt edici hâle getirilebilir. Bu temsillerin birleştirilmesiyle modelin ayrıştırma gücü artar ve daha yüksek tanıma performansı elde edilir. Tez kapsamında hazırlanan Topluluk Öğrenmeli Parça-Tabanlı mimari Şekil 23' te sunulmuştur.



Şekil 23. Topluluk Öğrenmeli Parça-Tabanlı Ağ Mimarisi

Şekil 23' te gösterildiği üzere, GEI öncelikle yatay olarak beş parçaya ayrılmaktadır. Her bir parça, kendisine özel bir CNN modeli tarafından işlenmekte ve bu modeller, birbirinden bağımsız olarak topluluk öğrenmesi sürecine tabi tutulmaktadır. Her bir topluluk öğrenmesi sürecinde,  $n$  adet veri örneği oluşturulmakta ve her bir parça için  $n$  adet homojen CNN modeli eğitilmektedir. Bu CNN modellerinin her biri,  $n$  farklı  $d$ -boyutlu öznitelik vektörü üretmektedir. TM  $n$  adet  $d$ -boyutlu öznitelik vektörünün birleştirilmesinde kullanılmaktadır. Öznitelik birleştirme, birden fazla öznitelik temsilini tek bir daha bilgilendirici temsil altında bütünleştirerek tanıma veya sınıflandırma görevlerinin başarımını artırmayı hedefler. Önerilen TM modeli, farklı bağımsız modellerden çıkarılan öznitelik vektörlerini birleştirmek için tam bağlantılı bir katman kullanmaktadır. Farklı kaynaklardan gelen özniteliklerin bir araya getirilmesiyle, model tamamlayıcı bilgileri etkili biçimde yakalayarak daha ayrıştırıcı bir temsil elde etmektedir. Tam bağlantılı katman aracılığıyla uygulanan doğrusal dönüşüm, modelin optimal özellik eşlemelerini öğrenmesine

olanak tanırken, temel karakteristikleri koruyup gereksiz tekrarları azaltmaktadır. Geleneksel topluluk öğrenme yöntemlerinin çoğunluk oylaması veya ağırlıklı ortalama gibi yaklaşımlara dayanmasının aksine, bu yöntem doğrudan birleştirilmiş öznitelik uzayını iyileştirmeye odaklanmakta ve bu sayede daha sağlam kararlar alınmasını sağlamaktadır. Son olarak, beş farklı vücut bölgesine ait sağlam ve zenginleştirilmiş temsiller öznitelik seviyesinde birleştirilerek tanıma işlemi gerçekleştirilmektedir. Tablo 3’ te Topluluk öğrenmeli parça-tabanlı ağ mimarisi’ nin detaylı algoritma adımları sunulmuştur.

Tablo 3. Topluluk öğrenmeli parça-tabanlı yürüyüş tanıma algoritması

<b>Algoritma</b>	
Eğitim:	<p>Girdi: GEI yürüyüş temsili  Çıktı: <i>rank-1</i> doğruluğu  1: <i>Parça eleme işleminin gerçekleştirilmesi.</i> Görünüme dayalı varyasyonlardan etkilenen parçalar, parça eleme işlemi ile eğitim aşamasının dışında tutulur.  2: <i>GEI’ nin beş yatay parçaya ayrılması.</i>  3: <i>Veri örnekleme.</i> Her yatay parça için <math>k</math>-katlı çapraz doğrulama veya önyükleme toplulaştırması ile <math>n</math> veri örnekleme oluşturulur.  4: <i>Topluluk öğrenme mimarisinin eğitilmesi.</i> Her bir parça için <math>n</math> adet homojen CNN eğitilir.</p>
Test:	<p>5: <i>TM’ nin eğitilmesi.</i> Homojen <math>n</math> adet CNN’ nin çıktıları üzerinden tek bir lineer katmana sahip TM eğitilir.  6: <i>Özniteliklerin elde edilmesi.</i> Tüm parçalar için TM tarafından üretilen en ayırt edici öznitelikler elde edilir.  7: <i>Birleştirme.</i> Tüm parça tabanlı öznitelikler bir araya getirilir.  8: <i>Doğruluk hesaplama.</i> İlgili tüm açılar için <i>rank-1</i> değeri hesaplanır.</p>

Tablo 3’ teki algoritmanın ilk adımı, parça eleme sürecini içermektedir. Görünüme dayalı varyasyonlardan etkilenen vücut parçaların eğitim sürecinden hariç tutulması, sınıflayıcıların genelleme performansını artırmasına olanak tanımaktadır. Ardından GEI, yatay olarak beş parçaya bölünmekte ve her bir parça bağımsız şekilde işlenmektedir. Veri örnekleme aşamasında, her yatay parça için çeşitliliği sağlamak amacıyla  $k$ -katlı çapraz doğrulama veya önyükleme toplulaştırması yöntemleri kullanılmaktadır.  $k$ -katlı çapraz doğrulama yönteminde, veri kümesi eşit büyüklükte  $k$  parçaya bölünmekte ve her seferinde  $k-1$  parça ile eğitim yapılırken geri kalan bir parça test için kullanılmaktadır. Algoritmada,  $k$  değeri veri örnekleme sayısı olan  $n$ ’ e karşılık gelmektedir. Diğer yandan, önyükleme toplulaştırması, yer değiştirmeli rastgele örnekleme yöntemi ile  $n$  adet eğitim veri örnekleme oluşturulmasını sağlar. Bu yöntem sayesinde bir veri örneği birden fazla eğitim alt

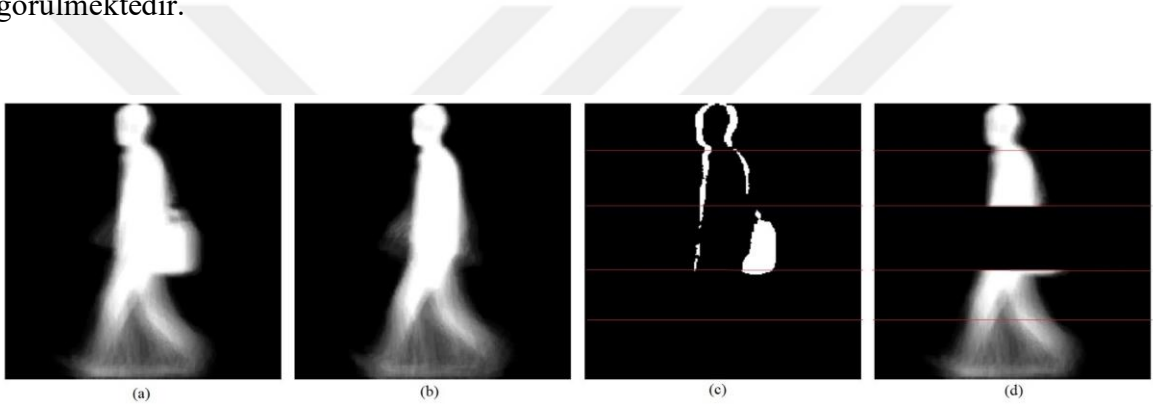
kümesinde yer alabilir. Daha sonra topluluk öğrenmesi çerçevesi uygulanmakta ve her yatay parça için  $n$  adet  $d$ -boyutlu öznitelik çıkarımı yapan CNN modelleri eğitilmektedir. Bu CNN'ler mimari olarak birbirinin aynısı olmakla birlikte farklı veri örnekleriyle eğitilmiştir. Test aşamasında, CNN'lerden elde edilen  $n$  farklı  $d$ -boyutlu çıktı, TM' yi eğitmekte kullanılır. TM, giriş boyutu  $n \times d$  ve çıkış boyutu  $d$  olan tek katmanlı bir lineer yapıdır ve bu katman, çıkarılan öznitelikleri optimize eder. Her parça için TM tarafından üretilen en ayırt edici öznitelikler öznitelik seviyesinde birleştirilerek daha güçlü bir temsil elde edilir. Elde edilen parça tabanlı temsillerin birleştirilmesiyle kapsamlı bir yürüyüş temsili oluşturulur ve bu temsil, her farklı görüş açısı için  $rank-1$  doğruluk değerinin hesaplanmasında kullanılır. Algoritma, yerel yürüyüş özniteliklerini etkili bir şekilde yakalarken, topluluk öğrenme yönteminden faydalanarak sınıflandırma başarımını artırmaktadır.

Hazırlanan topluluk öğrenme tabanlı mimaride, CNN modelleri için temel ağ olarak ConvNeXt-base mimarisi kullanılmıştır. Bunun sebebi bu mimarinin Çoklu Modalite Ağ Mimarisi üzerinden yapılan deneylerde elde edilen başarımların değerlendirilmelerinin sonucundan kaynaklanmaktadır. Ayrıca, performansı daha da artırmak amacıyla transfer öğrenme uygulanmaktadır. Veri örnekleme sürecinde  $n$  parametresi 5 (beş) olarak belirlenmiş olup bu değer, tek bir yatay parça için eğitilen toplam sınıflayıcı sayısını ifade etmektedir. Bununla birlikte,  $d$  boyutu 1024 olarak alınmış ve bu değer, ConvNeXt'in son katmanından elde edilen çıktı boyutunu temsil etmektedir.

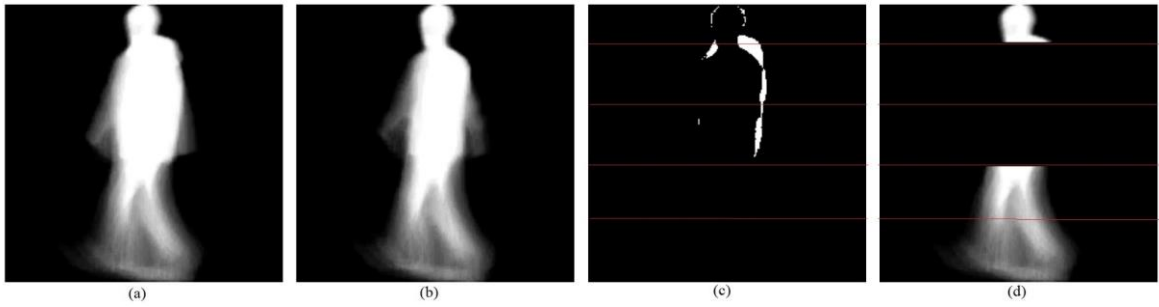
### 2.3.1. Parça Eleme İşlemi

Parça tabanlı yürüyüş tanıma, belirli vücut bölgelerinin kendine özgü hareket özellikleri sergilediği, kimlik belirlemede farklı düzeylerde katkı sağladığı ve gizlenmeleri durumunda tanıma performansında düşüşe neden olduğu gözlemlerine dayanarak daha önce birçok çalışmada ele alınmıştır. Bu doğrultuda, daha önceki bölümlerde detaylandırılan çalışmalarda (Rida vd., (2015); Rokanujjaman vd., (2012); (Choudhury & Tjahjadi, 2015); Bashir vd., (2010)) her bir yatay vücut bölmesinin ayrı ayrı tanımaya etkileri detaylıca araştırılmıştır. Bu çalışmalardan özellikle Ghebleh ve Ebrahimi Moghaddam (2018) tarafından önerilen yöntemde, galeri ve sorgu görüntülerinin her bir satırı için adaptif bir eşik değeri belirlenerek satırların geçerli veya geçersiz olarak değerlendirilmesini sağlayan adaptif bir aykırı değer tespiti yöntemi önerilmiştir. Bu yöntemde, giyim farklılıklarından etkilenen vücut bölgeleri tanıma sürecinden hariç tutulmuştur. Tez kapsamında da,

görünüme dayalı varyasyonların tanımayaya etkisini azaltmak amacıyla, GEI deseni öncelikle beş yatay parçaya ayrılmıştır. Ardından, giyim ve taşıma kaynaklı varyasyonlardan etkilenen parçalar tespit edilerek görüntüden çıkarılmıştır. Tanıma işlemi ise yalnızca varyasyondan etkilenmemiş, kalan vücut bölgeleri üzerinden gerçekleştirilmiştir. Parça eleme işlemi ile görünüme dayalı varyasyonlardan deforme olmuş vücut parçaları eğitim ve test aşamalarından hariç tutularak eğitimin daha istikrarlı gerçekleşmesi ve modelin sağlamlığını ile güvenilirliğini artırmak hedeflenmiştir. Şekil 24 ve Şekil 25' te, CASIA-B veri kümesinden alınmış görüntülerde sırasıyla taşıma ve giyim varyasyonları altında gerçekleştirilen parça eleme işlemlerine ait iki örnek sunulmuştur. Şekil 24' te tek bir parçanın tespit edilerek çıkarıldığı, Şekil 25' te ise iki parçanın tespit edilerek çıkarıldığı görülmektedir.



Şekil 24. Taşıma varyasyonu için parça eleme işlemi, soldan sağa sırasıyla varyasyonlu görüntü, varyasyonsuz görüntü, mutlak fark görüntüsü, parça elenmiş görüntü

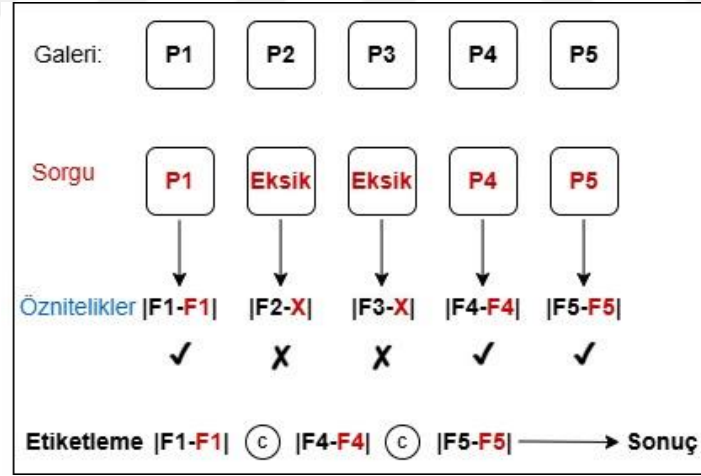


Şekil 25. Giyim varyasyonu için parça eleme işlemi, soldan sağa sırasıyla varyasyonlu görüntü, varyasyonsuz görüntü, mutlak fark görüntüsü, parça elenmiş görüntü

Şekil 24 ve Şekil 25'te, öncelikle varyasyonlu GEI ile aynı kişinin varyasyon içermeyen GEI deseni arasındaki mutlak fark görüntüsü hesaplanmaktadır. Elde edilen fark görüntüsü beş yatay parçaya bölünmekte ve her bir parça için varyasyon türüne bağlı olarak

manuel belirlenen eşik değerleri kullanılmaktadır (örneğin; deneysel çalışmada CASIA-B veri kümesindeki BG varyasyonu beş parça için eşik (th) değerleri sırasıyla th1:1100, th2:500, th3:400, th4:500 ve th5:500 olarak belirlenmiştir). Eşik değeri, yerel bölgedeki piksel değerlerinin toplamına dayalı olarak hesaplanmaktadır. Eşik değerini aşan parçalar, varyasyondan en çok etkilenen bölgeler olarak değerlendirilmekte ve bu parçalar tamamen siyah bölümlerle değiştirilerek analizden çıkarılmaktadır. Ardından tanıma işlemi yalnızca etkilenmeyen vücut parçaları kullanılarak gerçekleştirilmektedir. Örneğin, Şekil 24’ te dört vücut parçası üzerinden tanıma gerçekleştirilirken, Şekil 25’ te üç vücut parçası üzerinden gerçekleştirilir.

Parça eleme işleminin tanıma aşamasında, görünümü dayalı varyasyonlarda, her bir yatay parça varlığı açısından ayrı ayrı değerlendirilmiştir. Bir GEI deseninin parça eleme işlemi sebebiyle mevcut olmayan bir yatay parçası bulunması halinde, o parçaya yönelik herhangi bir öznitelik çıkarımı gerçekleştirilmemiş, bunun yerine mevcut olan yatay parçalardan öznitelikler çıkarılarak tanıma işlemi yalnızca bu öznitelikler üzerinden sürdürülmüştür. Bu sürece ilişkin bir örnek Şekil 26’ da sunulmuştur.



Şekil 26. Görünüme dayalı varyasyonlar için parça eleme işlemli tanıma süreci örneği

Şekil 26, GEI yürüyüş desenine ait bazı yatay parçaların eksik olduğu bir durumda gerçekleştirilen tanıma sürecini gösterilmektedir. Şekilde, galeri görüntüsü beş yatay parçadan (P1–P5) oluşurken, bazı parçaların parça eleme işlemi nedeniyle eksiltilmesinden kaynaklı sorgu görüntüsü yalnızca üç yatay parçadan (P1, P4, P5) oluşmaktadır. Öznitelik çıkarımı yalnızca mevcut olan bu üç parça için gerçekleştirilmiş ve eksik parçalara ait öznitelikler dikkate alınmamıştır. İlk olarak, mevcut galeri ve sorgu GEI parçalarından, en

ayırt edici parça tabanlı öznitelikler çıkarılmıştır. Elde edilen tüm öznitelikler, sorgu görüntüsündekilerin varlığına dayanarak galeridekilerle karşılaştırılmakta, başarılı ve başarısız eşleşmeler belirlenmiştir (şekilde başarılı eşleşmeler onay işaretiyle (✓), başarısız eşleşmeler ise çarpı işaretiyle (✗) temsil edilmiştir). Ardından, sorgu görüntüsünde eksik olan parçalara karşılık gelen galeri öznitelikleri (F2, F3) elenmiştir. Son aşamada, mevcut üç yatay parçaya ait öznitelikler birleştirilerek (c) tanıma işlemi gerçekleştirilmekte ve etiketleme süreci yalnızca başarılı eşleşen öznitelikler dikkate alınarak yapılmıştır.

## 2.4. Benzerlik Hesaplaması ve Değerlendirme Kriteri

Bu bölümde yürüyüş tanıma probleminin test aşamasının nasıl gerçekleştiği, kapalı küme ve açık küme terimlerinin açıklaması, test aşamasında kullanılan hesaplama yöntemi ve değerlendirme parametresi detaylandırılmıştır.

### 2.4.1. Kosinüs Benzerliği

Yürüyüş tanıma problemi iki farklı tanıma kümesi problemi üzerinde yoğunlaşmıştır; kapalı küme ve açık küme. Kapalı küme problemine odaklanan yöntemler eğitim aşamasında kayıt altına alınmış ve etiketlenmiş olan bilinen kişilerin yürüyüşlerini tanımak üzere eğitilmektedir. Ancak açık küme tanınmanın amacı, eğitim sürecinde kayıtlı olmayan ve daha önce görülmemiş sınıflara karşı başarılı bir şekilde çalışabilmektir. Örneğin, eğer sistem yalnızca 24 bireyin verileriyle eğitildiyse, test aşamasında sisteme yeni bir birey sunulduğunda bu birey derin öğrenme algoritması tarafından yanlışlıkla bu 24 kişiden biri olarak tanınacaktır. Bu durum, derin öğrenme ağlarının kapalı doğasından kaynaklanmaktadır (Alotaibi & Mahmood, 2017). Ancak, yürüyüş tanıma problemine yönelik yapılan son teknoloji yöntemler sıklıkla açık küme tanıma problemine odaklanmışlardır. Bu tez kapsamında, literatürdeki kullanım sıklığına uygun olarak eğitim ve test kümeleri birbiriyle kesişmeyen, açık küme problemine yönelik bir tanıma işlemi gerçekleştirilmiştir. Bu amaçla CASIA-B veri setinin eğitim ve test kümeleri mevcut 124 kişiden ilk 74 kişi eğitim, kalan 50 kişi ise test kümesi olacak şekilde bölümlendirilmiştir. Test kümesi ise galeri ve sorgu kümeleri olarak yeniden ikiye ayrılmıştır. Galeri kümesi, test kümesi içerisindeki referans ve etiketleri bilinen kişilerden oluşurken; sorgu kümesi, tanınması gereken, etiketi bilinmeyen verilerden oluşmaktadır. CASIA-B veri seti için NM

#01-04 aralığındaki diziler galeri ve NM #01-02, BG #01-02 ile CL #01-02 dizileri sırasıyla sorgu kümeleri olarak alınmıştır. CASIA-B için yapılan eğitim ve test kümeleri bölümlendirmesi Tablo 4' te sunulmuştur.

Tablo 4. CASIA-B eğitim ve test küme bölümlenmesi

Eğitim kümesi	Test kümesi	
	Galeri kümesi	Sorgu kümesi
ID: 001-074	ID: 075-124	ID: 075-124
Dizi: NM #01-06, BG #01-02, CL #01-02	Dizi: NM #01-04	Dizi: NM #05-06, BG #01-02, CL #01-02

CASIA-C veri seti için eğitim aşamasında, ilk 24 ve 62 kişi eğitim kümesi olarak kullanılırken, kalan 53 kişi test kümesi olarak kullanılmıştır. Her denek için, nm dizileri galeri kümesi olarak belirlenirken, fs, fq ve fb dizileri sırasıyla sorgu kümeleri olarak belirlenmiştir. Outdoor-Gait veri seti test aşamasında eğitim kümesi için ilk 69 birey, test kümesi için kalan 69 bireyler kullanılmıştır. Hem galeri hem de sorgu kümeleri, veri setindeki her koşuldan en az iki video dizisi içerecek şekilde ayarlanmışlardır.

Test aşamasının son adımı olan tanıma sürecinde bir benzerlik hesaplaması yapılarak sınıflandırma işlemi gerçekleştirilmektedir. Sınıflandırma işlemi sırasıyla şu şekilde gerçekleştirilir; galeri ve sorgu kümelerine ait tüm görüntülerin öznitelikleri çıkarılır, sorgu kümesine ait görüntüler ile galeri kümesindeki görüntüler arasında bir benzerlik değeri hesaplanır, en yüksek benzerliği veren vektör çiftindeki galeri görüntüsünün etiketi sorgu görüntüsünün etiketi olarak kabul edilir. Tez çalışması kapsamında Kosinüs uzaklığı benzerlik hesaplaması için kullanılmıştır. Kosinüs uzaklığı,

$$\cos(g, p) = \frac{gp}{\|g\| \|p\|} = \frac{\sum_{i=1}^d g_i p_i}{\sqrt{\sum_{i=1}^d (g_i)^2} \sqrt{\sum_{i=1}^d (p_i)^2}} \quad (7)$$

biçiminde hesaplanır. Burada  $g$  galeri görüntüsünden elde edilen  $1 \times d$  boyutundaki öznitelik vektörü iken,  $p$  sorgu görüntüsünden elde edilen  $1 \times d$  boyutundaki öznitelik vektörüdür.

### 2.4.2. Rank-1 Doğruluk Oranı

Tez kapsamında yapılan tüm deneylerde, önerilen modellerin etkinliğini değerlendirmek için ortalama *rank*-1 doğruluk parametresi kullanılmıştır. *rank*-1 doğruluğu, özdeş açılar hariç, sorgu dizisindeki tüm görüntülerin galerideki tüm görüntülerle karşılaştırılması sonucunda, kimliği doğru şekilde belirlenen örneklerin oranı olarak tanımlanır. Ortalama *rank*-1 doğruluğu ise belirli bir açı için elde edilen tüm *rank*-1 doğruluk değerlerinin toplamının, açı sayısına bölünmesiyle hesaplanır. Örneğin, CASIA-B veri seti için bu sayı 11 (onbir)' dir.  $C_A$  ile gösterilmiş ortalama *rank*-1 doğruluk parametresi,

$$C_A = \frac{1}{11} \sum_{a=0}^{10} C_a \quad (8)$$

şeklinde hesaplanır. Burada  $C_A$ ,  $a$  görüş açısındaki özdeş açı değeri hariç *rank*-1 doğruluğudur.

### 3. BULGULAR VE TARTIŞMA

Bu bölümde, tez kapsamında hazırlanan, farklı modalitelerin ve ağ mimarilerinin kombinlemesini ve bunların kapsamlı bir performans değerlendirmesini sağlayan birinci Çoklu Modalite Ağ Mimarisi ile bu değerlendirmelerden yola çıkılarak tanımda en etkili bulunan modalite ve ağ çeşidinin kullanılmasıyla hazırlanan ikinci Topluluk Öğrenmeli Parça-Tabanlı Ağ Mimarisi' nin farklı laboratuvar ve dış ortam yürüyüş veri setleri üzerinden detaylı bir başarımlı değerlendirme yapılmıştır. Deneylerde Çoklu Modalite Ağ Mimarisi için Tensorflow' un Keras kütüphanesi kullanılırken, Topluluk Öğrenmeli Parça-Tabanlı Ağ Mimarisi' nde Pytorch kullanılmıştır. Mimarilerin eğitiminde kullanılan ince ayarlanmış tüm ağ yapıları Stokastik Gradyan Azalma (Stochastic Gradient Descent, SGD) optimizasyon algoritması, 0.0001 öğrenme oranı ve 0.9 momentum değeriyle kullanılmıştır. Ayrıca, çıkış katmanı softmax aktivasyon fonksiyonuna sahiptir ve kayıp fonksiyonu olarak çapraz entropi seçilmiştir. Tüm modaliteler 224×224 boyutlarına yeniden ölçeklendirilmiştir. Bellek ve zaman maliyetlerini azaltmak amacıyla, her bir silüet ve optik akış için eğitim verisi sayısı 30 (otuz) ile sınırlandırılmıştır. Bu doğrultuda, bu eğitim kümelerinden rastgele 30' ar örnek seçilerek kullanılmıştır. Modeller, Nvidia GeForce RTX 3060 GPU' ları kullanılarak eğitilmiş, deneyler Windows 10 ortamında gerçekleştirilmiştir.

#### 3.1. Çoklu Modalite Ağ Mimarisi Başarı Değerlendirmesi

Bu bölümde daha önceki bölümlerde detaylandırılan iki farklı modalitenin — GEI ve silüetler (SH), GEI ve optik akışlar (OF), GEI ve baş-bacak bölgelerinin birleştirilmiş görüntüsü (HConL), bacak (Leg) ve baş (Head) bölgeleri— ağın iki farklı kolundan verildiği Çoklu Modalite Ağ Mimarisinin CASIA-B ve Outdoor-Gait veri setleri üzerindeki *rank-1* doğruluk oranlarının kapsamlı analizi sunulmuştur. Kullanılan veri seti üzerinde adil bir karşılaştırma yapabilmek için, tüm kombinasyonların ağ dallarında özdeş olarak kullanılmak üzere özellikle son teknoloji üç farklı CNN mimarisi — EfficientNetB0, MobileNetV1, ConvNeXt-base — ile ek olarak, VGG16 ve ResNet50 ağları üzerinden başarımlı değerlendirilmesine tabi tutulmuşlardır. Bu kombinasyonlar sırasıyla;

- GEI' ler ve SH' ler:  $CNN_{GEI} + CNN_{SH}$ ,
- GEI' ler ve OF' ler:  $CNN_{GEI} + CNN_{OF}$ ,

- GEI' ler ve HConL' ler:  $CNN_{GEI} + CNN_{HConL}$ ,
- Head' ler ve Leg' ler:  $CNN_H + CNN_L$ ,
- GEI' ler ve RGB' ler: GEI+RGB ile GEI' ler ve maskelenmiş RGB' ler: GEI+MRGB

olacak şekilde adlandırılmışlardır.

### 3.1.1. CASIA-B Veri Seti Performans Değerlendirmesi

Bu bölümde CASIA-B veri seti üzerinden yapılan analizler detaylandırılmıştır. CASIA-B veri seti NM varyasyonu, özdeş açılı haric ortalama *rank-1* doğruluk oranlarının farklı mimariler üzerinden performans karşılaştırması Tablo 5' te sunulmuştur.

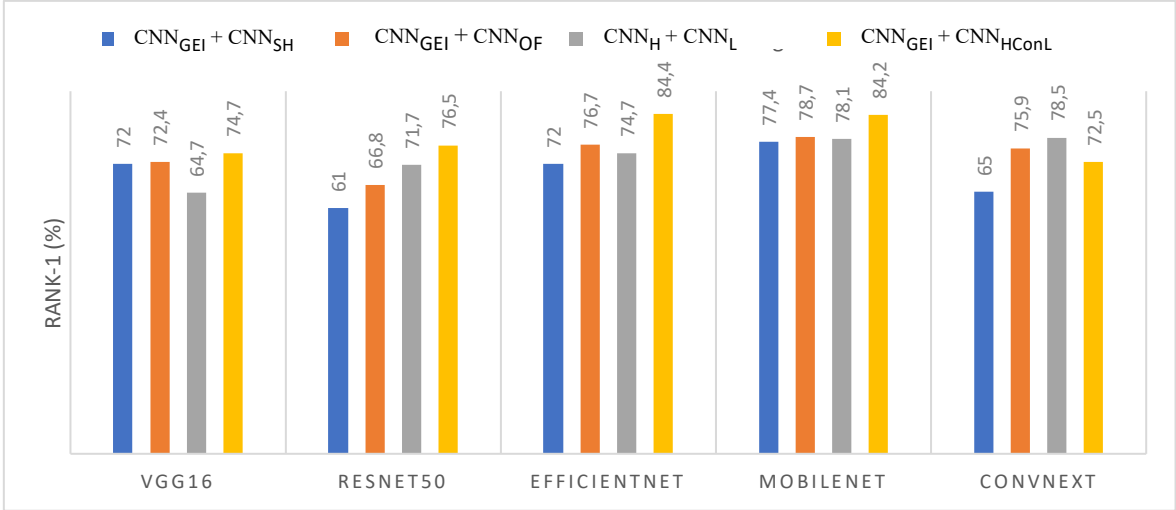
Tablo 5.  $CNN_{GEI} + CNN_{SH}$ ,  $CNN_{GEI} + CNN_{OF}$ ,  $CNN_H + CNN_L$  ve  $CNN_{GEI} + CNN_{HConL}$  kombinasyonlarının, CASIA-B veri seti NM varyasyonu, *rank-1* doğruluğu üzerinden karşılaştırması (özdeş açılar hariç)

Galeri NM#01-04		0° -180°										ort		
Sorgu		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°		
NM #05 -06	$CNN_{GEI}$	VGG16	54.8	72.8	73.9	73.8	68.8	65.8	70.1	73.3	75.1	74.7	54.3	68.8
		Resnet50	52.8	67.9	76.8	73.8	67.3	63.6	68.1	74.2	78.8	67.5	51.6	67.5
		EfficientNet	54.8	73.6	77.9	75.5	72.4	69.9	75.0	75.3	78.8	74.0	59.1	71.5
		MobileNet	59.3	79.4	83.4	85.2	76.9	72.3	80.0	84.3	84.6	80.9	60.2	76.9
		ConvNeXt	60.7	71.8	78.9	77.0	72.0	66.3	70.9	75.4	78.5	73.0	59.6	71.2
	$CNN_{GEI} + CNN_{SH}$	VGG16	60.6	76.2	78.2	75.9	72.3	67.3	71.4	76.0	77.9	76.9	60.3	72.0
		Resnet50	54.7	64.1	68.2	61.1	57.4	51.5	59.2	67.4	69.1	66.5	52.7	61.0
		EfficientNet	61.8	73.9	78.2	76.7	72.0	67.4	72.6	75.7	77.8	72.7	<b>63.2</b>	72.0
		MobileNet	<b>62.5</b>	<b>79.1</b>	<b>85.3</b>	<b>84.7</b>	<b>78.0</b>	<b>73.6</b>	<b>77.8</b>	<b>85.1</b>	<b>85.2</b>	<b>79.4</b>	61.0	<b>77.4</b>
		ConvNeXt	59.4	61.4	65.3	70.4	68.7	62.5	66.9	70.1	68.9	63.6	58.1	65.0
	$CNN_{GEI} + CNN_{OF}$	VGG16	60.5	76.0	76.7	75.8	71.7	69.2	72.5	75.9	77.9	80.0	60.2	72.4
		Resnet50	57.0	65.2	70.5	70.9	67.7	62.7	68.2	72.9	75.6	68.8	55.1	66.8
		EfficientNet	67.4	<b>79.0</b>	79.4	79.7	75.9	<b>75.0</b>	76.4	80.6	82.6	<b>80.1</b>	67.7	76.7
		MobileNet	67.5	78.8	<b>83.5</b>	<b>85.6</b>	<b>80.6</b>	74.1	<b>80.5</b>	<b>84.7</b>	<b>84.1</b>	79.4	66.5	<b>78.7</b>
		ConvNeXt	<b>70.9</b>	77.5	80.1	77.8	75.9	72.1	73.5	77.1	80.2	78.4	<b>71.2</b>	75.9
	$CNN_H + CNN_L$	VGG16	54.3	66.9	72.4	72.7	66.5	60.6	66.2	71.4	73.4	61.5	45.6	64.7
		Resnet50	57.4	73.5	81.7	80.0	71.6	68.7	73.8	77.6	75.5	71.2	57.9	71.7
		EfficientNet	62.7	80.0	83.6	78.7	74.5	71.7	76.4	79.8	82.1	73.2	59.1	74.7
		MobileNet	68.1	<b>82.2</b>	<b>87.6</b>	83.2	78.8	<b>74.8</b>	<b>79.1</b>	80.9	85.1	78.4	61.4	78.1
		ConvNeXt	<b>70.5</b>	80.8	85.9	<b>86.5</b>	<b>78.9</b>	74.0	76.7	<b>81.5</b>	<b>85.6</b>	<b>80.5</b>	<b>62.7</b>	<b>78.5</b>
$CNN_{GEI} + CNN_{HConL}$	VGG16	64.0	78.4	82.5	78.4	74.7	70.8	74.8	78.7	82.4	78.3	58.8	74.7	
	Resnet50	64.0	80.6	87.2	82.1	73.7	71.6	76.1	82.3	88.0	76.0	60.1	76.5	
	EfficientNet	<b>76.5</b>	<b>88.2</b>	<b>90.7</b>	87.6	82.0	<b>82.5</b>	83.7	87.2	<b>91.5</b>	<b>87.3</b>	71.2	<b>84.4</b>	
	MobileNet	71.0	87.3	90.5	<b>90.1</b>	<b>82.6</b>	81.5	<b>84.7</b>	<b>88.0</b>	90.8	87.1	<b>73.0</b>	84.2	
	ConvNeXt	61.4	72.1	76.9	79.5	74.2	68.6	71.3	78.0	81.2	74.7	60.1	72.5	

Tablo 5 incelendiğinde, GEI yürüyüş temsilinin tek başına en yüksek performansa MobileNet ile ulaştığı ve farklı ağlarda ulaştığı tanıma performanslarının, her bir çoklu modalite ağı üzerinden sağlanmış kombinasyon versiyonlarında farklı oranlarda değiştiği görülmektedir. Bu performans değişikliği  $CNN_{GEI} + CNN_{SH}$  kombinasyonunda sırasıyla; VGG16 ağı için artış, ResNet50 için düşüş, EfficientNet için artış, MobileNet için artış ve ConvNeXt için düşüş şeklinde seyretmiştir.  $CNN_{GEI} + CNN_{OF}$  kombinasyonunda bu durum; VGG16' da artış, ResNet50' de düşüş, EfficientNet' te artış, MobileNet' te artış ve ConvNeXt' te artış şeklindedir.  $CNN_H + CNN_L$  kombinasyonu için ise düşüş, artış, artış, artış, artış şeklinde seyretmiş ve özellikle, kollarda ConvNeXt' in olduğu mimaride bu artış değeri %7.3' e ulaşmıştır.  $CNN_{GEI} + CNN_{HConL}$  için ise performans sırasıyla VGG16 ağında artış, ResNet50' de artış, EfficientNet' te artış, MobileNet' te artış ve ConvNeXt' te artış şeklinde olmuştur. Böylece, HConL modalitesinin tüm ağlar için artış sağladığı, bunun özellikle EfficientNet ve MobileNet' te ciddi oranda fark edilebilir olduğu gözlemlenmiştir.

Tablo 5' teki sonuçlar sadece kombinasyonlar bazında karşılaştırıldığında,  $CNN_{GEI} + CNN_{OF}$  kombinasyonu, neredeyse tüm çapraz görüş *rank-1* sonuçları açısından,  $CNN_{GEI} + CNN_{SH}$  kombinasyonuna kıyasla, daha yüksek doğruluk oranı sunmaktadır. Bu durum, VGG16, ResNet50, EfficientNet, MobileNet ve ConvNeXt olmak üzere tüm ağlar için geçerli olmuştur. Ağların performansları değerlendirildiğinde, MobileNet' in hem SH hem de OF modalitelerinde çoğu zaman en iyi sonuçları verdiği, bu durumun ortalama değerlerde de açıkça görüldüğü ifade edilebilir. Ancak, ConvNeXt'in OF modalitesiyle  $0^\circ$  ve  $180^\circ$  görüş açıları altında elde ettiği tanıma oranı, MobileNet' in oldukça üzerindedir. Parça tabanlı modalite kombinasyonları incelendiğinde,  $CNN_H + CNN_L$  kombinasyonu,  $CNN_{GEI} + CNN_{SH}$  ve  $CNN_{GEI} + CNN_{OF}$  kombinasyonlarıyla benzer sonuçlar üretmektedir. Bununla birlikte,  $CNN_{GEI} + CNN_{HConL}$  kombinasyonu, ConvNeXt dışında tüm ağların performansını ciddi ölçüde artırmıştır. ConvNeXt ise en yüksek performansına  $CNN_H + CNN_L$  kombinasyonu ile ulaşmıştır. Şekil 27, Tablo 5' teki sonuçlar üzerinden görselleştirilmiştir.

Şekil 27' den de görülebileceği üzere, NM varyasyonu altında, EfficientNet ve MobileNet tabanlı  $CNN_{GEI} + CNN_{HConL}$  kombinasyonu, diğer ağ tabanlı kombinasyonlar arasında en başarılı sonuçları elde etmiştir. Bu kombinasyonu,  $CNN_{GEI} + CNN_{OF}$  kombinasyonu takip etmiştir. Ancak, ConvNeXt ağı için bu durum parça tabanlı  $CNN_H + CNN_L$  kombinasyonu ile geçerli olmuştur.



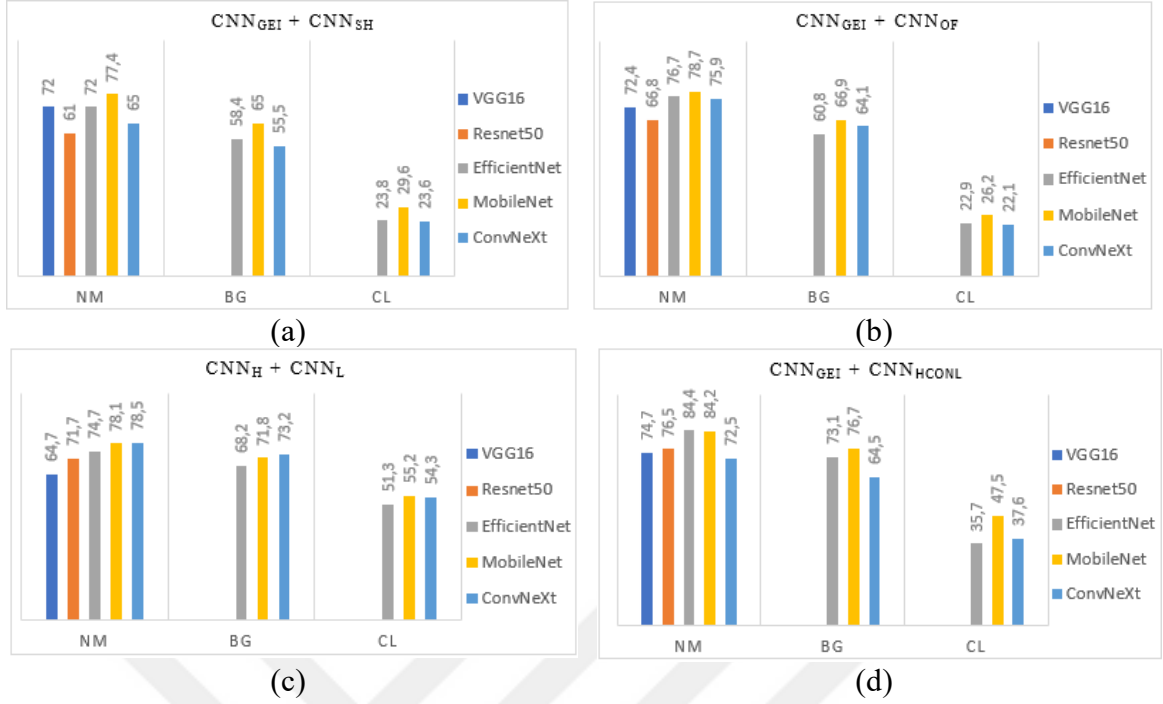
Şekil 27. CASIA-B, NM varyasyonu, kombinasyonların performans karşılaştırması

Tablo 6. CNN<sub>GEI</sub> + CNN<sub>SH</sub>, CNN<sub>GEI</sub> + CNN<sub>OF</sub>, CNN<sub>H</sub> + CNN<sub>L</sub> ve CNN<sub>GEI</sub> + CNN<sub>HConL</sub>, CASIA-B, BG ve CL varyasyonu, rank-1 (%) karşılaştırması (özdeş açılar hariç)

Galeri NM#1-4			0° -180°												ort
Sorgu			0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°		
BG #1-2	CNN <sub>GEI</sub> +	EfficientNet	<b>56.5</b>	61.8	62.4	58.6	55.7	50.3	56.2	60.5	65.4	61.5	<b>53.9</b>	58.4	
		MobileNet	56.1	<b>68.8</b>	<b>70.0</b>	<b>67.4</b>	<b>64.7</b>	<b>59.2</b>	<b>63.8</b>	<b>69.9</b>	<b>73.2</b>	<b>69.8</b>	52.6	<b>65.0</b>	
	CNN <sub>SH</sub>	ConvNeXt	52.3	54.8	59.8	61.3	55.7	52.7	52.8	55.6	59.6	56.8	49.4	55.5	
		CNN <sub>GEI</sub> +	EfficientNet	58.1	66.8	63.9	60.0	59.0	54.3	59.4	63.0	65.2	65.0	54.4	60.8
	CNN <sub>OF</sub>	MobileNet	61.7	<b>69.5</b>	71.2	<b>66.0</b>	<b>66.5</b>	<b>60.4</b>	<b>65.4</b>	<b>70.2</b>	<b>75.0</b>	<b>68.8</b>	<b>61.8</b>	<b>66.9</b>	
		ConvNeXt	<b>63.0</b>	69.0	<b>71.6</b>	65.4	62.7	57.9	57.9	64.8	65.8	66.2	61.1	64.1	
	CNN <sub>H</sub> +	EfficientNet	58.3	73.0	75.7	73.6	67.9	64.9	68.1	72.5	71.2	69.8	54.8	68.2	
		MobileNet	<b>65.1</b>	<b>77.7</b>	<b>82.5</b>	<b>77.0</b>	73.1	66.3	69.1	73.8	75.0	<b>73.1</b>	56.8	71.8	
	CNN <sub>L</sub>	ConvNeXt	64.4	76.8	80.3	76.8	<b>74.3</b>	<b>71.4</b>	<b>72.6</b>	<b>77.9</b>	<b>79.7</b>	71.8	<b>59.7</b>	<b>73.2</b>	
		CNN <sub>GEI</sub> +	EfficientNet	67.4	75.4	78.6	76.8	73.3	70.0	71.1	78.4	79.6	73.1	60.0	73.1
	CNN <sub>HConL</sub>	MobileNet	<b>69.1</b>	<b>79.9</b>	<b>81.6</b>	<b>79.0</b>	<b>74.6</b>	<b>72.2</b>	<b>75.2</b>	<b>81.5</b>	<b>82.2</b>	<b>80.2</b>	<b>67.8</b>	<b>76.7</b>	
		ConvNeXt	56.6	65.6	68.8	68.9	65.8	60.7	64.2	69.0	71.9	66.4	52.2	64.5	
CL #1-2	CNN <sub>GEI</sub> +	EfficientNet	<b>20.1</b>	26.0	26.8	27.2	22.1	20.2	24.1	25.2	27.1	23.8	19.2	23.8	
		MobileNet	19.3	<b>29.0</b>	<b>34.9</b>	<b>35.9</b>	<b>36.1</b>	<b>31.5</b>	<b>31.0</b>	<b>33.2</b>	<b>28.9</b>	<b>25.6</b>	<b>19.8</b>	<b>29.6</b>	
	CNN <sub>SH</sub>	ConvNeXt	16.5	17.6	24.2	23.1	27.4	29.7	30.5	28.6	27.1	19.2	16.2	23.6	
		CNN <sub>GEI</sub> +	EfficientNet	<b>24.4</b>	28.1	28.6	25.6	20.0	20.9	19.6	20.6	22.1	22.7	<b>19.3</b>	22.9
	CNN <sub>OF</sub>	MobileNet	20.8	<b>31.1</b>	<b>30.6</b>	<b>31.5</b>	<b>27.9</b>	<b>29.9</b>	<b>24.5</b>	<b>25.5</b>	<b>23.1</b>	<b>24.0</b>	19.2	<b>26.2</b>	
		ConvNeXt	22.0	24.0	28.0	25.0	24.3	22.8	20.6	21.1	22.1	19.1	15.2	22.1	
	CNN <sub>H</sub> +	EfficientNet	40.5	55.8	59.6	59.0	52.7	53.4	52.9	55.9	50.3	45.3	39.2	51.3	
		MobileNet	43.8	<b>62.5</b>	<b>65.1</b>	<b>61.2</b>	57.1	55.0	<b>58.2</b>	57.1	53.9	<b>54.1</b>	<b>39.8</b>	<b>55.2</b>	
	CNN <sub>L</sub>	ConvNeXt	<b>44.7</b>	56.9	59.4	58.9	<b>59.0</b>	<b>57.9</b>	55.4	<b>58.3</b>	<b>57.0</b>	50.9	39.2	54.3	
		CNN <sub>GEI</sub> +	EfficientNet	30.0	38.9	43.9	42.6	37.0	35.0	35.4	34.4	34.6	34.4	26.1	35.7
	CNN <sub>HConL</sub>	MobileNet	<b>36.8</b>	<b>54.8</b>	<b>56.4</b>	<b>52.3</b>	<b>51.9</b>	<b>49.2</b>	<b>47.0</b>	<b>50.1</b>	<b>46.1</b>	<b>43.6</b>	<b>34.5</b>	<b>47.5</b>	
		ConvNeXt	32.8	40.1	42.6	39.3	41.0	39.5	35.7	36.9	40.4	36.4	29.4	37.6	

Tablo 6’ da, hazırlanan tüm çoklu modalite ağ mimarileri CASIA-B veri seti, BG ve CL varyasyonları altında test edilmiştir. Ancak, Tablo 5 ‘te NM varyasyonu için VGG16 ve ResNet50 tabanlı mimarilerin oldukça düşük tanıma oranları sergilemesi sebebiyle, BG ve CL varyasyonları için yalnızca EfficientNet, MobileNet ve ConvNeXt’e ait tanıma oranları sunulmuştur. Tablo 6’ dan çıkarılabilecek sonuçlara göre, BG varyasyonu altında,  $CNN_{GEI} + CNN_{OF}$  kombinasyonu, neredeyse tüm çapraz açılarda *rank-1* sonuçları açısından,  $CNN_{GEI} + CNN_{SH}$  kombinasyonuna kıyasla tüm ağlar üzerinden daha başarılı bir performans sergilemiştir. Bu iki kombinasyon dikkate alındığında, EfficientNet ve MobileNet’in performansları birbirine yakın olmakla birlikte, her iki kombinasyonda da en yüksek ortalama *rank-1* değeri MobileNet ile elde edilmiştir.  $CNN_H + CNN_L$  kombinasyonu incelendiğinde, tüm ağlar için tanıma başarısının anlamlı ölçüde arttığı gözlemlenmektedir. Ayrıca,  $CNN_{GEI} + CNN_{HConL}$  kombinasyonu için bu başarı artışı devam etmekte ancak, bu eğilim ConvNeXt ağı için tersine dönmektedir. CL varyasyonu altında ise, tüm ağlar bazında  $CNN_{GEI} + CNN_{SH}$  ve  $CNN_{GEI} + CNN_{OF}$  kombinasyonlarının tanıma oranlarının önemli ölçüde azaldığı görülmektedir. Bununla birlikte, parça tabanlı modalite kombinasyonlarında tanıma oranlarında bir artış meydana geldiği anlaşılmaktadır. Özellikle MobileNet tabanlı  $CNN_H + CNN_L$  kombinasyonu, CL varyasyonu altında en yüksek ortalama *rank-1* değerine ulaşmaktadır. Bu sonuç parça tabanlı modalitelerin varyasyonlara karşı daha dayanıklı olduğunu ispatlamaktadır.

CASIA-B veri seti NM, BG ve CL varyasyonları altında, her kombinasyona ait ortalama *rank-1* görselleri Şekil 28’ de sunulmuştur. Şekil 28 incelendiğinde,  $CNN_{GEI} + CNN_{SH}$  kombinasyonunun MobileNet tabanlı olarak en iyi sonuçları verdiği gözlemlenmektedir. Bu durum,  $CNN_{GEI} + CNN_{OF}$  kombinasyonu için de geçerlidir.  $CNN_H + CNN_L$  kombinasyonunda ise MobileNet ve ConvNeXt benzer performanslar sergileyerek EfficientNet’e kıyasla daha iyi sonuçlara ulaşmaktadır. Son kombinasyon olan  $CNN_{GEI} + CNN_{HConL}$ , NM varyasyonu altında en yüksek performansı EfficientNet ve MobileNet tabanlı olarak sağlarken; BG ve CL varyasyonlarında yalnızca MobileNet ile yüksek performans sergilemektedir. Sonuç olarak, BG varyasyonu için en iyi sonuç MobileNet tabanlı  $CNN_{GEI} + CNN_{HConL}$  ile ikinci sırada ConvNeXt tabanlı  $CNN_H + CNN_L$  kombinasyonu tarafından elde edilirken; CL varyasyonu için en iyi MobileNet ve ConvNeXt tabanlı  $CNN_H + CNN_L$  kombinasyonu ile elde edilmiştir. Ayrıca parça tabanlı modalitelerin varyasyonlara karşı global tabanlı modalitelere göre daha dayanıklı oldukları sonucuna varılmıştır.



Şekil 28. Her kombinasyonun NM, BG, CL varyasyonları altında ortalama *rank-1* karşılaştırması (a)  $CNN_{GEI} + CNN_{SH}$ , (b)  $CNN_{GEI} + CNN_{OF}$ , (c)  $CNN_H + CNN_L$ , (d)  $CNN_{GEI} + CNN_{HConL}$

Yukarıda yapılan analizler özellikle, başta  $CNN_{GEI} + CNN_{HConL}$  kombinasyonu olmak üzere, farklı modalitelerin birleştirilmesiyle CASIA-B veri seti üzerinde iyi performanslar elde edildiğini göstermiştir. Bu kapsamda, önerilen bazı kombinasyonlar, CASIA-B veri seti üzerinde literatürdeki güncel bazı son teknoloji yöntemlerle karşılaştırılmıştır. Bu amaçla üç farklı karşılaştırma grubu oluşturulmuştur. İlk karşılaştırma grubu, tüm çapraz görüş açıları için tanıma oranlarını sunan ve NM varyasyonu için Tablo 4 ile aynı deneysel ayarları kullanan güncel bir yöntem olan GaitNet (Song vd., 2019) ile yapılmıştır. Karşılaştırma için, önerilen çok modlu ağlardan  $CNN_{GEI} + CNN_{HConL}$  kombinasyonunun EfficientNet (Eff+HConL) ve MobileNet (Mobile+HConL) tabanlı versiyonları seçilmiştir. İkinci karşılaştırma grubu, tüm görünüm tabanlı varyasyonlar (NM, BG, CL) için ortalama *rank-1* doğruluğu (%) (özdeş görüş açısı hariç tutularak) üzerinden yapılmıştır. Karşılaştırma için seçilen yöntemler GeiNet (Shiraga vd., 2016), DeepCNN (Wu vd., 2016), GaitNet (Song vd., 2019) ve PoseGait (Liao vd., 2020) olup, bu yöntemlerin deneysel ayarları Tablo 2 ile aynıdır. Karşılaştırma için önerilen çok modlu ağlar arasından, NM varyasyonu için EfficientNet tabanlı  $CNN_{GEI} + CNN_{HConL}$  (Eff+HConL) ve MobileNet tabanlı  $CNN_{GEI} + CNN_{HConL}$  (Mobile+HConL), BG varyasyonu için Mobile+HConL, ve CL varyasyonu için MobileNet tabanlı  $CNN_H + CNN_L$

(Mobile+H+L) seçilmiştir. Son karşılaştırma, NM koşulları altında ilk 74 birey ile eğitilmiş üretici bir yöntem olan MGan (He vd., 2018) ile yapılmıştır. İlk karşılaştırma grubu için elde edilen tüm açı bazında tanıma oranları Tablo 7’ de sunulmuştur.

Tablo 7. CASIA-B veri seti NM varyasyonu önerilen bazı çoklu modalite mimarilerinin GaitNet (Song vd., 2019) ile çapraz görünüm tanıma oranlarının karşılaştırması

		Sorgu Açısı NM#5-6											ort	
		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°		
Galeri Açısı NM #1-4	0°	GaitNet	<b>100</b>	91.0	86.0	<b>77.0</b>	<b>69.0</b>	<b>67.0</b>	63.0	<b>74.0</b>	70.0	75.0	84.0	75.6
		Eff+HConL	<b>100</b>	<b>99.0</b>	<b>89.0</b>	74.0	59.0	55.0	<b>65.0</b>	63.0	<b>77.0</b>	<b>89.0</b>	<b>95.0</b>	<b>76.5</b>
		Mobile +HConL	<b>100</b>	93.0	86.0	65.0	47.0	50.0	54.0	66.0	73.0	85.0	91.0	71.0
	18°	GaitNet	<b>98.0</b>	<b>100</b>	<b>100</b>	<b>99.0</b>	<b>87.0</b>	<b>86.0</b>	<b>89.0</b>	<b>88.0</b>	<b>91.0</b>	90.0	85.0	<b>91.3</b>
		Eff+HConL	<b>98.0</b>	<b>100</b>	99.0	94.0	78.0	78.0	78.0	80.0	85.0	<b>96.0</b>	<b>96.0</b>	88.2
		Mobile +HConL	97.0	99.0	99.0	92.0	73.0	74.0	82.0	78.0	90.0	95.0	93.0	87.3
	36°	GaitNet	<b>88.0</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>96.0</b>	<b>90.0</b>	90.0	90.0	91.0	85.0	<b>82.0</b>	<b>91.2</b>
		Eff+HConL	86.0	<b>100</b>	<b>100</b>	<b>100</b>	<b>96.0</b>	86.0	<b>92.0</b>	86.0	<b>93.0</b>	<b>91.0</b>	77.0	90.7
		Mobile +HConL	86.0	<b>100</b>	<b>100</b>	<b>100</b>	<b>96.0</b>	<b>90.0</b>	91.0	<b>92.0</b>	91.0	85.0	74.0	90.5
	54°	GaitNet	<b>83.0</b>	<b>97.0</b>	<b>100</b>	<b>100</b>	99.0	97.0	96.0	<b>96.0</b>	<b>95.0</b>	<b>87.0</b>	<b>79.0</b>	<b>92.9</b>
		Eff+HConL	70.0	95.0	98.0	<b>100</b>	99.0	98.0	96.0	95.0	85.0	76.0	64.0	87.6
		Mobile +HConL	76.0	96.0	<b>100</b>	<b>100</b>	<b>100</b>	<b>99.0</b>	<b>98.0</b>	<b>96.0</b>	91.0	77.0	68.0	90.1
	72°	GaitNet	<b>79.0</b>	<b>89.0</b>	95.0	97.0	<b>100</b>	<b>100</b>	<b>100</b>	<b>99.0</b>	<b>95.0</b>	<b>92.0</b>	<b>79.0</b>	<b>92.5</b>
		Eff+HConL	55.0	76.0	<b>96.0</b>	98.0	<b>100</b>	99.0	97.0	94.0	87.0	64.0	54.0	82.0
		Mobile +HConL	49.0	82.0	95.0	<b>99.0</b>	<b>100</b>	<b>100</b>	99.0	95.0	<b>95.0</b>	64.0	48.0	82.6
	90°	GaitNet	<b>78.0</b>	<b>85.0</b>	91.0	95.0	<b>99.0</b>	<b>100</b>	<b>100</b>	<b>99.0</b>	<b>97.0</b>	<b>89.0</b>	<b>77.0</b>	<b>91.0</b>
		Eff+HConL	56.0	71.0	<b>94.0</b>	<b>99.0</b>	<b>99.0</b>	99.0	<b>100</b>	98.0	88.0	68.0	52.0	82.5
		Mobile +HConL	52.0	76.0	92.0	97.0	<b>99.0</b>	99.0	99.0	96.0	92.0	65.0	47.0	81.5
	108°	GaitNet	<b>73.0</b>	<b>83.0</b>	88.0	95.0	<b>99.0</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>99.0</b>	<b>95.0</b>	<b>86.0</b>	<b>91.8</b>
		Eff+HConL	56.0	75.0	<b>91.0</b>	96.0	<b>99.0</b>	99.0	99.0	97.0	96.0	73.0	55.0	83.7
		Mobile +HConL	53.0	79.0	<b>91.0</b>	<b>98.0</b>	<b>99.0</b>	99.0	99.0	97.0	96.0	81.0	54.0	84.7
	126°	GaitNet	<b>80.0</b>	<b>88.0</b>	90.0	94.0	<b>99.0</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>99.0</b>	<b>98.0</b>	<b>90.0</b>	<b>93.8</b>
		Eff+HConL	59.0	80.0	88.0	<b>98.0</b>	<b>99.0</b>	98.0	99.0	98.0	97.0	88.0	66.0	87.2
		Mobile +HConL	60.0	74.0	<b>94.0</b>	97.0	<b>99.0</b>	<b>100</b>	99.0	98.0	98.0	94.0	65.0	88.0
	144°	GaitNet	<b>75.0</b>	86.0	91.0	92.0	<b>95.0</b>	<b>99.0</b>	<b>99.0</b>	<b>99.0</b>	<b>100</b>	<b>100</b>	<b>93.0</b>	<b>92.9</b>
		Eff+HConL	74.0	<b>92.0</b>	<b>95.0</b>	<b>94.0</b>	93.0	90.0	98.0	98.0	98.0	95.0	86.0	91.5
		Mobile +HConL	74.0	87.0	92.0	93.0	94.0	94.0	98.0	98.0	98.0	96.0	82.0	90.8
	162°	GaitNet	83.0	90.0	<b>90.0</b>	<b>94.0</b>	<b>95.0</b>	<b>96.0</b>	<b>98.0</b>	<b>100</b>	<b>100</b>	<b>100</b>	95.0	<b>94.1</b>
		Eff+HConL	<b>93.0</b>	94.0	85.0	87.0	74.0	75.0	79.0	92.0	98.0	98.0	<b>96.0</b>	87.3
		Mobile +HConL	<b>93.0</b>	<b>95.0</b>	<b>90.0</b>	77.0	72.0	77.0	81.0	95.0	96.0	98.0	95.0	87.1
	180°	GaitNet	89.0	86.0	<b>83.0</b>	<b>76.0</b>	<b>74.0</b>	<b>76.0</b>	<b>76.0</b>	<b>81.0</b>	<b>85.0</b>	93.0	<b>99.0</b>	<b>81.9</b>
		Eff+HConL	<b>95.0</b>	88.0	73.0	66.0	55.0	44.0	52.0	68.0	78.0	93.0	98.0	71.2
		Mobile +HConL	94.0	<b>89.0</b>	75.0	59.0	52.0	53.0	58.0	74.0	81.0	<b>95.0</b>	98.0	73.0

Tablo 7’ de sunulan bulgular, önerilen çok modlu ağların (Eff+HConL ve Mobile+HConL), GaitNet yöntemiyle oldukça yakın ve karşılaştırılabilir performans düzeyleri sergilediğini göstermektedir. Ancak, açılar arası fark arttığında elde edilen tanıma oranlarının düşüklüğü bu durumun ortalama değere yansımaları engellediği sonucuna ulaştırmaktadır. Genel olarak, GaitNet’ e göre, birbirine yakın (örneğin, galeri açısı 18° iken simetriği 54°) ya da simetrik açılar arasında (örneğin, galeri açısı 0° iken simetriği 180° ya da tersi durumlarda) performans artışı gözlemlenmektedir. Ortalama *rank-1* doğruluğu (%) (özdeş görüş açısı hariç tutularak) üzerinden yapılan ikinci karşılaştırma grubu için değerlendirme sonuçları Tablo 8 ‘de sunulmuştur.

Tablo 8. Önerilen bazı çok modlu ağların, CASIA-B veri seti NM, BG, CL varyasyonları, literatürde önde gelen yöntemler ile ortalama *rank-1* (%) üzerinden (özdeş açılar hariç) karşılaştırması

Galeri NM#1-4		0° -180°										ort	
Sorgu		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	
NM #5-6	GeiNet (Shiraga vd., 2016)	45.8	57.6	67.1	66.9	56.3	48.3	58.3	68.4	69.4	59.0	46.5	58.5
	DeepCNN (Wu vd., 2016)	<b>82.6</b>	90.3	<b>96.1</b>	<b>94.3</b>	90.1	87.4	89.9	<b>94.0</b>	<b>94.7</b>	91.3	78.5	<b>89.9</b>
	GaitNet (Song vd., 2019)	75.6	<b>91.3</b>	91.2	92.9	<b>92.5</b>	<b>91.0</b>	<b>91.8</b>	93.8	92.9	<b>94.1</b>	<b>81.9</b>	<b>89.9</b>
	PoseGait (Liao vd., 2020)	55.3	69.6	73.9	75.0	68.0	68.2	71.1	72.9	76.1	70.4	55.4	68.7
	Eff+HConL	76.5	88.2	90.7	87.6	82.0	82.5	83.7	87.2	91.5	87.3	71.2	84.4
	Mobile+HConL	71.0	87.3	90.5	90.1	82.6	81.5	84.7	88.0	90.8	87.1	73.0	84.2
BG #1-2	DeepCNN (Wu vd., 2016)	64.2	<b>80.6</b>	<b>82.7</b>	76.9	64.8	63.1	68.0	76.9	<b>82.2</b>	75.4	61.3	72.4
	PoseGait (Liao vd., 2020)	35.3	47.2	52.4	46.9	45.5	43.9	46.1	48.1	49.4	43.6	31.1	44.5
	Mobile+HConL	<b>69.1</b>	79.9	81.6	<b>79.0</b>	<b>74.6</b>	<b>72.2</b>	<b>75.2</b>	<b>81.5</b>	<b>82.2</b>	<b>80.2</b>	<b>67.8</b>	<b>76.7</b>
CL #1-2	DeepCNN (Wu vd., 2016)	37.7	57.2	<b>66.6</b>	61.1	55.2	54.6	55.2	<b>59.1</b>	<b>58.9</b>	48.8	39.4	54.0
	PoseGait (Liao vd., 2020)	24.3	29.7	41.3	38.8	38.2	38.5	41.6	44.9	42.2	33.4	22.5	35.9
	Mobile+H+L	<b>43.8</b>	<b>62.5</b>	65.1	<b>61.2</b>	<b>57.1</b>	<b>55.0</b>	<b>58.2</b>	57.1	53.9	<b>54.1</b>	<b>39.8</b>	<b>55.2</b>

Tablo 8 ‘den açıkça görülebileceği üzere, Eff+HConL ve Mobile+HConL çok modlu ağları, NM koşulları altında GeiNet ve model tabanlı bir yöntem olan PoseGait’e kıyasla daha yüksek ortalama tanıma oranları elde etmiştir. Ancak, ortalama tanıma başarımı açısından en iyi sonuçlara DeepCNN ve GaitNet ulaşmıştır. BG ve CL varyasyonları altında ise sırasıyla Mobile+HConL ve Mobile+H+L kombinasyonları en yüksek performansa ulaşmıştır. Bu sonuçlar görünüme dayalı varyasyonlar için parça tabanlı yaklaşımların performansı artırmada daha etkili olabileceği anlamına gelmektedir. Son grup, üretici bir yöntem olan MGan (He vd., 2018) ile NM varyasyonu altında, çalışmada yalnızca 54°, 90° ve 126° çapraz-açıları için sunulan başarı değerleri üzerinden, performans karşılaştırması Tablo 9’ da sunulmuştur.

Tablo 9. Çok modlu ağların, CASIA-B, NM varyasyonu MGan ile karşılaştırılması

Metotlar	54°	90°	126°	ort
MGan (He vd., 2018)	84.2	72.3	83.0	79.8
Eff+HConL	87.6	<b>82.5</b>	87.2	85.8
Mobile+HConL	<b>90.1</b>	81.5	<b>88.0</b>	<b>86.5</b>

Tablo 9’ da sunulan başarı oranlarına göre Mobile+HConL çok modlu ağı, ilgili görüş açıları için ortalama *rank-1* doğruluğu parametresinde en yüksek tanıma başarısına ulaşmıştır. Buradan yola çıkılarak, hazırlanan çoklu modalite ağ mimarilerinden şimdiye kadar üç kategoride de yapılan karşılaştırmalar sonucunda literatürdeki önde gelen bazı yürüyüş tanıma yöntemleriyle karşılaştırılabilir sonuçlara ulaşanların yalnızca parça tabanlı modalitelere dayalı mimarilerin olduğu görülmektedir.

Yürüyüş tanıma işleminde, kıyafet ve arka planın öğrenmeyi etkileyebilme ihtimali sebebi ile literatürde direkt kullanımı fazla yaygın olmayan, daha önceki bölümlerde detaylandırılmış RGB yürüyüş görüntüleri ile arka planı elimine etmeye yönelik hazırlanan, maskelenmiş RGB yürüyüş görüntülerinin CASIA-B veri seti üzerinden ayrı bir performans analizi, diğer tüm modalitelerden hariç tutularak bu kısımda yapılmıştır. Diğer modaliteler üzerinden elde edilen sonuçlara dayanarak ağ mimarisi olarak MobileNetV1 seçilmiştir. Analizin ilk aşamasında, GEI girdi verisi ile RGB ve maskelenmiş RGB yürüyüş görüntülerinin ayrı ayrı MobileNetV1 üzerinden elde edilen tanıma başarıları değerlendirilmiştir. Ardından, GEI’ nin tanıma başarısına katkıları, hazırlanan çoklu modalite ağ mimarisi üzerinden GEI’ nin öznelikleri ile ayrı ayrı birleştirme (concatenate) işlemine tabi tutularak hesaplanmıştır. Tablo 10’ da; GEI, RGB, maskelenmiş RGB (MRGB) ile GEI ve RGB (GEI+RGB), GEI ve MRGB (GEI+MRGB) çoklu modalite kombinasyonları üzerinden elde edilmiş, ortalama *rank-1* doğruluk oranları sunulmuştur.

Tablo 10. CASIA-B veri seti RGB ve maskelenmiş RGB (MRGB) bazlı tekli ve çoklu modalite mimarilerinin *rank-1* doğruluk oranı karşılaştırması (özdeş açılar hariç)

Galeri NM#01-04		0° -180°											ort
Sorgu		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	
NM #05-06	GEI	59.3	79.4	83.4	85.2	76.9	72.3	80.0	84.3	84.6	80.9	60.2	76.9
	RGB	86.0	79.8	78.1	87.9	81.8	86.8	82.4	86.8	72.4	77.9	84.4	82.2
	MRGB	86.6	80.7	78.5	89.4	79.2	87.6	87.5	87.6	72.7	83.9	85.6	83.5
	GEI+RGB	92.6	93.2	<b>95.7</b>	96.8	93.1	<b>93.0</b>	92.6	<b>96.5</b>	92.7	94.0	<b>92.8</b>	93.9
	GEI+MRGB	<b>94.1</b>	<b>95.2</b>	95.1	<b>97.3</b>	<b>94.4</b>	92.4	<b>95.1</b>	96.1	<b>94.7</b>	<b>95.7</b>	92.0	<b>94.7</b>
BG #01-02	GEI+MRGB	89.5	90.9	89.3	93.3	86.2	87.3	89.1	91.8	88.8	91.3	88.8	89.7
CL #01-02	GEI+MRGB	46.1	52.4	48.9	48.5	36.0	32.8	33.1	35.3	30.1	33.3	29.7	38.7

Tablo 10 incelendiğinde NM varyasyonunda RGB modalitesinin beklendiği üzere, özellikle kıyafet gibi detaylı ve renkli verilerin tanımaya dahil olması sebebiyle, GEI' ye göre oldukça yüksek tanıma performansına ulaştığı görülmektedir. Maskelenmiş RGB (MRGB) ise arka plan etkisinin ortadan kaldırılmasıyla, RGB modalitesine göre %1.3' lük bir performans artışı sağlamıştır. GEI ile RGB ve GEI ile MRGB modalitelerinin çoklu modalite ağları üzerinden özneliklerinin birleştirilmesiyle ise toplam tanıma başarısı ciddi oranda artmıştır. GEI' nin tanıma performansındaki bu artış; RGB modalitesinde %17 olurken, MRGB' de %17.8 olmuştur. Çoklu modalite GEI+MRGB ağının, BG varyasyonunda tanıma performansı %5 oranında azalsa da bu düşüş, CL varyasyonunda çok ciddi oranda olmuştur. Buna sebep olarak, RGB modalitelerinin, GEI, SH ve OF modaliteleri gibi parça tabanlı olmamalarından, varyasyona duyarlı hale gelmeleri gösterilebilir.

Son olarak, hazırlanan çoklu modalite GEI+MRGB ağı, literatürde kapalı veri seti üzerinde RGB çerçeveleri üzerinden yürüyüş tanıma gerçekleştiren, son teknoloji bir yöntem olan HGRBOL (Khan vd., 2023) ile karşılaştırılmıştır. Yöntemde, optik akışlar ve iyileştirilmiş RGB çerçevelerinden, ince ayarlanmış EfficientNet-B0 modeli ile çıkarılan öznelikler birleştirilerek tanıma gerçekleştirilmiştir. HGRBOL yöntemi ile hazırlanan GEI+MRGB ağının, CASIA-B veri seti NM varyasyonu üzerinden gerçekleştirilen performans analizi Tablo 11' de sunulmuştur. Tablo'dan hazırlanan MRGB ve GEI ile eğitilmiş MobileNetV1 tabanlı çoklu modalite ağının, birçok çapraz görüş açısında ve ortalamada açık veri seti üzerinde eğitilmesine rağmen HGRBOL yönteminden daha iyi performansa sahip olduğu görülmektedir.

Tablo 11. Çoklu modalite ağının CASIA-B veri seti NM varyasyonu HGRBOL (Khan vd., 2023) yöntemi ile performans karşılaştırması (özdeş açılar hariç)

Galeri NM#01-04		0° -180°											ort
Sorgu		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	
NM #05-06	HGRBOL	94.1	94.8	95.3	90.2	95.0	95.6	94.8	94.3	94.8	95.7	94.4	94.4
	GEI+MRGB	94.1	<b>95.2</b>	95.1	<b>97.3</b>	94.4	92.4	<b>95.1</b>	<b>96.1</b>	94.7	95.7	92.0	<b>94.7</b>

### 3.1.2. Outdoor-Gait Veri Seti Performans Değerlendirmesi

Bu bölümde hazırlanan çoklu modalite ağ mimarilerinin kendi aralarında ve son teknoloji yöntemlerin performansları ile karşılaştırıldığı bir dış ortam veri seti olan Outdoor-

Gait' e ait analiz sonuçları sunulmuştur. Bu veri setinde, CASIA-B' den farklı olarak, galeri görüntülerinde BG ve CL varyasyonları da kullanılmıştır. Farklı ağ tabanlı modalite kombinasyonlarının ortalama *rank*-1 doğruluğu üzerinden karşılaştırma sonuçları Tablo 12' de gösterilmiştir. Tablo 12' den açıkça görüldüğü üzere, CNN<sub>GEI</sub> + CNN<sub>HConL</sub> kombinasyonu CASIA-B' deki sonuçlara benzer olarak tüm ağlar arasında en iyi performansı sergilemiştir. Ayrıca, MobileNet tabanlı CNN<sub>GEI</sub> + CNN<sub>HConL</sub> kombinasyonu, diğer ağlarla kıyaslandığında en yüksek başarıya ulaşmıştır.

Tablo 12. Outdoor-Gait veri seti üzerinde önerilen çok modlu ağların tüm varyasyonlar altında ortalama *rank*-1 doğruluğu (%) üzerinden karşılaştırması

		Galeri	NM			BG			CL			ort
		Sorgu	NM	BG	CL	NM	BG	CL	NM	BG	CL	
CNN <sub>GEI</sub>	EfficientNet		58.94	48.55	40.34	47.58	63.28	34.42	39.98	35.02	61.59	47.74
+	MobileNet		76.09	67.39	55.31	63.28	81.40	<b>55.19</b>	55.55	<b>54.59</b>	75.12	64.88
CNN <sub>SH</sub>	ConvNeXt		<b>81.16</b>	<b>67.99</b>	<b>58.94</b>	<b>65.94</b>	<b>86.71</b>	51.81	<b>56.88</b>	53.02	<b>84.54</b>	<b>67.44</b>
CNN <sub>H</sub>	EfficientNet		58.94	44.44	40.94	42.75	65.94	27.17	36.83	25.48	63.28	45.08
+	MobileNet		77.54	<b>75.24</b>	<b>72.34</b>	<b>70.17</b>	<b>86.96</b>	<b>58.57</b>	<b>67.87</b>	<b>63.77</b>	<b>85.02</b>	<b>73.05</b>
CNN <sub>L</sub>	ConvNeXt		<b>80.19</b>	72.95	58.82	66.67	82.13	52.54	52.54	47.58	84.54	66.44
CNN <sub>GEI</sub>	EfficientNet		77.29	61.95	51.45	64.61	79.23	44.32	49.52	47.22	82.37	61.99
+	MobileNet		87.44	<b>81.52</b>	<b>71.01</b>	<b>76.81</b>	93.24	<b>64.61</b>	<b>67.75</b>	<b>64.37</b>	88.41	<b>77.24</b>
CNN <sub>HconL</sub>	ConvNeXt		<b>92.51</b>	79.59	69.08	73.91	<b>95.41</b>	57.49	63.41	54.71	<b>95.17</b>	75.70

Hazırlanan çoklu modalite ağ mimarileri arasından, MobileNet tabanlı CNN<sub>GEI</sub> + CNN<sub>HConL</sub> (Mobile+HConL) ve ConvNeXt tabanlı CNN<sub>GEI</sub> + CNN<sub>HConL</sub> (ConvNeXt+HConL) kombinasyonları, sırasıyla görünüm tabanlı GaitNet (Song vd., 2019) ve model tabanlı 3DHuman (Meng vd., 2023) yöntemleriyle Tablo 13 ve Tablo 14' te karşılaştırılmıştır.

Tablo 13. Outdoor-Gait veri seti, ortalama *rank*-1 doğruluğu (%) üzerinden bazı çok modlu ağlar ile görünüm tabanlı GaitNet (Song vd., 2019) yönteminin karşılaştırması

	Galeri Sorgu	NM			BG			CL			ort
		NM	BG	CL	NM	BG	CL	NM	BG	CL	
GaitNet		<b>96.9</b>	<b>89.1</b>	60.2	<b>92.0</b>	<b>97.1</b>	59.7	58.7	55.4	<b>97.3</b>	<b>78.5</b>
Mobile+HConL		87.4	81.5	<b>71.0</b>	76.8	93.2	<b>64.6</b>	<b>67.7</b>	<b>64.3</b>	88.4	77.2
ConvNeXt+HConL		92.5	79.6	69.1	73.9	95.4	57.5	63.4	54.7	95.2	75.7

Tablo 14. Outdoor-Gait veri seti, ortalama rank-1 doğruluğu (%) üzerinden bazı çok modlu ağlar ile model tabanlı 3DHuman (Meng vd., 2023) yönteminin karşılaştırması

	Galeri	NM			BG			CL			ort
	Sorgu	NM	BG	CL	NM	BG	CL	NM	BG	CL	
3DHuman		90.0	<b>83.6</b>	<b>80.9</b>	<b>77.5</b>	90.3	<b>72.6</b>	<b>72.2</b>	<b>66.4</b>	88.8	80.2
ConvNeXt+HConL		<b>92.5</b>	79.6	69.1	73.9	<b>95.4</b>	57.5	63.4	54.7	<b>95.2</b>	75.7

Tablo 13 değerlendirildiğinde, Mobile+HConL kombinasyonunun, CL varyasyonunun sorgu görüntüsü olduğu; NM-CL, BG-CL, CL-NM ve CL-BG eşleşmelerinde GaitNet' e kıyasla daha yüksek performans gösterdiği görülmektedir. Ortalama değer açısından ise sonuçlar birbirine oldukça yakındır. Tablo 14' te, ConvNeXt tabanlı  $CNN_{GEI} + CNN_{HConL}$  kombinasyonunun; NM-NM, BG-BG ve CL-CL galeri-sorgu görüntü eşleşmelerinde yüksek tanıma oranına ulaştığı gözlemlense de bu durum, ortalama tanıma başarımına yansımamıştır. Ancak, 3DHuman (Meng vd., 2023) model tabanlı yöntemi, 3B insan vücudu yeniden yapılandırma süreci nedeniyle daha yüksek bir hesaplama maliyetine sahiptir.

### 3.1.3. MobileNet Hesaplama Maliyeti Analizi

Önceki bölümlerde çoklu modalite ağları üzerinden hazırlanan çeşitli kombinasyonlar arasında, MobileNet tabanlı mimarilerin sıklıkla daha başarılı performansa ulaştığı görülmüştür. MobileNet, sınırlı kaynaklarla çalışırken doğruluğu en üst düzeye çıkarmayı hedefleyen bir yapıdır. Bu nedenle; küçük boyutu, düşük güç tüketimi, yüksek hızı ve maliyet etkinliği ile öne çıkmaktadır. Bu bağlamda, çok modaliteli MobileNet tabanlı herhangi bir kombinasyon dikkate alınarak, GaitSet (Chao vd., 2019) ve GaitNet (Song vd., 2019) yöntemleri ile bir maliyet karşılaştırması gerçekleştirilmiştir. GaitSet, yürüyüş tanıma oranını önemli ölçüde artırmış olsa da karmaşık ağ mimarisi, yüksek sayıda parametre ve FLOP içermesi ile dikkat çekmektedir. Bu yöntemin toplamda 2.59M parametresi, 8.6G FLOP değeri ve 15872 boyutunda bir nihai öznitelik vektörü bulunmaktadır (Song vd., 2022). Parametre ve FLOP değerleri eğitim aşamasının hesaplama maliyetini belirlerken, nihai öznitelik vektörünün boyutu test aşamasında ciddi bir hesaplama yükü getirir. Benzer şekilde, GaitNet omurga ağ yapısı olarak yüksek parametre sayısına sahip Tam Konvolüsyonel Ağ (FCN) (Long vd., 2015) kullanılmaktadır. Bu ağların, çok modlu MobileNet (mm-MobileNet) ile karşılaştırmalı sonuçları Tablo 15' te sunulmuştur.

Tablo 15. Ağların maliyet analizi. Boyut, nihai öznitelik vektörünün boyutudur

Network	FLOP(G)	Param(M)	Boyut
GaitSet	8.59	<b>2.59</b>	15872
FCN-Alexnet	-	57	-
FCN-VGG16	-	134	-
FCN-GoogLeNet	-	6	-
mm-MobileNet	<b>1.15</b>	3.3	<b>1024+1024</b>

Tablo 15 incelendiğinde, mm-MobileNet'in sahip olduğu parametre sayısının GaitSet ile oldukça yakın olduğu, ancak FLOP değerinin GaitSet'e kıyasla yaklaşık 8 kat daha az olduğu görülmektedir. Ayrıca, parametre sayısı FCN tabanlı ağlara kıyasla belirgin ölçüde daha düşüktür. Bununla birlikte, test aşaması için kritik öneme sahip olan nihai öznitelik boyutu, GaitSet' inkiye göre ciddi düzeyde daha küçüktür.

### 3.2. Topluluk Öğrenmeli Parça-Tabanlı Ağ Mimarisi Başarı Değerlendirmesi

Bu bölümde daha önceki bölümlerde detaylandırılan topluluk öğrenmeli parça-tabanlı ağ mimarisinin detaylı performans analizleri gerçekleştirilmiştir. Çoklu modalite ağ yapısı üzerinden yapılan detaylı performans analizlerinin sonuçlarına göre yürüyüş tanımada parça-tabanlı modaliteler en etkili modalite türleri olarak tespit edilmiş, bu modalite türleri özellikle görünüm tabanlı varyasyonlara dayanıklılıkta ön plana çıkmışlardır. Ağ kollarında kullanılan ince ayarlanmış VGG16, ResNet50, EfficientNetB0, MobileNetV1, ve ConvNeXt-base mimarilerinden ise parça-tabanlı modalitelere dayalı MobileNetV1 ve ConvNeXt-base en başarılı mimariler olarak öne çıkmış ve hazırlanan topluluk öğrenmeli parça-tabanlı ağ mimarisi için ConvNeXt-base omurga ağ olarak seçilmiştir. Bu kapsamda sonraki bölümlerde adım adım parça bazlı performans değerlendirmelerinden yola çıkılarak nihai topluluk öğrenmeli parça-tabanlı ağ mimarisinin performans analizine erişilmiştir.

#### 3.2.1. Parça Bazında Başarı Değerlendirmeleri

Nihai topluluk öğrenmeli parça-tabanlı ağ mimarisi performans analizinden önce, literatürde sıkça önerilen insan vücut yürüyüş özelliğinin yatay dört parçaya bölünmesinin tanımada en etkili olabilecek bölümlendirme şekli (Zhang vd., 2019) olabileceğinden yola çıkılarak, dört parçalı bir GEI görüntüsünün parça bazında detaylı tanıma analizi gerçekleştirilmiştir. Bu kapsamda GEI yürüyüş özelliği ilk olarak dört yatay

parçaya bölümlendirilmiş ve baş bölgesinden başlanarak sırasıyla baş, göğüs, bel ve bacak bölgeleri parça-1, parça-2, parça-3 ve parça-4 olarak adlandırılmıştır. Sonrasında her bir parçaya ait bireysel tanıma oranları incelenmiş ardından bu parçaların birleştirilmesiyle elde edilen dört parçalı bütün mimarinin tanıma performansı analiz edilmiştir. İkinci olarak ise yatay bölümlendirme beş parça olarak gerçekleştirilmiş ve beş parçalı bütün mimarinin tanıma performansı analize dahil edilmiştir. Tablo 16’ da, CASIA-B veri seti NM, BG ve CL varyasyonları altında, Tablo 4’ teki eğitim ve test küme bölümlenmesine göre ayarlanmış deney ortamında, her bir yatay parçanın bireysel tanıma oranları, parça-1, parça-2, parça-3, parça-4; dört yatay parçanın birleştirilmesiyle elde edilen bütün mimarinin tanıma oranı 4part; beş yatay parçanın birleştirilmesiyle elde edilen bütün mimarinin tanıma oranı 5part adıyla adlandırılarak her birinin *rank-1* doğruluğu üzerinden performans karşılaştırması sunulmuştur.

Tablo 16. CASIA-B veri seti tüm varyasyonlar parça ve bütün bazlı mimarilerin ortalama *rank-1* (%) oranları karşılaştırması

Galeri NM#1-4		0° -180°											ort
Sorgu		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	
NM	parça-1	46.6	51.9	57.5	62.5	59.8	58.3	59.2	61.0	62.0	52.9	43.1	55.9
#5-6	parça-2	49.9	58.6	61.0	64.8	59.8	49.8	52.7	62.3	64.0	58.9	48.1	57.3
	parça-3	50.0	60.3	67.5	66.4	66.0	57.0	64.5	67.8	67.6	63.5	50.2	61.9
	parça-4	40.9	56.8	59.6	62.4	58.3	53.9	57.9	57.0	54.3	50.4	35.7	53.4
	4part	83.5	93.8	95.6	95.7	90.5	85.6	90.5	92.2	95.3	91.4	81.5	90.5
	5part	<b>84.7</b>	<b>93.9</b>	<b>96.9</b>	94.7	89.6	<b>86.2</b>	88.9	<b>92.8</b>	<b>95.6</b>	<b>91.8</b>	<b>83.2</b>	<b>90.8</b>
BG	4part	79.2	86.6	87.7	85.4	78.8	73.1	77.2	83.9	86.9	81.8	75.0	81.4
#1-2	5part	<b>80.0</b>	<b>88.5</b>	86.2	83.2	77.4	<b>74.2</b>	<b>79.5</b>	<b>86.0</b>	<b>89.8</b>	<b>85.6</b>	<b>77.1</b>	<b>82.5</b>
CL	4part	37.7	48.5	51.9	53.8	51.1	46.4	46.4	42.8	42.1	39.9	32.2	44.8
#1-2	5part	<b>46.1</b>	<b>56.3</b>	<b>61.6</b>	<b>60.4</b>	<b>55.5</b>	<b>50.4</b>	<b>53.1</b>	<b>47.5</b>	<b>49.4</b>	<b>47.9</b>	<b>40.0</b>	<b>51.7</b>

Tablo 16 incelendiğinde, dört yatay vücut parçasının her birinin tanımadaki farklı oranlarda etkiye sahip oldukları ve bireysel tanımadaki oldukça düşük performans gösterdikleri açıktır. Ancak bu parçalar 4part bütün mimarisi üzerinden bir araya getirilerek tanıma sağlandığında *rank-1* değerinde ciddi bir artış yaşanmıştır. Bu artış oranı parça-bazlı mimarinin yürüyüş tanımadaki performansı ne derece artırdığını ortaya koymaktadır. Buna ek olarak, beş parçalı 5part bütün mimarisi, 4part mimarisine göre daha yüksek tanıma oranına ulaşmıştır. Bu durum, BG ve CL varyasyonlarında çok daha fark edilebilir bir etki göstermiştir. Özellikle CL varyasyonunda tanıma oranındaki artış %5’ in üzerine çıkmıştır. Bu iyileşme, varyasyondan etkilenen bölgenin daha dar bir alana sıkıştırılması ve

etkilenmeyen parça sayısının artmasıyla ilişkilendirilebilir. Bu sebeple tez çalışması kapsamında, ana yaklaşım olarak beş parçalı bir mimari benimsenmiş, varyasyondan etkilenen bölgelerin daha dar bir alana sıkıştırılmasıyla etkilenmeyen bölgeler üzerinden yürüyüş tanıma sağlanarak doğruluğun artırılması hedeflenmiştir. Ayrıca, bu yaklaşım ile önceki bölümlerde ayrıntılı biçimde açıklanan parça eleme işleminin etkinliği artırılarak tanıma performansının daha da iyileştirilmesi amaçlanmıştır.

### 3.2.2. Parça Eleme İşleminin Tanıma Etkisi

Parça eleme işlemi, hazırlanan topluluk öğrenmeli parça-tabanlı ağ mimarisinin hem eğitim hem de test aşamasında kullanılmıştır. Bu sebeple, bu işlem sadece görünüm tabanlı varyasyonlara sahip görüntülerin doğru tanınmasına etkili olmakla kalmamış, varyasyonlu parçaların eğitime dahil olmasını engelleyerek, varyasyon barındırmayan görüntülerin de doğru sınıflandırılmasına katkıda bulunmuştur. Parça eleme işleminin tanıma etkisini analiz edebilmek amacıyla Tablo 17’de CASIA-B veri seti, tüm varyasyonlar altında, 5 parçalı mimarinin parça elemeli (5part+elm) ve elemesiz (5part) versiyonlarının *rank-1* doğruluğu üzerinden tanıma oranları sunulmuştur.

Tablo 17. CASIA-B veri seti tüm varyasyonlar parça elemeli (5part+elm) ve elemesiz 5 parçalı öğrenme (5part) *rank-1*(%) oranları

Galeri NM#1-4		0° -180°											ort
Sorgu		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	
NM	5part	84.7	93.9	96.9	94.7	89.6	86.2	88.9	92.8	95.6	91.8	83.2	90.8
#5-6	5part+elm	<b>85.3</b>	92.7	96.9	<b>96.1</b>	<b>90.8</b>	<b>87.1</b>	<b>89.9</b>	<b>94.6</b>	<b>95.8</b>	<b>93.3</b>	83.1	<b>91.4</b>
BG	5part	80.0	88.5	86.2	83.2	77.4	74.2	79.5	86.0	89.8	85.6	77.1	82.5
#1-2	5part+elm	78.1	88.1	<b>88.4</b>	<b>87.2</b>	<b>80.0</b>	<b>77.0</b>	79.0	<b>86.6</b>	<b>90.5</b>	84.8	75.8	<b>83.2</b>
CL	5part	46.1	56.3	61.6	60.4	55.5	50.4	53.1	47.5	49.4	47.9	40.0	51.7
#1-2	5part+elm	<b>52.5</b>	<b>68.1</b>	<b>71.7</b>	<b>67.0</b>	<b>66.0</b>	<b>62.4</b>	<b>61.4</b>	<b>61.6</b>	<b>64.8</b>	<b>61.5</b>	<b>46.5</b>	<b>62.1</b>

Tablo 17 incelendiğinde, 5 parçalı mimarinin (5part) hem eğitim hem de test aşamalarına parça çıkarma işlemi (5part+elm) uygulandığında tanıma başarısında anlamlı bir iyileşme olduğu açıktır. Bu iyileşme, NM varyasyonu altında %0.6, BG varyasyonu altında %0.7 ve CL varyasyonunda çok daha belirgin olup %10 oranının üzerinde bir artış ile sağlanmıştır. Bu artış, varyasyon kaynaklı bozulmalara sahip vücut bölgelerinin eğitim sürecinden çıkarılmasıyla daha kararlı bir öğrenmenin gerçekleşmesine ve test aşamasında

daha tutarlı benzerlik hesaplamalarının yapılabilmesine bağlanabilir. Buna ek olarak, 5 parçalı bir mimariye parça çıkarma işleminin uygulanması, dar bir yatay alana sıkıştırılmış alakasız nesnelere içeren vücut bölgelerinin elenmesini sağlayarak modelin dayanıklılığını ve güvenilirliğini artırmıştır.

### 3.2.3. Topluluk Öğrenmeli Parça-Tabanlı Ağ Mimarisinin Performans Değerlendirmesi

Bu bölümde, tez kapsamında hazırlanan ikinci ve son topluluk öğrenmeli parça-tabanlı ağ mimarisinin elde edilen nihai deney sonuçlarının detaylı analizi yapılmıştır. Bu amaçla, başta yürüyüş tanıma için kapsamlı bir laboratuvar veri seti olan CASIA-B ile iki dış ortam veri seti CASIA-C ve Outdoor-Gait kullanılmıştır. Bu veri setleri için eğitim, test ve galeri-sorgu kümelenmeleri bölüm 2.5.1’de detaylandırıldığı şekilde yapılmıştır. Mimarinin, ilk veri seti CASIA-B üzerindeki tüm varyasyonlar altındaki *rank-1* doğruluğu analizleri Tablo 18’de sunulmuştur.

Tablo 18. CASIA-B tüm varyasyonlar, farklı parça tabanlı mimarilerin *rank-1* oranı (%) üzerinden karşılaştırılması (özdeş açılar hariç)

Galeri NM#1-4		0° - 180°										ort	
Sorgu		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	
NM	4part	83.5	93.8	95.6	95.7	90.5	85.6	90.5	92.2	95.3	91.4	81.5	90.5
#5-6	5part	84.7	93.9	96.9	94.7	89.6	86.2	88.9	92.8	95.6	91.8	83.2	90.8
	5part+elm	85.3	92.7	96.9	<b>96.1</b>	90.8	87.1	89.9	<b>94.6</b>	95.8	93.3	83.1	91.4
	5part+Tpl (Kkatlı)	87.7	<b>95.5</b>	<b>97.9</b>	95.7	<b>92.5</b>	88.3	89.8	93.2	<b>96.9</b>	94.2	84.2	92.3
	5part+Tpl (Önyük)	86.5	94.9	96.9	94.1	89.8	87.3	89.7	92.5	95.3	93.6	84.3	91.3
	5part+elm+Tpl (Kkatlı)	87.3	94.8	97.5	96.0	92.0	<b>88.6</b>	90.6	93.7	<b>96.9</b>	94.0	85.1	<b>92.4</b>
	5part+elm+Tpl (Önyük)	<b>88.3</b>	94.6	97.3	94.9	90.8	88.5	<b>90.7</b>	93.4	96.1	<b>94.6</b>	<b>85.7</b>	92.3
BG	4part	79.2	86.6	87.7	85.4	78.8	73.1	77.2	83.9	86.9	81.8	75.0	81.4
#1-2	5part	80.0	88.5	86.2	83.2	77.4	74.2	79.5	86.0	89.8	85.6	77.1	82.5
	5part+elm	78.1	88.1	88.4	87.2	80.0	77.0	79.0	86.6	90.5	84.8	75.8	83.2
	5part+Tpl (Kkatlı)	<b>80.1</b>	88.4	88.8	87.0	78.8	73.6	78.1	86.6	90.7	85.9	78.2	83.3
	5part+Tpl (Önyük)	78.8	88.6	<b>89.5</b>	86.4	79.7	73.7	79.5	85.8	89.4	<b>86.7</b>	77.8	83.3
	5part+elm+Tpl (Kkatlı)	79.6	87.5	87.9	<b>87.6</b>	80.0	76.5	81.1	<b>86.7</b>	91.2	86.0	<b>78.9</b>	83.9
	5part+elm+Tpl (Önyük)	77.9	<b>89.4</b>	89.3	87.1	<b>80.1</b>	<b>78.1</b>	<b>82.2</b>	86.6	<b>92.0</b>	85.9	78.6	<b>84.3</b>
CL	4part	37.7	48.5	51.9	53.8	51.1	46.4	46.4	42.8	42.1	39.9	32.2	44.8
#1-2	5part	46.1	56.3	61.6	60.4	55.5	50.4	53.1	47.5	49.4	47.9	40.0	51.7
	5part+elm	52.5	68.1	<b>71.7</b>	67.0	<b>66.0</b>	<b>62.4</b>	<b>61.4</b>	<b>61.6</b>	<b>64.8</b>	<b>61.5</b>	46.5	<b>62.1</b>
	5part+Tpl (Kkatlı)	43.0	54.0	56.9	55.8	50.4	46.4	45.0	43.4	44.7	45.1	38.5	47.6
	5part+Tpl (Önyük)	47.9	58.7	62.5	61.2	53.9	50.2	51.2	48.9	52.5	52.7	44.7	53.1
	5part+elm+Tpl (Kkatlı)	50.4	68.3	70.5	<b>67.4</b>	62.2	57.7	58.9	60.6	63.3	58.6	46.9	60.4
	5part+elm+Tpl (Önyük)	<b>53.4</b>	<b>68.7</b>	68.0	65.0	61.9	58.4	60.1	61.0	63.2	59.2	<b>47.9</b>	60.6

Tablo 18’de, daha önceki bölümlerde detaylı analizleri yapılan 4 parçalı (4part), 5 parçalı (5part) ve parça eleme işlemi uygulanmış 5 parçalı (5part+elm) mimarilerin de

sonuçları, iki farklı yöntemeye dayanan — $k$ -katlı çapraz doğrulama (Kkatlı) ve önyükleme toplulaştırması (Önyük) —topluluk öğrenmeli parça tabanlı mimarilerin ( $5part+Tpl$ ) tanıma başarılarıyla kolay karşılaştırılabilmesi amacıyla tabloya yeniden dahil edilmiştir. Tabloda, parça eleme ve topluluk öğrenmeyi bir arada kullanan, tez kapsamında önerilen topluluk öğrenmeli parça-tabanlı ağ mimarisinin ( $5part+elm+Tpl$ ) nihai tanıma oranları her kategorinin son iki satırında sunulmuştur.

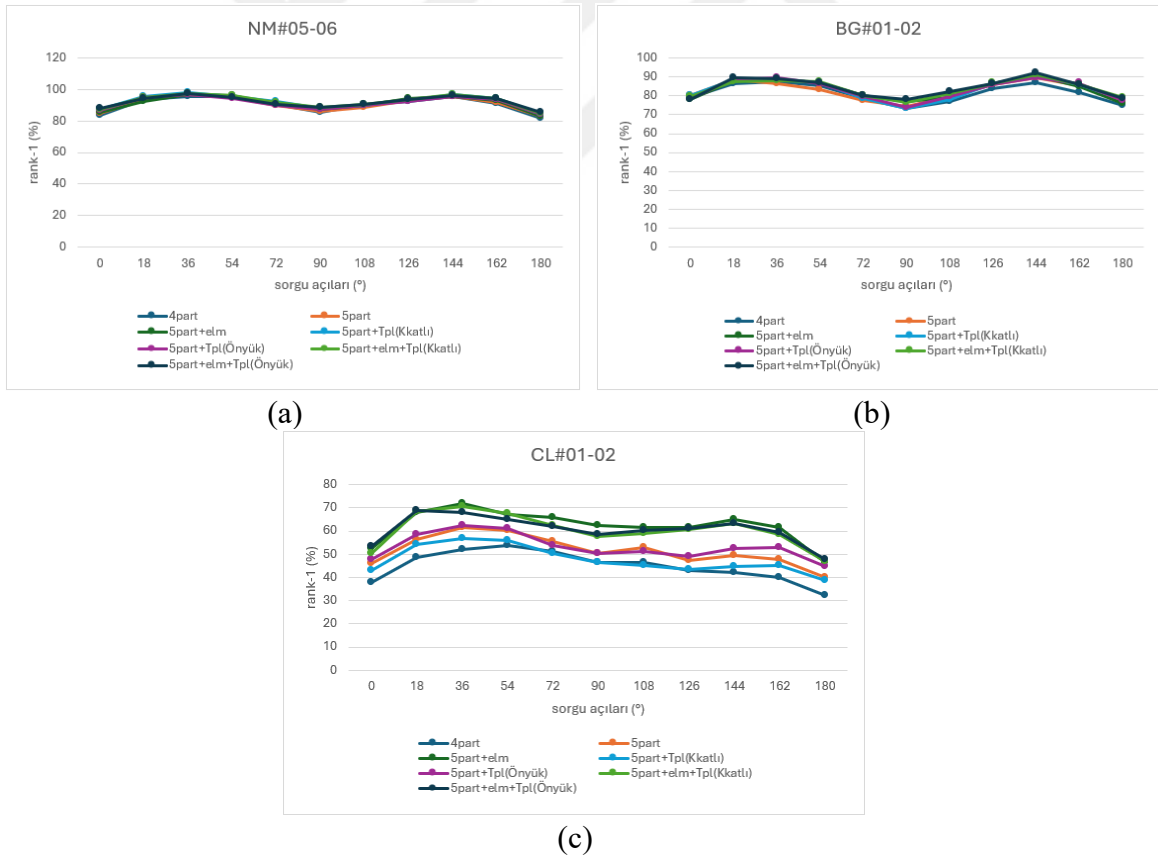
Tablo 18' den, beş parçalı öğrenme yaklaşımına topluluk öğrenmesi ( $5part+Tpl$ ) entegre edilerek anlamlı performans artışları elde edildiği gözlemlenmektedir. Özellikle,  $k$ -katlı çapraz doğrulama ( $5part+Tpl(Kkatlı)$ ) ve önyükleme toplulaştırma ( $5part+Tpl(Önyük)$ ) yöntemleriyle elde edilen tanıma oranları, standart beş parçalı öğrenme yaklaşımına ( $5part$ ) kıyasla belirgin şekilde yüksektir.  $5part+Tpl(Kkatlı)$  yöntemi, NM varyasyonu altında  $5part$ 'a göre daha yüksek bir tanıma başarımı sergilese de, BG varyasyonunda bu başarı  $5part+Tpl(Önyük)$  ile eşit düzeyde kalmaktadır. CL varyasyonunda ise  $5part+Tpl(Önyük)$ ' ten daha düşük bir performans göstermiştir. CL varyasyonundaki bu düşüş,  $k$  alt kümeyle bölünerek oluşturulan veri alt kümelerinde varyasyondan etkilenmeyen örnek sayısının, yer değiştirmeli rastgele örnekleme yöntemiyle elde edilen veri kümelerine kıyasla daha az sayıda olabilmesiyle açıklanabilir.

Tablo 18, parça eleme işleminin topluluk öğrenme yaklaşımlarına ( $5part+elm+Tpl$ ) entegre edilmesinin,  $5part+Tpl$  yönteminin tanıma performansını daha da artırdığını göstermektedir. Bu iyileşme,  $k$ -katlı çapraz doğrulama ve önyükleme toplulaştırma yaklaşımları kapsamında her bir varyasyon için farklı oranlarda gerçekleşmiştir. Özellikle, önyükleme toplulaştırma yaklaşımı ( $5part+elm+Tpl(Önyük)$ ) için NM ve BG varyasyonlarında,  $k$ -katlı çapraz doğrulama yaklaşımı ( $5part+elm+Tpl(Kkatlı)$ ) için CL varyasyonunda bu etki daha belirgin hale gelmiştir. Söz konusu iyileşme oranları sırasıyla %1, %1 ve %12.8 olarak gözlemlenmiştir.

Tablo 18 incelendiğinde, parça çıkarma işlemi uygulanmış beş parçalı öğrenme yaklaşımı ( $5part+elm$ ) üzerinde topluluk öğrenme yaklaşımlarının etkisi değerlendirildiğinde, NM ve BG varyasyonlarında beklenen performans artışı gözlemlenmişken, CL varyasyonunda yaklaşık %2' lik bir düşüş meydana gelmiştir. Bu durum, CL varyasyonunda değişimden etkilenen parça sayısının daha fazla olması ve topluluk öğrenmesinin geriye kalan parçalara katkısının sınırlı kalmasıyla açıklanabilir. Bununla birlikte, her iki topluluk öğrenme yaklaşımının da ( $5part+elm+Tpl$ ) tüm varyasyonlarda —NM, BG ve CL— birbirine yakın tanıma başarımları elde ettiği

söylenbilir. Sadece NM varyasyonunda,  $k$ -katlı çapraz doğrulama yöntemi, önyükleme toplulaştırma yöntemine kıyasla bir miktar daha iyi sonuç vermiştir. Sonuç olarak, 5part+elm+Tpl(Kkatlı) ve 5part+elm+Tpl(Önyük) yöntemleri, NM ve BG varyasyonlarında en yüksek performansa ulaşırken; CL varyasyonunda en yüksek doğruluk oranı 5part+elm yöntemiyle elde edilmiştir. Topluluk öğrenme yaklaşımları arasında ise, 5part+elm+Tpl (Önyük) yöntemi yalnızca CL varyasyonunun  $0^\circ$  ve  $180^\circ$  görüş açılarında en iyi sonuçları vermiştir.

Tablo 18' deki sonuçlar üzerinden elde edilen Şekil 29, tüm yöntemlerin NM varyasyonunda benzer performans sergilediğini, BG varyasyonunda yöntemler arası performans farklılıklarının belirginleşmeye başladığını ve CL varyasyonunda ise açık bir ayrışmanın ortaya çıktığını göstermektedir. Bu ayrışma, parça eleme işleminin varyasyonlar üzerindeki etkisinden kaynaklanmakta olup topluluk öğrenme kararlarının tutarlılığını etkilemekte ve gözlemlenen performans farklarına katkı sağlamaktadır.

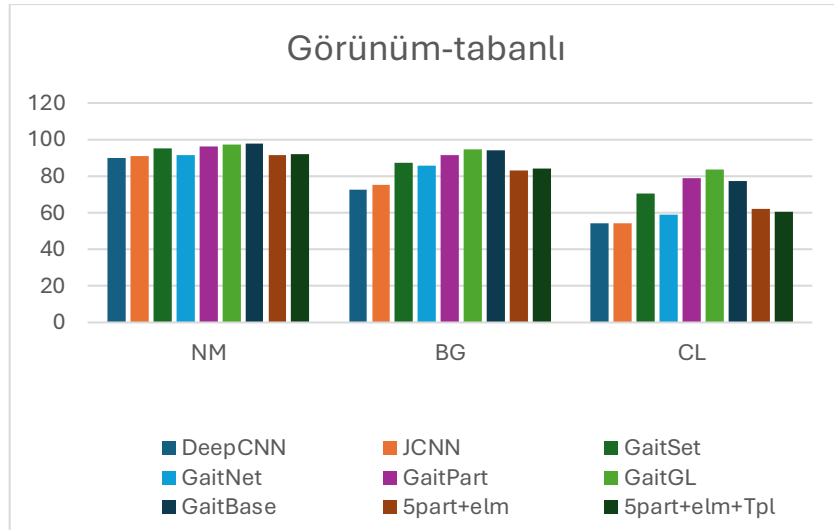


Şekil 29. CASIA-B veri seti tüm varyasyonlar hazırlanan parça tabanlı öğrenme mimarilerinin karşılaştırılması (a) NM, (b) BG, (c) CL

Tablo 19’ da, tez kapsamında hazırlanan parça eleme süreçli ağ mimarisinin topluluk öğrenme içeren (5part+elm+Tpl) ve içermeyen (5part+elm) versiyonlarının, literatürdeki son teknoloji görünüm tabanlı ve model tabanlı yöntemlerle karşılaştırmalı bir analizi sunulmuştur. Karşılaştırma için görünüm tabanlı yöntemlerden DeepCNN (Wu vd., 2016), JCNN (Zhang vd., 2019), GaitSet (Chao vd., 2019), GaitNet (Zhang vd., 2019), GaitPart (Fan vd., 2020), GaitGL (Lin vd., 2021) ve GaitBase (Fan vd., 2023) seçilmiştir. Ayrıca, bu yöntemlerle performans karşılaştırmasına yönelik grafik Şekil 30’ da sunulmuştur.

Tablo 19. CASIA-B veri seti, tüm varyasyonlar, görünüm ve model tabanlı yöntemler ile ortalama *rank-1* (%) üzerinden karşılaştırma (özdeş açılar hariç)

Tür	Yöntemler	NM	BG	CL
Görünüm-tabanlı	DeepCNN (Wu vd., 2016)	89.9	72.4	54.0
	JCNN (Zhang vd., 2019)	91.2	75.0	54.0
	GaitSet (Chao vd., 2019)	95.0	87.2	70.4
	GaitNet (Zhang vd., 2019)	91.6	85.7	58.9
	GaitPart (Fan vd., 2020)	96.2	91.5	78.7
	GaitGL (Lin vd., 2021)	97.4	<b>94.5</b>	<b>83.6</b>
	GaitBase (Fan vd., 2023)	<b>97.6</b>	94.0	77.4
	Önerilen	5part+elm	91.4	83.2
	5part+elm+Tpl	92.3	84.3	60.6
Model-tabanlı	PoseGait (Liao vd., 2020)	68.7	44.5	36.0
	GaitGraph (Teepe vd., 2021)	87.7	74.8	66.3
	GaitGraph2 (Teepe vd., 2022)	82.0	73.2	63.6
	ResGait (Gao vd., 2023)	89.6	79.2	<b>70.2</b>
	GaitDLF (Wei vd., 2024)	84.9	70.7	67.9
	MSRG (Chen vd., 2024)	89.0	75.6	70.0
	Önerilen	5part+elm	91.4	83.2
	5part+elm+Tpl	<b>92.3</b>	<b>84.3</b>	60.6

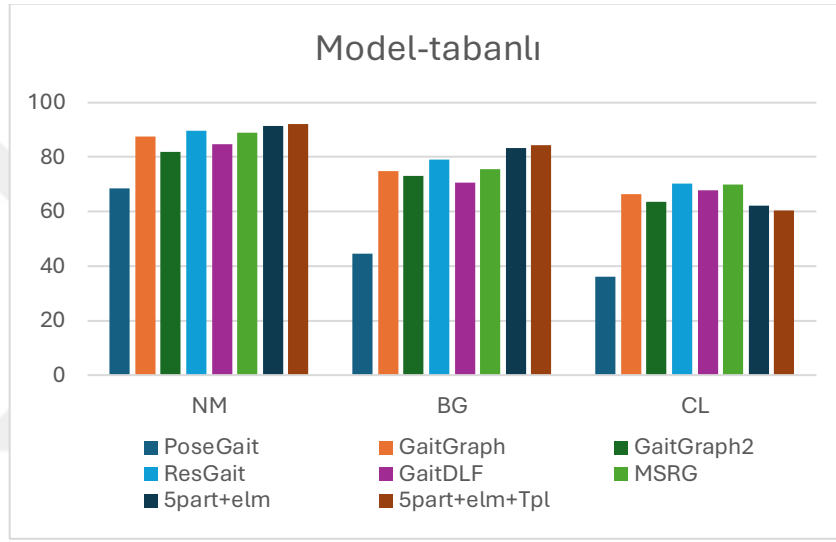


Şekil 30. CASIA-B veri seti, görünüm tabanlı yöntemlerin *rank-1* oranları karşılaştırması

Tablo 19’ da elde edilen sonuçlar ve Şekil 30, önerilen topluluk öğrenmeli parça-tabanlı ağ mimarisinin (5part+elm+Tpl), küresel özniteliklere odaklanan görünüm tabanlı DeepCNN, JCNN ve GaitNet yöntemlerinden daha iyi performans sergilediğini, ancak GaitSet yöntemine göre daha düşük bir performansa sahip olduğunu göstermektedir. GaitSet, giriş verisi bir dizi yürüyüş silüetini kullanarak, son öznitelik temsilini elde edebilmek için HPM yönteminden faydalanmaktadır. Bu sayede, çok boyutlu parça düzeyinde öznitelikler çıkararak hem yerel hem de küresel yürüyüş bilgilerini toplar. Öte yandan, önerilen yöntemin performansı, yerel özniteliklere odaklanan GaitPart, GaitGL ve GaitBase yöntemlerine kıyasla daha düşüktür. Bu yöntemler, yürüyüşe özgü en anlamlı yerel özellikleri etkin bir şekilde çıkarmak üzere tasarlanmıştır. Örneğin, GaitPart öznitelik haritasını yatay olarak birçok parçaya ayırarak parça düzeyinde uzaysal özniteliklerin ayrıntılı şekilde öğrenilmesini sağlayan Odaklı Konvolüsyon (Focal Convolution, FConv) Katmanı kullanır. GaitGL, hem küresel hem de yerel bilgileri içeren daha kapsamlı öznitelikler çıkarmayı hedefleyen yenilikçi bir GLFE modülü sunarak komşu parçalar arasındaki ilişkileri de yakalar. GaitBase ise silüetlerin 3B öznitelik haritasını yatay olarak birden fazla parçaya ayırarak, bu parçaları daha ayırt edilebilir bir öznitelik uzayına yansıtan bir çerçeve kullanmaktadır. Bu yaklaşımlarda, yürüyüşe özgü desenlere özel olarak tasarlanan modüllerin varlığı, tanıma başarımını önemli ölçüde artırmıştır. Buna karşın, tez kapsamında önerilen topluluk öğrenmeli parça-tabanlı ağ mimarisi daha az karmaşıklığa sahip olup uygulaması daha kolaydır. Yürüyüşe özgü modüllere bağlı kalmadan, parça tabanlı yöntemlerde topluluk öğrenmesinin temel etkisini ortaya koyan genel bir çerçeve sunmaktadır. Elde edilen sonuçlar, topluluk öğrenmesinin uygulanmasıyla performansın artırılabilirliğini göstermektedir. Ayrıca, topluluk öğrenmesi yaklaşımlarının yürüyüşe özgü desenler için özel olarak tasarlanan modüllerle zenginleştirilmesi, performansın daha da artırılmasına katkı sağlayabilir.

Tablo 19 ‘da tez kapsamında önerilen topluluk öğrenmeli parça-tabanlı ağ mimarisinin performansını son teknoloji model tabanlı yöntemlerle karşılaştırmak için PoseGait (Liao vd., 2020), GaitGraph (Teepe vd., 2021), GaitGraph2 (Teepe vd., 2022), ResGait (Gao vd., 2023), GaitDLF (Wei vd., 2024) ve MSRG (Chen vd., 2024) yöntemleri seçilmiştir. Tablo incelendiğinde önerilen mimari, model tabanlı yöntemlere göre NM ve BG varyasyonlarında en iyi performansı sergilerken; CL varyasyonunda ise diğer yöntemlerle karşılaştırılabilir sonuçlar elde etmiştir. Özellikle, yakın zamandaki en popüler yöntem olan GaitGraph ile karşılaştırıldığında, önerilen topluluk öğrenme yaklaşımı NM varyasyonunda %4.6 ve BG

varyasyonunda %9.5 oranında iyileşme sağlamış, CL varyasyonunda ise %60.6 doğruluk oranına ulaşmıştır. Benzer şekilde, en güncel yöntemler arasında yer alan GaitDLF ve MSRSG ile kıyaslandığında, önerilen topluluk öğrenmeli parça-tabanlı ağ mimarisi, CASIA-B veri setinde 74 denekten oluşan eğitim kümesi üzerinde NM ve BG varyasyonlarında güçlü bir performans sergilemiştir. CL varyasyonundaki sınırlı iyileşme ise model tabanlı özniteliklerin, giyim varyasyonuna bağlı bozulmalara karşı daha dayanıklı olmasından kaynaklanıyor olabilir. Tablo 19’ dan yola çıkılarak hazırlanan performans karşılaştırma grafiği Şekil 31’ de sunulmuştur.



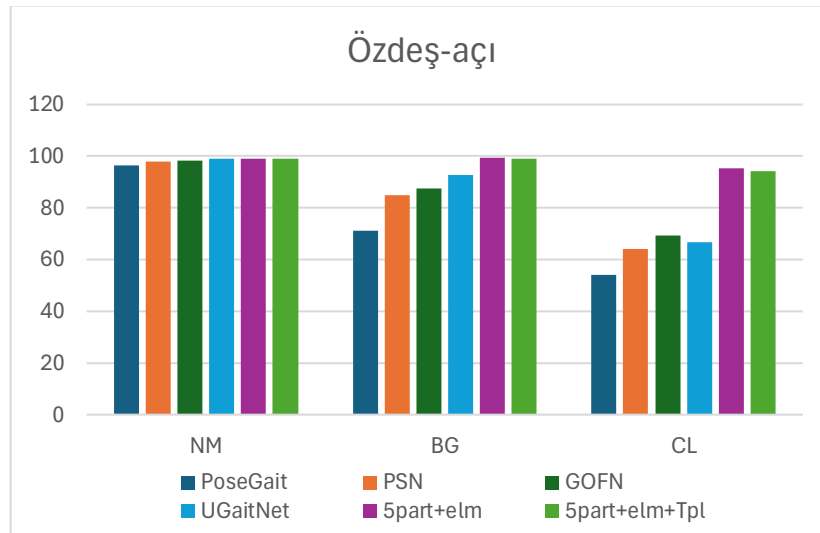
Şekil 31. CASIA-B veri seti, model tabanlı yöntemlerin *rank-1* oranları karşılaştırması

Tablo 20’ de, tez kapsamında hazırlanan ağ yapısı son olarak, özdeş açılı koşulu altında literatürde önde gelen çalışmalarla karşılaştırılmıştır. Bu amaçla, literatürde CASIA-B veri seti özdeş açılı performans sonuçlarını sunan PoseGait, PSN (Xu vd.,2021), GOFN (Ye vd., 2023) ve UGaitNet (Marin-Jimenez vd., 2021) çalışmaları seçilmiştir. Tablo’ya göre NM varyasyonunda tüm yöntemler, eğitim veri setinde 62 veya 74 birey kullanmış olsa da, oldukça benzer performans sergilemişlerdir. Yöntemler arasındaki küçük fark, büyük olasılıkla eğitim veri setinin boyutundaki farklılıktan kaynaklanmaktadır. Ancak bu fark, BG ve CL varyasyonlarında belirgin şekilde artmaktadır. Bu durum kısmen eğitimde kullanılan birey sayısındaki farklılığa bağlanabilir ancak, temel etken kullanılan yaklaşımlar arasındaki farklılıklardır. Örneğin, PoseGait, GOFN ve UGaitNet küresel yaklaşımlar benimserken, PSN eklem noktalarının yerel desenlerinden yararlanarak kısmen yerel bir yaklaşım sunmaktadır. Buna karşın, önerilen 5part+elm yöntemi, her yatay parçanın desenine ayrı ayrı

odaklanarak ve parça çıkarma sürecinden faydalanarak BG ve CL varyasyonlarında tanıma başarımını anlamlı ölçüde artırmıştır. Ayrıca, 5part+elm yönteminde topluluk öğrenmesinin tanıma performansı üzerindeki etkisi değerlendirildiğinde, Tablo 18’deki sonuçlarla benzer şekilde, özellikle CL varyasyonunda parça çıkarma sürecinin topluluk kararlarının tutarlılığı üzerinde önemli bir etkisi olduğu gözlemlenmiştir. Tablo 20’deki sonuçlara dayanarak oluşturulan grafiksel analiz Şekil 32’de sunulmuştur.

Tablo 20. CASIA-B veri seti tüm varyasyonlar, önerilen yöntemlerin özdeş açılı altında tanıma doğruluğu (%) performans karşılaştırması

Galeri NM#1-4			0° -180°										ort	
Sorgu			0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	
NM #5-6	62id	PoseGait	96.0	96.8	96.0	96.8	96.0	97.6	97.6	94.4	96.8	97.6	97.6	96.6
		PSN	99.2	99.2	98.4	99.2	97.6	98.4	97.6	96.8	97.6	97.6	99.2	98.1
		GOFN	98.4	99.2	99.2	97.6	97.6	96.8	96.8	99.2	99.2	98.4	97.6	98.2
	74id	UGaitNet	99	100	100	99	100	100	100	98	98	96	98	98.9
		5part+elm	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>98</b>	<b>97</b>	<b>98</b>	<b>98</b>	<b>99.2</b>
		5part+elm+Tpl	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>99</b>	<b>100</b>	<b>98</b>	<b>98</b>	<b>98</b>	<b>99.2</b>
BG #1-2	62id	PoseGait	74.2	75.8	77.4	76.6	69.4	70.2	71.0	69.4	74.2	65.3	60.5	71.3
		PSN	86.3	84.7	83.1	88.7	90.3	86.3	90.3	83.9	84.7	76.6	80.7	85.0
		GOFN	84.7	88.7	92.7	91.1	85.5	80.7	87.1	88.7	91.9	91.1	79.9	87.5
	74id	UGaitNet	98	97	95	94	88	90	88	91	93	95	91	92.7
		5part+elm	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>98</b>	<b>98</b>	<b>98</b>	<b>98</b>	<b>99.3</b>
		5part+elm+Tpl	<b>100</b>	99	99	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>98</b>	<b>98</b>	<b>98</b>	97	99.0
CL #1-2	62id	PoseGait	46.8	48.4	57.3	61.3	58.1	56.5	59.7	54.8	55.7	58.1	39.5	54.2
		PSN	64.5	68.6	70.2	71.0	68.6	64.5	62.9	56.5	59.7	59.7	60.5	64.2
		GOFN	67.7	72.8	77.6	73.4	67.0	67.7	66.1	68.5	69.3	68.5	64.5	69.4
	74id	UGaitNet	59	64	70	67	68	70	71	69	71	68	57	66.7
		5part+elm	<b>97</b>	<b>97</b>	<b>98</b>	<b>95</b>	<b>98</b>	95	93	<b>93</b>	<b>94</b>	<b>95</b>	<b>94</b>	<b>95.4</b>
		5part+elm+Tpl	<b>97</b>	96	94	93	97	<b>96</b>	<b>94</b>	<b>93</b>	92	93	90	94.1



Şekil 32. CASIA-B veri seti önerilen yöntemlerin mevcut yöntemler ile özdeş açılı altında tanıma doğruluğu karşılaştırması

Tez kapsamında önerilen topluluk öğrenmeli parça-tabanlı ağ mimarisinin performansı ikinci olarak bir kızılötesi ve dış ortam veri seti olan CASIA-C üzerinde test edilmiştir. Karşılaştırma için Tablo 21’ de literatürde bu veri seti üzerinde analiz yapmış olan PSN, GOFN ve GaitSTAR (Bilal vd., 2024) yöntemleri seçilmiştir. Bu veri seti için topluluk öğrenme yöntemi olarak  $k$ -katlı çapraz doğrulama kullanılmıştır.

Tablo 21. Önerilen yöntemin CASIA-C veri seti, mevcut yöntemlerle tanıma doğruluğu karşılaştırması

Eğitim	Yöntemler	fb	fq	fs	ort
24id	PSN (Xu vd.,2021)	38.1	58.6	56.0	50.9
	GOFN (Ye vd., 2023)	39.5	64.2	54.3	52.7
	GaitSTAR (Bilal vd., 2024)	40.7	65.3	56.4	54.1
	5part	84.2	96.3	95.4	92.0
	5part+Tpl	<b>93.5</b>	<b>99.1</b>	<b>96.3</b>	<b>96.3</b>
62id	PSN (Xu vd.,2021)	42.3	63.6	60.0	55.3
	GOFN (Ye vd., 2023)	43.5	69.3	58.2	57.0
	GaitSTAR (Ye vd., 2023)	44.6	71.5	59.1	58.4
	5part	97.2	<b>99.1</b>	97.2	97.8
	5part+Tpl	<b>99.1</b>	<b>99.1</b>	<b>98.1</b>	<b>98.8</b>

Tablo 21’ de sunulduğu üzere, tez kapsamında önerilen 5part+Tpl yöntemi, 24 kişi ile eğitildiğinde %96.3 ve 62 kişi ile eğitildiğinde %98.8 oranında tanıma doğruluğu elde ederek en yüksek performansa sahip olmuştur. Tabloya ilişkin yapılan analiz, topluluk öğrenme yaklaşımının, hem 24 hem de 62 kişi ile eğitilen 5part yönteminin tanıma başarımını önemli ölçüde artırdığını ortaya koymuştur. Özellikle, ortalama (ort) tanıma doğruluk oranındaki artış, 24 kişi için %4.3, 62 kişi için ise %1 olarak gözlemlenmiştir. Literatürdeki diğer çalışmalarla karşılaştırıldığında, önerilen topluluk öğrenme yöntemi, PSN, GOFN ve GaitSTAR yöntemlerinden daha başarılı bir performans sergilemiştir. Bu durum, PSN yönteminin model tabanlı bir yaklaşımı benimsemesi nedeniyle termal (kızılötesi) görüntüler için uygun olmayabileceği gerekçesiyle açıklanabilir. Benzer şekilde, termal görüntülerden optik akış özellikleri çıkaran GOFN ve GaitSTAR yöntemlerinde tanıma süreci benzer sınırlamalara maruz kalmış olabilir. Buna karşılık, önerilen yöntem GEI özniteliklerinden faydalanmakta olup, bu da termal görüntüler için daha uygun bir girdi temsili sunmuştur. Ayrıca, Tablo 20’ deki sonuçlarla paralel şekilde, önerilen yöntemin yürüyüş deseninin her yatay bölümünü ayrı ayrı analiz edebilme yetisi, özdeş görüş açısı koşullarında iyi performans elde etmesini sağlamıştır.

Topluluk öğrenmeli parça-tabanlı ağ mimarisinin performans analizi son olarak, bir dış ortam veri seti olan Outdoor-Gait üzerinde gerçekleştirilmiştir. Tablo 22’ de bu veri seti üzerinde literatürden seçilen GaitNet (Song vd., 2019) ve model tabanlı 3DHuman (Meng vd., 2023) çalışmaları ile hazırlanan mimarilerin performans karşılaştırması sunulmuştur. 5part+Ens yöntemi için topluluk öğrenmesi yaklaşımı olarak  $k$ -katlı çapraz doğrulama tercih edilmiştir.

Tablo 22. Önerilen yöntemin Outdoor-Gait veri seti üzerinden mevcut yöntemlerle karşılaştırması

	Galeri Sorgu	NM			BG			CL			ort
		NM	BG	CL	NM	BG	CL	NM	BG	CL	
GaitNet		96.9	89.1	60.2	<b>92.0</b>	97.1	59.7	58.7	55.4	97.3	78.5
3Dhuman		90.0	83.6	80.9	77.5	90.3	72.6	72.2	66.4	88.8	80.2
5part		<b>98.3</b>	90.8	<b>85.6</b>	88.4	98.1	74.6	<b>80.4</b>	72.9	97.8	87.4
5part+Ens		97.8	<b>91.1</b>	84.5	89.7	<b>98.5</b>	<b>77.4</b>	79.3	<b>75.1</b>	<b>98.3</b>	<b>88.0</b>

Tablo 22’ den görülebileceği üzere, 5part+Ens yaklaşımı, GaitNet ve 3Dhuman gibi diğer çalışmalara kıyasla en yüksek tanıma başarımına ulaşmıştır. Ayrıca, topluluk öğrenmesi, 5part yaklaşımı için ortalama tanıma başarı oranı olan  $rank-1$  değerinde genel bir iyileşme sağlamıştır. Ancak, bazı varyasyonlarda topluluk öğrenmesi performansı artırmada yetersiz kalmıştır. Örneğin; galeri kümesinin NM, sorgu kümesinin ise NM veya CL olduğu durumlarda ya da galeri kümesinin CL, sorgu kümesinin NM olduğu senaryolarda topluluk öğrenmesi tanıma oranını yükseltmemiştir. Ayrıca, Tablo 20’ de sunulan CASIA-B ve Tablo 21’ de yer alan özdeş aç CASIA-C veri setleri—ki Outdoor-Gait veri kümesi de özdeş aç temelli bir veri kümesidir—ile karşılaştırıldığında, tüm galeri ve sorgu senaryoları için tatmin edici bir performans elde edilememiştir. Buna sebep olarak, Outdoor-Gait veri setinin yalnızca görünüm tabanlı varyasyonları değil, aynı zamanda karmaşık arka plan sahne varyasyonlarını de içermesi gösterilebilir. Bu durum, temel sınıflandırıcıların topluluk öğrenmesi sırasında birbirinden oldukça farklı kararlar vermesine neden olarak, sağlam özniteliklerin çıkarılmasını engellemiş olabilir.

#### 4. SONUÇLAR

Bu tez kapsamında yapılan arařtırmalar ile derin öğrenmeye dayalı kullanılan girdi deseni ve mimari farklılıklarının yürüyüş tanımada etkinliklerinin doğru bir şekilde belirlenmesi için detaylı bir analiz sunulmuştur. Deneylerde bu amaçla, hem laboratuvar ortamında hem de doğal ortamlarda üretilmiş yürüyüş veri setleri kullanılmıştır. Yürüyüş tanıma yaklaşımları genellikle tanıma sürecini tek bir yürüyüş deseni üzerinden gerçekleştirmeye odaklanmıştır. Bu çalışmada ise ilk olarak farklı yürüyüş desenleri modalite olarak adlandırılmış ve bu modalitelerin birlikte kullanılması durumunda elde edilen tanıma performansları araştırılmıştır. Adil ve güvenilir sonuçlar elde edebilmek amacıyla hazırlanan ilk Çoklu Modalite Ağ Mimarisi'nde beş farklı yürüyüş desenine dayalı girdi kombinasyonu, literatürde başarısı ispatlanmış beş farklı ağ mimarisi kullanılarak yürüyüş tanıma performansı üzerinden, farklı yürüyüş varyasyonları altında detaylı bir karşılaştırma analizine tabi tutulmuştur. Bu mimaride kullanılan modaliteler arka planda iki temel gruba ait tutulmuştur. Bunlar global modaliteler ve yerel modaliteler olarak adlandırılmıştır. Global modalitelerde kişiye ait yürüyüş desenine bütünsel bir yaklaşım sunulurken, yerel modalitelerde yürüyüş deseninin yatay vücut bölmelerine odaklanılmıştır. Deneysel sonuçlar, farklı modalite kombinasyonlarının varyasyonlara bağlı olarak değişen performanslara sahip olduğunu gösterse de, geçmişteki bazı çalışmalara da paralel olarak, belirli vücut bölgelerinin kendine özgü bir tanıma performansı sergilediği görülmüştür. Bu nedenle global modaliteler üzerinden yapılan kombinasyonlara göre yerel modaliteler üzerinden yapılan kombinasyonların daha yüksek performansa eriştiği sonucuna varılmıştır. Bu performans artışı, özellikle varyasyonlu görüntülerin tanınmalarında daha belirgin olmuştur. Kullanılan son teknoloji ağ mimarilerinden ise MobileNet ve ConvNext tanımada ön plana çıkan iki mimari olarak gözlemlenmiştir. MobileNet, hem tanıma performansı hem de düşük hesaplama maliyeti ile ön plana çıkarken; ConvNext, yerel vücut parçalarına dayalı tanımada daha yüksek performansa erişmiştir.

Tez kapsamında hazırlanan ilk Çoklu Modalite Ağ Mimarisi' nin detaylı analizinden yola çıkarak tamamen yerel yürüyüş desenlerine odaklanan, ikinci Topluluk Öğrenmeli Parça-Tabanlı Ağ Mimarisi tez kapsamında önerilmiştir. Bu mimaride, hazırlanan birinci mimarideki yerel vücut parçalarına dayalı kombinasyonlar daha da genişletilmiş, bu parçalardan üretilen özniteliklerin sağlamlığını artırabilmek amacıyla yeni bir parça tabanlı

tanıma yaklaşımı, topluluk öğrenme yöntemleri kullanılarak sunulmuştur. Yatay ekseninde beş farklı bölgeye ayrılmış vücut deseninin her bir yerel parçası için ayrı bir CNN modeli eğitilmiştir. Bu CNN'ler tarafından üretilen özniteliklerin daha da güçlendirilmesi amacıyla topluluk öğrenmesi uygulanmış, ardından toplu CNN çıktılarının tek ve daha sağlam bir özellik temsiline dönüştürülmesini sağlayan bir Toplayıcı Model (TM) önerilmiştir. Geleneksel topluluk yöntemlerinin çoğunluk oylaması ya da ağırlıklı ortalama gibi yaklaşımlarına kıyasla, TM yöntemi doğrudan birleşik özellik uzayını optimize ederek daha kararlı yürüyüş tanıma imkânı sunmuştur. Ayrıca TM, birden fazla modelin birbirinden farklı ancak anlamlı bilgiler sunduğu senaryolarda, daha kapsamlı bir özellik temsili elde edilmesini sağlamış ve gözlemlenen performans artışına katkıda bulunmuştur. Öte yandan, görünüm temelli değişimlerin etkisini azaltmak amacıyla tez kapsamında bir parça eleme işlemi önerilmiştir. Bu işlemde, varyasyon içeren ve içermeyen görüntüler arasındaki fark görüntüsü hesaplanarak, değişimden en çok etkilenen yerel vücut parçaları belirlenmiş ve her parçaya özgü önceden tanımlanmış bir eşik değeri kullanılarak, hem eğitim hem de test aşamalarından bu parçalar çıkarılmıştır. Bu işlem ile eğitimin daha istikrarlı gerçekleşmesi ve modelin tanıma sağlamlığı ile güvenilirliğini artırmak hedeflenmiştir. Tanıma işlemi, nihai olarak elde edilen, parça tabanlı en ayırt edici öznitelikler üzerinden gerçekleştirilmiştir. Laboratuvar ve dış ortam veri setleri üzerinde yapılan deneysel çalışmalar, topluluk öğrenmesi ve parça eleme işleminin birlikte kullanımının, parça tabanlı yürüyüş tanıma performansını önemli ölçüde artırdığını ispatlamıştır. Ayrıca, bu çalışma ile önerilen yöntemin basitliği ve uygulanabilirliğinin kolaylığı açısından, literatürdeki diğer önde gelen parça tabanlı yaklaşımlara, topluluk öğrenme kullanılarak performans artışı sağlayabilecekleri yeni bir yol sunduğu düşünülmektedir.

## 5. ÖNERİLER

Doktora tezi kapsamında yürütülen bu çalışmada, parça-tabanlı yürüyüş tanıma modellerinin tanıma performansını artırmaya yönelik basit ve etkili bir topluluk öğrenmeli yaklaşım sunulmuş, topluluk öğrenme modellerinin çıktılarını daha kararlı ve optimize şekilde bir araya getirebilecek ikinci bir model eğitilmiştir. Konuya katkı sağlamak isteyen araştırmacı ve uygulayıcılara yönelik öneriler aşağıda sunulmuştur:

- Hem küresel bilgiyi hem de yerel bilgiler arasındaki ilişkileri modelleyebilen modüllerin önerilen yönteme entegre edilmesi yöntemin genellenebilirliğini iyileştirecektir.
- Literatürde yürüyüş tanımada önde gelen GaitSet, GaitPart, GaitGL, GaitBase gibi parça tabanlı yöntemlerin süreçleri arasına önerilen yöntemin topluluk öğrenme aşamasının entegresi genel performansı artırabilir.
- Özellikle, varyasyonlar için önerilen parça eleme işlemi, elde olan bilgi kaybını en aza indirecek şekilde güncellenebilirse, çıkarılacak öznelik bilgisi artırılarak daha performanslı bir tanıma işlemi gerçekleştirilebilir.

## 6. KAYNAKLAR

- Alotaibi, M., & Mahmood, A. (2017). Improved gait recognition based on specialized deep convolutional neural network. *Computer Vision and Image Understanding*, 164, 103-110.
- Alotaibi, M., & Mahmood, A. (2017). Reducing covariate factors of gait recognition using feature selection and dictionary-based sparse coding. *Signal, Image and Video Processing*, 11, 131-1138.
- An, W., Yu, S., Makihara, Y., Wu, X., Xu, C., Yu, Y., & Yagi, Y. (2020). Performance evaluation of model-based gait on multi-view very large population database with pose sequences. *IEEE transactions on biometrics, behavior, and identity science*, 24(4), 421-430.
- Ariyanto, G., & Nixon, M. S. (2012). Marionette mass-spring model for 3d gait biometrics. *In IEEE 5th IAPR International Conference on Biometrics (ICB)*, s. 354-359.
- Bashir, K., Xiang, T., & Gong, S. (2010). Cross view gait recognition using correlation strength. *In Bmvc*, s. 1-11.
- Bashir, K., Xiang, T., & Gong, S. (2010). Gait recognition without subject cooperation. *Pattern Recognition Letters*, 31(13), 2052-2060.
- Ben, X., Gong, C., Zhang, P., Jia, X., Wu, Q., & Meng, W. (2019). Coupled patch alignment for matching cross-view gaits. *IEEE Transactions on Image Processing*, 28(6), 3142-3157.
- Ben, X., Zhang, P., Meng, W., Yan, R., Yang, M., Liu, W., & Zhang, H. (2016). On the distance metric learning between cross-domain gaits. *Neurocomputing*, 208, 153-164.
- Bilal, M., Jianbiao, H., Mushtaq, H., Asim, M., Ali, G., & ElAffendi, M. (2024). GaitSTAR: Spatial–Temporal Attention-Based Feature-Reweighting Architecture for Human Gait Recognition. *Mathematics*, 12(16), 2458.
- Castro, F. M., Marin-Jimenez, M. J., Guil, N., Lopez-Tapia, S., & de la Blanca, N. P. (2017). Evaluation of CNN architectures for gait recognition based on optical flow maps. *In 2017 IEEE international conference of the biometrics special interest group (BIOSIG)*, s. 1-5.
- Chao, H., He, Y., Zhang, J., & Feng, J. (2019). Gaitset: Regarding gait as a set for cross-view gait recognition. *In Proceedings of the AAAI conference on artificial intelligence*, 33(01), s. 8126-8133.
- Chen, C., Liang, J., Zhao, H., Hu, H., & Tian, J. (2009). Frame difference energy image for gait recognition with incomplete silhouettes. *Pattern Recognition Letters*, 30(11), 977-984.

- Chen, G., Chen, X., Zheng, C., Wang, J., Liu, X., & Han, Y. (2024). Spatiotemporal smoothing aggregation enhanced multi-scale residual deep graph convolutional networks for skeleton-based gait recognition. *Applied Intelligence*, 54(8), 6154-6174.
- Chen, J., Wang, Z., Zheng, C., Zeng, K., Zou, Q., & Cui, L. (2023). GaitAMR: Cross-view gait recognition via aggregated multi-feature representation. *Information Sciences*, 636, 118920.
- Chen, X., Luo, X., Weng, J., Luo, W., Li, H., & Tian, Q. (2021). Multi-view gait image generation for cross-view gait recognition. *IEEE Transactions on Image Processing*, 30, 3041-3055.
- Choudhury, S. D., & Tjahjadi, T. (2012). Silhouette-based gait recognition using Procrustes shape analysis and elliptic Fourier descriptors. *Pattern Recognition*, 45(9), 3414-3426.
- Choudhury, S. D., & Tjahjadi, T. (2015). Robust view-invariant multiscale gait recognition. *Pattern Recognition*, 48(3), 798-811.
- Dockstader, S. L., Berg, M. J., & Tekalp, A. M. (2003). Stochastic kinematic modeling and feature extraction for gait analysis. *IEEE Transactions on Image Processing*, 12(8), 962-976.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint, 2010*, 11929.
- Ekinci, M., & Gedikli, E. (2005). A novel approach on silhouette based human motion analysis for gait recognition. In *International Symposium on Visual Computing*, s. 219-226.
- Ekinci, M., Aykut, M., & Gedikli, E. (2007). Gait recognition by applying multiple projections and kernel PCA. In *5th Machine Learning and Data Mining in Pattern Recognition International Conference (MLDM)*, s. 727-741.
- Elharrouss, O., Almaadeed, N., Al-Maadeed, S., & Bouridane, A. (2021). Gait recognition for person re-identification. *The Journal of Supercomputing*, 77, 3653-3672.
- Fan, C., Liang, J., Shen, C., Hou, S., Huang, Y., & Yu, S. (2023). Opengait: Revisiting gait recognition towards better practicality. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, s. 9707-9716.
- Fan, C., Peng, Y., Cao, C., Liu, X., Hou, S., Chi, J., Huang, Y., Li, Q., & He, Z. (2020). Gaitpart: Temporal part-based model for gait recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, s. 14225-14233.
- Farneback, G. (2003). Two-frame motion estimation based on polynomial expansion. In *Proceedings Image Analysis: 13th Scandinavian Conference, SCIA*, s. 363-370.

- Feng, Y., Li, Y., & Luo, J. (2016). Learning effective gait features using LSTM. *23rd international conference on pattern recognition (ICPR)*, s. 325-330.
- Fu, H., Gao, J., & Liu, H. (2023). Human pose estimation and action recognition for fitness movements. *Computers & Graphics, 116*, s. 418-426.
- Gao, S., Tan, Z., Ning, J., Hou, B., & Li, L. (2023). ResGait: gait feature refinement based on residual structure for gait recognition. *The Visual Computer, 39*(8), 3455-3466.
- Ghebleh, A., & Ebrahimi Moghaddam, M. (2018). Clothing-invariant human gait recognition using an adaptive outlier detection method. *Multimedia Tools and Applications, 77*, 8237-8257.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM, 63*(11), 139-144.
- Gross, R., & Shi, J. (2001). The CMU Motion of Body (MOBO) Database. *Robotics Institute*.
- Gul, S., Malik, M. I., Khan, G. M., & Shafait, F. (2021). Multi-view gait recognition system using spatio-temporal features and deep learning. *Expert Systems with Applications, 179*, 115057.
- Gupta, A., & Semwal, V. B. (2020). Multiple task human gait analysis and identification: ensemble learning approach. *Emotion and Information Processing: A Practical Approach*, 85-197.
- Han, F., Li, X., Zhao, J., & Shen, F. (2022). A unified perspective of classification-based loss and distance-based loss for cross-view gait recognition. *Pattern Recognition, 125*, 108519.
- Han, J., & Bhanu, B. (2005). Individual recognition using gait energy image. *IEEE transactions on pattern analysis and machine intelligence, 28*(2), 316-322.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, s.770-778.
- He, Y., Zhang, J., Shan, H., & Wang, L. (2018). Multi-task GANs for view-specific feature learning in gait recognition. *IEEE Transactions on Information Forensics and Security, 14*(1), 102-113.
- Hermans, A., Beyer, L., & Leibe, B. (2017). In defense of the triplet loss for person re-identification. *arXiv preprint, 07737*, 1703.
- Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science, 313*(5786), 504-507.
- Hofmann, M., Geiger, J., Bachmann, S., Schuller, B., & Rigoll, G. (2014). The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects

- and traits. *Journal of Visual Communication and Image Representation*, 25(1), 195-206.
- Hossain, M. A., Makihara, Y., Wang, J., & Yagi, Y. (2010). Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition*, 43(6), 2281-2291.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., . . . Hartwig, A. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint, 1704*, 04861.
- Isaac, E. R., Elias, S., Rajagopalan, S., & Easwarakumar, K. S. (2017). View-invariant gait recognition through genetic template segmentation. *IEEE signal processing letters*, 24(8), 1188-1192.
- Iwama, H., Okumura, M., Makihara, Y., & Yagi, Y. (2012). The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Transactions on Information Forensics and Security*, 7(5), 1511-1521.
- Iwashita, Y., Uchino, K., & Kurazume, R. (2013). Gait-based person identification robust to changes in appearance. *Sensors*, 13(6), 7884-7901.
- Johansson, G. (1975). Visual motion perception. *Scientific American*, 232(6), 76-89.
- Kale, A., Sundaresan, A., Rajagopalan, A. N., Cuntoor, N. P., Roy-Chowdhury, A. K., Kruger, V., & Chellappa, R. (2004). Identification of humans using gait. *IEEE Transactions on image processing*, 13(9), 1163-1173.
- Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. *In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, s. 1725-1732.
- Kazmaier, J., & Van Vuuren, J. H. (2022). The power of ensemble learning in sentiment analysis. *Expert Systems with Applications*, 187, 115819.
- Khan, M. A., Arshad, H., Khan, W. Z., Alhaisoni, M., Tariq, U., Hussein, H. S., & Elashry, A. (2023). HGRBOL2: human gait recognition for biometric application using Bayesian optimization and extreme learning machine. *Future Generation Computer Systems*, 143, 337-348.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- Kusakunniran, W., Wu, Q., Li, H., & Zhang, J. (2009). Multiple views gait recognition using view transformation model based on optimized gait energy image. *In 12th IEEE international conference on computer vision workshops, ICCV workshops*, s. 1058-1064.

- Lam, T. H., & Lee, R. S. (2006). A new representation for human gait recognition: Motion silhouettes image (MSI). *In International conference on biometrics*, s. 612-618.
- Lam, T. H., Cheung, K. H., & Liu, J. N. (2011). Gait flow image: A silhouette-based gait representation for human identification. *Pattern recognition*, 44(4), 973-987.
- Lam, T. H., Lee, R. S., & Zhang, D. (2007). Human gait recognition by the fusion of motion and static spatio-temporal templates. *Pattern Recognition*, 40(9), 2563-2573.
- Lee, L., & Grimson, W. E. (2002). Gait analysis for recognition and classification. *In Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, s. 155-162.
- Lee, S., Liu, Y., & Collins, R. (2007). Shape variation-based frieze pattern for robust gait recognition. *In 2007 IEEE Conference on Computer Vision and Pattern Recognition*, s. 1-8.
- Li, H., Qiu, Y., Zhao, H., Zhan, J., Chen, R., Wei, T., & Huang, Z. (2022). GaitSlice: A gait recognition model based on spatio-temporal slice features. *Pattern Recognition*, 124, 108453.
- Li, N., & Zhao, X. (2022). A strong and robust skeleton-based gait recognition method with gait periodicity priors. *IEEE Transactions on Multimedia*, 25, 3046-3058.
- Liang, D., Fan, G., Lin, G., Chen, W., Pan, X., & Zhu, H. (2019). Three-stream convolutional neural network with multi-task and ensemble learning for 3d action recognition. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, s. 0-0.
- Liao, R., Yu, S., An, W., & Huang, Y. (2020). A model-based gait recognition method with body pose and human prior knowledge. *Pattern Recognition*, 98, 107069.
- Lin, B., Zhang, S., & Yu, X. (2021). Gait recognition via effective global-local feature representation and local temporal aggregation. *In Proceedings of the IEEE/CVF international conference on computer vision*, s. 14648-14656.
- Lishani, A. O., Boubchir, L., & Bouridane, A. (2014). Haralick features for GEI-based human gait recognition. *In 26th IEEE International Conference on Microelectronics (ICM)*, s. 36-39.
- Lishani, A. O., Boubchir, L., Khalifa, E., & Bouridane, A. (2016). Gabor filter bank-based GEI features for human Gait recognition. *In 39th IEEE International Conference on Telecommunications and Signal Processing (TSP)*, s. 648-651.
- Lishani, A. O., Boubchir, L., Khalifa, E., & Bouridane, A. (2017). Human gait recognition based on Haralick features. *Signal, Image and Video Processing*, 11, 1123-1130.
- Liu, L., Wang, X., Bao, Q., & Li, X. (2024). Behavior detection and evaluation based on multi-frame MobileNet. *Multimedia Tools and Applications*, 83(6), 15733-15750.

- Liu, N., Lu, J., & Tan, Y. P. (2011). Joint subspace learning for view-invariant gait recognition. *IEEE Signal processing letters*, 18(7), 431-434.
- Liu, N., Lu, J., Yang, G., & Tan, Y. P. (2013). Robust gait recognition via discriminative set matching. *Journal of Visual Communication and Image Representation*, 24(4), 439-447.
- Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., & Song, L. (2017). Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, s. 212-220.
- Liu, Z., & Sarkar, S. (2004). Simplest representation yet for gait recognition: Averaged silhouette. In *IEEE Proceedings of the 17th International Conference on Pattern Recognition*, 4, s. 211-214.
- Liu, Z., & Sarkar, S. (2006). Improved gait recognition by gait dynamics normalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(6), 863-876.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., . . . Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, s. 10012-10022.
- Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, s. 11976-11986.
- Liu, Z., Peng, J., Guo, X., Chen, S., & Liu, L. (2024). Breast cancer classification method based on improved VGG16 using mammography images. *Journal of Radiation Research and Applied Sciences*, 172(2), 100885.
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, s. 3431-3440.
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, s. 3431-3440.
- Lu, H., Plataniotis, K. N., & Venetsanopoulos, A. N. (2008). MPCA: Multilinear principal component analysis of tensor objects. *IEEE transactions on Neural Networks*, 19(1), 18-39.
- Lu, J., & Zhang, E. (2007). Gait recognition for human identification based on ICA and fuzzy SVM through multiple views fusion. *Pattern Recognition Letters*, 28(16), 2401-2411.
- Lu, J., & Zhang, E. (2007). Gait recognition for human identification based on ICA and fuzzy SVM through multiple views fusion. *Pattern Recognition Letters*, 28(16), 2401-2411.

- Ma, K., Fu, Y., Zheng, D., Cao, C., Hu, X., & Huang, Y. (2023). Dynamic aggregated network for gait recognition. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, s. 22076-22085.
- Makihara, Y., Mannami, H., & Yagi, Y. (2010). Gait analysis of gender and age using a large-scale multi-view gait database. *In Asian Conference on Computer Vision*, s. 440-451.
- Makihara, Y., Mansur, A., Muramatsu, D., Uddin, Z., & Yagi, Y. (2015 ). Multi-view discriminant analysis with tensor representation and its application to cross-view gait recognition. *In 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, s. 1-8.
- Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T., & Yagi, Y. (2006). Gait recognition using a view transformation model in the frequency domain. *In Computer Vision–ECCV 2006: 9th European Conference on Computer Vision*, s. 151-163.
- Mansur, A., Makihara, Y., Muramatsu, D., & Yagi, Y. (2014). Cross-view gait recognition using view-dependent discriminative analysis. *In IEEE International joint conference on biometrics*, s. 1-8.
- Marín-Jiménez, M. J., Castro, F. M., Delgado-Escañó, R., Kalogeiton, V., & Guil, N. (2021). UGaitNet: Multimodal gait recognition with missing input modalities. *IEEE Transactions on Information Forensics and Security*, 16, 5452-5462.
- Meng, C., He, X., Tan, Z., & Luan, L. (2023). Gait recognition based on 3D human body reconstruction and multi-granular feature fusion. *The Journal of Supercomputing*, 79(11), 12106-12125.
- Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University-Computer and Information Sciences*, 35(2), 757-774.
- Muramatsu, D., Makihara, Y., & Yagi, Y. (2015). View transformation model incorporating quality measures for cross-view gait recognition. *IEEE transactions on cybernetics*, 46(7), 1602-1615.
- Muramatsu, D., Shiraishi, A., Makihara, Y., & Yagi, Y. (2012 ). Arbitrary view transformation model for gait person authentication. *In IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, s. 85-90.
- Murray, M. P., Drought, A. B., & Kory, R. C. (1964). Walking patterns of normal men. *JBJS*, 46(2), 335-360.
- Nandy, A., Chakraborty, R., & Chakraborty, P. (2016). Cloth invariant gait recognition using pooled segmented statistical features. *Neurocomputing*, 191, 117-140.
- Nixon, M. S., & Carter, J. N. (2006). Automatic recognition by gait. *Proceedings of the IEEE*, 94(11), 2013-2024.

- Niyogi, S. A., & Adelson, E. H. (1994). Analyzing and recognizing walking figures in XYT. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, s. 469-474.
- Niyogi, S. A., & Adelson, E. H. (1994). Analyzing gait with spatiotemporal surfaces. *In Proceedings of IEEE Workshop on Motion of Non-rigid and Articulated Objects*, s. 64-69.
- Pan, X., Xie, H., Zhang, N., & Li, S. (2024). GaitLRDF: gait recognition via local relevant feature representation and discriminative feature learning. *Applied Intelligence*, 23, 12476-12491.
- Rida, I., Almaadeed, N., & Almaadeed, S. (2019). Robust gait recognition: a comprehensive survey. *IET Biometrics*, 8(1), 14-28.
- Rida, I., Almaadeed, S., & Bouridane, A. (2016). Gait recognition based on modified phase-only correlation. *Signal, Image and Video Processing*, 10, 463-470.
- Rida, I., Jiang, X., & Marcialis, G. L. (2015). Human body part selection by group lasso of motion for model-free gait recognition. *IEEE Signal Processing Letters*, 23(1), 154-158.
- Rokanujjaman, M., Hossain, M. A., & Islam, M. R. (2012). Effective part selection for part-based gait identification. *In 2012 7th IEEE International Conference on Electrical and Computer Engineering*.
- Rokanujjaman, M., Islam, M. S., Hossain, M. A., Islam, M. R., Makihara, Y., & Yagi, Y. (2015). 'Effective part-based gait identification using frequency-domain gait entropy. *Mult Tools and Appl*, 74(9), 3099–3120.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., & L., F.-F. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115, 211-252.
- Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 8(4), 1249.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, s. 4510-4520.
- Sarkar, S., Phillips, P. J., Liu, Z., Vega, I. R., Grother, P., & Bowyer, K. W. (2005). The humanoid gait challenge problem: Data sets, performance, and analysis. *IEEE transactions on pattern analysis and machine intelligence*, 27(2), 162-177.
- Semwal, V. B., Gupta, A., & Lalwani, P. (2021). An optimized hybrid deep learning model using ensemble learning approach for human walking activities recognition. *The Journal of Supercomputing*, 77(11), 12256-12279.

- Sepas-Moghaddam, A., & Etemad, A. (2020). View-invariant gait recognition with attentive recurrent learning of partial representations. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(1), 124-137.
- Sepas-Moghaddam, A., Ghorbani, S., Troje, N. F., & Etemad, A. (2021). Gait recognition using multi-scale partial representation transformation with capsules. *In IEEE 2020 25th international conference on pattern recognition (ICPR)*, s. 8045-8052.
- Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T., & Yagi, Y. (2016). Geinet: View-invariant gait recognition using a convolutional neural network. *In 2016 international conference on biometrics (ICB)*, s. 1-8.
- Shutler, J. D., Grant, M. G., Nixon, M. S., & Carter, J. N. (2004). On a large sequence-based human gait database. *In Applications and science in soft computing*, s. 339-346.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint, 1409*, 1556.
- Song, C., Huang, Y., Huang, Y., Jia, N., & Wang, L. (2019). Gaitnet: An end-to-end network for gait based human identification. *Pattern recognition*, 96, s. 106988.
- Song, X., Huang, Y., Shan, C., Wang, J., & Chen, Y. (2022). Distilled light gaitset: Towards scalable gait recognition. *Pattern Recognition Letters*, 157, 27-34.
- Sun, J., Wang, Y., Li, J., Wan, W., Cheng, D., & Zhang, H. (2018). View-invariant gait recognition based on kinect skeleton feature. *Multimedia Tools and Applications*, 77, 24909-24935.
- Sundaresan, A., RoyChowdhury, A., & Chellappa, R. (2003). A hidden markov model based framework for recognition of humans from gait sequences. *In IEEE Proceedings 2003 international conference on image processing*, s. II-93.
- Tafazzoli, F., & Safabakhsh, R. (2010). Model-based human gait recognition using leg and arm movements. *Engineering applications of artificial intelligence*, 23(8), 1237-1246.
- Takemura, N., Makihara, Y., Muramatsu, D., Echigo, T., & Yagi, Y. (2017). On input/output architectures for convolutional neural network-based cross-view gait recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(9), 2708-2719.
- Takemura, N., Makihara, Y., Muramatsu, D., Echigo, T., & Yagi, Y. (2018). Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ transactions on Computer Vision and Applications*, 10, 1-14.
- Tan, D., Huang, K., Yu, S., & Tan, T. (2006). Efficient night gait recognition based on template matching. *18th international conference on pattern recognition (ICPR'06)*, s. 1000-1003.
- Tan, M., & Le, Q. (2019, May 19). *EfficientNet: Improving Accuracy and Efficiency through AutoML and Model Scaling*. Google AI:

<https://research.google/blog/efficientnet-improving-accuracy-and-efficiency-through-automl-and-model-scaling/> adresinden alındı

- Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *In International conference on machine learning*, s. 6105-6114.
- Tao, D., Li, X., Wu, X., & Maybank, S. J. (2007). General tensor discriminant analysis and gabor features for gait recognition. *IEEE transactions on pattern analysis and machine intelligence*, 29(10), 1700-1715.
- Teepe, T., Gilg, J., Herzog, F., Hörmann, S., & Rigoll, G. (2022). Towards a deeper understanding of skeleton-based gait recognition. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, s. 1569-1577.
- Teepe, T., Khan, A., Gilg, J., Herzog, F., Hörmann, S., & Rigoll, G. (2021). Gaitgraph: Graph convolutional network for skeleton-based gait recognition. *IEEE international conference on image processing (ICIP)*, s. 2314-2318.
- Thapar, D., Nigam, A., Aggarwal, D., & Agarwal, P. (2018). VGR-net: A view invariant gait recognition network. *In 2018 IEEE 4th international conference on identity, security, and behavior analysis (ISBA)*, s. 1-8.
- Tsuji, A., Makihara, Y., & Yagi, Y. (2010). Silhouette transformation based on walking speed for gait identification. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 717-722.
- Uddin, M. Z., Ngo, T. T., Makihara, Y., Takemura, N., Li, X., Muramatsu, D., & Yagi, Y. (2018). The ou-isir large population gait database with real-life carried object and its performance evaluation. *IPSJ Transactions on Computer Vision and Applications*, 10, 1-11.
- Wang, C., Zhang, J., Pu, J., Yuan, X., & Wang, L. (2010). Chrono-gait image: A novel temporal template for gait recognition. *Computer Vision—ECCV 2010: 11th European Conference on Computer Vision*, s. 257-270.
- Wang, L., Ning, H., Tan, T., & Hu, W. (2004). Fusion of static and dynamic body biometrics for gait recognition. *IEEE Transactions on circuits and systems for video technology*, 14(2), 149-158.
- Wang, L., Tan, T., Ning, H., & Hu, W. (2003). Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence TPAMI*, 25(12), 1505-1518.
- Wang, L., Tan, T., Ning, H., & Hu, W. (2003). Silhouette analysis-based gait recognition for human identification. *IEEE transactions on pattern analysis and machine intelligence*, 25(12), 1505-1518.
- Wang, X., & Yan, W. Q. (2020). Cross-view gait recognition through ensemble learning. *Neural computing and applications*, 32, 7275-7287.

- Wang, X., & Yan, W. Q. (2020). Human gait recognition based on frame-by-frame gait energy images and convolutional long short-term memory. *International journal of neural systems*, 30(01), 1950027.
- Wang, X., & Yan, W. Q. (2021). Non-local gait feature extraction and human identification. *Multimedia Tools and Applications*, 80, 6065-6078.
- Wang, X., Wang, J., & Yan, K. (2018). Gait recognition based on Gabor wavelets and (2D) 2 PCA. *Multimedia Tools and Applications*, 77, 12545-12561.
- Wang, X., Zhang, J., & Yan, W. Q. (2020). Gait recognition using multichannel convolution neural networks. *Neural computing and applications*, 32(18), 14275-14285.
- Wei, S., Liu, W., Wei, F., Wang, C., & Xiong, N. N. (2024). Gaitdlf: global and local fusion for skeleton-based gait recognition in the wild. *The Journal of Supercomputing*, 80(12), 17606-17632.
- Wei, T., Liu, M., Zhao, H., & Li, H. (2024). Gmsn: an efficient multi-scale feature extraction network for gait recognition. *Expert Systems with Applications*, 252, 124250.
- Wen, Y., Zhang, K., Li, Z., & Qiao, Y. (2016). A discriminative feature learning approach for deep face recognition. In *Computer vision—ECCV 14th European conference*, s. 499-515.
- Wolf, T., Babae, M., & Rigoll, G. (2016). Multi-view gait recognition using 3D convolutional neural networks. In *IEEE international conference on image processing (ICIP)*, s. 4165-4169.
- Wu, Z., Huang, Y., Wang, L., Wang, X., & Tan, T. (2016). A comprehensive study on cross-view gait based human identification with deep cnns. *IEEE transactions on pattern analysis and machine intelligence*, 39(2), 209-226.
- Xiong, J., Zou, S., Tang, J., & Tjahjadi, T. (2024). MCDGait: multimodal co-learning distillation network with spatial-temporal graph reasoning for gait recognition in the wild. *The Visual Computer*, 40(10), 7221-7234.
- Xiuhui, W., & Yan, K. (2021). Gait classification through CNN-based ensemble learning. *Multimedia Tools and Applications*, 80(1), 1565-1581.
- Xu, C., Makihara, Y., Ogi, G., Li, X., Yagi, Y., & Lu, J. (2017). The OU-ISIR gait database comprising the large population dataset with age and performance evaluation of age estimation. *IPSJ Transactions on Computer Vision and Applications*, 9, 1-14.
- Xu, K., Jiang, X., & Sun, T. (2021). Gait recognition based on local graphical skeleton descriptor with pairwise similarity network. *IEEE Transactions on Multimedia*, 24, 3265-3275.
- Xu, Y., Cheng, J., Wang, L., Xia, H., Liu, F., & Tao, D. (2018). Ensemble one-dimensional convolution neural networks for skeleton-based action recognition. *IEEE Signal Processing Letters*, 25(7), 1044-1048.

- Ye, H., Sun, T., & Xu, K. (2023). Gait recognition based on gait optical flow network with inherent feature pyramid. *Applied Sciences*, 13(19), 10975.
- Ye, M., Lan, X., Leng, Q., & Shen, J. (2020). Cross-modality person re-identification via modality-aware collaborative ensemble learning. *IEEE Transactions on Image Processing*, 29, 9387-9399.
- Yu, S., Chen, H., Garcia Reyes, E. B., & Poh, N. (2017). Gaitgan: Invariant gait feature extraction using generative adversarial networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, s. 30-37.
- Yu, S., Chen, H., Wang, Q., Shen, L., & Huang, Y. (2017). Invariant feature extraction for gait recognition using only one uniform model. *Neurocomputing*, 239, 81-93.
- Yu, S., Tan, D., & Tan, T. (2006). A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. *18th international conference on pattern recognition (ICPR'06)*, s. 441-444.
- Yu, S., Wang, Q., & Huang, Y. (2013). A large RGB-D gait dataset and the baseline algorithm. *In Biometric Recognition: 8th Chinese Conference*, s. 417-424.
- Zhang, C., Liu, W., Ma, H., & Fu, H. (2016). Siamese neural network based gait recognition for human identification. *In IEEE international conference on acoustics, speech and signal processing (ICASSP)*, s. 2832-2836.
- Zhang, R., Vogler, C., & Metaxas, D. (2007). Human gait recognition at sagittal plane. *Image and vision computing*. 25(3), 321-330.
- Zhang, Y., Huang, Y., Wang, L., & Yu, S. (2019). A comprehensive study on gait biometrics using a joint CNN-based method. *Pattern Recognition*, 93, 228-236.
- Zhang, Y., Huang, Y., Yu, S., & Wang, L. (2019). Cross-view gait recognition by discriminative feature learning. *IEEE Transactions on Image Processing*, 29, 1001-1015.
- Zhang, Z., Tran, L., Yin, X., Atoum, Y., Liu, X., Wan, J., & Wang, N. (2019). Gait recognition via disentangled representation learning. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, s. 4710-4719.
- Zhao, L., Guo, L., Zhang, R., Xie, X., & Ye, X. (2022). mmGaitSet: multimodal based gait recognition for countering carrying and clothing changes. *Applied Intelligence*, 52(2), 2023-2036.
- Zheng, J., Liu, X., Liu, W., He, L., Yan, C., & Mei, T. (2022). Gait recognition in the wild with dense 3d representations and a benchmark. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, s. 20228-20237.
- Zheng, S., Zhang, J., Huang, K., He, R., & Tan, T. (2011). Robust view transformation model for gait recognition. *In 8th IEEE international conference on image processing*, s. 2073-2076.

Zhu, Z., Guo, X., Yang, T., Huang, J., Deng, J., Huang, G., & Zhou, J. (2021). Gait recognition in the wild: A benchmark. *In Proceedings of the IEEE/CVF international conference on computer vision*, s. 14789-14799.



## ÖZGEÇMİŞ

Büşranur YAPRAK, İlköğrenimini Şükrüpaşa İlköğretim Okulu'nda bitirdikten sonra lise eğitimini 2010 yılında Erzurum Mehmet Akif Ersoy Lisesi'nde tamamladı. 2010-2015 yılları arasında Kayseri Erciyes Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği bölümünde lisans eğitimini tamamladı. 2017 yılında Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim dalında tezli yüksek lisans programına başladı. 2019 yılında "Meta-Sezgisel Yaklaşımlara Dayalı İmge Bölütleme" başlıklı yüksek lisans tezini sunmasının ardından doktora eğitimine başladı. 2016 yılından itibaren Gümüşhane Üniversitesi Yazılım Mühendisliği Bölümü'nde araştırma görevlisi olarak görev yapmaktadır. İyi derece İngilizce bilmektedir.

- Yaprak, B., & Gedikli, E. (2025). Enhancing part-based gait recognition via ensemble learning and feature fusion. *Pattern Analysis and Applications*, 28, 98.
- Yaprak, B., & Gedikli, E. (2024). Different gait combinations based on multi-modal deep CNN architectures. *Multimedia Tools and Applications*, 83, 83403–83425.
- Yaprak, B., & Gedikli, E. (2024). In Different Scenarios MobileNet-V1 for Cross-view Gait Recognition. *In International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, s. 1-5. Craiova, Romania.
- Yaprak, B., Gedikli, E., Bingöl, Ö., & Doğan, R. Ö. (2024). Cross-view Gait Recognition Based on Fine-Tuned Deep Networks. *32nd Signal Processing and Communications Applications Conference (SIU)*, s. 1-4. Mersin, Türkiye.