Bilgisayar Bilimleri ve Mühendisliği Dergisi

Cilt-15
Sayı-1
2022

Silüet ve RGB Görüntüleri ile Derin Evrişimsel Sinir Ağları Kullanarak Yürüyüşten Kimlik Tanıma

Gait Recognition from Silhouette and RGB Images Using Deep Convolutional Neural Networks

Selin GÖK IŞIK
ASELSAN A.Ş.,
Mikroelektronik, Güdüm ve Elektro-Optik Sektör Başkanlığı,
Ankara, Türkiye
selingok@aselsan.com.tr
ORCID: 0000-0001-7914-4530

Hazım Kemal EKENEL
İstanbul Teknik Üniversitesi,
Bilgisayar Mühendisliği Bölümü,
İstanbul, Türkiye
ekenel@itu.edu.tr
ORCID: 0000-0003-3697-8548

Öz

Günümüzde kişi tanıma için kullanılan birçok biyometrik özellik vardır. Göz, iris, kulak, parmak izi, DNA gibi fiziksel biyometrik özelliklerden farklı olarak davranışsal biyometrik özelliklerimiz zamanla öğrenilir ve gelişirler. Yürüyüş, yakın mesafeden görüntü edinimi gerektiren fiziksel biyometrilerin aksine, uzak mesafeden kaydedilmiş görüntüler üzerinden kişiyi tanımayı sağlamaktadır. Bu makalede, yürüyüşten kişi tanıma problemi için derin öğrenme yaklaşımı kullanan görünüm tabanlı bir yöntem önerilmiştir. Yöntemin her adımı farklı parametrelerle analiz edilmiştir. Çalışmada, yürüyüş tanıma probleminde yaygın kullanılan girdiler olan ikili insan silüetinin ve yürüyüş enerji imgesinin kişi tanıma başarımına etkileri incelenmiştir. Ayrıca yöntemi pratik uygulamalarda kullanıma daha uygun hale getirebilmek için insan silüeti çıkarma, yürüyüş döngüsü hesaplama gibi ön işleme adımları kaldırılmış ve doğrudan RGB çerçeveleri girdi olarak kullanılmıştır. Ek olarak transfer öğrenmenin başarıma katkısı gözlemlenmiş, bu amaçla popüler bir nesne tanıma modeli CASIA-B yürüyüş veri kümesi ile ince ayarlanmıştır. Yürüyüş dizisini temsil edecek öznitelik vektörünün elde edilmesi aşamasında çerçevelerden çıkarılan öznitelik vektörleri arasında farklı birleştirme yöntemleri denenmiş ve başarıları karşılaştırılmıştır. Önerilen yaklaşımın başarımı hem bu alanda sıkça kullanılan CASIA-B ve OU-ISIR Büyük Popülasyon yürüyüş veri kümelerinde hem de gerçek hayattan toplanmış yürüyüş verileri içeren PRID-2011 kişiyi yeniden tanıma veri kümesi üzerinde deneyler yapılarak ölçülmüştür. Açı farklılıklarının etkisini gözlemlemek için deneyler özdeş ve çapraz görünüm koşulları için tekrarlanmıştır. Derin öğrenme yaklaşımı kullanılarak elde ettiğimiz sonuçlar geleneksel yöntemlerin

sonuçlarına göre daha başarılı bulunmuştur.

Anahtar Sözcükler: Derin Öğrenme, Yürüyüş Tanıma, Transfer Öğrenme, Çapraz Görünüm, Biyometri

Abstract

There have been many different biometric features used for human identification. Unlike physical biometric features such as eyes, iris, ears, fingerprints, and DNA, our behavioral biometric features are learned and developed over time. Gait trait, unlike physical biometrics, allows identification from distance. In this paper, we present an appearance-based gait recognition approach which utilizes deep learning. Experiments were carried out with binary human silhouettes and gait energy images, which are widely used for gait recognition. In addition, to make the method more suitable for use in practical applications, pre-processing steps such as human silhouette extraction and gait cycle calculation were removed and the RGB frames are used directly as the input. A popular object recognition model was fine-tuned on the CASIA-B gait dataset to observe the contribution of transfer learning to the performance. To obtain the feature vector that represents the gait sequence, different fusion methods were applied on the feature vectors which are extracted from individual frames. The performance of the proposed approach was assessed by conducting experiments on both the widely used CASIA-B and OU-ISIR Large Population gait datasets, as well as the PRID-2011 person re-identification dataset, which included gait data collected from real life. Experiments were repeated for both identical view and cross view conditions to

Gönderme, düzeltme ve kabul tarihi: 19.02.2022 - 20.04.2022 - 21.04.2022
Makale türü: Araştırma

observe the effect of angle differences. The achieved results using the deep learning approach were found to be more successful than the ones obtained by the traditional methods.

Keywords*:* Deep Learning, Gait Recognition, Transfer Learning, Cross-view, Biometric

1. Giriş

Günümüzde birçok farklı biyometrik özellik kişi tanıma probleminde kullanılmaktadır. Biyometrik özellikler iki temel gruba ayrılırlar; fiziksel özellikler ve davranışsal özellikler. Bireylerin doğdukları andan itibaren sahip oldukları temel özellikler fiziksel özelliklerdir. Göz, kulak, iris, retina, parmak izi, DNA bu özelliklere verilebilecek örneklerdendir. Davranışsal özellikler, zamanla gelişebilen ve dış faktörlerden etkilenebilen özellikler olarak tanımlanmaktadır. Konuşma, yürüyüş, imza davranışsal özelliklerin birkaç örneğidir.

Fiziksel biyometriler üzerimizde taşıdığımız özellikler olduğundan kaybolmaları, unutulmaları veya bir başkasına aktarılmaları mümkün değildir. Kopyalanma ya da çalınma riski vardır ancak bunların gerçekleşmesi çok kolay değildir. Fiziksel biyometrik özellikleri kullanarak tanımlama işlemini gerçekleştirmek için çoğunlukla biyometrik tarayıcılara ihtiyaç duyulmaktadır. Ek olarak bu bilgilerin toplanabilmesi için kişinin iş birliğine ihtiyaç duyulmaktadır. Davranışsal biyometrileri kopyalamak, taklit etmek çok daha zordur hatta bazı davranışsal biyometriler için pek mümkün değildir. Örneğin her bireyin bilgisayarda yazı yazarken klavyedeki tuşlara basma hızı, basma kuvveti ya da klavye üzerindeki el hareketi kişiye has bir strateji içermektedir. Bunun taklit edilmesi oldukça güçtür. Aynı şekilde çalışmamızda kullandığımız yürüyüş davranışsal biyometrisi de hız, salınım, adım uzunluğu, kol salınımı gibi birçok parametre için kişiden kişiye farklılık gösterir. Bu parametrelerin hepsini aynı anda birinin yürüyüşüne benzetmek ve kopyalamak kolay değildir. Ayrıca fiziksel biyometrik özelliklerin aksine kişilerin iş birliğine ihtiyaç duyulmadan bilgi toplanabilmektedir. Kişi bir cihaza parmağını okutmak ya da bir tarayıcıya yüzünü göstermek zorunda kalmaksızın gündelik yaşamına devam ederken bu bilgilere erişilebilmektedir. Öte yandan bireyin davranışları hastalık, yorgunluk, farklı ruh hali gibi durumlarda değişiklik gösterebileceğinden dolayı kişinin standart bir davranışını belirleyebilmek için kişiye ait çok sayıda veriye ihtiyaç duyulmaktadır.

Davranışsal biyometri kişi tanımada henüz fiziksel biyometriler kadar yaygın kullanılmasa da kişiye özel olması ve kopyalanmasının daha zor olması nedeniyle kimlik doğrulama problemlerinde tercih edilebilir. Aynı zamanda kritik bilgilere erişim gerektiğinde fizyolojik biyometrik özelliklerin yanında ek bir doğrulama sistemi olarak davranışsal biyometrinin de kullanılması sistemi daha güvenli hale getirecektir.

Bu makalede önceki çalışmamızı [1] genişleterek, yürüyüşten kimlik tanıma probleminin önerdiğimiz yaklaşımla çözümünde farklı parametrelerin kullanımının sonuçlara etkisi deneysel olarak kapsamlı bir şekilde araştırılmıştır. Çalışmamızın katkıları kısaca aşağıdaki gibi özetlenebilir:

- Önceden ImageNet [2] veri kümesi üzerinde eğitilmiş nesne tanıma modeli, CASIA-B [3] yürüyüş veri kümesi ile ince ayarlanmıştır. OU-ISIR Büyük Popülasyon (BP) [4] veri kümesi üzerinde yapılan deneylerle de transfer öğrenmenin katkısı gözlemlenmiştir.
- İkili insan silüeti, yürüyüş enerji imgesi ve RGB görüntüler olmak üzere üç farklı tipte insan vücudu görselleri üzerinden öznitelikler çıkarılarak farklı girdi tiplerinin sonuçlara etkisi araştırılmıştır.
- Önerilen yaklaşımda elde edilen çok sayıda öznitelik vektörlerinin birleştirilmesi ve yürüyüşü temsil edecek tek bir öznitelik vektörü elde edilmesi aşamasında farklı örnekleme yöntemleri denenmiş ve sonuçları karşılaştırılmıştır.
- Önerilen yaklaşım hem kontrollü ortamda kaydedilmiş yürüyüş verileri üzerinde hem de gündelik hayattan toplanmış yürüyüş verileri üzerinde test edilmiş ve sonuçlar tartışılmıştır.
- Deneyler hem özdeş görünüm hem çapraz görünüm durumları için tekrarlanmıştır.
- Yürüyüş tanıma problemi için önerilen derin öğrenme yaklaşımları ile gerçekleştirilen deneylerin sonuçları ve geleneksel yöntemler ile gerçekleştirilen deneylerin sonuçları karşılaştırılmıştır.

Çalışmamızda literatürde yaygın kullanılan derin evrişimsel sinir ağı olan VGG16 [5] mimarisi kullanılmıştır. Ağ, CASIA-B [3] yürüyüş veri kümesi ile ince ayarlanmıştır. Sonraki adımda ikili insan silüeti, yürüyüş enerji imgesi ve RGB görseller gibi farklı girdi tiplerinden öznitelikler çıkarılmıştır. Yürüyüşü temsil edecek öznitelik vektörleri eleman tabanlı maksimum örnekleme, eleman tabanlı ortalama ve eleman tabanlı medyan yöntemleri olmak üzere üç farklı yöntemle birleştirilmiştir. Sınıflandırma ise galeri ve sorgu dizilerinin temsil vektörleri arasındaki maksimum benzerlik sonucuna göre yapılmıştır.

Bu makalenin ilk bölümünde yürüyüşten kimlik tanıma probleminin tanımına ve literatürde bu konu üzerindeki çalışmalara yer verilmiştir. İkinci bölümünde çalışmada önerilen yönteme bir genel bakış sağlanmış ve alt adımlar detaylı bir şekilde açıklanmıştır. Üçüncü bölümde, deneylerde kullanılan veri kümeleri hakkında bilgiler, gerçekleştirilen deneyler ve elde edilen sonuçlar sunulmuştur. Son bölüm ise sonuç ve olası gelecek çalışmaları içermektedir.

2. Literatür Araştırması

Bu bölüm, yürüyüşten kişi tanıma problemi için geçmiş çalışmalarda araştırmacıların kullandıkları temel yaklaşımların özetini içermektedir. Çalışmalar genel hatlarıyla iki gruba ayrılabilir: Model tabanlı yaklaşımlar ve görünüm tabanlı yaklaşımlar.

Yürüyüş sınıflandırma probleminde, bir yürüyüş dizisinden yürüyüşü temsil edecek öznitelikler çıkarılır. Bu öznitelikler model tabanlı yaklaşımlarda insan vücudu modellerinden çıkarılmaktadır. İnsan vücudunun modellenmesinde farklı yöntemler kullanılabilir. Çalışma [6]'da yazarlar insanı anatomik oranları ve yörünge tabanlı kinematik özellikleri

kullanıp iki boyutlu olarak modellemişlerdir. Çalışma [7]'de, kemikler birbirine bağlı sarkaçlar olarak modellenmiş ve bunların açısal hareketlerini çıkarmak için Fourier serisi kullanılmıştır. Çalışma [8]'de ilk olarak insan silüetlerinden dokuz vücut noktası çıkarılmış ve bunlar birleştirilerek iki boyutlu (2B) çubuk görüntüler elde edilmiştir. Daha sonra 2B çubuk görüntüler kullanılarak öznitelikler çıkarılmıştır. Çalışma [9]'da insan uzuvlarının yönelimindeki değişiklikler modellenmiştir. Vücut duruşu, vücuttaki bölümlerin eğimi ve diğer bölümlerle arasındaki mesafe ölçümü ile modellenmiştir. Mesafe ölçümleri göğüs kafesinin yüksekliğiyle normalize edilmiştir. Bu tip model tabanlı yaklaşımların en güçlü yanı açı farklılıklarından, görüntülerde kişilerin yan yana gelmesinden kaynaklı görüntünün bir kısmındaki kapanmalardan, ölçek etkilerinden çok fazla etkilenmemeleridir. Ancak kullanılan videonun çözünürlüğü yaklaşımın başarısını doğrudan etkilemektedir. Dolasıyla düşük çözünürlüklü görüntüler üzerinde çok başarılı sonuçlar elde edilememektedir.

Görünüm tabanlı yaklaşımlar kişinin vücut şeklini ya da hareketlerini doğrudan görüntülerden işlerler. Bu yaklaşımlarda en yaygın kullanım silüet tabanlı bir gösterim elde etmek ve bu gösterim üzerinden öznitelikler çıkarmaktır. Literatürde silüetlerden elde edilen çeşitli yürüyüş temsilleri vardır. Çalışma [10] ve [11]'de yazarlar farklı kıyafet seçimlerine karşı dayanıklı bir yöntem geliştirmek amacıyla yürüyüş özniteliklerini insan iskeletlerinden çıkarmışlardır. Bir yürüme döngüsü boyunca elde edilen ikili insan silüetlerinin matematiksel ortalaması alınarak yürüyüş enerji imgesi (YEİ) elde edilir [12]. YEİ hesaplaması oldukça kolay olan ve yüksek başarım elde edilen bir gösterim biçimidir. Ancak zamansal bilgileri kaybeder. Çalışma [13]'te de ikili insan silüetleri ve YEİ kullanmışlardır ancak YEİ'de kaybolan zamansal bilgileri her bir çerçeve üzerinden yakalayabilmek için LSTM [14] modelinden faydalanılmıştır. Yaklaşımlarını temel aldığımız çalışma [15]'te yazarlar, yürüme döngüsünden elde edilen insan silüetlerinden öznitelik çıkarmışlardır. Çalışma [16]'da yürüyüş döngüsünü hesaplama maliyetini ortadan kaldırmak amacıyla birkaç farklı sabit döngü uzunluğu seçilerek yürüyüş tanınmaya çalışılmıştır. Benzer şekilde [17]'de yürüyüş döngüsü hesaplanmamış ve yürüyüş dizisindeki tüm çerçeveler kullanılmıştır. Çalışma [18]'de bir silüet görünümünün konum bilgisini içerdiği ve dolayısıyla yürüyüşü zamansal bilgi elde edebilmek için bir dizi olarak kabul edilmiştir. Çalışma [19]'da yazarlar yürüyüş esnasında insan vücudunun farklı bölümlerinin çeşitli görsel görünümlere ve farklı hareket şekillerine sahip olduğunu gözlemlemişlerdir. Bunun üzerine yazarlar kişilere ait silüet görüntülerini parçalara ayırarak yürüyüşü analiz etmişlerdir. Silüet görüntüler veya silüet tabanlı gösterimler dışında da çalışma [20]'de öznitelikler RGB görsellerden çıkarılmıştır. Yazarlar zamansal değişiklikleri modellemek için LSTM [14] modeli kullanmışlardır. Model tabanlı yaklaşımlara kıyasla, görünüm tabanlı yaklaşımlar daha hızlı çalışırlar ve uygulamaları daha basittir. Bununla birlikte, silüet tabanlı yürüyüşten kişi tanıma yaklaşımlarının çoğu için statik şekil bilgisinin dinamik bilgiden daha önemli bir rol oynayabileceği düşünülmektedir [21]. Ayrıca açı farklılıklarından, kişinin kılık-kıyafetinden ve ölçekleme gibi faktörlerden etkilenirler.

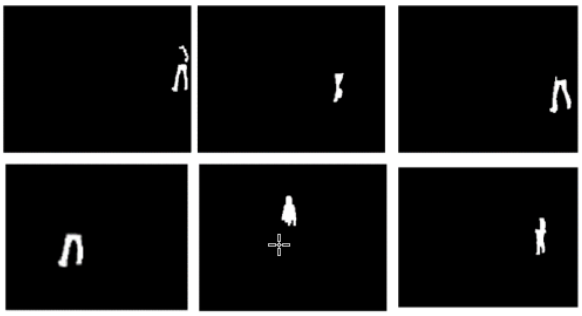
3. Önerilen Yöntem

Bu bölümde çalışmamızda kullandığımız yaklaşım anlatılmıştır. Önerilen yöntem toplam beş adımdan oluşmaktadır. Bu adımlar sırasıyla; veri kümeleri üzerinde uygulanan ön işleme adımları, modelin ince ayarlanma süreci, derin ağlarda eğitilmiş model yardımıyla öznitelik çıkarma, öznitelik birleştirme yöntemleri ve sınıflandırmadan oluşmaktadır. Her adım alt başlıklarda detaylıca anlatılmıştır.

3.1 Veri Kümelerinin Düzenlenmesi

CASIA-B [3] veri kümesi RGB video kayıtları ile birlikte ayrıca silüet görüntülerini de içermektedir. Her çerçevede arka plan çıkarma işlemi yapılarak kişiler bulundukları ortamdan soyutlanmıştır. Modelin eğitimi sırasında kişinin kıyafetlerinin renklerine, desenlerine odaklanma riskini ortadan kaldırmak amacıyla ikili insan silüeti görüntüleri çıkarılmıştır. Silüet çıkarma işlemi hata yapılma olasılığı yüksek bir işlemdir. Olası bölütleme hataları durumunda Şekil 1'deki gibi hatalı silüet görüntüleri elde edilebilmektedir. CASIA-B [3] veri kümesi ışık, yansıma, kıyafet deseni gibi birçok faktörden ötürü hatalı bölütlenmiş ikili silüet görüntüleri içermektedir. Modelin eğitimi öncesinde bu verileri temizlemek amacıyla bir ön işleme adımı uygulanmıştır.

Yürüyüş vücudumuzun tamamını kullanarak gerçekleştirdiğimiz bir eylemdir. Bu eylem gerçekleşirken baş, omuz, kollar, kalça ve ayakların her biri yürüyüşü öznel kılan özellikler hakkında bilgi içerir. Dolayısıyla çalışmamızda ilk amacımız tam ve bütün bir vücut şekline sahip görseller elde etmektir. Bunu gerçekleştirebilmek, veri kümesindeki hatalı görselleri temizlemek için basit bir yol izlenmiştir. Bir insan vücudunun tamamını içeren bir çerçevenin içinde en az kaç beyaz piksel bulunabileceği yönünde deneysel sayılar önerilmiştir. Daha sonra veri kümesinde bu beyaz piksel sayısının altında kalan görseller incelenmiş olup istenmeyen verileri elemek için 800 piksel sayısı çalışmalarımız için yeterli ve uygun bulunmuştur. Şekil-1'de 800'den daha az beyaz piksel içerdiği için veri kümesinden çıkarılan görsellerin birkaç örneği görülmektedir. Hatalı silüetler sadece vücudun belirli parçasını içerip tamamını içermeyen silüetlerden oluşmamaktadır. Kıyafetin arka plan rengine benzer çizgiler içermesi nedeniyle kesik kesik elde edilen silüetler, yerdeki yansımanın da insana ait olduğu sanılıp insan vücuduyla birleştirilmiş olarak sunulan silüetler yine en az 800 beyaz piksel içermekte ancak doğru bölütlenmiş insan vücudu görüntüleri olmamaktadır. Bu tip hatalı görseller veri kümesinden temizlenememiş olup, modelin eğitimi ve testi sırasında kullanılmışlardır.



Şekil-1: CASIA-B veri kümesinden temizlenen hatalı veri örnekleri.

3.1.1 RGB Girdilerin Üretilmesi

İkili insan silüeti görüntüleri ve silüet tabanlı yürüyüş enerji imgeleri elde edilirken kişinin kıyafet, aksesuar gibi görünümlerini etkileyecek faktörlerin bir kısmından soyutlandırılmış bir veri elde edilir. Ancak bu görüntüleri elde etmek hem hesaplama maliyeti hem de zaman açısından pratik uygulamalarda tercih edilmeyebilir. Ayrıca, arka plan çıkarma, insan vücudunu bölütleme gibi ön işleme adımları hataya çok açıktır. Arka planın karmaşıklığı, kişinin kıyafetinin arka planla benzer renkler içermesi ya da çok karmaşık desenler içermesi elde edilecek ikili insan silüeti görüntüsünün şeklinin bozulması konusunda risk taşımaktadır. Bu tip olumsuzlukları en aza indirmek ve pratik uygulamalarda ön işleme adımlarını mümkün olabildiğince azaltıp ham veriyi en hızlı yoldan işleyebilmek amacıyla RGB görseller üzerinde de çalışılmıştır. CASIA-B [3] yürüyüş veri kümesi ikili insan silüetleri ile birlikte RGB videoları da kullanımımıza sunduğu için bu veri kümesinden deneylerde kullanılmak üzere Şekil-2'deki adımlar izlenerek RGB girdiler üretilmiştir.

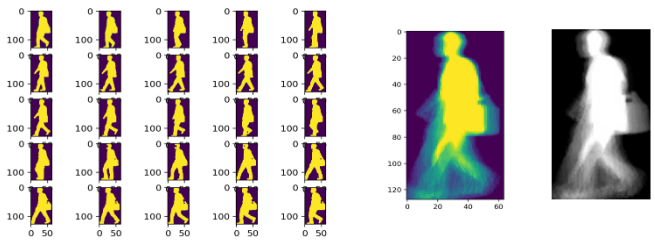
Öncelikle video çerçeveleri üzerinde insan vücudunu tespit edebilmek amacıyla literatürde poz tahmini için eğitilmiş DeepHRNet [22, 23, 24] modelinden faydalanılmıştır. İlk olarak çerçeve üzerinde kişinin tespiti yapılmıştır. Daha sonra ikili insan silüetlerinin elde edilişindeki adımlarla benzer adımlar izlenmiştir. Kişi çerçeve içinden kırpılmıştır. Modele girdi olarak beslenebilmesi için uygun boyutlara yani 128x88 boyutuna getirilmiştir. Veri kümeleri arasındaki yürüyüş yönü farklılığını ortadan kaldırmak için çerçevenin dikey eksene göre yansıması alınmıştır. Bu işlemler CASIA-B [3] veri kümesinde bulunan tüm yürüyüş videoları üzerinde tekrarlanmış ve RGB görüntülerden oluşan bir yürüyüş veri seti elde edilmiştir.



Şekil-2: RGB girdilerin elde edilme adımları.

3.1.2 Yürüyüş Enerji İmgelerinin Üretilmesi

Çalışmamızda farklı girdi türleri ile deneyler gerçekleştirilmiştir. Silüet ve RGB görüntülere ek olarak elimizde bulunan silüetlerin ortalaması alınarak elde edilen yürüyüş enerji imgeleri de farklı bir girdi tipi olarak modele verilmiştir. Şekil-3'te ikili insan siluetlerinin ortalamasıyla elde edilen bir YEİ örneği verilmiştir. Deneyler birer kez de ham yürüyüş videolarından doğrudan elde edilmiş RGB insan vücudu görselleri üzerinde tekrarlanmıştır.



Şekil-3: İkili insan silüetlerinin ortalaması ile elde edilen YEİ örneği.

Yürüyüş, benzer örüntülerin tekrarlandığı sürekli bir eylemdir. Örüntülerin tekrara dönmediği tekil hali bir yürüyüş döngüsü olarak adlandırılır. Sağ veya sol ayak topuğunun yere ilk teması ile başlayıp, aynı ayağın tekrar yere değmesine kadar geçen sürenin tamamı literatürde yürüyüş döngüsü veya yürüyüş periyodu olarak adlandırılmaktadır. Yapılan hesaplamalar sonucu tespit edilen yürüyüş döngüsü üzerinden kimlik tanıma işlemleri gerçekleştirilir. Çalışmamızda yürüyüş döngüsü hesaplamaları yapılmamıştır. Onun yerine bir yürüyüş döngüsünün bir saniye civarında sürdüğü kabul edilip, elimizdeki veri tabanlarında her 30 çerçevede bir yürüyüş döngüsünün olduğu varsayılmıştır. Yürüyüş enerji imgeleri, ikili insan silüeti içeren her 30 çerçevenin Denklem 1'deki gibi matematiksel ortalaması alınarak elde edilmiştir.

$$YE\dot{I}(x,y) = \frac{1}{N}\sum_{t=1}^{N} B(x,y,t) \quad (1)$$

Denklem 1'de N bir yürüme döngüsünde bulunan toplam çerçeve sayısıdır ve bizim çalışmamızda bu değer 30 olarak kabul edilmiştir. $B(x,y,t)$, ikili insan silüeti görüntülerinden oluşan bir yürüyüş dizisinin tamamıdır. t, yürüyüş dizisindeki her bir ikili insan silüetini, (x,y) de iki boyutlu görüntü koordinatları temsil etmektedir.

3.2 Veri Artırma

Makine öğrenmesinde modellerin başarısı ile kullanılan verilerin miktarı ve kalitesi doğru orantılıdır. Bu nedenle çalışmamızda geliştirilen modelin performansını yükseltmek amacıyla veri çoğaltma yöntemlerine başvurulmuştur.

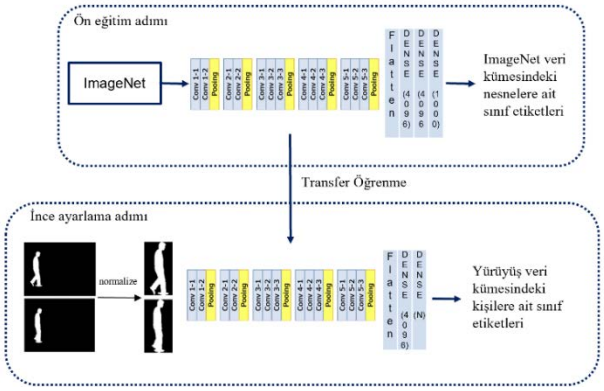
Veri artırma yöntemleri görüntü veya zaman serisi *(time-series)* seviyelerinde yapılabilmektedir. Görüntü seviyesinde veri artırımı genelde döndürme, kaydırma, ölçeklendirme, kırpma, kesme gibi dönüşümlerle yapılmaktadır. Ancak yürüyüş dizisine ait çerçeveler üzerinde bu işlemleri yapmak yürüyüş örüntüsüne zarar verir. Dolayısıyla bu dönüşümler yerine yürüyüş örüntüsünü bozmadan, sadece benzer örüntülerde veriler üretilerek veri miktarı arttırılmaya çalışılmıştır. Bunu gerçekleştirebilmek için her yürüyüş dizisi üzerinde %50 oranında örtüşümle veri miktarı iki katına çıkarılmıştır. %50 oranında örtüşüm; 30 çerçevelik bir yürüyüş dizisinin son 15 çerçevesinin, bir sonraki yürüyüş dizisini elde etmek üzere tekrar kullanılmasına karşılık gelmektedir. Şekil-4'te yeni yürüyüş örüntülerinin oluşturulma süreci gösterilmektedir.



Şekil-4: Yürüyüş enerji imgelerinin çoğaltılma işlemi

3.3 Eğitim ve İnce Ayarlama

Derin öğrenmede belirli bir problem için çok büyük veri kümeleri ile eğitilmiş modeller başka problemlerin çözümü için de kullanılabilir. Farklı transfer öğrenme stratejileriyle önceden eğitilmiş modelin kazandığı bilgiler yeni modele aktarılabilir. Çalışmamızda, daha önce nesne tanıma problemi için, 1000 sınıfa ait yaklaşık 1,2 milyon görseli içeren ImageNet [2] veri kümesi ile eğitilmiş VGG16 [5] modelinden faydalanılmıştır. Önceden eğitilmiş bu modelin doğrudan öznitelik çıkarıcı olarak kullanması ile modeli yürüyüşten kimlik tanıma problemine uyarlamak amacıyla ince ayarlama yapılmasının katkısı araştırılmıştır.

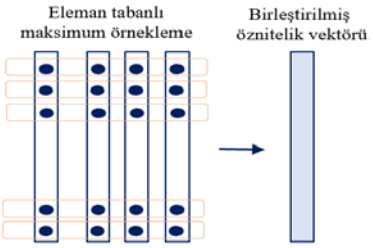


Şekil-5: İnce ayarlı modelin elde edilme süreci.

Önceden eğitilmiş VGG16 [5] modeli sınıflandırma ile ilişkili tam bağlantılı katmanları olmadan, sadece öznitelik çıkarıcı bölümüyle yüklenmiştir. Daha sonra sınıflandırma bölümü için iki tam bağlantıya sahip katman eklenmiştir. Bunlardan ilki 4096 nörona sahiptir. Son tam bağlantılı katman yürüyüş verileri içeren eğitim kümesindeki kişi sayısı kadar nörona sahip olacaktır. Deneylerimizde CASIA-B [3] veri kümesine ait 124 kişi ve yine CASIA-B [3] veri kümesine ait 74 kişi içeren iki farklı eğitim kümesi ile deneyler gerçekleştirilmiştir. Bu nedenle eğitimden önce, kullanılacak veri kümesine göre son tam bağlı katmanın sayısı 124 veya 74 olarak güncellenmiştir. Bu çalışmada elimizde yeterli yürüyüş verisi olması göz önünde bulundurularak tam bağlı katmanlar ve son üç evrişim bloğu yürüyüş verileriyle eğitilmiştir. İnce ayarlanmış modelin elde ediliş süreci Şekil 5'te gösterilmiştir.

3.4 Öznitelik Çıkarımı

Çalışmamızda bir yürüyüş dizisini temsil edecek öznitelik vektörlerini elde etmek için VGG16 [5] evrişimli sinir ağı mimarisi kullanılmıştır. VGG16 [5] mimarisi çalışmamızda sınıflandırıcı olarak değil de öznitelik çıkarıcı olarak kullanıldığından standart olarak kabul ettiği 224x224 girdi boyutundan bağımsız çalışılmıştır. Bu çalışmada insan vücudundaki en/boy oranlarını daha iyi koruyabilmek, daha hızlı eğitim ve belleğin daha verimli kullanılabilmesi adına 128x88 girdi boyutuna sahip görseller kullanılmıştır. Girdinin ağa verilmesinden sonra, mimariye sonradan eklenen, sınıflandırma katmanından önceki son tam bağlı katmandan öznitelik çıkarımı yapılmıştır. İlgili katman 4096 nöron içerdiğinden elde edilen vektörün boyutu 4096 olmuştur. Bu vektör ağa girdi olarak verilen çerçevenin bir temsili niteliğindedir. Her kişiye ait yürüyüş dizilerinin çerçeveleri



Şekil-6: Eleman bazında maksimum örnekleme işlemi

teker teker ağa girdi olarak verilmiş ve hepsine ait öznitelik vektörleri elde edilmiştir. Kişilere ait öznitelik vektörlerinin sayısı, yürüyüş dizilerinin uzunluğu ile birebir orantılıdır. A kişisine ait beş YEİ var ise beş öznitelik vektörü, 45 ikili insan silüeti görüntüsü varsa 45 öznitelik vektörü elde edilmiş olur.

3.5 Öznitelik Birleştirimi

Yürüyüş eylemi; bir başlangıcı, gerçekleşme süresi ve bitişi olan karmaşık bir uzamsal-zamansal motor kontrol davranışıdır [25]. Çalışmamızda her bir yürüyüşü temsil edecek tek bir öznitelik vektörü elde edilmiştir. VGG16 [5] mimarisinin son tam bağlı katmanı olan sınıflandırıcı katmanından elde edilen 4096 boyutlu öznitelik vektörlerini tek vektörde birleştirmek üzere üç farklı yöntem denenmiştir. Kişilerin yürüyüş dizilerine ait çerçeve sayısına bağlı olarak farklı sayıda öznitelik vektörleri elde edilmiştir. Bu vektörler Şekil-6'daki gibi eleman bazında maksimum, eleman bazında ortalama ve eleman bazında medyan yöntemleri ile birleştirilmişlerdir. Böylece elde edilen öznitelik vektörüne bir yürüyüş dizisindeki her çerçevenin katkısı olmuştur.

Denklem 2'de eleman bazında maksimum örnekleme yöntemi ile *N* tane çerçeveden elde edilen, 4096 boyutlu birleştirilmiş öznitelik vektörünün (böv) hesaplama formülü verilmiştir. Denklemlerdeki *N*, yürüyüş dizisine ait toplam çerçeve sayısıdır. fv , her bir çerçeve için VGG16 [5] mimarisinin son tam bağlı katmanından elde edilmiş 4096 boyutlu öznitelik vektörünü temsil etmektedir.

$$\text{böv}_{\mathbf{i}} = \text{maksimum}_{j=1}^{N}\ fv_{i,j}, i = 0 \dots 4095 \qquad (2)$$

Denklem 3'te eleman bazında ortalama yöntemi ile *N* tane çerçeveden elde edilen, 4096 boyutlu birleştirilmiş öznitelik vektörünün hesaplama formülü verilmiştir.

$$\text{böv}_{\mathbf{i}} = \text{ortalama}_{j=1}^{N}\ fv_{i,j}, i = 0 \dots 4095 \qquad (3)$$

Denklem 4'te eleman bazında medyan yöntemi ile *N* tane çerçeveden elde edilen, 4096 boyutlu birleştirilmiş öznitelik vektörünün hesaplama formülü verilmiştir.

$$\text{böv}_{\mathbf{i}} = \text{medyan}_{j=1}^{N} fv_{i,j}, i = 0 \dots 4095 \qquad (4)$$

3.1 Sınıflandırma

Yöntemimizin son aşaması test verisini sınıflandırma adımıdır. Test kümesi içerisinde galeri ve sorgu kümelerini barındırır. Galeri kümesi, test kümesindeki kişilerin referans yürüyüş verilerini içermektedir. Sorgu kümesi ise tanınması gereken kişiye ait yürüyüş verilerini içermektedir. Önerilen yöntemin son adımında kişilere ait referans ve sorgu görüntüleri arasında bir benzerlik hesaplanarak sınıflandırma işlemi gerçekleştirilir. Önerdiğimiz yöntemin sınıflandırma adımı için sırasıyla aşağıdaki adımlar uygulanmıştır:

- Galeri kümesindeki tüm yürüyüş dizilerine ait öznitelik vektörleri çıkarılır.
- Sorgu yürüyüş dizisine ait öznitelik vektörü çıkartılır.
- Sorgu dizisine ait öznitelik vektörü ile tüm galeri dizilerinin öznitelik vektörleri arasındaki kosinüs benzerliği hesaplanır.
- En yüksek benzerliğe sahip olan vektör çiftinden galeri dizisine ait yürüyüş dizisinin etiketi, sorgu dizisinin etiketi kabul edilir.

İki vektör arasındaki benzerlik Öklid, Manhattan, Minkowski uzaklığı gibi çeşitli uzaklık hesaplama yöntemleri ile de bulunabilir. Bizim çalışmamızda iki vektör arasındaki benzerliğin hesaplanması için Denklem 5'te formülü verilen kosinüs benzerliği kullanılmıştır.

$$\text{kosinüs}\,(\mathbf{g},\mathbf{p}) = \frac{\mathbf{gp}}{\|\mathbf{g}\|\|\mathbf{p}\|} = \frac{\sum_{i=1}^{4096} g_i p_i}{\sqrt{\sum_{i=1}^{4096} (g_i)^2}\sqrt{\sum_{i=1}^{4096} (p_i)^2}} \quad (5)$$

Denklemdeki $\mathbf{g}$, galeri yürüyüş dizisinden elde edilen 1x4096 boyutlu öznitelik vektörünü temsil etmektedir. $\mathbf{p}$, sorgu yürüyüş dizisinden elde edilen 1x4096 boyutlu öznitelik vektörünün temsilidir. g_i ve p_i bu vektörlere ait elemanlardır.

4. Deneyler ve Sonuçlar

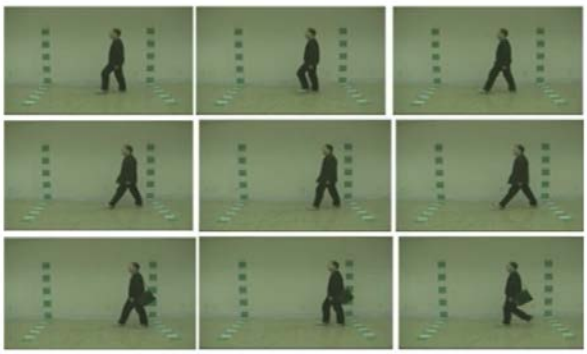
Bu bölümde deneylerde kullanılan veri kümeleri, gerçekleştirilen deneyler, deney düzenekleri ve bu deneylerin sonuçları sunulmuştur.

4.1 Veri Kümeleri

Bu çalışma kapsamında iki yürüyüş veri kümesi ve bir tane kişiyi yeniden tanıma veri kümesi kullanılmıştır. CASIA-B [3] yürüyüş veri kümesi bir kişiye ait çok fazla yürüyüş örneği içermesi nedeniyle çalışmalara dahil edilmiştir. OU-ISIR BP [4] yürüyüş verisi çok sayıda kişi içermesi ve modelin fazla sayıda insan üzerindeki başarısını test edebilmek amacıyla seçilmiştir. PRID-2011 [26] veri kümesi ise gündelik yaşamda karşılaşma ihtimalimizin daha yüksek olduğu yürüyüş videosu örneklerini içermesi nedeniyle deneylerimizde kullanılmıştır. Her veri kümesi detaylıca alt başlıklarda açıklanmıştır.

4.1.1 CASIA-B Veri Kümesi

Çalışmamızda modelimizin yürüyüş veri kümeleri üzerindeki başarısını ölçebilmek için literatürde sık kullanılan CASIA-B [3] veri kümesinden faydalanılmıştır. CASIA-B [3] veri kümesi içerdiği her birey için farklı varyasyonlarda ve çok geniş bir açı yelpazesinden kaydedilen yürüyüş görüntülerine sahiptir. Veri kümesi 124 kişiye ait farklı varyasyonlarda yürüyüş videoları içermektedir. Her bir kişinin normal yürüyüş videosu, çanta taşırken kaydedilmiş yürüyüş videosu ve üzerinde mont/ceket varken kaydedilmiş yürüyüş videosu bulunmaktadır. Kişilerden düz bir duvar önünde bir başlangıç noktasından bitiş noktasına kadar yürümeleri istenmiştir. Bu eylem gerçekleştirilirken, odanın 11 farklı noktasına 18 derecelik açı farkıyla (0, 18, 36, 54, 72, 90, 108, 126, 144, 162, 180) sırayla dizilmiş 11 kamera yürüyüşü eş zamanlı olarak kaydetmiştir. Her kişiden önce duvarın önünde, kameralar sol taraflarında kalacak şekilde altı kere normal bir yürüyüş gerçekleştirmeleri istenmiştir. Kameralar aynı anda bu altı yürüyüşü kaydetmiştir. Daha sonra kişilerden montlarını giymeleri ve montla birlikte iki kere daha yine aynı yönde yürümeleri istenmiştir. Kameralar bu görüntüleri de kaydetmiştir. Son olarak her bireyden birer çanta (sırt çantası, kol çantası, el çantası) seçmeleri ve onu taşıyarak yürümeleri istenmiştir. Bu son görüntülerle birlikte kameralar her bir kişi için toplam on video kaydı gerçekleştirmiştir. Bu da 10 video x 11 açı x 124 kişi = 13640 video kaydını temsil etmektedir. Hem RGB yürüyüş videoları hem de videolardan çıkarılmış ikili insan silüeti görüntüleri ile paylaşıma sunulan bu veri kümesi oldukça büyük bir yürüyüş veri kümesidir. Video çerçeve boyutu 320x240'tır ve videolar saniyede 25 çerçeve içermektedir. Video süreleri, kişilerin yürüme hızına bağlı olarak 1-5 saniye arasında değişiklik göstermektedir. CASIA-B veri kümesine ait örnek görüntüler Şekil-7'de verilmiştir.



Şekil-7: CASIA-B veri kümesine ait örnek görüntüler.

4.1.2 OU-ISIR Büyük Popülasyon Veri Kümesi

Çok yaygın kullanımı olan bir diğer yürüyüş veri kümesi Osaka Üniversitesi Bilimsel ve Endüstriyel Araştırma Enstitüsü (OU-ISIR) tarafından oluşturulan Büyük Popülasyon veri kümesidir. OU-ISIR BP [4] veri kümesi modelin çok büyük sayıda kişi içeren bir veri kümesi üzerindeki başarısını gözlemlemek için seçilmiştir. Veri kümesi 4000'den fazla kişiyi içermektedir. Her kişiden aynı koşullar altında, elleri boş bir şekilde, kendi kıyafet seçimleriyle düz bir rotada yürümeleri istenmiştir. Kişilerin yürüdükleri rota uzunluğu on metre olsa da başındaki ve sonundaki üç metrede kişilerin hızlanma ve yavaşlama yapacakları düşünülerek ortadaki dört metreyi kaydedecek iki kamera ile yürüyüşlerinin kaydı gerçekleştirilmiştir. Kaydedilen yürüyüşler 55, 65, 75 ve 85 derece olmak üzere dört farklı açıya göre gruplandırılmıştır. Her açı için kişilere ait iki yürüyüş dizisi mevcuttur. Video çerçevelerinin boyutu 640x480'dir. Bir saniyede 30 çerçeve kaydedilmektedir.

OU-ISIR BP [4] veri kümesinde yürüyüşler sadece ikili insan silüetleri formunda paylaşılmaktadır. RGB video görüntüleri

kişisel verilerin gizliliğini korumak amacı ile paylaşılmamıştır. İkili insan silüetlerinin video dizilerinden çıkarılması işleminde bölütleme hatalarının oluşma riski çok yüksektir. Bu nedenle yazarlar çizge-kesme tabanlı [27] bir bölütleme algoritması yardımıyla arka plan çıkarma işlemi yapmışlardır. Olası bölütleme hatalarını gidermek için veri setinin tamamı üzerinde en az iki kez gözle kontroller ve elle düzenlemeler yapılmıştır. Bu nedenle, CASIA-B [3] veri kümesine kıyasla çok daha kaliteli ve doğru bölütlenmiş ikili insan silüetleri görülmektedir. Veri kümesine ait örnek görüntüler Şekil-8'de verilmiştir.

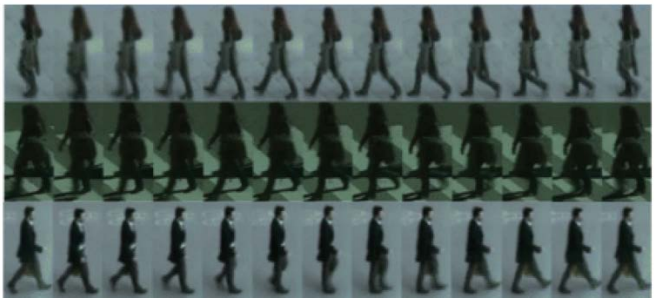


Şekil-8: OU-ISIR BP veri kümesine ait örnek görüntüler.

4.1.3 PRID 2011 Veri Kümesi

Literatürde bulunan yürüyüş veri kümeleri kısmen kontrollü ortamlarda kaydedilmiş yürüyüş görüntüleri içermektedir. Kişiler belirli bir A noktasından B noktasına sadece yürüyüş eylemini gerçekleştirerek geçiş yapmaktadır. Ancak günlük hayatta yürüme eylemi ile birlikte başka eylemler de geçekleştirebiliriz. Yürürken telefonla görüşmek, bir şeyler atıştırmak, çantanın içine bakmak gibi çoğu zaman üst bedenimizin farklı eylemlerle hareket düzenini bozabilecek durumlar mevcuttur. Dolasıyla önerdiğimiz yaklaşımın günlük hayatta karşılaşacağımız yürüyüş örnekleri üzerindeki başarısını ölçmek üzere PRID 2011 [26] veri kümesi üzerinde deneyler gerçekleştirilmiştir.

PRID 2011 [26] veri kümesi, kişiyi yeniden tanıma problemlerinde kullanılmak üzere toplanmıştır. Veri kümesi iki statik kamera tarafından Şekil-9'da görüldüğü gibi açık alanda yürüyen insanların görüntülerini içermektedir. A kamerasından 385, B kamerasından 749 kişi kayda alınmıştır. Bu kişilerin 200 tanesi hem A hem de B kamerasının görüş alanına girmiştir. Veri kümesinde her kişiye ait minimum beş ardışık çerçeve bulunmaktadır. Beş görüntü, bir kişinin yürüyüşünden tanınması için çok az bir veri miktarı olsa da veri kümesi genelinde 30 ardışık çerçeveden daha az çerçeveye sahip kişi sayısı az olduğundan bu problem göz ardı edilmiştir. Veri kümesindeki kişiler çerçevelerden kırpılmış, normalize edilmiş ve 128x64 boyutlarında paylaşıma sunulmuştur. PRID 2011 [26] veri kümesi yürüyüş yönü farklılıkları içermektedir.



Şekil-9: PRID 2011 veri kümesine ait örnek görüntüler.

4.2 Uygulama Detayları

Çalışmamızda Python betik dili kullanılmıştır. Kullanılan Python sürümü 3.8.8 ve açık kaynak kodlu Keras derin öğrenme kütüphanesinin sürümü 2.4.3'tür. Modellerin eğitimi ve deneyler NVIDIA GeForce GTX TITAN X kartı üzerinde gerçekleştirilmiştir. Eğitimlerde ilk öğrenme oranı 0.0001, eksilme parametresi 0.000002 olarak seçilmiştir. Veri kümesi üzerinden her bir geçiş sonrasında öğrenme oranı Denklem 5'teki gibi zamana dayalı bir öğrenme oranı planlaması ile güncellenmiştir.

$$\text{Öğrenme Oranı} = \text{Öğrenme Oranı} * 1/(1 + \text{eksilme} * \text{tur}) \quad (5)$$

Ağın eğitiminde Adam optimizer [28] kullanılmıştır. Ağa her seferinde verilen veri boyutu 128'dir ve veri kümesi toplam 50 kez ağa gösterilmiştir.

4.3 Ön Eğitimli Model - İnce Ayarlı Model Karşılaştırma

Çalışmamızda gerçekleştirilen ilk aşamada iki ayrı model öznitelik çıkarıcı olarak deneylerde kullanılmış ve OU-ISIR BP [4] veri kümesi üzerinde test edilmiştir.

İlk model, nesne tanıma probleminde kullanılmak üzere ImageNet [2] veri kümesi üzerinde önceden eğitilmiş VGG16 [5] modelidir. ImageNet [2] veri kümesi 1000 sınıfa ait, yaklaşık 1.2 milyon görüntü içermektedir. İkinci model, önceden eğitilmiş nesne tanıma problemine özgü modelin CASIA-B [3] veri kümesine ait yürüyüş verileri ile ince ayarlanmasıyla elde edilmiş bir modeldir.

İnce ayarlama aşamasında kullanılan eğitim kümesi, CASIA-B [3] veri kümesindeki 124 kişinin ilk dört normal yürüyüş videolarına ait ikili insan silüetlerinden oluşmaktadır. Test kümesi OU-ISIR BP [4] veri kümesidir. Bu yürüyüş veri kümesinde her kişi için iki yürüyüş dizisi örneği bulunmaktaydı. Bir tanesi galeri diğeri sorgu yürüyüş dizisi olarak kullanılmıştır. Veri kümelerinde kişiler birbirlerine zıt yönde yürümektedir. Yürüme yönü farklılığını ortadan kaldırmak için CASIA-B [3] veri kümesine ait her insan siluetinin dikey eksene göre simetrisi alınarak iki veri kümesinde de aynı yönde yürüyüş görüntüleri elde edilmiştir. Ayrıca iki yürüyüş veri kümesindeki görüntülerin kaydedildiği açılar birbirinden farklıdır. Yüksek açı farklılıklarının görünüm tabanlı yaklaşıma olumsuz etkisini azaltmak için iki veri kümesindeki birbirine en yakın açılar seçilmiş ve model onlarla eğitilip, eğitimde kullanılan açıya en yakın açıyla da test edilmiştir. Örneğin; CASIA-B [3] veri kümesindeki kişiler 0, 18, 36, 54, 72, 90, 108, 126, 144, 162, 180 derecelerinden kaydedilirken, OU-ISIR BP veri kümesinde 55, 65, 75, 85 dereceden kaydedilmiş yürüyüş

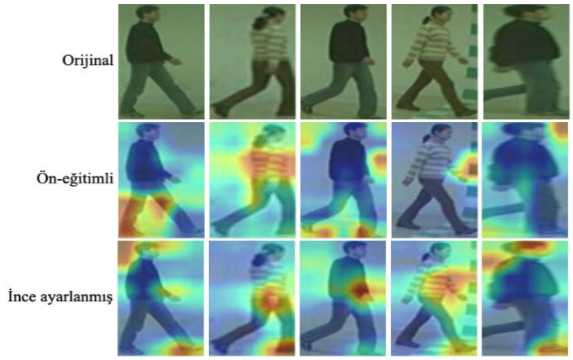
örnekleri içermektedir. Dolayısıyla CASIA-B [3] veri kümesindeki 54 dereceyle eğitilen model, OU-ISIR BP [4] veri kümesindeki 55 dereceden kaydedilmiş yürüyüşler üzerinde test edilmiştir. CASIA-B'deki 72 dereceye ait görüntülerle eğitilen modelin başarısı, OU-ISIR BP [4] veri kümesindeki 65 ve 75 dereceden kaydedilmiş görüntüler için ayrı ayrı test edilmiştir.

Çizelge 4.1: OU-ISIR Büyük Popülasyon veri setinde transfer öğrenmenin sonuçlara katkısı.

	CASIA-B eğitim seti	V-54	V-72		V-90
		V-55	V-65	V-75	V-85
Rank-1	Ön-eğitimli	83,23	85,26	87,66	88,69
	İnce ayarlı	92,73	91,16	92,04	93,67
Rank-5	Ön-eğitimli	90,87	92,43	93,59	94,92
	İnce ayarlı	96,80	95,16	96,46	97,48

Modellerin yürüyüşten kişi tanıma başarımlarını ölçmek için Rank-1 ve Rank-5 metrikleri seçilmiştir. Yaklaşımımızda galeri ve sorgu dizilerinden elde edilen öznitelik vektörleri arasında benzerlik skorları hesaplanmıştır. Rank-1 metriği için en yüksek benzerliğe sahip olan vektör çiftinin aynı kişiye ait olması beklenirken, Rank-5 metriği için en yüksek benzerlik skoruna sahip ilk 5 öznitelik vektör çiftlerinden birinin doğru eşleşmesi beklenmektedir. Ön eğitimli model ve yürüyüşe daha fazla adapte olması için ince ayarlanmış model ile gerçekleştirilen deneyin sonuçları Çizelge 4.1'de verilmiştir.

Sonuçlar incelendiğinde özellikle Rank-1 metriği için ince ayarlı modelin başarısının %5 ila %9,5 oranında iyileşme sağladığı görülmektedir.



Şekil-10: Grad-CAM yaklaşımı ile görüntüdeki önemli bölgelerin görselleştirilmesi.

Rank-5 metriği için ön eğitimli modelin başarısı %90 üzerinde, ince ayarlı model ile elde edilen sonuçlar %95'in üzerinde doğrululuk oranına sahiptir. Bu sonuçlar transfer öğrenmenin çalışmamıza yaptığı olumlu katkıyı göstermektedir.

Transfer öğrenmenin katkısını gözlemlemek için bir de her iki modelin de tahminlerini yapmadan önce odaklandıkları bölgeleri görselleştiren Gradyan-ağırlıklı Sınıf Aktivasyon Haritalaması [29] (Grad-CAM) yaklaşımından faydalanılmıştır. Bu yaklaşım, modelin kararını etkileyen, görüntüdeki önemli bölgelerin kaba bir yerelleştirme haritalamasını üretmek için

Çizelge 4.2: CASIA-B veri setinde silüet girdi tipinin özdeş görünüm ve çapraz görünüm deneyleri doğru sınıflandırma oranı sonuçları.

		Sorgu Açısı										
		0	**18**	**36**	**54**	**72**	**90**	**108**	**126**	**144**	**162**	**180**
Galeri Açısı	**0**	100	100	84	56	38	36	42	60	74	90	90
	18	94	100	100	96	66	44	60	70	76	94	84
	36	78	100	100	100	98	76	82	88	82	88	66
	54	60	92	100	100	100	98	98	96	86	68	54
	72	36	68	94	100	100	100	100	100	84	56	36
	90	34	46	70	98	100	100	100	100	76	44	30
	108	24	54	80	96	98	98	100	100	92	52	34
	126	60	72	88	92	90	92	100	100	100	88	60
	144	68	84	90	92	80	74	100	100	100	96	74
	162	88	94	82	62	38	34	64	80	100	100	98
	180	94	86	58	40	38	32	36	60	84	100	100

Çizelge 4.3: CASIA-B veri setinde YEİ girdi tipinin özdeş görünüm ve çapraz görünüm deneyleri doğru sınıflandırma oranı sonuçları.

		Sorgu Açısı										
		0	**18**	**36**	**54**	**72**	**90**	**108**	**126**	**144**	**162**	**180**
Galeri Açısı	**0**	96	6	46	32	18	28	22	36	40	58	64
	18	70	100	92	66	42	38	46	56	60	52	44
	36	48	84	100	92	74	70	64	68	54	50	26
	54	40	58	88	98	94	72	70	70	48	30	14
	72	30	44	72	88	96	84	68	64	44	30	18
	90	20	26	58	74	92	92	92	78	46	28	14
	108	22	32	52	64	80	96	96	86	50	44	12
	126	32	46	58	54	56	80	80	84	88	58	32
	144	38	54	52	46	40	52	52	90	94	78	46
	162	42	38	32	16	22	36	36	56	78	94	76
	180	54	36	26	18	14	22	22	34	54	74	94

evrişimsel sinir ağının son evrişim katmanındaki sınıfa özgü gradyan bilgisini kullanmaktadır. Şekil-10'da kırmızı alanlar modelin en çok odaklandığı bölgeleri, mavi ile işaretlenmiş alanlar ise modelin en az ilgilendiği alanları temsil etmektedir. Önceden eğitilmiş ve ince ayarlanmış modellerin en çok odaklandığı bölgelere bakılacak olursa, ince ayarlanmış modelin yürüyüş hakkında daha fazla bilgi içeren bölgelere yani baş, omuz, eller, ayaklar gibi bölgelere odaklandığı görülmektedir. Önceden eğitilmiş model ise zaman zaman arka plan gibi yürüyüş probleminden daha anlamsız bölgelere odaklandığını görülmektedir.

4.4 Farklı Girdi Deneyleri

Literatürde yürüyüşten kimlik tanıma probleminde silüet tabanlı görüntüler yaygın kullanılmaktadır. Çalışmamızda da ikili insan silüetleri ve yine silüet tabanlı yürüyüş enerji imgeleri ile deneyler gerçekleştirilmiştir. Bunlara ek olarak pek yaygın kullanımı olmayan RGB görüntüler ile de yürüyüşten kimlik tanımaya çalışılmıştır. RGB görüntüler ikili insan silüeti görüntülerine kıyasla daha fazla bilgi içermektedir. Ancak derin ağlarda modelin bu görüntüler üzerindeki arka plana, kişinin kıyafetlerine, renklere, desenlere odaklanma ihtimali de vardır. Bu da kişinin yürüyüşünden değil de görünümünden tanınmış olma riskini taşımaktadır. Ancak bir başka taraftan ikili insan silüetini oluşturmak, günlük hayatta uygulamaya geçirilmesi çok pratik olmayan bir işlemdir. Arka plan çıkarma işlemi veya bölütleme işlemi zaman alan ve hataya oldukça açık işlemlerdir. Videolardan elde edilen RGB çerçevelerden çıkarmanın önerdiğimiz yaklaşımın başarımına katkısı araştırılmıştır.

CASIA-B [3] yürüyüş veri kümesi ikili insan silüetleri ile birlikte RGB videoları da kullanımımıza sunduğu için deney bu veri kümesi üzerinde gerçekleştirilmiştir. Deneyde modelin ince ayarlanma aşamasında kullanılması üzere CASIA-B [3] veri kümesinin ilk 74 kişisi seçilmiştir. İlk 74 kişiye ait tüm yürüyüş videoları eğitim seti olarak kullanılmıştır. Modelin başarısını ölçmek için kullanılacak test setinde ise CASIA-B [3] veri kümesindeki son 50 kişi bulunmaktadır. Test kümesindeki kişilere ait referans görüntüler içeren galeri kümesi, test kümesindeki her kişinin ilk dört normal yürüyüş videosundan oluşmaktadır. Sorgu kümesi ise bu kişilerin son iki normal yürüyüş videosundan oluşmaktadır. Veri setlerinde bulunan tüm videolardan ikili insan silüetleri, yürüyüş enerji imgeleri ve RGB görüntüler oluşturulmuştur ve her bir girdi tipi için ayrı bir model eğitilmiştir.

Görünüm tabanlı çözüm yöntemlerinde açı farklılıkları modelin başarısını doğrudan etkilemektedir. Aynı açıdan kaydedilmiş (özdeş görünüm) görüntülere bakarak yürüyüşten kişiyi tahmin etmek çok zor bir problem değildir. Ancak farklı açılardan kaydedilmiş (çapraz görünüm) görüntülere bakarak yürüyüşten kimlik tanıma problemi daha zor bir problemdir. Çalışmamızda farklı girdi tipleri için hem özdeş hem çapraz görünüm koşullarında deneyler yapılmıştır.

İkili insan silüeti, YEİ ve RGB görüntüler için detaylı sonuçlar sırasıyla Çizelge 4.2, Çizelge 4.3 ve Çizelge 4.4'te raporlanmıştır. Her girdi tipinin çapraz görünüm için ortalama başarımları ise Çizelge 4.5'te verilmiştir. YEİ üretilirken veri miktarındaki ciddi azalmanın sonuçlara etkisi olumsuz olmuştur ve en düşük başarılar bu girdi tipinde elde edilmiştir. İkili insan silüetinin doğrudan kullanılması YEİ'lerine göre başarıları %16,6-31,6 oranında arttırmıştır. RGB görseller kullanıldığında bu iyileşme %29,6-51,2 aralığına yükselmiştir. Ancak burada modelin yürüyüşten çok kıyafetlere ve arka plana odaklanmış olma riski mevcuttur.

Çizelge 4.4: CASIA-B veri setinde RGB girdi tipinin özdeş görünüm ve çapraz görünüm deneyleri doğru sınıflandırma oranı sonuçları.

		Sorgu Açısı										
		0	**18**	**36**	**54**	**72**	**90**	**108**	**126**	**144**	**162**	**180**
Galeri Açısı	**0**	98	100	98	98	94	86	82	82	82	84	84
	18	100	100	100	100	98	92	82	82	84	86	84
	36	94	100	100	100	98	98	94	90	84	86	82
	54	92	96	98	98	100	98	96	94	92	84	76
	72	86	98	100	100	100	98	98	98	96	86	80
	90	84	94	100	100	100	98	100	96	96	86	82
	108	86	88	98	100	100	100	100	100	98	90	88
	126	88	92	96	98	98	98	98	100	98	96	88
	144	86	92	92	96	98	98	98	100	100	98	94
	162	88	90	92	92	86	94	94	98	98	100	100
	180	90	86	90	92	86	88	92	94	92	98	100

Çizelge 4.5: Farklı girdi tiplerinin çapraz görünüm için ortalama başarı (%) sonuçları. (Özdeş görünümler ortalamaya dahil edilmemiştir.)

		Sorgu Açısı										
		0	**18**	**36**	**54**	**72**	**90**	**108**	**126**	**144**	**162**	**180**
Galeri Açısı (0-180 derece)	Silüet	63,6	79,6	84,6	83,2	74,6	68,4	78,2	85,4	85,4	77,6	62,6
	YEİ	40	48	57,6	55	53	51,8	55,2	63,8	56,2	49,2	34,6
	RGB	**89,4**	**93,6**	**96,4**	**97,6**	**95,8**	**95**	**93,4**	**93,4**	**92**	**86,4**	**85,8**

4.5 Farklı Öznitelik Birleştirim Yöntemleri Karşılaştırma

Önerilen yaklaşımda bir yürüyüş dizisine ait tüm çerçeveler VGG16 [5] modeline sırayla verilmektedir. Her bir çerçeve için VGG16 [5] mimarisinin 4096 nörona sahip son tam bağlantılı katmanından öznitelik çıkarımı yapılmaktadır. Böylece bir yürüyüş dizisine ait çerçeve sayısı kadar 4096 boyutlu öznitelik vektörleri elde edilmektedir. Yürüyüşü temsil edecek

çok sayıda vektörle işlemler yapmak yerine, bu özniteliklerin bir örnekleme yöntemi ile birleştirilmesi önerilmiştir. Bunu gerçekleştirebilmek için elde edilmiş tüm öznitelik vektörleri arasında eleman bazında maksimum örnekleme ile birleştirme, eleman bazında ortalama ve medyan örnekleme ile birleştirme yöntemleri denenmiştir.

Deneyde her girdi tipi için ayrı bir model eğitilmiştir. Modellerin eğitim kümelerinde CASIA-B [3] veri kümesindeki ilk 74 kişiye ait tüm açılardan kaydedilmiş yürüyüş videoları bulunmaktadır. Test kümelerinde ise son 50 kişi mevcuttur. Her bir kişiye ait ilk dört normal yürüyüş videoları galeri kümesine, son iki normal yürüyüş videoları sorgu kümelerine dahil edilmiştir. Üç tip görüntü için de üç farklı örnekleme yöntemi ile deney tekrarlanmış ve özdeş görünüme ait ortalama sonuçlar Çizelge 4.6'da çapraz görünüme ait ortalama sonuçlar ise Çizelge 4.7'de raporlanmıştır.

Çizelge 4.6: Özdeş görünümde farklı birleştirme yöntemlerinin üç görüntü tipi üzerindeki ortalama başarıları (%).

	Maksimum	Medyan	Ortalama
Silüet	100	99,82	99,82
YEİ	95,82	94,36	96,54
RGB	99,45	99,09	96,54

Çizelge 4.7: Çapraz görünümde farklı birleştirme yöntemlerinin üç görüntü tipi üzerindeki ortalama başarıları (%).

	Maksimum	Medyan	Ortalama
Silüet	76,65	66,79	75,69
YEİ	51,31	47,54	51,04
RGB	92,62	92,16	92,65

Özdeş görünüm sonuçları üç birleştirme yönteminde de birbirine oldukça yakındır ve iki girdi tipinde en başarılı sonuçlar eleman bazında maksimum için elde edilmiştir. Çapraz görünüm sonuçları incelendiğinde medyan yöntemi diğer iki yönteme göre daha az başarılı sonuçlar üretmiştir. Çapraz görünüm sonuçları detaylı analiz edildiğinde, eleman bazında maksimum örnekleme yönteminin ikili insan silüeti görüntülerinde 11 açının 9'unda, RGB görüntülerde 8'inde, YEİ tipinde 11 açının 7'sinde diğer yöntemlerden daha iyi veya eşit performans göstermiştir. Dolayısıyla çalışmalarımızda eleman bazında maksimum örnekleme yöntemi ile birleştirme yönteminden yararlanılmıştır.

4.6 PRID 2011 Veri Kümesinde Yürüyüşten Kişi Tanıma

Yürüyüşten kimlik tanıma görevi pratik hayatta güvenlik sistemlerinde de kullanılabilir. Ancak gündelik yaşamda karşımıza çıkan yürüyüş verilerinin, literatürde yaygın kullanılan CASIA-B [3], OU-ISIR BP [4], TUM GAID [30], SOTON [31] gibi yürüyüş veri kümelerinde bulunan verilerden çok farklı olduğu söylenebilir. Örneğin bireyin bir hedefe doğru yürüyüş hareketini gerçekleştirdiği sırada aynı anda birden fazla eylemi de gerçekleştirdiği görülebilir. Telefonla konuşmak, bir şeyler atıştırmak, montunu çıkarmak gibi. Bu durumlar göz önünde bulundurulduğunda önerilen yaklaşımın kontrolsüz ortamda kaydedilmiş veriler üzerindeki başarısı ölçülmek istenmiştir. PRID 2011 [26] veri kümesi gündelik yaşamdan kaydedilen görüntüler içermesi nedeniyle yaklaşım bu veri kümesi üzerinde de test edilmiştir.

Deneyde kullanılan model daha önce farklı girdi tipleriyle gerçekleştirilen deneydeki RGB görüntülerle eğitilen modeldir. Yani CASIA-B veri kümesine ait ilk 74 kişinin tüm yürüyüş videolarından elde edilmiş RGB görüntülerle elde edilen model, PRID 2011 [26] veri kümesindeki RGB görüntüler ile test edilmiştir. PRID 2011 [26] veri kümesinde her iki kamerada da kayda girmiş olan ilk 200 kişi üzerinde çalışılmıştır. A kamerasından kayda girmiş 200 kişi galeri kümesine dahil edilmişken B kamerasından kayda giren 200 kişiye ait görüntüler sorgu kümesine dahil edilmiştir ve bu düzenek AtoB düzeneği olarak adlandırılmıştır. Literatürde PRID 2011 [26] veri kümesi için test setine ait iki resmi düzenek paylaşıldığı için galeri ve sorgu kümesindeki veriler yer değiştirilerek deney bir kez de bu şekilde tekrarlanmıştır ve bu düzenek BtoA olarak adlandırılmıştır. Deneyin sonuçları Çizelge 4.8'de verilmiştir.

Sonuçlar incelendiğinde, başarıların oldukça düşük olduğu görülmektedir ancak bu beklenen bir durumdur. Veri kümesi içerisinde yürüyüş yönü farklılıkları, yürüme eylemi ile birlikte paralel gerçekleştirilen başka eylemlerin gerçekleştirilmesi, taşınan poşet, valiz gibi eşyaların bacaklardaki yürüyüş verilerinin tam ve doğru alınamamasına sebep olması gibi durumlar başarıyı olumsuz etkilemektedir. Bu durum da pratik uygulamalarda yürüyüşten kimlik tanıma problemi için değişen koşullara daha dayanıklı ve sistematik çözümler araştırılması gerektiğini doğrulamaktadır.

Çizelge 4.8: PRID 2011 veri kümesi üzerinde gerçekleştirilen deneyin sonuçları. (%).

	Galeri-Sorgu	Rank-1	Rank-5
PRID 2011	AtoB	37,5	48,5
	BtoA	36,5	50,5

4.7 Farklı Yaklaşımların Karşılaştırılması

Literatürde yürüyüşten kimlik tanıma problemi farklı yaklaşımlarla ele alınmıştır. Bizim çalışmamızda yürüyüş verilerinden öznitelik çıkarımında derin evrişimsel sinir ağlarından faydalanılmıştır. Genel olarak, derin öğrenme modellerinin performansı büyük ölçüde eğitim setlerinin ölçeğine bağlıdır. Bu bölümde listelenen çalışmalarda literatürde yaygın kullanılan ve eğitim setinin büyük ölçekli olduğu kabul edilebilecek bir deney düzeneği kullanılmıştır. Geleneksel yöntemlerle yürüyüşü analiz edip öznitelik çıkaran yaklaşımlarla derin öğrenme tabanlı yaklaşımların deneysel sonuçları Çizelge 4.9'da karşılaştırmaya sunulmuştur. Bizim çalışmamız hariç diğer tüm çalışmaların sonuçları doğrudan orijinal makalelerinden alınmıştır. İlk üç satırda bizim yöntemimizin farklı girdi tipleri için sonuçları yer almaktadır. Bu sonuçlar elde edilirken kullanılan model, farklı girdi tipleri deneyindeki RGB modelle ve PRID 2011 deneyindeki modelle aynı modeldir. Çalışma [32] farklı ön işleme yaklaşımları ve farklı ESA tabanlı modeller ile çapraz görünüm senaryolarını temel alan kapsamlı bir değerlendirme çalışmasıdır. Bu makaleye ait yaklaşımlardan grup seyrekliğini (CNN-LB GS) ve üç evrişim katmanı (CNN-LB 3CL) kullanarak gerçekleştirilmiş deneylerin sonuçları karşılaştırmaya sunulmuştur. CNN-LB GS [32], CNN-LB 3CL [32], GaitNet [20], 3DCNN [33] ve Image Set [34] çalışmalarında bizim çalışmamızda olduğu gibi derin öğrenme yaklaşımı ile yürüyüşten kimlik tanıma problemine

çözüm üretilmeye çalışılmıştır. Ancak ViDP [35] çalışmasında geleneksel yöntemler kullanılarak görünümden etkilenmeyen öznitelikler çıkarılmış ve yürüyüş sınıflandırması gerçekleştirilmiştir.

Sonuçlar incelendiğinde derin öğrenme yaklaşımından faydalanarak gerçekleştirilmiş deneylerin hepsinin geleneksel yöntemlere göre daha başarılı sonuçlar ürettiği görülmektedir. Ayrıca GaitNet [20] çalışmasının ve bizim önerdiğimiz yaklaşımdaki RGB deneylerinin sonuçlarına bakarak, derin öğrenme yaklaşımında silüet tabanlı görüntüler yerine RGB görüntülerin kullanılmasının modelin başarısını oldukça yüksek oranlara taşıdığı söylenebilir.

5. Sonuç

Bu çalışmada, derin evrişimsel sinir ağları kullanılarak yürüyüşten kimlik tanıma problemi analiz edilmiştir. Önceden eğitilmiş evrişimsel sinir ağı mimarisinin ve ince ayarlanmış evrişimli sinir ağı mimarisinin yürüyüşten kimlik tanıma performansı üzerindeki etkisi araştırılmıştır ve yürüyüş verileri ile ince ayarlanmış modelin daha iyi sonuçlar verdiği gözlemlenmiştir. Ayrıca daha hızlı ve daha düşük maliyetli bir çözüme ulaşmak için ikili insan silüeti ve silüet tabanlı girdi tipleri kullanmak yerine, bazı ön işleme aşamaları kaldırılarak ham videolardan elde edilen RGB girdi tipiyle de deneyler gerçekleştirilmiştir. Yürüyüş döngülerini hesaplamadan ve ikili insan silüetlerini çıkarmadan, doğrudan RGB görüntülerini kullanarak, çapraz görünüm deneylerinde %86'nın üzerinde başarı elde edilmiştir. Yürüyüşten kimlik tahmini aşamasında VGG16 [5] mimarisinin son tam bağlantılı katmanından çıkarılan öznitelik vektörleri arasında eleman tabanlı maksimum örnekleme yöntemiyle birleştirilen vektörler kullanılmıştır. Ve bu yöntemle elde edilen birleştirilmiş öznitelik vektörü özdeş görünüm ve çapraz görünüm için en iyi sonuçları vermiştir. Literatürde belirli şartlar ve kurallar çerçevesinde toplanmış yürüyüş görüntüleri üzerinde gerçekleştirilen çalışmalar dışında bir de gündelik hayattan yürüyüş verileri içeren PRID 2011 [26] veri kümesi üzerinde yürüyüşten kimlik tanıma problemi için önerdiğimiz yaklaşım denenmiştir. Ancak bu veri kümesi üzerindeki başarı oranları diğer veri kümeleri üzerinde elde edilen sonuçlar kadar yüksek çıkmamıştır. Bu da pratik uygulamalarda yürüyüşten kimlik tanıma probleminin çözümü için gerçek dünya koşullarına, açı ve ölçek farklılıklarına daha dayanıklı çözümlerin araştırılması gerektiğini göstermektedir. Son olarak, derin öğrenme temelli yaklaşımlar ve geleneksel yöntemlerle gerçekleştirilen çalışmalara ait deneysel sonuçlar karşılaştırılmış ve derin öğrenme yaklaşımlarının geleneksel yöntemlere kıyasla daha iyi sonuç verdiği gözlemlenmiştir.

Yürüyüşten kimlik tanıma probleminde zaman boyutu olan dizi şeklindeki bir veriyle uğraştığımızdan ileriki çalışmalarda özyinelemeli sinir ağı mimarilerinden faydalanarak daha başarılı sonuçlar üretilebilir. RGB tipi görüntülerin silüet tabanlı yöntemlere göre daha fazla veri içermesi ancak istenilmeyen bilgilerin de model tarafından öğrenme riskini önlemek amacıyla gelecek çalışmalarda RGB görüntüler üzerinde bölütleme işlemi yapılarak kişilerin arka planla olan ilişkisi ortadan kaldırılabilir. Bunlara ek olarak çeşitli maskeleme teknikleri ile kıyafetlerden elde edilen şekil ve renk bilgilerinin de modele olan katkısı azaltılabilir. Böylece modelin arka plana veya görseldeki kıyafetlerin desenlerine odaklanma riski ortadan kaldırılacaktır.

Çizelge 4.9: CASIA-B veri setinde benzer ve farklı yaklaşımlarla ortalama doğruluk (%) karşılaştırması. (Özdeş görünüm durumlarının sonuçları hesaplanan ortalamalara dahil edilmemiştir.)

		Sorgu Açısı										
		0	**18**	**36**	**54**	**72**	**90**	**108**	**126**	**144**	**162**	**180**
Galeri Açısı (0-180 derece)	**Silüet**	63,6	79,6	84,6	83,2	74,6	68,4	78,2	85,4	85,4	77,6	62.6
	YEİ	40	48	57,6	55	53	51,8	55,2	63,8	56,2	49,2	34.6
	RGB	89,4	**93,6**	**96,4**	**97,6**	**95,8**	**95**	93,4	93,4	**92**	**86,4**	85.8
	CNN-LB 3CL [32]	79,1	88,4	95,7	92,8	89,1	87,0	89,3	92,1	94,4	89,4	75.4
	CNN-LB GS [32]	82,6	90,3	96,1	94,3	90,1	87,4	89,9	94,0	94,7	91,3	78,5
	GaitNet [20]	**91,2**	92,0	90,5	95,6	86,9	92,6	**93,5**	**96,0**	90,9	88,8	**89,0**
	3DCNN [33]	61.8	76,2	85,2	86,0	77,9	71,8	74,4	80,8	83,5	76,5	58,3
	Image Set [34]	-	-	-	78,0	-	60	-	75,0	-	-	-
	ViDP [35]	-	-	-	64,2	-	60,4	-	65,0	-	-	-

Kaynakça

[1] S. G. Işık and H. K. Ekenel, "Deep Convolutional Feature-based Gait Recognition Using Silhouettes and RGB Images," 2021 6th International Conference on Computer Science and Engineering (UBMK), 2021, pp. 336-341, doi: 10.1109/UBMK52708.2021.9559026.

[2] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C. ve Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge, International Journal of Computer Vision, 115(3), 211–252.

[3] Yu, S., Tan, D. ve 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), cilt 4, s.441–444.

[4] Iwama, H., Okumura, M., Makihara, Y. ve Yagi, Y. (2012). The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition, IEEE Trans. on Information Forensics and Security, 7, Issue 5, 1511–1521.

[5] Simonyan, K. ve Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition, 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings

[6] Tafazzoli, F. ve Safabakhsh, R. (2010). Model-based human gait recognition using leg and arm movements, Engineering Applications of Artificial Intelligence, 23(8), 1237–1246.

[7] Cunado, D., Nixon, M.S., Carter, J.N., 2003. Automatic extraction and description of human gait models for recognition purposes. Computer Vision and Image Understanding 90 (1), 1–41.

[8] Yoo, J.H., Hwang, D., Moon, K.Y. ve Nixon, M.S. (2008). Automated Human Recognition by Gait using Neural Network, 2008 First Workshops on Image Processing Theory, Tools and Applications, s.1–6.

[9] Zhang, R., Vogler, Ch., Metaxas, D., 2007. Human gait recognition at sagittal plane. Image and Vision Computing 25, 321–330.

[10] R. Liao, C. Cao, E. B. Garcia, S. Yu, and Y. Huang, "Posebased temporal-spatial network (PTSN) for gait recognition with carrying and clothing variations," in Chinese Conference on Biometric Recognition. Springer, 2017, pp. 474–483.

[11] W. An, S. Yu, Y. Makihara, X. Wu, C. Xu, Y. Yu, R. Liao, and Y. Yagi, "Performance evaluation of model-based gait on multi-view very large population database with pose sequences," IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 2, no. 4, pp. 421–430, 2020.

[12] Lam, T., Lee, R. ve Zhang, D. (2007). Human gait recognition by the fusion of motion and static spatio-temporal templates, Pattern Recognition, 40, 2563–2573.

[13] Y. Zhang, Y. Huang, S. Yu, and L. Wang, "Cross-view gait recognition by discriminative feature learning," IEEE Transactions on Image Processing, vol. 29, pp. 1001–1015, 2019.

[14] Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-term Memory. Neural computation. 9. 1735-80.

[15] Zhang, X., Sun, S., Li, C., Zhao, X. ve Hu, Y. (2017). DeepGait: A Learning Deep Convolutional Representation for Gait Recognition, J. Zhou, Y. Wang, Z. Sun, Y. Xu, L. Shen, J. Feng, S. Shan, Y. Qiao, Z. Guo ve S. Yu, (düzenleyenler), Biometric Recognition, Springer International Publishing, Cham, s.447–456.

[16] Wang, K., Liu, L., Lee, Y., Ding, X. ve Lin, J. (2019). Nonstandard Periodic Gait Energy Image for Gait Recognition and Data Augmentation, Z. Lin, L. Wang, J. Yang, G. Shi, T. Tan, N. Zheng, X. Chen ve Y. Zhang, (düzenleyenler), Pattern Recognition and Computer Vision, Springer International Publishing, Cham, s.197–208.

[17] Wu, Z., Huang, Y. ve Wang, L. (2015). Learning Representative Deep Features for Image Set Analysis, IEEE Transactions on Multimedia, 17(11), 1960–1968.

[18] Chao, Hanqing & He, Yiwei & Zhang, Junping & Feng, Jianfeng. (2019). GaitSet: Regarding Gait as a Set for Cross-View Gait Recognition. Proceedings of the AAAI Conference on Artificial Intelligence.

[19] Fan, Chao & Peng, Yunjie & Cao, Chunshui & Liu, Xu & Hou, Saihui & Chi, Jiannan & Huang, Yongzhen & Li, Qing & He, Zhiqiang. (2020). GaitPart: Temporal Part-Based Model for Gait Recognition.

[20] Ziyuan Zhang, Luan Tran, Xi Yin, Yousef Atoum, Jian Wan, Nanxin Wang, and Xiaoming Liu. Gait recognition via disentangled representation learning. In CVPR, 2019.

[21] Zhong Li, Jiulong Xiong, Xiangbin Ye, "A new gait energy image based on mask processing for pedestrian gait recognition," Proc. SPIE 11321, 2019 International Conference on Image and Video Processing, and Artificial Intelligence, 113212A (27 November 2019).

[22] Sun, K., Xiao, B., Liu, D. ve Wang, J. (2019). Deep High-Resolution Representation Learning for Human Pose Estimation, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), s.5686–5696.

[23] Xiao, B., Wu, H. ve Wei, Y. (2018). Simple Baselines for Human Pose Estimation and Tracking, V. Ferrari, M. Hebert, C. Sminchisescu ve Y. Weiss, (düzenleyenler), Computer Vision – ECCV 2018, Springer International Publishing, Cham, s.472–487.

[24] Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W. ve Xiao, B. (2021). Deep High-Resolution Representation Learning for Visual Recognition, IEEE Transactions on Pattern Analysis & Machine Intelligence, 43(10), 3349–3364.

[25] Adler, A., Youmaran, R. ve Loyka, S. (2009). Towards a measure of biometric feature information, Formal Pattern Analysis & Applications, 12, 261–270.

[26] Hirzer, M., Beleznai, C., Roth, P.M. ve Bischof, H. (2011). Person Re-Identification by Descriptive and Discriminative Classification, Proc. Scandinavian Conference on Image Analysis (SCIA).

[27] Y. Boykov and M. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images," in Proc. Int. Conf. Computer Vision, 2001, pp. 105–112.

[28] Kingma, D. ve Ba, J. (2014). Adam: A Method for Stochastic Optimization, International Conference on Learning Representations.

[29] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. ve Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, 2017 IEEE International Conference on Computer Vision (ICCV), s.618–626.

[30] Hofmann, M., Geiger, J., Bachmann, S., Schuller, B. ve 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.

[31] Shutler, J.D., Grant, M.G., Nixon, M.S. ve Carter, J.N. (2004). On a Large Sequence-Based Human Gait Database, Applications and Science in Soft Computing, Springer Berlin Heidelberg, Berlin, Heidelberg, s.339–346.

[32] Z. Wu, Y. Huang, L. Wang, X. Wang and T. Tan, "A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 2, pp. 209-226, 2017.

[33] Xing, W., Li, Y. ve Zhang, S. (2018). View-invariant gait recognition method by three-dimensional convolutional neural network, Journal of Electronic Imaging, 27, 1.

[34] Wu, Z., Huang, Y. ve Wang, L. (2015). Learning Representative Deep Features for Image Set Analysis, IEEE Transactions on Multimedia, 17(11), 1960–1968.

[35] Hu, M., Wang, Y., Zhang, Z., Little, J.J. ve Huang, D. (2013). View-Invariant Discriminative Projection for Multi-View Gait-Based Human Identification, IEEE Transactions on Information Forensics and Security, 8(12), 2034–2045.