Fırat Üniv. Müh. Bil. Dergisi
30(3), 121-129, 2018

Science and Eng. J of Fırat Univ.
30(3), 121-129, 2018

Dış Ortam Görüntülerindeki İnsan Hareketlerinin Hibrit Derin Öğrenme Yöntemleri Kullanarak Sınıflandırılması

Özlem ALGUR[1], Vedat TÜMEN[2], Özal YILDIRIM[2*]
[1]Elektrik - Elektronik Mühendisliği Bölümü, Munzur Üniversitesi, Tunceli, Türkiye
[2]Bilgisayar Mühendisliği Bölümü, Munzur Üniversitesi, Tunceli, Türkiye
[*]oyildirim@munzur.edu.tr

(Geliş/Received: 23.03.2018; Kabul/Accepted: 03.09.2018)

Özet

Bu makale çalışmasında, dış ortam görüntülerinde yer alan insan hareketlerinin otomatik sınıflandırılması için hibrit bir derin öğrenme yaklaşımı önerilmiştir. İlk olarak, dış ortamdan çekilen görüntü içerisindeki kişilerin tespiti sağlanmıştır. Bu amaçla, literatürde yaygın olarak kullanılan önceden eğitilmiş derin nesne tespit aracı olan YOLO kullanılmıştır. Dış ortam görüntülerinin elde edilmesinde Google street view platformu tercih edilmiştir. Daha sonra tespit edilen kişiler için hareket sınıfları oluşturulmuştur. Bu hareket sınıfları; sağa yürüme, sola yürüme, ayakta durma ve oturma şeklindedir. Böylece dış ortam görüntülerinden tespit edilen kişiler için kapsamlı bir veri seti oluşturulmuştur. Sınıfları belirlenen verilerin otomatik olarak tanınması işlemi için bir konvolüsyonel sinir ağı (KSA) modeli tasarlanmıştır. Eğitimi tamamlanan bu model, YOLO nesne tespit sistemi ile hibrit bir şekilde kullanılarak giriş görüntüsü içerisindeki kişi hareketlerinin otomatik olarak tanınmasını sağlamıştır. Makale kapsamında, dört sınıflı bir veri seti oluşturularak önerilen sistemin performans değerlendirmeleri yapılmıştır.

Anahtar Kelimeler: Derin öğrenme, İnsan hareket tespiti, Görüntü işleme, Nesne tespiti.

Classification of Human Actions in Outdoor Images Using Hybrid Deep Learning Methods

Abstract

In this paper, a hybrid deep learning approach is proposed for automatic classification of human actions in outdoor images. Firstly, the detection of the humans in the image taken from the outdoor environment is provided. For this purpose, YOLO, a pre-trained deep object detection tool widely used in the literature, is used. The Google street view platform was preferred to acquire outdoor images. Action classes were created for the detected humans. These action classes; walking to the right, walking to the left, standing and sitting. Thus, a comprehensive dataset has been created for humans identified from the outdoor images. A convolutional neural network (CNN) model is designed for the automatic recognition of classes. With this designed CNN model, automatic recognition of the human actions is provided. This trained model is used in a hybrid manner with the YOLO object detection system to automatically recognize the human actions in the input images.Within the scope of the article, a four-class dataset was created and performance evaluations of the proposed system were made.

Keywords: Deep learning, Human action recognition, Image processing, Object detection.

1. Giriş

Görüntü verileri içerisindeki kişilerin tespit edilmesi ve aynı zamanda bu kişilerin gerçekleştirdiği hareketlerin belirlenmesi insan makine etkileşimi çalışmaları için önemli bir problem alanıdır. Örneğin, video görüntüleri içerisindeki insan hareketleri, içerik tabanlı video arama çalışmalarında kullanılan etkili bir yaklaşımdır [1]. Yol güzergâhındaki yayaların algılanması [2-4] ve bu yayaların hareket yönlerinin tespiti; güvenlik sistemleri, sürücüsüz araçlar ve robotik sistemlerin gelişimi açısından oldukça önemlidir. Araçlara ya da benzer donanımlara yerleştirilebilecek bir görüntüleme sistemi ile insan hareketlerinin otomatik tespit edilmesi sağlanabilir. Kara yollarında araç hareket halindeyken gerçek görüntülerden insanların tespiti ve buna ilaveten hareketlerinin gerçek zamanlı algılanması, araçlarda erken uyarı sistemlerinin oluşturulması gibi birçok uygulama alanı için büyük önem arz etmektedir.

İnsan hareketi tanımı yapılırken çeşitli seviyelerde soyutlamalar kullanılmaktadır [5]. Örneğin sağ ayağın önde olması gibi atomik eylemlerin bir döngü içerisinde gerçekleşmesi ile (koşma gibi) bir hareket meydana gelmektedir. Genel tanımıyla insan vücudunun ardışık manevralarının sonucunda ortaya çıkan durum olarak ifade edilebilir [6]. Hareket tespiti için öncelikle kişilerin tespitinin gerçekleştirilmesi ve tüm görüntüden ayrılması gerekmektedir. Tespit edilen kişi görüntüleri üzerinde özellik çıkarımı ve sınıflandırma yöntemleri ile bu kişilerin hareketlerine dair anlamlı çıkarımlar yapılabilir [7-9]. Fakat hareket tanıma oldukça karmaşık bir problemdir. Karmaşık arka planlar, görüntü açısı, görüntüdeki kapanıklıklar doğru tanımayı etkileyen zorluklardır.

İnsan hareket tespiti ile ilgili yapılan çalışmalar incelendiğinde, video görüntüleri üzerinde yapılan çalışmalar yoğunluktadır. Yu ve Yuan [10] çalışmalarında, videolardaki belli insan hareketlerinin sınıflandırılması işlemini gerçekleştirmiştir. Geliştirilen algoritma ile iki farklı veriseti kullanılarak yapılan bu çalışmada insan hareketleri olan el sallama, el sıkışma ve boks yapma gibi belirli hareketler sınıflandırılmış ve MSRII veri seti üzerindeki ortalama başarımları %61.30 olmuştur.

Baccoubhe ve diğ. [11] yapmış oldukları çalışmada; insan hareketlerini herhangi bir ön bilgi kullanmadan sınıflandırmayı öğrenen, tamamen otomatikleştirilmiş derin bir model önermişlerdir. KSA'ların üç boyuta genişletilmesine dayanan sistemin ilk basamağı, otomatik olarak uzay-zamansal özellikleri öğrenmektir. Daha sonra tekrarlanan bir sinir ağı her bir zaman aşımı için öğrenilen özelliklerin zamansal gelişimini göz önüne alarak her diziyi sınıflandırmak üzere eğitilmiştir. KTH veri setinde yapılan deneysel sonuçlar; önerilen yaklaşımın mevcut derin modellerden daha iyi performans gösterdiğini ve en iyi ilgili çalışmalarla karşılaştırılabilir sonuçlar verdiğini göstermişlerdir.

Eroğlu ve diğ. [12] video işleme teknikleri ile görüntülerdeki ilginç noktaların tespitini sağlayarak bu noktalar üzerinde öznitelik çıkarımı gerçekleştirmişlerdir. Elde edilen öznitelikler üzerinde k-en yakın komşu sınıflandırma yöntemi kullanarak insan harekelerini sınıflandırmışlardır. Yapılan çalışmada zıplayarak ilerleme ve koşma videoları için başarı %83 oranında olmuştur.

Niebles ve diğ. [13] farklı özellik çıkarımı modelleri kullanarak KTH, Weizmann ve buz pateni veri setleri üzerinde video etiketlemeden sonra yapmış oldukları sınıflandırmada ortalama olarak sırasıyla %83.33, %90 ve %80.76 tanıma performansı sağlamışlardır. Schüldt ve diğ. [14] çalışmalarında 25 insana ait dört farklı senaryo ile yürüme, zıplama, koşma, boks, el sallama ve alkışlama hareketlerini sınıflandırma işlemini gerçekleştirmişlerdir. Çalışmalarında, destek vektör makineleri ve görüntülerde tespit edilen yerel benzerliklerin eşleştirilmesi yöntemlerini kullanmışlardır.

Bahsedilen bu çalışmalar el ile ayarlanmış özellik çıkarım metotları ve sığ yapılı sınıflandırıcılar içermektedir. Makine öğrenmesi alanında son zamanlarda ön plana çıkan derin öğrenme algoritmaları hem özellik çıkarımı hem de sınıflandırma yapısını bir arada bulunduran etkili yaklaşımlar olmuştur [19, 25-31]. İnsan hareketi tanımada derin öğrenme kullanan çalışmalar ön plana çıkmaktadır.

Ji ve diğ. [15] derin öğrenme kullanarak insan hareketlerinin otomatik tanınmasını sağlamışlardır. Çalışmada, özellik belirleme sürecini otomatikleştiren ham girişlere doğrudan etki edebilen derin bir 3D-KSA modeli önermişlerdir. Bu model ile KTH veri seti üzerinde ortalama %90.2 tanıma başarımı raporlamışlardır.

Rahmani ve diğ. [16] insan hareketlerinin tanınmasında büyük problem teşkil eden bilinmeyen ve görülmeyen açılar için derin bir ağ önermişlerdir. Simonyan ve Zisserman [17] video görüntülerindeki hareketlerin tespiti için iki yönlü bir KSA yapısı önermişlerdir.

Bu çalışmada, dış ortam görüntüleri içerisinde yer alan insan hareketlerinin otomatik tespitinde derin öğrenme modellerinin hibrit bir kullanımı önerilmiştir. Google Street View ortamından elde edilen görüntülerde, daha önceden eğitilmiş nesne tespit aracı olan YOLO (You Only Look Once) ile kişilerin tespiti yapılmıştır. Elde edilen birçok kişi görüntüsü dış ortamdaki hareketlerine göre dört sınıfta etiketlendirilmiştir. Etiketlendirilmiş veri setleri üzerinde önerilen KSA modelinin eğitimleri gerçekleştirilmiştir. Eğitilen modelin test verileri üzerindeki tanıma performansları

değerlendirilmiş böylece farklı görüntü boyutları ve farklı veri bölümleme oranları kullanılmıştır.

Önerilen modelin avantajları şu şekilde özetlenebilir;

- El ile ayarlanan özellik çıkarımları ve sığ yapılı sınıflandırıcılar yerine tamamen uçtan-uca ham görüntüler üzerinde çalışan bir tanıma sistemidir.
- Görüntü içerisindeki kişilerin tespiti önceden eğitilmiş bir ağ modeli ile otomatik olarak yapılmaktadır. Böylece dış ortam görüntülerindeki birçok kişinin tespiti yüksek bir performansla sağlanmaktadır.
- Derin öğrenme modellerinin hibrit bir şekilde kullanarak istenen amaca yönelik tanıma işlemlerinin yapılması açısından literatüre katkı sunmaktadır.
- Dış ortamda kullanılabilecek tanıma sistemleri için önemli bilgiler sağlamaktadır.

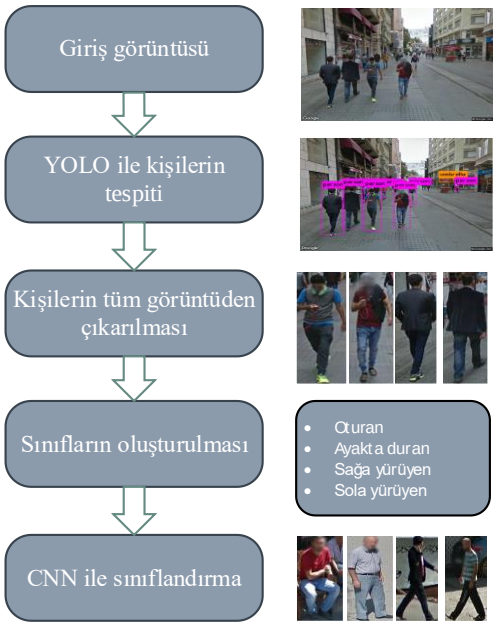
2. Derin Öğrenme

Derin öğrenme makine öğrenmesi alanının en güncel konusudur. Derin öğrenme, yapay sinir ağları olarak adlandırılan beynin yapısından ve işlevinden esinlenmiş olan algoritmalarla alakalı makine öğrenmesinin bir alt dalıdır. Makine öğrenmesi, derin öğrenme ve yapay zekâ ifadeleri birbirleri ile ilişkili kavramlardır [18]. Derin öğrenme, büyük miktarlarda, etiketlenmemiş eğitim verilerinden öznitelik çıkarım yöntemleri ile özelliklerin saptamasını yapabilen sistemler oluşturmak için geliştirilmiştir. Derin öğrenme; görüntü sınıflandırma, imge işleme, video analizi ve sınıflama, ses tanıma ve işleme ve doğal dil öğrenme süreci gibi günümüzde birçok alanda yoğun bir şekilde kullanmaktadır [19].

Derin öğrenme mimarileri giriş verilerinin düşük seviyeli temsillerinden (çizgi, kenarlar, renk, doku vb.) yüksek seviyeli temsillerine kadar çeşitli seviyelerde özelliklerini elde ederek bu verilerin işlenmesini sağlar. Derin öğrenmenin en sık kullanılan yapıları, Derin Sinir Ağları, Otomatik Kodlayıcılar ve Boltzmann Makinelerinin türevleri olan Kısıtlı Boltzmann Makineleridir [20]. Derin öğrenmeyi geleneksel yapay sinir ağı modellerinden ayıran en önemli özelliklerinden biri çok fazla sayıda katmanla çalışabilmesidir. Geleneksel yapay sinir ağlarında her hücre kendinden bir önceki ve bir sonraki katmanlarda bulunan tüm hücrelere bağlı durumdadır. Her bağlantıda hesap gerektiren matematiksel işlemler mevcuttur. Katman sayısında ve hücre sayısında artışın olabilmesi için yüksek miktarda CPU gücüne ihtiyaç duyulmaktadır. Derin bir ağ yapısının oluşturulması için kişisel bilgisayarlar içerisindeki CPU gücü tam anlamıyla yeterli değildir. Bunun yerine günümüzde güçlendirilmiş GPU tabanlı ve yüksek hafızalı işlemciler üzerinden çalışılmaktadır.

3. Önerilen Tanıma Sistemi

Çalışmada, dış ortam görüntülerinden insan hareketlerinin tespiti için derin öğrenme tabanlı hibrit bir tanıma sistemi geliştirilmiştir. Bu tanıma sistemi ile Google street view uygulamasından elde edilen görüntülerdeki insanların hareketlerinin belirlenmesi sağlanmıştır. Önerilen sistemin çalışma adımları Şekil 1'de gösterildiği gibidir.



Şekil 1. Geliştirilen tanıma sisteminin çalışma adımları

Google street view uygulamasından elde edilen bazı örnek görüntüler Şekil 2'de verilmiştir. Öncelikle görüntülerin elde edilmesi ve bu görüntüler için bir veri tabanı oluşturulması sağlanmıştır. Daha sonraki aşamada toplanan bu görüntüler üzerinde nesne tespit aracı olan YOLO ile kişiler tespit edilmiştir. Tespit edilen kişiler için sınıf ayırımı yapılmıştır. Bu sınıflar; sağa

hareket, sola hareket, ayakta duran ve oturan şeklinde belirlenmiştir.



Şekil 2. Google street view uygulamasından elde edilen bazı görüntüler

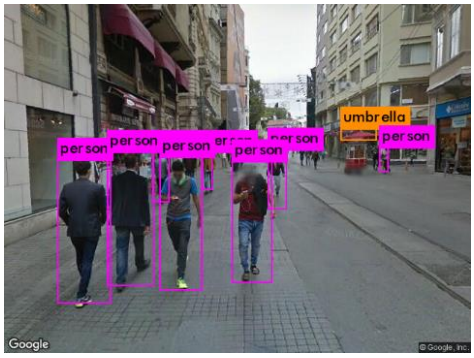
3.1. Kişilerin tespit edilmesi

Görüntü içerisindeki insanların tespiti işlemi, görüntü işleme alanında nesne tespiti problemi olarak ele alınmaktadır. Nesne tespitine dayalı birçok çalışma mevcuttur fakat bunlardan en iyisi ve en hızlısı Redmon ve diğ. [21] tarafından sunulmuş olan You Only Look Once (YOLO) sistemidir. YOLO sınırlayıcı kutuları ve sınıf olasılıklarını tahmin etmek için tek bir sinir ağı kullanan bir derin KSA ağıdır. YOLO'nun mimarisi KSA tabanlı olup hızlı bir yapıya sahiptir. Görüntüleri gerçek zamanlı olarak saniyede 45 kare hızında işler. Ağın daha küçük bir versiyonu olan Fast YOLO, saniyede 155 kare işleme özelliğine sahiptir [22]. Mevcut nesne tespiti sistemlerine kıyasla, YOLO daha fazla lokalizasyon hatası yapar, fakat arka planda yanlış pozitifleri tahmin etme olasılığı daha düşüktür. Son olarak YOLO, nesnelerin en genel temsillerini öğrenir. Öznitelik çıkarımı için 7×7×1024, aktivasyon haritası elde etmek için 24 konvolüsyon katmanı kullanılır. Bu görüntü bilgileri iki tam bağlı katmana sahip tahmin bölümüne geçirilir. Olasılık ve nesne koordinat bilgileri son katman olan tam bağlı katmanı tarafında üretilmektedir [23].

Bu çalışmada, orijinal görüntüler içerisinden kişilerin tespit edilmesi için YOLO aracı kullanılmıştır. Şekil 3'te, görüntü içerisindeki kişilerin tespit edilmesi ve tüm görüntüden ayrılmasına ait aşamalar gösterilmiştir.



(a) Orijinal görüntü



(b) Kişilerin tespit edilmesi



(c) Kişilerin bölümlenmiş halleri

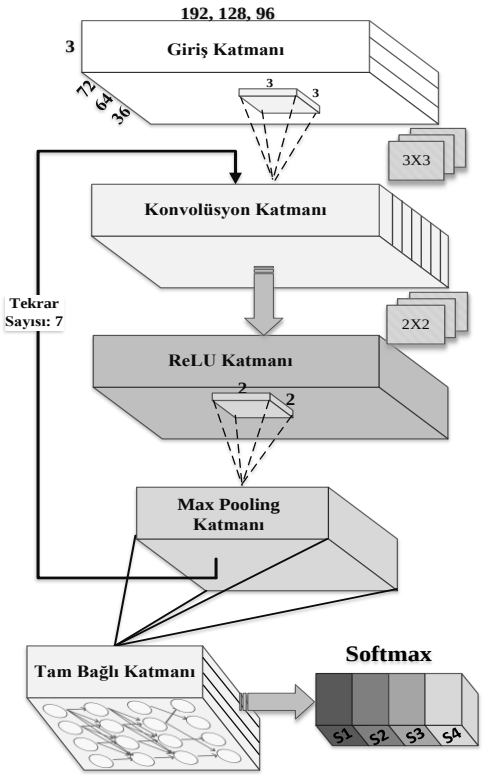
Şekil 3. Hareket görüntülerinin elde edilme aşamaları

3.2. KSA sınıflandırma modeli

Veri tabanında bulunan tüm imgelerin YOLO sisteminden geçirildikten sonra elde edilen insan görüntülerinin sınıflandırılması için KSA yöntemi kullanılmıştır. Özellik çıkarma yöntemi için ham görüntü piksellerini kullanarak sınıflandırma işlemi yapılmaktadır. Bu yapıların başlıca avantajı görüntüye ait çeşitli seviyelerde özelliklerin hiyerarşik olarak elde edilmesidir. İlk katmanlarda köşeler gibi düşük seviyeli özellikler çıkartılır, sonraki katmanlarda ise daha karmaşık şekiller gibi yüksek seviyeli özelikler çıkarılır. KSA, en düşük sayıda önişlem ile görüntülerin

işlenmesi ve doğrudan görsel kalıplar çıkarmak için geliştirilmiştir. KSA'ların, diğer sınıflandırma algoritmalardan daha iyi bir performans göstermesinin nedenlerinden birisi öznitelik çoğaltma ve özetleme katmanlarına sahip olmasıdır. Bu özellikler yardımıyla mevcut veriyi daha iyi sınıflandırmakta ve tanımaktadır. Günümüzde KSA mimarileri kullanarak geliştirilmiş birçok popüler model bulunmaktadır. Bunlar; LeNet [24], AlexNet [25], GoogLeNet [26], VGGNet [27] ve ResNet [28] gibi modellerdir. Bu modeller KSA tabanlı uygulamalar geliştirerek görüntü işleme, sınıflama alanında büyük başarımlar elde etmişlerdir.

KSA katmanları gelişmekle birlikte temel olarak dört katmandan oluşmaktadır. Bu katmanlar; konvolüsyon katmanı, aktivasyon katmanı, alt-örnekleme (pooling) katmanı ve tam bağlı katmandır.



Şekil 4. Önerilen KSA modelinin blok gösterimi

Bu çalışmamızda önerdiğimiz bir KSA model ile günümüzde görüntü tanımada oldukça başarılı olan Vgg Net [27] ve Lenet [24] KSA modellerinin görüntüleri tanımadaki başarımları karşılaştırılmıştır.

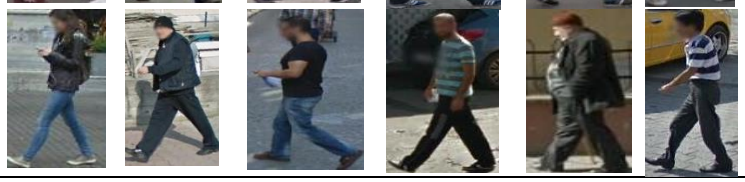
Dış ortam görüntülerindeki insan hareketlerinin otomatik tanınması için yapılan bu çalışmada önerilen KSA modeli Şekil 4'de verilmiştir. Önerilen derin KSA modelinde ilk katmanda ham görüntüler alınır, 3×3'lük birimler alınarak rastgele oluşturulan bir konvolüsyon katmanından geçirilir. ReLU katmanında bir aktivasyon fonksiyonu değerlendirilir ve ardından oluşan büyük verideki en anlamlı değerlerin bulunması için pooling katmanı ile veri sayısı azaltılmaktadır. Bu işlemlerden sonra tam-bağlı katmanı ile verilerin çaprazlanması sağlanır ve son olarak softmax katmanı ile verilerin sınıflarının tahmini sağlanmaktadır.

4. Deneysel Sonuçlar

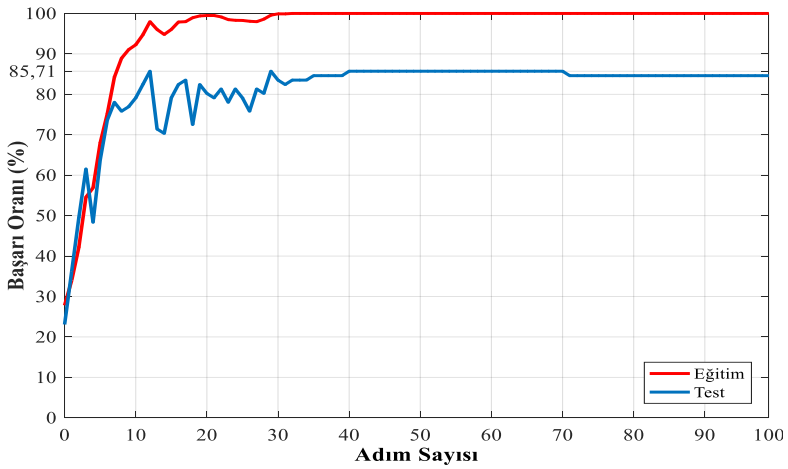
Google street view uygulamasından toplamda 1000 adet görüntü alınarak kapsamlı bir veri tabanı oluşturulmuştur. Bu görüntüler içerisindeki kişiler nesne tespit aracı olan YOLO ile tespit edilip bölümlenmiştir. Böylece 2000 adet kişi görüntüsü elde edilmiştir. Belirlenen yükseklik ve genişlik ölçülerine uygun olmayan görüntüler çıkarılıp tekrardan veritabanı güncellenmiştir. Güncellenen veritabanındaki görüntülerin ortalama genişlik ve yüksekliklerine göre üç farklı boyut için deneysel çalışmalar yapılmıştır. Bu veri kümeleri; yükseklik × genişlik × renk olmak üzere sırasıyla 192×72×3, 128×64×3 ve 96×36×3 şeklinde düzenlenmiştir.

Çalışmada kullanılan veri sınıfları ve sayıları ile ilgili bilgiler Tablo1'de verilmiştir. Çalışmada, 234 tane ayakta duran, 198 tane oturan, 249 tane sola yönelen ve 228 tane sağa yönelen olmak üzere toplamda 909 tane insan hareket verisi bulunmaktadır.

Tablo 1. Çalışmada kullanılan sınıflar ve bu sınıflara ait bazı örnek görüntüler

Sınıf	Görüntü Örnekleri	Veri Sayısı
S1- Oturan		198
S2- Ayakta duran		234
S3- Sağa yürüyen		228
S4-Sola yürüyen		249

Çalışmada, belli oranlarda eğitim, doğrulama ve test verileri alınarak ayrı sonuçlar elde edilmiştir. Görüntüler eğitim, doğrulama ve test olacak şekilde belli oranlarda üç parçaya bölünmüştür. Bu bölümleme yapıları eğitim-doğrulama ve test olmak üzere; %80-%10-%10, %70-%15-%15 ve %50-%25-%25 şeklinde seçilmiştir. Şekil 5'de, %80 eğitim, %10 doğrulama ve %10 test şeklinde bölünen veriseti üzerinde önerilen KSA modeli ile elde edilen eğitim ve doğrulama sonuçlarına ait performans grafiği gösterilmiştir. Derin KSA modelinin eğitimi 100 adım sürdürülmüştür. Grafikten görülebileceği gibi modelin eğitimi 30. adımda tamamlanırken doğrulama aşaması tam olarak başarıya ulaşamayıp 80. adımda modelin başarımı %85.71 olarak ölçülmüştür.



Şekil 5. KSA tabanlı modelimizin başarım sonuçları

Önerilen modelde elde edilen sonuçlar, diğer modellerde elde edilen sonuçlar ile karşılaştırılmıştır. Önerilen modelin başarısının daha yüksek olduğu tespit edilmiştir. Eğitim,

doğrulama ve test verileri oranı değiştirilerek en iyi sonuçların elde edilmesi sağlanmıştır. Tablo 2' de, deneysel çalışma için farklı veri bölümleme oranları ile elde edilen sınıflandırma sonuçları verilmiştir.

Tablo 2. Deneysel çalışmadan elde edilen sonuçların karşılaştırılması

Görüntü boyutu	Modeller	Sınıflandırma başarımı (%)		
		Veri setleri Eğitim – Doğrulama – Test (%)		
		80 – 10 – 10	70 – 15 – 15	50 – 25 – 25
192 × 72 × 3	**Önerilen KSA**	**85.71**	**78.10**	**75.00**
	Vgg Net	83.52	77.37	69.74
	Lenet	82.42	67.15	71.05
128 × 64 × 3	**Önerilen KSA**	**71.43**	74.45	68.86
	Vgg Net	70.33	**75.94**	**69.30**
	Lenet	71.43	74.45	67.54
96 × 36 × 3	**Önerilen KSA**	**72.53**	**77.37**	69.30
	Vgg Net	72.53	74.45	**71.05**
	Lenet	64.84	73.72	70.8

Tablo 2’ de görüldüğü gibi dört sınıflı veriler üzerinde önerilen KSA modeli diğer modellere kıyasla daha iyi sonuçlar elde etmiştir. Önerilen KSA modelinde %80 eğitim, %10 doğrulama ve %10 test veri seti üzerinde %85.71 ile en iyi sonuç gözlemlenmiştir. Diğer görüntü boyutları ve veri setlerinde önerilen KSA modeli, diğer modellere göre daha yüksek sonuç vermiştir. Vgg Net modeli görüntü boyutunun 128×64×3 olduğu ve eğitim, doğrulama ve test oranlarının sırasıyla %70, %15, %15 ve %50, %25, %25 olarak ayarlandığı verilerde ağırlıklarını daha uygun düzenleyebildiğini göstermiştir.

Eğitilen KSA modelinin test verileri üzerinde elde ettiği sınıflama sonuçlarına ait bir doğruluk matrisi verilmiştir. Tablo 3’ te önerilen KSA modelinin 4 sınıfa ait test verileri üzerinde elde ettiği doğruluk matrisi verilmiştir.

Tablo 3. KSA modelinin test verileri üzerinde elde ettiği doğruluk matrisi (%80-%10-%10)

Sınıflar	**S1**	**S2**	**S3**	**S4**
S1	**20**	2	3	0
S2	3	**21**	1	0
S3	2	1	**16**	1
S4	0	0	0	**21**
Ortalama Tanıma Performansı=%85.71				

Doğruluk matrisinden görüldüğü gibi, ayakta duran sınıfa atanması gereken imge sayısının 25 olması gerekirken 2 imge oturan sınıfa, 3 imge de sola yürüyen sınıfa yanlış olarak atılarak ayakta duran sınıf için 20 imge doğru tespit edilmiştir. Oturan sınıfa atanması gereken imge sayısının 25 olması gerekirken 3 imge ayakta duran sınıfa, 1 imge de sola yürüyen sınıfa yanlış olarak atılarak oturan sınıfa 21 imge atanmıştır. Sola yürüyen sınıfa atanması gereken imge sayısının 20 olması gerekirken yanlış bir şekilde imge ataması yapılarak imgelerin iki tanesi ayakta duran sınıfa, bir tanesi oturan sınıfa, bir tanesi de sağa giden sınıfa atanarak sola yürüyen sınıfa 16 imge atanmıştır. Sağa yürüyen sınıf için herhangi bir yanlış atama yapılmayıp 21 imge doğru olarak sınıflandırılmıştır.

5. Sonuçlar

Bu çalışmada iki aşamalı derin öğrenme tabanlı bir insan hareketi tanıma sistemi önerilmiştir. Daha önceden eğitilmiş nesne tespit aracı olan YOLO ile kişiler tespit edilmiş ve derin öğrenme mimarilerinden KSA kullanılarak kişi hareketlerinin sınıflandırılması sağlanmıştır. Geliştirilen hibrit derin öğrenme yöntemi ile insan hareketlerini içeren görüntülerin otomatik tanınması işlemi yapılmıştır. Çalışmada geliştirilen KSA modeli ile popüler derin öğrenme modelleri karşılaştırılmış ve önerilen yöntemin daha yüksek doğruluk oranları ile sonuçlar ürettiği görülmüştür.

Önerilen modelinin yeterli derecede eğitim ve parametre düzenlemesinden sonra hareket tespit sistemleri, insan takip sistemleri, otonom araçlar için sürücü destek sistemleri gibi birçok alanda kullanılması sağlanabilir.

6. Kaynaklar

1. Dhamsania, C. J., & Ratanpara, T. V. (2016). A survey on Human action recognition from videos. Online International Conference on In Green Engineering and Technologies, 1-5.
2. Geronimo, D., Lopez, A. M., Sappa, A. D., & Graf, T. (2010). Survey of pedestrian detection for advanced driver assistance systems. IEEE transactions on pattern analysis and machine intelligence, **32**(7): 1239-1258.
3. Ouyang, W., Zeng, X., & Wang, X. (2016). Learning mutual visibility relationship for pedestrian detection with a deep model. International Journal of Computer Vision, **120**(1): 14-27.
4. Aguilar, W. G., Luna, M. A., Moya, J. F., Abad, V., Parra, H., & Ruiz, H. (2017). Pedestrian detection for UAVs using cascade classifiers with meanshift. In Semantic Computing (ICSC), 509-514.
5. Thomas B. Moeslund, Adrian Hilton, Volker Krüger, (2006). A survey of advances in vision-based human motion capture and analysis, Computer Vision and Image Understanding, **104** (2–3) 90–126.
6. Cheng, G., Wan, Y., Saudagar, A. N., Namuduri, K., & Buckles, B. P. (2015). Advances in human action recognition: A survey. arXiv preprint arXiv:1501.05964.
7. Aggarwal, J. K., & Ryoo, M. S. (2011). Human activity analysis: A review. ACM Computing Surveys, **43**(3): 16.
8. Chaudhary, A., Raheja, J. L., Das, K., & Raheja, S. (2011). A survey on hand gesture recognition in context of soft computing. In International Conference on Computer Science and Information Technology **133**: 46-55.
9. Weinland, D., Ronfard, R., Boyer, E., (2011). A survey of vision-based methods for action representation, segmentation and recognition. Computer Vision and Image Understanding **115**: 224–241.
10. Yu, G., & Yuan, J. (2015). Fast action proposals for human action detection and search. In Proceedings of the IEEE conference on computer vision and pattern recognition, 1302-1311.
11. Baccouche, M., Mamalet, F., Wolf, C., Garcia, C., & Baskurt, A. (2011). Sequential deep learning for human action recognition. In International Workshop on Human Behavior Understanding, **7065**: 29-39.
12. Eroglu, H., Gokce, C. O., & Ilk, H. G. (2014, April). Human action recognition via spatial and temporal methods. In Signal Processing and Communications Applications Conference (SIU), 2014 22nd, 104-107.
13. Niebles, J. C., Wang, H., & Fei-Fei, L. (2008). Unsupervised learning of human action categories using spatial-temporal words. International journal of computer vision, **79**(3): 299-318.
14. Schuldt, C., Laptev, I., & Caputo, B. (2004). Recognizing human actions: a local SVM approach. In Pattern Recognition, **3**: 32-36.
15. Ji, S., Xu, W., Yang, M., & Yu, K. (2013). 3D convolutional neural networks for human action recognition. IEEE transactions on pattern analysis and machine intelligence, **35**(1): 221-231.
16. Rahmani, H., Mian, A., & Shah, M. (2017). Learning a deep model for human action recognition from novel viewpoints. IEEE Transactions On Pattern Analysis and Machine Intelligence, **40**(3), 667-681.
17. Simonyan, K., & Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. In Advances In Neural Information Processing Systems, **27**: 68-576.
18. İnternet Erişimi (2018): http://fortune.com/ai-artificial-intelligence-deep-machine-learning/
19. Coşkun, M., Yildirim, Ö., Uçar, A. & Demir, Y. (2017). An Overview of Popular Deep Learning Methods. European Journal of Technique, **7**(2): 165-176.
20. Gehring, J., Miao, Y., Metze, F., & Waibel, A. (2013). Extracting deep bottleneck features using stacked auto-encoders. In Acoustics, Speech and Signal Processing (ICASSP), 3377-3381.
21. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779-788.
22. Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. arXiv preprint.
23. Koçer, S., & Akdağ, A. (2017). Convolutional neural network based handgun detection. In Computer Science and Engineering (UBMK), 89-93.
24. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11): 2278-2324.
25. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems **25**: 1097-1105.
26. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. Computer Vision and Pattern Recognition, 1-9.
27. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

28. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Computer Vision and Pattern Recognition, 770-778.

29. Yildirim Ö. (2018). A novel wavelet sequences based on deep bidirectional LSTM network model for ECG signal classification; Computers in Biology and Medicine; 2018; **96**: 189 – 202.

30. Yildirim, Ö., San Tan, R., & Acharya, U. R. (2018). An efficient compression of ECG signals using deep convolutional autoencoders. Cognitive Systems Research, **52**:198-211.

31. Uçar, A., Demir, Y., & Güzeliş, C. (2017). Object recognition and detection with deep learning for autonomous driving applications. Simulation, **93**(9), 759-769.