



Geliş/Received : 04.08.2018 & Kabul/Accepted : 31.10.2018 & Yayınlanma/Published (online) : 23.12.2018

## Derin Öğrenme ile İnsansız Hava Aracı Görüntülerinden Yaya Tespiti

Suat TORAMAN

Fırat Üniversitesi, Enformatik Bölümü, Elazığ, Türkiye

### Özet

Bu çalışmada, insansız hava araçlarından (İHA) elde edilen görüntüler kullanılarak yaya tespitine yönelik bir uygulama gerçekleştirilmiştir. Bunun için, elde edilen İHA görüntülerinden, derin öğrenme yöntemi yardımıyla özellik çıkarımı yapılmıştır. İHA'lerden alınan görüntülerin işlenmesinde karşılaşılan zorluklardan biride, büyük veri kümelerinin sınıflandırmasıdır. Bu çalışmada, bu zorluğun üstesinden gelmek için Evrişimsel Sinir Ağları (ESA) kullanılmıştır. Bir diğer zorluk ise bazı veri türlerinin azlığından dolayı kaliteli bir eğitim sürecinin gerçekleştirilememesidir. Bu nedenle, eğitimin etkinliğini artırabilmek için resim çoğaltma yöntemi uygulanmıştır. Önerilen yöntem ile İHA'dan elde edilen yaya, bisikletli, araba, ağaç ve sokak lambası resimleri istenen boyutlarda ayarlanarak ESA modellerinden AlexNet ve VGG16'ya giriş verisi olarak verilerek özellik çıkarımı gerçekleştirilmiştir. Çıkarılan özellikler Destek Vektör Makinesi (DVM) ile sınıflandırılmıştır. Sınıflandırma işlemi sayesinde hem yaya ile diğer öğelerin ayrımı gerçekleştirilirken hem de AlexNet ile VGG16'nın performansları karşılaştırılmıştır.

**Anahtar Kelimeler:** Yaya bulma, İHA, Evrişimsel sinir ağları, Sınıflandırma

## Pedestrian Detection with Deep Learning from Unmanned Aerial Imagery

### Abstract

In this study, an application for pedestrian detection was carried out using images obtained from Unmanned Aerial Vehicles (UAV). For this purpose, the feature extraction was performed from the UAV imagery by using the deep learning method. One of the challenges encountered in processing images from UAV is the classification of large datasets. In this study, Convolution Neural Networks (CNN) have been used to overcome this challenge. Another difficulty is that a quality training process can not be achieved due to the lack of some data types. For this reason, image augmentation has been applied to increase the efficiency of education. With the proposed method, images of pedestrians, bicycles, cars, trees and street lamps obtained from the UAV were adjusted to the required dimensions. Then, these images were given as input data to AlexNet and VGG16, and feature extraction was performed. The extracted features were classified by Support Vector Machine (SVM). With the classification process, while both pedestrian detection was carried out and the performances of AlexNet and VGG16 were compared.

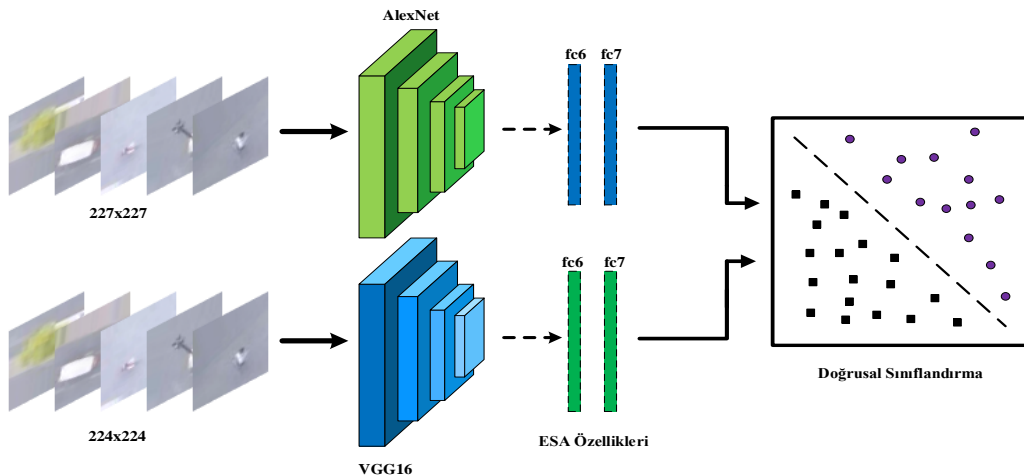
**Keywords:** Pedestrian detection, UAV, Convolutional neural networks, Classification

## 1. Giriş

Karışık, yoğun ortamlarda, hareketli nesnelere veya hareketli platformlarda nesne belirleme, nesne takibi gibi konular hala önemli zorluklar içermektedir. Son zamanlarda İHA'lar ve otonom araç sistemlerine olan ilginin artmasıyla bu alandaki çalışmalarda oldukça fazlaşmıştır. Özellikle gerçek zamanlı nesne takibi ve sınıflandırılması gibi konularda derin öğrenmenin başarılı olması bu ilginin derin öğrenme tarafına doğru artmasındaki etkenlerdendir. Ayrıca son birkaç yıl içinde, derin ESA'ların yüksek doğruluk ve hızları, onların nesne tespiti ve nesnelerin sınıflandırılması gibi uygulamalar için oldukça başarılı yaklaşımlar olduğu gösterilmiştir [1,2]. Fakat donanımsal anlamda GPU'ların derin öğrenmede çok büyük etkileri olmasına rağmen gerçek görüntü işleme uygulamasını her alanda uygulamak kolay değildir. İHA'lardan alınan büyük ölçekli görüntülerin işlenmesi için çeşitli çevrim dışı sistemler kullanılmaktadır. Ekin sınıflandırma, tütün belirleme, ağaç sayma ve sınıflandırma gibi uygulamalar örnek olarak verilebilir [3–6]. İHA'lar sadece görüntünün işlenmesi amaç ile de kullanılmamaktadır. Bunun dışında arama kurtarma, trafik akışını denetleme, kişi takibi gibi görüntü aktarım işlerinde de etkin bir şekilde kullanılmaktadır [7–10].

Bu çalışmada, son zamanlarda yapılan uygulamalarda yoğun bir şekilde kullanılan derin öğrenme modellerinden olan AlexNet ve VGG16 kullanılmıştır. Bu modeller, farklı çalışma gruplarının geliştirdiği ESA modelleridir. 2012

yılındaki ImageNet yarışmasını kazanan, milyonlarca resim ile eğitilmiş AlexNet, ile 2014 yılındaki yarışmada ikinci olan VGG16, 1000 farklı sınıflandırma yapacak şekilde tasarlanmış ESA modelleridir [11, 12]. Sınıflandırma yapılacak her uygulamada veya tasarlanacak her modelde milyonlarca veri setini elde etmek veya bu kadar büyük bir veri kümesini işlemek mümkün olmayabilir. Bu durumda önceden eğitilmiş bir modeli kullanmak daha uygun olacaktır. Bu durumu ifade eden yapıya ise Transfer Öğrenme denilmektedir [13]. Oluşturulan ağı eğitilmesi için yeteri veri yoksa önceden eğitilmiş ESA'ları kullanarak veri kümesinden özellik vektörü çıkarılabilir. Bu çalışmada da, önceden eğitilmiş ESA modellerinden olan AlexNet ve VGG16 mimarisi kullanılarak, İHA'lardan elde edilen görüntülerin sınıflandırıldığı bir yöntem önerilmiştir. İHA görüntüleri AlexNet ve VGG16 mimarisine uygun boyuta dönüştürülmüş ve daha sonra resim çoğaltma işlemi ile veri kümesi arttırılmıştır. AlexNet ve VGG16'ya giriş olarak verilen resimlerden 4096 boyutlu özellik vektörleri elde edilmiştir. Elde edilen özellik vektörleri ayrı ayrı DVM ile sınıflandırılmıştır. Bu çalışmanın amacı, İHA görüntülerini kullanarak transfer öğrenme yöntemi yardımıyla, yaya tanıma uygulaması gerçekleştirmek ve kullanılan ESA modellerinin performanslarını karşılaştırmaktır. Ayrıca bu çalışmanın diğer bir amacı, transfer öğrenme yönteminin ve resim çoğaltma işleminin nesne tanımadaki etkinliğini incelemektir. Çalışmanın akış diyagramı Şekil 1'de gösterilmiştir.



Şekil 1. Çalışmanın akış diyagramı

## 2. Materyal Metot

Veri seti oluşturmak için Stanford Üniversitesi drone veri setinden alınan video görüntüleri kullanılmıştır. Video görüntüleri Stanford Üniversitesi kampüsündeki belirli noktalarda ve belirli sürelerde alınan görüntüleri kapsamaktadır [14, 15]. Çalışmada ise, bu video görüntülerinden belirli nesnelerin resimleri alınmıştır. Alınan resimler yaya, bisikletli, araba, ağaç ve sokak lambalarından oluşmaktadır. Her bir resim kümesi

için belirli sayıda örnek alınmıştır. Derin öğrenmede kaliteli bir eğitim için çok sayıda veri kullanımı oldukça önemlidir. Bu nedenle, veri çoğaltma işlemi ile eğitim veri seti artırılmıştır. Böylece her bir gruptan yaklaşık 408 resim elde edilecek şekilde bir resim çoğaltma işlemi gerçekleştirilmiştir. Sonuç olarak 2040 resimden oluşan bir veri seti oluşturulmuştur. Video görüntülerinden elde edilen ve resim çoğaltma işlemi uygulanan resim örnekleri Şekil 2’de gösterilmiştir.



**Şekil 2.** Resim çoğaltma ile elde edilen resim örnekleri (yukarıdan aşağı; ağaç, araba, sokak lambası, bisikletli, yaya)

### 2.1. Önceden eğitilmiş ESA Modeli ile Özellik Çıkarma ve Sınıflandırma

Milyonlarca resim kullanılarak eğitilmiş çeşitli derin öğrenme modelleri bulunmaktadır. GoogleNet, VGG, AlexNet, ResNet bunlara verilebilecek örneklerden bazılarıdır. Bu derin öğrenme yöntemlerinin her biri farklı model kullanarak eğitim işlemi gerçekleştirmektedir. Bu çalışmada kullanılan derin öğrenme modelleri AlexNet ve VGG16’dır. AlexNet, bir milyondan fazla resim kullanılarak eğitilmiş bir modeldir. Ayrıca bu model ile resimler 1000 farklı kategoriye ayrılabilir. AlexNet, 8 katmanlı bir yapıdan

oluşan bir model sunmaktadır. AlexNet modelinde beş konvolüsyon katmanı (convolution layer) ve 3 tam bağlı katman (fully connected layer) bulunmaktadır. Konvolüsyon katmanlarında farklı çekirdek boyutları (11x11, 5x5, 3x3) kullanılmaktadır. Konvolüsyon katmanlarından sonra maksimum havuz katmanları (max-pooling layer) yer almaktadır [11]. VGG16 modeli de milyonlarca resim ile eğitilmiştir. AlexNet gibi 1000 farklı sınıfı ayırtabilir. VGG16’da 5 konvolüsyon katmanına sahiptir. Fakat VGG16’da, AlexNet gibi değişken çekirdek boyutları yerine sabit 3x3’lük yapı kullanılmıştır [12]. Her iki modelde

özellik çıkarımı için fc6 ve fc7 katmanları kullanılmıştır. Her bir resim fc6 ve fc7 katmanlarının çıkışı olan 4096 boyutunda bir özellik vektörü ile temsil edilmektedir [16]. Sınıflandırma işlemi için bir makine öğrenme algoritması olan ve risk minimizasyonu mantığı ile çalışan Destek Vektör Makinesi ile iki sınıflı bir kümeyi ayırabilecek en uygun düzlemin bulunması hedeflenmektedir [17]. Bu amaçla, Liblinear kütüphaneli DVM sınıflandırıcısı kullanılmıştır. Homojen haritalama özelliğine sahip Liblinear kütüphaneli DVM sınıflandırıcısı kullanılmasının nedeni, küçük miktarlardaki eğitim verisi için etkili bir sınıflandırma gerçekleştirebilmesi söylenebilir [18, 19]. Liblinear kütüphaneli DVM'nin parametrelerinden  $C$  [ $10^{-4}, \dots, 10^3$ ] aralığında incelenmiştir.

### 3. Deneysel Sonuçlar ve Tartışma

Intel Core i5-4200 CPU ve 8 GB bellek sahip bilgisayar ile yapılan uygulamadaki işlemler için MATLAB kullanılmıştır. Resimlerin AlexNet'e giriş olarak verilebilmesi için 227x227, VGG16'ya giriş olarak verilebilmesi için 224x224 boyutlu hale dönüştürülmesi gerekmektedir. Yeniden boyutlandırılan veri kümesindeki her bir resimden AlexNet ve VGG16 modelinin fc6- fc7 katman çıkışından derin spektrum özellikleri elde edilmiştir. Fc6 ve fc7 katmanları çıkış olarak 4096 boyutlu bir özellik vektörü üretmektedir. Üretilen özellik vektörleri DVM sınıflandırıcısı ile sınıflandırılmıştır. Tablo 1'de görüldüğü gibi AlexNet'te fc6 ile %99.78, fc7 ile %99.84 doğruluk elde edilirken, VGG16'da fc6 ile %99.88, fc7 ile %99.78 oranında sınıflama doğruluğu elde edilmiştir. Fc6 ve fc7 katmanlarının çıkışından elde edilen doğrulukların benzer olması, özellik çıkarımının önceki katmanlarda etkili bir şekilde yapıldığını göstermektedir. AlexNet ile VGG16 modelleri arasında oluşan küçük fark ise gerçekleştirilen eğitim süreci ile açıklanabilir. Ayrıca derin özellik vektörlerinin DVM ile sınıflandırılması öncesinde veri seti çapraz doğrulama yöntemi ile 4 eşit parçaya bölünmüştür. Çapraz doğrulama öncesi veri seçimi rastgele yapılmıştır. Bu işlem 10 kez uygulanmıştır. Tablo 1'deki değerler, 10 kez tekrarlanan sınıflandırma işleminin ortalama değerlerini göstermektedir.

**Tablo 1.** AlexNet ve VGG16 ESA modellerinin yaya tespiti sınıflandırma sonuçları

ESA modeli	Fc6 (Ort. Doğruluk+ Std)	Fc7 (Ort. Doğruluk + Std)
AlexNet	%99.78 ± 0.234	%99.84 ± 0.241
VGG16	%99.88 ± 0.189	%99.78 ± 0.171

Masanori ve ark yaptığı çalışmada [20], ağaç türlerini sınıflandırma işlemi gerçekleştirmişlerdir. İHA'dan alınan görüntüleri GoogleNet mimarisini kullanarak bir ayırım işlemi gerçekleştirmişlerdir. Bu çalışmada da transfer öğrenme yönteminden faydalanılmıştır. Sonuç olarak 7 ağaç türünü ortalama %89 ile sınıflandırmışlardır. Diğer birçok çalışmada [10, 21–25], genellikle gerçek zamanlı görüntüler veya yüksek çözünürlüklü uydu görüntüleri sınıflandırılmaktadır. Bu işlemlerde çoğunlukla hızlı ve etkili bir eğitim gerçekleştirmek için GPU destekli donanım kullanılmaktadır. Mevcut çalışmada, CPU destekli bir uygulama gerçekleştirilmiştir. Bu durum derin öğrenme gibi büyük veri kümeleri ile yapılan işlemlerde hız ve işlem zamanı noktasında oldukça önemlidir.

### 4. Sonuç ve Öneriler

Bu çalışmada, İHA'lardan elde edilen çeşitli görüntülerden yayaların belirlenmesine yönelik bir çalışma gerçekleştirilmiştir. Çalışma;

- Video görüntülerinden yaya, bisikletli, ağaç, araba ve sokak lambası resimlerinin elde edilmesi
- Elde edilen resimlerin giriş vektörü olarak AlexNet ve VGG16'ya uygun hale getirilmesi için yeniden boyutlandırması
- Yeniden boyutlandırılan resimlerin resim çoğaltma yöntemi ile artırılması
- Çoğaltılan resimlerin AlexNet ve VGG16 ile özellik vektörlerinin elde edilmesi
- Son olarak özellik vektörlerinin DVM sınıflandırıcısı ile sınıflandırılması aşamalarından oluşmaktadır.

Ayrıca çalışma ile farklı ESA modellerinin performansları karşılaştırılmış ve transfer öğrenmenin farklı türdeki verileri ayırt edebilmedeki yeteneği gösterilmiştir. Çalışmada uygulanan diğer bir yöntem ise resim çoğaltmadır.



Eğitim verisinin yeterli olmadığı durumlarda kullanılabilir faydalı bir özellik olduğu vurgulanmıştır. İleri dönük olarak yapılacak çalışmalarda, farklı ESA modellerinin incelenmesi ve karşılaştırması düşünülmektedir. Ayrıca, daha fazla işlem gücüne sahip donanım alt yapısı ile büyük ve karmaşık veri setleri kullanılarak mevcut çalışmadaki sınırlılıkların giderilmesi ve farklı gerçek zamanlı (yaya takibi, bisikletli takibi, nesne sınıflandırma) uygulamalarında yapılması planlanmaktadır.

### Kaynaklar

- [1] Radovic M, Adarkwa O, Wang Q (2017) Object Recognition in Aerial Images Using Convolutional Neural Networks. *J Imaging* 3:21
- [2] Li, L., Fan, Y., Huang, X., & Tian L (2016) Real-time UAV weed scout for selective weed control by adaptive robust control and machine learning algorithm. *Am Soc Agric Biol Eng Annu Int Meet ASABE*
- [3] Hung C, Xu Z, Sukkarieh S (2014) Feature learning based approach for weed classification using high resolution aerial images from a digital camera mounted on a UAV. *Remote Sens* 6:12037–12054
- [4] Zarjam P, Epps J, Chen F, Lovell NH (2013) Estimating cognitive workload using wavelet entropy-based features during an arithmetic task. *Comput Biol Med* 43:2186–1295
- [5] Chen SW, Shivakumar SS, Dcunha S, et al (2017) Counting Apples and Oranges With Deep Learning: A Data-Driven Approach. *IEEE Robot Autom Lett* 2:781–788
- [6] Li W, Fu H, Yu L, Cracknell A (2016) Deep Learning Based Oil Palm Tree Detection and Counting for High-Resolution Remote Sensing Images. *Remote Sens* 9:22
- [7] Kim NV, Chervonenkis MA (2015) Situation control of unmanned aerial vehicles for road traffic monitoring. *Mod Appl Sci* 9:1–13
- [8] Bejiga MB, Zeggada A, Nouffidj A, Melgani F (2017) A convolutional neural network approach for assisting avalanche search and rescue operations with UAV imagery. *Remote Sens* 9:
- [9] Kalal Z, Mikolajczyk K, Matas J (2011) Tracking-Learning-Detection. *IEEE Trans Pattern Anal Mach Intell* 34:1409–1422
- [10] De Smedt F, Hulens D, Goedeme T (2015) On-board real-time tracking of pedestrians on a UAV. *IEEE Comput Soc Conf Comput Vis Pattern Recognit Work* 2015–Octob:1–8
- [11] Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Adv Neural Inf Process Syst* 1–9
- [12] Simonyan K, Zisserman A (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. 1–14
- [13] Kaya A, Keçeli AS, Can AB Akciğer nodül özelliklerinin tahmininde çeşitli sınıflama stratejilerinin incelenmesi. *Gazi Üniversitesi Mimar Mühendislik Fakültesi Derg* (2018), <https://doi.org/10.17341/gazimmfd416530>
- [14] A. Robicquet, A. Sadeghian, A. Alahi SS (2016) Learning Social Etiquette: Human Trajectory Prediction In Crowded Scenes. In: *European Conference on Computer Vision (ECCV)*
- [15] (2016) Stanford Drone Dataset. [http://cvgl.stanford.edu/projects/uav\\_data/](http://cvgl.stanford.edu/projects/uav_data/). Accessed 20 Jul 2018
- [16] Zhou Y, Nejati H, Do T-T, et al (2016) Image-based Vehicle Analysis using Deep Neural Network: A Systematic Study
- [17] Khazae A, Ebrahimzadeh A (2010) Classification of electrocardiogram signals with support vector machines and genetic algorithms using power spectral features. *Biomed Signal Process Control* 5:252–263
- [18] Vedaldi A, Zisserman A (2010) Efficient Additive Kernels via Explicit Feature Maps. *Proc {CVPR}* 34:480–492
- [19] Fan R-E, Chang K-W, Hsieh C-J, et al (2008) LIBLINEAR: A Library for Large Linear Classification. *J Mach Learn Res* 9:1871–1874
- [20] Onishi M, Ise T (2018) Automatic classification of trees using a UAV onboard camera and deep learning

- [21] Sheppard C, Rahnemoonfar M Real - time Scene Understanding for UAV Imagery based on Deep Convolutional Neural Networks. 5–8
- [22] Aguilar WG, Luna MA, Moya JF, et al (2017) Pedestrian Detection for UAVs Using Cascade Classifiers with Meanshift. Proc - IEEE 11th Int Conf Semant Comput ICSC 2017 509–514
- [23] Ma Y, Wu X, Yu G, et al (2016) Pedestrian Detection and Tracking from Low-Resolution Unmanned Aerial Vehicle Thermal Imagery. Sensors 16:446
- [24] Xu Y, Yu G, Wang Y, et al (2017) Car detection from low-altitude UAV imagery with the faster R-CNN. J Adv Transp 2017:
- [25] Ammour N, Alhichri H, Bazi Y, et al (2017) Deep learning approach for car detection in UAV imagery. Remote Sens 9: