

## Robotik Uygulamalar İçin Derin Öğrenme Tabanlı Nesne Tespiti ve Sınıflandırması

Ferdi ÖZBİLGİN<sup>1\*</sup>, Cengiz TEPE<sup>2</sup>

<sup>1</sup>Giresun Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Giresun, Türkiye

<sup>2</sup>Ondokuz Mayıs Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Samsun, Türkiye

Geliş Tarihi: 08.05.2020

Kabul Tarihi: 08.06.2020

\*Sorumlu Yazar: ferdi.ozbilgin@giresun.edu.tr

### Öz

Görüntü içerisindeki nesnelerin tespit edilmesi ve sınıflandırılma uygulamaları her geçen gün artmaktadır. Bu çalışmada da robotik uygulamalarda da kullanılabilir bir nesne tespiti ve sınıflandırılması uygulaması gerçekleştirilmiştir. Alexnet Evrimsel Sinir Ağları (ESA) mimarisi ve Bölgesel Evrimsel Sinir Ağları (B-ESA) algoritması ile gerçekleştirilen çalışmada yedi farklı nesne sınıfı seçilmiştir. Veri setindeki 684 eğitim verisi etiketlenerek ağı eğitilmesinde kullanılmıştır. 226 test görüntüsü eğitilen ağda test edilmesi sonucunda her sınıfa ait doğru tahmin değerleri ve toplam doğruluk değerleri bulunmuştur. Sınıflara ait tahminlerde en düşük %85,74 ve en yüksek %100 değerlerine ulaşılmıştır. Tüm test verileri için doğruluk değeri %93,81 bulunmuştur.

**Anahtar Kelimeler:** Nesne tespiti, Sınıflandırma, Derin öğrenme.

## Deep Learning Based Object Detection and Classification for Robotic Applications

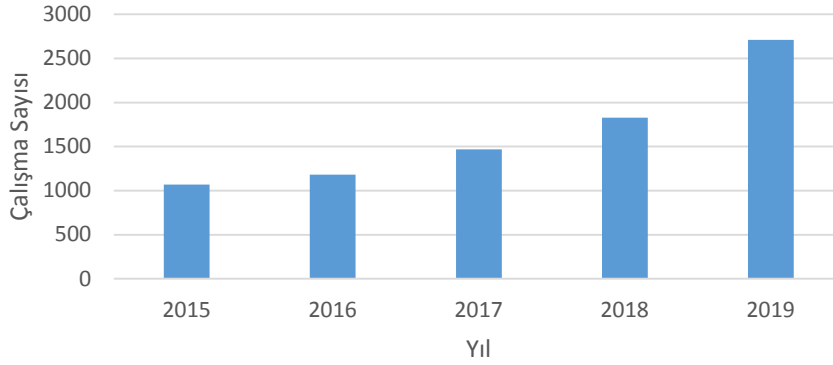
### Abstract

The detection and classification applications of the objects in the image are increasing day by day. In this study, an object detection and classification application, which can also be used in robotic applications, has been realized. Seven different object classes were selected in the study conducted with Alexnet Evolutionary Neural Networks (CNN) architecture and Regional Convolutional Neural Networks (R-CNN) algorithm. 684 training data in the data set were labelled and used to train the network. As a result of testing 226 test images in the trained network, correct predictive values and total accuracy values of each class were found. The lowest estimates of 85.74% and the highest 100% were reached in the estimates of the classes. The accuracy value was 93.81% for all test data.

**Keywords:** Object detection, Classification, Deep learning.

## 1. Giriş

Popülerliği sürekli artmakta olan nesne tespiti ve sınıflandırılması, dijital görüntü ve videolardan insanlar, binalar ve arabalar gibi nesnelere algılayan ve tanımlayan bir bilgisayar teknolojisi (Liu ve ark., 2020). Nesne tespitinin amacı bilgisayarla görme uygulamalarının ihtiyaç duyduğu en temel bilgi parçalarından birini sağlayan hesaplama tekniklerini geliştirmektir (Zou ve ark., 2019). Nesne algılama birçok bilgisayarla görme görevinin temelini oluşturmakta ve birçok alanda da kullanılmaktadır. Görüntü bölütleme (Hariharan ve ark., 2014; Khoreva ve ark., 2019; Voigtlaender ve ark., 2019), görüntüyü metin olarak ifade etme (Karpathy ve Fei-Fei, 2015) ve nesne takibi (Tang ve ark., 2019; Wang ve ark., 2019) bu alanlar arasında ifade edilebilir. Şekil 1’de son beş yıldaki nesne tespiti ile alakalı çalışma sayıları verilmekte ve bu sayılarda her yıl artış olduğu anlaşılmaktadır. Bu durum da nesne tespitinin giderek popüler hale geldiğinin bir göstergesidir.



**Şekil 1.** Son beş yıldaki yıllara göre nesne tespiti ile alakalı çalışma sayıları (Veriler Google Akademik Gelişmiş arama bölümünden “object detection” ismi yer alan çalışmalar aratılarak elde edilmiştir.)

Bir nesneyi algılama, konumunu tespit etme ve sınıflandırılması robotik uygulamalarda da sıklıkla kullanılmaktadır. Robotlar çevrelerindeki nesnelere tespit edip sınıflandırmasıyla birlikte daha otonom hale gelmektedirler. Üzerlerine bağlanacak bir kamera ile farklı sensörlere ihtiyaç olmadan nesne hakkında daha detaylı bilgi elde edilebilecektir (Çoban et al., 2019).

Nesne algılama çalışmaları ViolaJones Detektörü, HOG Detektörü ve Deforme Edilebilir Parça Tabanlı Model gibi geleneksel yöntemler ile yapıldığı gibi 2012 yılından itibaren Bölgesel Evrişimsel Sinir Ağları (B-ESA), SPPNet ve Hızlı B-ESA gibi derin öğrenme metotları ile de yapılmaktadır (Zou ve ark., 2019). Derin öğrenme çalışmaları arttıkça Evrişimsel Sinir Ağları (ESA) performansını da artırmaktadır. ESA mimarileri, nöron adı verilen ağırlık ve bias parametreleri sayesinde oldukça fazla şey öğrenebilen yapılardır. Bu yapılara Alexnet, ZF Net, VGG Net, Googlenet, Microsoft Resnet ve Dense Net ESA mimarileri yöntem olarak örnek verilebilir.

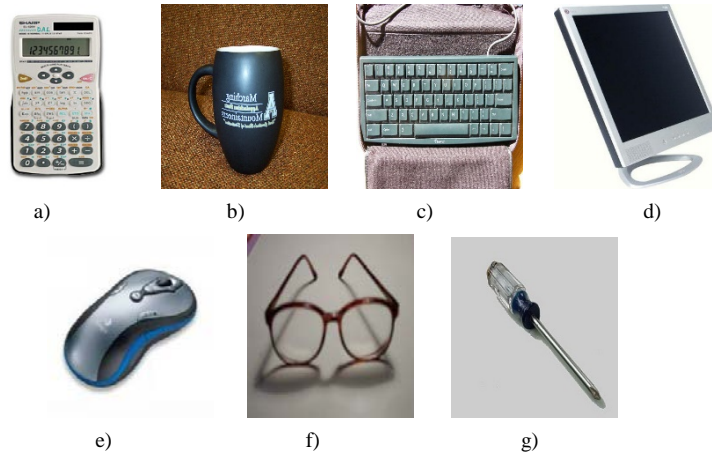
Literatürde nesne tespiti ve sınıflandırılması ile ilgili birçok çalışma mevcuttur. Alexnet mimarisi kullanarak çöp tespiti (Lee ve ark., 2019), elektrik panosu üzerindeki malzemelerin tespiti (Lv ve Tan, 2019), araç tespiti (Espinosa ve ark., 2017), tenis topu tespiti (Gu ve ark., 2017), yaya algılama (Tomè ve ark., 2016) çalışmaları ve buna benzer çalışmalar gerçekleştirilmiştir. Diğer ESA mimarileri kullanılarak yapılan çalışmalar (Fu ve ark., 2017; Haque ve ark., 2019; Huang ve ark., 2019; Kang ve ark., 2017) mevcut olduğu gibi robotik uygulamalarda da nesne tespiti ile ilgili çalışmalar yapılmaktadır (Do ve ark., 2018; Li, 2017; Lu ve ark., 2017).

Bu çalışmada robotik uygulamalarda da kullanılacak Alexnet ESA mimarisi ile 7 farklı sınıfa ait görüntüler RCNN öğrenme metodu ile MATLAB programlama dili kullanılarak eğitilmiştir. İnternet ortamından temin edilen görseller test edilerek sistem başarısı belirlenmiştir.

## 2. Materyal ve Metot

### 2.1. Veri Seti

Çalışmada Caltech 256 Nesne Kategorisi veri seti (Griffin ve ark., 2007) kullanılmıştır. Toplamda 256 farklı nesne bulunan veri seti 30607 görüntü içermektedir. Nesne algılamak için veri seti içerisinde yedi farklı sınıfa ait nesne seçilmiştir. Bu nesnelere ait görsel Şekil 2’de verilmiştir. Yedi farklı sınıfa ait toplam 684 görüntü çalışmada kullanılmıştır.



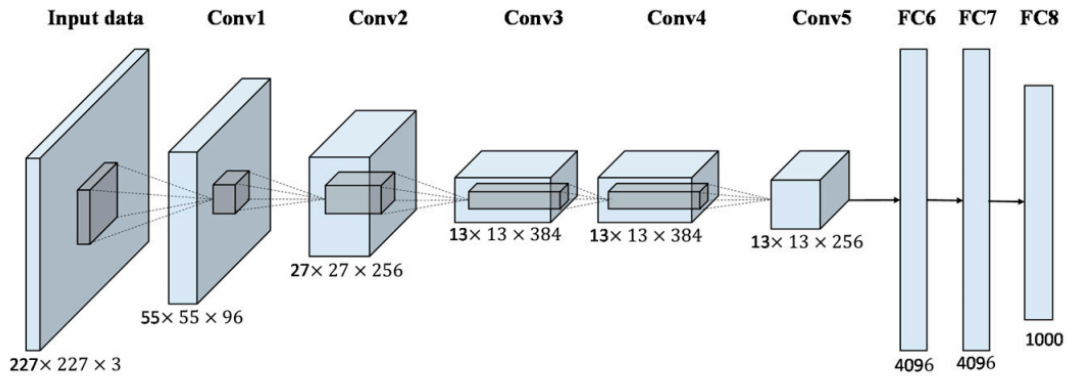
**Şekil 2.** Çalışmada kullanılan veri sınıfları a) Hesap makinesi b) Bardak c) Klavye d) Monitör e) Mouse f) Gözlük g) Tornavida

## 2.2. Evrişimsel Sinir Ağları (ESA)

ESA, bilgisayarla görme ve örüntü tanımda nesnelerin sınıflandırılması, bölütlenmesi ve tespiti için kullanılmaktadır. Oldukça başarılı sonuçlar veren bu yöntem, girdi verilerinin hiyerarşik öğrenme prensibine dayanmaktadır. Girdiler katmanlar aracılığıyla işlendikçe, elde edilmesi istenen bilginin ayrışma seviyesi artar. Katmanların derinleşmesiyle bilgi daha ayrıntılı olarak öğrenilir.

ESA mimarisi evrişimsel katman, havuzlama katmanı ve Doğrultmuş Doğrusal Birim (Rectified Linear Unit-Relu) olmak üzere üç ana katmandan oluşmaktadır (Hou ve ark., 2016). **Evrişimsel Katman**, ESA'nın ana katmanıdır. Girdiyi birleştirerek sonucunu bir sonraki katmana aktarır. Görüntüye uyguladığı bazı filtreler ile özellik çıkarmayı sağlar. **Havuzlama Katmanı**, bir katmandaki nöron kümelerinin çıktılarını bir sonraki katmandaki tek bir nörona birleştirerek verilerin boyutunu azaltır. Havuzlama işlemi maksimum veya ortalama olarak gerçekleştirilir. Aktivasyon fonksiyonları, ESA'lar için önemli bir faktördür. **Relu**, en yaygın kullanılan aktivasyon işlemidir. Eşitlik 1'de matematiksel gösterimi verilmiştir. Bu katmanın en önemli özelliği girişteki negatif değerleri sıfıra çekmesidir. Ağ bu katmanla daha hızlı öğrenme gerçekleştirmektedir. ESA mimarisinde yer alan **Tam bağlantılı katman** ise sınıflandırma işlemi için kullanılmaktadır.

$$f(x) = \begin{cases} 0 & \text{eğer } x < 0 \\ x & \text{eğer } x \geq 0 \end{cases} \quad (1)$$

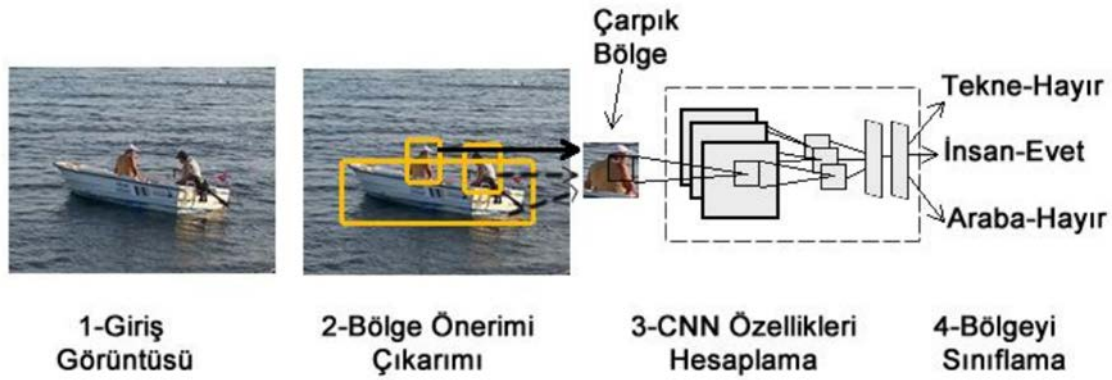


Şekil 3. Alexnet mimarisi (Han ve ark., 2017)

Çalışmada Şekil 3'teki Alexnet (Krizhevsky ve ark., 2012) ESA mimarisi kullanılmıştır. Alex Krizhevsky, Ilya Sutskever ve Geoffrey Hinton tarafından geliştirilen bu mimari ESA modellerinin ve derin öğrenmenin daha popüler olmasını sağlamıştır. Alexnet 5 evrişim katmanı ve 3 tam bağlantılı katmandan oluşmaktadır.

### 2.3. ESA ile Nesne Tanıma

ESA görüntü sınıflandırmaya ve görüntü içerisindeki nesnenin olasılığını tahmin eder. Görüntü içerisindeki nesnenin algılanması ile olasılığını ve bulunduğu yeri tahmin etmek için görüntü içerisinde çok sayıda bölge ve bu bölgelerin yoğunluklarını hesaplamak gerekir. Bu nedenle B-ESA gibi algoritmalara ihtiyaç vardır. Girshick ve arkadaşları (Jia ve ark., 2014) çok sayıda bölgenin oluşturacağı zorluğu azaltmak için B-ESA algoritmasını geliştirmişlerdir. Şekil 4'te örnek bir B-ESA mimarisi verilmiştir.



Şekil 4. B-ESA mimarisi (Kutlu, 2019)

Dört bölümden oluşan mimaride giriş görüntüsü içerisinde nesne barındıran olası 2000 bölgeye ayrılmaktadır. Ardından her bölge için ESA uygulanır. ESA çıkışında bulunan Destek Vektör Makinesi ile de bölge sınıflandırması yapılır. Bu çalışmada nesne tanıma Alexnet ESA mimarisi kullanılarak B-ESA algoritması ile gerçekleştirilmiştir.

### 3. Bulgular ve Tartışma

Bu çalışmada, robotik uygulamalarda da kullanılabilecek bir nesne tespiti ve sınıflandırılması amaçlanmıştır. Alexnet ESA mimarisi ve B-ESA algoritması kullanılarak MATLAB programlama dili kullanılarak gerçekleştirilmiştir. Çalışma AMD Ryzen 73750H 2.3 GHz işlemci, NVIDIA Geforce GTX 1050 3 GB ekran kartı ve 16 GB RAM değerlerine sahip bilgisayarda tamamlanmıştır.

Çalışmanın ilk aşamasında verilerin etiketlenmesi aşaması gerçekleştirilmiştir. Etiketleme işlemi manuel olarak MATLAB Image Labeling uygulaması ile yapılmıştır. Tablo 1'de etiket isimleri, eğitim verisi ve test verisi sayıları verilmiştir. Yedi farklı sınıfa ait toplam 684 görüntü ağı eğitilmesi için kullanılmıştır. Toplam verinin %25'ini oluşturan test verisi ise eğitilen ağı test edilmesini sağlamaktadır. Test verileri internet ortamından temin edilmiştir.

**Tablo 1.** Eğitim ve test verileri için görüntü sayıları.

Sınıf Numarası	Etiket	Eğitim Verisi Sayısı	Test Verisi Sayısı
1	Hesap makinesi	100	33
2	Bardak	87	29
3	Klavye	85	28
4	Monitör	133	44
5	Mouse	94	31
6	Gözlük	83	27
7	Tornavida	102	34

B-ESA algoritması ile eğitilen ağın eğitilme parametreleri 20 epoch ve öğrenme hızı 0.001 olarak belirlenmiştir. 20 epoch değerine 1820 iterasyonda ve yaklaşık 36 dakikada ulaşılmıştır.

Ağın eğitilmesi sonrasında test görüntüleri eğitilen ağda test edilmiştir. Çıktı görüntülerinden bazıları Şekil 5'deki gibidir. Mouse, tornavida ve hesap makinesi sınıfına ait görseller doğru bir şekilde sınıflandırılmış ve hepsi de yaklaşık 0.99 güven değerleri arasında tespit edilmiştir.

**Şekil 5.** Test edilen bazı görseller

Test edilen görüntülere göre karışıklık matrisi Tablo 2'de verilmiştir. Tabloda her bir etiket değerine göre ağ çıktısı sınıflandırma sonuçları gösterilmektedir.

**Tablo 2.** Çalışmaya ait karışıklık matrisi.

	Hesap makinesi	Bardak	Klavye	Monitör	Mouse	Gözlük	Tornavida
Hesap makinesi	<b>30</b>	0	3	0	0	0	0
Bardak	0	<b>27</b>	0	0	2	0	0
Klavye	4	0	<b>24</b>	0	0	0	0
Monitör	0	0	0	<b>44</b>	0	0	0
Mouse	0	3	1	0	<b>27</b>	0	0
Gözlük	0	0	0	0	0	<b>27</b>	0
Tornavida	0	0	1	0	0	0	<b>33</b>

Karışıklık matrisine göre her bir sınıfa ait doğruluk değerleri ve toplam doğruluk değerleri Tablo 3'te verildiği gibidir. Nesne sınıflandırmasında en yüksek doğruluk %100 ile monitör ve gözlük etiketlerine ait iken en düşük doğruluk ise %85,71 ile klavye etiketinde saptanmıştır. Test edilen toplam 226 görüntü içindeki nesnelere %93,81 doğrulukla sınıflandırılmıştır.

**Tablo 3.** Test verilerinin doğruluk değerleri.

Etiket	Doğru tahmin	Yanlış Tahmin	Doğru Tahmin Yüzdesi
Hesap makinesi	30	3	%90,91
Bardak	27	2	%93,10
Klavye	24	4	%85,71
Monitör	44	0	%100,0
Mouse	27	4	%87,10
Gözlük	27	0	%100,0
Tornavida	33	1	%97,06
Toplam	212	14	%93,81

Gu ve arkadaşları yaptıkları çalışmada (Lee ve ark., 2019) kendi oluşturdukları veri setini kullanarak Alexnet mimarisi ile tenis topu tespitini yapmıştır. Test edilen 30 test görüntüsünün tamamı da doğru bir şekilde sınıflandırılmıştır. Lee ve arkadaşları (Lee ve ark., 2019) Alexnet mimarisini kullanarak geri dönüşüm için plastik şişe ve alüminyum folyo poşeti tespitini gerçekleştirmişlerdir. Kendilerinin oluşturduğu veri setinde nesnelere birçok açıdan görüntüleri yer alan 1000 görsel kullanılmış ve test edilen görüntülerin tamamı başarılı bir şekilde sınıflandırılmıştır. Lv ve arkadaşlarına ait çalışmada (Lv ve ark., 2019) kendileri oluşturdukları veri setinde Alexnet mimarisi ile elektrik panosu ve üzerindeki gerilim göstergesi, uyarı ışıkları ve elektrik değerleri göstergesi tespiti gerçekleştirmişlerdir. Sınıflara ait 47 test görüntüdeki nesnelere tamamı doğru bir şekilde sınıflandırılmıştır. Alexnet mimarisi kullanarak yüksek doğrulukla tamamlanan benzer çalışmalar da (Do ve ark., 2018; Espinosa ve ark., 2017; Fu ve ark., 2017; Haque ve ark., 2019; Huang ve ark., 2019; Kang ve ark., 2017; Li, 2017; Lu ve ark., 2017; Lv ve Tan, 2019; Tomè ve ark., 2016)

vardır. Kang ve ark., (2017) YouTubeObjects veri setini ve Googlenet ve VGG ESA mimarilerini kullanarak en yüksek %93 başarımları sağlamıştır. Haque ve ark., (2019) PASCAL VOC2010 veri setini ve VGG ile birlikte Resnet kullanarak en yüksek %93,7 başarımları elde etmişlerdir.

Bu çalışmada da Alexnet mimarisi kullanılmış ve diğer çalışmalardan farklı olarak ev ve ofis ortamlarındaki nesnelere tespiti ve sınıflandırılması yapılmıştır. Doğru sınıflandırma başarımları ise diğer çalışmalardan daha yüksek olarak %93,81 şeklinde tespit edilmiştir.

#### 4. Sonuçlar ve Öneriler

Bu çalışmada robotik uygulamalar için de kullanılacak ESA tabanlı nesne tespiti işlemi gerçekleştirilmiştir. Yedi farklı nesne sınıfına ait görüntüler kullanılmış ve test edilmiştir. Her nesneye ait bulunan doğruluk değerlerine bakılarak çalışmanın başarılı olduğu sonucu varılabilmektedir.

Nesne tanıma ve sınıflandırılması çeşitli alanlarda kullanılabildiği gibi robotik alanlarda da kullanılmaktadır. Endüstrideki gelişmeler ile birlikte robotların da sanayideki önemi sürekli artmaktadır. Tut-yerleştir ve paketleme gibi görevlere sahip robotların nesnelere tanımları da gerekmektedir.

Çalışmanın sonraki aşamasında hata değerini ve ağırlık eğitilme süresini azaltmak için farklı mimaride ESA ve daha hızlı bölge tabanlı ESA algoritmaları kullanılabilir. Robot üzerine entegre edilecek kamera ile de gerçek zamanlı nesne tanıma yapılabilir.

#### Kaynaklar

- Çoban, M., Cubukcu, B., Yüzgeç, U. 2019. Nesne Takibi Yapan Robot Uygulaması (RasPiBot) Object Tracking Robot Application (RasPiBot).
- Do, T.-T., Nguyen, A., Reid, I. 2018. Affordancenet: An end-to-end deep learning approach for object affordance detection. 2018 IEEE international conference on robotics and automation (ICRA). IEEE. pp. 1-5.
- Espinosa, J.E., Velastin, S.A., Branch, J.W. 2017. Vehicle detection using alex net and faster R-CNN deep learning models: a comparative study. International Visual Informatics Conference. Springer. pp. 3-15.
- Fu, C.-Y., Liu, W., Ranga, A., Tyagi, A., Berg, A.C. 2017. Dssd: Deconvolutional single shot detector. arXiv preprint arXiv:1701.06659.
- Griffin, G., Holub, A., Perona, P. 2007. Caltech-256 object category dataset.
- Gu, S., Ding, L., Yang, Y., Chen, X. 2017. A new deep learning method based on AlexNet model and SSD model for tennis ball recognition. 2017 IEEE 10th International Workshop on Computational Intelligence and Applications (IWCIA). IEEE. pp. 159-164.
- Han, X., Zhong, Y., Cao, L., Zhang, L. 2017. Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification. Remote Sensing, 9(8), 848.
- Haque, M.F., Lim, H.-Y., Kang, D.-S. 2019. Object Detection Based on VGG with ResNet Network. 2019 International Conference on Electronics, Information, and Communication (ICEIC). IEEE. pp. 1-3.
- Hariharan, B., Arbeláez, P., Girshick, R., Malik, J. 2014. Simultaneous detection and segmentation. European Conference on Computer Vision. Springer. pp. 297-312.



- Hou, L., Wu, Q., Sun, Q., Yang, H., Li, P. 2016. Fruit recognition based on convolution neural network. 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). IEEE. pp. 18-22.
- Huang, L., Ren, K., Fan, C., Deng, H. 2019. A Lite Asymmetric DenseNet for effective object detection based on convolutional neural networks (CNN). Optoelectronic Imaging and Multimedia Technology VI. International Society for Optics and Photonics. pp. 111871T.
- Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T. 2014. Caffe: Convolutional architecture for fast feature embedding. Proceedings of the 22nd ACM international conference on Multimedia. pp. 675-678.
- Kang, K., Li, H., Yan, J., Zeng, X., Yang, B., Xiao, T., Zhang, C., Wang, Z., Wang, R., Wang, X. 2017. T-cnn: Tubelets with convolutional neural networks for object detection from videos. IEEE Transactions on Circuits and Systems for Video Technology, 28(10), 2896-2907.
- Karpathy, A., Fei-Fei, L. 2015. Deep visual-semantic alignments for generating image descriptions. Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3128-3137.
- Khoreva, A., Benenson, R., Ilg, E., Brox, T., Schiele, B. 2019. Lucid data dreaming for video object segmentation. International Journal of Computer Vision, 127(9), 1175-1197.
- Krizhevsky, A., Sutskever, I., Hinton, G.E. 2012. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems. pp. 1097-1105.
- Kutlu, Ö. 2019. İnsansız hava aracı ile elde edilen görüntülerin derin öğrenme yöntemleri ile analizi.
- Lee, S.-H., Yeh, C.-H., Hou, T.-W., Yang, C.-S. 2019. A Lightweight Neural Network Based on AlexNet-SSD Model for Garbage Detection. Proceedings of the 2019 3rd High Performance Computing and Cluster Technologies Conference. pp. 274-278.
- Li, B. 2017. 3d fully convolutional network for vehicle detection in point cloud. 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE. pp. 1513-1518.
- Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., Pietikäinen, M. 2020. Deep learning for generic object detection: A survey. International journal of computer vision, 128(2), 261-318.
- Lu, K., An, X., Li, J., He, H. 2017. Efficient deep network for vision-based object detection in robotic applications. Neurocomputing, 245, 31-45.
- Lv, L., Tan, Y. 2019. Detection of cabinet in equipment floor based on AlexNet and SSD model. The Journal of Engineering, 2019(15), 605-608.
- Tang, P., Wang, C., Wang, X., Liu, W., Zeng, W., Wang, J. 2019. Object detection in videos by high quality object linking. IEEE transactions on pattern analysis and machine intelligence.
- Tomè, D., Monti, F., Baroffio, L., Bondi, L., Tagliasacchi, M., Tubaro, S. 2016. Deep convolutional neural networks for pedestrian detection. Signal processing: image communication, 47, 482-489.
- Voigtlaender, P., Chai, Y., Schroff, F., Adam, H., Leibe, B., Chen, L.-C. 2019. Feelvos: Fast end-to-end embedding learning for video object segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9481-9490.
- Wang, B., Tang, S., Xiao, J.-B., Yan, Q.-F., Zhang, Y.-D. 2019. Detection and tracking based tubelet generation for video object detection. Journal of Visual Communication and Image Representation, 58, 102-111.
- Zou, Z., Shi, Z., Guo, Y., Ye, J. 2019. Object detection in 20 years: A survey. arXiv preprint arXiv:1905.05055.