Avrupa Bilim ve Teknoloji Dergisi
Sayı 25, S. 159-171, Ağustos 2021
© Telif hakkı EJOSAT'a aittir
Derleme Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
No. 25, pp. 159-171, August 2021
Copyright © 2021 EJOSAT
Review Article

Derin Öğrenme Teknikleri İle Nesne Tespiti Ve Takibi Üzerine Bir İnceleme

Fatma Gülşah Tan[1*], Asım Sinan Yüksel[2], Erdal Aydemir[3], Mevlüt Ersoy[4]

[1*] Süleyman Demirel Üniversitesi, Rektörlük, Bilgi İşlem Daire Başkanlığı, Isparta, Türkiye, (ORCID: 0000-0002-2748-0396), gulsahtan@sdu.edu.tr
[2] Süleyman Demirel Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Isparta, Türkiye, (ORCID: 0000-0003-1986-5269), asimyuksel@sdu.edu.tr
[3] Süleyman Demirel Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, Isparta, Türkiye, (ORCID: 0000-0003-4834-725X), erdalaydemir@sdu.edu.tr
[4] Süleyman Demirel Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Isparta, Türkiye, (ORCID: 0000-0003-2963-7729),mevlutersoy@sdu.edu.tr

(İlk Geliş Tarihi 15 Şubat 2021 ve Kabul Tarihi 27 Haziran 2021)

(**DOI:** 10.31590/ejosat.878552)

ATIF/REFERENCE: Tan, F.G., Yüksel, A.S., Aydemir, E. & Ersoy, M. (2021). Derin Öğrenme Teknikleri ile Nesne Tespiti ve Takibi Üzerine Bir İnceleme. *Avrupa Bilim ve Teknoloji Dergisi*, (25), 159-171.

Öz

Derin öğrenme, son zamanlarda insan hatalarını en aza indirmesiyle popüler olan yapay zekâ yaklaşımlarındandır. Derin öğrenme teknikleri birçok alanda büyük miktardaki veri kullanımı ile başarılı bir şekilde algılama, yorumlama yapabilme yeteneğine sahiptir. Özellikle görüntü işleme alanında birikmiş etiketli verilerdeki hızlı artış derin öğrenme algoritmalarına yönelmeyi zorunlu hale getirmiştir. Bu alanlardaki verilerin giderek artmasıyla büyük verilerden yararlı bilgiyi ayırmak ve metin, görüntü, ses dosyalarına anlam kazandırmak amacıyla derin öğrenme yöntemleri kullanılmaktadır. Son yıllarda, nesne tespiti ve nesne takibi alanında yapılan çalışmalarda artış görülmektedir. Videolar gibi durağan olmayan görüntüler üzerinde tespit ve analiz sonrasında takip edilecek olan bir nesne varsa anlamlı bilgiler çıkarmak daha zor olmaktadır. Bu gibi durumlarda derin öğrenme algoritmalarının kullanılması görüntü işleme problemlerinin kolaylıkla çözüme kavuşturulabilmesini sağlamaktadır. Bu çalışmanın amacı; derin öğrenme ile nesne tespiti ve takibi konusunda yapılan uygulamaları incelemek, son gelişmeleri anlatmak, popüler kütüphaneler, veri setleri, algoritmalar hakkında bilgi vererek bu alanda çalışacak olan araştırmacılara yardımcı olmaktır.

Anahtar Kelimeler: Derin Öğrenme, Görüntü İşleme, Nesne Tespiti, Nesne Takibi, Sinir Ağları.

A Review On Object Detection And Tracking With Deep Learning Techniques

Abstract

Deep learning is one of the artificial intelligence approaches that has recently become popular for minimizing human error. Deep learning techniques have the ability to successfully detect and interpret with the use of large amounts of data in many areas. Especially, the rapid increase in labeled data accumulated in the field of image processing has made it necessary to turn to deep learning algorithms. With the increasing data in these areas, deep learning methods are used to separate useful information from big data and to give meaning to text, images and audio files. In recent years, there has been an increase in the studies conducted in the field of object detection and object tracking. If there is an object to be followed after detection and analysis on non-stationary images such as videos, it is more difficult to extract meaningful information. In such cases, the use of deep learning algorithms enables image processing problems to be solved easily. The aim of this study is to examine the applications of deep learning and object detection and tracking, to explain the latest developments, to help researchers who will work in this field by giving information about popular libraries, data sets, algorithms.

Keywords: Deep Learning, Image Processing, Object Detection, Object Tracking, Neural Networks.

[*] Corresponding Author: gulsahtan@sdu.edu.tr

1. Giriş

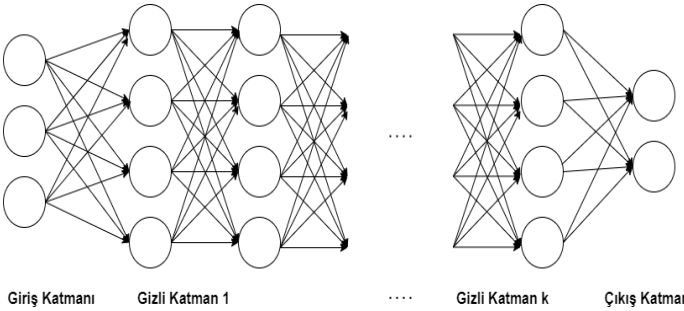
Son yıllarda gelişen teknoloji sayesinde görüntü ve videolardaki artış, arka planda etiketsiz yığılmış verilerin artmasına da neden olmuştur. Bundan dolayı bu verilerden anlamlı sonuçlar çıkarmak, verilerin analizini yapmak önemli bir konu haline gelmiştir. Bu problemi çözmek için görüntü işleme algoritmaları kullanılmaktadır. Görüntü işleme algoritmaları yararlı bilgiyi seçmek; metin, görüntü, ses dosyalarına anlam kazandırmak; yüz tanıma, plaka tespiti, nesne tespiti gibi birçok alanda kullanılmaktadır (Luo vd., 2014). Veri miktarındaki artış ile verilerin karmaşıklığı arttıkça geleneksel yöntemler yetersiz kalmaya başlamıştır. Görüntü işleme algoritmalarının yetersiz kaldığı uygulamalarda son yıllarda popüler hale gelen, güçlü öğrenme yeteneğine sahip ve yüksek doğruluk oranı sağlayan derin öğrenme yöntemleri kullanılmaya başlanmıştır.

Derin öğrenme yöntemleri endüstri (Wallach vd., 2015), sağlık (Havaei vd., 2017), askeri sistemler (Yang, 2018), (Verma ve Dhillon, 2017), (Fu, 2018) ve akademik alanlar başta olmak üzere birçok araştırma ve uygulamada; görüntü işleme ve büyük miktardaki verileri başarılı bir şekilde yorumlama yeteneğine sahiptir. Özellikle görüntü işleme alanında, nesne tespiti ve takibi son yıllarda üzerinde çalışılan popüler bir konu haline gelmiştir. Durağan olmayan görüntülerde nesne tespitinin yapılması zor bir süreçtir. Derin öğrenme, nesne tespiti ve takibini yüksek performansta gerçekleştirmek için birden fazla yöntem barındırmaktadır. Belirlenmiş bir nesneye ait şekil, konum, renk, hız gibi özellikler kullanılarak görüntü ve videoda ilgilenilen nesnenin tespiti ve takibi yapılabilir (Wang, 1998) ya da arka planın videodan çıkarılması, takip edilecek nesnenin özniteliklerinin belirlenmesi, nesnenin sınıflandırılması gibi işlemlerle nesne takibi ve tespiti yapılabilmektedir (Avidan, 2004). Bu noktada probleme göre doğru algoritmanın, mimarinin seçimi önem kazanmaktadır.

Bu çalışmada, derin öğrenme yöntemleri kullanılarak nesne tespiti ve takibi alanında son yıllarda yapılan farklı kategorilerdeki çalışmalar incelenmiştir. Kullanılan ağlar, mimariler, veri setleri, algoritmalar, kütüphaneler hakkında bilgiler verilerek bu alanda çalışacak olan araştırmacılara yardımcı olmak amaçlanmıştır.

2. Derin Öğrenme

Derin öğrenme; doğrusal olmayan katmanlardan yararlanarak ses dosyalarını metne dönüştürme, görüntü analizi, görüntü sınıflandırma, verileri işleme gibi işlemler için kullanılan makine öğrenmesi yöntemlerindendir. Derin öğrenme katmanlardan oluşur ve her katman önceki katmandan gelen çıktıyı kendisine girdi olarak alır (Deng ve Yu, 2013). Derin öğrenme katmanlarında girdi katmanı ilk katman, çıkış katmanı ise son katmandır. Giriş ve çıkış katmanı arasındaki katmanlara da gizli katmanlar denilmektedir. Şekil 1'de derin ağ mimarisi gösterilmiştir.



Şekil 1. Derin ağ mimarisi (Amidi ve Amidi, 2020)

Yapay zekâ çalışmalarında kullanılan makine öğrenimi yöntemi yapay zekânın alt dalı olarak kabul edilmektedir. Derin öğrenme de makine öğreniminin alt dalıdır. Şekil 2'de yapay zekâ, makine öğrenmesi ve derin öğrenmeye ait ilişki gösterilmiştir.



Şekil 2. Yapay zekâ, makine öğrenmesi, derin öğrenme kronolojisi

Microsoft, Apple, Google, Facebook gibi büyük ölçekli teknoloji ve yazılım şirketleri kendi ürünlerinde derin öğrenme yöntemleri kullanmaktadırlar (Cbinsights, 2019). Bu uygulamalarda kullandıkları teknikleri açık hale getirerek bu alanda ilerlemek isteyen kullanıcılara destek olmaktadırlar.

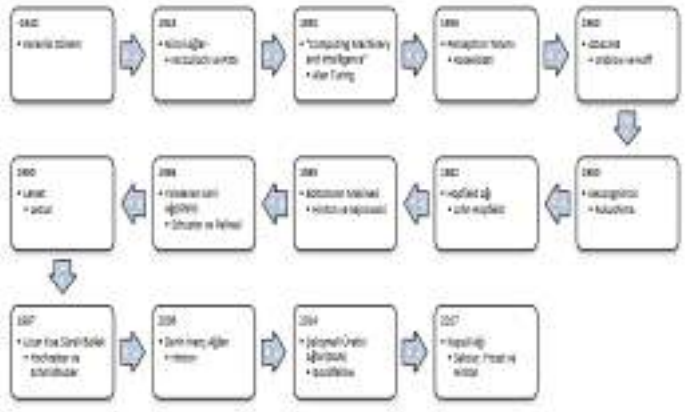
2.1. Derin Öğrenmenin Evrimi

Derin öğrenme, beyin nöronlarının mimarisinden etkilenilerek oluşmuştur. Bu keşifle birlikte çok katmanlı sinir ağlarını eğitmek için araştırmacılar büyük çaba sarf etmişlerdir (Bolouri, 1995). Derin öğrenme kavramı ilk olarak, yapay sinir ağları hakkındaki çalışmaları sırasında Igor Aizenberg ve arkadaşları tarafından 2000'li yıllarda kullanılmıştır. Yapay zekâ ve derin öğrenmenin geçmişine bakıldığında dönemsel kırılmalar olduğu görülmektedir. Dönemsel olarak bu gelişmeler incelendiğinde;

1940 ve 60'li yıllarda derin olmayan sinir ağlarından bahsedilmiş ve uygulamalar yapılmıştır. 1943 yılında bir sinir ağına ait ilk matematiksel model Warren McCulloch ve Walter Pitts tarafından geliştirilmiştir (McCulloch and Pitts, 1988). 1950 yılında Alan Turing (Turing, 2012), makine öğrenimi kavramının gelişmesini ve ileride nasıl etkiler yaratacağını öngörmüştür. "Computing Machinery and Intelligence" adlı çalışmasında genetik algoritmalardan bahsetmiştir. 1960 ve 80'li yıllarda derin öğrenme ağları, geri yayılma, ilk yapay zekâ kışı ve konvolüsyonel sinir ağları kavramları ortaya çıkmıştır. 1965 yılında Alexey Ivakhnenko ve VG Lapa, derin öğrenme ile ilgili teorileri kullanarak grup veri işleme metodunu geliştirmiş ve ilk derin öğrenme ağlarını oluşturmuşlardır (Ivakhnenko and Lapa, 1965). 1980'li yıllarda Kunihiko Fukushima, Neocognitron adını verdiği el yazısı karakterleri, görüntü tanıma için çok katmanlı

evrişimsel yapay sinir ağını önermiştir (Fukushima, 1980). 1985 yılında Terry Sejnowski, NetTalk oluşturmak için İngilizce kelimeleri telaffuz eden programının sürecini geliştirmiştir. 1986 yılında Geoffrey Hinton, David Rumelhart ve Ronald J. Williams, kelime tahmini ve şekil tanımada gelişmeler kaydederek sinir ağlarının nasıl hızlı ve kapsamlı şekilde geliştirilebileceğini göstermiştir (Sejnowsk ve Rosenberg, 1986). 1989 yılında Yann LeCun, elle yazılmış rakamları okuyabilmek için konvolüsyonel sinir ağlarını geri yayılım kuramlarıyla birleştirerek elle yazılmış çekleri ve posta kodlarını okumak için kullanmıştır (LeCun vd., 1989).

1990'lı yıllar ikinci yapay zekâ kışı olarak bilinmektedir. Bu dönemden günümüze kadar baktığımızda denetimsiz derin öğrenme, denetimli derin öğrenme ve modern derin öğrenme kavramları ortaya çıkmıştır. 1993 yılında Jürgen Schmidhuber, çok derin öğrenme kavramı ile sinir ağlarının derinliği ve karmaşıklığında büyük bir adım atmıştır. 1995 yılında Corinna Cortes ve Vladimir Vapnik, benzer iki gruba ait verileri sınıflandırmak için destek vektör makinelerini geliştirmişlerdir (Cortes ve Vapnik, 1995). 1997 yılında Jürgen Schmidhuber ve Sepp Hochreiter, sinir ağlarının verimliliğini arttırmaya yönelik, uzun vadeli bağımlılık problemlerini ortadan kaldıran uzun kısa süreli bellek adını verdikleri sinir ağı modelini önermişlerdir. 2000'li yıllarda, Hinton ve Salakhutdinov'un çalışmalarında çok katmanlı bir sinir ağının katmanının tek seferde önceden nasıl eğitebileceği gösterilmiş ve derin öğrenme kavramı popülerlik kazanmaya başlamıştır (Salakhutdinov ve Hinton, 2009). 2009 yılında Fei-Fei Li, denetimli öğrenme ile sinir ağlarını eğitmek için kullanılabilecek 14 milyondan fazla etiketli verilerin bulunduğu çok büyük ve ücretsiz olan ImageNet'i başlatmıştır (Deng vd., 2009). 2014 yılında, Facebook kendi ortamında paylaşılan fotoğraflarda kullanıcıları otomatik olarak bulup etiketlemek için derin öğrenme ile geliştirilen DeepFace sistemini kullanmıştır. Derin öğrenme algoritmaları %97.35 oranındaki hassasiyetle insanlara rakip bir gelişme göstermiştir. 2016 yılında Google'ın AlphaGo programı, profesyonel bir Go oyuncusu olan Lee Sedol'u yenmiştir. AlphaGo, makine öğrenmesi ve ağaç arama teknikleri kullanılarak DeepMind tarafından geliştirilmiştir (Gibney, 2016). 1940'lı yıllarda başlayıp günümüzde hala devam eden yapay zekânın eksikliklerini makine öğrenmesi tamamlamıştır. 2010 yılından sonra büyük verinin işlenmesi ve yapay zekâ evriminde makine öğrenmesinin eksik kaldığı yerlerin tamamlanması derin öğrenme ile gerçekleştirilmiştir. Şekil 3'te derin öğrenme zaman çizelgesi gösterilmiştir.



Şekil 3. Derin öğrenme zaman çizelgesi

2.2. Derin Öğrenmenin Uygulama Alanları

Derin öğrenme verinin olduğu her alanda başarılı bir şekilde uygulanabilir. Derin öğrenmenin uygulandığı popüler alanlara doğal dil işleme, sinyal ve görüntü işleme, video işleme, bilgisayar oyunları, robotlar ve otonom arabalar, ses, konuşma, yüz, metin tanıma, ilaç tasarımı, biyobilişim, duygu analizi, dijital pazarlama, nesne algılama, sanat gösterilebilir.

Sinyal, video ve görüntü işleme alanlarında, Sun vd. (2017) görüntü işleme ile bilgisayarlı tomografi görüntü tanısından akciğer nodülünü tespit etmeyi amaçlamışlardır. Geliştirilen modelin performansını hesaplamak için el ile belirledikleri özellikleri kullanarak geleneksel bilgisayar destekli tanı sistemleriyle karşılaştırmışlardır. Çalışma sonucunda akciğer nodülü tespitinde istenilen performansı elde ettiklerini belirtmişlerdir. Tan ve Cao (2020), derin öğrenme yöntemleri ile geliştirilen video analiz uygulamalarının özellikle mobil platformlardaki enerji tüketimini ve hesaplama yükünü azaltmak için sezgisel tabanlı bir model önermişlerdir. Akıllı telefonlarda yaptıkları deneyler sonucunda önerdikleri modelin etkinliğini kanıtladıklarını belirtmişlerdir. Yousefi-Azar ve Hamey (2017), e-mail veri kümesini kullanarak tek belge özetleme yöntemlerini önermişlerdir. Amaçları terim frekansı girdisinden bir özellik alanı hesaplamaktır. Önerdikleri derin oto-kodlayıcı sayesinde daha ayırt edici bir özellik alanı sağladıklarını ve %11.2 oranında iyileşme olduğunu söylemişlerdir.

Sorin vd. (2020), çalışmalarında son yıllarda doğal dil işleme teknikleri ile radyoloji alanında yapılan çalışmaları incelemişlerdir. Pulmoner emboli ve kırıklar gibi tanıların işaretlenmesinin ve görüntüleme protokollerinin otomatik seçiminin doğal dil işleme ile iyi performans gösterdiğini söylemişlerdir. Biyobilişim ve sağlık alanında, Havaei vd. (2017) derin sinir ağlarını kullanarak beyin tümörlerini segmente eden bir yöntem önermişlerdir. Konvolüsyonel sinir ağı tabanlı bu uygulamaların başarılı sonuçlar verdiğini belirtmişlerdir. Toğaçar ve Ergen (2019), derin öğrenme hakkında literatür taraması yapmışlar ve biyomedikal alanında yapılan çalışmalar hakkında bilgiler sunmuşlardır. Derin öğrenme modelleri ve kütüphanelerini geleneksel yöntemlerle kıyaslayarak elde ettikleri performansları göstermişlerdir. Toğaçar vd. (2019), erişime açık akciğer röntgen filmleri üzerinde konvolüsyonel sinir ağı modeli kullanarak zatürre hastalığını erken tespit etmeye çalışmışlardır. Farklı sınıflandırma algoritmaları kullanarak modelinin başarı oranını test etmişler ve destek vektör makineleri ile %95.8 yüksek başarı oranı elde etmişlerdir.

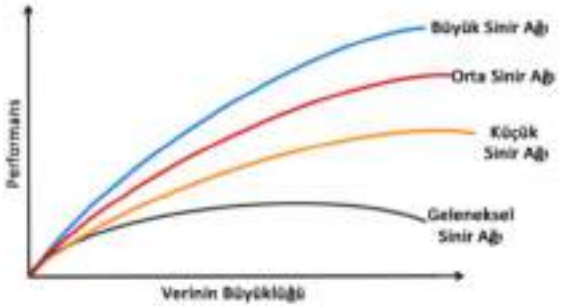
Singla vd. (2017), çalışmalarında e-ticaret sitelerindeki verileri kullanarak duygu analizi yapmışlardır. Müşteri gereksinimlerinin belirlenmesinde ve tüketicinin ürün hakkındaki geri bildirimlerinin daha hızlı alınmasında derin öğrenme yöntemlerinin büyük rol oynadığını belirtmişlerdir. You vd. (2016), cezaevlerinden gelen kısa mesajları güvenli, güvenli olmayan şeklinde sınıflandırmak için derinlemesine öğrenmeye dayalı tekrarlayan sinir ağı modeli kullanmışlardır. Duygu analizinden esinlenerek uygulanan bu model ile %92.7 doğruluk oranı elde etmişlerdir. Sarkar ve Bruyn (2020), çalışmalarında girdi olarak ham verileri kullanarak uzun kısa süreli bellek sinir ağı tabanlı bir tahmin modeli önermişlerdir. Bu model ile dijital pazarlama alanında müşteri davranışlarını büyük bir doğrulukla tahmin edebildiklerini belirtmişlerdir.

Robotik alanda, Mikada vd. (2020) YOLOv3 ve konvolüsyonel sinir ağı kullanarak cerrahi robotların iğnenin penetrasyon ve çekme durumunu tahmin ettiği bir yöntem önermişlerdir. İnsan iş birliğine dayalı yapılan deney sonuçları ile önerilen yönteminin etkinliğini göstermişlerdir.

Ses, konuşma, yüz ve metin tanıma alanlarında, Kim (2020), patentini aldığı uygulamasında yapay sinir ağı tabanlı konuşma tanıma modeli geliştirmiştir. Prasad vd. (2019), çalışmalarında derin öğrenme modellerinden olan VGG-Face kullanarak farklı koşullar altına (farklı açı pozları, ışık vb.) yüz tanıma işlemini başarılı bir şekilde yaptıkları belirtmişlerdir. Brocardo vd. (2017), Twitter ve Enron e-posta veri seti üzerinde Gaussian - Bernoulli derin inanç ağına dayanarak yazar doğrulama çalışması yapmışlardır. %8.21 ile %16.73 arasında değişen hata oranlarıyla umut verici sonuçlar elde etmişlerdir.

2.3. Derin Öğrenmenin Mimarisi

Derin öğrenme kaliteli sonuçlar çıkarmak için çok büyük verilere ihtiyaç duymaktadır. Öğrenme işleminin en yüksek başarı ile yapılabilmesi için sistemin eğitilmesi gerekmektedir. Derin öğrenmeyi makine öğrenmesinden ayıran fark çok katmanlı bir sinir ağına sahip olmasıdır. Öğrenme aşaması katmanlardan oluşur. Daha az ayırt edici özellikler alt katmanlarda, alt katmanların birleştirilmesiyle oluşan üst katmanlarda ise daha fazla ayırt edici özellikler bulunmaktadır. Problemin durumuna göre kullanılacak sinir ağı modeli değişmektedir. Şekil 4'te büyük katmanlara sahip bir sinir ağından daha fazla performans alındığına dair şekil gösterilmiştir.



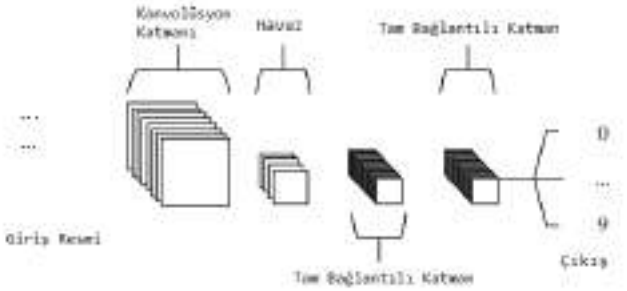
Şekil 4. Derin öğrenmede veri büyüklüğü-performans ilişkisi (Leijnen ve Veen, 2020)

Derin öğrenmede kullanılan mimariler;

- Konvolüsyonel sinir ağları,
- Tekrarlayan sinir ağları,
- Sınırlı boltzmann makineleri,
- Uzun kısa vadeli bellek ağları,
- Derin oto-kodlayıcılar,
- Derin inanç ağlarıdır.

Konvolüsyonel Sinir Ağları (Convolutional Neural Network - CNN)

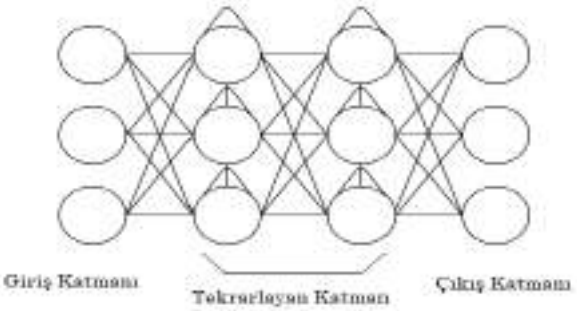
İleri beslemeli, yapay sinir ağı tabanlı, denetimli bir sinir ağıdır. Girdi olarak verilen görüntüdeki nesneleri sınıflandırmak, kümelemek, tanımak için kullanılır (O'Shea ve Nash, 2015). Sabit boyutlu girişler ve çıkışlar vardır. Konvolüsyonel sinir ağı mimarisi tabanlı geliştirilen yöntemlere AlexNet, VGG, GoogleLeNet, Clarifai örnek verilebilir. Şekil 5'te konvolüsyonel sinir ağına ait mimari gösterilmiştir.



Şekil 5. Konvolüsyonel sinir ağı mimarisi (O'Shea ve Nash, 2015)

Tekrarlayan/Yineleyen Sinir Ağları (Recurrent Neural Network – RNN)

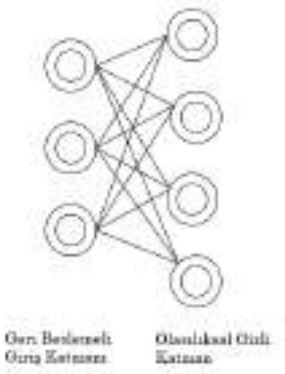
İleri beslemeli yapay sinir ağından farklı olarak kendi girdi dizisini isteğe bağlı olarak işleyen, keyfi giriş ve çıkışları olan denetimli bir sinir ağıdır (Medsker ve Jain, 2001). Amacı ardışık bilgileri kullanarak eğitim modelini oluşturmaktır. Tekrarlayan sinir ağı mimarisi tabanlı geliştirilen yöntemlere Uzun Kısa Süreli Bellek, Kapılı Tekrarlayan Hücre, İki Yönlü Tekrarlı Sinir Ağı, İki Yönlü Uzun Kısa Süreli Bellek örnek verilebilir. Şekil 6'da tekrarlayan sinir ağına ait mimari gösterilmiştir.



Şekil 6. Tekrarlayan sinir ağı mimarisi (Leijnen ve Veen, 2020)

Sınırlı Boltzmann Makineleri (Restricted Boltzman Machine – RBM)

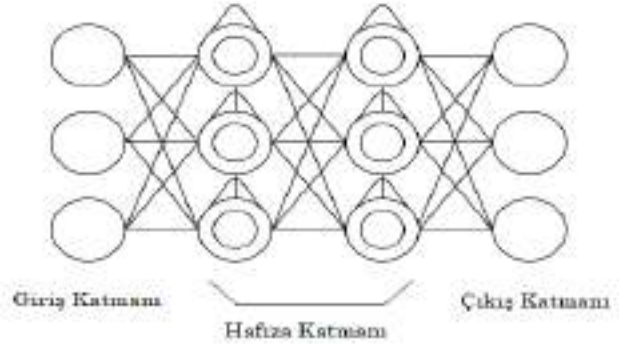
Regresyon, sınıflandırma, modelleme, boyutsallık azaltma gibi alanlarda kullanılan iki katmanlı denetimli ve denetimsiz öğrenme türüne sahip bir sinir ağıdır (Shetty vd., 2020). Katmanlar arasındaki düğümler arasında bağlantı yoktur. Sınırlı Boltzmann makineleri mimarisi tabanlı yöntemlere Derin İnanç Ağları, Derin Boltzmann Makineleri örnek verilebilir. Şekil 7'de sınırlı boltzmann makinelerine ait mimari gösterilmiştir.



Şekil 7. Sınırlı boltzmann makineleri sinir ağı mimarisi (Leijnen ve Veen, 2020)

Uzun Kısa Süreli Bellek Ağları (Long-Short Term Memory – LSTM)

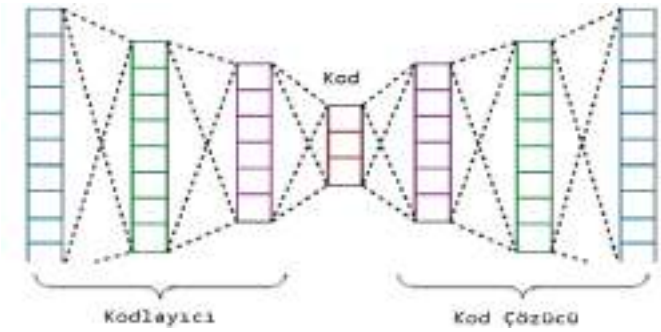
İleri beslemeli sinir ağlarından farklı olarak geri bildirim bağlantılarına dayalı, tekrarlayan yapay sinir ağı mimarisidir (Shetty vd., 2020). Veri işleme, sınıflandırma, tahmin yapma işlemleri için kullanılabilir. Uzun kısa süreli bellek ağları, tekrarlayan sinir ağlarına hafızanın eklenmesiyle oluşmuştur. Bu sayede hafızayı kontrolsüz bir şekilde güncellemek yerine daha seçici davranarak karalı bir şekilde günceller (Hochreiter ve Schmidhuber, 1997). Şekil 8'de uzun kısa süreli bellek ağlarına ait mimari gösterilmiştir.



Şekil 8. Uzun kısa süreli bellek ağı mimarisi (Leijnen ve Veen, 2020)

Derin Oto-Kodlayıcılar (Deep Autoencoders – DAE)

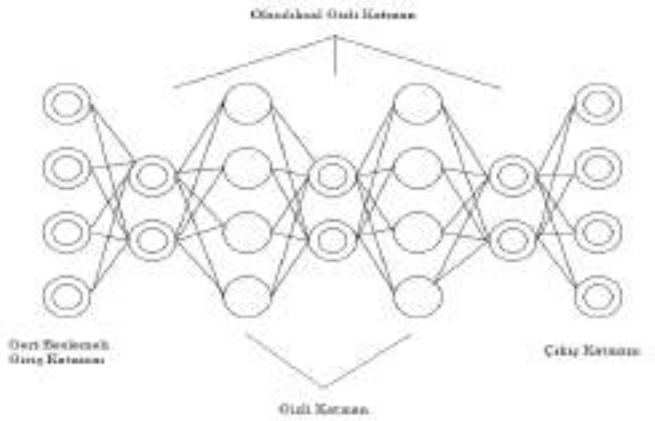
Kodlama ve kod çözücü olarak 2 kısımdan oluşan, eğitim sırasında kendi etiketlerini kullanarak girdi katmanındaki verileri çıkış katmanına gönderen ileri beslemeli, tekrarlanmayan sinir ağı mimarisidir (Şeker vd., 2017). Derin oto-kodlayıcı mimarisi tabanlı geliştirilen yöntemlere Seyrek Oto-Kodlayıcı, Gürültü Temizlemeli Oto-Kodlayıcı, Varyasyonel Oto-Kodlayıcı örnek verilebilir. Şekil 9'da derin oto-kodlayıcılara ait mimari gösterilmiştir.



Şekil 9. Derin oto-kodlayıcı sinir ağı mimarisi (Liu vd., 2017)

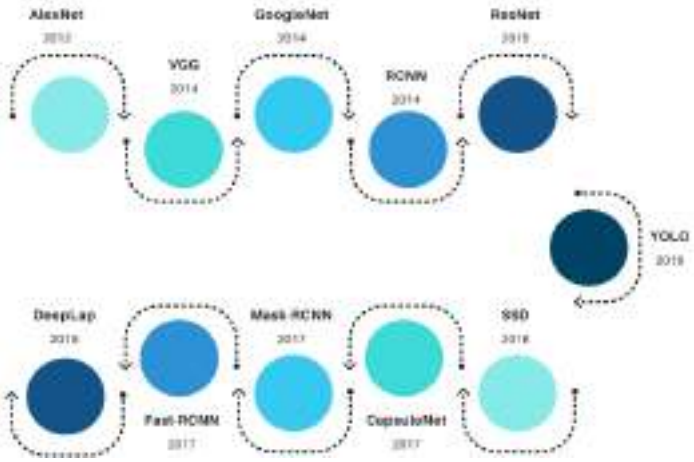
Derin İnanç Ağları (Deep Belief Networks – DBN)

Değişken giriş çıkış boyutuna sahip, katmanların önceki ve sonraki katmanlarla bağlantılı olduğu, düğümler arasında ise yatay iletişimin bulunmadığı bir sinir ağı modelidir. Son katmanda denetimsiz öğrenme ile kümeleme, softmax katmanı ile sınıflandırma yapılabilir (Hinton vd., 2006). Şekil 10'da derin inanç ağlarına ait mimari gösterilmiştir.



Şekil 10. Derin inanç ağları mimarisi (Leijnen ve Veen, 2020)

Derin öğrenme mimarileri ile geliştirilen yöntemlerin zaman içerisindeki gelişimi Şekil 11'de gösterilmiştir.



Şekil 11. Derin öğrenme mimarileri ile geliştirilen yöntemler (Saleem vd., 2019)

2.4. Derin Öğrenmede Kullanılan Kütüphaneler

Derin öğrenme çalışmalarında kullanılan birden fazla kütüphane ve yazılım bulunmaktadır. Bunlara ait bilgiler Tablo 1'de gösterilmiştir.

Tablo 1. Derin öğrenmede kullanılan kütüphaneler

Kütüphane	Dil	Yıl	Açıklama	GitHub Yıldız Sayısı	GPU Çalışma Zamanı Performansı
PyTorch (Collobert vd., 2017)	Python	2017	GPU desteği, hız ve esneklik	45k	2.88
Tensorflow (TensorFlow, 2020)	Python,C++	2020	Otomatik resim yazısı oluşturma yazılımı, tek bir API ile birden fazla GPU ve CPU'ya dağıtma olanağı	152k	5.57
Keras (Keras: The Python Deep Learning API, 2020)	Python	2020	Aktivasyon fonksiyonları ve optimize ediciler	50.4k	-
Digits (NVIDIA DIGITS, 2017)	C++	2017	Çoklu GPU üzerinde verimli, tamamen etkileşimli	4k	-
Theano (Theano Development Team, 2016)	Python	2016	CPU ve GPU üzerinde verimli bir şekilde çalışır, matematiksel hesaplamalar kolay	9.3k	1.40
Caffe (Jia vd., 2014)	Python, C++	2014	Görüntü sınıflandırma ve görüntü bölümleme, GPU desteği	31.2k	2.45
Deep learning 4j (DeepLearning4j, 2020)	Java, C, C++, Python	2020	Görselleştirme aracı	11.9k	-

Bu kütüphanelerden GPU desteği veren Theano, TensorFlow, Caffe ve Torch kullanılarak bazı veri kümeleri ve modeller üzerinden çalışma zamanı performansı karşılaştırıldığında (Yuret, 2016), Theano'nun en hızlı çalıştığı ardından Caffe, Torch'un ve en yavaş çalışanın ise TensorFlow kütüphanesi olduğu görülmüştür. Github yıldız sayılarına bakıldığında her açıdan, TensorFlow liderdir. Keras, Caffe, PyTorch ilk beş sırada yer almaktadır.

3. Derin Öğrenme, Nesne Tespiti ve Nesne Takibi ile İlgili Çalışmalar

Derin öğrenme büyük verileri başarılı bir şekilde analiz edebilmekte ve verinin olduğu her alana uygulanabilmektedir. Bu bölümde son yıllarda derin öğrenme, nesne tespiti-takibi alanında yapılan çalışmalar incelenerek kullanılan mimariler, önerilen yöntemler, kütüphaneler ve ağlar hakkında bilgiler verilmiştir.

Bewley vd. (2016), çalışmalarında Kalman filtresi ve Hungarian algoritması kullanarak çevrimiçi ve gerçek zamanlı nesne takip uygulamaları için bir yaklaşım önermişlerdir. Faster R-CNN ağı ile geliştirdikleri bu modelin izleme performansını %18.9 iyileştirdiğini belirtmişlerdir.

Yu vd. (2016), çalışmalarında öznitelik çıkarımı için GoogleNet ve Kalman filtresi kullanarak yüksek performanslı algılama ve derin öğrenme tabanlı görünüm özelliğini keşfetmişlerdir. Çoklu nesne izleme için geliştirdikleri Faster R-CNN tabanlı bu model ile hem çevrimiçi hem de çevrimdışı ortamda önemli ölçüde daha iyi sonuçlar aldıklarını belirtmişlerdir.

Liu vd. (2017), çalışmalarında yaygın olarak kullanılan derin öğrenme mimarilerini ve uygulamalarını inceleyerek, farklı tiplerde derin sinir ağlarını araştırmışlar ve son gelişmeleri özetleyerek sunmuşlardır. Oto-kodlayıcı, konvolüsyonel sinir ağı, derin inanç ağı ve sınırlı Boltzmann makinesi olmak üzere dört derin öğrenme mimarisi hakkında güncel bir genel bakış sunulmaktadır. Çalışma sonucunda, derin öğrenme teknolojileri ve uygulamaları hakkında en son gelişmeleri kazandırmak amaçlanmaktadır.

Lu vd. (2017), çalışmalarında uzun kısa süreli bellek ağı ile ardışık çerçeveler arasındaki nesne ilişkisini modellemedeki sorunları ele almışlardır. SSD ağı ve Hungarian algoritması kullanarak geliştirdikleri bu modelde nesnelerin özellikleri arasındaki eşleşme hatalarını en aza indirmişlerdir. Geleneksel video nesnesi algılama yöntemleriyle karşılaştırdıklarında yöntemlerinin daha iyi performans gösterdiğini belirtmişlerdir.

Perez vd. (2017), çalışmalarında otomatik pornografik içerik algılamayı araştırmışlardır. Pornografi saptamaya doğal bir yaklaşım sunan çalışma önce çıplaklığı tespit etme ve sonra içeriği daha fazla filtrelemek için uygun eşikleri tanımlama aşamalarından oluşur. Bu tür çözümler yaygın olarak renk, doku ve insan geometrisi gibi insan derisi özelliklerini kullanır. Pornografi tespitinde hareket bilgisinin haritalama sorununu hafiflettiğini belirtmişlerdir. Derin öğrenme mimarileri ile hareket bilgisi kullanıldığında otomatik pornografi tespitinde başarı oranı artmaktadır. Optik akış ve MPEG hareket vektörlerini kullanarak statik (resim) ve dinamik (hareket) bilgileri birleştirmenin yeni yollarını önermişlerdir. Her iki yöntemin de eşdeğer doğruluklar sağladığını, ancak MPEG hareket vektörlerinin daha verimli bir uygulamaya izin verdiğini söylemişlerdir. Deneylerde iki veri seti kabul edilmiştir: Pornografi-800 ve Pornografi-2k. Önerilen yöntem, 800 zorlu test vakası veri kümesinde %97.9 sınıflandırma doğruluğu, son teknolojiye kıyasla %64.4 hata azalımı vermektedir.

Brunetti vd. (2018), çalışmalarında video gözetimi, insan-makine etkileşimi ve analizi konularında sınıflandırma stratejilerini esas alarak bir anket çalışması sunmuşlardır. Son zamanlarda yaya tespit ve takibinde kullanılan konvolüsyon sinir ağları dahil derin öğrenme metodolojilerinin analiz çalışmalarına yer vermişlerdir. Video izleme, birkaç sensör kullanarak zaman içinde tek veya birden fazla nesneyi bulmayı ve takip etmeyi sağlayan karmaşık bir süreçtir. Hem edinim hem de işleme sistemlerinde dikkate değer bir iyileştirme ihtiyacı nedeniyle, literatürde izleme ile ilgili birçok çalışma bulunabilir. Bu çalışmada yaya tespit ve takip sistemi üzerine bir araştırma

sunulmuştur. Analiz edilen çalışmalar, yaya tespit çalışmalarına modern yaklaşımların nasıl araştırılması ve kullanılması gerektiğini vurgulamaktadır. Bununla birlikte, rapor edilen çalışmalar otomatik yaya tespitinde cesaret verici sonuçlar göstermektedir. Ancak daha fazla mimarinin uygulanması ve test edilmesi gerektiğini belirtmişlerdir.

Chaudhary vd. (2018) çalışmalarında, videolarda bulunan birden fazla anormal etkinliği otomatik olarak algılamayı amaçlamışlardır. Suç oranının katlanarak artması nedeniyle, alışveriş merkezlerinde, istasyonlarda, okullarda, havaalanlarında gözetleme sistemleri kurulmaktadır. Bu nedenle, akıllı gözetim sistemine olan büyük bir talepten dolayı önerilen çerçeve ile hareketli nesne algılama, nesne izleme ve etkinlik tanıma için davranış anlama modeli gerçekleştirmişlerdir. Önerilen çerçevenin, birden fazla anormal etkinliği %90'a kadar algılama doğruluğu ile başarılı bir şekilde tespit edebildiğini kanıtlamışlardır. Kim vd. (2018), çalışmalarında önceden eğitilmiş YOLOv2 ile gerçek zamanlı çoklu yaya takibi gerçekleştirmişlerdir. Klasik konvolüsyonel sinir ağı modeli haricinde önceden eğitilmiş bu model sayesinde çevrimiçi öğrenme süresini azaltmışlardır. İki taraflı eşleşme için Greedy algoritmasından yararlanmışlardır. Deneyler sonucunda önerilen algoritmanın kullanılan geleneksel yöntemlerden daha iyi performans sağladığını belirtmişlerdir.

Zhang vd. (2019), çoklu domuz algılama uygulamalarındaki ışık, şekil bozuklukları, tıkanmalar gibi sorunları ele alarak domuzların manuel olarak işaretlenmesini veya fiziksel olarak tanımlanmasını gerektirmeyen gerçek zamanlı tespit yapan bir model önermişlerdir. Konvolüsyonel sinir ağı tabanlı ve Hungarian algoritması kullanılarak gerçekleştirilen bu model ile ticari bir çiftlikten alınan veri kümesiyle yapılan deneylerin zorlu koşullar altında birden fazla domuzu başarılı bir şekilde algılayıp izleyebildiklerini belirtmişlerdir.

Ciaparrone vd. (2020), çalışmalarında tek kameralı videolarda çoklu nesne takibi görevini çözmek için derin öğrenme modellerini kullanan çalışmalar hakkında kapsamlı bir anket sunmaktadır. MOT (Multiple Object Tracking- Çoklu Nesne İzleme) algoritmalarında dört ana adım tanımlanmıştır ve bu aşamaların her birinde derin öğrenmenin nasıl kullanıldığına dair derinlemesine bir inceleme sunulmaktadır. Genel bir MOT boru hattını karakterize etmek için dört ana adım gösterilmiştir: algılama, özellik çıkarma, sonsuz hesaplama, ilişkilendirme. Konvolüsyonel sinir ağları uzaysal örüntü çıkarmada en son teknolojiyi oluşturur ve tekrarlayan sinir ağları görüntü sınıflandırma veya nesne tespiti gibi görevlerde kullanılır. Uzun kısa süreli bellek ses sinyalleri, zamansal seri ve metin gibi sıralı verileri işlemek için kullanılır. MOTChallenge, MOTChallenge1 çoklu nesne takibi için en yaygın kullanılan kriterdir.

Cheng vd. (2020), çalışmalarında video dizilerindeki eğitim süresini kısaltan ve nesneyi izlemek için kullanılan bir tahmin ağı tasarlamışlardır. İlk on karenin nesne konumlarını, tahmin ağını ayarlamak için kullanmışlardır. Tahmin ağı, uyarlanabilir giriş özelliklerini nesne görünümü varyasyonlarını yakalamak için rastgele maskeler üreten üretken çekişmeli ağ çerçevesine entegre etmişlerdir. Son olarak, izleme performansını iyileştirmek için mekansal bir dikkat mekanizması sunmaktadırlar. Önerilen ağ, uzun bir zamansal aralık boyunca nesnelerin en sağlam özelliklerini koruyan maskeyi tanımlayabilir. İki büyük ölçekli kıyaslama üzerinde yapılan kapsamlı deneyler, önerilen algoritmanın son teknoloji yöntemlere göre olumlu performans gösterdiğini göstermektedir.

Iliadis vd. (2020) çalışmalarında, video sıkıştırma algılaması için DeepBinaryMask olarak adlandırılan bir kodlayıcı-kod çözücü sinir ağı modeli önermişlerdir. Video sıkıştırma algılamasında, bir kare, kodlanmış maske sayısına eşit sayıda video karesinin yeniden oluşturulduğu bir dizi kodlanmış maske kullanılarak elde edilir. Önerilen çerçeve, algılama matrisinin video rekonstrüksiyonu ile eğitildiği uçtan uca bir modeldir. Kodlayıcı, algılama matrisinin ikili elemanlarını öğrenerek bir video bloğunu basınç ölçümleriyle eşler. Kod çözücü, çok katmanlı algılayıcı ağının birkaç gizli katmanı aracılığıyla bir video yamasından bir video bloğuna kadar ölçümleri eşlemek üzere eğitilir. Öngörülen video blokları, bilinmeyen video sekansını kurtarmak için birlikte birleştirilir. Yeniden yapılanma performansının, eğitimli algılama maskesi kullanıldığında, çok çeşitli sıkıştırıcı algılama rekonstrüksiyon algoritmalarına göre daha iyi performans sağladığını belirtmişlerdir.

Kanna vd. (2020), çalışmalarında tanığın sağladığı kimlik işaretlerine dayanarak, suçlu gözetim kameralarından suçluları tespit etmek için derin öğrenme kullanarak yeni bir akıllı metodoloji önermişlerdir. Bu yöntem, nesneyi video verilerinden kişi olarak bulmak için cinsiyet, gömlek deseni ve gözlük durumunu girdi olarak alır. Bu yöntemin performansı, video karesindeki kişiyi tanımlamada %87'lik bir doğruluk sağlamıştır. Li vd. (2020) çalışmalarında, bitki hastalıkları ve zararlılarının zamanında tespiti ve tahıl üretiminin arttırılması için video algılama sistemi kurmuşlardır. Gelecekte gerçek zamanlı bitki hastalıkları ve zararlıları video algılama sistemi ile tespit etmek için özel bir yapıya sahip derin bir öğrenme tabanlı video algılama mimarisi önermişlerdir. Önce videoyu durağan kareye dönüştürerek çerçeveyi algılama için hareketsiz görüntü detektörüne göndermişler ve son olarak kareleri videoya sentezlemişlerdir. Hareketsiz görüntü detektöründe, çerçeve olarak Faster R-CNN, göreceli olarak bulanık videoları tespit etmek için görüntü eğitim modellerini kullanmışlardır. Ayrıca deneylerde video algılama kalitesini etkili bir şekilde yansıtan bir makine öğrenme sınıflandırıcısına dayanan bir dizi video tabanlı değerlendirme ölçütü önerilmiştir.

Nawaratne vd. (2020) çalışmalarında, gerçek zamanlı video gözetimi için anomali tespiti ve lokalizasyon zorluklarını ve sınırlamalarını ele almak için artımlı mekânsal-zaman öğrenen (ISTL) modelini önermişlerdir. ISTL, zaman içinde gelişen yeni anomalileri ve normalliği sürekli olarak güncellemek ve ayırt etmek için bulanık kümelenme ile aktif öğrenmeyi kullanan denetimsiz bir derin öğrenme yaklaşımıdır. ISTL, üç karşılaştırma veri seti kullanılarak doğruluk, sağlamlık, hesaplama yükü ve bağlamsal göstergeler üzerinde gösterilmiş ve değerlendirilmiştir. Bu çalışmanın deney sonuçlarına göre gerçek zamanlı video gözetimi için kullanılabilir ve uygun olduğunu belirtmişlerdir.

Yang vd. (2020), çalışmalarında bulut bağlantılı araçlar ve engelli yayalara yardım gibi akıllı şehir kullanım durumlarını desteklemek amacıyla trafik kavşaklarındaki nesnelerin gerçek zamanlı olarak algılanması ve izlenmesi sorununu ele almaktadır. Bu çalışmada, veriler bulutla geliştirilmiş açık yazılım tanımlı test aracı (COSMOS) ile test edilmiştir. Yayaların güvenilir bir şekilde tespit edilmesinin araçlara göre çok daha zor olduğu bilinmektedir. Bu çalışmada bu sorunun çözümü için tespit algoritması (Yolo, MaskR-CNN ve Tek Çekim Algılama) ve çoklu nesne izleme algoritması (Deep-Sort, Multicut ve Deep) performansına genel bir bakış sunmuşlardır. Çoklu nesne takip doğruluk oranları (MOTA) otomobiller için yaklaşık %73.2 ve yayalar için yaklaşık %2.8'dir. Araçların

tespiti/izlenmesi için 33.3 ms'lik bir gecikme hedefi belirlemişler ve böyle bir gecikme, 10 km/s hızla hareket eden araçlar için 0.1 metrelik mesafelere karşılık gelen araç konumlarının yararlı bir şekilde örneklenmesini sağlayacağını iddia etmişlerdir. Ayrıca kavşaklardaki yayalar araçlardan 3-10 kat daha yavaş hareket ettiğinden, yaya algılama gecikmelerinin araç gecikmelerinden 3-10 kat daha fazla olmasına izin verilebileceği sonucuna varılmıştır.

4. Derin Öğrenme ile Nesne Tespiti ve Nesne Takibi

Görüntülerden anlamlı bilgiler çıkarmak için görüntü işleme yöntemleri kullanılır. Videolar üzerinde tespit ve analiz sonrasında takip edilecek olan bir nesne varsa anlamlı bilgiler çıkarmak daha zordur. Bu gibi durumlarda derin öğrenme algoritmalarının kullanılması görüntü işleme problemlerinin kolaylıkla çözüme kavuşturulabilmesini sağlamaktadır. Görüntüler üzerinde nesnelerin hızlı bir şekilde tespitini gerçekleştiren ilk algoritma Viola Jones (Viola ve Jones, 2001) olmuştur.

Derin öğrenme ile nesne tespiti ve takibi alanında yapılan çalışmalar incelendiğinde genel olarak dört adım uygulanmaktadır. Bunlar;

- Ön işleme aşaması,
- Nesnenin tespiti,
- Nesnenin sınıflandırılması
- Nesnenin takibidir.

4.1. Ön İşleme

Belirlenen kategorilere göre oluşturulan veri setlerinde verinin işlenemeye hazır hale getirilmesi için yapılan ön işlemlerdir. Ön işleme, sistematik hataları düzeltmeye veya telafi etmeye çalışan sonraki analizler ve adımlar için veri hazırlayan işlemlerden oluşur. Dijital görüntüler, geometrik, atmosferik gibi çeşitli düzeltmelere tabi tutulur. Tüm bu düzeltmeler her durumda uygulanmayabilir. Bu hatalar sistematiktir ve kullanıma ulaşmadan önce kaldırılabilir.

4.2. Nesne Tespiti

Videolar üzerinde belirlenen nesne tespiti aşaması iki adımdan oluşmaktadır. Birinci adımda nesne belirgin hale getirilir. İkinci adımda ise nesne arka plandan ayrılır. Görünüm ve şekle ilişkin özellikler genellikle nesne tespitinin temeli olarak kullanılır. Nesne tespiti, nesnenin izlenmesi için uygun bir algoritmanın seçimini belirleyen aşamadır. Nesne tespit aşaması sonraki aşamaların başarısını etkilediğinden daha fazla önem taşımaktadır.

Nesnenin tespiti ve takibinde kullanılan metotlar;

- R-CNN (Region Based Convolutional Networks),
- YOLO (You Only Look Once),
- SSD (Single Shot Multibox Detector),
- Mask R-CNN (Ozbaysar ve Borandag, 2018),
- Fast R-CNN,
- Faster R-CNN'dir.

R-CNN, iki yaklaşımı birleştirerek nesne tespiti yapmaktadır. Birincisi nesneye ait bölge önerileri çıkartmak, ikincisi ise bölge önerilerine konvolüsyonel sinir ağı uygulamaktır (Girshick vd., 2016). Çıkarılan bölge önerilerine Destek Vektör Makineleri (DVM) kullanılarak sınıflandırma yapılmaktadır (Daş vd., 2019).

YOLO, tek bir sinir ağı ile problemi uçtan uça optimize ederek doğrudan algılama performansına göre nesne tespiti yapmaktadır (Redmon vd., 2016). Özellikle son yıllarda nesne tespiti konusunda popüler olan algoritmalardandır. Nesne tespitinde hız önemli bir kavramdır. YOLO'yu diğer algoritmalardan ayıran gerçek zamanlı nesne tespiti yapmasıdır.

SSD, tek bir derin sinir ağı kullanarak tek seferde nesne tanıma işlemi yapmakta ve Faster R-CNN'den daha hızlı çalışmaktadır (Liu vd., 2016). Diğer yöntemlerdeki eksiklikleri gidermek için bazı iyileştirilmeler yapılmıştır. Bu iyileştirmeler, SSD'nin daha düşük çözünürlüklü görüntüler kullanarak Faster R-CNN'nin doğruluğunu eşleştirmesini sağlayarak hızı yükseltir. SSD nesne algılama 2 bölümden oluşur:

- Özellik haritasının çıkarılması ve
- Nesneleri algılamak için evrişim filtrelerini uygulama.

Mask R-CNN, bölgesel tabalı nesne tespit algoritmalarının en gelişmişidir. Verimli bir şekilde görüntülerdeki nesneleri algılarken nesnenin olduğu tüm pikselleri belirleyerek (Omeroglu vd., 2019) her örnek için yüksek kalitede segmantasyon maskesi oluşturur (He vd., 2017). Faster R-CNN'i nesne tanıma aşamasında kullanmaktadır.

Fast R-CNN, derin konvolüsyonel ağlar kullanarak önceki çalışmalardan yararlanıp nesne tespitini daha hızlı ve verimli bir şekilde yapmaktadır. Gerçek zamanlı nesne tespitinde başarısız olan R-CNN modelinin hız sorununu çözmek için geliştirilmiştir (Girshick, 2015). Fast R-CNN'nin R-CNN'den farkı ilk aşamada konvolüsyonel sinir ağı uygulaması ve oluşan özellik haritasına göre bölge önerilerini çıkartmasıdır. Sınıflandırma işlemi Softmax katmanı kullanılarak yapılır.

Faster R-CNN, bölge önerileri yaparak nesne tespiti gerçekleştiren R-CNN'deki çalışma zamanını azaltmayı amaçlayan Region Proposal Network (RPN) önerisiyle ortaya çıkmıştır (Ren vd., 2016). RPN ile Fast R-CNN modellerinin birleştirilmiş halidir. Nesnenin bulunduğu yeri giriş verisindeki görüntüyü sınıflandırma aşamasında bir kez kullanarak hesaplar. Nesne tespitinin performansını geliştirmek için evrişimsel katmanlar paylaşılır ve derin bir konvolüsyonel ağ eğitilebilir (Cömert vd., 2019).

4.3. Nesne Sınıflandırma

Nesneleri özelliklerine göre gruplandırmak için sınıflandırma işlemi uygulanmaktadır. Sınıflandırma işlemi yapılırken nesnenin kendine has öznitelikleri kullanılmaktadır (Hanbay ve Üzen, 2017). Doku, eğim, zaman, renk ve hareket nesne sınıflandırmada en yaygın kullanılan özniteliklerdir (Fan vd., 2016). Renge göre nesne sınıflandırmak etkili bir yöntemdir fakat ışık değişimine hassas olduğu için hesaplama zamanı yüksektir. Doku hesaplama zamanı ve doğruluk oranı yüksek bir öznitelik tipidir. İnsan takibi için eğim ve hareket öznitelik tipi kullanılabilir. Gürültüye karşı hassastır. Hareket ve zamansal değişimlerde hesaplama zamanı fazla olduğu için maliyet yüksektir.

Tablo 2'de sınıflandırma yöntemlerinin özniteliklerine göre kıyaslanma durumları gösterilmiştir.

Tablo 2. Sınıflandırma için kullanılan özniteliklerin kıyaslanması (Javed ve Shah, 2002; Hu vd., 2004)

Sınıflandırma Yöntemi	Maliyet	Başarı Oranı	Yorum
Şekil Tabanlı	Düşük	Orta	Işık değişimine hassas. Farklı şekillerde oluşan nesnelerde zayıf.
Hareket Tabanlı	Yüksek	Orta	Farklı karelerdeki nesneleri tanımada başarılı.
Renk Tabanlı	Yüksek	Yüksek	Işık değişimine duyarlı.
Doku Tabanlı	Yüksek	Yüksek	Öznitelik çıkarma aşaması fazla, maliyet yüksek.

4.4. Nesne Takibi

Nesneleri izlemek amacıyla kullanılan birçok farklı algoritma vardır. Bazı algoritmalar yalnızca tek nesne takibini gerçekleştirirken, bazıları birden fazla nesnenin izlenmesi için kullanılmaktadır. Nesne izleme için kullanılan algoritmalar geniş gruplara veya kategorilere ayrılabilir. Yaygın olarak kullanılan nesne takip yöntemleri 3 gruba ayrılmaktadır. Tablo 3'te kullanılan nesne takip yöntemleri ve özellikleri gösterilmiştir.

Tablo 3. Nesne takibi için kullanılan yöntemler (Hanbay ve Üzen, 2017)

Nesne Takip Yöntemi	Kullanımı	İşlem Zamanı	Başarı
Nokta tabanlı	Deterministik model Olasılık modeli	Düşük-Orta Orta-Yüksek	Orta Orta
Çekirdek tabanlı	Çoklu görünüme dayalı Şablon tabanlı	Düşük-Orta Orta	Düşük Orta-Yüksek
Siluet tabanlı	Kontur gelişimi Şekil eşleştirme	Orta Yüksek	Orta-Yüksek Yüksek

Nokta tabanlı yöntem, algılanan nesneler nokta olarak temsil edildiğinde kullanılır ve aydınlatma değişikliklerine duyarlı değildir. Çekirdek tabanlı yöntem ile nesneyi temsil etmek için kare, dikdörtgen gibi şekiller kullanılır. Nesne algılama yalnızca nesne sahnede ilk göründüğünde gereklidir. Katı olmayan nesnelerde arka planın tespit edilen çerçeve içerisinde görünmesi sorunu ortaya çıkabilir. Siluet tabanlı yöntemler nesnenin, silueti kullanılacağı zaman tercih edilir. Nesne algılama yalnızca nesne sahnede ilk göründüğünde gereklidir. Çok çeşitli nesnelerde başarılı bir şekilde takip yapabilir. Nesne takibinde kullanılan fakat yaygın olmayan diğer yollara, bölge tabanlı izleme (Deori ve Meitei, 2014), kontur tabanlı izleme (Jacob and Anitha, 2012) ve sınır tabanlı izleme (Ojha ve Sakhare, 2015) örnek verilebilir.

4.5. Nesne Tespiti ve Takibi Veri Setleri

Uygulamanın içeriğine göre kullanılabilecek veri setleri değişmektedir. ImageNet (Deng vd., 2009), WordNet tabanlı yaklaşık 50 milyon etiketlenmiş temiz görüntü içeren veri kümesidir. Hiyerarşinin her bir düğümü binlerce resimle tasvir edilmektedir. ImageNet ölçek ve çeşitlilik açısından mevcut görüntü veri setlerinden çok daha büyüktür. Nesne tanıma, otomatik nesne kümeleme, görüntü sınıflandırma gibi uygulamalarda başarılı sonuçlar vermektedir.

Caltech101/256 (Griffin vd., 2007), MSRC (Shotton vd., 2006), PASCAL (Everingham vd., 2008), TinyImage (Torralba vd., 2008) küçük görüntülere ait veri setleri içermektedir. Buradaki görüntüler WordNet veri tabanında da saklanmaktadır. Wordnet'ten gelen anlamsal bilgilerle etiketleme gürültüsünün etkilerini en aza indirmek için en yakın komşu yöntemiyle birlikte kullanılır. Nesne sınıflandırması uygulamalarında özellikle insan gibi yaygın olarak kullanılan nesnelerde başarılı sonuçlar alınmaktadır.

ESP (Ahn ve Dabbish, 2004), çevrimiçi oyun aracılığı ile kullanıcılar tarafından etiketlenen resimlerden oluşmaktadır. İnsanlar oyun sırasında onlar için anlamlı etiketler ile görüntülerin içeriklerini belirlerler. Görüntü etiketleme sorununu ele alarak eğlenceli bir şekilde webde bulunan tüm görüntüleri etiketleyebileceklerini tahmin etmektedirler. Çok az bir miktarda etiket ve görüntü kümesi haricinde halka açık değildir.

Lotus Hill ve LabelMe (Russell vd., 2008; Yao vd., 2007), yaklaşık 200 kategoriye sahiptir. Nesnelere ait etiket, konum gibi detaylı özellikleri içermektedir. ImageNet'e kıyasla daha az kategori ve veri setine sahiptir. LabelMe çizim arayüzü sayesinde etiketlenen görüntü özelliklerinin web üzerinde anında paylaşılmasına olanak tanır. Bu şekilde birçok platformdan etiketli görüntü kümeleri toplanmış olur. Lotus Hill, diğer veri etiketleme araçlarından daha farklı bir yöntem izleyerek 3 seviyede bilgiyi gruplamayı amaçlamıştır. Bu araç sayesinde genel amaçlar için kullanılabilecek 636.748'den fazla ek açıklamalı görüntü ve video karesinden oluşan bir veri tabanı oluşturmuşlardır.

MS-COCO (Lin vd., 2014), PASCAL-VOC (Everingham vd., 2010), KITTI (Geiger vd., 2012) silah, tabanca gibi küçük nesne algılama uygulamalarında kullanılan popüler veri kümelerindendir. MS-COCO nesne algılamadaki sahne sorununun ele alınmasıyla yeniden oluşturulan bir veri kümesidir. 91 kategori, yaklaşık 328 bin görüntü ve 2,5 milyon etiketli örnekle nesne, kategori algılama, kümeleme sınıflandırma uygulamalarında kullanılmaktadır. KITTI, otonom bir sürüş platformu kullanarak oluşturulmuş, görüş görevi içeren, nesne algılama uygulamalarında kullanılan veri kümesidir. Veri kümesi, 3B sınırlayıcı kutularla açıklanmış 7481 eğitim görüntüsü içermektedir.

Pornografi-800 (Avila vd. 2013) ve Pornografi-2k (Moreira vd., 2016) ise cinsel içerikleri tespit etmek için kullanılabilecek veri kümesidir. Pornografi-2k Pornografi 800'ün bir uzantısıdır. Pornografi 800, yaklaşık 80 saat, 800 video, 400 pornografik ve 400 pornografik olmayan içerir. Pornografi içeriğine sahip videolar, bu tür içerik üzerinde uzmanlaşmış, çok çeşitli türlerde

örnekler arayan ve farklı etnik kökenlerden aktörlerle oluşturulmuş videolardan alınmıştır. Pornografi-2k, 6 saniyeden 33 dakikaya kadar değişen yaklaşık 140 saat, 1000 pornografik ve 1000 pornografik olmayan videodan oluşur. Herkese açık bir veri setidir.

5. Tartışma ve Sonuç

Bu çalışmada derin öğrenme yöntemleri kullanılarak nesne tespiti için yapılan uygulamalar incelenmiştir. Özellikle Konvolüsyonel Sinir Ağı tabanlı yöntemler ve Tekrarlı Sinir Ağı tabanlı yöntemlere ağırlık verilmiştir. Bu çalışmalara göre kullanılan mimariler, ağlar, kütüphaneler, algoritmalar, veri setleri ve modeller hakkında bilgiler sunulmuştur. Literatüre göre nesne tespit etmek için yaygın olarak kullanılan adımlar ön işleme, nesne tespiti, nesne sınıflandırma ve nesne takibidir. Ön işleme aşaması, tespit edilecek nesneye ait veri setinin bir sonraki adım için hazır hale getirilmesi işlemlerinden oluşur. Belirlenen nesneye göre kullanılacak olan veri seti kümesi de değişmektedir. Örneğin küçük nesneler için TinyImage veri seti kullanılabilirken, cinsel unsurları tespit etmek için Pornografi-800 kullanılabilir. Oluşturulacak olan model, nesneye ait kendi çektiğimiz görüntüler kullanılarak eğitilebilir. İncelenen çalışmalarda nesne tespiti aşamasında en çok YOLO, Faster R-CNN ve SSD algoritmaları kullanılmıştır. Modeli eğitirken verinin fazla olması başarıyı arttıracaktır fakat işlem zamanını da arttırdığı için yavaşlamaya neden olacaktır. Özellikle R-CNN ve Fast R-CNN'nin çalışma yöntemine bakıldığında ilk aşamada nesneleri sınıflandırıp ardından konvolüsyonel sinir ağı uyguladığı için yavaş çalışmaktadırlar. Bu algoritmalardaki hız sorununu çözmek için ise Faster R-CNN kullanılmaya başlanmıştır. SSD ve Faster R-CNN karşılaştırıldığında ise SSD, tek bir sinir ağı kullanarak nesne tespiti yaptığı için Faster R-CNN'den daha hızlı çalışmaktadır. Faster R-CNN'nin başarı oranı SSD algoritmasına göre daha yüksektir. Hem yüksek başarı oranı hem de hızın önemli olduğu uygulamalarda ise son yıllarda popüler olan YOLO kullanılmaktadır. Nesne sınıflandırma aşamasında tespit edilecek nesnenin özniteliklerine göre sınıflandırma yöntemi de değişiklik göstermektedir. Renk ve doku tabanlı yöntemlerin daha yüksek başarı oranı sağladığı görülmüştür. Nesne takibi sırasında ise Kalman Filtresi yöntemi daha sık kullanılmıştır. İzlenecek nesneye göre uygulanacak sinir ağı modeli de değişiklik gösterecektir. Çoklu nesne izleme yapılacaksa Deep-Sort, Deep, Multicut algoritmaları, uygunsuz görüntülerin tespiti için de ResNet-50 kullanılması önerilmiştir.

Kaynakça

Amidi, A., Amidi, S. (2020). Derin Öğrenme El Kitabı. Derin Öğrenme El Kitabı: https://stanford.edu/~shervine/l/tr/teaching/cs-229/cheatsheet-deep-learning adresinden alındı.

Avidan, S. (2004). Support Vector Tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1064-1072, doi:10.1109/TPAMI.2004.53.

Avila, S., Thome, N., Cord, M., Valle, E., De A. Araújo, A. (2013). Pooling in image representation: The visual codeword point of view. Computer Vision and Image Understanding, 453-465, doi: 10.1016/j.cviu.2012.09.007.

Bewley, A., Ge, Z., Ott, L., Ramos, F., Upcroft, B. (2016). Simple online and realtime tracking. Proceedings - International Conference on Image Processing, ICIP, 3464–3468, doi: 10.1109/ICIP.2017.8296962.

Bolouri, H. (1995). Book Review: Fundamentals of Neural Networks — Architectures, Algorithms, and Applications: L. FAUSETT. International Journal of Electrical Engineering Education, doi: 10.1177/002072099503200320.

Brocardo, M., Traore, I., Woungang, I., Obaidat, M. (2017). Authorship verification using deep belief network systems. International Journal of Communication Systems, doi: 10.1002/dac.3259.

Brunetti, A., Buongiorno, D., Trotta, G., Bevilacqua, V. (2018). Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. Neurocomputing, doi: 10.1016/j.neucom.2018.01.092.

Cbinsights, C. (2019). The Race For AI: Here Are The Tech Giants Rushing To Snap Up Artificial Intelligence Startups. 2020 tarihinde CBINSIGHTS adresinden alındı.

Chaudhary, S., Khan, M., Bhatnagar, C. (2018). Multiple Anomalous Activity Detection in Videos. Procedia Computer Science, 336-345, doi: 10.1016/j.procs.2017.12.045.

Cheng, X., Song, C., Gu, Y., Chen, B. (2020). Learning Attention for Object Tracking with Adversarial Learning Network, doi: 10.21203/rs.3.rs-15512/v3.

Ciaparrone, G., Luque Sánchez, F., Tabik, S., Troiano, L., Tagliaferri, R., Herrera, F. (2020). Deep learning in video multi-object tracking: A survey. Neurocomputing, 61-88, doi: 10.1016/j.neucom.2019.11.023.

Collobert, R., Farabet, C., Kavukcuoğlu, K. (2017). Torch | Scientific computing for LuaJIT. NIPS Workshop on Machine Learning Open Source Software.

Cortes, C., Vapnik, V. (1995). Support-Vector Networks. Machine Learning, doi: 10.1023/A:1022627411411.

Cömert, O., Hekim, M., Adem, K. (2019). Faster R-CNN Kullanarak Elmalarda Çürük Tespiti. Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi, 335-341, doi: 10.29137/umagd.469929.

Daş, R., Polat, B., Tuna, G. (2019). Derin Öğrenme ile Resim ve Videolarda Nesnelerin Tanınması ve Takibi. Fırat Üniversitesi Müh. Bil. Dergisi, 571-581, doi: 10.35234/fumbd.608778.

DeepLearning4j. (2020). DeepLearning4j: https://deeplearning4j.org/ adresinden alındı.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Li, F. (2009). ImageNet: a Large-Scale Hierarchical Image Database. 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25, doi: 10.1109/cvpr.2009.5206848.

Deng, L., Yu, D. (2013). Deep learning: Methods and applications. Foundations and Trends in Signal Processing, 197-387, doi: 10.1561/2000000039.

Deori, B., Meitei, D. (2014). A survey on moving object tracking in video. International Journal on Information Theory, 31-46, doi: 10.5121/ijit.2014.3304.

Everingham, M., Van Gool, L., Williams, C., Winn, J., Zisserman, A. (2008). The PASCAL Visual Object Classes Challenge 2008 (VOC) Results. http://www.pascal-network.org/ challenges/VOC/voc2008/workshop/index.html adresinden alındı.

Fu, K., Li, Y., Sun, H., Yang, X., Xu, G., Li, Y., Sun, X. A. (2018). Ship Rotation Detection Model in Remote Sensing Images Based on Feature Fusion Pyramid Network and Deep Reinforcement Learning. Remote Sens., 10, 1922.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, 193-202, doi: 10.1007/BF00344251.

Geiger, A., Lenz, P., Urtasun, R. (2012). Are we ready for autonomous driving? the KITTI vision benchmark suite. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 3354-3361, doi: 10.1109/CVPR.2012.6248074.

Gibney, E. (2016). Google AI algorithm masters ancient game of Go. Nature, doi: 10.1038/529445a.

Girshick, R. (2015). Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 1440-1448, doi: 10.1109/ICCV.2015.169.

Girshick, R., Donahue, J., Darrell, T., Malik, J. (2016). Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 142 – 158, doi: 10.1109/TPAMI.2015.2437384.

Griffin, G., Holub, A., Perona, P. (2007). Caltech-256 object category dataset. Caltech mimeo.

Hanbay, K., Üzen, H. (2017). Nesne tespit ve takip metotları: Kapsamlı bir derleme. Tr. Doğa ve Fen Dergisi.

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., . . . Larochelle, H. (2017). Brain tumor segmentation with Deep Neural Networks. Medical Image Analysis, 18-31, doi: 10.1016/j.media.2016.05.004.

He, K., Gkioxari, G., Dollar, P., Girshic, R. (2017). Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2961-2969, doi: 10.1109/ICCV.2017.322.

Hinton, G., Osindero, S., Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. Neural Computation, 1527-1554, doi: 10.1162/neco.2006.18.7.1527.

Hochreiter, S., Schmidhuber, J. (1997). LONG SHORT-TERM MEMORY. Neural Computation, 1735-1780, doi: 10.1162/neco.1997.9.8.1735.

Iliadis, M., Spinoulas, L., Katsaggelos, A. (2020). DeepBinaryMask: Learning a binary mask for video compressive sensing. Digital Signal Processing: A Review Journal, doi: 10.1016/j.dsp.2019.102591.

Ivakhnenko, A., Lapa, V. (1965). Cybernetic predicting devices. CCM Information Corporation.

Jacob, A., Anitha, J. (2012). Inspection of various object tracking techniques. International Journal of Engineering and Innovative Technology, 118-124, doi: 10.17605/OSF.IO/Y5K3H.

Jia, Y., Shelhame, E., Donahue, J., Karayev, S., Long, J., Girshick, R., . . . Darrell, T. (2014). Caffe: Convolutional Architecture for Fast Feature Embedding. Computer Science > Computer Vision and Pattern Recognition, doi: 10.1145/2647868.2654889.

Keras: The Python Deep Learning API. (2020). Keras: The Python Deep Learning API: https://keras.io/ adresinden alındı.

Kim, D. (2020). Deeplearning Method For Voice Recognition Model And Voice Recognition Device Based On Artificial Neural Network.

Kim, S., Nam, J., Ko, B. (2018). Online Tracker Optimization for Multi-Pedestrian Tracking Using a Moving Vehicle Camera. IEEE Access, 48675-48687, doi: 10.1109/ACCESS.2018.2867621.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4), 541-551.

Leijnen, S., Veen, F. (2020). The Neural Network Zoo. Proceedings, 9.

Li, D., Wang, R., Xie, C., Liu, L., Zhang, J., Li, R., . . . Liu, W. (2020). A recognition method for rice plant diseases and pests video detection based on deep convolutional neural network. Sensors, Switzerland, doi: 10.3390/s20030578.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., . . . Zitnick, C. (2014). Microsoft COCO: Common Objects in Context. European Conference on Computer Vision, 740-755.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., C. Berg, A. (2016). SSD: Single Shot MultiBox Detector. European Conference on Computer Vision, 21-37, doi: 10.1007/978-3-319-46448-0_2.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F. (2017). A survey of deep neural network architectures and their applications. Neurocomputing, doi: 10.1016/j.neucom.2016.12.038.

Lu, Y., Lu, C., Tang, C. (2017). Online Video Object Detection Using Association LSTM. Proceedings of the IEEE International Conference on Computer Vision, 2344-2352, doi: 10.1109/ICCV.2017.257.

Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., Zhao, X., Kim, T.-K. (2014). Multiple Object Tracking: A Literature Review. Computer Vision and Pattern Recognition, doi: 10.1016/j.artint.2020.103448.

Mark, E., Van Gool, L., Williams, C., Winn, J., Zisserman, A. (2010). The pascal visual object classes (VOC) challenge. International Journal of Computer Vision, 303-338, doi: 10.1007/s11263-009-0275-4.

McCulloch, W., Pitts, W. (1988). A logical calculus of the ideas immanent in nervous activity. Neurocomputing, 15-27, doi: 10.1016/s0092-8240(05)80006-0.

Medsker, L., Jain, L. (2001). Recurrent Neural Network Design and Applications. Boca Raton London New York Washington, D.C.: CRC Press.

Mikada, T., Kanno, T., Kawase, T., Miyazaki, T., Kawashima, K. (2020). Suturing Support by Human Cooperative Robot Control Using Deep Learning. IEEE Access, 167739-167746, doi: 10.1109/ACCESS.2020.3023786.

Moreira, D., Avila, S., Perez, M., Moraes, D., Testoni, V., Valle, E., . . . Rocha, A. (2016). Pornography classification: the hidden clues. Forensic Science International, 46-61, doi: 10.1016/j.forsciint.2016.09.010.

Nawaratne, R., Alahakoon, D., De Silva, D., Yu, X. (2020). Spatiotemporal anomaly detection using deep learning for real-time video surveillance. IEEE Transactions on Industrial Informatics, doi: 10.1109/TII.2019.2938527.

NVIDIA DIGITS. (2020). NVIDIA DEVELOPER: https://developer.nvidia.com/digits adresinden alındı.

Omeroglu, A., Kumbasar, N., Oral, E., Ozbek, I. (2019). Mask R-CNN Algoritması ile Hangar Tespiti. 27th Signal Processing and Communications Applications Conference (SIU), doi: 10.1109/siu.2019.8806552.

O'Shea, K., Nash, R. (2015). An Introduction to Convolutional Neural Networks. Neural and Evolutionary Computing.

Ojha, S., Sakhare, S. (2015). Image processing techniques for object tracking in video surveillance-a survey. In Pervasive

Computing (ICPC), 2015 International Conference on, 1-6, doi: 10.1109/PERVASIVE.2015.7087180.

Ozbaysar, E., Borandag, E. (2018). Vehicle plate tracking system. 26th IEEE Signal Processing and Communications Applications Conference, 1-4, doi: 10.1109/SIU.2018.8404648.

Perez, M., Avila, S., Moreira, D., Moraes, D., Testoni, V., Valle, E., . . . Rocha, A. (2017). Video pornography detection through deep learning techniques and motion information. Neurocomputing, 279-293, doi: 10.1016/j.neucom.2016.12.017.

Prasad, P., Pathak, R., Gunjan, V., Rao, H. (2019). Deep Learning Based Representation for Face Recognition. ICCCE, 419-424, doi: 10.1007/978-981-13-8715-9_50.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779-788, doi: 10.1109/CVPR.2016.91.

Ren, S., He, K., Girshick, R., Sun, J. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1137 – 1149, doi: 10.1109/TPAMI.2016.2577031.

Russell, B., Torralba, A., Murphy, K., Freeman, W. (2008). LabelMe: A database and web-based tool for image annotation. International Journal of Computer Vision, 157-173, doi: 10.1007/s11263-007-0090-8.

Salakhutdinov, R., Hinton, G. (2009). Replicated softmax: An undirected topic model. Advances in Neural Information Processing Systems 22 - Proceedings of the 2009 Conference, 1607-1614.

Saleem, M., Potgieter, J., Arif, K. (2019). Plant Disease Detection and Classification by Deep Learning. Plants, 8(11):468, doi: doi:10.3390/plants8110468.

Sarkar, M., Bruyn, A. (2020). LSTM Response Models for Direct Marketing Analytics:Replacing Feature Engineering with Deep Learning. Forthcoming in Journal of Interactive Marketing, doi: 10.2139/ssrn.3601025.

Sejnowsk, T., R. Rosenberg, C. (1986). NETtalk: a parallel network that learns to read aloud. The Johns Hopkins University Electrical Engineering and Computer Science Technical Report.

Shetty, D., Varma, J., Navi, S., Ahmed, M. (2020). Diving Deep into Deep Learning: History, Evolution, Types and Applications. The International Journal on Media Management, 2278-3075, doi: 10.35940/ijitee.A4865.019320.

Shotton, J., Winn, J., Rother, C., Criminisi, A. (2006). TextonBoost: Joint Appearance, Shape and Context Modeling for Multi-class Object Recognition and Segmentation. European Conference on Computer Vision, 1-15, doi: 10.1007/11744023_1.

Singla, Z., Randhawa, S., Jain, S. (2017). Statistical and sentiment analysis of consumer product reviews. 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 1-6, doi: 10.1109/ICCCNT.2017.8203960.

Sorin, V., Barash, Y., Konen, E., Klang, E. (2020). Deep Learning for Natural Language Processing in Radiology—Fundamentals and a Systematic Review. Journal of the American College of Radiology, 639-648.

Sun, W., Zheng, B., Qian, W. (2017). Automatic feature learning using multichannel ROI based on deep structured algorithms for computerized lung cancer diagnosis. Computers in Biology and Medicine, 530-539, doi: 10.1016/j.compbiomed.2017.04.006.

Şeker, A., Diri, B., Balık, H. (2017). Derin Öğrenme Yöntemleri ve Uygulamaları Hakkında Bir İnceleme. Gazi Mühendislik Bilimleri Dergisi, 47-64.

TensorFlow. (2020). TensorFlow: https://www.tensorflow.org/ adresinden alındı.

Theano Development Team. (2016). Theano: A Python framework for fast computation of mathematical expressions. arXiv e-prints, vol. abs/1605.02688.

Toğaçar, M., Ergen, B. (2019). Biyomedikal Görüntülerde Derin Öğrenme ile Mevcut Yöntemlerin Kıyaslanması. Fırat Üniversitesi Mühendislik Bilimleri Dergisi.

Toğaçar, M., Ergen, B., Sertkaya, M. (2019). Zatürre Hastalığının Derin Öğrenme Modeli ile Tespiti. Fırat Üniversitesi Mühendislik Bilimleri Dergisi.

Torralba, A., Fergus, R., Freeman, W. (2008). 80 million tiny images: A large data set for nonparametric object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1958-1970, doi: 10.1109/TPAMI.2008.128.

Turing, A. (2012). Computing machinery and intelligence. Machine Intelligence: Perspectives on the Computational Model, 433-460, 10.1093/mind/lix.236.433.

Verma, G., Dhillon, A. (2017). A Handheld Gun Detection using Faster R-CNN Deep Learning. ACM International Conference Proceeding Series, 84-88. doi:10.1145/3154979.3154988.

Vignesh Kanna, J., Ebenezer Raj, S., Meena, M., Meghana, S., Mansoor Roomi, S. (2020). Deep Learning Based Video Analytics for Person Tracking. International Conference on Emerging Trends in Information Technology and Engineering, ic-ETITE 2020, doi: 10.1109/ic-ETITE47903.2020.173.

Viola, P., Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, doi: 10.1109/cvpr.2001.990517.

Von Ahn, L., Dabbish, L. (2004). Labeling images with a computer game. Conference on Human Factors in Computing Systems - Proceedings, 319-326, doi: 10.1145/985692.985733.

Wallach, I., Dzamba, M., Heifets, A. (2015). AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery.

Wang, D. (1998). Unsupervised video segmentation based on watersheds and temporal tracking. IEEE Transactions on Circuits and Systems for Video Technology, 539-546, doi: 10.1109/76.718501.

Yang, S., Bailey, E., Yang, Z., Ostrometzky, J., Zussman, G., Seskar, I., Kostic, Z. (2020). COSMOS Smart Intersection: Edge Compute and Communications for Bird's Eye Object Tracking. IEEE Annual Conference on Pervasive Computing and Communications Workshops (PerCom), 1-7, doi: 10.1109/PerComWorkshops48775.2020.9156225.

Yang, X., Sun, H., Fu, K., Yang, J., Sun, X., Yan, M., Guo, Z. (2018). Automatic Ship Detection in Remote Sensing Images from Google Earth of Complex Scenes Based on Multiscale Rotation Dense Feature Pyramid Networks. Remote Sens., 10, 132.

Yao, B., Yang, X., Zhu, S. (2007). Introduction to a large-scale general purpose ground truth database: Methodology, annotation tool and benchmarks. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 169-183, doi: 10.1007/978-3-540-74198-5_14.

You, L., Li, Y., Wang, Y., Zhang, J., Yang, Y. (2016). A deep learning based RNNs model for automatic security audit of short messages. 2016 16th International Symposium on Communications and Information Technologies, ISCIT 2016, doi: 10.1109/ISCIT.2016.7751626.

Yousefi-Azar, M., Hamey, L. (2017). Text summarization using unsupervised deep learning. Expert Systems with Applications, doi: 10.1016/j.eswa.2016.10.017.

Yu, F., Li, W., Li, Q., Liu, Y., Shi, X., Yan, J. (2016). POI: Multiple object tracking with high performance detection and appearance feature. European Conferance on Computer Vision, 36-42, doi: 10.1007/978-3-319-48881-3_3.

Yuret, D. (2016). Julia ve Knet ile Derin Öğrenmeye Giriş. Julia ve Knet ile Derin Öğrenmeye Giriş: http://www.denizyuret.com/2016/09/julia-ve-knet-ile-derin-ogrenmeye-giris.html adresinden alındı.

Zhang, L., Gray, H., Ye, X., Collins, L., Allinson, N. (2019). Automatic individual pig detection and tracking in pig farms. Sensors (Switzerland), doi: 10.3390/s19051188.

Zhao, D., Fu, H., Xiao, L., Wu, T., Dai, B. (2018). Multi-object tracking with correlation filter for autonomous vehicle. Sensors (Switzerland), doi: 10.3390/s18072004.