

T.C.
AKDENİZ ÜNİVERSİTESİ



**VIDEO GÖRÜNTÜLERİ ÜZERİNDE DERİN ÖĞRENME TABANLI
ANORMAL DURUM TESPİTİ VE DERİNLİK TAHMİNİ**

Nuri ÖZÇELİK

**FEN BİLİMLERİ ENSTİTÜSÜ
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ
ANABİLİM DALI
YÜKSEK LİSANS TEZİ**

TEMMUZ 2021

ANTALYA

T.C.
AKDENİZ ÜNİVERSİTESİ



**VIDEO GÖRÜNTÜLERİ ÜZERİNDE DERİN ÖĞRENME TABANLI
ANORMAL DURUM TESPİTİ VE DERİNLİK TAHMİNİ**

Nuri ÖZÇELİK

**FEN BİLİMLERİ ENSTİTÜSÜ
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ
ANABİLİM DALI
YÜKSEK LİSANS TEZİ**

TEMMUZ 2021

ANTALYA

T.C.
AKDENİZ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

VIDEO GÖRÜNTÜLERİ ÜZERİNDE DERİN ÖĞRENME TABANLI
ANORMAL DURUM TESPİTİ VE DERİNLİK TAHMİNİ

Nuri ÖZÇELİK

ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ

ANABİLİM DALI

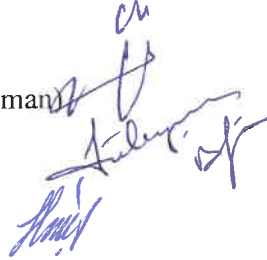
YÜKSEK LİSANS TEZİ

Bu tez 01.07./2021 tarihinde jüri tarafından Oybirliği / Oyçokluğu ile kabul edilmiştir.

Doç. Dr. Övünç POLAT (Danışman)

Doç. Dr. Süleyman BİLGİN

Dr. Öğr. Üyesi Hakan AKTAŞ



ÖZET

VIDEO GÖRÜNTÜLERİ ÜZERİNDE DERİN ÖĞRENME TABANLI ANORMAL DURUM TESPİTİ VE DERİNLİK TAHMİNİ

Nuri ÖZÇELİK

Yüksek Lisans Tezi, Elektrik Elektronik Mühendisliği Anabilim Dalı

Danışman: Doç. Dr. Övünç POLAT

Temmuz 2021; 43 sayfa

Bu tez çalışmasında, derin öğrenme metotları kullanılarak, güvenlik kameralarına yansıyan anormal durumların tespit edilmesi, anormal durumun görüntü içerisinde konumlandırılması ve derinlik tahmin modeli ile anormal durumun mesafesinin kestirimi amaçlanmıştır. Anormal durumun gerçekleştiği anda tespit edilebilmesi amacıyla 3 boyutlu evrişimsel sinir ağı ve devamında tam bağlantılı sinir ağı kullanılmıştır. Anormal durum tespit sisteminde bulunan 3 boyutlu evrişimsel sinir ağı modeli sports-1m veri seti ile önceden eğitilmiş ve eğitim sonucu elde edilen ağırlıklar anormal durum tespiti için özellik çıkartma sürecinde kullanılmıştır. 3 boyutlu evrişimsel sinir ağı modelinden elde edilen özellikler tam bağlantılı sinir ağına uygulanmıştır. Konumlandırma işlemi için bir veri seti oluşturulmuştur ve bu veri seti ile konumlandırma modeli eğitilmiştir. Girdi videosundan belirli aralıklarla elde edilen görüntüler konumlandırma tahmini için MobileNet V2 modeli ve bu modele sonradan eklenen yukarı örnekleme katmanlarına uygulanmıştır, bu sayede anormal durum gerçekleştiği esnada ilgili anormal durum görüntü içerisinde konumlandırılabilmiştir. Konumlandırma modeline uygulanan girdi görüntüleri aynı zamanda derinlik tahmin modeline de uygulanmaktadır. Derinlik tahmini için kullanılan sistemde ImageNet veri seti ile önceden eğitilmiş olan DenseNet-169 modeli enkoder olarak çalışmaktadır. DenseNet-169 modelinin çıkışına yukarı örnekleme yapacak olan dekoder modeli ilave edilmiştir ve 320x240 boyutunda derinlik haritası üretilmiştir. Derinlik tahmin modeli NYU-Depth V2 veri seti ile önceden eğitilmiş bir modeldir. Tez kapsamında geliştirilen sistemde, anormal durum gerçekleştiği anda konumlandırma modeli çalışarak anormal durumu konumlandırmakta ve derinlik tahmin modeli ile olayın gerçekleştiği mesafe tahmin edilmektedir. Tez çalışmasında, video görüntüleri kullanılarak anormal durum tespiti, anormal durum konumlandırma ve anormal durumun mesafesinin tahmini aynı anda gerçekleştirilmektedir, bu özelliği tez çalışmasını özgün bir çalışma yapmaktadır. Çalışmada anormal durum tespiti aşamasında farklı optimizasyon yöntemlerinin etkisi, konumlandırma aşamasında farklı filtre boyutlarının etkisi gibi sistem performansına etki eden ilgili parametrelerin belirlenmesi işlemleri de gerçekleştirilmiş ve sonuçları ilgili bölümde tartışılmıştır. Eğitim sonucu elde edilen veriler ve bu eğitim parametrelerine bağlı olarak farklı video görüntüleri üzerinde elde edilen sonuçlar da bulgular ve tartışma kısmında ele alınmıştır.

Tez çalışması sonucu geliştirilen sistem, otonom araçlarda seyir halinde iken gerçekleşen trafik kazası vb. anormal durumu tespit edebilecek, konumlandırabilecek ve

mesafesini tahmin edebilecektir, ayrıca otobüs durakları, meydanlar, kavşaklar ve diğer asayişin sağlanması gerektiği alanlarda kullanılabilir.

ANAHTAR KELİMELER: Anormal durum lokalizasyon, Anormal durum tahmini, Derin öğrenme, Derinlik tahmini, Yapay sinir ağları

JÜRİ: Doç. Dr. Övünç POLAT

Doç. Dr. Süleyman BİLGİN

Dr. Öğr. Üyesi Hakan AKTAŞ

ABSTRACT
**DEEP LEARNING BASED ABNORMAL SITUATION DETECTION AND
DEPTH ESTIMATION ON VIDEO IMAGES**

Nuri ÖZÇELİK

MSc Thesis in Electrical and Electronics Engineering

Supervisor: Assoc. Prof. Dr. Övünç POLAT

July 2021; 43 pages

This thesis, it is aimed to detect abnormal situations recorded by security cameras, locate the abnormal situation in the image, and estimate the distance of the abnormal situation with the depth estimation model by using the deep learning methods. In order to detect the abnormal situation as soon as it occurs, a 3D convolutional neural network and in the continuation a fully connected neural network are used. The 3D convolutional neural network model in the abnormal situation detection system was pre-trained with the sports-1m dataset and the weights obtained as a result of the training were used in the feature extraction process for abnormal situation detection. The features obtained from the 3D convolutional neural network model were applied to the fully connected neural network. A dataset was created for the positioning process and the localization model was trained with this dataset. The images obtained from the input video at certain intervals were applied to the MobileNet V2 model and the upsampling layers were added to this model for localization estimation so that the related anomaly could be located in the image during the abnormal situation. Input images applied to the localization model are also applied to the depth estimation model. In the system used for depth estimation, the DenseNet-169 model, which was pre-trained with the ImageNet dataset, works as an encoder. Upsampling decoder model was added to the output of the DenseNet-169 model and a depth map of 320x240 size was produced. The depth estimation model is a pre-trained model with the NYU-Depth V2 dataset. In the system developed within the scope of the thesis, the positioning model works as soon as the abnormal situation occurs, and the depth estimation model estimates the distance at which the event occurs. In the thesis study, abnormal situation detection, abnormal situation localization, and estimation of the distance of the abnormal situation are performed simultaneously using video images, this feature makes the thesis work a unique study. In the study, the determination of the relevant parameters that affect the system performance, such as the effect of different optimization methods in the anomaly detection stage, the effect of different filter sizes in the localization stage, were also carried out and the results were discussed in the relevant section. The data obtained as a result of the training and the results obtained on different video images depending on these training parameters are also discussed in the discussion section.

The system can be detected as a result of the thesis study, traffic accidents, etc. while driving in autonomous vehicles. It will be able to detect, locate and estimate the abnormal situation, and can also be used in bus stops, squares, intersections, and other areas where security is required.

KEYWORDS: Anomaly localization, Anomaly prediction, Artificial neural networks, Deep learning, Depth estimation

COMMITTEE: Assoc. Prof. Dr. Övünç POLAT

Assoc. Prof. Dr. Süleyman BİLGİN

Assist. Prof. Dr. Hakan AKTAŞ

ÖNSÖZ

Tez konusunun belirlenmesinden sonuçlanmasına kadar bütün süreçte bilgi ve tecrübesiyle bana yol gösteren, değerli bilgi birikimini ve vakitlerini esirgemeyen danışman hocam Sayın Doç. Dr. Övünç POLAT'a sonsuz teşekkürlerimi ve saygılarımı sunarım.

Tez çalışmam ve yüksek lisans eğitimim için gerekli zamanı ve desteği benden esirgemeyen değerli yöneticim Sayın Fırat ATAŞ'a saygı ve teşekkürlerimi sunarım.

Hayatım boyunca koşulsuz şartsız beni destekleyerek olduğum kişi olmamı sağlayan çok sevgili anneme, aynı şekilde çok sevgili ablama sevgi ve teşekkürlerimi sunarım.

İÇİNDEKİLER

ÖZET.....	i
ABSTRACT.....	iii
ÖNSÖZ	v
AKADEMİK BEYAN	viii
SİMGELER VE KISALTMALAR.....	ix
ŞEKİLLER DİZİNİ.....	xi
ÇİZELGELER DİZİNİ	xiii
1. GİRİŞ	1
2. KAYNAK TARAMASI	3
2.1. Kalabalık Analizinde Kullanılan Yaklaşımlar	3
2.1.1. Kalabalık yoğunluğu tahmini, takibi ve davranışların tanımlanması	3
2.1.2. Kalabalık ortamlarda gerçekleşen anormal olayların tespiti	8
2.2. Kalabalık Ortam Veri setleri	9
2.2.1. Ucf-crime veri seti	9
2.2.2. Ucsd veri seti	10
2.2.3. Umn veri seti.....	10
2.2.4. Pets veri seti.....	11
2.2.5. Subway veri seti.....	11
2.2.6. Avenue veri seti	11
2.2.7. U-turn veri seti.....	11
2.2.8. Web veri seti.....	12
2.2.9. Caviar veri seti	12
2.3. Derinlik Tespiti	12
2.3.1. Derinlik tespitinde kullanılan literatürdeki yöntemler.....	12
2.3.2. Derinlik tespiti için kullanılan literatürdeki veri setleri.....	13
2.4. Görüntü İçerisinde Nesne Lokalizasyonu	14
3. MATERYAL VE METOT	16
3.1. Yapay Sinir Ağları.....	16
3.2. Evrişimsel Sinir Ağları.....	17
3.2.1. Evrişimsel sinir ağları hiper-parametreleri	17
3.2.1.1. Evrişimsel sinir ağlarında alıcı alan.....	18

3.2.1.2. Evrişimsel sinir ağlarında adım	18
3.2.1.3. Evrişimsel Sinir ağlarında sıfır dolgu	18
3.3. Anormal Durum Tespiti	19
3.4. Derinlik Tahmini	21
3.5. Anomali Lokalizasyon	22
3.6. Geliştirilen Sistem	25
4. BULGULAR VE TARTIŞMA	27
5. SONUÇLAR	37
6. KAYNAKLAR	38
ÖZGEÇMİŞ	

AKADEMİK BEYAN

Yüksek Lisans Tezi olarak sunduğum “Video Görüntüleri Üzerinde Derin Öğrenme Tabanlı Anormal Durum Tespiti Ve Derinlik Tahmini” adlı bu çalışmanın, akademik kurallar ve etik değerlere uygun olarak yazıldığını belirtir, bu tez çalışmasında bana ait olmayan tüm bilgilerin kaynağını gösterdiğimi beyan ederim.

01/07/2021

Nuri ÖZÇELİK



SİMGELER VE KISALTMALAR

Simgeler

b	: Bias değeri
B_a	: Anormal örneklerin bulunduğu çanta
BC	: Binari Çapraz Entropi
B_n	: Normal örneklerin bulunduğu çanta
$f(V_a^i)$: Tahmin edilen anormal skor
$f(V_n^i)$: Tahmin edilen normal skor
k	: Örnek sayısı
min_w	: Minimum ağırlık değeri
n	: Toplam anomali örnek sayısı
w	: Ağırlık değeri
y_i	: Her örneğin etiketi
$f(y_{true})$: Hedef lokalizasyon değeri
$f(y_{pred})$: Tahmin edilen lokalizasyon değeri

Kısaltmalar

3D	: 3 Boyutlu
ACAN	: Attention-Based Context Aggregation Network
C3D	: 3 Boyutlu Evrişim (Convolutional 3D)
CAM	: Durum Dikkat Modülü (Context Attention Module)
CNN	: Evrişimsel Sinir Ağı
COVID-19	: Koronavirüs hastalığı 2019
D-INCSFA	: Derin Artımlı Yavaş Özellik Analizi
DORN	: Derin Sıralı Regresyon Ağı (Deep Ordinal Regression Network)

HOG	: Gradyan. Oryantasyon Histogram (Histogram of Oriented Gradients)
KLT	: Kanade–Lucas–Tomasi
KNN	: K-En Yakın Komşu
MCCNN	: Çok Kanallı Evrişimsel Sinir Ağı
MIT	: Massachusetts Teknoloji Enstitüsü
MIL	: Multiple Instance Learning
MOC	: Mixture of CNN
NWPU	: Northwestern Polytechnical University
R-CNN	: Bölge Tabanlı Evrişimsel Sinir Ağları
RELU	: Doğrultulmuş Lineer Birim
RESNET	: Artık Değerli Sinir Ağı (Residual Neural Network)
RGB	: Kırmızı- Yeşil- Mavi
SIFT	: Scale-Invariant Feature Transform
SOM	: Kendini Düzenleyen Haritalar (Self-Organizing Maps)
UCSD	: Kaliforniya San Diago Üniversitesi
YOLO	: Sadece Bir Kez Bak (You Only Look Once)

ŞEKİLLER DİZİNİ

Şekil 1.1. Kalabalıkla Alakalı Araştırma Yaklaşımları	1
Şekil 2.1. Shao vd. (2015) Görünüm ve hareket bilgisini girdi olarak alan çift kanal derin model.....	4
Şekil 2.2. Boominahtan vd. (2016) Sığ ve derin sinir ağından oluşan yoğunluk ve kişi sayısı tahmin modeli	5
Şekil 2.3. Zhang vd. (2016) Farklı filtre boyutlarına sahip MCCNN yapısı	5
Şekil 2.4. a) Kumagai vd. (2017) Farklı yoğunluk ve ölçeğe sahip, ayrıca engel arkasında kalan kalabalık insanların görüntüsü. b) MoC-CNN Modeli	6
Şekil 2.5. Ding vd. (2020) Tanıtılan kalabalık sayımı için simetrik mimari. Turuncu renkli kısım enkoder, mor kısım ise dekoder kısmını ifade etmektedir	7
Şekil 2.6. Wang vd. (2020) Farklı renk, ışık, perspektif, ışık koşulları içeren NWPU veri setinden bazı görseller.	8
Şekil 2.7. Veri setinden patlama anormal durumu örneği	10
Şekil 2.8. UCSD Veri setinden Peds1 verisetine ait yaya ve araç görüntüsü	10
Şekil 2.9. a) UMN Veri seti normal durumları. b) UMN Veri seti kaçışma durumları.	11
Şekil 2.10. HOG metodu uygulanmış görseller (N. Dalal ve B. Triggs, 2005).....	14
Şekil 3.1. Basit bir yapay sinir ağı gösterimi.	16
Şekil 3.2. Basit bir yapay sinir ağı örneği (Karahana M. 2015)	17
Şekil 3.3. Evrimsel sinir ağı yapısı (Gülcü, A. , Kuş, Z. 2019)	18
Şekil 3.4. Anormal durum tespit mimarisi (W. Sultani vd. 2018).....	19
Şekil 3.5. C3D Mimarisi Du Tran vd. (2014).....	20
Şekil 3.6. Girdi resmi, katmanları ve derinlik haritası (Alhashim, I. vd. (2018)).....	21
Şekil 3.7. Girdi resmi, tahmin edilen derinlik haritası, DORN modelinin çıktısı (Alhashim, I. vd. (2018)).....	22
Şekil 3.8. Dice katsayısının gösterimi(Zou vd. 2004)	23
Şekil 3.9. LabelIMG Programı.....	24
Şekil 3.10. Patlama ve trafik kazası etiketleme işlemi.....	24
Şekil 3.11. Geliştirilen modelin akış diyagramı	25

Şekil 4.1. Dice katsayısı lokalizasyon performansı	27
Şekil 4.2. Konumlandırma ilk eğitim aşaması boyunca öğrenme oranı	27
Şekil 4.3. Konumlandırma ilk eğitim aşaması boyunca kayıp fonksiyonu çıktısı.....	28
Şekil 4.4. Konumlandırma ikinci eğitim aşaması boyunca dice katsayısı çıktısı	28
Şekil 4.5. Konumlandırma ikinci eğitim aşaması boyunca öğrenme katsayısı çıktısı....	28
Şekil 4.6. Konumlandırma ikinci eğitim aşaması boyunca kayıp fonksiyonu çıktısı.....	29
Şekil 4.7. Konumlandırma üçüncü eğitim aşaması boyunca dice katsayısı çıktısı.....	29
Şekil 4.8. Konumlandırma üçüncü eğitim aşaması boyunca öğrenme katsayısı çıktısı ...	29
Şekil 4.9. Konumlandırma üçüncü eğitim aşaması boyunca kayıp fonksiyonu çıktısı ..	30
Şekil 4.10. Konumlandırma üçüncü eğitim aşaması boyunca dice katsayısı çıktısı.....	30
Şekil 4.11. Konumlandırma dördüncü eğitim aşaması boyunca kayıp fonksiyonu.....	31
Şekil 4.12. Yanlış ve doğru konumlandırma sonuçlarının aynı anda olduğu görüntü....	31
Şekil 4.13. Anormal durum tahmin modeli kayıp fonksiyonu grafiği.....	32
Şekil 4.14. Anormal durum tahmin modeli adadelta optimizör için kayıp fonksiyonu grafiği	32
Şekil 4.15. Farklı lambda değeri için kayıp fonksiyonunun grafiği	33
Şekil 4.16. Anormal durum gerçekleşme anı.....	33
Şekil 4.17. Anormal durum gerçekleştiği anda oluşan sinyal.....	33
Şekil 4.18. Anormal durum tahmin modeli ROC grafiği	34
Şekil 4.19. Girdi resmi ve tahmin edilen derinlik haritası	34
Şekil 4.20. Orta ve yakın mesafe tahminleri.....	35
Şekil 4.21. Anormal durum gerçekleştiği anda tespiti, konumlandırılması ve derinlik haritasında gösterimi	35

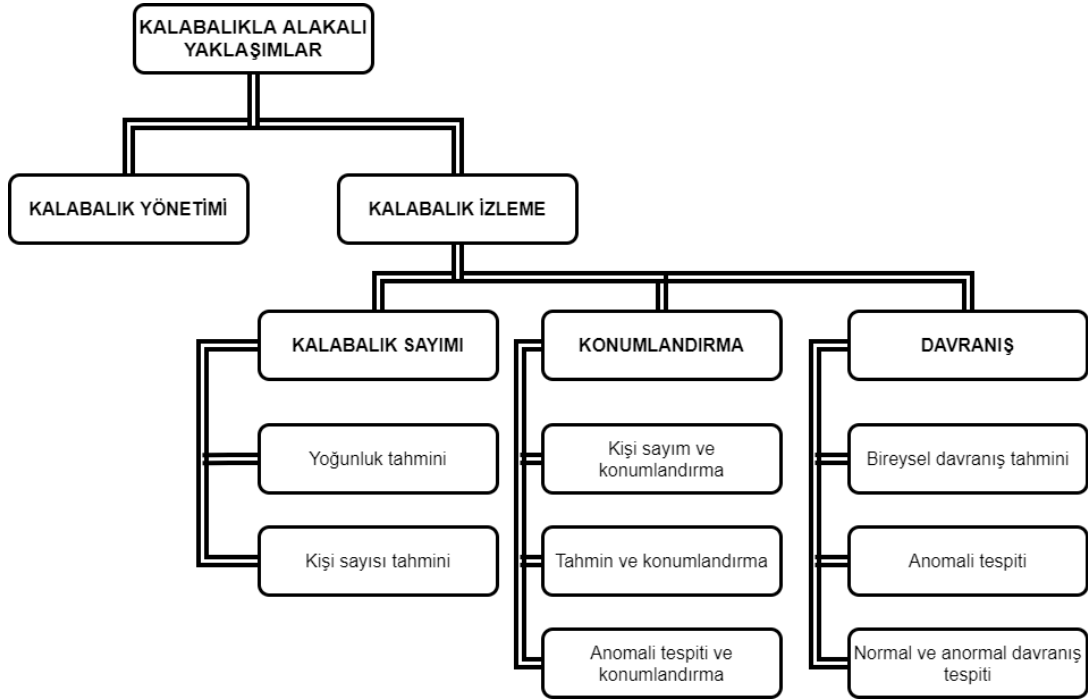
ÇİZELGELER DİZİNİ

Çizelge 2.1. Derin öğrenme tabanlı monoküler derinlik tahmini metotları (FC: Tam bağlantılı; CNN: Konvolüsyonel sinir ağları)	13
--	----

1. GİRİŞ

Kalabalık, aynı veya farklı bir takım insanın ortak hedefler doğrultusunda motive edilerek bir grup oluşturmasıdır (Khan vd. 2020). Kalabalık iki kategoriye bölünmüştür: (1) Yapılandırılmış kalabalık ve (2) yapılandırılmamış kalabalık. Yapılandırılmış kalabalıkta insanlar aynı yönde hareket ederler (ör. yoldaki trafik akışı). Diğer taraftan, yapılandırılmamış kalabalıkta insanlar farklı zamanlarda farklı yönlerde doğru hareket ederler (ör. Tren istasyonu) (Choundary vd. 2017).

Kalabalık ortamların izlenmesi ve kalabalık ortamlarda dâhil olmak üzere halka açık alanlarda asayişin sağlanması son yıllarda büyük önem kazanmıştır. Bu nedenle sokaklar, caddeler, kavşaklar, banklar, alışveriş merkezleri vb. kamusal alanlarda güvenlik kamerası kullanımı giderek artmıştır. Ancak, izleme tarafındaki kapasite aynı oranda artmamıştır. Bunun sonucunda güvenlik kamerası sayısı ve izleme ortamları arasında verimsiz bir oran ortaya çıkmıştır. Güvenlik kameralarının en önemli görevlerinden birisi trafik kazası, hırsızlık, illegal aktiviteler gibi anormal durumları kayıt altına almasıdır. Gündelik hayatta kayıt altına alınan bu görüntülerin izlenmesi için çeşitli kalabalık izleme yöntemleri geliştirilmiştir. Bu yöntemler ana başlıklar halinde Şekil 1.1’de gösterilmiştir.



Şekil 1.1 Kalabalıkla Alakalı Araştırma Yaklaşımları

Kalabalık izleme yöntemleri güvenlik takibi, afet yönetimi, trafik takibi, kamusal alanların tasarımı gibi çok geniş bir kullanım alanına sahiptir. Bu çeşitli uygulama alanları araştırmacıları araştırma konusunda cesaretlendirmiştir ve bu konuda birçok araştırma yapılmıştır. Kalabalık izleme konusu çalışmaya açık bir alandır. Kalabalık izleme uygulama başlıklarından bazıları, kalabalık sayım uygulamaları, yoğunluk tahmini, kişi takibi, sahne anlamlandırma, görsel içerisinde konumlandırma, davranış analizidir. Bu başlıklar altında gerçekleştirilen çalışmalar kaynak taraması başlığı altında ayrıca

detaylandırılacaktır. Ayrıca içinde bulunduğunuz COVID-19 virüsü salgını döneminde kalabalık alanların yönetiminin ne kadar önemli olduğu bir kez daha anlaşılmıştır. Sosyal mesafenin korunması, kamusal alanlardaki yoğunluğun azaltılması gibi hayati önem taşıyan sınırlamaların takibi de kalabalık izleme yöntemleri ile çözülebilecek problemler arasındadır.

Genelde de anormal durumlar normal durumlara göre nadiren gerçekleşirler. Ayrıca anormal durum ve normal durum tanımları kişiye, duruma ve zaman göre değişebilecek kavramlardır. Bu iki kavram arasında ki sınır çizgileri çok belirgin olmamakla beraber muğlaktır. Ayrıca kalabalık ortamlarda anormal durum tespitinin sağlanması için normal durumların veya anormal durumların tek tek belirlenmesi, işlenmesi çok fazla iş gücü ve zaman gerektirecek bir süreçtir. Anormal durumlar bahsedildiği gibi normal durumlara göre nadir oluşurlar, bu nedenle akıllara şöyle bir yöntem gelebilir, tüm normal durumların belirlenerek bu durumun dışında kalan durumların anormal durum olarak tespit edilmesi. Bu her ne kadar mantıklı gelse de tüm normal durumları hesaba katmak ve bunları tek tek etiketlemek çok zor hatta imkânsıza yakındır. Buna ek olarak anormallik tespiti kaba seviyede bir video filtreleme olarak düşünülebilir. İlerleyen tekniklerle anormallik bir kere tespit edildiğinde özel teknikler ile bu anormallik sınıflandırılabilir. (Sultani vd. 2018) Anormal durum tespiti sonucu elde edilen değer tek başına değerli bir bilgi olsa da bu bilgiye ilaveten anormal durumun konumu ve anormal durumun gerçekleştiği mesafenin tahmin edilmesi de önemlidir. Görüntü işleme algoritmalarının donanımsal iyileşmelere bağlı olarak hız kazandığı bu dönemde nesne konumlandırma ve tespiti için farklı farklı yöntemler kullanılmaktadır. Nesne konumlandırma için nesnenin görüntü içerisinde, x ve y ekseninde ki konumu, genişliği ve yüksekliğini tahmin eden modeller geliştirilmiştir. Koordinat tahmini yanı sıra problemi segmentasyon problemi olarak ele alan ve çıktı olarak bir görüntü haritası oluşturarak nesnenin konumunu tahmin eden modeller de geliştirilmiştir. Literatür de bulunan konumlandırma algoritmaları devam eden kaynak taraması bölümünde ele alınmıştır. Derinlik tahmini problemi görüntü işleme alanında en çok ilgilenen problemlerden birisidir. Derinlik tahmini için ilk başlarda stereo kamera düzenekleri kullanılsa da günümüzde tek kamera ve ondan elde edilen görüntüler üzerinden derinlik tahmini gerçekleştirebilen modeller geliştirilmiştir. Derinlik tahmini otonom araçlar ve robotik alanında sıkça kullanılan bir yöntemdir ve literatürde bulunan çalışmalar incelenerek kullanılan yöntem hakkında detaylı bilgiler verilmiştir.

Metotların geliştirilmesinin zorluğunun yanı sıra diğer bir zorlukta konu ile alakalı kaliteli veri setinin bulunmasıdır. Veri setleri genelde gerçekçi olmayan, sabit ışık koşullarında oluşturulmuş ve gerçek hayat şartlarını pek yansıtmayan veri setleridir. Doğru veri setinin belirlenmesi de en az doğru metodun kullanılması kadar hayati önem taşımaktadır. Literatürde kullanılan veri setleri de ayrıntılarıyla anlatılacaktır. Sonraki bölümde, çalışmada kullanılacak olan yöntemler hakkında yapılan kaynak taraması ele alınmıştır. Bölüm 3'de uygulamayı gerçekleştirmek için kullanılan materyal ve metotlardan bahsedilmiştir. Bölüm 4'de bulgular açıklanmış ve elde edilen değerler tartışılmıştır. Bölüm 5'de ise elde edilen bulgulara bağlı olarak sonuçlar açıklanmış ve ileriki çalışmalar için önerilerde bulunulmuştur.

2. KAYNAK TARAMASI

Kalabalık grupların, ortamların analizinin gerçekleştirilmesi için yapılmış bazı araştırmalar ve bu araştırmalar sonucunda yayınlanmış bazı yayınlar bulunmaktadır. Bu kapsamda, ulusal ve uluslararası kaynak taraması gerçekleştirilmiş ve bazı bulgular edinilmiştir.

2.1. Kalabalık Analizinde Kullanılan Yaklaşımlar

2.1.1. Kalabalık yoğunluğu tahmini, takibi ve davranışların tanımlanması

Mao vd. (2010) kalabalık yoğunluğunu tahmin etmek için bir model önermişleridir. Modelde her bir hareket video görüntüleri içerisinde hareket segmentasyonu yöntemi ile alınmakta, daha sonra sekiz farklı düşük seviye özellik çıkarımı yapılmaktadır. Bu özellikler, Haris Köşe Tespiti, KLT özellik noktaları, kontur sayısı, kontur çevresi vb. özellik çıkarma metotları ile hesaplanmaktadır. Daha sonra perspektif kaynaklı hataları engellemek için görsel üzerinde geometrik düzeltmeler yapılmakta ve sonunda çoklu regresyon modeli kullanılarak yaya sayısı tahmin edilmektedir.

Li vd. (2010) yaptıkları çalışmada arka planı silmişler ve bunu optik akış yöntemini ile birleştirmişlerdir. Elde ettikleri ön plan görüntüsünden bir dizi yeni özellik çıkartmışlar ve farklı kalabalıkları sınıflandırmak için bu özellikler SOM algoritmasına uygulamışlardır. 500 adet resimlik bir veri setinde 85.6% başarımla elde etmişlerdir.

Kim vd. (2010) çalışmalarında yoğunluk kestirimi için optik akış yöntemini kullanmışlar ve yoğunluk ile beraber kaç kişilik bir grubun hareketli olduğunu kenar piksel sayılarına göre hesaplamışlardır. Yoğunlukları farklı 5 farklı seviye başarılı şekilde sınıflandırılmıştır.

Fiaschi vd. (2012) kalabalıktaki kişi sayısının tahmini için bir çalışma gerçekleştirmişlerdir. Bu çalışmada sistem görüntü ve yoğunluk haritasını girdi olarak almakta ve Rastgele Karar Ağaçları ile görüntüdeki kişi sayısını tahmin etmektedir.

Bunster vd. (2012) yaptıkları çalışmada otomatik otobüs filosu planlama ve sevkiyatı için mevcut yolcu taleplerini doğru şekilde yönetecek bir sistem tasarımının otobüs duraklarındaki insan sayısının bilinmesine bağlı olduğunu düşünmüşlerdir. Bu nedenle otobüs duraklarında bekleyen insan sayısını tahmin edecek bir metod geliştirmişlerdir. Görüntüler farklı ışık ve perspektif durumlarından etkilenmemesi için düzeltilmiştir, daha sonra, alandaki insan sayısının bulunması için 4 yaklaşım uygulanmıştır, bu yaklaşımlardan iki tanesi(standart lineer regresyon modeli, lineer diskriminant analizi) parametrik diğer iki tanesi(olasılıksal sinir ağları, k-en yakın komşu) ise parametrik değildir. Bu yaklaşım ayrıca sokak ve caddeler gibi kamusal alanlarda da başarılı sonuç vermiştir.

Chu vd. (2013) yaptıkları çalışmalarında çoklu kernel izleme yöntemiyle kalabalık içerisinde bir kişi veya nesneyi başarılı şekilde takip etmişlerdir. Daha önceleri geliştirilen sistemler sadece 1 adet kernel içerdikleri için takip edilecek nesnenin önüne bir engel gelmesi durumunda başarısız olmaktaydılar. Bunun üstesinden gelmek için kendi

içlerinde ilişkili olan çoklu kernel metodunu geliştirdiler. Bu nedenle, metot nesne önünde engel olması durumunda da başarılı sonuç vermiştir. Ayrıca birden fazla nesne veya kişi takibinde de umut verici sonuçlar vermiştir.

Chen vd. (2013) araştırmalarında insan yaş tahmini, vücut/yüz poz-duruş tahmini, kalabalık yoğunluk tahmini gibi problemlerin çözümünde genelde regresyon metodu üzerinde durulduğunu, bu metot ile yüksek boyutsal vektör biçimli özelliklerin öğrenme fonksiyonu ile haritalandırıldığından bahsetmiştir. Bu haritalama ve öğrenme işlemi eğitim seti seyrek ve dengesiz dağılıma sahip olduğunda başarısız olmaktadır. Bu başarısızlığın önüne geçecek, sadece dağınık ve dengesiz veri setleri üzerinde çalışan ve başarılı sonuçlar veren bir metot önermişlerdir.

Ping vd. (2013) araştırmalarında önceki yöntemlerde genelde doku tabanlı veya piksel tabanlı sistemlerin kullanıldığından bahsetmiş ve bu iki yöntemi bir araya getirecek bir yöntem belirlemişlerdir. Öncelikle görüntüyü 8 bloka ayırmışlardır, 8 parçaya ayrılan bu görüntü üzerinde en küçük kareler regresyonu, gri seviye matris bileşimi ile doku analizi gerçekleştirmişler ve destek vektör makineleri modeli ile yoğunluk tahmini yapmışlardır.

Pham vd. (2015) çalışmalarında her parça arasındaki özellikleri ve göreceli konumlarını öğrenen bir yaklaşım geliştirmişlerdir. Gaussian kernel yoğunluk tahmini aracılığı ile bahsedilen yoğunluk harita parçalarını üretmişler ve bu parçaları kullanarak rastgele karar ağaçlarına uygulamışlardır.

Wang vd. (2015) var olan yöntemlerin genelde el yordamı ile öz nitelik çıkarttığını ve zahmetli olduğunu düşünmüşlerdir. Ayrıca SIFT (D. G. Lowe, 1999), HOG (N. Dalal ve B. Triggs, 2005) vb. yöntemler yoğun kalabalık ortamlarda başarısız olmaktadır. Bu doğrultuda yazarlar bu sorunların üstesinden gelecek bir derin öğrenme yöntemi önermişlerdir. Evrimsel sinir ağı ile uçtan uca eğitilebilir bir yapı oluşturulmuş ve sayım işlemi için gerekli öz nitelikleri öğrenecek şekilde veri seti üzerinden eğitilmiştir. Modelin doğruluğunu arttırmak için veri seti içerisine pozitif örneklerin yanı sıra negatif örnekler de ilave edilmiş ve arka planda bulunan bina, ağaç vb. nesnelere sistemin etkilenmesinin önüne geçilmiştir.

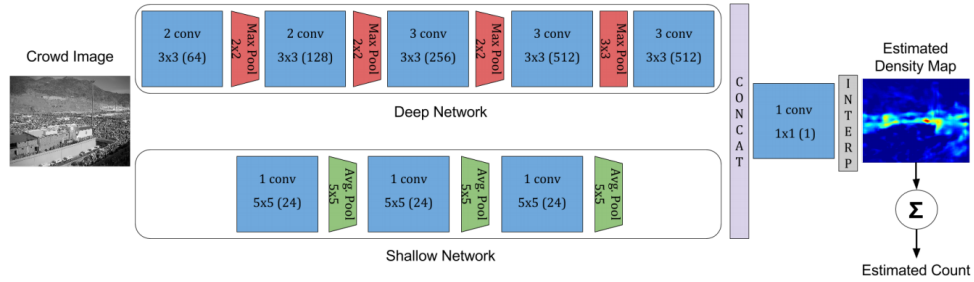
Shao vd. (2015) çoklu görev yapabilen derin model geliştirmişlerdir ve bu model kalabalığı daha iyi anlayabilmek için görünüm ve hareket özelliklerini bir arada öğrenecek bir model geliştirmişlerdir. Çoklu görev yapabilen derin sinir ağı Şekil 2.1'de gösterilmiştir.



Şekil 2.1. Shao vd. (2015) Görünüm ve hareket bilgisini girdi olarak alan çift kanal derin model

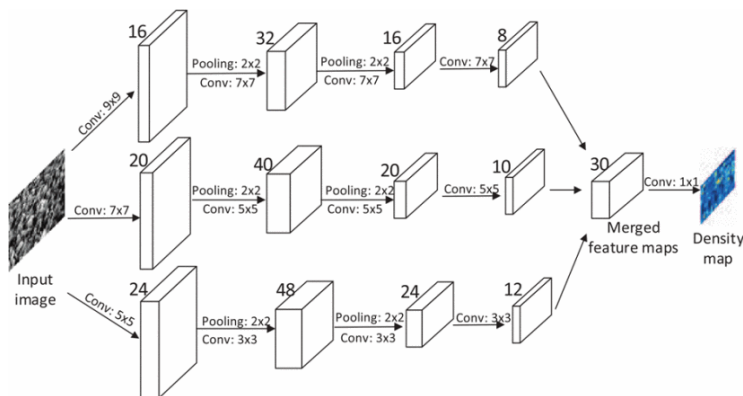
Şekil 2.1’de görüldüğü üzere derin öğrenme modeli hareket bilgisini ve görünüm bilgisini girdi olarak almakta ve bu girdiyi evrişim, ortaklama, normalizasyon katmanlarından geçirdikten sonra tahmin etme aşaması için tam bağlantılı sinir ağı ile bağlamıştır. Model 1000 video ve 8257 kalabalık sahne içeren bir veri seti ile eğitilmiştir. Bu yöntem temel özellik çıkarım metotlarına göre daha iyi bir performans sergilemiştir.

Boominahtan vd. (2016) çalışmalarında CrowdNet olarak adlandırdıkları bir derin evrişimsel ağ yapısı önermişlerdir ve önerdikleri ağ yapısı Şekil 2.2’de gösterilmiştir. Statik resimlerden kalabalığın yoğunluk tahmini ve kalabalıktaki kişi sayısının tahmini gerçekleştirebilen bir modeldir. Model derin ve sığ ağların, ağın çıkışındaki tam bağlantılı katmanda birleştirildiği bir yapıya sahiptir. Derin ağ modeli VGG-16 olarak bilinen modeli temel almaktadır. Sığ ağ modeli ise 24 adet 5x5 filtre boyutuna sahip 3 evrişim ve 3 ortaklama katmanından oluşmaktadır. Model yüz vücut gibi yüksek seviye anlamsal bilgileri ve düşük seviye bilgileri yakalayabilecek kapasitededir. Çoğu kalabalık veri seti sınırlı sayıda görüntü içerdiği için veri artırma tekniği veri setine uygulanmıştır. Model UCF_CC_50 veri seti üzerinde test edilmiş ve iyi sonuçlar vermiştir.



Şekil 2.2. Boominahtan vd. (2016) Sığ ve derin sinir ağından oluşan yoğunluk ve kişi sayısı tahmin modeli

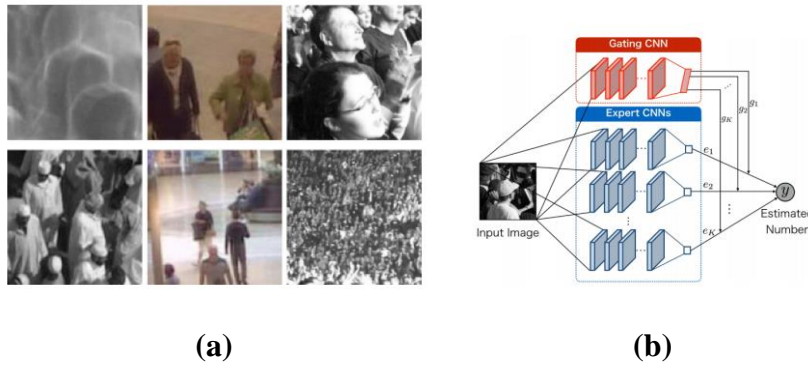
Zhang vd. (2016) çalışmalarında boyutu, çözünürlüğü, perspektifi fark etmeksizin herhangi bir resimden yoğunluk ve kişi sayısı tahmini yapabilen bir sistem önermişlerdir. Sahip olduğu farklı boyutlardaki filtreler görüntüdeki özellikleri öğrenen yapılarıdır, farklı boyutlarda olması dolayısıyla farklı kalabalık yoğunluklarında çalışabilmektedir. Ayrıca bu özelliği sayesinde herhangi boyuttaki görsel ile rahat çalışabilir. Bahsedilen çok kanallı model Şekil 2.3’de gösterilmiştir.



Şekil 2.3. Zhang vd. (2016) Farklı filtre boyutlarına sahip MCCNN yapısı

Kang vd. (2016) bilgisayar görüşü sistemleri sıklıkla yan bilgileri de içermektedir ve bu bilgiler problemin çözümü için çok yardımcı olabilir. Örneğin, kalabalık sayımında kameranın perspektifi, açısı, yüksekliği vb. parametreler sahnede ki insanın skalası hakkında bilgiler içerebilir. Bu yan bilgiler ile görüntüleri bir araya getiren, bu yan bilgiler ile ağırlıklarını optimize eden bir sistem önermişlerdir. Bu tip bir modeli eğitecek veri seti olmadığı için kamera yüksekliği, kamera açısı, kamera görüntüsü vb. parametreleri içeren bir veri seti oluşturulmuştur.

Kumagai vd. (2017) yaptıkları çalışmada kalabalık içerisindeki kişilerin yüksek doğrulukla sayımını hedeflemişlerdir. Kalabalık sayımının 3 zorlu problemini ele almışlardır, bunlar, sayılacak nesnenin engel arkasında kalması tam görünmemesi, küçük nesnelerin az piksel ile ifade edilmesi ve hedef nesnenin ölçüğe göre boyutunun değişmesidir. Farklı ölçeklerde insan görüntülerinden oluşan veri seti görüntüleri ve kalabalık görsellerinin ölçek ve yoğunluğuna göre uzman evrimsel sinir ağları ve kapı sinir ağları ile oluşturulmuş MoC-CNN modeli Şekil 2.4'de gösterilmiştir.



Şekil 2.4. a) Kumagai vd. (2017) Farklı yoğunluk ve ölçüğe sahip, ayrıca engel arkasında kalan kalabalık insanların görüntüsü. **b)** MoC-CNN Modeli

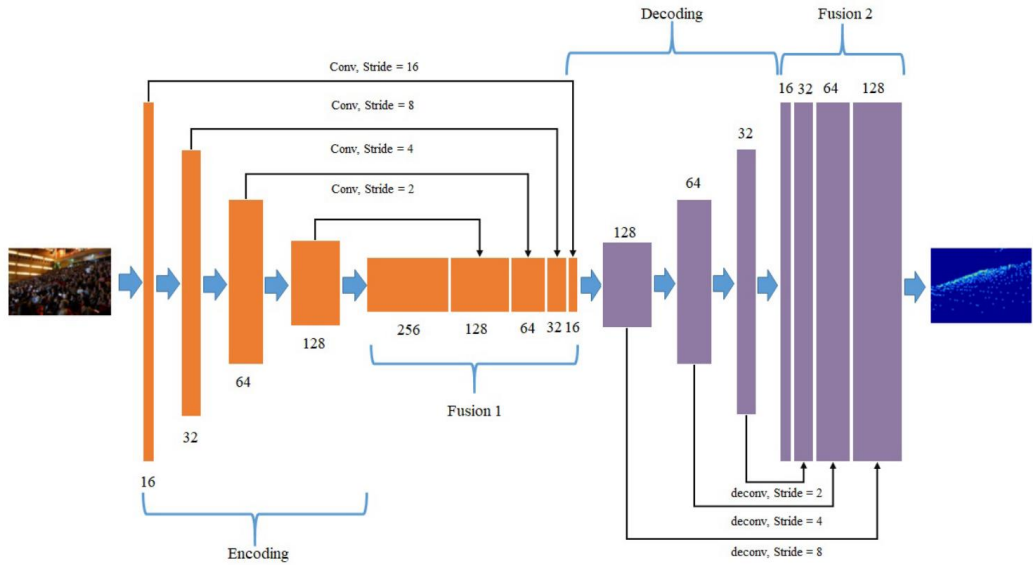
Bu zamana kadar geliştirilen metotların bazıları bu problemlerin üstesinden gelirken geliştirilen bazı metotların bu problemlerin çözümünde yetersiz kaldığını ifade etmişlerdir. Örneğin, görüntüdeki yoğun bölgelerin tespiti için regresyon temelli metotların kullanımı, tespit temelli metotların kullanımına oranla daha iyi sonuçlar vermektedir. Bahsedilen 3 problemin çözümü için evrimsel sinir ağlarının bir karışımını içeren bir model önermişlerdir. Bu model, 2 farklı bölümden oluşmaktadır. Uzman evrimsel sinir ağları ve kapı evrimsel sinir ağları. Uzman sinir ağları ve kapı sinir ağları hedeflenen çıktılar doğrultusunda doğru veriler ile eğitilmektedir. Uzman sinir ağları ve kapı sinir ağları ilgili yoğunluk veri setine göre eğitildikten sonra girdi resminin kalabalık yoğunluğu ve nesne ölçüğüne göre hangi uzman sinir ağının çıkışının aktif olacağı kapı sinir ağı tarafından belirlenmektedir. Bu şekilde değişken yoğunluk ve ölçekli kalabalık görsellerinin sayımı yüksek doğruluk ile gerçekleştirilmiştir. Modelin testleri UCF_CC_50 veri seti ve Mall veri seti üzerinde gerçekleştirilmiştir.

Li vd. (2018) yaptıkları çalışmada sıkışık, yoğun ortamlarda ki insan sayısını ve yoğunluk haritasını tahmin etmeye çalışmışlardır. Tanıttıkları CSRNet algoritması 2 ana husustan oluşmaktadır: Evrimsel Sinir Ağı (CNN) ve 2 boyutlu özellik çıkarımı için genişlemiş filtrelerle sahip Evrimsel Sinir Ağı. Uçtan uca eğitilebilir olan bu yapıyı 4 farklı veri seti(ShanghaiTech veri seti, UCF CC 50 veri seti, WorldEXPO'10 veri seti ve

UCSD veri seti) üzerinde test etmişler ve iyi sonuçlar elde etmişlerdir.

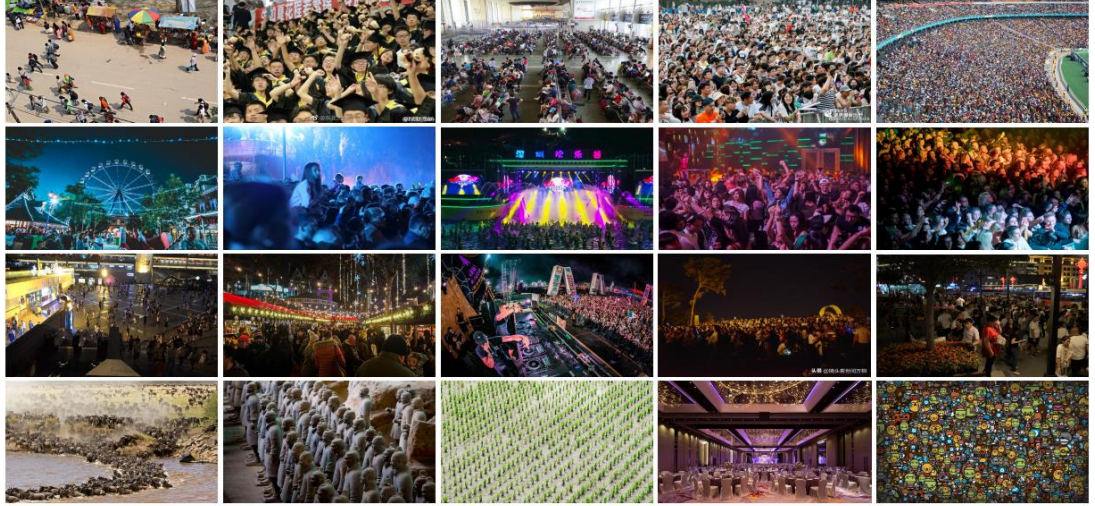
Miao vd. (2019) gerçekleştirdikleri çalışmada zamansal ve mekânsal bilgileri işleyerek kalabalık sayımı yapabilecek bir sistem geliştirmişlerdir. Geliştirdikleri metod video görüntülerinden elde ettiği her bir görüntüyü S ve T olarak isimlendirdikleri iki farklı CNN ağına uygulamakta ve bu iki ağdan iki farklı kalabalık tahmini elde etmektedir. Bu ağlardan S olarak isimlendirilen 2 boyutlu evrişim, T olarak isimlendirilen ise 3 boyutlu evrişim işlemi uygulayarak özellik çıkarımı yapmaktadır. Bu iki modelin ürettiği yoğunluk haritası Fusion olarak adlandırılan diğer bir modele girdi olarak alınmakta ve nihai kalabalık yoğunluk tahmin haritası elde edilmektedir. Bu metod ile var olan sistemlerde video görüntülerinin her bir karesinin ayrı ayrı ele alınması sonucu oluşan performans kayıplarının önüne geçilmesi amaçlanmıştır. Her bir ağ ayrı ayrı eğitilmiştir ve 2 farklı veri seti üzerinde test edilmiştir.

Ding vd. (2020) çalışmalarında kalabalık ortamların yoğunluk haritasını çıkaracak bir sistem tasarlamışlardır. Kalabalık sayımının önemli bir görev olduğundan bahsetmişler ve doğru kalabalık sayım sistemine ulaşmak için kamera açısı, karmaşık arka plan, görüntüdeki engeller, farklı ölçekli görüntüler gibi problemlerin aşılması gerektiğini belirtmişlerdir. Bu kapsamda bu sorunları çözmek için Encoder-Decoder CNN yapısı önerilmiştir. Modele ait mimari Şekil 2.5’de gösterilmiştir.



Şekil 2.5. Ding vd. (2020) Tanıtılan kalabalık sayımı için simetrik mimari. Turuncu renkli kısım enkoder, mor kısım ise dekoder kısmını ifade etmektedir

Wang vd. (2020) çalışmalarında genel çalışmalardan farklı olarak mimari geliştirmek yerine veri seti hazırlama ve performansını test etme üzerine çalışmışlardır. Gerçekleştirilen CNN metodlarında kullanılan veri setlerinin küçük ve ihtiyacı karşılamadığı belirtilmiştir. Bu sorunu aşmak için 5109 resim ve 2,133,375 etiketli veri seti oluşturulmuştur. Oluşturulan bu veri seti gerçekleştirilen geçmiş çalışmalar üzerinde test edilmiş sonuçları çalışmada karşılaştırılmıştır. Bahsedilen NWPU kalabalık veri seti içerisinde bulunan görsellerden bazıları Şekil 2.6’ da gösterilmiştir.



Şekil 2.6. Wang vd. (2020) Farklı renk, ışık, perspektif, ışık koşulları içeren NWPU veri setinden bazı görseller

2.1.2. Kalabalık ortamlarda gerçekleşen anormal olayların tespiti

Önceki bölümlerde bahsedilen, kalabalık ortamlarda kişilerin takibi, sayısının tahmini gibi konuların yanında kalabalık ortamlarda gerçekleşen anormal durumların da tespiti önemli çalışma alanlarının başında gelmektedir. Araştırmacılar geleneksel görüntü işleme tekniklerinin yanı sıra derin öğrenme yöntemlerini de kullanarak kalabalık ortamlarda anormal durumların tespit edilmesini sağlamıştır.

Kratz ve Nishino (2009) çalışmalarında zaman-mekânsal davranış değişimlerini dağıtık saklı Markov modelleri ile tespit ederek anormal davranışları görüntü içerisinde belirleyebilen bir sistem tasarımı gerçekleştirmişleridir. Bu yöntem sayesinde kalabalık içerisinde gerçekleşen belirledikleri ani davranışları tespit edebilmişlerdir.

Jiang vd. (2009) çalışmalarında kalabalık sahnelerdeki kişisel davranışlarda gerçekleşen anormal durumları tespit etmeye çalışmışlardır. Bunun için gözetimsiz yöntemler ile tespit yaklaşımı kullanılmıştır ve sistem otomatik olarak bağlamsal bilgileri kalabalık video içerisinde çıkartmakta ve bu bağlamsal bilgiye karşılık gelen anormal davranışlar video içerisinde tespit edilebilmektedir.

Saligrama ve Chen (2012) çalışmalarında Neyman-Pearson karakterizasyonu yöntemini temel almışlardır. Bu yöntemle beraber elde edilen özneliklerin K-en yakın komşuluk değerlerini hesaplamışlardır. Yöntemin genel çerçevesi Şekil 2,7'de gösterilmiştir. Her özellik tanımlayıcısı için her bir uzay-zamansal konumda düzgün tahminler elde etmek için farklı ölçeklerdeki Uzay-Zamansal filtreler uygulanır. Her konum için yerel KNN mesafesi, eğitim ve test videosu için hesaplanır. Bu yerel KNN mesafeleri, test ve eğitim videosu için bileşik bir puan oluşturmak üzere toplanır. Farklı videolar üzerinde yöntemlerini denemiş ve başarılı sonuçlar elde etmişlerdir.

Huang G. vd. (2017) çalışmalarında geleneksel anormal durum tespit yöntemlerinde kullanılan ön işlem ve özellik çıkartma süreçlerini de kapsayan bir yöntem önermişlerdir. Ham veri üzerinde çalışarak anormal durumları tespit edebilecek bir derin

öğrenme modeli geliştirmişleridir. Bu modellerine Derin artımlı yavaş özellik analiz ağı (D-IncSFA) ismini vermişlerdir. Bu yöntemler kalabalıkta oluşan panik gibi anormal durumları doğrulukla tespit edebilmişlerdir.

Sabokrou vd. (2017) çalışmalarında kalabalık sahneler içeren görüntüleri tanımlamak için kademeli bir sinir ağı modeli önermişlerdir. Bu kademeli yapı iki aşamadan oluşmaktadır. İlk aşamada 3D auto-encoder ikinci aşamada ise 3D CNN mimarisi uygulanmaktadır. Sınıflandırma aşamasında ise Gaussian sınıflandırıcı kullanılmaktadır.

Narasimhan vd. (2018) çalışmalarında video görüntülerinde oluşan anormal durumları tespit etmek için zamansal ve mekânsal özellikleri çıkartacak yerel ve genel tanımlayıcılar kullanmışlardır ve bu özellik çıkarımı üzerine Gauss sınıflandırıcı uygulamışlar ve sonuçta oluşan değerleri bir araya getirerek bir anormallik skoru elde etmeye çalışmışlardır.

Sabokrou vd. (2018) önceden bahsedilen çalışmalarda kullanılan ve yama-temelli metotlarda verimsiz olduğu görülen bu ağların yerine yeni bir ağ yapısı önermişlerdir. Yöntemlerinde önceden eğitilmiş bir tam bağlantılı CNN ile eğitmeye müsait bir ağ kullanmışlardır ve transfer öğrenmesi metodu ile anormallik sınıflandırmışlardır.

Lunescu vd. (2019), Xu vd. (2020) anormal durum tespiti problemini tek sınıf veya ikili sınıflandırma aksine çok sınıflı bir sınıflandırma problemi olarak ele almışlardır. Önceki metotlarda dikkatlerini çeken, anormallikleri ifade eden alt bölgelerinin efektif şekilde tanımlanabilmesidir. Ancak bu yaklaşımların çoğunda optik akış, gradyan gibi bir dizi geniş giriş ve çıkışa sahiptir. Bu durum Xu vd. (2020) yaklaşımına ilham vermiştir ve geliştirilen metot ham girdiyi alıp örneğin anormallik olup olmadığını tanımlamaktadır.

Bazı araştırmalar bu problemleri regresyon problemi olarak ele almıştır. Burada amaç video karesi veya bölümünde olan anormallik durumunun derecelendirilmesi ve skorun tahmin edilmesidir. Bunun bir örneği de Sultani vd. (2018) yaptıkları çalışmadır. Zayıf etiketlenmiş videoların çoklu örnek öğrenimi metodu kullanılarak anormalliklerin tanınmasıdır. Zamana bağlı özellikleri öğrenmesi için yöntemlerinde 3D CNN ağı kullanmışlardır. Sonra bu zamana bağlı özellikleri tam bağlı sinir ağı katmanına beslemişler ve anormallik skoru elde etmişlerdir. Hatanın geri beslemesi için çoklu örnek öğrenimi metodu kullanmışlardır, bu sayede model veriler zayıf etiketlenmiş olsa bile anormal bölümleri öğrenebilmektedir.

2.2.Kalabalık Ortam Veri setleri

2.2.1. Ucf-Crime veri seti

Bu veri seti geniş ölçekli bir anormal durum veri setidir. Sultani vd. (2018) tarafından tanımlanan veri setinde 13 farklı gerçek hayat anormal durumları bulunmaktadır. Bu anormal durumlar; Kötü davranış, Alıkoyma, Kundaklama, Taciz-Saldırı, Kaza, Gasp, Patlama, Kavga, Soygun, Vurulma, Hırsızlık, Yağma ve Vandalizm olmak üzere 13 farklı durumdur. Bu veri seti YouTube ve LiveLeak gibi kaynaklar üzerinden farklı arama terimleri ve farklı dillerde (ör. Çince, İngilizce, Fransızca vb.) araştırmalar sonucu elde

edilmiştir. Şaka videoları, CCTV kamera ile çekilmeyen, el kamerası ile kayıt edilen, oynanmış olan videolar veri seti içerisine dâhil edilmemişlerdir. Aynı zamanda anormal durumun net anlaşılmadığı videolar da veri setinden kaldırılmıştır. 950 anormal, 950 normal durum olmak üzere 1900 video veri setinde bulunmaktadır. Veri seti eğitim seti için 800 normal ve 810 anormal video, test seti için 150 normal, 140 anormal olmak üzere iki bölüme ayrılmıştır. Veri setine ait bir patlama anı Şekil 2.7’de gösterilmiştir.



Şekil 2.7. Veri setinden patlama anormal durumu örneği

2.2.2. Ucsd veri seti

UCSD veri seti (S.V.C. Lab 2014) sabit kameralar ile iki farklı sahneden, Peds1 ve Peds2, kayıt altına alınmış video görüntülerinden oluşmaktadır. Normal olaylar yayaların yürüdüğü sahnelerdir, anormal olaylar ise bisiklet sürücülerinin, kaykaycılarının, patencilerin, tekerlekli sandalyeli insanların bulunduğu sahnelerdir. Yayaların beklenmedik alanlardaki hareketleri de ayrıca anormallik olarak düşünülebilir. Peds1 veri seti 34 eğitim, 36 test veri setinden oluşmaktadır. Peds2 veri seti, 16 eğitim, 12 test veri setinden oluşmaktadır. Şekil 2.8’den de görüleceği üzere yayaların yürümesi normal iken taşıt bulunması anormal durum olarak ele alınabilir.



Şekil 2.8. UCSD Veri setinden Peds1 verisine ait yaya ve araç görüntüsü

2.2.3. Umn veri seti

Bu veri seti panik durumlarının kayıt altına alınması ile oluşturulmuştur. UMN veri seti (UMN 2006), 3 farklı sahneden oluşmaktadır: çimenlik, kapalı alan ve plaza. Video çözünürlükleri ise 320x240 pikseldir. Videolarda iki farklı olay vardır, yürüyen insanlar ve koşan insanlar. Normal sahneler farklı yönlerde yürüyen insanları içerirken anormal sahneler kaçışan insanları içermektedir. Eğitim ve test videoları bu videolardan seçilerek oluşturulabilir. UMN veri setine ait 3 farklı alandan alınan görüntüler Şekil 2.9’da verilmiştir. Anormal durum olarak insanların kaçışması gösterilmiştir.



(a)



(b)

Şekil 2.9. a) UMN Veri seti normal durumları. **b)** UMN Veri seti kaçışma durumları

2.2.4. Pets veri seti

PETS Veri seti (PETS, Vellore 2009) normal ve anormal olmak üzere iki farklı kategoriden oluşmaktadır. Bu veri setinde normal durum yayaların yürüdüğü durum, anormal durum ise yayaların kaçtığı durum olarak ele alınmıştır. Eğitim ve test veri setleri ise bu durumlar arasından elde edilebilir. Veri seti 576x768 çözünürlüğe sahiptir.

2.2.5. Subway veri seti

Metro giriş ve çıkış kapılarının güvenlik kameralarından elde edilen video görüntüleri ile oluşturulmuş bir veri setidir. Giriş kapısı görüntüleri 1 saat 36 dakika uzunluğunda 144249 frame içermektedir. Çıkış kapısı görüntüleri ise 43 dakika uzunluğunda 64900 frame içermektedir. Video görüntülerinin çözünürlüğü 512x84 pikseldir. Video görüntülerinde ki normal durum çıkış kapısından çıkış yapıp turnikeden çıkarak kapıdan uzaklaşanlar ve giriş kapısından giriş yaparak turnikeye yönelen ve platforma giriş yapanlardır. Anormal durumlar ise giriş tarafından çıkış, çıkış tarafından giriş yapmaya çalışanlardır. Veri seti 66 nadir anormal durum içermektedir.

2.2.6. Avenue veri seti

AVENUE veri seti (Lu vd. 2013) CUHK kampüs caddesinde yaşanan olayların kayıt altına alınması ile elde edilmiştir. 16 eğitim, 21 test video klibinden oluşmaktadır. Toplamda 30652 frame içermektedir, 15328 tanesi eğitim geri kalanı ise test için kullanılmaktadır. Eğitim videoları normal durumları, test videoları ise normal ve anormal durumları içermektedir.

2.2.7. U-turn veri seti

U-TURN veri seti tramvayların geçtiği, araçların farklı yönlere sürdüğü, yayaların geçtiği bir kavşaktan elde edilen görüntülerden oluşmaktadır. Anormal durumlar araçların ve tramvayın illegal U dönüşü yaptığı durumları ifade etmektedir. Normal olaylar ise

rutin trafik akış durumudur.

2.2.8. Web veri seti

WEB veri seti (Mehran vd. 2009) GettyImages, ThoughtEquity gibi farklı internet sitelerinden elde edilmiş kalabalık görüntülerinden oluşmaktadır. Veri seti 12 farklı senaryo içermektedir, normal durum senaryoları, yürüyen insanlar, maraton koşucuları, anormal insanlar ise kaçışma, panik, protestocular, kavga vb. olaylardır.

2.2.9. Caviar veri seti

CAVIAR veri seti (R.B. Fisher 2004) alenen kullanıma açık bir veri setidir. Veri seti, aktivite tanıma, anormal durum tespiti gibi durumlar için kullanılabilir. 384x288 piksel çözünürlüğünde 25fps video görüntülerinden oluşmaktadır. Yürüyen, kavga eden, alışveriş merkezine giren ve çıkan insan görüntüleri vb. görüntülerden oluşturulmuştur.

2.3. Derinlik Tespiti

2.3.1. Derinlik tespitinde kullanılan literatürdeki yöntemler

Derinlik tahmini bilgisayar görüşü alanında üzerinde uğraşılan temel alanlardan birisidir. Robotik, sahne anlama, 3 boyutlu yeniden oluşturma ve tıbbi alanda kullanım potansiyeli çok fazla olan bir yöntemdir. Monoküler derinlik tahmini, tek bir resim içerisinde derinlik bilgisi olmadığı için zor bir yöntemdir. Genellikle klasik yöntemler de derinlik tahmini için çift kamera, çok açılı geometri yöntemleri gibi yöntemler kullanılmaktadır. Son zamanlarda CNN mimarilerinde kat edilen yol ve veri setlerinin halka açık hale gelmesi sayesinde monoküler derinlik tahmini metodları da yol kat etmişlerdir.

Scharstein D. ve Pal C. (2007) çalışmalarında Markov Random Field modeli yerine Conditional Random Field modelini önermişlerdir. Bunun için de çok sayıda stereo görüntü içeren bir veri seti oluşturmuşlar ve bu veri setinden CRF parametrelerini elde etmişlerdir. Bu sayede elle çıkarılan MRF parametreleri yerine otomatik çıkarılan CRF parametreleri tanıtılmıştır.

Ji R. vd. (2019) çalışmalarında monoküler resimlerden derinlik tahmini yapabilmek için CNN mimarisini kullanan bir yöntemi önermişlerdir. Maliyetli olan veri toplama kısmını hafifletmek için de yarı gözetimli tartışmalı öğrenme metodunu kullanmışlardır.

Godard C. vd. (2019) çalışmalarında var olan karmaşık çalışmalara kıyasla daha basit bir monoküler resimlerden derinlik tahmini modeli tanıtılmışlardır. Çalışmalarında derin öğrenme modellerinden U-Net modelini temel alan bir model geliştirerek Depth network adını vermişlerdir. Ayrıca başka bir modelde pose network adını verdikleri modeldir. Bu iki model sayesinde derinlik tahmini resimler üzerinde uygulanmaktadır.

Kuznetsov, Y. vd. çalışmalarında resimlerden derinlik haritası tahmini için bir model önermişlerdir. Modelde enkoder ve dekoder mimarisi ile derinlik haritası tahmini yapan model tanıtılmıştır. Model KITTI veri seti ile eğitilmiştir.

Chen, Y. vd. (2019) çalışmalarında ACAN adını verdikleri bir model tanıtmışlardır. Modellerinde enkoder olarak ResNet mimarisini kullanmışlardır. ResNet mimarisinden elde edilen özellikler CAM modeline uygulanmış ve piksel ve resim seviyesinde özellik çıkarımı yapmışlardır. Son olarak tanıtılan metotta tahminler tahmin edilen derinlik değerlerine dönüştürmüştür.

Alhashim, I. vd. (2018) çalışmalarında transfer öğrenme ile yüksek kaliteli derinlik tahminini amaçlamışlardır. Önceden eğitilmiş DenseNet-169 mimarisi enkoder olarak kullanılmıştır. Dekoder katmanında ise evrişim katmanları ile yukarı örnekleme yapılmaktadır. Literatürde kullanılan farklı derinlik tahmin metodları Çizelge 2.1’de özetlenmiştir.

Çizelge 2.1. Derin öğrenme tabanlı monoküler derinlik tahmini metotları (FC: Tam bağlantılı; CNN: Konvolüsyonel sinir ağları)

Metot	Mimari	Kategori
EMDEOM(Dos Santos Rosa N. vd. 2019)	FC	
ACAN(Lee, J.H vd. 2019)	Encoder-Decoder	
DenseDepth(Teed, Z vd. 2018)	Encoder-Decoder	
DORN(Fu, H vd. 2018)	CNN	Gözetimli
VNL(Goldman, M vd. 2019)	Encoder-Decoder	
BTS(Guizilini, V vd. 2019)	Encoder-Decoder	
DeepV2D(Andraghetti, L vd. 2019)	CNN	
LISM(Zhao, S vd. 2019)	Encoder-Decoder	
monoResMatch(Tosi, F vd. 2019)	CNN	
PackNet-SfM(Chen, W vd. 2016)	CNN	Öz-Gözetimli
VOMonodepth(Xie, J vd. 2016)	Auto-Decoder	
monodepth2(Godard, C vd. 2019)	CNN	
GASDA(Garg, R vd. 2016)	CNN	Yarı-Gözetimli

2.3.2. Derinlik tespiti için kullanılan literatürdeki veri setleri

NYU-v2 veri seti derinlik tahmini için oluşturulmuştur (Silberman, N vd. 2012). Veri seti 1449 RGB etiketli resimden oluşmuştur. 3 farklı şehirden 464 sahneden 407 bin kare içermektedir. Veri setleri kapalı alan sahnelerde derinlik tahmini, segmentasyonu ve

sınıflandırması için kullanılabilir.

Make3D veri seti 400 ve 134 eğitim ve test için dış ortam resmi içermektedir(Saxena, A vd. 2008). Ayrıca, veri seti farklı tiplerde dış ortam, iç ortam ve sentetik sahneler içermektedir.

KITTI veri seti RGB stereo resimlerden oluşmaktadır(Geiger, A vd. 2012). 394 yol sahnesi içermektedir.

Pandora veri seti 250 bin yüksek çözünürlüklü RGB resim ve ona karşılık gelen derinlik resimlerinden oluşmaktadır (Borghini, G vd. 2017).

SceneFlow veri seti ilk geniş ölçekli sentetik veri setidir (Mayer, N vd. 2016). 39 bin stereo resim ve onlara karşılık gelen derinlik, optik akış maskeleri içermektedir.

2.4. Görüntü İçerisinde Nesne Lokalizasyonu

Nesne tespiti ve konumlandırılması görüntü işleme alanında üzerine en fazla çalışılan alanların başında gelmektedir. Nesne konumlandırma, belirli bir nesnenin görüntü içerisinde ki konumunun tespit edilmesi işlemidir. Bu işlemi gerçekleştirebilmek için birçok matematiksel model ve teknik geliştirilmiştir ve cevap aranan soru; Resimdeki nesne nerededir? Geçtiğimiz yıllarda görüntü işleme alanında nesne tespiti ve konumlandırma gibi alanlara ilginin arttığı Google verilerinden gözlemlenebilir. Şekil 2.10'da gösterildiği üzere matematiksel işlem gücünün artmasına bağlı donanımsal iyileşmelerden de kaynaklı yıllara göre yayın sayısı artarak devam etmektedir.

Görüntü içerisinde nesne tespiti için geliştirilen sistemlerden birisi de Viola Jones Dedektörleri'dir. P. Viola ve M. Jones, herhangi bir kısıtlama(ten rengi vb.) olmaksızın gerçek zamanlı yüz tespiti ve konumlandırması yapmıştır. P. Viola ve M. Jones, (2001) çalışmalarında ilk olarak hızlı hesaplama yapmayı sağlayan integral görüntü adını verdikleri yeni bir görüntü formatını tanıtmışlardır. Daha sonra, verimli bir sınıflandırma sağlayan Adaboost algoritmasına dayalı yeni bir sınıflandırma algoritmasını tanıtmışlardır. Ayrıca, arka plandan hedef görseli ayırt edilebilmek için kaskad diye adlandırdıkları karmaşık bir sınıflandırıcı yöntemi geliştirmişlerdir.

N. Dalal ve B. Triggs (2005) çalışmalarında Histogram of Oriented Gradienr (HOG) adını verdikleri özellik dönüşüm yöntemi tanıtmışlardır. Bu yöntem çeşitli nesnelere tespit etmek için kullanılsa da öncelikle yaya tespiti problemi için kullanılmıştır. Bu yöntem,1800 farklı etiketli insan resmi içeren geniş bir resim veri tabanı üzerinde eğitilerek test edilmiştir. İnsan görüntüsüne ait piksellerin histogram gradyanını içeren görseller Şekil 2.10'da gösterilmiştir.



Şekil 2.10. HOG metodu uygulanmış görseller (N. Dalal ve B. Triggs, 2005)

Girshick, R.B. vd. (2014) çalışmalarında R-CNN modelini tanıtmışlardır. Bölge tabanlı CNN modelinde girdi resminden bölgesel özellikler elde edildikten sonra bu özellikler bir CNN modeline uygulanır. Model sonucunda sınıflandırma ve konumlandırma işlemi gerçekleştirilir. S. Ren VD. (2015) çalışmalarında Fast R-CNN modelini tanıtmışlardır ve R-CNN modelinin sahip olduğu eksik alanları gidermişlerdir. Bu eksiklikler, hız, performans ve kaynak tüketimi gibi parametrelerdir.

YOLO-V3 ağı (Redmon, J. ve Farhadi A. 2018), YOLO (Redmon J. vd. 2016) ve YOLO-V2 (Redmon J. vd. 2017) ağları üzerine çıkmış bir mimaridir. YOLO mimarisi nesne tespiti problemine regresyon bakış açısı ile yaklaşmıştır. YOLO mimarisi R-CNN mimarisinin aksine nesne tespiti için herhangi bir muhtemel bölgeye ihtiyaç duymamaktadır. Nesne tespitine regresyon problemi olarak yaklaşması sayesinde her sınıf için sınır çizgilerini ve sınıf değerlerini aynı anda tahmin edebilmektedir. Bu sayede hız ve performans konusunda artış sağlamıştır.

3. MATERYAL VE METOT

Bu tez kapsamında, güvenlik kameraları tarafından kayıt edilen görüntüler üzerinden anormal durumu tespit edebilecek, anormal durumu konumlandırabilecek ve anormal durumun mesafesini tahmin edebilecek bir sistem derin öğrenme metotları kullanılarak gerçekleştirilmiştir. Derin sinir ağları eğitim aşaması için çok sayıda veriye ihtiyaç duymaktadır. Derin sinir ağlarında kullanılan parametre sayısı ve veri kümesi boyutu tahmin başarısına pozitif etki etmektedir fakat parametre sayısı ve veri sayısı arttıkça sistemin hesaplama yükü artmaktadır. Bu tez kapsamında kullanılan modelleri eğitmek için Intel core i7 işlemciye ve RTX 2070 8GB ekran kartına sahip bir bilgisayar kullanılmıştır. Yazılım Python dili ile geliştirilmiş ve yazılımda tensorflow, scikit, opencv, keras kütüphaneleri ağırlıklı olmak üzere farklı yazılım kütüphaneleri kullanılmıştır.

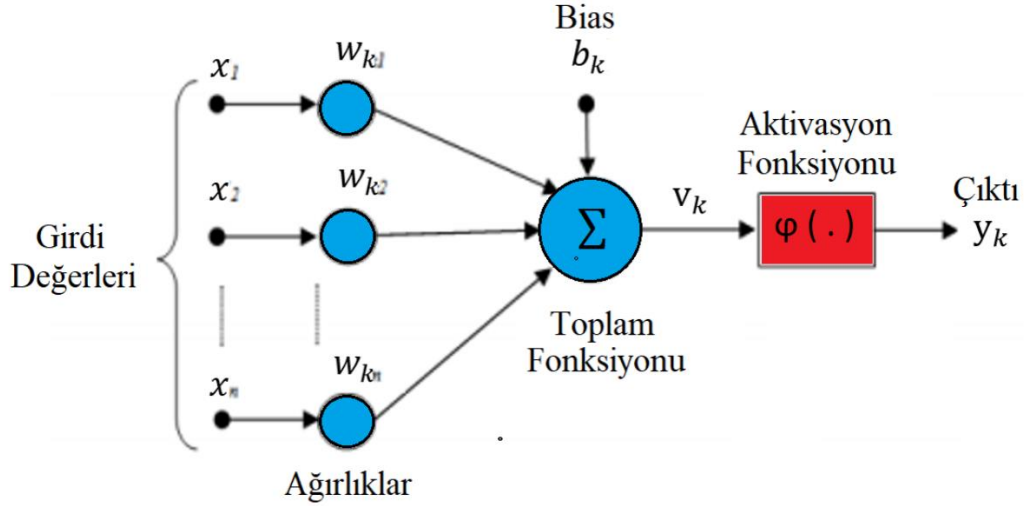
3.1.Yapay Sinir Ağları

Memeli beyinde farklı bölgeler farklı görevleri icra eder. Serebral korteks memeli beyinde ki dış katmandır. Onu yaklaşık 2 ile 5mm kalınlıkta bir kâğıt olarak düşünebiliriz, kendi üzerine katlanmış kompakt yapıda geniş yüzey alanına sahip bir yapısı vardır. Korteks memeli beyinde ki en geniş ve en karmaşık parçadır. Çok sayıda sinir hücresi, nöron içerir. İnsanlardaki Serebral korteks yaklaşık 10^{10} nöron içerir. Bu nöronlar aksondan geçerek sinapsis ile biten dallar ile bağlıdır. Sinapsisler dentritlere bağlıdır (Mehlig 2019). Beynin bu nöron ve nöronların bağlantı yapısının matematiksel olarak modellenmesi sonucu elde edilen sistem yapay sinir ağlarıdır. Yapay sinir ağlarında ki en küçük birime de sinir adı verilir ve her katman sisteme göre farklı sayıda sinir içerebilir. Şekil 3.1 yapay sinir ağlarının katmanlardan oluştuğunu ve tahmin için bilginin akış yönünü göstermektedir.



Şekil 3.1. Basit bir yapay sinir ağı gösterimi

Yapay sinir ağları, insan beyinde olduğu gibi nöronlara sahiptir, bu nöronlar ağırlıklı ismi verilen değişkenler ile birbirlerine bağlanmışlardır. Sistemin lineer olmaması için her nöron bir aktivasyon fonksiyonuna sahiptir ve bu aktivasyon fonksiyonları girdi olarak aldıkları sayıyı işleyerek başka bir sayıyı çıktı olarak üretir. Bu aktivasyon fonksiyonlarının farklı çeşitleri bulunmaktadır. Farklı sistemlerde ve veri setlerinde farklı aktivasyon fonksiyonlarının kullanılması gerekmektedir. Genelde kullanılan aktivasyon fonksiyonları Sigmoid, ReLu, Tanh fonksiyonlarıdır. Yapay sinir ağlarının temel birimi olan bir sinire ait girdileri ve girdiler üzerinde uygulanan matematiksel işlemleri ifade eden Şekil 3.2 basit bir yapay sinir ağını göstermektedir.



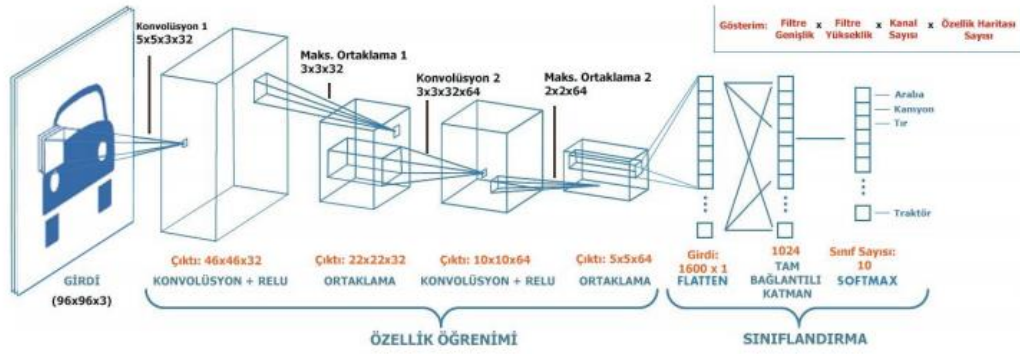
Şekil 3.2. Basit bir yapay sinir ağı örneği (Karahan M. 2015)

3.2. Evrişimsel Sinir Ağları

Evrişimsel sinir ağları yapı olarak yapay sinir ağlarına çok benzer, evrişimsel yapay sinir ağları da öğrenebilen ağırlıklar ve bias değerleri içermektedir. Evrişimsel sinir ağları, Conv olarak adlandırılan katman içerisinde bulunan filtreler ile girdi görselini tarar ve bu filtre içerisinde bulunan ağırlıkları istenen çıktıya göre ve kayıp fonksiyonuna göre ayarlar. Buna ek olarak evrişimsel sinir ağları bazı hiper-parametreler de içermektedir.

3.2.1. Evrişimsel sinir ağları hiper-parametreleri

Girdi katmanı evrişimsel sinir ağlarının ilk katmanıdır ve bu katman görüntünün alındığı katmandır. Resim sayısal matrislere dönüştürülerek evrişim katmanına aktarılır. Evrişim katmanlarında, gizli nöronlar, evrişimsel filtreler ile değiştirilir. Burada yapay sinir ağlarında bulunan ağırlık terimi için değil, her bir filtrenin kendi ağırlığının olduğu bir filtre ailesi için ağırlık güncellemesi yapılır. Evrişim katmanları, işlenen görüntü için yükseklik, genişlik ve derinlik olmak üzere nöronları üç boyutlu olarak ayarlar. Ortaklama katmanı ise literatürde ayrıca pooling katmanı olarak bilinir. Bu katman boyut azaltma işlemi için kullanılır. Bu boyut azaltma genel olarak özellik kaybına neden olsa da ağı eğitiminin hızlandırılması için gereklidir. Ortaklama katmanında öğrenilecek parametre bulunmaz, sabit ortaklama boyutu ve adım hiper-parametreleri vardır. Yaygın olarak kullanılan ortaklama yöntemi maksimum ortaklama ve ortalama ortaklama yöntemidir. Maksimum ortaklama yönteminde belirli bir çerçeve içerisinde kalan değerlerden en büyük olan değerler seçilerek boyut azaltma işlemi gerçekleştirilir. Bu sayede özellik azalsa bile sonucu etkileyen en büyük özellikler kalmış olur. Tam bağlantılı katman ise evrişimsel sinir ağlarının son katmanıdır. Bu katman temel yapay sinir ağı mimarisine sahiptir. Bu katmanda elde edilen çıktıya göre hata değeri hesaplanır ve istenilen sonuca göre ağırlıklar yeniden hesaplanarak güncellenir. Temel bir evrişimsel sinir ağı mimarisi ve bazı hiper-parametreleri Şekil 3.3’de gösterilmiştir.



Şekil 3.3. Evrişimsel sinir ağı yapısı (Gülcü, A. , Kuş, Z. 2019)

3.2.1.1. Evrişimsel sinir ağlarında alıcı alan

CNN'lerdeki filtreler, tüm görüntüyü tipik bir evrişim işleminden geçirir. Filtre boyutu görüntüden çok daha küçük olduğu için, çözmesi gereken ağırlıkların sayısı önemli ölçüde azaltılır. Filtrelerin boyutu alıcı alan boyutuna göre belirlenir. Bu süreç, kedi ve maymun gibi hayvanların korteksindeki algılayıcı alanların arkasındaki bilimden esinlenerek, yakınlardaki görüntü bölgelerinin beyindeki aynı veya yakın nöronlarla eşleştirildiği gerçeğinden esinlenerek oluşturulmuştur (Hupel D, Wiesel T. 1962).

3.2.1.2. Evrişimsel sinir ağlarında adım

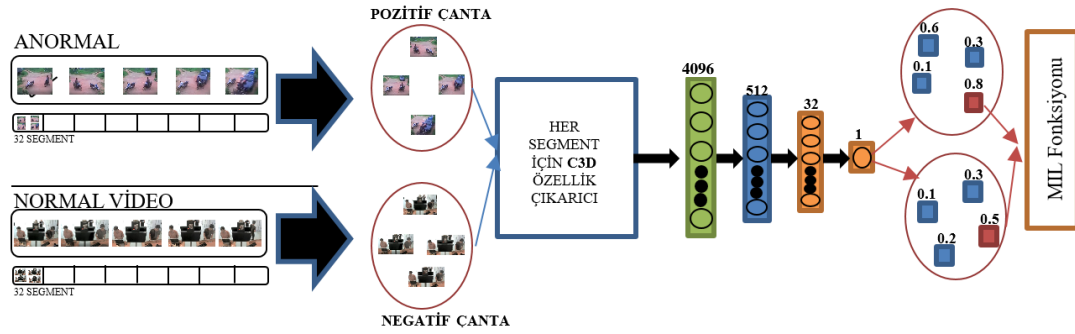
Derinlik, bir filtreye giren düzlemlerin sayısıyla belirlenirken, adım, evrişim gerçekleştirildiğinde görüntü üzerinde ki filtrenin kayacağı piksel değerini belirler (S. Albawi vd. 2017). Filtre genişliği, yüksekliği, derinliği ve adımı, 3D evrişim katmanını oluşturmak için kullanılır. Bir birim adım mesafesi, görüntü üzerinde evrişim maruz kalan bölgenin bir birim yanında diğer bir evrişim işleminin yapılacağını ifade eder. Düşük adım değerleri, filtrelerin daha sık örnek alıp nokta çarpımı yapacağı anlamına gelir, bunun neticesinde işlem yükü ve parametre sayısı artacaktır fakat üretilen çıktının çözünürlüğü artacağından eğitim uzun fakat yüksek doğrulukta olacaktır. Buna karşın, daha büyük adım değerleri, daha hızlı öğrenme sağlayacaktır fakat filtre girdi görüntüsü üzerinde fazla piksel atlayarak ilerleyeceğinden daha düşük çözünürlüklü bir çıktı üretecek ve bunun sonucunda bazı girdiler için karmaşıklık veya tanıyamama durumu gerçekleşecektir. Bu sebeple adım değeri büyük bir öneme sahiptir ve seçilirken istenen doğruluk oranı ve işlem gücü göz önüne alınmalıdır.

3.2.1.3. Evrişimsel sinir ağlarında sıfır dolgu

Sıfır dolgu, bazen girdi hacminin sınırının çevresinde dolgu yapmamız gerekebilir. Sıfır dolgu yapmanın güzel bir özelliği çıktı hacimlerinin boyutunu kontrol etmemize izin verir. 2 boyutlu bir görüntü için, sıfır dolgu 1 alınrsa, üst ve alt satıra bir, sola ve sağa bir sütun sıfır dolgusu ekler; böylece hem görüntünün yüksekliğini hem de genişliğini 2 artırır(S. Albawi vd. 2017). Sıfırdan büyük bir değere dolgu yapar Görüntünün kenarındaki bilgileri korumak için yararlı bir yöntemdir. Aynı zamanda çıkış hacmi olarak adlandırılan evrişim katmanlarının çıktısının boyutlarını korur.

3.3. Anormal Durum Tespiti

Anormal durumların tespit edilebilmesi için Şekil 3.4 ve Şekil 3.5’de gösterilen çalışmalar referans alınmıştır. Bu çalışmalar içerisinde W. Sultani vd. (2018) tarafından gerçekleştirilen araştırmada zayıf etiketleme metodu kullanılmıştır. Zayıf etiketleme metodu sayesinde anormal ve normal durum veri seti oluşturulurken videolar saniye saniye etiketlenmez, bunun yerine içerisinde anormal bir durum olduğu bilinen video anormal, içerisinde anormal bir durum olmadığı bilinen video ise normal olarak etiketlenir. Şekil 3.4’de gösterilen çalışmada anormal durum özellik çıkarım işlemi için C3D modeli kullanılmıştır. C3D modeli Du Tran vd. (2014) tarafından video görüntüleri içerisinden hareket tespiti ve sınıflandırılması için geliştirilmiş bir modeldir. C3D modelinin genel görünümü Şekil 3.5’de gösterilmiştir. Şekil 3.5’den de görüleceği üzere C3D modeli farklı sayıda evrişim katmanına ve çıktı katmanında 4096 sinire sahip bir modeldir.



Şekil 3.4. Anormal durum tespit mimarisi (W. Sultani vd. 2018)

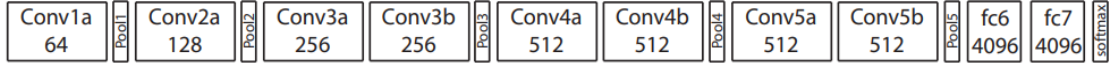
W. Sultani vd. (2018) çalışmalarında ağır eğitim aşamasında kullanmak için MIL fonksiyonunu geliştirmişlerdir. Bu fonksiyon sayesinde zayıf etiketlenmiş veri setleri ile modellerini eğitebilmişlerdir. Burada problem bir sınıflandırma problemi yerine bir regresyon problemi olarak ele alınmıştır ve regresyon problemlerinde destek vektör makinalarında olduğu gibi denklem 3.1’de gösterilen hinge kayıp fonksiyonu kullanılmaktadır.

$$\min_w \left[\frac{1}{k} \sum_{i=1}^k \max(0, 1 - y_i(w \cdot \phi(x) - b)) \right] + \frac{1}{2} \|w\|^2 \quad (3.1)$$

Burada k örneklerin sayısı, y_i her örneğin etiketi, $\phi(x)$ video segmentinin özneteliği, b bias ve w öğrenilecek ağırlığı ifade etmektedir. Hinge kayıp fonksiyonunda denklemden anlaşılacağı üzere videolardan elde edilen görüntü bloklarının tek tek etiketlenmesi gerekmektedir. Bu problemin üstesinden gelmek için denklem 3.2’de gösterilen MIL fonksiyonu önerilmiştir.

$$l(B_a, B_n) = \max\left(0, 1 - \max_{i \in B_a} f(V_a^i) + \max_{i \in B_n} f(V_n^i)\right) + \lambda_1 \sum_i^{(n-1)} \left(f(V_a^i) - f(V_a^{i+1})\right)^2 + \lambda_2 \sum_i^{(n)} f(V_a^i) \quad (3.2)$$

Bu yöntem sayesinde anormal durum tespiti regresyon problemi olarak ele alınmıştır. Sistemde, anormal durum içeren video bloğunun anormal durum içermeyen video bloğuna göre daha yüksek skor alması istenmektedir. Ranking loss yaklaşımı sayesinde anormal durum içeren video blokları kullanılarak elde edilen tahmin değerleri için ağ cesaretlendirilir, anormal durum içermeyen video blokları kullanılarak elde edilen tahmin değerlerine göre ağ cezalandırılır. Bu yöntemde çantada bulunan maksimum skor ele alındığından video bloğu seviyesinde etiketleme yerine video seviyesinde etiketleme kullanılabilir. Yani, 32 video bloğu için elde edilen tahmin değerleri anormal ve normal durum videoları için kıyaslanarak normal durum tahmin değerleri sıfıra yaklaştırılır, anormal durum tahmin değerleri bire yaklaştırılır. Anormal durumların bulunduğu çanta pozitif çanta, anormal durum bulunmayan çanta negatif çanta olarak isimlendirilir. Burada pozitif çantadaki en yüksek anormal durum skoru doğru pozitif, negatif çantadaki en yüksek anormal durum skoru ise yanlış pozitif olarak ele alınır. Negatif çantada bulunan yanlış pozitif değeri normal olup, anormal duruma en çok benzeyen durumlara karşılık gelmektedir. Bu durumun normal olduğu bilindiği için ileride yanlış alarm üretmemek amacıyla sistemden istenen bu iki değer arasındaki farkın açılması ve bu iki değer birbirinden ayrıştırılmasıdır. Anormal durum genelde kısa bir sürede ve normal duruma oranla çok daha seyrek gerçekleşmektedir, buna ek olarak anormal durum segmentler arasında geçiş yapabilir. Bu iki durumun üstesinden gelmek için temporal smoothnes ve sparsity-term ismi verilen yöntemler geliştirilmiştir.



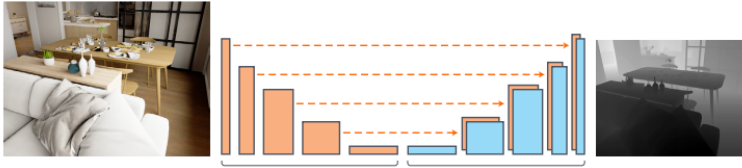
Şekil 3.5. C3D Mimarisi Du Tran vd. (2014)

Geliştirilen sistemde, video görüntüsü işlenmemiş şekilde modele uygulanır. Modele uygulanan video görüntüsü içerisinde bulunan bütün frameler bir diziye aktarılır ve toplam frame sayısı kayıt edilir. Diziye aktarılan frameler sliding window metodu ile 16 kare içeren klipler olacak şekilde kayıt edilir. Örneğin, 1000 frame içeren bir video 64 klibe ayrılır ve her klip 16 kare içermektedir. Elde edilen klipleri C3D ağına uygulayabilmek için kliplerin bulunduğu dizi 16x112x112x3 boyutuna ölçeklenerek bir diziye aktarılır ve bu dizi özellik çıkarım işlemi için C3D ağına uygulanır. C3D modeli Sports1M veriseti ile önceden eğitilmiş bir modeldir. C3D ağının çıkışından elde edilen tahmin değerleri bir dizi içerisinde tutulur ve tam bağlantılı sinir ağını eğitme aşamasında kullanılır. Tam bağlantılı sinir ağını MIL fonksiyonu ile eğitmek için veriseti 32 video bloğuna ayrıştırılarak gereken çantalar oluşturulur. Her bir çanta 32x4096 boyutunda bir diziden oluşmaktadır. Burada 32 değeri daha önceden belirlenmiş sabit bir değerdir, 4096 değeri ise C3D ağının çıkışından tam bağlantılı ağın girişine geçiş aşamasında kullanılan sinir sayısını ifade etmektedir. Bu sayede C3D modelinden elde edilen özellikler tam bağlantılı sinir ağına uygulanmıştır. Tam bağlantılı sinir ağı 32 farklı özellik için 1 anormal durum tahmini gerçekleştirilmiştir ve bu tahmin değerleri extrapolate fonksiyonu

ile bütün videoya ölçeklenmiştir. Anormal durum tespiti için kullanılan bu model UCF Crime veri seti ile eğitilmiştir. Tez kapsamında gerçekleştirilen çalışmalarda farklı lamda değerleri ve Adagrad ve Adadelta optimizasyon algoritmaları denenmiştir. Denemeler sonucu en iyi değeri veren optimizasyon algoritması Sultani vd. (2018) tarafından da bahsedilen Adagrad algoritmasıdır. Bu nedenle anormal durum tahmininde Adagrad algoritması kullanılmıştır. Diğer optimizasyon algoritması ile kıyaslama sonuçları bulgular ve tartışma bölümünde verilmiştir. Sparsity ve smoothnes parametreleri farklı değerler için denenmiştir ve bu değerler λ_1 ve λ_2 için $8 \cdot 10^{-5}$ alınmıştır. Tüm bu anormal durumları tespit eden anormal durum modeli 63,328,577 adet parametreye sahiptir.

3.4.Derinlik Tahmini

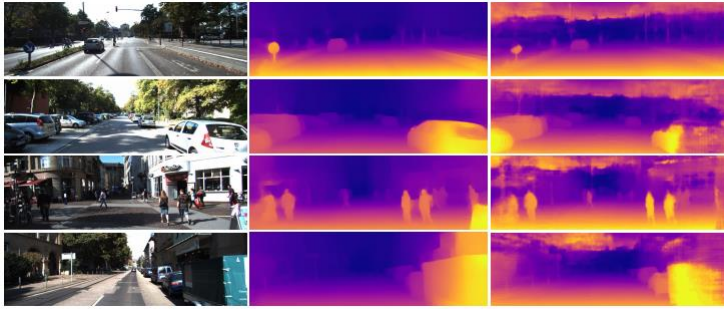
Görüntülerden yüksek doğrulukla derinlik tahmini çoğu uygulama için önemli bir adımdır. Video görüntülerinden derinlik tahmininin başlıca uygulama alanları, otonom araçlar ve sahne anlamlandırma uygulamalarıdır. Uygulamada var olan metotlar genellikle düşük çözünürlüklü ve bulanık haritalar oluşturmaktadır. Burada RGB resimden derinlik tahmini için yüksek çözünürlüklü derinlik haritası oluşturabilen bir metot kullanılmıştır.



Şekil 3.6. Girdi resmi, katmanları ve derinlik haritası (Alhashim, I. vd. (2018))

Uygulamada Alhashim, I. vd. (2018) tarafından tanıtılan ve Şekil 3.6'da gösterilen metot referans alınmıştır. Model şekil 3.6'da gösterildiği üzere enkoder katmanı ve dekoder katmanından oluşmaktadır. Uygulamanın enkoder katmanında ImageNet veri seti üzerinde önceden eğitilmiş olan DenseNet-169 modeli kullanılmıştır. DenseNet-169 modeli, çok katmanlı evrimsel sinir ağlarındaki çok sayıdaki parametrenin azaltılması ve kompakt bir yapıda bir modelin oluşturulması için Huang G. vd. (2017) tarafından geliştirilmiştir. DenseNet-169 modeli 169 katmana sahip olmasına rağmen diğer modeller ile kıyaslandığında çok daha az sayıda parametreye sahiptir. Modelin öne çıkan özelliği dense block(yoğun blok) olarak adlandırılan katmanlarıdır. Bu katmanlar normal bir evrimsel sinir ağına ki evrim katmanına benzer fakat burada her katman bir önceki katmanlara bağlıdır ve o katmanlardan bilgi akışı diğer katmanlara doğru sağlanır. Bu sayede her katman bilgiyi bir biri ile paylaşmış olur ve bu bilgi paylaşımı sayesinde gradyan yok olması problemi de ortadan kaldırılmış olur. Enkoder katmanında öznetelik çıkartmak için kullanılan DenseNet-169 ağından elde edilen değerler yukarı örnekleme katmanlarına girdi olarak uygulanır ve nihayetinde çıktı olarak derinlik haritası tahmin edilir. Ağı eğitim aşaması için en önemli etkenlerden biri olan kayıp fonksiyonu da bu yapı için özel olarak geliştirilmiştir. Kayıp fonksiyonu ağı eğitim hızı ve tahmin doğruluğu üzerinde direk etkiye sahiptir Huang G. vd. (2017). Literatürde ki çalışmada kayıp fonksiyonu, derinlik haritasını tahmin ederken aynı zamanda resimdeki nesnelere sınır detaylarını bozmaması için cezalandırarak şekilde geliştirilmiştir. Uygulamada kullanılacak olan model NYU Depth v2 veri seti üzerinde eğitilmiştir ve eğitim ağırlıkları

modelde direkt olarak kullanılmıştır. Anormal durum tahmini için hazırlanan her klipten bir adet frame alınarak bilgisayara kayıt edilir ve bu görüntüler derinlik modeline uygulanır. Derinlik modeli 42,657,689 adet parametreye sahiptir ve çıktı olarak 32x24x320x1 boyutunda bir dizi derinlik haritası döndürür. Derinlik haritasına elde edilen mesafe bilgisi kameranın görüş açısına bağlı olarak değişebilmektedir, geniş açıyı gören bir kamerada uzak bilgisi 500m olurken dar alanda çalışan bir kamerada uzak bilgisi 50m mesafeye karşılık gelebilmektedir.



Şekil 3.7. Girdi resmi, tahmin edilen derinlik haritası, DORN modelinin çıktısı (Alhashim, I. vd. (2018))

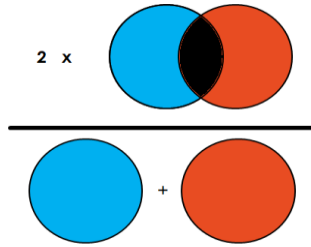
Tez kapsamında kullanılan ağırlıklar Alhashim, I. vd. (2018) tarafından eğitim sonucu oluşturulmuştur ve bu ağırlıklar oluşturulurken 4 adet 12 GB RAM belleğe sahip NVIDIA TITAN XP ekran kartları kullanılmıştır. Eğitilen model, NYU Depth v2 veri seti üzerinde 1M iterasyon gitmiştir ve bu eğitimin tamamlanması 20 saat sürmüştür. Alhashim, I. vd. (2018) tarafından elde edilen çıktı görselleri Şekil 3.7’de gösterilmiştir. Normal kullanım için geliştirilen bilgisayarların bu iş yükü altından kalkması pek mümkün görünmemektedir. Bu tez kapsamında Alhashim, I. vd. (2018) tarafından oluşturulan ağırlıkların kullanılmasının nedeni hesaplama maliyetinin yüksek olmasıdır.

3.5. Anomali Lokalizasyon

Anormal durum gerçekleştiği esnada anormal durumun görüntü içeriğinde ki konumunu tespit etmek için lokalizasyon işlemi gerçekleştirilmiştir. Görüntü işleme alanında kullanılan bazı terimler kafa karıştırıcı olmaktadır. Bu terimler; resim sınıflandırma, nesne konumlandırma ve nesne tespittir. Resim sınıflandırma, görüntünün ait olduğu sınıfın tespit edilmesi işlemidir. Nesne konumlandırma ise resimde bulunan bir veya daha fazla nesnenin etrafında sınır çizgisi çizerek konumunun tespit edilmesi işlemidir. Nesne tespiti ise en zor olan işlemidir, bu iki özelliğin birleştiği bu teknikte resim içerisinde ki nesnenin hem sınıfı hem de konumu belirlenmektedir.

Bu tez çalışmasında, video içerisinde gerçekleşen anormal durumun konumlandırılması için MobileNetV2 modelini temel alan bir model kullanılmıştır. Bu amaç için alternatif birçok yapay sinir ağı modeli bulunsa da genelde bu modeller yüksek hesaplama maliyetlerine ve kaynak tüketimlerine sahiptirler ve bu hesaplama maliyeti mobil ve gömülü cihazların hesaplama kapasitelerinin çok ötesindedir. Bu sorunun üstesinden gelmek için MobileNetV2 modeli geliştirilmiştir ve daha düşük hesaplama işlemleri ve bellek tüketimine rağmen başarılı konumlandırma sonuçları verebilmektedir. Uygulamada kullanılan model de MobileNetV2 modeline ek olarak çıktı katmanında

yukarı örnekleme, evrişim ve normalizasyon yöntemleri uygulanmıştır. Bu işlemler sayesinde normalde 7x7x160 olan ağ çıktısının boyutu öncelikle 14x14x352 daha sonra da 28x28x256 boyutuna yükseltilmiştir. Ağın çıktı katmanında ise 28x28 boyutunda bir çıktı elde edilmiştir. Bu çıktı değeri 28x28 boyutuna sahip bir haritaya karşılık gelmektedir ve bu harita üzerinden konumlandırma yapılmıştır. Çıktı haritası için farklı boyutlar denenmiştir fakat en başarılı çıktı boyutunun 28x28 olduğu gözlemlenmiştir. Burada lokalizasyon problemi segmentasyon temelli bir çözüm ile ele alındığı için özel bir kayıp fonksiyonuna ihtiyaç duymaktadır. Bunun nedeni, piksel temelli bir performans ölçümüne ihtiyaç duyulmasıdır. Piksel doğruluğunun yüksek olması her zaman segmentasyon başarısının yüksek olacağı anlamına gelmeyeceği için dice katsayısı ismi verilen bir fonksiyon geliştirilmiştir ve fonksiyonun görsel gösterimi Şekil 3.8’de gösterilmiştir. Dice katsayısı, örtüşen alanın iki katının toplam piksel sayısına bölümüdür. Sistemde çıktı, bir ve sıfır değerleri ile ifade edildiği için binary crossentropy fonksiyonu performansı arttırmak için kayıp fonksiyonuna eklenmiştir.



Şekil 3.8. Dice katsayısının gösterimi(Zou vd. 2004)

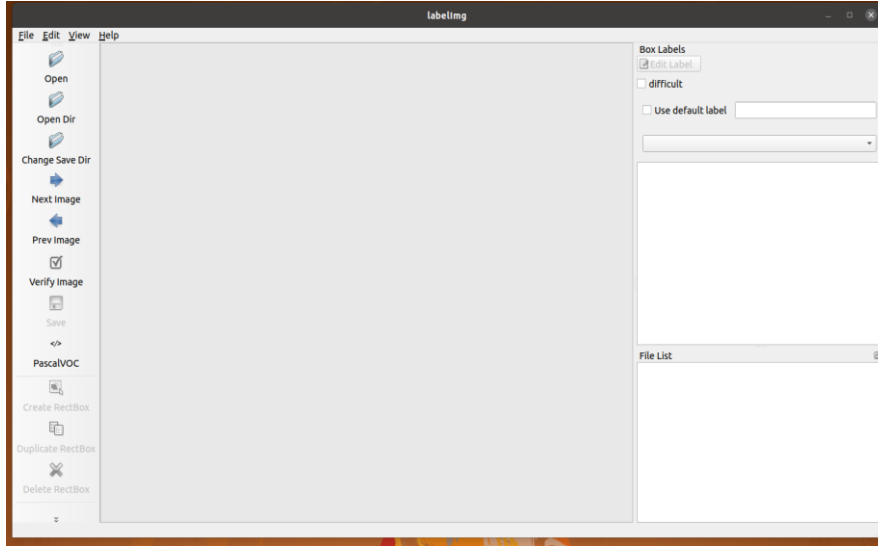
Ağın eğitim süresine ve tahmin performansına direkt olarak etkisi olan kayıp fonksiyonu denklem 3.3 de gösterilmiştir.

$$loss = BC(f(y_{true}), f(y_{pred})) - \log\left(\frac{2(\sum y_{true} * y_{pred})}{(\sum y_{true} + y_{pred})}\right) \quad (3.3)$$

Formülden anlaşılacağı üzere dice katsayısı segmentasyon örtüştükçe 1’e yakınsayacaktır ve bunun neticesinde kayıp fonksiyonunun hataya etkisi de sıfıra yakınsayacaktır. Modelin eğitim aşamasında performansı da bu şekilde ölçülmüştür.

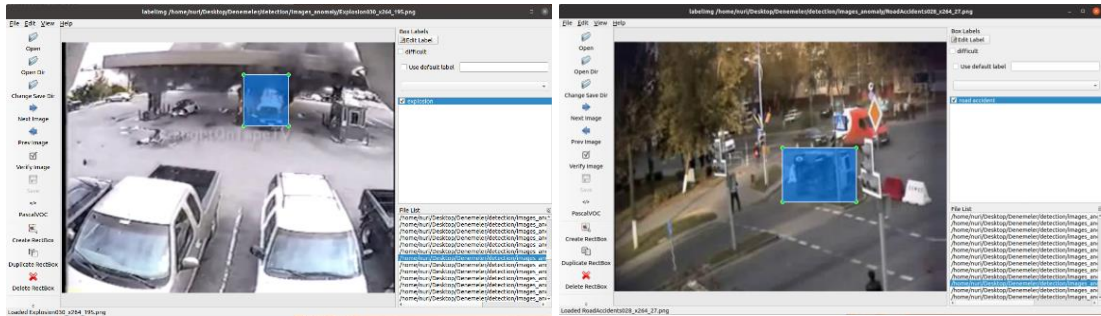
Uygulamada, anormal durum tahmini için kayıt altına alınan görüntüler orijinal boyutlarında modele uygulanır ve bu görüntüler daha sonra 224x224 boyutuna yeniden boyutlandırılır. Yeniden boyutlandırılan görüntü 0-255 arasında değerlere sahiptir. Daha sonra görüntüler tip dönüşümü ile float tipinde görüntülere dönüştürülerek -1 ve 1 değerleri arasında float değerlere sahip olur. Tip dönüşümü ve yeniden boyutlandırma işlemi yapılan görüntüler MobileNetV2 modeline ve bu modelin devamına eklenen yukarı örnekleme modeline uygulanarak 28x28x1 boyutunda bir görüntü haritası oluşturması sağlanır. Elde edilen bu görüntü haritası değerleri 0,5 sınır değerine göre 1 ve 0 değerleri arasında yeniden binary tipine dönüştürülür. Görüntü haritası içerisinde bulunan 1 değerleri görüntüde aranan nesnenin muhtemel olarak bulunduğu noktalara karşılık gelmektedir. Bu görüntü haritasında kontur noktaları 1 ve 0 değerlerine göre tespit edilerek görüntü üzerine sınır çizgilerinin çizilmesi sağlanır. Tüm bu işlemleri gerçekleştiren konumlandırma modeli 3,803,265 adet parametreye sahiptir.

Model oluşturulduktan sonra eğitim aşamasına geçilebilmesi için veri toplamak gerekmektedir. Veri toplama işlemi için arayüzü Şekil 3.9’da gösterilen LabelImg uygulaması kullanılmıştır. LabelImg uygulaması, Python yazılım dili ile Qt grafik kütüphanesini kullanarak yazılmış, MIT lisansı ile lisanslanmış ücretsiz bir etiketleme aracıdır (Tzotalin 2015) .



Şekil 3.9. LabelIMG Programı

Uygulama ile elimizde bulunan görüntüler sınıflandırılabilir ve sınır çizgileri ile nesne konumlandırma uygulaması için etiketlenebilir. Uygulama YOLO ve PascalVOC formatlarında çıktı verileri üretebilmektedir. Uygulamamızda PascalVOC formatında etiketlenmiş veriler kullanılmıştır ve elimizde bulunan görüntüler bu formatta etiketlenmiştir. Veri etiketleme aşaması öncesi anormal durum içeren resimleri elde etmek amacıyla tüm anormal durum videolarını işleyen ve bu video görüntülerinden anormal durum görüntülerini ayırarak bu görüntüleri PNG formatında kayıt eden bir kod yazılmıştır. Bu kod veri seti üzerinde çalıştırılarak patlama ve trafik kazası veri setleri içerisinde anormal durum içeren görüntüler ayrı bir klasöre kayıt edilmiştir. Eğitim aşaması için toplam 2226 adet anormal durum içeren görüntü elde edilmiş ve içlerinde ki anormal durumlar tek tek etiketlenmiştir. Etiketleme işlemi neticesinde her görüntü için XML formatında bir dosya oluşturulmaktadır. Bu dosya içerisinde dosya adı, klasör adı, resim boyutları, etiket ve kare içine alınan konum değerleri bulunmaktadır.

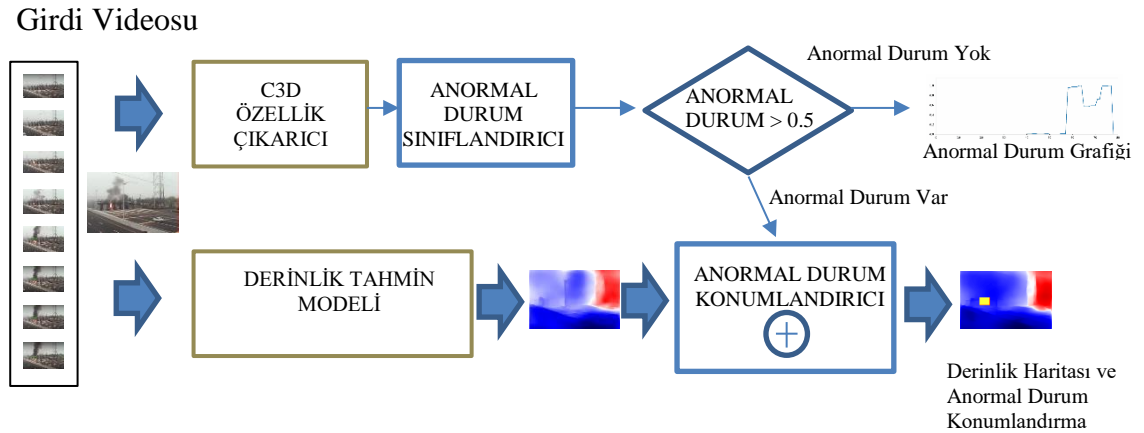


Şekil 3.10. Patlama ve trafik kazası etiketleme işlemi

Veri setleri Şekil 3.10’da gösterildiği şekilde etiketlendikten sonra eğitim aşamasına geçilmiştir ve oluşturulan veri seti üzerinde eğitim işlemi gerçekleştirilmiştir. Eğitim aşaması tamamlanan model anormal durum tespit ve derinlik tahmin modeline entegre edilmiştir.

3.6. Geliştirilen Sistem

Anormal durumların tespit edilmesi için geliştirilen sistemde, video görüntüsü uygulamaya yüklenir ve video görüntüsü, önce kliplere daha sonra da video bloklarına ayrılır. Elde edilen bu video blokları anormal durum tespit işlemi için C3D modeline girdi olarak verilir. C3D modelinden elde edilen çıktı 4096 adet çıktı değeri tam bağlantılı sinir ağına uygulanır ve videonun ilgili bloğu için anormal durum tespit işlemi gerçekleştirilir. Eğer videonun ilgili anında anormal durum tespit edilirse anormal durumun görüntü içerisinde konumlandırılması için ilgili video bloğunun ilgili kareleri MobileNetV2 modeline uygulanır ve anormal durumun gerçekleştiği anda anormal durumun görüntü içerisinde nerede gerçekleştiğine dair tahmin işlemi gerçekleştirilir ve anormal durum kare içerisine alınarak görüntü üzerine işlenir. Anormal durumun gerçekleştiği mesafesinin tespit edilebilmesi için, elde edilen video bloklarından bazı görüntü kareleri kayıt altına alınır ve bu görüntüler DenseNet-169 modeli ile bu modele eklenmiş yukarı örnekleme katmanlarına uygulanır. Bu sayede her girdi görüntüsü için bir derinlik haritası tahmin edilir. Son olarak derinlik haritası üzerine önceden tahmin edilen anormal durum konumu işaretlenerek muhtemel mesafe kestirim işlemi gerçekleştirilir. Bu çalışmada önerilen model Şekil 3.11’de gösterilmiştir.



Şekil 3.11. Geliştirilen modelin akış diyagramı

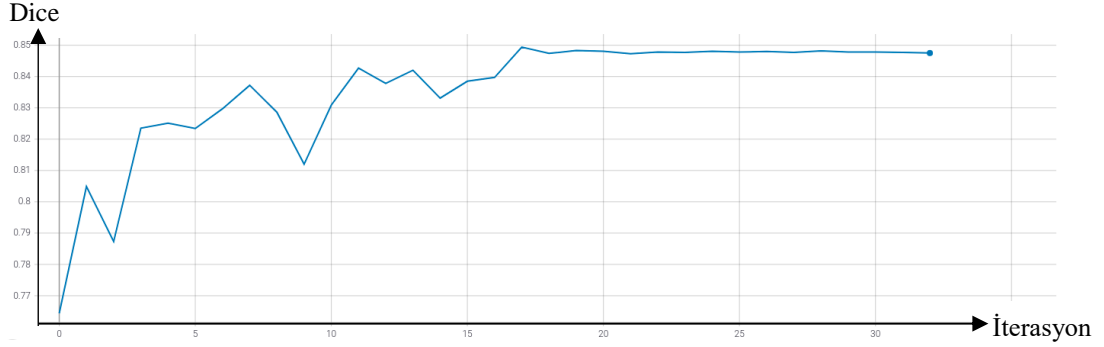
Şekil 3.11’de gösterilen sistemde kullanılan veri setlerini ve yöntemleri kısaca özetlemek gerekirse, anormal durum tespit işlemi için özellik çıkarıcı olarak Sports1M veri seti üzerinde daha önceden eğitilmiş olan C3D modeli kullanılmıştır. Bu modelden elde edilen değerler Anormal durum sınıflandırıcı olarak gösterilen tam bağlantılı sinir ağına uygulanır. Tam bağlantılı sinir ağı Ucf-Crime veri seti ile metaryal ve metot kısmında bahsedilen bilgisayar ile eğitilmiştir. Derinlik tahmin modeli ise özelleştirilmiş bir modeldir. DenseNet-169 modeli ile bu modelin sonuna eklenen yukarı örnekleme katmanından oluşmaktadır. Bu modelde kullanılan DenseNet-169 mimarisi ImageNet veri seti ile önceden eğitilmiş bir modeldir. Tez çalışmasında önerilen sistemde kullanılan

derinlik tahmin modeli ise NYU-Depth V2 veri seti ile eğitilmiş bir modeldir. Anormal durum konumlandırıcı modeli ise MobileNetV2 modeli ve bu modelin sonuna eklenmiş yukarı örnekleme katmanlarından oluşmaktadır. Bu modeli eğitmekte kullanılan veriseti anormal durumlar etiketlenerek yoğun bir çalışma sonucunda oluşturulmuştur. Bu üç modelin bir araya getirilmesi ile Şekil 3.11’de gösterilen sistem tez çalışması kapsamında geliştirilmiştir.

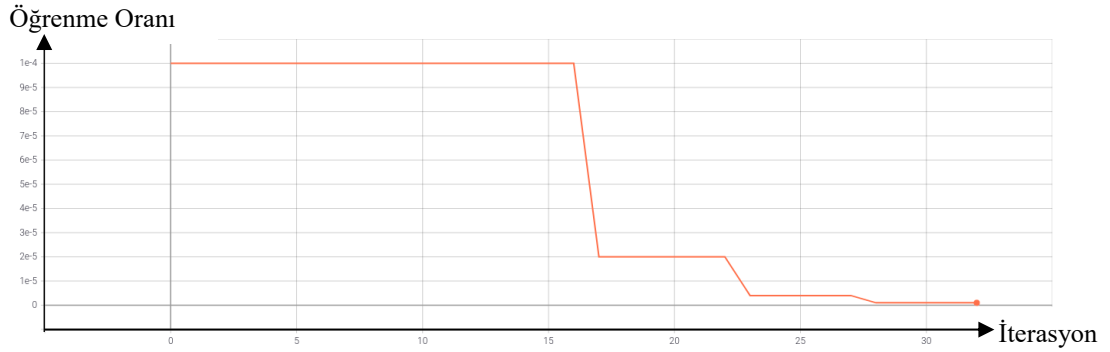
4. BULGULAR VE TARTIŞMA

Sistem tasarlandıktan sonra veri setleri ile eğitim aşamasına geçilmiştir. Eğitim aşamasına anormal durum konumlandırma modelinin eğitim işlemi ile başlanmıştır. Modelin eğitim işlemi sırasında göz önüne alınan parametreler şunlardır; modelin çıktı evrişim katmanı filtre sayısı, modelin toplam parametre sayısı, batch boyutu, alfa parametresi, öğrenme oranı ve dice katsayısıdır.

Batch boyutu 8, alfa değeri 1, evrişim katmanı filtre sayısı 256, toplam parametre sayısı 3,803,265 olacak şekilde eğitim gerçekleştirilmiştir. Eğitim aşaması neticesinde segmentasyon performansı Şekil 4.1’de belirtilmiştir.

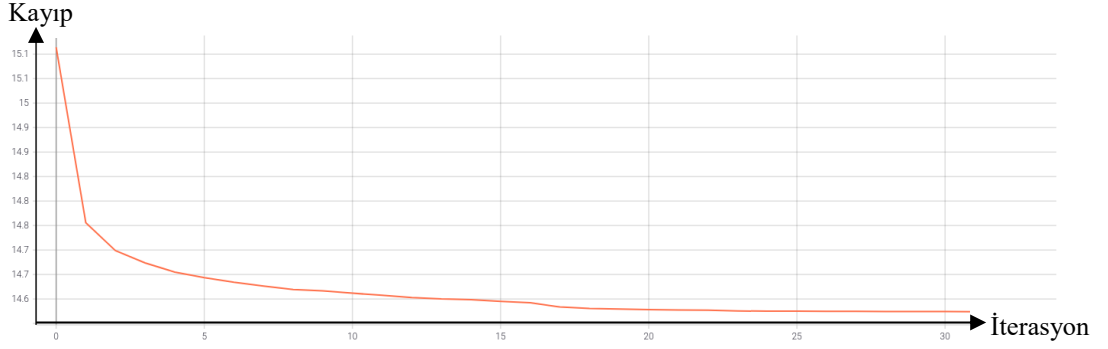


Şekil 4.1. Dice katsayısı lokalizasyon performansı



Şekil 4.2. Konumlandırma ilk eğitim aşaması boyunca öğrenme oranı

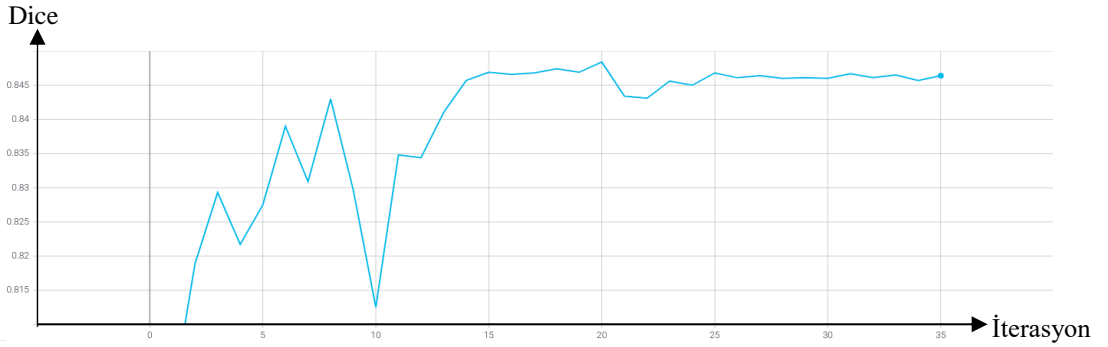
Model, 17 iterasyonda 0.85 dice katsayısı değerine ulaşmıştır daha sonra model plato (plateaus) da sıkışmış ve eğitim performansını arttıramamıştır. Platodan kurtulmak için öğrenme oranı eğitim süresi boyunca Şekil 4.2’de görüleceği üzere aktif şekilde değiştirilmiştir, bu sayede platodan kurtulma ve tahminde iyileşme hedeflenmiştir fakat 0,85 değeri aşamamıştır.



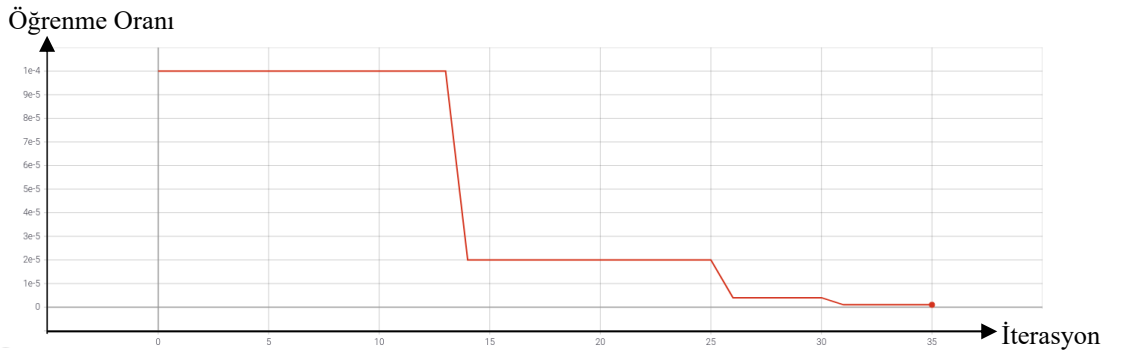
Şekil 4.3. Konumlandırma ilk eğitim aşaması boyunca kayıp fonksiyonu çıktısı

Eğitim süresince kayıp fonksiyonunda ki değişim Şekil 4.3’de gösterilmiştir. Dice katsayısının değişmemesi nedeniyle toplam kayıp fonksiyonun çıktısı da plato yapmıştır. Kayıp fonksiyonu değeri 33 iterasyon sonucunda 14.5 değerine sabitlenmiştir. Tahmin performansının artırılması için parametreler değiştirilerek yeniden eğitim işlemi gerçekleştirilmiştir.

Batch boyutu 8, alfa değeri 0,5, evrişim katmanı filtre sayısı 256, toplam parametre sayısı 3,803,265 olacak şekilde eğitim tekrar gerçekleştirilmiştir.



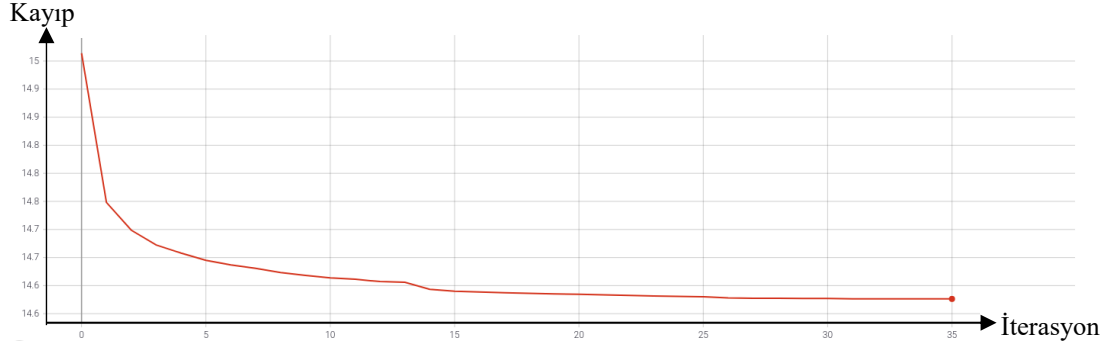
Şekil 4.4. Konumlandırma ikinci eğitim aşaması boyunca dice katsayısı çıktısı



Şekil 4.5. Konumlandırma ikinci eğitim aşaması boyunca öğrenme katsayısı çıktısı

Alfa değeri MobileNetV2 modelinde ağın genişlik katsayısı olarak bilinmektedir (Howard A. G. vd. 2017). Bu katsayısının artması ağın her katmanında ki filtre sayısını artırır, azaltılması ise her katmandaki filtre sayısını azaltır. Alfa değeri 0,5 yapılarak model yeniden eğitildiğinde, Şekil 4.4’de görüleceği üzere ağın platoya ulaşması daha

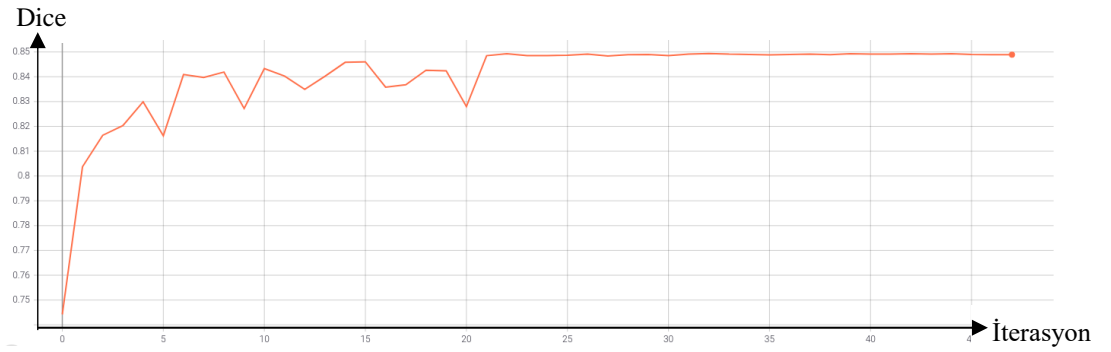
uzun sürmüştür ve iterasyon ilerledikçe modelin dice katsayısında değer azalmaları gözlemlenmiştir. Şekil 4.5’de görüleceği üzere öğrenme oranı yine aktif olarak değiştirilmiştir.



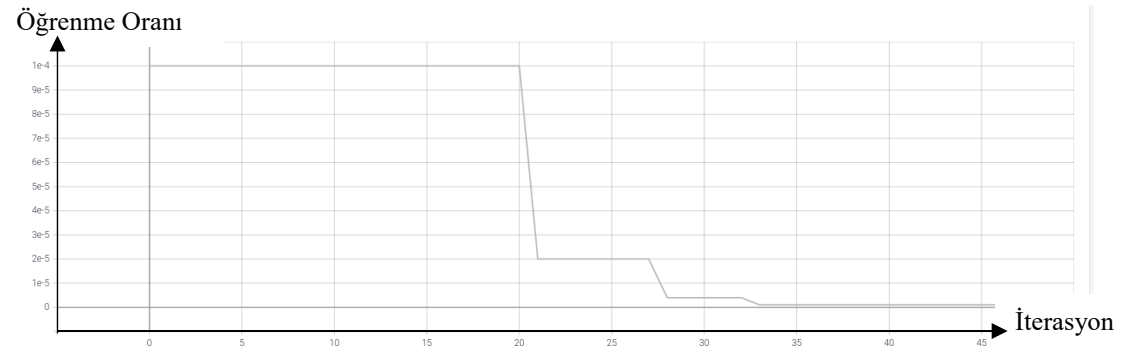
Şekil 4.6. Konumlandırma ikinci eğitim aşaması boyunca kayıp fonksiyonu çıktısı

Alfa değerindeki 0,5 puanlık azalma Şekil 4.6’da görüleceği üzere ağırlık eğitimi süresince kayıp fonksiyonunda belirgin bir değişime neden olamamıştır.

Batch boyutu 16, alfa değeri 1, evrişim katmanı filtre sayısı 256, toplam parametre sayısı 3,803,265 olacak şekilde sadece batch boyutu artırılarak eğitim işlemi tekrarlanmıştır. Batch boyutu ağırlık güncelleme işlemi gerçekleştirilmeden önce ağırlık gönderilen örnek sayısıdır ve bu değer artırılarak 2 katına çıkarılmıştır.

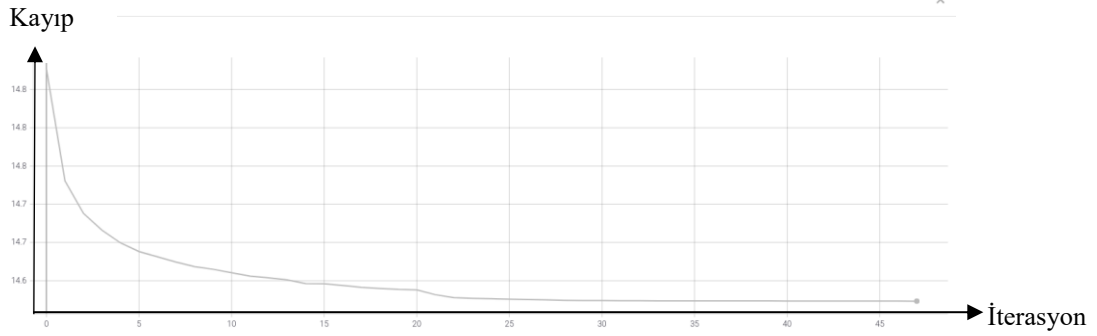


Şekil 4.7. Konumlandırma üçüncü eğitim aşaması boyunca dice katsayısı çıktısı



Şekil 4.8. Konumlandırma üçüncü eğitim aşaması boyunca öğrenme katsayısı çıktısı

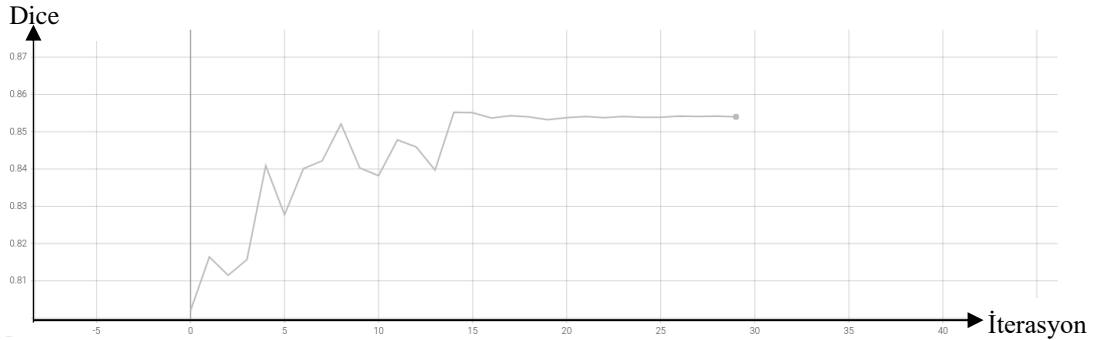
Ağa tek seferde daha fazla örnek göndermek iterasyon sayısını arttırarak eğitim aşamasının uzamasına neden olmuştur. Önceki eğitimlerde genelde 15 iterasyonda azalan öğrenme oranı Şekil 4.8’de görüleceği üzere 20 iterasyon sonra azalmaya başlamıştır. Öğrenme oranının değişimi Şekil 4.7’de gösterilen dice katsayısının değişimine bağlıdır.



Şekil 4.9. Konumlandırma üçüncü eğitim aşaması boyunca kayıp fonksiyonu çıktısı

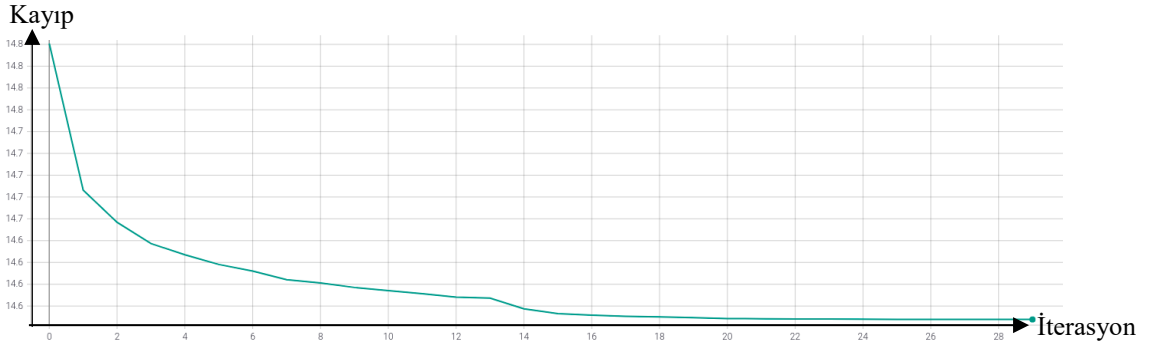
İlk iki eğitim aşamasında da gerçekleşen dice katsayısının platoya takılması problemi nedeniyle Şekil 4.9’da görüleceği üzere kayıp fonksiyonu belirli bir noktadan sonra sifıra yakınsayamamaktadır.

Batch boyutu 8, alfa değeri 1, evrişim katmanı filtre sayısı 512, toplam parametre sayısı 9,780,609 olacak şekilde eğitim tekrar gerçekleştirilmiştir. Son yukarı örnekleme katmanında ki filtre sayısı arttırılarak eğitim performansı gözlemlenmiştir. Filtre sayısı arttırılınca modelin toplam parametre sayısı 3,803,265 iken 9,780,609 olmuştur ve yaklaşık 2,57 katına çıkmıştır. Örnekler ağa uygulanmış ve aşağıda ki bulgular gözlemlenmiştir.



Şekil 4.10. Konumlandırma üçüncü eğitim aşaması boyunca dice katsayısı çıktısı

Parametre sayısı artmasına bağlı olarak ağın dice katsayısı performansı artmıştır. İlk üç eğitim sonucu dice katsayısı 0,85 değerine ulaşırken bu model Şekil 4.10’da görüleceği üzere 0,855 değerine ulaşmıştır. Öğrenme katsayısı aktif olarak bu eğitim aşamasında da değiştirilmiştir.



Şekil 4.11. Konumlandırma dördüncü eğitim aşaması boyunca kayıp fonksiyonu

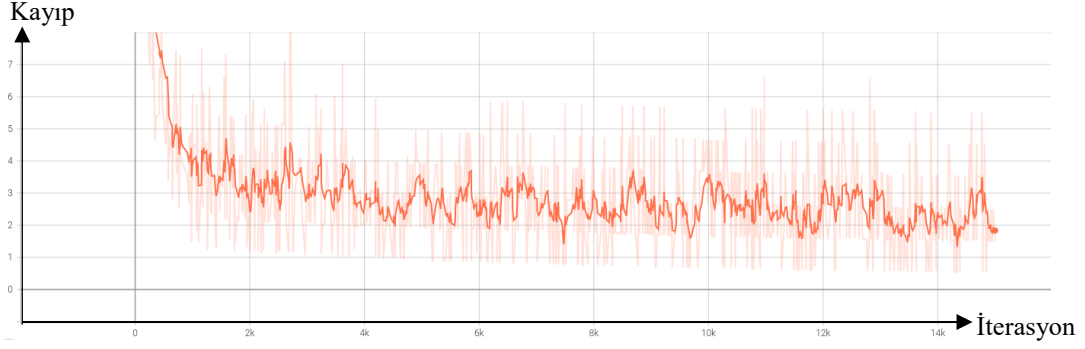
Filtre sayısının arttırılmasına bağlı olarak parametre sayısının arttırılması Şekil 4.11’de görüleceği üzere kayıp fonksiyonun sifıra yakınsamasına yardımcı olmamıştır. Parametre sayısının artmasına bağlı olarak eğitim süresi uzamıştır. Konumlandırma modelinin performansının arttırılması için ayrıca veri setinde bulunan örnek sayısının arttırılması gerekmektedir. Modelin parametreleri değiştirilerek elde edilen en büyük dice katsayı değeri 0,855’dir. Buna bağlı olarak örnek bir görüntü girişi için Şekil 4.12’de görüleceği üzere konumlandırma hataları gerçekleşmektedir.



Şekil 4.12. Yanlış ve doğru konumlandırma sonuçlarının aynı anda olduğu görüntü

Konumlandırma modelinin eğitim aşaması tamamlanarak anormal durum tespit modelinin eğitim işlemine geçilmiştir. Anormal durumun video içerisinden tespit edilebilmesi için eğitilen modelde üç farklı eğitim aşaması denenerek en iyi model kayıt altına alınmıştır.

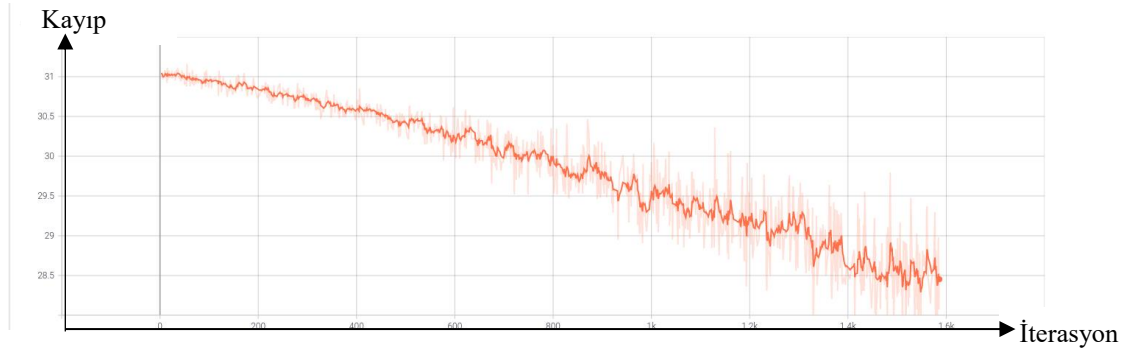
Anormal durum tespiti için optimizör fonksiyonu Adagrad, lambda değeri ise $8 \cdot 10^{-5}$ seçilmiştir ve ağı eğitim verileri uygulanmıştır. Eğitim verileri Ucf-Crime veri setinden elde edilmiştir.



Şekil 4.13. Anormal durum tahmin modeli kayıp fonksiyonu grafiği

Anormal durum tespiti için gerçekleştirilen ilk eğitim aşaması Şekil 4.13’de görüleceği üzere 15 bin iterasyon sürmüştür ve kayıp fonksiyonu değeri 35 değerinden başlayarak bazı video blokları için 1 değerinin de altına inmiş ve sıfıra yakınsamıştır.

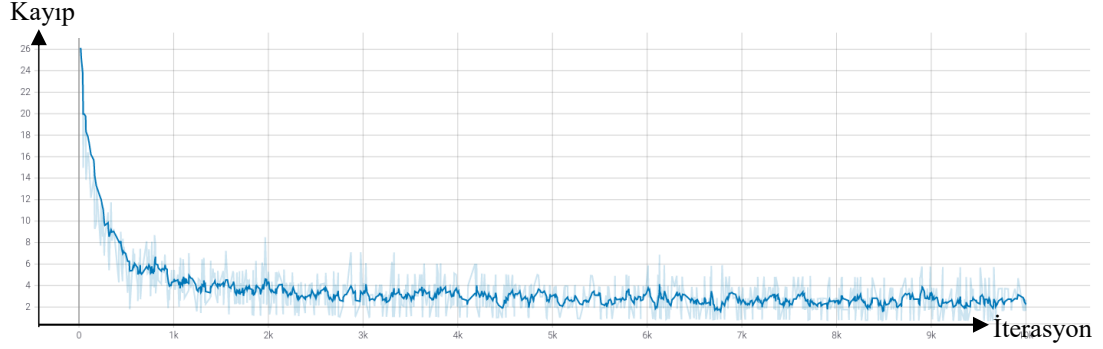
Anormal durum tespiti için optimzer fonksiyonu Adadelata, lambda değeri ise $8 \cdot 10^{-5}$ seçilmiştir ve ağa eğitim verileri uygulanmıştır. Eğitim performansını arttırmak için optimize edici fonksiyon olarak Adagrad fonksiyonu yerine Adadelata fonksiyonu denenmiş ve performansı net ölçebilmek için eğitim oranı parametresi sabit tutulmuştur.



Şekil 4.14. Anormal durum tahmin modeli adadelata optimzer için kayıp fonksiyonu

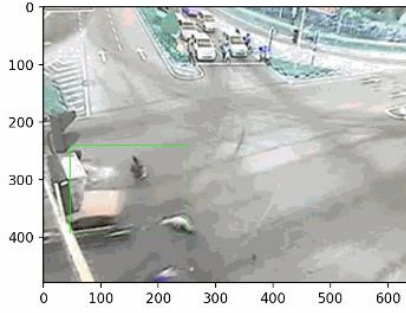
Adadelata optimzer ile eğitim sonucu 1600 iterasyon gidilmiştir fakat Şekil 4.14’de görüleceği üzere istenen performans sağlanamadığı için eğitim yarıda kesilmiştir. Adagrad optimzer fonksiyonunun Şekil 4.13’de 35 iterasyonda ulaştığı değere adadelata optimzer fonksiyonu Şekil 4.14’de görüleceği üzere 1600 iterasyonda ulaşamamıştır.

Anormal durum tespiti için bu eğitim aşamasında optimzer fonksiyonu Adagrad, lambda değeri ise $4 \cdot 10^{-5}$ seçilmiştir. Video görüntüsü alınarak önce video kliplere daha sonra da video bloklarına ayrılmaktadır. Anormal durum tespiti video blokları üzerinden gerçekleştirildiğinden video blokları arasında geçiş yapan bir anormal durum olduğunda bu durumu tespit etmek için lambda değeri diye adlandırılan bir parametre kayıp fonksiyonuna eklenmiştir. Lamda değeri, aynı zamanda kısa süreli gerçekleşen ve tespiti zor olan anormal durumların tespit edilmesinde de kullanılır. Bu parametrenin artırılması sonucu kayıp fonksiyondaki çarpan etkisi de artmıştır ve bu sayede video blokları arasındaki geçişlerde gerçekleşen anormal durum tespit hassasiyetinin artırılmasını sağlamıştır.



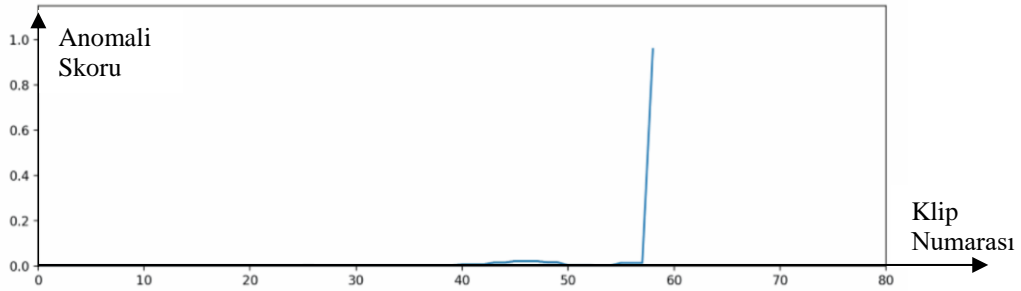
Şekil 4.15. Farklı lambda değeri için kayıp fonksiyonunun grafiği

Lambda değerinin azaltılması sonucu Şekil 4.15’de görüleceği üzere kayıp fonksiyonunda ki salınım değerinin azaldığı gözlemlenmiştir. Bu da lambda değerinin azaltılmasının hassasiyeti azalttığı anlamına gelmektedir.



Şekil 4.16. Anormal durum gerçekleşme anı

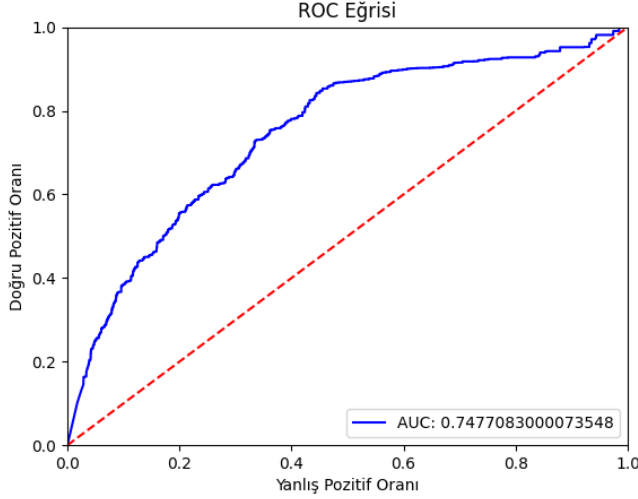
Anormal durum tespit modeli çalıştığında Şekil 4.16’de gösterilen görüntüye benzer bir görüntü üretmektedir. Şekil 4.17’de görüldüğü üzere anormal durum gerçekleştiği anda sistem onu tespit etmiş ve sinyal vermiştir.



Şekil 4.17. Anormal durum gerçekleştiği anda oluşan sinyal

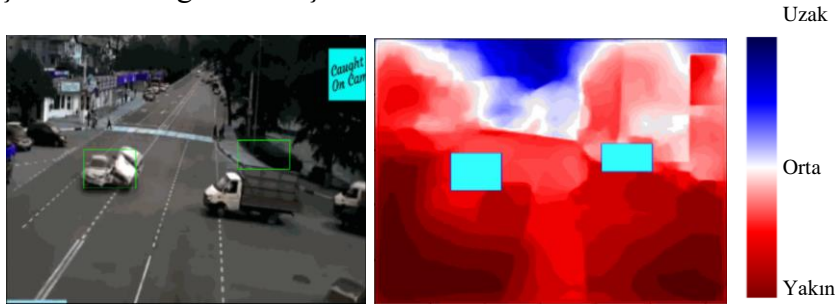
Anormal durum tahmin modeli test veri seti üzerinde denenmiştir. Test veri seti anormal durum videolarının adımları, anormal durumun gerçekleştiği sahneleri içeren etiketli veri setidir. Veri seti ile gerçekleştirilen test sonucunda elde edilen ROC grafiği Şekil 4.18’de gösterilmiştir. Bu eğri doğru tahmin edilen verilerin yanlış tahmin edilen verilere oranını göstermektedir. Burada başarılı bir modelin 1 değerine yaklaşması istenmektedir. Bu eğri altında kalan alan ise AUC olarak adlandırılır ve iki grubun ne

kadar başarılı şekilde ayırt edilebildiğinin bir ölçütüdür. Modelimizde AUC değeri 0.74 olarak elde edilmiştir. Sultani vd. (2018) çalışmalarında AUC değerini 0.7541 olarak elde etmişlerdir. Bu çalışma kapsamında Şekil 4.18’de görüleceği üzere AUC değeri 0.7477 olarak elde edilmiştir. Bu sonucun elde edilmesinin en etkili nedeni eğitim aşamasında kullanılan iterasyon sayısıdır.



Şekil 4.18. Anormal durum tahmin modeli ROC grafiği

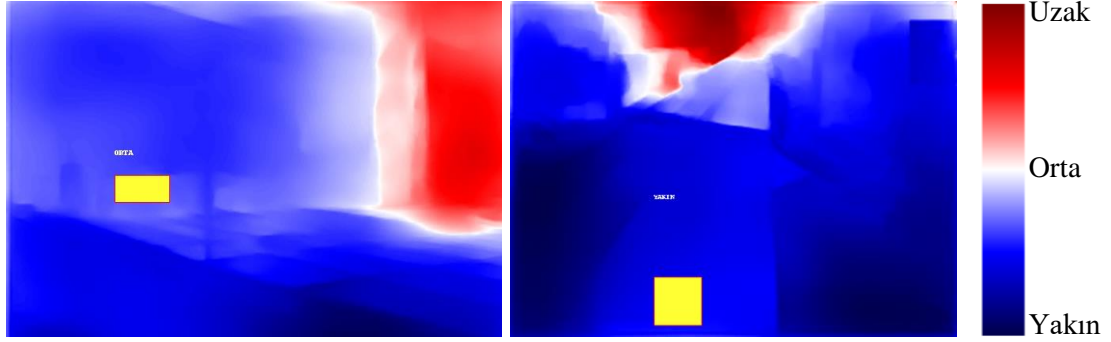
Derinlik tahmini için geliştirilen modelde önceden eğitilmiş ağırlıklar kullanılmıştır. Sistem tarafından girdi görseline göre derinlik haritası oluşturmaktadır. Derinlik veri setleri genelde araç içi veya araç üzeri kameralardan alınan görüntülerden oluşmaktadır. Sistemde kullanılan derinlik modelinin çıktısı ve konumlandırma sonucu Şekil 4.19’da gösterilmiştir.



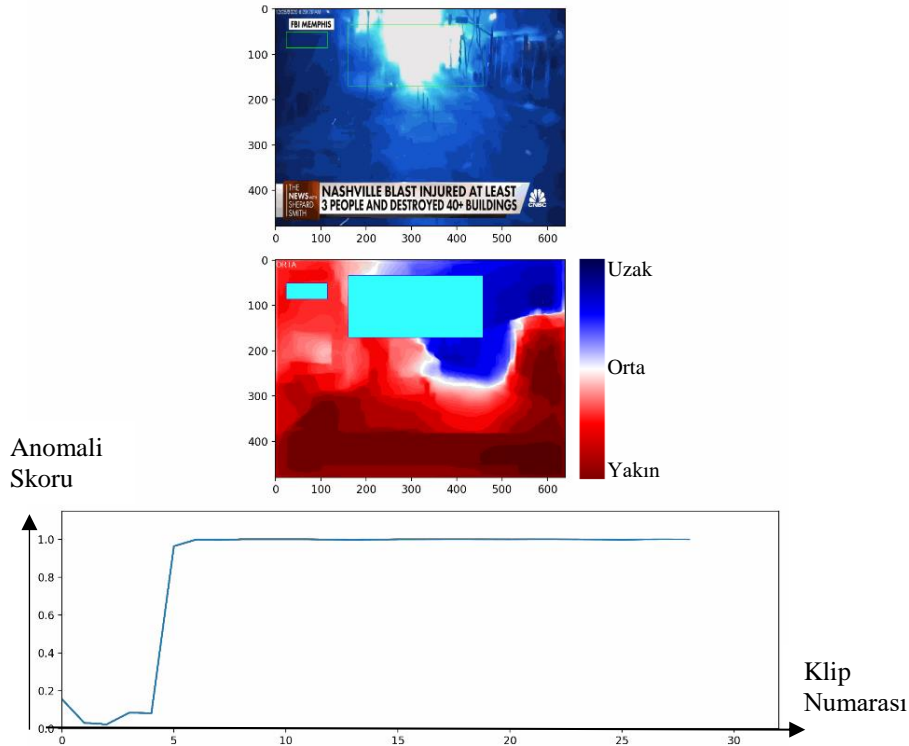
Şekil 4.19. Girdi resmi ve tahmin edilen derinlik haritası

Derinlik modelinin ürettiği derinlik haritasından elde edilen veriler doğrultusunda derinlik haritası üzerine 3 farklı derinlik sınıfı tahmini yazdırılmaktadır. Bu mesafe sınıflandırma işlemi 3 farklı sınıf için oluşturulmuştur. Bu sınıflar YAKIN, ORTA, UZAK olacak şekilde belirlenmiştir. Bazı derinlik sınıfı tahminleri Şekil 4.20’de gösterilmiştir. Sistemin mesafe tahmin aralığı, sistem tarafından üretilen derinlik haritaları incelenerek belirlenmiştir. Bu sınıflandırma işlemi için yapay sinir ağı kullanılacak olursa performansta artış gözlemlenebilir. Şekil 4.19 ve Şekil 4.20’de gösterilen derinlik haritalarından görüleceği üzere Şekil 4.19’da kırmızı yakın sınıfını temsil ederken, Şekil 4.20’de mavi yakın sınıfını temsil etmektedir. Bu değişimin nedeni

çıkı olarak üretilen GIF dosyasının oluşturulmasında kullanılan kütüphanedir. Kullanılan GIF kütüphanesi dosyayı GIF formatında kayıt ederken tersini olarak kayıt etmektedir. Bu durumun mesafe sınıflandırması üzerinde herhangi bir etkisi bulunmamaktadır. Derinlik tahmini modelinin performansının artırılması için sadece güvenlik kameralarından elde edilen görüntülerden oluşan bir derinlik haritası veri seti oluşturulabilir.



Şekil 4.20. Orta ve yakın mesafe tahminleri



Şekil 4.21. Anormal durum gerçekleştiği anda tespiti, konumlandırılması ve derinlik haritasında gösterimi

Model çalıştırıldığında Şekil 4.21'de gösterilen çıktıyı üretmektedir. Bu çıktı bilgisayarda gif formatında kayıt edilmektedir. Derinlik, konumlandırma ve anormal durum tespit değeri gif üzerinde anlık olarak gösterilmektedir. Şekil 4.21'de gösterilen görüntü Ucf-Crime veri içeriğinden farklı olarak web ortamından rastgele elde edilmiş bir görüntüdür, bu görüntüyü kullanmaktaki amaç test veriseti dışında bulunan görüntüler

ile modeli sınınamaktır. Şekil 4.21’de de görüleceği üzere tam patlama olduğu anda sistem patlamayı tespit etmiş, konumlandırmış ve mesafesini ORTA olarak tahmin etmiştir. Sistemde kullanılan modellerin işlem yaparken harcadığı zaman yazılımsal olarak ölçülmüştür. Örnek olarak 43 saniyelik bir video görüntüsü sisteme uygulanmıştır. Materyal ve metot başlığında bahsedilen sistem için anormal durum tespiti 8,86 saniye, lokalizasyon tahmini 8,03 saniye ve derinlik tahmini 21.34 saniye sürmüştür. Yazılımın 43 saniyelik bir video için harcadığı toplam süre 38.2 saniyedir. Görüleceği üzere en fazla zaman alan süreç derinlik tahminidir. Güvenlik kamerası gibi sabit uygulamalarda derinlik tahmin sıklığı düşürülerek modelin hızı arttırılabilir ve bu sayede daha hızlı tahmin işlemleri gerçekleştirilebilir.

5. SONUÇLAR

Anormal durum tahmini için C3D modeli ve tam bağlantılı yapay sinir ağı kullanılmıştır. Tam bağlantılı sinir ağının eğitimi için 100GB boyuta sahip Ucf-Crime veri seti kullanılmıştır. Veri setinin ağa uygulanması için videolardan özellikler C3D modeli ile elde edilerek dosya ortamında kayıt altına alınmıştır. Bu sayede eğitim aşamasında RAM bellekte tutulacak olan veri boyutu kontrol altına alınmıştır ve bu sayede veriler dosya formatında tutularak bellek yetersizliği gibi problemler ile karşılaşmamıştır. Anormal durum tahmini için eğitilen model de genel olarak başarı performansı iyi olsa da ani frame değişimleri, objenin kameraya çok yaklaşması, ani ışık parlamaları gibi beklenmedik durumlar anormal durum olarak tespit edilebilmektedir. Anormal durum tespiti için daha iyi donanımlar ile eğitim iterasyon sayısı artırılarak performansın artırılması sağlanabilir. Anormal durum tespitinde düşük çözünürlüklü bulanık görüntülerde yanlış anormal durum tespiti sayısı yüksek çözünürlüklü resimlere göre daha fazladır. Anormal durumun konumlandırılması için oluşturulan veri setinin performansa büyük etkisi olduğu gözlemlenmiştir. Veri setinin çeşitlendirilmesi durumunda konumlandırma performansının artacağı öngörülmektedir. Hatalar olmasına rağmen çoğu anormal durum konumlandırılabilmiştir. Derinlik tahmin algoritması da açık alan da derinlik algısını tam oluşturamasa da yakın uzak ilişkisi hakkında oldukça bilgi vericidir. Derinlik tahmin modelinin performansının artırılması için de yeni bir veri seti ile eğitim gerçekleştirilebilir. Konumlandırma tahmini için farklı filtre değerleri, batch boyutları denenmiştir ve en iyi sonuç filtre sayısı artırıldığında elde edilmiştir. Anormal durum tespiti için ise Adadelta ve Adagrad optimizasyon algoritmaları denenmiştir ve en iyi sonuç Adagrad algoritması ile elde edilmiştir. Anormal durumu tespit eden, anormal durumu konumlandıran ve bu konuma göre mesafe kestirimini yapan bir sistem bu tez çalışması kapsamında gerçekleştirilmiştir. Bu üç işlemi de bir arada yapması bu çalışmayı diğer çalışmalardan ayıran en önemli özelliğidir.

Akaryakıt istasyonları, duraklar, meydanlar gibi anormal durumların gerçekleşmesinin yüksek ihtimal olduğu ve sürekli gözlem altında tutulması gereken alanlarda kullanılabilecek bir sistem geliştirilmiştir.

6. KAYNAKLAR

- A. Saxena, M. Sun and A. Y. Ng, 2009. Make3D: Learning 3D Scene Structure from a Single Still Image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 824-840
- Alhashim, I., & Wonka, P. 2018. High Quality Monocular Depth Estimation via Transfer Learning. *ArXiv, abs/1812.11941*.
- Andraghetti, L., Myriokefalitakis, P., Dovesi, P.L., Luque, B., Poggi, M., Pieropan, A., & Mattocchia, S. 2019. Enhancing Self-Supervised Monocular Depth Estimation with Traditional Visual Odometry. *2019 International Conference on 3D Vision (3DV)*, 424-433.
- Boominathan, L., Kruthiventi, S., & Babu, R.V. 2016. CrowdNet: A Deep Convolutional Network for Dense Crowd Counting. *Proceedings of the 24th ACM international conference on Multimedia*, 640– 644
- Borghi, G., Venturelli, M., Vezzani, R., & Cucchiara, R. 2017. POSEidon: Face-from-Depth for Driver Pose Estimation. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5494-5503.
- Choudhary, S., Ojha, N., & Singh, V. 2017. Real-time crowd behavior detection using SIFT feature extraction technique in video sequences. *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, 936-940.
- Chu, C., Hwang, J., Pai, H., & Lan, K. 2013. Tracking Human Under Occlusion Based on Adaptive Multiple Kernels With Projected Gradients. *IEEE Transactions on Multimedia*, 15, 1602-1615.
- Chen, K., Gong, S., Xiang, T., & Loy, C.C. 2013. Cumulative Attribute Space for Age and Crowd Density Estimation. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2467-2474.
- Chen, W., Fu, Z., Yang, D., & Deng, J. 2016. Single-Image Depth Perception in the Wild. *ArXiv, abs/1604.03901*.
- Chen, Y., Zhao, H., & Hu, Z. 2021. Attention-based Context Aggregation Network for Monocular Depth Estimation. *Int. J. Mach. Learn. Cybern.*, 12, 1583-1596.
- Dalal, N., & Triggs, B. 2005. Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1, 886-893 vol. 1.
- D. G. Lowe, 1999 "Object recognition from local scale-invariant features," *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1150-1157
- Dos Santos Rosa, N., Guizilini, V., & Grassi, V. 2019. Sparse-to-Continuous: Enhancing Monocular Depth Estimation using Occupancy Maps. *2019 19th International Conference on Advanced Robotics (ICAR)*, 793-800.
- Fan Jiang, Ying Wu & A. K. Katsaggelos, 2019 Detecting contextual anomalies of crowd motion in surveillance video. *2009 16th IEEE International Conference on Image Processing (ICIP)*, 1117-1120.
- Fu, H., Gong, M., Wang, C., Batmanghelich, K., & Tao, D. 2018. Deep Ordinal

- Regression Network for Monocular Depth Estimation. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2002-2011.
- Garcia-Bunster, G., Torres-Torriti, M., & Oberli, C. 2012. Crowded pedestrian counting at bus stops from perspective transformations of foreground areas. *Iet Computer Vision*, 6, 296-305.
- Garg, R., Kumar, B.V., Carneiro, G., & Reid, I. 2016. Unsupervised CNN for Single View Depth Estimation: Geometry to the Rescue. *ECCV*.
- Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? The KITTI vision benchmark suite. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 3354-3361.
- Girshick, R.B., Donahue, J., Darrell, T., & Malik, J. 2014. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 580-587.
- Godard, C.; Mac Aodha, O.; Firman, M.; Brostow, G.J. Digging into self-supervised monocular depth estimation. 2019. *In Proceedings of the 2019 IEEE International Conference on Computer Vision, Seoul, Korea*, 3828–3838.
- Goldman, M., Hassner, T., & Avidan, S. 2019. Learn Stereo, Infer Mono: Siamese Networks for Self-Supervised, Monocular, Depth Estimation. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2886-2895.
- Guizilini, V., Ambrus, R., Pillai, S., & Gaidon, A. 2020. 3D Packing for Self-Supervised Monocular Depth Estimation. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2482-2491.
- Gülcü, A., Kuş, Z. 2019 Konvolüsyonel Sinir Ağlarında Hiper-Parametre Optimizasyonu Yöntemlerinin İncelenmesi. *Gazi University Journal of Science Part C: Design and Technology* 7, 503-522
- Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. ArXiv, abs/1704.04861.
- Huang, G., Liu, Z., & Weinberger, K.Q. 2017. Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261-2269.
- Hu, X., Hu, S., Huang, Y., Zhang, H., & Wu, H. 2016. Video anomaly detection using deep incremental slow feature analysis network. *IET Comput. Vis.*, 10, 258-265.
- Hubel, D., & Wiesel, T. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160.
- Ionescu, R.T., Khan, F., Georgescu, M., & Shao, L. 2019. Object-Centric Auto-Encoders and Dummy Anomalies for Abnormal Event Detection in Video. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7834-7843.
- Shao, J., Kang, K., Loy, C.C., & Wang, X. 2015. Deeply learned attributes for crowded scene understanding. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4657-4666.

- Ji, R., Li, K., Wang, Y., Sun, X., Guo, F., Guo, X., Wu, Y., Huang, F., & Luo, J. 2020. Semi-Supervised Adversarial Monocular Depth Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42, 2410-2422.
- K. Ping, P. Bo, Z. Wenying and L. Shuai, 2013. Research on central issues of crowd density estimation. *10th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, 143-145
- K. Xu, T. Sun and X. Jiang, 2020. Video Anomaly Detection and Localization Based on an Adaptive Intra-Frame Classification Network. *IEEE Transactions on Multimedia*, 394-406
- Kang, D., Dhar, D., & Chan, A.B. 2016. Crowd Counting by Adapting Convolutional Neural Networks with Side Information. *ArXiv, abs/1611.06748*.
- Karahan M., 2015 A Case Study on Forecasting of Tourism Demand With Artificial Neural Network Method, *Suleyman Demirel University The Journal of Faculty of Economics and Administrative Sciences*, 20(2), 195-209
- Khan, Akbar & Shah, Jawad & Kadir, Kushsairy & Albattah, Waleed & Khan, Faizullah. 2020. Crowd Monitoring and Localization Using Deep Convolutional Neural Network: A Review. *Applied Sciences*. 10 p.
- Kumagai, S., Hotta, K., & Kurita, T. 2017. Mixture of Counting CNNs: Adaptive Integration of CNNs Specialized to Specific Appearance for Crowd Counting. *ArXiv, abs/1703.09393*.
- Kuznetsov, Y., Stückler, J., & Leibe, B. 2017. Semi-Supervised Deep Learning for Monocular Depth Map Prediction. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2215-2223.
- L. Fiaschi, U. Koethe, R. Nair and F. A. Hamprecht, 2012. Learning to count with regression forest and structured labels. *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, Tsukuba, 2685-2688.
- L. Kratz and K. Nishino, 2009. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 1446-1453
- Lee, J.H., Han, M., Ko, D.W., & Suh, I. 2019. From Big to Small: Multi-Scale Local Planar Guidance for Monocular Depth Estimation. *ArXiv, abs/1907.10326*.
- Li, Y., Zhang, X., & Chen, D. 2018. CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1091-1100.
- Lu, C., Shi, J., & Jia, J. (2013). Abnormal Event Detection at 150 FPS in MATLAB. *2013 IEEE International Conference on Computer Vision*, 2720-2727.
- Mayer, N., Ilg, E., Häusser, P., Fischer, P., Cremers, D., Dosovitskiy, A., & Brox, T. 2016. A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4040-4048.
- Mehlig, B. (2019). Artificial Neural Networks. *ArXiv, abs/1901.05639*.
- Mehran, R., Oyama, A., & Shah, M. 2009. Abnormal crowd behavior detection using

- social force model. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 935-942.
- Narasimhan, Medhini & Kamath S, Sowmya. 2018. Dynamic video anomaly detection and localization using sparse denoising autoencoders. *Multimedia Tools and Applications*. 77.
- PETS, Vellore,. 2009. Performance Evaluation of Tracking and Surveillance (PETS) 2009 Benchmark Data. Multisensor Sequences Containing Different Crowd Activities.
- Pham, V., Kozakaya, T., Yamaguchi, O., & Okada, R. 2015. COUNT Forest: CO-Voting Uncertain Number of Targets Using Random Forest for Crowd Density Estimation. *2015 IEEE International Conference on Computer Vision (ICCV)*, 3253-3261.
- R. B. Fisher, 2004 The PETS04 Surveillance Ground-Truth Data Sets. *Proc. Sixth IEEE Int. Work. on Performance Evaluation of Tracking and Surveillance (PETS04)*, 1–5, CAVIAR Dataset. <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>
- Redmon, J., & Farhadi, A. 2017. YOLO9000: Better, Faster, Stronger. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6517-6525.
- Redmon, J., & Farhadi, A. 2018. YOLOv3: An Incremental Improvement. *ArXiv, abs/1804.02767*.
- Redmon, J., Divvala, S., Girshick, R.B., & Farhadi, A. 2016. You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788.
- Ren, S., He, K., Girshick, R.B., & Sun, J. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39, 1137-1149.
- S. Albawi, T. A. Mohammed and S. Al-Zawi, 2017 "Understanding of a convolutional neural network," *2017 International Conference on Engineering and Technology (ICET)*, 1-6
- S. V. C. Lab, UCSD anomaly data set. 2014. <http://www.svcl.ucsd.edu/projects/anomaly/dataset.html>
- Sabokrou, M., Fayyaz, M., Fathy, M., & Klette, R. 2017. Deep-Cascade: Cascading 3D Deep Neural Networks for Fast Anomaly Detection and Localization in Crowded Scenes. *IEEE Transactions on Image Processing*, 26, 1992-2004.
- Sabokrou, M., Fayyaz, M., Fathy, M., & Klette, R. 2018. Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes. *ArXiv, abs/1609.00866*.
- Scharstein, D., & Pal, C. 2007. Learning Conditional Random Fields for Stereo. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 1-8.
- Silberman, N.; Hoiem, D.; Kohli, P.; Fergus, R. 2012. Indoor segmentation and support inference from rgb-d images. *European Conference on Computer Vision*, Florence, Italy, 746–760.
- Sultani, W., Chen, C., & Shah, M. 2018. Real-World Anomaly Detection in Surveillance Videos. *2018 IEEE/CVF Conference on Computer Vision and Pattern*

- Recognition*, 6479-6488.
- Teed, Z., & Deng, J. 2020. DeepV2D: Video to Depth with Differentiable Structure from Motion. *ArXiv, abs/1812.04605*.
- Tosi, F., Aleotti, F., Poggi, M., & Mattocchia, S. 2019. Learning Monocular Depth Estimation Infusing Traditional Stereo Knowledge. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9791-9801.
- Tran, D., Bourdev, L.D., Fergus, R., Torresani, L., & Paluri, M. 2014. C3D: Generic Features for Video Analysis. *ArXiv, abs/1412.0767*.
- Tzotalin (2015). LabelImg Free Software: MIT License .
- UMN, Minneapolis, MN, USA. Unusual Crowd Activity Dataset of University of Minnesota, Department of Computer Science and Engineering, 2006.
- W. Li, X. Wu, K. Matsumoto and H. Zhao, 2010. A new approach of crowd density estimation. *TENCON 2010 - 2010 IEEE Region 10 Conference*, 200-203
- Wang, C., Zhang, H., Yang, L., Liu, S., & Cao, X. 2015. Deep People Counting in Extremely Dense Crowds. *Proceedings of the 23rd ACM international conference on Multimedia*. 1299-1302
- Wang, Q., Gao, J., Lin, W., & Li, X. 2021. NWPU-Crowd: A Large-Scale Benchmark for Crowd Counting and Localization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43, 2141-2149.
- X. Ding, F. He, Z. Lin, Y. Wang, H. Guo and Y. Huang, 2020. Crowd Density Estimation Using Fusion of Multi-Layer Features. *IEEE Transactions on Intelligent Transportation Systems*, 1-12
- Xie, J., Girshick, R.B., & Farhadi, A. 2016. Deep3D: Fully Automatic 2D-to-3D Video Conversion with Deep Convolutional Neural Networks. *ECCV*. 842–857.
- Viola, P.A., & Jones, M.J. 2001. Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 1*, I-I.
- Y. Zhang, D. Zhou, S. Chen, S. Gao and Y. Ma, 2016 Single-Image Crowd Counting via Multi-Column Convolutional Neural Network. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 589-597,
- Yaobin Mao, Junyan Tong and Wenbo Xiang, 2010. Estimation of crowd density using multi-local features and regression. *8th World Congress on Intelligent Control and Automation*. 6295-6300
- Miao, Y., Han, J., Gao, Y., & Zhang, B. 2019. ST-CNN: Spatial-Temporal Convolutional Neural Network for crowd counting in videos. *Pattern Recognit. Lett.*, 125, 113-118.
- Zhao, S., Fu, H., Gong, M., & Tao, D. 2019. Geometry-Aware Symmetric Domain Adaptation for Monocular Depth Estimation. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9780-9790.
- Zou, K., Warfield, S., Bharatha, A., Tempany, C., Kaus, M., Haker, S., Wells, W., Jolesz, F., & Kikinis, R. 2004. Statistical validation of image segmentation quality based

on a spatial overlap index. *Academic radiology*, 11 2, 178-89 .

ÖZGEÇMİŞ

NURİ ÖZÇELİK
nuriozcelik.07@gmail.com



ÖĞRENİM BİLGİLERİ

Yüksek Lisans 2018-2021	Akdeniz Üniversitesi Fen Bilimleri Enstitüsü, Elektrik-Elektronik Mühendisliği Anabilim Dalı, Antalya
Lisans 2012-2017	Gazi Üniversitesi Teknoloji Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Ankara

MESLEKİ VE İDARİ GÖREVLER

Donanım Tasarım ve Gömülü Yazılım Uzmanı 2018-Devam Ediyor	Cantek Group AR-GE Departmanı
--	----------------------------------