T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

# VİDEO GÖRÜNTÜLERİNDE KALABALIK ANALİZİ

Merve Ayyüce KIZRAK

DOKTORA TEZİ

Elektronik ve Haberleşme Mühendisliği Anabilim Dalı

Haberleşme Programı

Danışman

Doç. Dr. Bülent BOLAT

Haziran, 2021

**T.C.**

**YILDIZ TEKNİK ÜNİVERSİTESİ**

**FEN BİLİMLERİ ENSTİTÜSÜ**

# VİDEO GÖRÜNTÜLERİNDE KALABALIK ANALİZİ

Merve Ayyüce KIZRAK tarafından hazırlanan tez çalışması 30.06.2021 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Elektronik ve Haberleşme Mühendisliği Anabilim Dalı, Haberleşme Programı **DOKTORA TEZİ** olarak kabul edilmiştir.

Doç. Dr. Bülent BOLAT
Yıldız Teknik Üniversitesi
Danışman

**Jüri Üyeleri**

Doç. Dr. Bülent BOLAT, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. Lütfiye DURAK ATA, Üye
İstanbul Teknik Üniversitesi _____

Dr. Soner ÖZGÜNEL, Üye
Haliç Üniversitesi _____

Doç. Dr. Nihan KAHRAMAN, Üye
Yıldız Teknik Üniversitesi _____

Doç. Dr. Umut Engin AYTEN, Üye
Yıldız Teknik Üniversitesi _____

Danışmanım Doç. Dr. Bülent BOLAT sorumluluğunda tarafımca hazırlanan Video Görüntülerinde Kalabalık Analizi başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Merve Ayyüce KIZRAK

Bu çalışma, NVIDIA Grant Program tarafından Titan Xp GPU ile desteklenmiştir.

*Aileme*

# TEŞEKKÜR

Kasıtlı veya kasıtsız olarak doktora yolculuğumun bir kısmında bile yanımda olan herkese ve bana ilham kaynağı olan müzik ve ağaçlar gibi canlı cansız her şeye,

Meslek hayatımın önemli adımlarından biri olan doktora tezimi tamamlarken uzun, zorlu ve bir o kadar öğretici süreçte yolumu aydınlatan, bilgi birikimi ve tecrübeleriyle akıl hocalığı yapan saygıdeğer danışmanım Doç. Dr. Bülent Bolat'a,

Çalışmayı yürütebilmem için ihtiyacım olan donanım desteğini sağlayan NVIDIA şirketine,

Tez izleme komitemde yer alan; anlayışlı yaklaşımları, eleştiri ve yorumlarıyla zorlukların üstesinden gelmeme katkı sağlayan kıymetli hocalarım Prof. Dr. Lütfiye Durak Ata ve Dr. Soner Özgünel'e,

Doktoram süresince uzmanlık derslerimin bir kısmını aldığım ve bu alana yönelmem konusunda teşvikleri dolayısıyla sayın hocam Prof. Dr. Tülay Yıldırım'a

Hem doktora tezim hem de akademik kariyerim konusunda farkındalığımı ve motivasyonumu artıran, yorulduğumda beni cesaretlendirerek yola devam etmemi sağlayan sayın hocam Prof. Dr. Nafiz Arıca'ya,

Bu uzun süreçte görev yaptığım Haliç Üniversitesi, Bahçeşehir Üniversitesi ve T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi'nde tecrübelerinden faydalandığım meslektaşlarıma,

Gerek teknik konularda gerekse paha biçilemez manevi destekleriyle bana güç veren kıymetli arkadaşlarım Abdulmoniem Diri, Ayça Fulya Üstüntanır Dede, Başak Buluz Kömeçoğlu, Neslihan Çor, Yavuz Kömeçoğlu ve Zümrüt Müftüoğlu'na,

Hayatımın her anında sarsılmaz destekleri ve bana olan inançlarıyla yanımdan ayrılmayan çok sevgili anneme, babama ve kardeşlerime teşekkürlerimi sunarım.

Merve Ayyüce KIZRAK

# İÇİNDEKİLER


ŞEKİL LİSTESİ x

TABLO LİSTESİ xii

ÖZET xiii

ABSTRACT xv

1 GİRİŞ 1

1.1 Literatür Özeti ..... 2

1.1.1 Kalabalık Yoğunluk Kestirim, Davranış Tanıma ..... 5

1.1.2 Kalabalıktaki Alt Grupların veya Kişilerin Kestirimi ..... 8

1.1.3 Kalabalıkta Anormal Durumların Kestirimi ..... 9

1.1.4 Kalabalık Analizinde CNN Temelli Yaklaşımlar ..... 10

1.1.5 Kalabalık Analizinde Dikkat Temelli Yaklaşımlar ..... 13

1.2 Tezin Amacı ..... 14

1.3 Hipotez ..... 15

2 YAKLAŞIM VE YÖNTEMLER 16

2.1 Yaklaşım ..... 18

2.1.1 Yapay Sinir Ağları ..... 19

2.1.2 Yapay Öğrenme ..... 21

2.1.3 Derin Öğrenme ..... 23

2.2 Veri Kümeleri ..... 35

2.3 Yöntemler ..... 39

2.3.1 Yüksel Yoğunluklu Kalabalık için Derin ve Geniş CNN Modeli ..... 39

2.3.2 Kalabalık Sayımı için Küme Temelli İzleme ve Konum Tahmini ..... 43

2.3.3 Dikkat Temelli Kapsül Ağı ve İki Sütunlu CNN Kullanarak Kalabalık Yoğunluğu Tahmini ..... 47

2.4 Değerlendirme Metrikleri ..... 55
2.5 Deney Ortamı ..... 56
**3 SONUÇ VE ÖNERİLER** **58**
3.1 Sonuçlar ..... 58
3.1.1 Yüksel Yoğunluklu Kalabalık için Derin ve Geniş CNN Modeli Sonuçları ..... 58
3.1.2 Kalabalık Sayımı için Küme Temelli İzleme ve Konum Tahmini Sonuçları ..... 60
3.1.3 Dikkat Temelli Kapsül Ağı ve İki Sütunlu CNN Kullanarak Kalabalık Yoğunluğu Tahmini Sonuçları ..... 63
3.2 Tartışma ve Öneriler ..... 70
3.2.1 Gelecek Çalışmalar ..... 71
**KAYNAKÇA** **72**
**TEZDEN ÜRETİLMİŞ YAYINLAR** **84**

# KISALTMA LİSTESİ

| | |
|---|---|
| BM | Birleştirme Modülü |
| CapsNet | Kapsül Ağları |
| CNN | Convolutional Neural Networks (Evrişimli Sinir Ağları) |
| Covid-19 | CoronaVirus Disease 2019 |
| CRF | Conditional Random Field (Koşullu Rastgele Alanlar) |
| CUHK | The Chinese University of Hong Kong, Crowd Dataset |
| DWCNN | Deep and Wide Convolutional Neural Network |
| FFAM | Feature Fusion Attention Module |
| FFANet | Feature Fusion Attention Networks |
| GAN | Generative Adversarial Networks (Üretici Çekişmeli Ağlar) |
| GPU | Graphic Processing Unit (Grafik İşlem Birimi) |
| KLT | Kanade–Lucas–Tomasi |
| KNN | K-Nearest Neighborhood |
| MAE | Mean Absolute Error (Ortalama Mutlak Hata) |
| MALL | Mall Crowd Dataset |
| MCNN | Multi-Column CNN (İki Sütunlu Evrişimli Sinir Ağları) |
| MSE | Mean Square Error (Ortalama Karesel Hata) |
| PETS2009 | Crowd Image Analysis |
| POM | Probabilistic Occupancy Map (Olasılık Doluluk Haritalama) |
| RB | Resudial Block |
| ShanghaiTech | ShanghaiTech University, A large-scale crowd counting dataset |
| SSIM | Structural Similarity Index (Yapısal Benzerlik Endeksi) |
| SURF | Speeded Up Robust Features (Hızlandırılmış Güçlü Özellikleri) |
| UCF_CC_50 | University of Central California, Action Recognition Data Set |
| UCF-QNRF | University of Central California, A Large Crowd Counting Data Set |
| UCSD | University of San Diego, Anomaly Detection Dataset |
| WCS | World Coordinate System (Dünya Koordinat Sistemi) |

| | |
|---|---|
| YOLO | You Only Look Ones |
| YSA | Yapay Sinir Ağları |
| YZ | Yapay Zekâ |

# ŞEKİL LİSTESİ

**Şekil 1.1** Kalabalık analizi [1] ..... 1
**Şekil 1.2** Kalabalıkla ilgili araştırma yaklaşımlarının taksonomisi [16] ..... 4
**Şekil 2.1** Biyolojik bir sinir hücresi [134] ..... 20
**Şekil 2.2** Matematiksel bir sinir hücresi modeli [134] ..... 20
**Şekil 2.3** İleri ve geri yayılım gösterimi [137] ..... 23
**Şekil 2.4** İleri/geri yayılım ve ağırlıkların güncellenmesi gösterimi [137] ..... 23
**Şekil 2.5** Basitleştirilmiş derin YSA modeli ..... 24
**Şekil 2.6** Memeli canlıların görsel hiyerarşi özellikleri [140] ..... 26
**Şekil 2.7** CNN'lerde öğrenilen özniteliklerin görselleştirmesi [136]. ..... 28
**Şekil 2.8** Görselden çıkarılan öznitelikle ve hesaplanan tahmin [144] ..... 29
**Şekil 2.9** Evrişim katmanı terminolojisi [136] ..... 29
**Şekil 2.10** LeNet CNN modeli [142] ..... 31
**Şekil 2.11** CapsNet anlaşmalı yönlendirme stratejisi [146] ..... 34
**Şekil 2.12** Basitleştirilmiş CapsNet gösterimi [146] ..... 35
**Şekil 2.13** PETS2009 veri kümesi hiyerarşik yapısı [150] ..... 38
**Şekil 2.14** Paralel derin ağ modülü [15] ..... 41
**Şekil 2.15** Sığ ağ modülü [15] ..... 42
**Şekil 2.16** Derin ve geniş CNN Modeli [15] ..... 42
**Şekil 2.17** Problem ifadesi ve çözüm hiyerarşisi [16] ..... 44
**Şekil 2.18** Küme temelli kalabalık izleme yaklaşımı [16] ..... 44
**Şekil 2.19** Gauss YOLOv3 derin öğrenme modeli mimarisi [152] ..... 45
**Şekil 2.20** Önerilen algoritmanın tahmin kutusundaki öğeler [152] ..... 46
**Şekil 2.21** KNN algoritmasının temel gösterimi [16] ..... 46
**Şekil 2.22** Önerilen modelin temel modülü [157] ..... 49
**Şekil 2.23** Evrişim dikkat modülü [157] ..... 50
**Şekil 2.24** İki farklı ezme fonksiyonunun karşılaştırılması ..... 51
**Şekil 2.25** Kapsül dikkat modülü [157] ..... 52
**Şekil 2.26** Önerilen dikkat temelli kapsül ağı ve iki sütunlu CNN modelin genel mimari temsili [157] ..... 54
**Şekil 3.1** Shanghai Tech Part A veri kümesinden örnekler [82] ..... 59

**Şekil 3.2** Sekans 1 için yürüme/koşma davranışı için elde edilen örnek çıktı zamanı serisi [16]..... 61
**Şekil 3.3** Yürüme davranışından koşma davranışına değişen bir grup insan [16] .. 62
**Şekil 3.4** Soldan sağa doğru sütun sırasıyla UCF-QNRF, UCF_CC_50, UCSD, ShangaiTech Part A ve WorldExpo'10 veri kümelerinde referans değer, kestirilen kalabalık yoğunluk haritası ve kişi sayısı sonuçları ..... 64

# TABLO LİSTESİ

**Tablo 2.1** Kalabalık analizi için hazırlanan veri kümeleri ..... 36
**Tablo 2.2** PETS2009 veri kümesi detayları [150] ..... 38
**Tablo 2.3** PETS2009 veri kümesinde davranış analizi için hazırlanan video dizilerinin özeti [16] ..... 39
**Tablo 3.1** Seçilen parametreler ve optimizasyon yöntemleri [15] ..... 60
**Tablo 3.2** ShanghaiTech Part A veri kümesindeki DWCNN yönteminin başarım karşılaştırılması [15] ..... 60
**Tablo 3.3** Algılama sonuçları: yürüme / koşma davranışı [16] ..... 62
**Tablo 3.4** Dört farklı video sekansı için önerilen yöntemin doğruluk başarım karşılaştırması ..... 62
**Tablo 3.5** UCF-QNRF veri kümesinde karşılaştırma sonuçları ..... 65
**Tablo 3.6** UCF_CC_50 veri kümesinde karşılaştırma sonuçları ..... 65
**Tablo 3.7** UCSD veri kümesinde karşılaştırma sonuçları ..... 66
**Tablo 3.8** ShangaiTech Part A veri kümesinde karşılaştırma sonuçları ..... 66
**Tablo 3.9** ShangaiTech Part A veri kümesinde yoğunluk haritası kalitesi için karşılaştırma sonuçları ..... 67
**Tablo 3.10** WorldExpo'10 veri kümesinde karşılaştırma sonuçları ..... 67
**Tablo 3.11** Yoğunluk haritası üretimi için Gauss çekirdek ablasyon değerinin sonuçları ..... 68
**Tablo 3.12** UCF-QNRF, UCF_CC_50, UCSD, ShangaiTech Part A ve WorldExpo'10 veri kümelerindeki ablasyon modeli yapılarının sonuçları ..... 69

# ÖZET

## Video Görüntülerinde Kalabalık Analizi

Merve Ayyüce KIZRAK

Elektronik ve Haberleşme Mühendisliği Anabilim Dalı

Doktora Tezi

Danışman: Doç. Dr. Bülent BOLAT

Dünya; aşırı nüfus artışının yanı sıra terör, savaş ve salgın gibi sınavlardan geçerken, kamuya açık alanların güvenliğinin sağlanması giderek zorlaşan bir konu haline gelmektedir. Video kayıtlarından elde edilen görüntüler kullanılarak kamuya açık alanların denetlenmesi, düzenlenmesi gibi pratik konularda kalabalık yoğunluk analizinin en uygun şekilde gerçekleştirilmesi amaçlanmaktadır. Kalabalık analizi, akademik ve pratik hayatta yaygın etkiye sahip, önemli ve güncel bir araştırma konusudur.

Bu çalışma dâhilinde yapay zekâ alt konularından olan derin öğrenme yaklaşımları kullanılarak özgün ve etkin bir yöntem önerilmektedir. Çalışmanın ilk sonuçları CNN (Evrişimli Sinir Ağları) temelli paralel mimari, Gauss-YOLOv3 (You Only Look Once) ve iyi bilinen KNN (K-En Yakın Komşuluk) gibi yöntemlerin kombinasyonları ile alındı. Kalabalık analizi için literatürde yer alan UCF-QNRF, UCF_CC_50, UCSD, ShanghaiTech Part A, WorldExpo'10 ve PETS2009 veri kümeleri hakkında detaylı bilgi verildi. Çalışmanın başarısının genelleştirilebilir olduğunu göstermek için bu veri kümeleri ile testler yapıldı.

Çalışmanın kalabalık davranış değişiminin kestirilmesi aşaması için PET2009 veri kümesindeki altı farklı video dizisi kullanıldı. Doğruluk başarım %83,2 ile %96,4

aralığında elde edildi. Bu sonuçların literatürdeki benzerleriyle karşılaştırılabilir düzeyde olduğu gösterildi.

Çalışmanın nihayetinde, dikkat temelli evrişim ve kapsül ağı modülüne sahip ve anlaşmalı yönlendirme algoritması kullanan iki sütunlu bir derin öğrenme mimarisi önerildi. Önerilen yöntemin başarımını kapsamlı bir şekilde göstermek için MAE (Ortalama Mutlak Hata) ve MSE (Ortalama Karesel Hata) değerlendirme metrikleri güncel çalışmalarla karşılaştırıldı ve kendilerine en yakın başarıma göre gerçekleştirilen iyileştirilme oranları hesaplandı. İyileştirme ile önerilen yöntem literatürdeki son teknoloji yaklaşımlarla karşılaştırıldığında bu problem için etkili ve genelleştirilebilir olduğunu göstermektedir. Elde edilen sonuçlar şöyledir: UCF-QNRF'de %2,18 MSE; UCF_CC_50'de %2,33 MAE ve %1,68 MSE; UCSD'de %8,46 MAE ve %7,03 MSE; ShanghaiTech Part A'de %4,69 MAE ve %3,94 MSE; WorldExpo'10'de %6,94 MAE.

Önerilen kapsül ağı temelli derin öğrenme mimarisi; davranış analizi için gelecekteki çalışmalarda optik akış gibi hareket bilgileri kullanılmaksızın, konum ve yönelim bilgilerinin kalabalık analizlerinde kullanılabileceğini göstermektedir.

**Anahtar Kelimeler:** Kalabalık analizi, kişi sayma, evrişimli sinir ağları, kapsül ağları, dikkat mekanizması.

# ABSTRACT

## Crowd Analysis in Video Images

Merve Ayyüce KIZRAK

Department of Electronics and Communication Engineering

Doctor of Philosophy Thesis

Advisor: Assoc. Prof. Dr. Bülent BOLAT

World; In addition to excessive population growth, while passing through tests such as terrorism, war, and epidemics, ensuring the security of public spaces is becoming an increasingly difficult issue. By using images obtained from video imaging, it is aimed to perform crowd density analysis most appropriately in practical matters such as controlling and arranging public spaces. Crowd analysis is an important and recent research subject that has a widespread impact on academic and practical life.

In this study, an original and effective method is proposed by using deep learning approaches, which is one of the sub-topics of artificial intelligence. The first results of the study were obtained with a combination of CNN-based parallel architecture, Gauss-YOLOv3, and well-known KNN methods. Detailed information was given about the datasets of UCF-QNRF, UCF_CC_50, UCSD, ShanghaiTech Part A, WorldExpo'10, and PETS2009 in the literature for crowd analysis. Tests were performed with these datasets to demonstrate that the success of the study was generalizable.

Six different video sequences in the PET2009 dataset were used to predict crowd behavior change in the study. Accuracy performance was between 83.2% and

96.4%. These results were shown to be comparable to their counterparts in the literature.

At the end of the study, a two-column deep learning architecture with attention-based convolution and capsule network module and using negotiated routing algorithm is proposed. To show the performance of the proposed method comprehensively, MAE and MSE evaluation metrics were compared with current studies and the improvement rates according to their closest performance were calculated. The proposed method with improvement shows that it is effective and generalizable for this problem when compared with the state-of-the-art approaches in the literature. The results obtained are as follows: 2.18% MSE in UCF-QNRF; 2.33% MAE and 1.68% MSE at UCF_CC_50; 8.46% MAE and 7.03% MSE in UCSD; 4.69% MAE and 3.94% MSE in ShanghaiTech Part A; 6.94% MAE at WorldExpo'10.

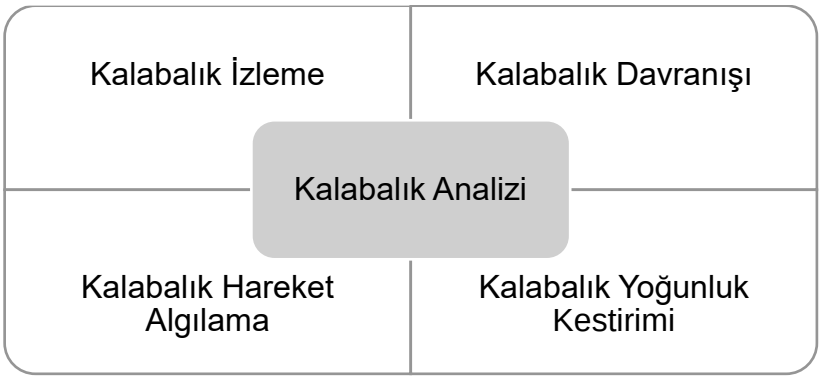
The proposed capsule network-based deep learning architecture; shows that position and orientation information can be used in crowd analysis without using motion information such as optical flow in future studies for behavioral analysis.

**Keywords:** Crowd analysis, people counting, convolutional neural networks, capsule networks, attention mechanism.

# 1
# GİRİŞ

Kalabalık, fiziksel olarak ortak bir konumu paylaşan bir grup olarak tanımlanmaktadır. Nüfus artışı, hızlı kentleşme insanları bir araya toplamakta ve oluşan kalabalıktan doğabilecek izdihamların önüne geçmek için planlama yapılmasını gerektirmektedir. Geçmiş yıllarda dünyanın pek çok yerindeki büyük şehirlerin meydanlarında yaşanan provakatif olayların önlenmesi ve denetlenmesi ön plandayken, 2020 yılı itibariyle dünyayı etkisi altına alan ve bir yılı aşkın süredir devam etmekte olan Covid-19 (CoronaVirus Disease 2019) pandemisi kalabalığa bakış açısını değiştirdi. Pandemi dolayısıyla kuyruklar, alışveriş alanları gibi noktalarda kişilerin sayılması ve temas durumlarının denetlenmesi dünya sağlığı açısından kritik bir hal aldı.

Kalabalık yönetimi, etkinlikler ve meydanlarda tıkanıklığı önlemek için akış analizi, kentsel planlama, kamu güvenliği yönetimi, afet yönetimi ve savunma gibi birçok konuyu kapsamaktadır. Kalabalık analizi ise kalabalık yönetimi için bir araçtır ve yoğunluk kestirimi, hareket algılama, davranış analizi ve izleme gibi alt konulara ayrılabilmektedir. Kalabalık analizi çalışmalarını Şekil 1.1'de gösterildiği gibi dört temel alt grupta incelemek mümkündür.




**Şekil 1.1** Kalabalık analizi [1]

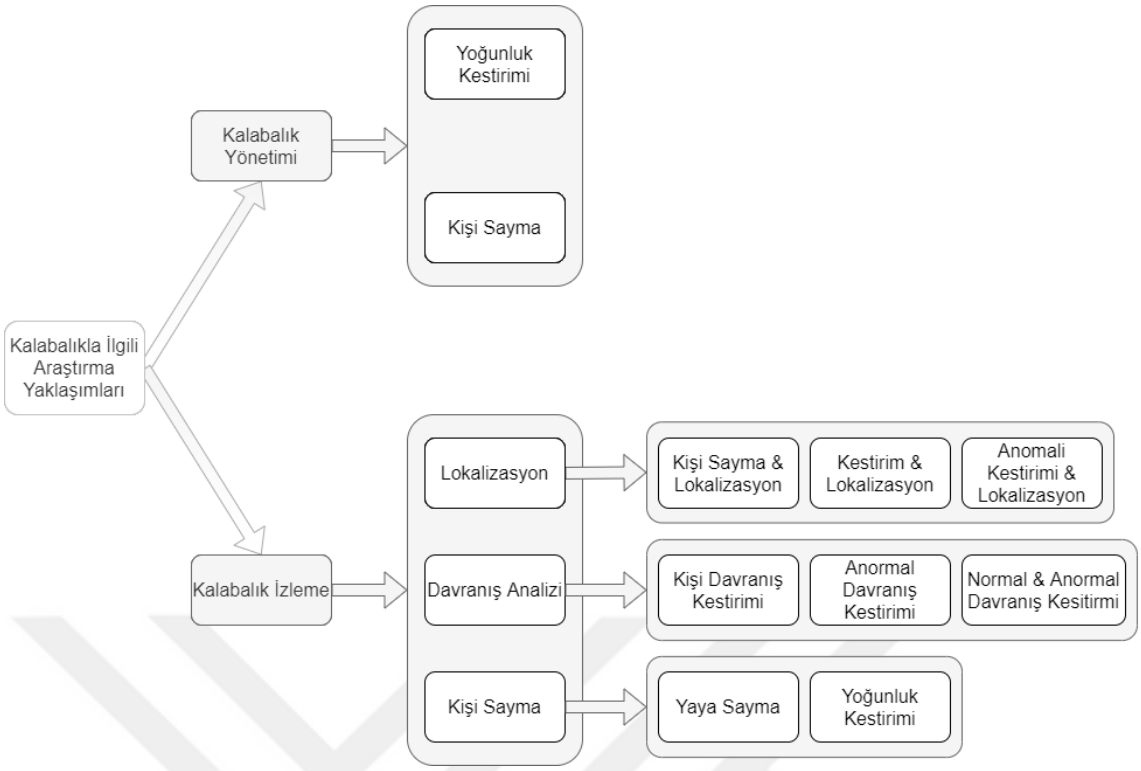
Kalabalık yoğunluk kestirimi ve davranış değişiminin algılanması bu doktora tezinin konusudur. Sağlam ve hızlı bir kalabalık analizi yaklaşımlarının bugün ve gelecekte yaygın olarak kullanılması planlanmaktadır.

## 1.1 Literatür Özeti

Kalabalık analizinde ele alınması gereken önemli konuların başında kalabalıktaki kişi sayısının belirlenmesi gelmektedir. Bunun için uzun yıllardır çalışmalar yürütülmektedir. 2006 yılı öncesine kadar geleneksel görüntü işleme yaklaşımlarıyla ve takip algoritmaları ile çözümler üretilmekteydi. Homografik, piksel ve renk temelli yaklaşımlar, genel nesne algılamasını sağlamak için her sensörden yerel görüntü özelliklerini ortak bir düzleme yansıtmak için kullanılmaktaydı. Bu tür yaklaşımlar özellikle ışık ve boyut gibi değişkenlere karşı kırılgandır [2]-[7]. Optik akış algoritması, Kalman filtresi gibi yaklaşımlar nispeten başarılı yaklaşımlar arasındadır. Ancak ışık değişimleri, insanların birbiri ile basitçe örtüşmüş olmaları veya kişilerin yanlarında büyük çantalar taşımaları gibi basit ve farklılık arzeden durumlar için bile kırılgan yapıları dolayısıyla pratikte çok uygulanamamaktadır. Bu tezin ilk araştırma evresinde temel bileşenler analizi, Kalman filtre ve optik akış algoritması ile deneyler yapılmış ancak başarılı sonuçlar elde edilememiştir. Kırılganlıklar tespit edildikten sonra bunun üstesinden gelebilmek için derin sinir ağları modellerine odaklanılmıştır. Derin bir sinir ağı, uçtan-uca çoklu görüntü bilgilerinden yararlanarak insanları algılamaya olanak sağlamaktadır. Ancak insan sayma probleminde önden, profilden ve örtüşen görüntüler yer almaktadır. Bu yüzden çalışmaların büyük bir kısmı bu problemleri çözmeye odaklanmaktadır. Bunun için ortalama alan değişimi çıkarımları, Olasılık Doluluk Haritalama (POM)'lar, olay yerinde bulunan insanların ikincil olasılık dağılımını tahmin etmeye odaklanan araştırmalar bulunmaktadır. Son çalışmalardan birinde çok görüntülü yaya algılama için Koşullu Rastgele Alanlar (CRF)'ler kullanılarak tartışılmaktadır [8]-[14]. Bilgisayarlı görü alanında alınan yol ve örüntü tanıma yeteneği dolayısıyla CNN temelli çalışmalar 2006 yılı sonrasında sıklıkla önerilmektedir.

Tez çalışmasının ikinci evresinde CNN temelli yaklaşımlar araştırılarak kalabalık yoğunluk kestirimi yapılması üzerinde durulmuştur. Yoğunluk haritası elde edilerek kişi sayısı tahmini son yıllarda pek çok araştırmacının ilgisini çeken bir konu haline gelmesi daha hızlı, başarılı ve özgün bir yaklaşım ihtiyacını doğurmaktadır. Bu nedenle özniteliklerin daha etkin çıkarılarak daha başarılı sonuçlar elde edilebilmesi için sıralı ve paralel bağlantılı çeşitli CNN modelleri tasarlanmaktadır [15].

Kalabalık düzenli bir gruptan meydana geliyorsa, belirli bir amaç için buluştukları anlamına gelebilir. Bazen kalabalık dağınık hareket edebilir, bazen de benzer düzenlere sahip hareketler benimseyebilir. Kalabalık izlemenin önemli amaçlarından biri de beklenmedik olayları tespit etmektir. Kalabalık izleme, insanların sıradan davranışlarından ziyade, çökme, yangın, patlama gibi ani durum değişikliklerinde güvenliği sağlamak için kritik bir konudur. Bunun için özellikle şehirlerin kalabalık noktalarında güvenlik kameralarından toplanan video verilerinin işlenmesi yoluyla gecikmeyi azaltarak daha güvenli ortamlar sağlamayı amaçlamaktadır. Kamera sistemlerinin kullanımı yaygınlaştıkça kalabalık izleme yöntemleri de gelişmektedir. Kalabalıkla ilgili araştırmaları yönteme ve sorunun ele alınış şekline göre sınıflandırmak mümkündür. Şekil 2.2, kalabalığa ilişkin çalışmalar üzerine literatürdeki araştırmaların bir taksonomisini göstermektedir [16].




**Şekil 1.2** Kalabalıkla ilgili araştırma yaklaşımlarının taksonomisi [16]

Kalabalıkla ilgili araştırmalar görüntüleri ve videoları kullanarak çeşitli sahnelerdeki insan sayısını, dağılımını ve davranışlarını incelemektedir. Kalabalık analizi ile insanların sayıları ve mekânsal dağılımlarının kestirilmesi, temel olarak algılama, izleme ve sayma olarak gruplandırılmaktadır. Kalabalık yoğunluk kestirimi, kalabalık farklı gruplardan oluşuyorsa bunun tespit edilmesi ve davranış değişim anlarının kestirimi için kişi sayısı, grup sayısı ve grup davranış değişim yönü ve zamanının tahmin edilebilmesine ihtiyaç duymaktadır. Ancak CNN özellikle kişilerin örtüşmesi, konum ve yönleri konusunda istenen başarılı sonuçları verememektedir. Literatürde bu konunun bir doyuma ulaştığı görülmektedir. Bunun için bu çalışmada literatürde daha önce bu konu üzerinde denenmemiş güncel ve teorik bir yaklaşım olan kapsül ağlarının elde edebildiği özniteliklerden yararlanılmıştır. Önerilen teorik yaklaşım deneysel sonuçlarla da pratik olarak desteklenmektedir.

Bu bölümde; video görüntülerinde kalabalık analizi için ihtiyaç duyulan temel tanım ve literatür özetine yer verilmektedir. Bu sayede araştırmacılara konuyla ilgili temel bilgiyi sağlaması hedeflenmektedir.

### 1.1.1 Kalabalık Yoğunluk Kestirim, Davranış Tanıma

2008'de Ali ve Shah tarafından yapılan çalışmada yoğunluğu yüksek kalabalık görüntülerinde duvar ve bariyer gibi statik ve insanların hareketliliğinin olduğu dinamik alanların ve yayaların takibi için ortalama kaydırma yönteminden faydalanılmıştır [17].

[18] ve [19]'de Mao vd. video görüntülerinin her bir çerçevesi arasındaki fark bilgisini kullanarak hareketli bölgeleri belirlenmiştir. Sekiz tane alt çerçeve kullanarak öznitelik çıkarılmıştır. Bu çalışmada köşe belirlemek için Harris filtre kullanılmıştır. Kabalık yoğunluğu için KLT (Kanade–Lucas–Tomasi) öznitelik haritası kullanılarak 3 farklı veri kümesinde başarımları karşılaştırılmıştır.

Guo vd. [20], 2010 yılında yayınlanan ve rastgele Markov süreç yöntemlerini kullanarak kamu güvenliğini sağlamak amacıyla kalabalık analizi çalışmasıdır. Geleneksel kenar bulma ve ön plan ile arka planın ayrıştırılması için optik akış yaklaşımından faydalanılmıştır. Önerilen model gerçek zamanlıdır.

Li vd. [21] ve [22] makalelerinde optik akış yöntemi kullanarak hareketli ve hareketsiz bölgeleri belirlemekle işe başlanmıştır. Ön plan görüntüsünde doku analizi için öznitelikleri kullanarak kendi kendine organize olabilen haritalama yöntemi ile kalabalık içindeki farklı alt gruplar belirlenmiştir.

Ge ve Collins [23]'de kalabalık kişi sayısı tahmini üzerine yoğunlaşmışlardır. Görüntü çerçeveleri arasındaki ilişkiden yola çıkarak görüntünün yoğunluk haritası çıkarılmıştır. SURF (Hızlandırılmış Güçlü Özellikleri) noktaları belirlenmiş ve kümeleme için K-ortalamalar, Haar dalgacık dönüşümü ve Gauss yöntemleri birlikte kullanılmıştır.

Kim vd. [24]'de 5 farklı seviyede yoğunluğa sahip görüntülerinde optik akış yöntemi ve kenar piksel yaklaşımı ile kalabalık seviyeleri sınıflandırılmıştır.

Hsu vd. 2011 yılında yayınlanan [25]'da frekans dağılımı ve kosinüs dönüşümü yaklaşımları ile farklı frekans aralıklarındaki öznitelikler çıkarılmıştır. Bu öznitelikleri destek vektör makineleri yöntemi kullanarak 5 farklı seviyedeki kalabalık yoğunluğu sınıflandırmak için kullanılmıştır.

Xiong vd. [26]'de gruplaşma koşma sınıflarına sahip grupların davranışlarını tespit etmek için görüntünün enerji bilgileri kullanılmıştır. Uyarlamalı Gauss bileşim modeli kullanılarak histogram üzerinden potansiyel enerji ve optik akış hesaplanmıştır. Kinetik enerji davranış değişimini göstermekte ve bunun içinden optik akıştan elde edilen bilgi kullanılmıştır.

Yu vd. [27] yüksek yoğunluktaki kalabalık akış analizi için dinamik doku, optik akış ve ilgi bölgesi özellikleri ile Gauss regresyon sinir ağları kullanılmıştır.

[28]'de Yang ve Zhoa, doku analizi için öznitelikleri Gauss ve Lucas-Kanade optik akış yöntemi kullanarak elde etmişlerdir. Farklı kalabalık gruplarını sınıflandırmak için kendi kendine organize olabilen haritalama yöntemi kullanılmıştır.

Subburaman vd. 2012 yılında yayınlanan [29] makalelerinde kalabalıktaki kişilerin kafa bölgelerini ilgi noktası olarak işaretledikleri çalışmada gradyan yönelimler kullanılarak ikili yerel örtüler hesaplanmıştır. Bu bilgileri kullanarak kalabalıktaki kişi sayısı tahmin edilmiştir.

Chan ve Vasconcelos [30], [31] ve [32] çalışmalarında Bayes, Gauss ve Poisson regresyon yaklaşımları kullanılarak görüntüdeki dinamik bölgelere bağlı olarak kişi yoğunluk tahmini gerçekleştirmişlerdir.

Wang vd. [33]'de görüntüdeki yerel bölgelerde değişen yoğunlukların kestirilmesi üzerine bir çalışmadır. Yerel ikili örüntü ve gri seviye eş-oluşum matrisi beraber kullanılarak öznitelikler elde edilmiştir. Uzamsal ve istatistiksel bilgilerden faydalanarak doku özellikleri pencere kaydırma yöntemiyle elde edilmiştir. Öznitelikleri sınıflandırarak olası kalabalık bölgelerin kestirimi için ise destek vektör makineleri yönteminden faydalanılmıştır.

Fradi ve Dugelay'ın 2012 ve 2013 yıllarında yayınlanan [34] ve [35] makalelerde yoğun kalabalık ve perspektif etkinin çok olduğu zorlu kalabalık görüntülerinde analizler gerçekleştirilmiştir. Hareketli bölgelerin belirlenmesi için çerçeve farkları yöntemi kullanılmıştır. Perspektif ve yoğunluk normalizasyonu ardından Gauss regresyon modeli ile kişi sayısı tahmini gerçekleştirilmiştir.

Tehranipour vd. [36]'de bulanık mantık karar modelleri kullanılarak kalabalık yoğunluk kestirimi yapılmıştır. Kalabalık içinde yoğunluğun arttığı bölgeler riski

yüksek sınıflardır. Bu bölgelerin tespiti için dalgacık dönüşümü, modifiye edilmiş piksel temelli yöntemler uygulanmıştır. Sınıflandırma aşamasında destek vektör makineleri kullanılmıştır.

Fradi vd. [37]'de kalabalık analiz yöntemi için ikili yerel örüntüleri ile alt uzay öğrenme yöntemleri birlikte kullanılmıştır. Lineer diskriminant analizi ve temel bileşenler analizi yöntemleri ile görüntülerin alt uzayları hesaplanmıştır. Bu alt uzay bilgileri K-en yakın komşuluk ve destek vektör makineleri yöntemlerinde öznitelik olarak kullanılmıştır.

Rao vd. [38]'da görüntü bloklara ayrılarak her birine ayrı ayrı optik akış uygulanmıştır. Kümeleme hiyerarşik olarak gerçekleştirilmiştir. Bu adımdan sonra blokların ve orijinal görüntülerin arasındaki ilişki Öklid uzaklığı ile incelenmiştir. Böylece bir yoğunluk haritası elde edilmiş ve buradan tahmini kişi sayısı kestirimi gerçekleştirilmiştir.

Yuan vd. [39] tarafından yapılan çalışmada mevcut çalışmalardan farklı bir yaklaşım olarak internet ağı ve radyo frekans bilgilerini kullanarak sinyal yoğunluğuna bakarak kalabalık kişi sayısı kestirimi gerçekleştirilmiştir. Bu yaklaşım mevcut çalışmalarla karşılaştırılabilecek derecede başarılıdır.

Karpagavalli ve Ramprasad [40] makalelerinde kalabalığın genel yoğunluğu ilgilenmeyen ve doğrudan kişi sayısı kestirimi gerçekleştirmeye odaklanmışlardır. Bunun için uyarlamalı k-Gauss modeli ile görüntülerdeki ön ve arka plan ayrıştırması gerçekleştirilmiştir. Bu sayede sınır bölgeler çıkarılarak sonuca ulaşılmıştır.

Wu vd. [41]'de Lucas-Kanade optik akış yöntemi ile kalabalıktaki dinamik bölgeler belirlenmiştir. Elde edilen öznitelikler destek vektör makineleri yöntemi ile sınıflandırılmış ve kalabalık takibi için başarımlar hesaplanmıştır.

Ping vd. [42]'de 8 blok olarak ayrılmış görüntüleri kullanarak istatistiksel çıkarımlar yapılmıştır. Gri seviye matris bileşimi ile en küçük kareler regresyonu yöntemlerini kullanarak doku analizi yapılmıştır. Kalabalık yoğunluk kestirimi için destek vektör makineleri yöntemi kullanılmıştır.

Yuan 2014 yılında yayınlanan [43] makale, yoğun kalabalık bölgelerdeki gruplaşmayı algılamak için kablosuz ağ sinyal paylaşımından faydalanıldığı bir başka çalışmadır. Bu sinyaller kullanılarak lokalizasyon yapılmış ve kalabalık içinde farklı gruplar haritalandırılmıştır.

### 1.1.2 Kalabalıktaki Alt Grupların veya Kişilerin Kestirimi

Kalabalık video görüntülerinden kişilerin ya da grupların belirlenmesi ve bunların takip edilmesi literatürde ayrıca detaylı olarak ele alınan bir konudur. Belirlenen kişi ve grupların takip edilebilmesi için çerçeveler arasındaki ilişkiden faydalanılmaktadır. Hareketliliği yüksek kalabalık gruplarında problem daha da karmaşık bir hal almaktadır. Khansari vd. [44] makalelerinde video görüntüleri kullanılarak grup takibi için dalgacık dönüşümü ve doku analizi kullanılmıştır. Kişi ve grupların hız ve hareket yönlerini bu özniteliklerden yola çıkarak hesaplanmıştır. 2011 yılından Rodriguez vd. [45] ve [46] çalışmalarında görüntülerdeki kişilerin kafa tespitine odaklanılmıştır. Enerji yoğunluğu bilgilerinden yola çıkarak takip işlemi gerçekleştirilmiştir.

Literatürde kalabalık görüntülerindeki alt grupların takibi için uyarlamalı ve Kalman filtre modelleri belli bir başarıya ulaştığı görülmüştür. Kalabalık izleme, algılama ve regresyon temelli yöntemler alt grupların belirlenmesi ve takip edilmesinde kullanılmıştır. Algılama temelli yaklaşımlar, kafa, yüz ve omuz bölgeleri sahnelerden algılamaya odaklanmaktadır. Subburaman vd. [39] makalelerinde kafa tanıma ve kişilerin sayılması için algılama yöntemi olarak renk bilgisi kullanılarak regresyon temelli bir model ile doğrusal ve doğrusal olmayan fonksiyonu öğrenilmiştir. Böylece öğrenilmiş olan model kullanılarak insan sayısı kestirimi gerçekleştirilmiştir.

Conte vd. [47] tarafından regresyon temelli bir yöntem önerilmiştir. İzleme temelli yöntem ile benzer uzantılarda ve kalıplarda kümelemeye dayalı bir yaklaşımdır. Farklı kamera özellikleri ve farklı zorluklara sahip veri kümelerinde deneysel sonuçları incelenmiştir.

[48]'da Antonini vd. video dizilerinde kişi sayan bir yaklaşım önerilmiştir. Birbirine benzer yörüngeler kombine edilmiş ve ayaları saymak için kullanılmıştır.

Kişilerin sayılması ve takip edilmesi yalnızca kamu güvenliği için değil sosyal alanların planlanması, pazarlama stratejilerinin belirlenmesi, afet ve olağanüstü hal yönetimi gibi pek çok uygulama alanı bulması dolayısıyla araştırmacıların ilgi odağında olan çalışma konuları arasındadır. Güvenlik ve gözetim talepleri Covid-19 pandemisiyle daha da artış göstermişken çeşitli alt gruplara ayrılarak incelenen bir araştırma geliştirme alanıdır. Kalabalık analizi, kalabalık yoğunluk tahmini, kişi sayma, kalabalık davranış analizi, kalabalık izleme, kalabalık bölütleme, vb. alt konulardan oluşmaktadır [49]-[65].

Kalabalık analizi genellenebilirliği zayıf olması dolayısıyla çok zorlu bir konudur. Bu yüzden problem özgü çözümler geliştirilmesi her ne kadar başarılı sonuçlar vermiş olsa da günümüz daha genelleştirilebilir ve kırılganlıkları azaltılmış modellere ihtiyaç duyulmaktadır. Son çalışmalar, yüzlerin, nesnelerin, eylemlerin ve sahne görüntülerinin nitelik temelli gösterimlerini içermektedir [66]-[75].

### 1.1.3 Kalabalıkta Anormal Durumların Kestirimi

Kalabalık analizinin bir diğer alt araştırma alanı da anormalliklerin kestirilmedir. [75]'da Jiang vd. tarafından uzamsal değişimleri eğiticisiz bir model önerilmiş ve doku analiziyle de birleştirerek anormallikleri kestirilmiştir.

Kratz ve Nishino [76]'de zaman-uzamsal hareket örüntüleri ile mekânsal değişimleri saklı Markov modelleri kullanarak hesaplamışlardır. Çalışmalarında hızlı değişen davranışları kestirmek için başarılı sonuçlar sunulmuştur.

Mahadevan vd. 2010 yılında yayınladıkları [77] makalelerinde dokuların birleşimi modelini önerdiler. Gauss bileşim modelinden faydalanarak dinamik doku bölgelerini buldular. Ardından uzamsal analizi ile normal değişimleri kestirdiler.

2011 ylında Reddy vd. [78] çalışmalarında ön plan görüntüsündeki hızlı değişimleri, değişim büyüklüğünü ve doku bilgisini hücre temelli bir analizle anormallerin kestiriminde başarılı sonuçlar elde etmişlerdir.

Saligrama ve Chen 2012'de yayınladıkları [79] makalede görüntü çerçeveleri için yerel öznitelikleri Neyman-Pearson karakterizasyonu ile hesaplamışlardır. Uzamsal değişimleri hesapladıkları bir filtre sonucuna k-en yakın komşuluk algoritması uygulanmıştır. Farklı 5 veri kümesinde elde ettikleri sonuçlar karşılaştırılmıştır.

2013 yılında Cau vd. [80] çalışmalarında kalabalığın genel davranışı içinde farklı kalıplarda hareket eden kişi ve grupları kestirmek üzerine çalışmışlardır. Bunun için Gauss bileşim modeli kullanılmıştır. Ancak bu çalışmayı diğerlerinden ayıran özelliği birikmiş mozaik farkları ve perspektif bozukluklarını düzenleme yaklaşımlarıyla hızlı hareket değişimlerini yakalamayı başarmışlardır.

Kalabalık analizi altında incelenen bu çalışmalar farklı problemlere farklı istatistiksel, yerel görüntü işleme ve makine öğrenmesi yaklaşımları ile problemi ele almaktadır. Ancak başarım, hız ve genelleştirilebilirlik açısından yetersizdir. 2006 yılından sonra derin öğrenme tekniklerinin akademik dünyada tekrar uygulanmaya başlanması ve nesne algılamaya dair paylaşılmaya başlayan büyük veri kümeleri sayesinde kalabalık analizinde derin öğrenme yaklaşımlarının kullanılmaya başlanmıştır. Günümüzde en etkin sonuçlar bu yöntemlerle elde edilmektedir.

### 1.1.4 Kalabalık Analizinde CNN Temelli Yaklaşımlar

Görüntü özellikleri ve kalabalık boyutu arasındaki regresyona bağlı olarak, kalabalıktaki insan sayısını tahmin etmek için regresyon temelli yaklaşımlar kullanılabilir. Regresyon yaklaşımı, CNN temelli yaklaşımlarla görüntüden yoğunluk haritalarını çıkarmak için kullanılmaktadır. 2015 yılında Zhang vd. [81] tarafından tek bir görüntüdeki insan sayısını tahmin etmek için CNN regresyon modelini kullanılmıştır. Zhang vd. 2016 yılında yayınladıkları makaleleri [82]'de, MCNN'yi (İki Sütunlu Evrişimli Sinir Ağları) kullanarak perspektiften kaynaklanan bozulmaları önlemek için çok sütunlu bir CNN modeli önermişlerdir. 2017'de Sam vd. [83] tarafından yapmış oldukları çalışmada en yüksek performansa sahip kitle sayma mimarisi olan Switch-CNN seçilmiş ve çok sütunlu bir CNN mimarisi üzerinden değerlendirmelerde bulunulmuştur. Küresel ve yerel bağlamsal bilgileri kullanarak kalabalık yoğunluğu sayılarını yüksek bir doğrulukla tahmin etmeye odaklanan Sindagi vd. [84] tarafından CNN modeliyle konteksten pramit elde etme yöntemi kullanılmıştır. 2018 yılında Shi vd. [85] genelleştirilebilir özellikler geliştirmek için negatif bir ilişki öğrenmeye odaklanmıştır. Bu yöntemle, güçlü genelleme yeteneklerine sahip ilgisiz regresörlerden oluşan öğrenme, içsel çeşitliliği yönetilerek sağlanmıştır. Zhu vd. tarafından 2019'da kalabalık yoğunluğunu hesaplamak için farklı regresyon ağları birlikte kullanılmıştır [86].

Regresyon temelli yaklaşımlar yoğunluk tahmininde başarılı olsalar da, düşük yoğunluklu durumlarda ve yerelleştirme gerektiğinde yeterince iyi performans göstermezler. CNN temelli yaklaşımlar, birçok alanda sınıflandırma, tanıma ve bölütleme gibi görevleri başarıyla gerçekleştirebilir. Yoğunluk tahmini, kalabalık sayma, yerelleştirme, izleme ve gözetim gibi kalabalık analizi çalışmalarında da kullanılırlar. Perspektif bozulmaları ve tek tip olmayan yoğunluk değişiklikleri gibi zorlukların üstesinden gelmek için mevcut çalışmalarda farklı CNN yaklaşımları da yer almaktadır.

Kalabalık video görüntülerinden insan sayısının tahmin edilmesi konusunda CNN yaklaşımı ilk kez Wang vd. [87] ve Fu vd. [88] tarafından kullanılmıştır. [87]'de yaya sayma görevi için CNN regresyon modeli kullanılırken, [88]' de yoğunluk görüntüsü üzerinde beş seviyede bir sınıflandırma işlemi gerçekleştirildi. Kişi sayma görevi için tek bir görüntü üzerinden yama temelli bir yaklaşım benimseyen [81], uçtan uca bir tahmin yöntemi önerilmektedir. En uygun regresyonun otomatik olarak seçilebildiği CNN yaklaşımı, insan sayma görevini diğer görevlerle birleştirmek için ilham verici bir yöntemde güncel çalışmalar arasındadır [89]. Shang vd. [90], sahneler arası sayma modeli ile değişen sahnelerde doğruluğu azaltma sorununun çözümünü vurgulamaktadır. Sığ ve derin CNN modellerinin bir kombinasyonunu öneren CrowdNet çalışmasında görüntüden anlamsal bilgi toplanarak başarılı sonuçlar elde edilmiştir [91]. Benzer şekilde, yoğunluk haritalamasını tahmin edebilen ve aynı anda farklı yoğunluk seviyelerinde sayılan insanları sınıflandırabilen kademeli bir CNN modeli önerilmiştir [92]. Mundhenk vd. [93] sayma, izleme ve algılama görevlerini birlikte yürütmek için insan sayma modeli tarafından üretilen yoğunluk haritalarını kullanmıştır. Daha üst düzey bir bilişsel görevde geçen insanların hızını hesaplamak için insanları saymayı birleştiren bir yöntem [94]'de önerilmiştir. Sindagi vd. [95] ve Kang vd. [96] tarafından hem insan sayımının hem de yoğunluk tahmininin ResNet temelli modelle aynı anda sınıflandırılabileceği yaklaşımlar önerilmiştir. Ilyas vd. 2019 yılında yayınladıkları çalışmada bağlama duyarlı bir ölçek toplama CNN temelli kalabalık sayma tekniğinde (CASA-Crowd) [97], odak noktası derinlik, değişen ölçek ve perspektifle değişen özellikleri kullanmışlardır. Ayrıca, bağlamsal bilgi elde etmek için değişen filtre boyutlarına sahip genişletilmiş bir evrişim kullanılmıştır. Öte yandan, farklı

genişleme oranları nedeniyle, perspektif bozulmasına neden olan alıcı alan boyutundaki bir değişimin üstesinden gelinmiştir.

Zhang vd. [98] CSRNet adlı bir yaklaşımı kullanmış ve ortaklama işlemlerinde değişiklikler yaparak öznitelik alanını genişletilmesi ve kalite yoğunluk haritalarının oluşturulması amaçlanmıştır. 2019 yılında Shi vd. [99] kalabalık tahmininde en yaygın perspektif problemini azaltmak için perspektife duyarlı CNN modeli önerilmiştir. Wan vd. [100], örnekler arasındaki ilişki bilgisini kullanarak bir rezidüel regresyon yaklaşımı kullanılmış, böylece genelleme kapasitesini artırmak için daha içsel öznitelikleri öğrenmesi amaçlanmıştır. Dai vd. [101] tarafından, değişen genişleme oranlarına dayanmak için üç evrişim bloğundan oluşan bir model kullanılan ve bağlamsal bilgileri yakalamak için diğer çalışmalardan daha küçük filtre boyutları seçilen bir yaklaşım önerilmiştir. Önceden eğitilmiş VGG-16 ağı da bu çalışmada temel model olarak kullanılmıştır. 2020 yılında Jiang vd. [102], önce çok seviyeli yoğunluk haritalarını uyarlamalı olarak öğrenen ve daha sonra kalabalıklardaki insan sayısını tahmin etmek için birleştiren çok seviyeli CNN mimarisi önermiştir. Oñoro-Rubio ve Lopez-Saste [103] tarafından, perspektif bozulmalarının üstesinden gelmek için farklı çözünürlüklerdeki öznitelikleri çıkarılarak, kalabalık dağılımı ve kalabalık sayısı hakkında perspektif bilgisi kullanmadan çalışan MCNN [82] modeline dayalı bir model önerilmiştir. Öte yandan, görüntünün özelliklerini öğrenmek için üç boyutlu filtreleme yaklaşımının kullanıldığı [104]'da çok sütunlu bir CNN yaklaşımı önerilmiş ve böylece farklı bilgi seviyelerini analiz etmesi amaçlanmıştır. Yang vd. [105]'de değişken yoğunluk haritaları için çok sütunlu bir CNN mimarisi önerilmiş ve bu şekilde daha başarılı sonuçlar elde edildiği öne sürülmüştür. Tian vd. [106]'de, farklı yoğunluktaki sahnelerde insanları saymak için yoğunluk haritalama kullanan bir yöntem önerilmiştir.

Öznitelik füzyon ağının kullanıldığı bir başka çalışmada, farklı yoğunluk seviyelerine ilişkin bilgileri daha etkin bir şekilde elde edilmiştir. Gao vd. tarafınan 2020'de [107] yayınlanan çalışmada, yüksek benzerlik ve görünümdeki perspektif değişikliklerinden kaynaklanan hataları azaltmak için bir perspektif kalabalık sayma CNN (PCCNet) yöntemi önerilmiştir. Sam vd. [108]'da, yoğun kalabalık

sayımı için her bir kişiyi kalabalığın içinde konumlandırılması için çok sütunlu bir CNN mimarisi kullanılmıştır. Bu çalışmada aynı zamanda kalabalıktaki insanların kafalarını bulmak için sınırlayıcı kutu yöntemi de kullanılmıştır. Performans, orta yoğunluklu görüntülerde başarılı olmakta ancak çok yoğun görüntülerde sağlam olmama sorunu göstermiştir. Guo vd. [109] ölçek çeşitliliği problemini çözmek için yüksek çözünürlüklü yoğunluk haritaları için ölçek toplama modülleri kullanmıştır. Shen vd. [110], görüntü bozulmasındaki GAN (Üretici Çekişmeli Ağlar) modellerinin başarısına dayanarak, kalabalık sayımı için yüksek kaliteli bir yoğunluk haritası üzerinde dört alt yamalı rakip çapraz ölçekli tutarlılık arayışını (ACSCP) kullanmıştır. Kontrast kaybını kullanarak, kaybı en aza indirmek için ana yoğunluk haritası ile birleştirilmiş görüntü yoğunluk haritası arasındaki mesafe hesaplanmıştır. Liu vd. [111] tarafından, hem regresyon hem de algısal saymayı içeren ve farklı görüntü pozisyonları için uygun sayma moduna uyarlamalı olarak karar veren birleşik bir yöntem önerilmiştir. Algılamaya dayalı yöntemlerdeki nesne dedektörleri, her bir kişinin konumunu belirleyebilmektedir. Böylece kalabalık sayma ve yerelleştirme [112]-[115] sağlanmaktadır. [115]'da önerilen model, [100]'daki kafa dedektörü eğitim modeline benzerdir. Genel nesne örneğini [116]'de önerildiği gibi modelde kullanmak yerine, bir ölçek haritası kullanarak ölçeğe duyarlı örnekler sunulmaktadır. Ölçekli haritalar, nesne ölçeklerini tahmin edebilmekte ve bu nedenle, bu, tüm ölçeklerde kapsamlı aramalar yapmaktan daha etkili bir yaklaşımdır.

### 1.1.5 Kalabalık Analizinde Dikkat Temelli Yaklaşımlar

Hu vd. [116]'de, görüntüdeki değerli özelliklere odaklanmak için SENet adlı bir dikkat mekanizması önerilmiştir [117] ve [118]'da kanal ve uzamsal dikkat mekanizmaları birleştirilerek daha başarılı sonuçlar elde edilmiştir. Dikkat modelleri ilk olarak makine çevirisi gibi doğal dil işleme alanında kullanılsa da, görüntü temelli nesne algılama, sınıflandırma, bölütleme ve yüz tanıma gibi çalışmalarda bir süredir önemli başarılar elde etmektedir [119]-[122]. [125], [116] ve [123]'ün avantajlarını birleştirerek, yöntem global bir bağlam modülü önermekte ve görüntü sınıflandırması da önemli bir başarı elde edilmektedir. Li vd. tarafından [124]'de, hedef sinir hücresi boyutlarını dinamik olarak seçebilen bir dikkat

mekanizmasının görüntülerde başarılı sonuçlar sağlayabileceğini gösterilmiştir. [126]'de iki CNN ağından oluşan ADCrowdNet, önce görüntünün kalabalık alanları tahmin edilmiş ve bu dikkat modeli daha sonra yüksek kaliteli yoğunluk haritaları oluşturulması için kullanılmıştır. 2020 yılında yayınlanan [127]'de kalabalık sayımı için bir öznitelik füzyon dikkat ağı (FFANet) önerilmiştir. FFANet, VGG16 ağı ile birlikte uygulanmış ve kalabalık görüntülerden çıkarılan özellikler birleştirilmiştir. Çok seviyeli özniteliklerle ilgili bilgi geliştirme operasyonları, RB (Resudial Block) ile daha da geliştirilmiş olan öznitelik fizyon dikkat modülü (FFAM) tarafından gerçekleştirilmiştir. Bu öznitelikler, bir yoğunluk haritası oluşturmak için sıkıştırma modülü tarafından birleştirilmiştir.

Tez çalışması kapsamında yapılan literatür incelemesinde, kalabalık analizi konusunda pek çok yaklaşım hakkında referans verilmektedir. Spesifik olarak, görüntüler üzerinden kalabalık analizi konusunda geleneksel yöntemlere değinilmekle birlikte, esasen derin öğrenme mimarileriyle kalabalık analizi konusuna katkıları araştırılmakta ve özetlenmektedir.

## 1.2 Tezin Amacı

Kalabalık yönetimi ile şehirlerin ve sosyal alanların planlanması, toplanma alanlarının simüle edilmesi, kentsel planlama, afet yönetimi gibi görevler gerçekleştirilebilmektedir. 2020 yılı ile birlikte dünyayı etkileyen ve bir yılı geçkin süredir devam eden Covid-19 pandemisi sürecinde de sosyal mesafe kuralları kapsamında kamu sağlığının güvenliği denetlenmektedir. Bahse konu analiz, denetleme ve düzenleme işlemlerinin gerçekleştirilmesi için önemli ve pratik yöntemlerin başında kamera görüntülerinden faydalanmak gelmektedir. Yaygın etkiye sahip bu amaç veya farklı hedefler için kalabalık analizi gerçekleştirilmesi ve bunun optimum şekilde yapılabilmesi bu tezin başlıca konusudur. Kalabalık analizi literatür incelemesi ışığında görülmektedir ki; probleme özgü yaklaşımlar geliştirilmesine ihtiyaç vardır.

Bu çalışmanın amacı yoğun kalabalık görüntüler de dâhil olmak üzere buradaki kişi sayısını kestirmektir. Bu sayede kamuya açık alanların tasarlanması, spor müsabakaları, mitingler gibi ortamların analizinin yapılması, müze ve kültür-sanat toplantı alanlarının tasarlanması, alışveriş merkezleri gibi dinamik ortamlarda

pazarlamaya yönelik de istatistiklerin çıkarılmasına destek olabilecek bir yaygın etki hedeflenmiştir. Kalabalık içinde oluşan alt grupların belirlenmesi ve hareket değişim anlarının tespit edilmesi de tezin kapsamında araştırılmış ve çalışılmıştır. Bu çalışmalar gelecekte kalabalık içindeki anormal davranış gösteren veya gösterecek olan kişilerin önceden kestirilmesine yönelik çalışmalara da ışık tutması amaçlanmıştır.

## 1.3 Hipotez

Video görüntülerinde kalabalık analizi konusundaki geniş literatür incelemesinde mevcut yöntemlerin genelleştirilebilir olmadığı ve pek çok alanda başarısını kanıtlamış olan CNN'ler ile sağlam ve etkin sonuçlara ulaşmada kısıtlar tespit edilmiştir. Bu bağlamda, video görüntülerinin doğası gereği zamansal bilginin, kalabalık davranışının doğası gereği ise uzamsal bilgilerin de dikkate alınacağı teorik bir yapay zekâ yaklaşımı önerilmektedir.

Bu yaklaşım ile CNN'lerin kırılganlıklarının üstesinden gelebilmesine odaklanılmaktadır. Önerilen teorik yaklaşımın deneysel sonuçları literatürdeki güncel çalışmalarla karşılaştırılmıştır.

# 2
# YAKLAŞIM VE YÖNTEMLER

Yapay zekâ (YZ) kavramının ilk kez kullanıldığı Dartmouth Yapay Zekâ Araştırma Projesi ekibinde yer alan bilim insanı Minsky[1] şöyle demiştir: *"Genelde beynimizin en iyi yaptığı şey, en az farkına vardığı şeydir."* Bu söz; sürekli öğrenerek gelişen beynin genellikle en iyi çalışan programlarını en derine gömmesi ve zamanı geldiğinde bunu açığa çıkarması olarak açıklanabilir. Buna içgüdü de denir. Genetik olarak aktarılabilir ve çoğu zaman nasıl yapıldığı açıklanamaz eylemler ve kararlar olarak tanımlanmaktadır. Cosmides ve Tooby makalelerinde bunu içgüdü körlüğü olarak tanımlar ve şu şekilde ifade ederler: *"Bir şey ne kadar doğal ve kolay görünürse, gerçekte durum o kadar terstir."*[2] Bilinç bu kararlara müdahale edemez çünkü bu gömülü karar sistemi çoğu zaman yaşamsal önem bildiren faaliyetleri yönetir. Bu davranışlar genelde sinirsel açıdan çok karmaşık yapıdalardır. Hızlı, doğru, etkili şekilde yapılan işlerin modellenmesi bu yüzden zordur. Hatta bu durum çözülmemiş problemler olarak tanımlanmaktadır [128]. Canlı özellikle de insan, öylesine derin sinir katmanlarında varoluştan bu yana kodlanmış bilgiler barındırır ki bu bilgileri nasıl işlettiğinin anlaşılması uzun zamandır bilgisayar bilimlerini besleyen tartışma konuları arasındadır.

Bilgisayar bilimlerinin alt konusu olarak tanımlanan YZ, bir yandan beyin-sinir bilim araştırmalarının ilerleyişi ve teknolojinin gelişmesiyle günlük hayatın hemen her noktasına yerleşmiş durumdadır. Ceplerde taşınan telefonlar insanları parmak izi ve yüz görüntülerinden faydalanarak tanıyabilmektedir. Finansal işlemleri gerçekleştirmek için bankalarda akıllı sohbet robotlarıyla muhatap oluyor ve onlar da bizleri sesimizden tanıyabiliyorlar. Giyilebilir teknolojiler sayesinde sağlık verileri gerçek zamanlı olarak bulut ortamında işlenebilmekte ve böylece canlıların

[1]Marvin Minsky'nin 1998 yılında yayınlanan The Society of Mind eserinden. Bkz: https://en.wikipedia.org/wiki/Society_of_Mind

[2]Leda Comides ve John Tooby'nin 1992 yılında yayınlanan, Cognitive Adaptions for Social Exchange eserinden ve Eagleman'ın 2013 basımlı Incognito, Beynin Gizli Hayatı eserinde de üzerinde durulmaktadır.

gelecekteki sağlık durumlarına ilişkin kestirimlerde bulunarak sağlık personelinin hastalıkları erken teşhis etmesine olanak tanımaktadır. İnternet kullanarak yapılan alışverişler, izlenen filmler, dinlenen müzikler kişilere ilgi alanına yönelik tavsiyeler sunarak hayatı hızlandırmakta ve kolaylaştırmaktadır. Üretilen verilerin işlenmesi ile kişiye yönelik ürün, ortam tasarlanabilmekte ve bununla birlikte kişiselleştirilmiş eğitim hizmetleri de sunulmaktadır. Açık ve kapalı alanlara yerleştirilen kameralarla insanların, taşıtların ve çevrenin durum analizleri yapılabilmektedir. Bunlar günümüzde kullanılan YZ teknolojilerine sahip ürünlere yalnızca birkaç örnektir. Bu tür faydalı çalışmaların yapılabiliyor olmasındaki en büyük pozitif etken ise veri miktarının artması ve veriye erişimin geçmişe kıyasla daha kolay ve kurallarının belirli olmasıdır.

Nicelik bakımından veri çoğalmış olsa bile yeterince çeşitliliğe sahip, yanlı olmayan, gürültülerden arındırılmış ve kapsayıcı bilgiye sahip olması beklenmektedir [129]. Diğer tüm sektörlerde olması gerektiği gibi çevre sağlığı ve güvenliğinin sağlanması için kalabalık analizi verisinin üretilmesi, önişlemler yapılması, depolanması ve paylaşılması yani veri yaşam döngüsü süresince etik ve hukuksal açıdan değerlendirilmeli, güncel düzenlemeler takip edilmelidir [130].

Bu bölümde; video görüntülerinde kalabalık analizi için ihtiyaç duyulan teknik yaklaşımlar özetlenmiştir. Bahse konu yaklaşımlar doğrudan ya da kısmen bu doktora tezi kapsamında önerilen yaklaşımı beslemektedir. Kalabalık analizinde daha önce uygulanmayan özgün bir mimarı bu tez çalışması sonucunda önerilmektedir. Bu mimarinin matematiksel detayları, probleme ilişkin yaklaşım neden sonuç ilişkisine dayandırılarak öne sürülmektedir. Yaklaşım ve Yöntemler bölümünde matematiksel ve derin öğrenme mimarisine yönelik detaylara yer verilmektedir.

## 2.1 Yaklaşım

*"Yapay zekâ, makinelere zekâ kazandırmaya adanmış etkinliktir; zekâysa, bir varlığın kendi ortamında doğru düzgün ve olan biteni öngörerek işlev görmesini sağlayan niteliktir."* der, Nilsson[3]. 1950 yılında Alan Turing'in bilgi işlem makineleri ve zekâ makalesinde *"makineler düşünebilir mi?"* sorusu ve yapmış olduğu Turing makinesi tanımı YZ'nın ilk tanımları arasındadır [131]. Öyle ki 1950'li yıllardan günümüze, yaşanan gelişmeler YZ tanımını sürekli olarak güncellememizi gerektirdi. Minsky dereceli problemler dizileri kavramı ile de aslında bir YZ tanımı yapmaktadır. 1980'lere gelindiğinde sembolik YZ tanımı kullanıldı [132]. Bu aslında bilinen adıyla uzman sistemler demektir. Tüm bu tanımların çıktığı kapı, tek bir program ile insansı şekilde sonsuz problemi çözebilme ve öğrenebilme yeteneğinin olması arzusuna dayanmaktadır. Ancak 2000'li yıllara kadar gelişmeler zaman zaman daha hızlı bazen de durma seviyesinde yavaş ilerledi. Henüz ilk tanımını kapsayacak türden bir YZ sistemi de üretilmedi.

Günümüzde probleme özel çözümler sunan çalışmalar yoğunluklu olarak kullanılmaktadır. Turing ve Minsky'nin hedeflediği tanıma ulaşmak için emeklenmeye devam edilmektedir. Bu amaca ilerlerken YZ tanımını dar kapsamlı, genel ve süper olmak üzere üçe ayırmak uygun olmaktadır.

- **Dar kapsamlı YZ (ANI):** YZ'nın ilk örnekleri sayılabilecek belirli görevleri yeri getirebilen yazılım ve/veya donanım bütünüdür. İstenen sonuçlara göre kullanılan teknik ve yöntemler ile kuralların dışına çıkılmamaktadır. Günlük hayatta insanın rutin işlerini kolaylaştırmak amaçlı tercih edilmektedir. Günümüzdeki pek çok YZ uygulaması bu tanım altında değerlendirilmektedir.
- **Genel YZ (AGI):** Öğrenme ve kendini geliştirme gibi canlıya dair yeteneklere sahip ve insansı performansa yakın sonuçlar üretmesi beklenen YZ teknolojilerine verilen isimdir.

[3] Nils J. Nilsson 2010 basımlı ve 2018'de Türkçeye çevrilen Yapay Zekâ Geçmişi ve Geleceği kitabında derin kavramların üzerinde durmaktadır.

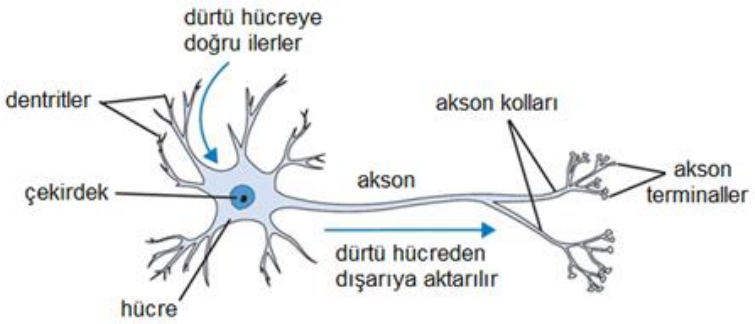
- **Süper YZ (ASI):** İnsanların becerilerinin ötesinde performans sergilemesi beklenen, öğrenen ve bağımsız kararlar alan ve uygulayabilen YZ teknolojilerini işaret etmektedir. Pek çok bilim insanı için 2050 yılına kadar YZ'nın bu seviyeye ulaşması öngörülmektedir. Henüz örneği bulunmayan bir tanımdır.

Bu tanımlar yapay öğrenme (makine öğrenmesi) kavramını beraberinde getirmektedir. Öğrenme işlemini modelleyen bu yapıyla, örneğin görüntüler için görsel korteks işlevinden yola çıkılarak özniteliklerin belirlenmesi sağlanmaktadır. Bu işlem için de veri miktarı ve bilgisi önemli hale gelmektedir. YZ'nın tanımı daha geniş iken yapay öğrenmenin tanımı veriye dayalı bir kapsamdadır. Büyük veriden çıkarılan örüntülerden faydalanarak algılama, tanıma, kestirim ve karar verme sistemleri geliştirilmektedir. Bu sistemlerin temelinde yapay sinir ağları teorisi yatmaktadır.

### 2.1.1 Yapay Sinir Ağları

1980 ve 1990'lı yıllar yapay sinir ağları (YSA) alanındaki araştırmaların günümüzde olduğu gibi çok atıf topladığı bir dönem olarak görülmektedir. Erken dönem çalışmalarda bu alanın adına sibernetik denmekteydi. Sibernetik; canlıların davranışlarını inceleyerek, matematiksel olarak tanımlanması ve benzer yapay modellerin üretilmesi olarak açıklanmaktadır. Eğitilebilen, adapte olabilen, kendi kendine organize olup öğrenebilen ve değerlendirme yapabilen YSA ile canlı beyninin öğrenme yapısı modellenmeye çalışmak amaçlansa dahi kıyaslanamaz derecede daha basit bir yaklaşımıdır.

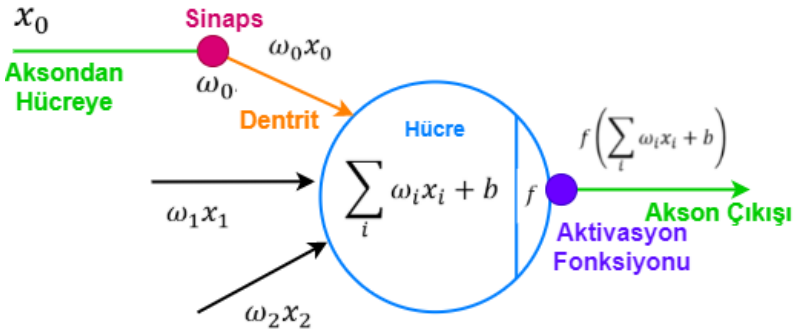
Yolculuk biyolojik bir sinir hücresi yapısının anlaşılmasıyla başlamaktadır. Sinir hücresinde iletilmesi istenen periyodik bir işaret çekirdek tarafından oluşturulmaktadır [133]. Akson yardımıyla dentritler üzerinden sinir hücreleri arasında iletim gerçekleştirilmektedir. Bu yapı Şekil 2.1'de gösterilmektedir.




**Şekil 2.1** Biyolojik bir sinir hücresi [134]

Taşınan işaret akson yolunca potansiyel değeri -40mV (Na+ iç) seviyesine eriştiğinde pozitif (+) değerli bir elektrik akım iletimi gerçekleştirmektedir. Potansiyel değer -90mV (K+ dış) olduğunda ise negatif (-) elektrik akıma yani bastırma işlevine sebep olmaktadır. Bu eşik değerinin üzerinde gerilim oluştuğunda sinir hücresi uyarılmakta, değerin altında kaldığındaysa durdurulmaktadır. Buna sinirsel hesaplama denmektedir [133].

Biyolojik sinir hücresinden yola çıkarak temel seviyede matematiksel bir model oluşturulduğunda dentritler boyunca hücreye giren bilgileri ve bu bilgilerin akson çıkışı ile belli bir aktivasyon değeri kullanılarak sonraki yapay sinir hücrelerine iletilmesi için tanımlanan model Şekil 2.2'de gösterilmekte ve denklem (2.1) ile ifade edilmektedir.




**Şekil 2.2** Matematiksel bir sinir hücresi modeli [134]

Yapay bir sinir ağı temel düzey yapısına algılayıcı denmektedir. Yapı doğrusal bir denklem ile ifade edilmektedir. Yapının kendisi bir sinir hücresini temsil etmektedir. Sayısını artırır ve katmanlar boyunca eklenmesiyle derin sinir ağı yapısı elde edilmesi mümkündür.

$$y = W \times x + b \tag{2.1}$$

Denklem (2.1)'de $W$ güncellenebilir ağırlıkları, $x$ giriş bilgisini ve $b$ önyargı değerini ifade etmektedir. Sinir hücrelerinin kendi katmanına etkisi ve her bir başka katmanın da genel mimariye etkisi hesaplanmaktadır. Sinir hücresi sayısı, katman sayısı, aktivasyon fonksiyonu seçimi, optimizasyon algoritması seçimi ya da başlangıçtaki ağırlık değerlerinin belirlenmesi gibi birçok parametre çıkış performansını doğrudan etkilemektedir. Yüzlerce katmanı ve milyonlarca sinir hücresi barındıran modeller de tasarlamak mümkündür. Tasarlanan YSA modelinin başarımı çıkışında hesaplanan kayıp fonksiyonunun sıfır değerine yakınsaması ile ölçülmektedir. Optimizasyon teknikleri ile kayıp değeri minimize edilmektedir. Bunu yapabilmek, ileri ve geri yayılım algoritmaları kullanmakla mümkün olmaktadır.

### 2.1.2 Yapay Öğrenme

Bilgisayarlara istenilenleri yaptırmak için ne istenildiğinin algoritmasını bilmek gerekir. Mevcut girdiden elde etmek istenen çıktı planlanabiliyorsa buna algoritma denir ve bu programlanabilir. En basit hali ile bir matematiksel işlemi programlamak kolaydır. Ancak bazen öyle problemlerle karşılaşılır ki canlının nasıl yaptığını açıklanamadığı ya da sezgisel yaklaşımlar gerektiren değişken durumlar olur. Bu halde sabit bir algoritma ile sonuca ulaşmak mümkün olmamaktadır. Toplanan verilerden faydalı anlamlar çıkarmak suretiyle kendi algoritmasını oluşturan sisteme yapay öğrenme denmektedir. Yapay öğrenme yaklaşımlarında başarılı olunabilmesinin en önemli kriteri verilerin işlemeye uygun şekilde üretilmesi, toplanması, depolanması ve önişlemler uygulanması aşamasıdır [135]. Bazı durumlarda verinin tamamının öznitelikleri elde edilemeyebilir ya da verinin kendisinde de eksiklikler olabilir. Veri kümesinin bu gibi özellik ve kısıtlarına göre tercih edilmesi gereken yapay öğrenme yaklaşımı farklılık göstermektedir.

Yapay öğrenme yaklaşımları temelde istatistik ve olasılık teorilerinden beslenmektedir. Bu tip bir yapının iyi çalışması için hacimce, sayıca ve hızca büyük veriye ihtiyaç duyulmaktadır. Tıp, pazarlama, finans, haberleşme, astronomi gibi büyük veri üreten spesifik alanlarda başarılı kestirimler yapılmaktadır. Büyük veri içinden anlam ve bilgi taşıyan örüntüler tanınarak katma değerli çıktılar üretilmektedir. Örneğin; rüzgâr enerjisinden elektrik enerjisi üreten modülün

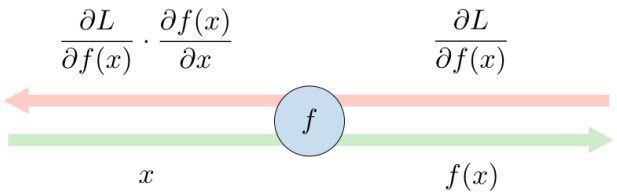
motorunun ne zaman arızalanacağını üzerindeki sensörlerden toplanan zaman serisi verilerden yola çıkarak geleceğe dair kestirim yapmak olanaklıdır. Böylece enerji, iş ve dolayısıyla maddi tasarruf sağlanmaktadır.

Özetle yapay öğrenme, YZ'nın alt konularından olup belli kurallara bağlı kalmadan ve algoritması açıklanamayan karmaşık problemlerde veriden faydalanarak çözüm üretmektir. Tüm bu işlemler kullanılan donanımın bellek ve zaman yeterliliklerinden de etkilenmektedir. Bu yüzden büyük veri işlenecek donanımları seçerken de dikkatli olmak gerekir. Yapay öğrenme yaklaşımları verinin özellikleri ve bu özelliklerden faydalanma şekline göre denetimli, denetimsiz ve pekiştirmeli öğrenme olmak üzere yeniden tanımlanmaktadır. Denetimli öğrenme etiketlenmiş veriye ihtiyaç duyarken denetimsiz öğrenme etiketli olmayan verilerin birbirlerine çeşitli matematiksel yaklaşımlarla benzerliğine ve yakınlığına göre gruplanmasına odaklanan bir alt alandır. Pekiştirmeli öğrenme ise veri kümesinin kısıtlı olduğu ve sabit olmadığı durumlarda tercih edilmektedir. Ödül ya da pekiştirme kavramı psikolojide de öğrenme kuramına[4] dayanmaktadır. Pekiştirmeli öğrenme diğer tüm yapay öğrenme modellerinden farklı olarak dinamik bir yapıya sahip ve çevre ile etkileşim içinde bir yapıya sahiptir [136].

#### 2.1.2.1 İleri Yayılım ve Geri Yayılım

Yapay öğrenme ve pek çok YSA modelleri giriş verisi, beklenen çıktılar ve çözümleme algoritmasının başarısı çıkışta elde edilen ve beklenen arasındaki farkın ölçümlenmesiyle elde edilmektedir. Elde dilen farkı minimize eden optimizasyon algoritması tercih edilerek işlemi ileri yönde tekrar başlatabilmek için geriye doğru tüm fonksiyonları türevi alınmaktadır. Beklenen değer ile elde edilen çıktı arasında fark minimum olana kadar ileri ve geri yönde bu işlemler iteratif olarak devam etmektedir. Şekil 2.3 ve Şekil 2.4'de gösterilmektedir.

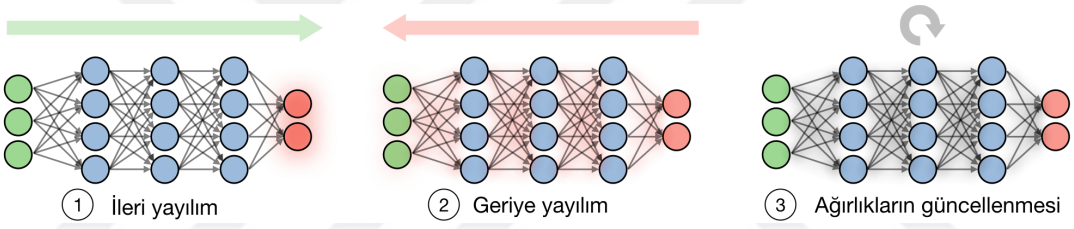
[4] Edward L. Thorndike, 1911'de hayvanlarda pekiştirmeli öğrenme üzerine çalışmalar yürütmüştür. Bkz. Animal Intelligence, New York: The Macmillan Co.




**Şekil 2.3** İleri ve geri yayılım gösterimi [137]

Geriye yayılım algoritması Rummelhart ve çalışma arkadaşlarının 1986 yılında öne sürdükleri ve bilinenin aksine öğrenme algoritması değil buna katkı sağlayan ve gradyanların nümerik olarak hesaplanması yöntemidir. Sinir ağındaki her birim geriye doğru türevlenerek giriş değerine ulaşmakta ve güncellenen ağırlıklar ile model tekrar ileri doğru yeni çıktıyı hesaplamaktadır. Türev işlemleri zincir kuralına göre denklem (2.2)'deki gibi hesaplanmaktadır [137].

$$w \leftarrow w - \alpha \frac{\vartheta L(z,y)}{\vartheta w} \tag{2.2}$$

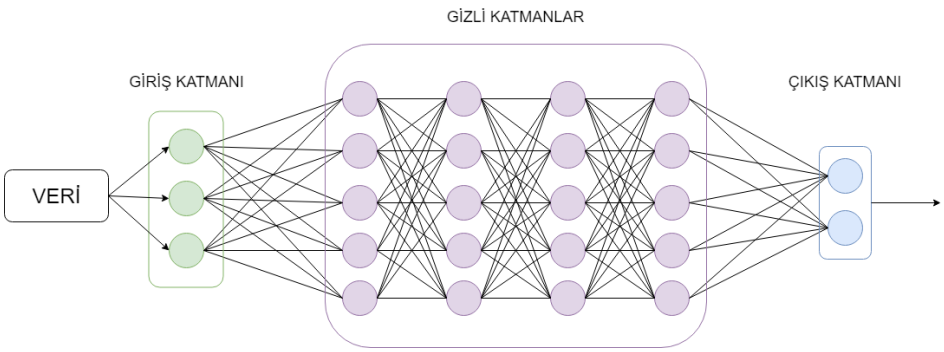



**Şekil 2.4** İleri/geri yayılım ve ağırlıkların güncellenmesi gösterimi [137]

### 2.1.3 Derin Öğrenme

Tüm bir model büyüklüğü, matematiksel işlem karmaşıklığına ve giriş verisinin boyutlarına bağlı olarak yüksek işlemci kapasitesi gerektiren donanımlara ihtiyaç duymaktadır. Bu ihtiyacı karşılamak üzere, merkezi işlem birimi, grafik işlem birimi ya da bazen daha güçlü olan tensor işlem birimleri[5] kullanılmaktadır. Bu donanımlara bulut servisleri kullanarak da erişim sağlamak mümkün olmaktadır. Bununla birlikle bellek ve YZ modelinin çalışacağı koşullara göre örneğin otonom bir robot üzerinde kullanılacaksa, gelişmiş bataryalara da ihtiyaç duyulmaktadır. Basitleştirilmiş bir tam bağlantılı derin sinir ağı yapısı Şekil 2.5'de gösterilmektedir.

[5] Tensör, düzenli bir x,y,z düzlemindeki ızgara yapısına yerleştirilen dizilere denir. Bu üç boyutlu diziler üzerinde işlem yapmayı sağlayan Google tarafından geliştirilen donanımların adı Tensör İşlem Birimi (TPU)'dur.




**Şekil 2.5** Basitleştirilmiş derin YSA modeli

Derin öğrenme her ne kadar 2006 yılından sonra popüler olup bu ismi almış olsa da YZ tarihi boyunca farklı tanımlamalar altında çalışılan bir araştırma alanıdır [139]. Özellikle YSA'nın eğitimi için verinin ve model boyutlarının artması, donanımların bu işlem gücünü karşılayabilir hale gelmesi çalışmaların üniversite laboratuvarlarından çıkarak projelendirilip gerçek hayatta karşılık bulmasına olanak tanımaktadır. Karmaşık problemlerin çözümünde klasik yaklaşımlardan daha başarılı sonuçlar verdiği de sayısız araştırma ışığında anlaşılmaktadır.

Derin öğrenmenin temel anlamda beyin benzetimi yapmak gibi bir amacı yoktur. Ancak sinirbilim ya da hesaplamalı sinirbilim kaynakları bazı derin öğrenme araştırmacılarının da ilgi odağı durumunda olabilmektedir. Çağdaş derin öğrenme bilgi kuramı, olasılık-istatistik, lineer cebir gibi birçok alandan beslenmektedir [136].

Günümüzde uygulama yöntemlerinin başında gelen derin öğrenme özetle; insanların çözebildiği zekâ gerektiren karmaşık problemleri çözmeyi hedefleyen veri, algoritma, model ve donanım bilgisi gerektiren sistemlerinin oluşturulması üzerine disiplinler arası bir çalışma alanıdır.

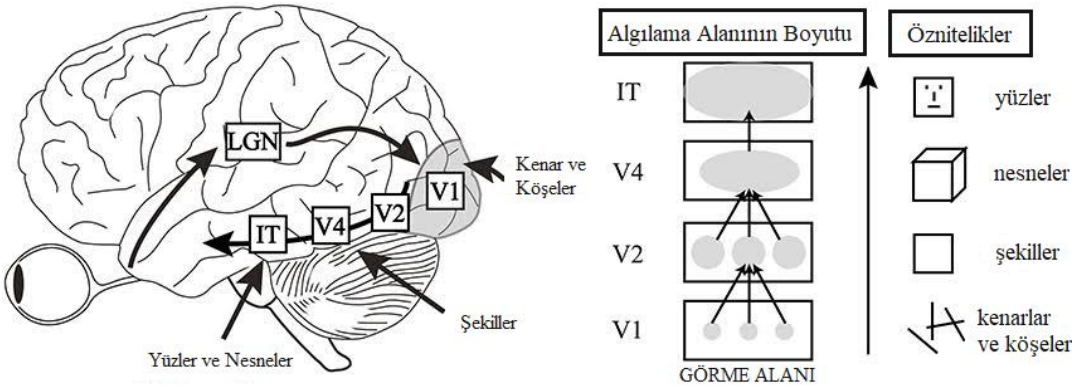
Derin öğrenmenin çeşitli alt konuları vardır. Bunlar uygulandıkları veriden anlam çıkarma şekillerine göre sınıflandırılabilir. Örneğin görüntüler üzerinden örüntüler çıkarmak için sık kullanılan yapılar CNN'dir. 2018 yılında ortaya atılan kapsül ağları bilgisayarlı görü çalışmalarına yeni bir boyut kazandırmıştır. Bu bölümde video görüntülerinde kalabalık analizi için başvurulan derin öğrenme yönemlerinin temel literatürü özetlenmektedir. Bu bağlamda tez kapsamında önerilen yaklaşım matematiksel anlamda modellenmektedir.

#### 2.1.3.1 Evrişimli Sinir Ağları

Derin öğrenme için her ne kadar beyin benzetimi amacı yoktur vurgusu yapılsa da sinirbilimden esinlendiği ya da sinirbilim ile temellendirdiği birtakım yaklaşımlar vardır. Bunların bir kısmı başarısızlığa uğramış olsa da CNN'ler, YSA yöntemlerine 1989[6] yılı itibariyle tekrar güven kazandıran ve çokça uygulama alanı bulan derin sinir ağı yöntemidir. Nörofizyolog Hubel ve Wiesel 1950'li yılların sonu itibari ile memeli canlıların görsel korteksleri üzerinde yaptıkları deneyler ile yayınladıkları sonuçlar Nobel ödülü ile taçlanmıştır [140]. Kediler üzerinde yaptıkları deneylerin[7] en önemli bulguları özellikle birincil görsel korteks olarak tanımlanan beynin V1 bölgesine aittir. V1 bölgesinin retinadan algılanan görüntünün uzamsal haritalaması, küçük ve karmaşık hücreler ile ışık ve uzamsal geçişlerin algılanması CNN yapısında önemli bilgi kaynağı olmaktadır. Ancak bununla beraber CNN yapısı memeli görsel algılama sisteminden çok farklı yönler de barındırmaktadır. Örneğin çoğunlukla görüntü algılama sensörleri yüksek çözünürlüklü görüntüyü olduğu gibi giriş bilgisi olarak kullanmaktadır. Ancak memelilerde görüntü fovea bölgesi dışında oldukça düşük çözünürlüklüdür. Giriş olarak tanımlanan görüntünün odaklanılan küçük bir bölgesi yüksek çözünürlüklü algılamaya sahiptir. Şekil 2.6'de bu yapı gösterilmektedir.

[6]LeCun ve çalışma arkadaşları tarafından 1989 yılında ABD Posta Servisi tarafından sağlanan el yazısı posta kodu rakamlarının tanınması için geriye yayılım algoritması kullanılarak yapay öğrenme uygulaması yapılmıştır.

[7]Hubel ve Wiesel kedilerin görsel korteksinden ölçümledikleri deneyin kayıtlarını YouTube üzerinde bulmak mümkündür. Bkz. https://www.youtube.com/watch?v=8VdFf3egwfg




**Şekil 2.6** Memeli canlıların görsel hiyerarşi özellikleri [140]

Gözler bir duyu organı olarak diğer duyumsal algılarımızdan da etkilenerek dışardan alınan girişleri yorumlamaktadır. İki boyutlu bir algı olarak tanımlansa da memeli canlılar görüşlerine bağlı olarak üç boyutlu uzaydaki konumlarına yönelik çıktılar da üretmektedir ve oldukça karmaşık, bilinmezliklerle dolu bir araştırma alanıdır. Memeli görü sistemi ile ilgili henüz bilinmeyen birçok konu vardır. Bu yüzden CNN'lerin tamamen memeli görü sistemi şeklinde çalıştığını iddia etmek doğru değildir.

1980 yılında Fukushima neocognitron makalesinde CNN mimarisinin bugünden biraz daha farklı kümeleme temelli ve denetimsiz öğrenme yaklaşımındaki bir sürümünden bahsetmektedir [141]. 1988 yılına gelindiğinden Hinton[8] zaman-gecikmeli sinir ağları makalesinde modeli eğitmek için ilkkez geriye yayılım algoritmasını kullandı [137]. Tüm bu bilgiler ışığında LeCun 1989 yılında bugün bilinen CNN mimarisini görüntülere uyguladı. CNN'lerin başarılı sonuçlar üretmesi emekleme aşamasındaki bilgisayarlı görü konusunu 2000'li yıllar itibari ile YZ gündeminin üst sıralarına taşımış oldu [142].

Temel olarak ızgara yapılı bir topolojiye ve yüksek boyutlu verilerdeki örüntüleri yakalama yeteneğine sahip CNN'ler ile 2012'de ImageNet[9] nesne tanıma yarışmasında Hinton'ın danışmanlığında Krizhevsky ve Sutskever önceki yılın

[8] Geoffrey Hinton, 1970'li yıllardan bu yana yapay zekâ alanında çalışmalarıyla 2018 Turing Ödülü'nü Yoshua Bengio ve Yann LeCun'la birlikte almaya hak kazanmış, Toronto Üniversitesi'nde profesör unvanına sahip bilim adamıdır.

[9] ImageNet, 2009 yılında Dr. Fei Fei Li tarafından ücretsiz olarak yayınlanan binlerce görüntü sınıfından meydana gelen, WordNet hiyerarşisine göre düzenlenmiş üzerinde her yıl nesne tanıma ve sınıflandırma yarışmaları düzenlenen bir görüntü veritabanıdır. Bkz. http://www.image-net.org/

sonuçlarını neredeyse ikiye katlayarak büyük bir başarı elde ettiler [143]. Ng[10] bir röportajında, bu çalışmada paralel grafik işlem birimlerinin kullanılmasını önerdiğinden bahsetmektedir. Böylelikle CNN'ler ve görüntü verilerinin gelişen donanım teknolojiyse birleşerek daha başarılı ve hızlı sonuçlar verdiği ispatlandı. CNN'ler en basit tanımı ile katmanları boyunca matris çarpımları yerine evrişim işlemine dayalı olan birden çok işlemi içinden barındıran katmanlara sahip sinir ağı yapısıdır. Bu yapının temeli olan evrişim zamansal $(t)$ ve uzamsal $(x)$ bilgisi barındıran matematiksel bir işlemdir. Zamanla yer değiştiren bir cismin ölçümlemesi yapılırken ölçüm kanalında gürültüler de kaydedilir ve zaman ilerledikçe gürültülerin fonksiyon içindeki önemi artar hale gelir ki bu istenmeyen bir durumdur. Bu gürültü etkisini azaltmak adına ölçümlerin ortalamalarını alan bir fonksiyon $w(a)$ evrişim işlemine dâhil edilmektedir. Ağırlık fonksiyonu olarak bilinen $w$ olasılık yoğunluk fonksiyonu şeklinde tanımlanmaktadır. Yapay öğrenme yaklaşımlarında $w$ yerine farklı fonksiyonlar da tanımlamak mümkündür. Tüm $(t)$ zamanlarında uzamsal bilgiye bu işlem uygulandığında denklem (2.3) elde edilmektedir. Evrişim işleminin gösterimi $*$ işareti ile gösterilmektedir. Burada; $x$ girdileri, $w$ tanımlanan çekirdek fonksiyonunu ve çıktı olarak hesaplanan $s$ öznitelik haritasını vermektedir.

$$s(t) = \int x(a)w(t-a)da \equiv s(t) = (x * w)(\mathrm{t}) \tag{2.3}$$

Zamandaki ölçümlerin periyodik ve ayrık olarak yapılması gerçek zamanda araştırmacıların problem ölçümlemede en çok kullandığı halidir. Bu yüzden zamanda ayrık şekilde evrişim işlemi denklem (2.4) gösterilmektedir. Farklı boyutlar içinde tanımlamaya elverişlidir. Örneğin standart bir görüntü iki boyutlu tanımlanmaktadır. Bu da evrişim işlemini iki eksen için tanımlamak gerekliliği getirmektedir. Evrişim sırabağımsız bir fonksiyon olması dolayısıyla çekirdek fonksiyonu ile giriş matrisinin işlemdeki sırası sonucu etkilememektedir. Giriş

[10] Andrew Ng, 2000'li yıllarda özellikle derin öğrenme alanında damgasını vuran çokça yapay zekâ şirketine sahip ve Stanford Üniversitesi Elektrik ve Bilgisayar Bilimlerinde profesör unvanına sahip bilim adamıdır. Bkz. Röportaj: https://www.youtube.com/watch?v=-eyhCTvrEtE

bilgisi $G$ ve çekirdek fonksiyonu Ç olarak tanımlanırsa çıkış $S$ zamanda ayrık örneklemler için $i$ ve $j$ eksenlerinde denklem (2.5)'deki gibi ifade edilmektedir.

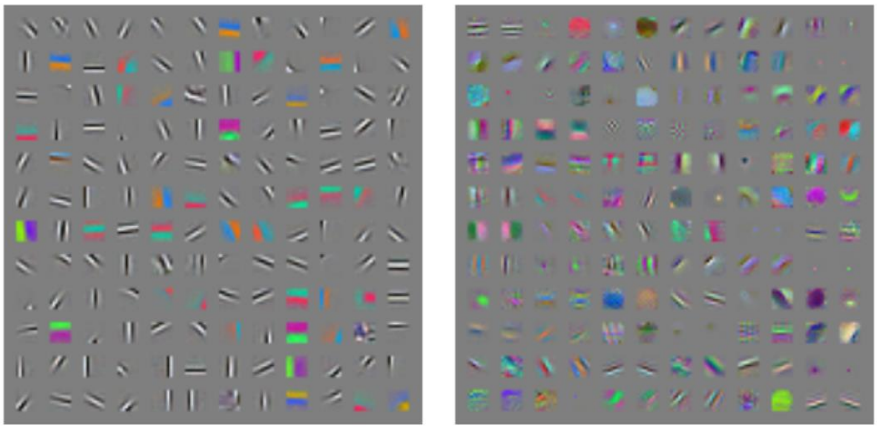
$$s(t) = (x * w)(\mathrm{t}) = \sum_{a=-\infty}^{\infty} x(a)w(t-a) \tag{2.4}$$

$$S(i,j) = (G * \text{Ç})(i,j) = \sum_{m}\sum_{n} G(m,n)\text{Ç}(i-m, j-n) \tag{2.5}$$

Bu denklem yapısından yola çıkarak yapay öğrenme modelleri çekirdek fonksiyonu eksende çevirmek yerine çok benzer bir başka matematiksel gösterim olan çapraz ilişki ifadesini benimsemektedir. Bu gösterim denklem (2.6)'da gösterilmektedir.

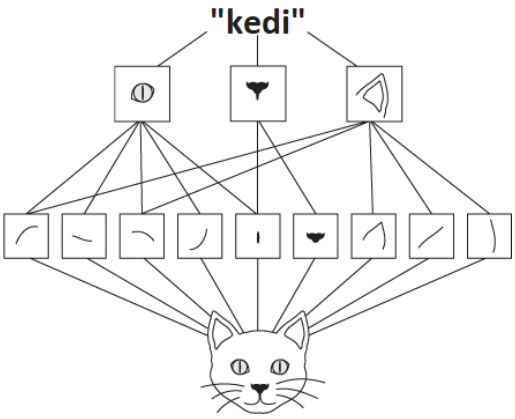
$$S(i,j) = (\text{Ç} * G)(i,j) = \sum_{m}\sum_{n} G(i+m, j+n)\text{Ç}(m,n) \tag{2.6}$$

Çapraz ilişki ve ters ilişki sinirbilim ile ilişkili Gabor fonksiyonlarının memelilerin görsel kortekste V1 hücresi olarak tanımlanan ağırlıkların yapısına benzerlik göstermektedir. Bu matematiksel yaklaşım ve CNN ile elde edilen öznitelik haritaları görselleştirildiğinde, görsel girdideki basit öznitelikler olarak tanımlanan kenar, köşe, parlaklık ve renk geçişleri gibi yüksek frekanslı uzamsal bilgileri çıkarmaktadır (Şekil 2.7).



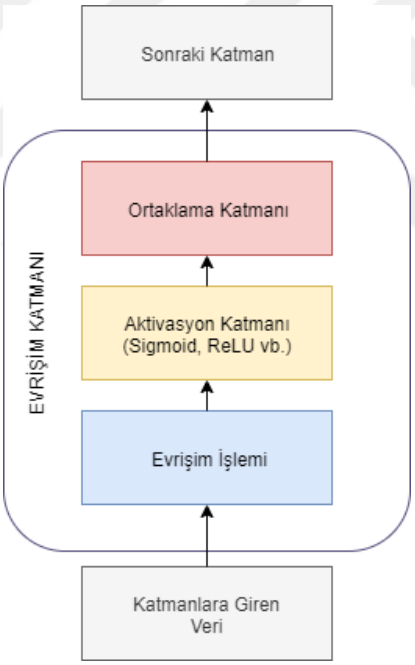
**Şekil 2.7** CNN'lerde öğrenilen özniteliklerin görselleştirmesi [136]

Gerçek dünyadaki görüntüler içindeki uzamsal hiyerarşiyi nesneler içindeki yerel örüntüleri tanımak konusunda CNN'ler oldukça tatmin edici çıktılar Şekil 2.8'dekine benzer olarak üretmektedir.




**Şekil 2.8** Görselden çıkarılan öznitelikle ve hesaplanan tahmin [144]

Bir CNN'in bileşenleri; evrişim işlemi ve bu işlemin çıkışına uygulanan doğrusal olmayan bir aktivasyon fonksiyonu ile devam etmektedir. Burada az da olsa bilgi kaybı yaratan boyut azaltma için ortaklama uygulanmaktadır. Bir evrişim katmanı terminolojisi Şekil 2.9'da gösterilmektedir [144].




**Şekil 2.9** Evrişim katmanı terminolojisi [136]

Ortaklama işlemi, değişmezlik özelliğinde ve doğrusal olmayan boyut azaltma sağlamaktadır. İşlem yükünü azaltmak temel amaçtır. Maksimum ortaklama fonksiyonu denklem (2.7)'de gösterilmektedir.

$$ap_{i,j}^{l} = arg \max_{i,j \in \Omega} \{ac_{i,j}^{l}\} \tag{2.7}$$

Burada Ω, görüntünün n×n boyutlu küçük bölgeleridir. Maksimum ortaklama, ortalama ortaklama yönteminden pratik olarak daha çok tercih edilmektedir. Maksimum ortaklama kullanılarak öznitelik haritası boyutunu yarıya indirmek mümkündür.

ReLU doğası gereği netagif değerler üretmediği için CNN modellerinde en sık kullanılan aktivasyon fonksiyonudur. Bu, ortaklama katmanının çıktısında kararsızlığa neden olabilir. Bu nedenle, veri dağılımını ayarlamak için bir Gauss normalizasyonu kullanmak gerekmektedir [136]. μ ve σ girdinin ortalama ve varyans değerleridir, γ ölçek parametresidir ve β kaydırma parametresi iken, m ve n giriş görüntüsünün yüksekliği ve genişliğidir. Tüm bu parametreler denklem (2.8)-(2.11)'de hesaplanmaktadır.

$$\mu \leftarrow \frac{1}{m*n}\sum_{i=1}^{m}\sum_{j=1}^{n} ap_{i,j}^{l} \tag{2.8}$$

$$\sigma_{\beta}^{2} \leftarrow \frac{1}{m*n}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(ap_{i,j}^{l}-\mu\right)^{2} \tag{2.9}$$

$$\hat{a}_{i,j}^{2} \leftarrow \frac{ap_{i,j}^{l}-\mu}{\sqrt{\sigma^{2}+\varepsilon}} \tag{2.10}$$

$$x_{i,j}^{l} \leftarrow \gamma \hat{a}_{i,j}^{l} + \beta \tag{2.11}$$

Tam bağlantılı katman, her sinir hücresinin bir önceki katmandaki her bir sinir hücresine bağlıdır. Softmax aktivasyon fonksiyonu genellikle ikiden daha fazla sınıflı problemlerde çıkış katmanında kullanılmaktadır. Elde edilen ileri seviye öznitelikler, tamamen bağlantılı bir katmana girdi olarak beslenmektedir. Tam bağlantılı katmanın fonksiyonu, sınıflandırmayı gerçekleştirmek için öznitlikleri yeniden yapılandırmaktır. Tam bağlı katmanın hesaplanması denklem (2.12)'de gösterilmektedir.

$$af_{j}^{FC_k} = relu\left(net_j\right) = relu\left(\sum_{i=1}^{l} w_{i,j}^{FC_k} * af_{i}^{FC_{k-1}} + b_{j}^{FC_k}\right) \tag{2.12}$$

Burada $af_{i}^{FC_k}$ k'inci tam bağlı katman çıktısıdır, $b_{j}^{FC_k}$ k'inci tam bağlı katmanın önyargı değeridir, $w_{i,j}^{FC_k}$ i. giriş sinir hücresi ile j. çıkış sinir hücresi arasındaki ağırlık, $l$ giriş sinir hücre sayısıdır. Çıktı katmanındaki sinir hücrelerinin sayısı aynı

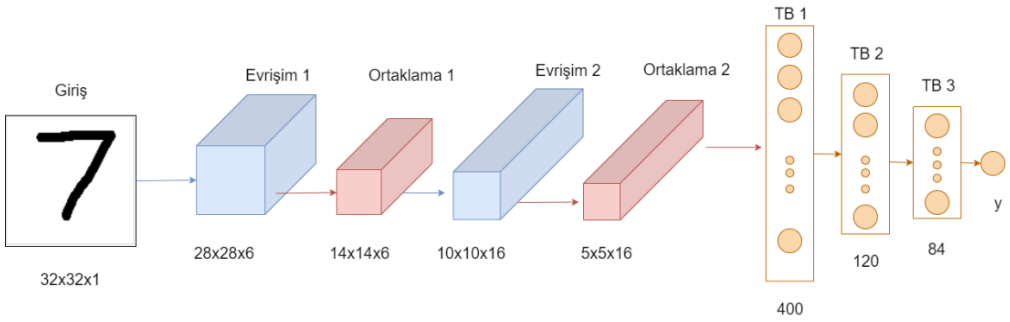
zamanda çok kategorili sınıflandırmanın kategori sayısını da belirlemektedir. Bu katmandaki bir sinir hücresinin çıktısı denlem (2.13) ve (2.14)'de gösterilmektedir.

$$o_j = softmax\left(h\left(af_j^{FC_L}\right)\right) \tag{2.13}$$

$$h(x_j) = \frac{1}{1 + e^{-\theta^T x}} \tag{2.14}$$

Burada $o_j$ sinir hücresi çıktısıdır ve $af_j^{FC_L}$ son tam bağlı katman çıktısıdır [136].

CNN modelleri AT&T şirketinde görevli bir grup araştırmacı tarafından ilk kez pratik bir uygulamada kullanılmıştır. Bu uygulama ile el yazısı rakamlarında tanınması amaçlanmış ve kullanılan model 2.10'da gösterilmektedir. Günümüzde derin öğrenme gelişimine devam eden popüler bir araştırma alanıdır.




**Şekil 2.10** LeNet CNN modeli [142]

Giriş görüntüsü olarak tek kanallı iki boyutlu gri seviyeli el yazısı rakamlar yer almaktadır [136]. Yükseklik ve genişlik bilgisi 32 olarak verilmektedir. İlk katmanda evrişim işleminin doğası gereği filtre sayısı kadar kanal oluşmakta ve yükseklik-genişlik azalmaktadır. Eğer yükseklik-genişlik bilgilerini sabit tutmak gerekirse evrişim işlemini uygulamadan önce giriş verisine piksel doldurma işlemi yapmak veya daha küçük boyut elde edilmesi gerekiyorsa adım kaydırma boyutunu büyük seçmek gerekmektedir. Ortaklama katmanında ise kanal sayısı sabit kalırken yükseklik ve genişlikte boyut azaltma yapılmaktadır. Bu işlem yükünü hafifleten ve giriş bilgisinden biraz kayıp yaratan matematiksel bir yaklaşımdır. Gruplanan değerlerin ortalaması alınarak, en büyüğü seçilerek ya da ortanca değeri seçilerek gerçekleştirilerek uygulanmaktadır.

Farklı veri tiplerinde uygulanabilen CNN'ler tıpkı diğer YSA mimarilerinde olduğu gibi başka birçok parametreye bağlı olarak çalışmaktadır. Filtre sayısı, aktivasyon fonksiyonu seçimi, optimizasyon algoritmasının seçimi, katman sayısı, verinin kalitesi ve yansızlığı, öneğitim gibi düzenlileştirme için dikkat edilmesi gereken parametreler vardır. Özellikle büyük veri kümelerinde eğitilmiş başarısı kanıtlanmış modelleri kullanarak çalışılması planlanan konu için transfer öğrenme de önemli yaklaşımlardandır. Genellikle kısıtlı veri ile çalışıldığında daha önceden eğitilmiş görüntülere ait temel öznitelikleri öğrenmiş olan ağırlıklar ile yeni bir probleme başlamak hız, işlem yükü ve doğruluk bakımından avantaj sağladığı bilinmektedir.

#### 2.1.3.2 Kapsül Ağları

CNN'lerin oryantasyon, örtüşme gibi durumlardaki yetersizliğine bir çözüm olarak CapsNet (kapsül ağları) ve dinamik yönlendirme algoritması (routing by agreement) önerilmektedir [145], [146]. Özellikle bilgisayarlı görü konularında açısal değerin değişmesi, duruş, yönelim ve konum gibi görüntü içindeki nesneye dair değişkenleri kapsül adı verilen bir grup sinir hücresi ile sağlanmaktadır. Nesneleri temsil eden kalınlık, ölçek gibi bilgileri ise anlaşmalı yönlendirme algoritması ile öğrenmesi önerilmektedir.

CapsNet, 2018 yılına kadar alışılagelinen sinir ağlarından farklı olarak kapsüllerden meydana gelecek şekilde tasarlandı ve bu model ilk olarak bilgisayarlı görü alanındaki bazı problemlere çözüm olarak öne sürüldü. Bir kapsül, görüntünün belirli bir bölgesinde berlirli bir nesneyi algılamayı öğrenmesi için küçük bir yapay sinir grubudur. Bu kapsülün uzunluğu, tahmin olasılığını temsil eden vektördür. Örneğin bu 8 boyutlu bir vektör ise nesneye dair poz yani kesin konum, döndürme, yön gibi parametrelerini temsil etmektedir. Nesnenin poz bilgilerindeki değişimler bu vektörlerin yön ve boyutlarında farklılaşarak yeniden hesaplanmaktadır. Aynı nesne için kapsülün uzunluğu değişmezken yönü değişmiş bir vektör elde edilir. Bu kapsüllerin eşdeğer olduğunu göstermektedir.

CapsNet'ler de CNN'ler gibi birden çok katmanlı şekilde tasarlanabilir. Giriş katmanındaki kapsüllere öncül kapsüller denir. Her bir öncül kapsül giriş görüntüsünün küçük bir bölgesini kullanmaktadır. Yönlendirme kapsülleri adı

verilen daha ileri katmanlardaki kapsüller, görüntü içindeki daha karmaşık ilişkileri algılamaktadır [145].

CapsNet'in $s_j$ girişi denklem (2.15)'de gösterilmektedir.

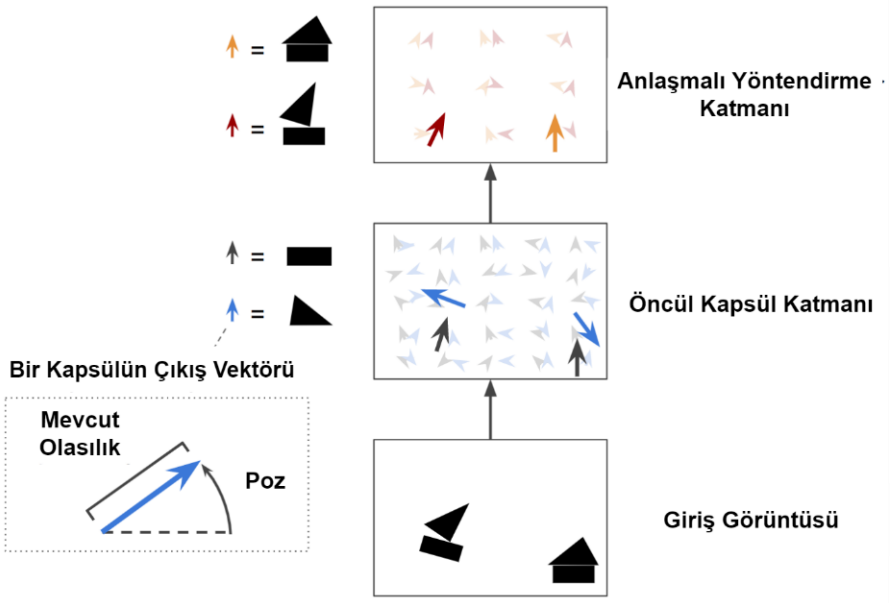
$$s_j = \sum_i c_{ij} \hat{p}_{j|i} \tag{2.15}$$

Bir kapsülde $s_j$ girdisinin toplam değeri, alt katmanın kapsülünden gelen tüm tahmin vektörleri $\hat{p}_{j|i}$'nin ağırlıklı bir toplamıdır. Buradaki $\hat{p}_i$, $W_{ij}$ ağırlık matrisi ile çarpılan alt katmanın çıktısıdır ve denklem (2.16)'deki gibi hesaplanmaktadır.

$$\hat{p}_{j|i} = W_{ij}.\hat{p}_i \tag{2.16}$$

$v_j$, $j$. kapsülün çıkış vektörü, $s_j$, $j$. kapsülün giriş vektörü, $\|s_j\|$ ise $s_j$ vektör uzunluğunun modüldür. Kapsül katmanının çıktı vektörünün uzunluğu, kapsül tarafından temsil edilen varlığın girdide mevcut olma olasılığını temsil etmektedir. Bu nedenle, kısa vektörün 0'a yakın bir uzunluğa sıkıştırılmasını ve uzun vektörün 1'den küçük ve 1'e yakın bir uzunluğa sabitlenmesini sağlamak için etkinleştirme fonksiyonu olarak doğrusal olmayan bir sıkıştırma fonksiyonu kullanılmaktadır.

İki katmanlı bir CapsNet ele alındığında Şekil 2.11'deki örnekte, birincil kapsül katmanı iki adet 5×5 kapsül haritasına sahipken, ikinci kapsül katmanı iki adet 3×3 kapsül haritasına sahiptir. Her kapsülden bir vektör elde edilmektedir. Şekil 2.11'deki her vektör, farklı bir kapsülün çıktısını temsil etmektedir. Mavi vektörler, üçgenleri algılamaya çalışan bir kapsülün çıktısını, siyah vektörler ise dikdörtgenleri algılamaya çalışan bir kapsülün çıktısını temsil etmektedir [146].




**Şekil 2.11** CapsNet anlaşmalı yönlendirme stratejisi [146]

Öncül kapsül katmanı, birkaç düzenli evrişimli katman kullanılarak oluşturulmaktadır. Orijinal çalışmada iki evrişim katmanı uygulandıktan sonra 8 boyutlu vektörler içeren 32 tane 6×6 öznitelik haritası elde edilmekte ve bunlar yeniden şekillendirilmektedir. Bu vektörlerin 0 ile 1 arasında bir uzunluğa sahip olmasını sağlamak için yeni bir sıkıştırma fonksiyonu kullanılmaktadır. Bu fonksiyon olasılık bilgisini barındırmakta ve öncül kapsül çıktısını vermektedir.

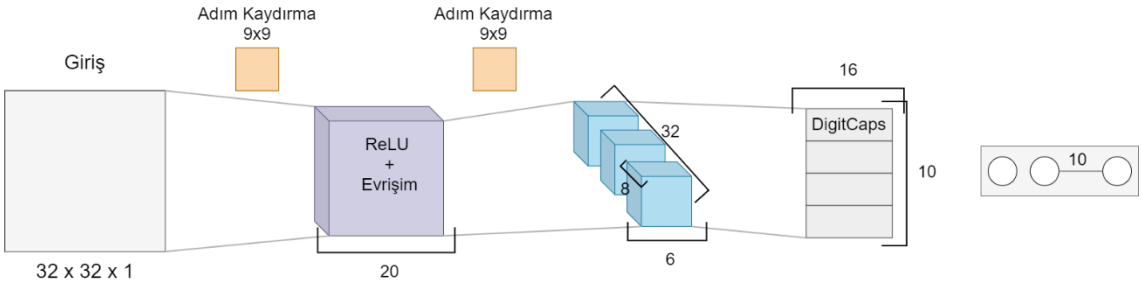
Sıkıştırma fonksiyon denklemi (2.17) gibi değiştirilmektedir [145].

$$v_j = Squash(s_j) = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \tag{2.17}$$

Öncül kapsül katmanı için başlangıç değeri 0 olarak seçilen $b_{ij}$ denklem (2.18)'deki gibi hesaplanmalıdır. Öncül kapsül katmanı çıktısı $c_{ij}$ birleştirme katsayısı denklem (2.19) olarak tanımlanmakta ve yinelemeli bir dinamik yönlendirme işlemi yapılmaktadır. Şekil 2.12'de basitleştirilmiş bir CapsNet gösterilmektedir.

$$b_{ij} + \hat{p}_{j|i}.v_j \Longrightarrow b_{ij} \tag{2.18}$$

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k exp\,(b_{ik})} \tag{2.19}$$




**Şekil 2.12** Basitleştirilmiş CapsNet gösterimi [146]

CapsNet ile klasik CNN'ler tarafından çözülemeyen; konum, yönelim, duruş ve açısal değerin değişmesi durumunda dahi bir grup sinir hücresinden oluşan kapsüller aracılığıyla nesnenin başarıyla tanınabilmesi için nesneyi temsil eden kalınlık, ölçek, kaydırma vb. özelliklerin anlaşmalı yönlendirme ile öğrenilmesi önerilmiştir [146].

## 2.2 Veri Kümeleri

Kalabalık analizi için çeşitli amaçlarla ve farklı alanlardan toplanan video ve statik görüntülerin yer aldığı açık veri kümeleri bulmak mümkündür. Özellikle kişisel verilerin korunması kanunu pek çok ülkede yeni veri kümelerinin çalışmalar işçin toplanmasını zorlaştırmaktadır. Bu yüzden bu çalışma kapsamında açık veri kümelerinden faydalanılmaktadır. Bunun iyi yanı ise literatürdeki başarılı çalışmalarla karşılaştıma yapabilme imkânı sunmasıdır. Veri kümeleri, kişi sayısını tahmin etme, yoğunluk tahmini, konum belirleme, davranış analizi gibi görevler için dış mekân ve iç mekân sahnelerinden meydana gelmektedir. Uygulamaların performansını test etmek ve diğer çalışmalarla karşılaştırmak için kullanılmaktadır. Kalabalık analizi çalışmaları için kullanılan veri kümeleri Tablo 2.1'de gösterilmektedir.

Bu çalışmanın farklı aşamalarında tablodaki veri kümelerinden UCF-QNRF, UCF_CC_50, UCSD, ShanghaiTech Part A, WorldExpo'10 ve PETS2009 veri kümeleri kullanıldı. Bu veri kümelerine yeniden boyutlandırma işlemi uygulandı. Bu şekilde önerilen modellere uygun girişler haline getirildi. Böylece farklı çözürnürlüklere sahip olan verilerden olumsuz etkilenmesinin önüne geçilmeye çalışıldı.

Tablo 2.1 yer alan MALL ve CUHK veri kümeleri ise kalabalık analizinde sıklıkla karşılaşılan ancak bu çalışmanın hiçbir aşamasında tercih edilmemektedir. Bunun

en önemli sebebi MALL veri kümesinin yalnızca iç mekân görüntülerinden ve CUHK veri kümesinin ise belirgin davranışlar için hazırlanmış olmasıdır.

**Tablo 2.1** Kalabalık analizi için hazırlanan veri kümeleri

| Veri Kümeleri | Tanımalama | Görüntü sayısı | Çözünürlük | Ort. | Toplam | Erişilebilirliik |
|---|---|---|---|---|---|---|
| **UCF-QNRF [147]** | Kişi Sayma ve Konum Belirleme | 1535 | Variable | 815 | - | Evet |
| **MALL [148]** | Kişi Sayma | 2000 | 320 × 240 | - | 62325 | Evet |
| **UCF_CC_50 [149]** | Yoğunluk Kestirimi | 50 | Variable | 1279 | 63974 | Evet |
| **UCSD [32]** | Kişi Sayma | 2000 | 238 × 158 | 25 | 49885 | Evet |
| **Shanghai Tech Part A, Part B [82]** | Kalabalıkta Kişi Sayma | 482 | 768×1024 | 501 | 241677 | Evet |
| **WorldExpo'10 [81]** | Kalabalıkta Kişi Sayma | 3980 | 576 × 720 | 50 | 199923 | Evet |
| **CUHK [64]** | Kalabalık Davranışı | 716 | 400×300<br>9000×6000 | 123 | 88488 | Evet |
| **PETS2009 [150]** | Kalabalık Davranışı | 728 | - | - | - | Evet |

Bu yüzden bu çalışmanın kapsamı dışında kalmaktalardır. Kullanılan veri kümelerinin temel özellikleri şöyle sıralanabilir:

- **UCF-QNRF** veri kümesi, kalabalık sayımı ve konum belirleme için hazırlanan diğerlerine kıyasla daha güncel bir veri kümesidir. Yüksek yoğunluklu kalabalık görüntülerine, etiketlerine ve çok çeşitli perspektif, yoğunluklar ve ışık varyasyonlarını içeren geniş bir sahne yelpazesine sahip olması nedeniyle zorludur. Ortalama yoğunluk, yani tüm görüntülerde piksel başına düşen kişi sayısı düşüktür, bu da yüksek kaliteli büyük görüntüleri ifade etmektedir. Piksel başına daha düşük yoğunluk, kısmen birçok yüksek yoğunluklu bölgenin yanı sıra sıfır yoğunluklu bölgelerin bulunduğu arka plan bölgelerinin dâhil edilmesinden kaynaklanmakatadır. Veri kümesi

doğada yakalanan gerçekçi senaryolar mevcut olduğu için binalar, bitki örtüsü, gökyüzü ve yolları içermektedir [147].

- **MALL** veri kümesi, bir alışveriş merkezindeki güvenlik kamerasından toplanan görüntülerden meydana gelmektedir. Genellikle iç mekânlarda insanları saymak için kullanılmaktadır. Zorlu aydınlatma koşullarına ve cam yüzey yansımalarına sahip, seyrek ve kalabalık arasında çeşitli kalabalık yoğunluklarının yanı sıra günün farklı zamanlarında farklı aktivite modellerini kapsamaktadır [148].
- **UCF_CC_50**, stadyumlar, meydanlar ve konser salonlarındaki gerçek organizasyon anları ve farklı yoğunluktaki kalabalıkların toplanma ve dağılma süreleri hakkında bilgi içeren yoğunluk haritalaması için kullanılan zorlu bir veri kümesidir [149].
- **UCSD** veri kümesi, insan sayma problemi için kamuya açık yerler olan sokaklardan alınan görüntülerden meydana gelmektedir. Bu görüntüler yaya yolu ile sınırlı farklı uzaklık ve konumlardaki kişileri içermektedir [32].
- **ShanghaiTech** veri kümesi, A ve B olarak adlandırılan iki bölümden meydana gelmektedir. Part A, internetten rastgele alınan görüntülerden ve Part B, Şangay'daki bir metropol caddesinden toplanan görüntüleri kapsamaktadır. Farklı kamera perspektiflerine ve aydınlatmalara sahip yüksek yoğunluklu görüntüler içermektedir [82].
- **WorldExpo'10**, sahneler arası kalabalık sayımı için hazırlanmış108 farklı kameradan alınan ve kuş bakışı açıyıda kapsayan bir veri kümesidir. Az hareketli büyük grupların varlığı görüntüdeki ön plan ve arka plan ayrımını yapmayı zorlaştırmaktadır. Ancak düşük yoğunluğu nedeniyle doğruluk değerlendirmesi için kullanımı yetersizdir [81].
- **CUHK**, havaalanları, alışveriş merkezleri, parklar ve caddelerden elde edilen davranışsal analiz için yayınlanan bir veri kümesidir. Grup algılama, grup durumu ve davranış analizi ve kalabalık video sınıflandırması amacıyla yayınlanmıştır [64].
- **PETS2009**, farklı kitle aktiviteleri içeren bir veri kümesidir. Veri kümesindeki görüntüler farklı kamera görünümlerine sahiptir. Veri kümesinin 3 ana başlık altında 8 alt kümesi vardır [150].




**Şekil 2.13** PETS2009 veri kümesi hiyerarşik yapısı [150]

Veri kümesindeki görüntülerin hiyerarşik yapısı Şekil 2.13'de gösterilmektedir. Diğer veri kümelerinden farklı olması dolayısıyla bu çalışmada kullanılan kısmına dair detaylar Tablo 2.2'de gösterilmektedir.

**Tablo 2.2** PETS2009 veri kümesi detayları [150]

| **S1: İnsan sayma ve yopunluk kestirimi** | | **S2: İnsan Takibi** | | **S3: Çoklu akış / Olay tanıma** | |
|---|---|---|---|---|---|
| **Küme adı** | **İçerik** | **Küme adı** | **İçerik** | **Küme adı** | **İçerik** |
| S1.L1 yürüme | Orta yoğunluklu kalabalık, kapalı hava | S2.L1 yürüme | Seyrek kalabalık | S3 Çoklu akış | Yoğun kalabalık, koşma |
| S1.L2 yürüme | Yüksek yoğunluklu kalabalık, kapalı hava | S2.L2 yürüme | Orta yoğunluklu kalabalık | S3 Olay tanıma | Yoğun kalabalık |
| S1.L3 koşma | Orta yoğunluklu kalabalık, parlak güneş ışığı ve gölgeler | S2.L3 yürüme | Yoğun kalabalık | | |

Veri kümesinin S3 adlı bölümü, 14-16, 14-27, 14-31 ve 14-33 zaman damgalarına sahip dört farklı diziden oluşmaktadır. Çalışmanın bu aşaması için S3, çoklu akış analizi ve olay tanıma alt veri kümeleri kullanıldı.

Bu videolar, koşma ve yürüyüşün yanı sıra çeşitli kalabalık davranışlarını içermektedir. Kullanılan veri kümesinin detayları Tablo 2.3'de sunulmaktadır.

**Tablo 2.3** PETS2009 veri kümesinde davranış analizi için hazırlanan video dizilerinin özeti [16]

| Davranış | Video | Sahne | Dizi |
|---|---|---|---|
| Yürüme/Koşma | S3-HL, 14-16, View-001-p1 | 107 | Seq1 |
| Yürüme/Koşma | S3-HL, 14-16, View-001-p2 | 114 | Seq2 |
| Yürüme/Koşma | S3-HL, 14-33, View-001 | 377 | Seq3 |
| Yürüme/Koşma | S3-HL, 14-31, View-001 | 130 | Seq5 |

## 2.3 Yöntemler

Kalabalık analizine ilişkin detaylı literatür araştırması bu tezin başlangıç aşamasından son anına kadar aralıklı olarak sürdürüldü. Böylece bilimsel dünyanın ilerleyiş yönü sebep-sonuç ve olumlu-olumsuz yönleri değerlendirildi. Bu sayede, kalabalık analizi için gerekeli diğer ilişkili alanlarda literatür ve yeni yöntemlerde takip edilerek probleme özgün yaklaşımlar geliştirildi. Probleme ilişkin yeni veri kümesi oluşturmak üzerinde düşünüldü. Ancak gerek kamuya açık alanlarda kamera kurulumu yapılması gerekse ilgili bölgelerden geçecek insanların bu tip bir çalışmada görüntülerinin bulunacağı bilgisinin verilmesi gibi zorluklar sebebiyle açık veri kümeleri araştırıldı. Hâlihazırda yayınlanmış olan açık veri kümelerinin çeşitliliği ve sayısı bu çalışma için uygun olduğu ve bu sayede farklı çalışmaların performanslarıyla da karşılaştırma yapılmasına olanak tanıdı. Kalabalık analizi için sıklıkla kullanılan 6 veri kümesi çalışmanın belli aşamalarında birlikte ya da ayrı ayrı olarak ele alındı. Tez kapsamında geliştirilen yöntemler çalışmanın bu aşamasında detaylandırılmaktadır. Kalabalık yoğunluk kestirimine odaklanan bu çalışmada kalabalık davranış değişimi de ele alınmaktadır.

### 2.3.1 Yüksel Yoğunluklu Kalabalık için Derin ve Geniş CNN Modeli

Literatürde incelemesinde, kalabalık analizi için sınıflandırma ve regresyon temelli CNN mimarileri vardır. Çok sayıda bilimsel çalışma evrişim katmanları ardından ortaklama ve son olarak tam bağlı katmanı kullanmaktadır. Ancak, görüntü ve video

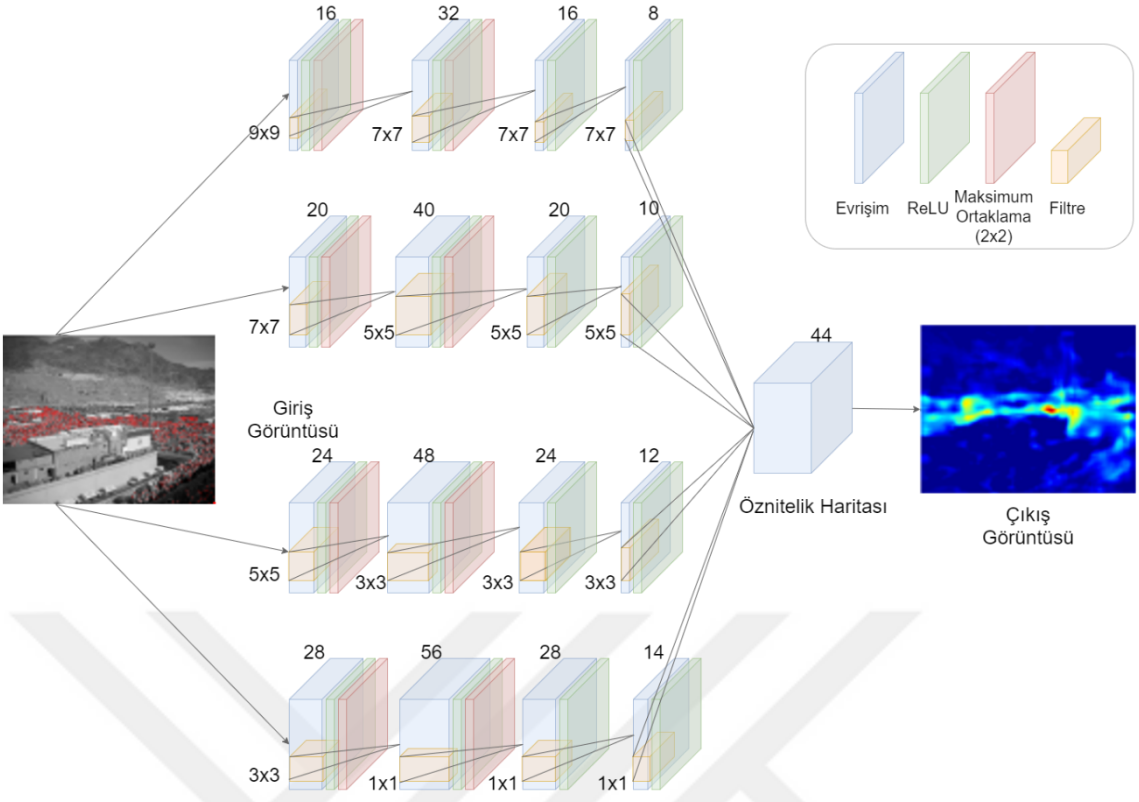
sınıflandırma sorunları için standart yegâne yapı yoktur. Dolayısıyla, sınıflandırma görevleri için en iyi CNN mimarisini tasarlamak henüz çözülmemiş bir sorundur.

Çalışmanın bu aşamasında, çok kalabalık bir veri kümesi üzerinde literatüre yeni bir model tasarlanarak başarılı bir kalabalık yoğunluğu analizi yapıldı. Standart yaklaşımlar, özellikle aşırı kalabalık görüntülerde insan sayısını bulmada tatmin edici sonuçlar elde etmek için mücadele etmektedir. Böylece probleme özgü derin ve geniş CNN modeli ile etkili sonuçlar elde edildi [15].

#### 2.3.1.1 Paralel Derin Ağ Modülü

Kalabalık sahnelerden anlamsal bilgi elde etmek çok zor bir problemdir. Çünkü görüntüler genellikle uzak bir mesafeden elde edilmekte ve çok çeşitli perspektif ve ışıklandırma bilgisine sahiptir. Bu nedenle bir kişinin baş veya vücut bölgelerini belirlemek çok zor bir iş haline gelmektedir [151]. Çalışmanın bu aşamasında, kalabalık görüntülere önce kalabalığın yerini belirlemek için iki paralel model uygulandı. Daha sonra ikinci bir paralel model kullanılarak insanların baş ve vücut kısımları belirlendi. Bunları birleştirerek bir yoğunluk haritası elde edildi. Evrişim ve maksimum ortaklama işlemleri, önerilen ağdaki farklı filtre boyutlarına paralel olarak gerçekleştirildi.

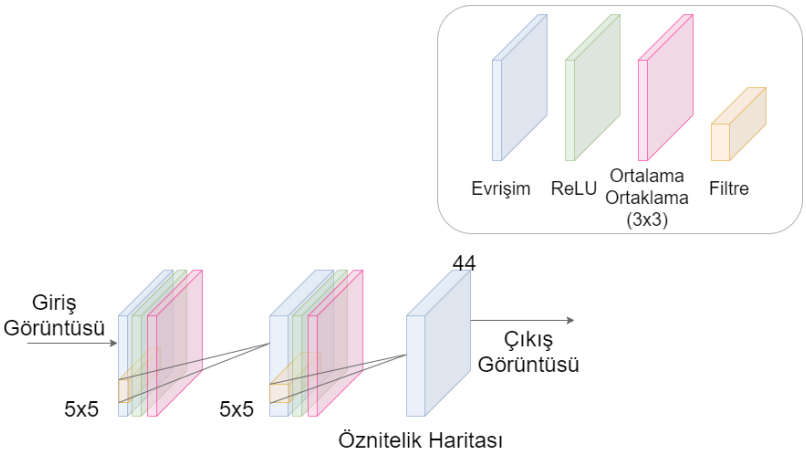
Önerilen kalabalık yoğunluk kestirimi yöntemi, diğer çalışmalardan farklı olarak özniteliklerin görüntüden farklı detaylarda hesaplanmasına ve böylece kişilerin baş ve vücut sayılarının bakış açısı etkisini azaltarak daha etkin hesaplanmasına yardımcı olmaktadır. Maksimum ortaklama filte boyutu (2×2) ve dört farklı (3×3, 5×5, 7×7, 9×9) boyutlu filtreli evrişim katmanları kullanılarak dört paralel modülden oluşan bir model oluşturuldu. Hesaplama maliyeti nedeniyle, bu tür görüntüler için küçük bir maksimum ortaklama filtre boyutu seçmek önemlidir. ReLU, sadece daha az hesaplama karmaşıklığı için değil, aynı zamanda kaybolan gradyan probleminin sağlamlığı için de aktivasyon fonksiyonu olarak kullanılır [144]. Tasarlanan model Şekil 2.14'de gösterilmektedir.




**Şekil 2.14** Paralel derin ağ modülü [15]

### 2.3.1.2 Sığ Ağ Modülü

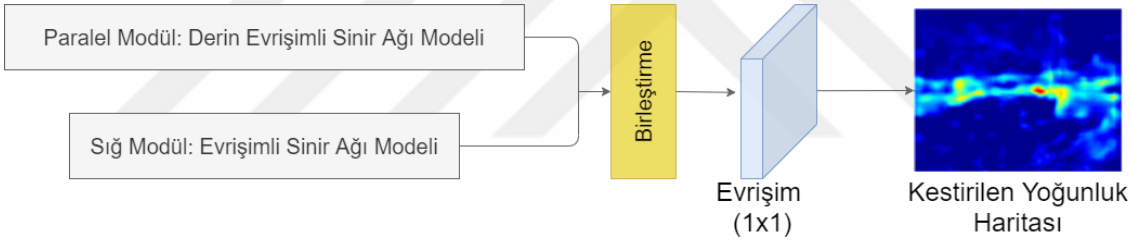
Boominathan vd. [91]'de sığ bir ağ ile derin bir sinir ağının birleştirilmesiyle tüm sistemin yoğun şekilde kalabalık görüntülerle daha başarılı hale geldiği bildirilmektedir. Üstelik çok katmanlı yapıların birleştirilmesi ile daha az katmanlı modeller sistem hızını arttırmaktadır. Bu modülün sığ olarak adlandırılmasının sebebi, yalnızca dört ardışık katmandan meydana gelmesidir. Evrişim katmanlarında, boyutları (5×5) olan filtreler ve ardından ortalama ortaklama işlemleri sırayla eklenen modül Şekil 2.15'de gösterilmektedir.




**Şekil 2.15** Sığ ağ modülü [15]

### 2.3.1.3 Paralel Modül Derin Ağ Modülü ve Sığ Ağ Modülünün Kombinasyonu

İki modülün bir (1×1) evrişim katmanı ile birbirine bağlandığı birleştirme modüldür. Tasarlanan model DWCNN (Deep and Wide Convolutional Neural Network) olarak adlandırıldı ve Şekil 2.16'de gösterildiği gibidir. Bu modelin çıktısı, kalabalık görüntünün tahmini yoğunluk haritasıdır.




**Şekil 2.16** Derin ve geniş CNN Modeli [15]

Elde edilen yoğunluk haritası ile her bir görüntü için gerçek değer arasında Öklid mesafesi (L2 normu) hesaplamaktadır. Kayıp fonksiyonu denklem (2.20) ile gösterilmektedir.

$$L(\theta) = \frac{1}{2N}\sum_{i=1}^{N} \|F(X_i;\theta) - F_i\|_2^2 \qquad (2.20)$$

Burada θ ağ tarafından öğrenilen parametreleri belirmekredir. N eğitim kümesindeki görüntülerin sayısıdır. $X_i$ giriş görüntüsüdür ve $F_i$ gerçek değerler yoğunluk haritasıdır. Öngörülen yoğunluk haritası $F(X_i;\theta)$ ile ifade edilmektedir.

#### 2.3.1.4 Düzenlileştirme

Düzenlileştirme, sistemin performansını iyileştirmek ve derin öğrenmede aşırı öğrenmeyi önlemek için modelden bağımsız bir yöntemdir. Eğitim kümesinde yeterli veri yoksa modelin fazla öğrenme ile beklenen başarısı azalabilmektedir. Literatürde farklı düzenlileştirme yöntemleri vardır. Çalışmanın bu noktasında transfer öğrenme yöntemi kullanılarak kısıtlı veri kümesi probleminin önüne geçilmesi hedeflendi. Model, özellikle ilk katmanda kenarlar ve şekiller gibi alt seviye öznitelikleri öğrenmektedir.

Yoğun şekilde kalabalık sahnelerdeki insan sayısını tahmin etmek için özgün yaklaşımlı bir derin sinir ağı modeli önerildi [15]. Önerilen ağ, ShanghaiTech PartA olan popüler bir kalabalık sahne veri kümesi üzerinde test edildi. DWCNN'nin performansını iyileştirmek için, ağın sığ kısmına transfer öğrenme yöntemi uygulandı. Önerilen yöntem özellikle güvenlik eksikliğinin önlenmesi ve çevrenin düzenlenmesi için önemlidir. Performans karşılaştırmaları Sonuç ve Öneriler bölümünde detaylandırılmaktadır.

### 2.3.2 Kalabalık Sayımı için Küme Temelli İzleme ve Konum Tahmini

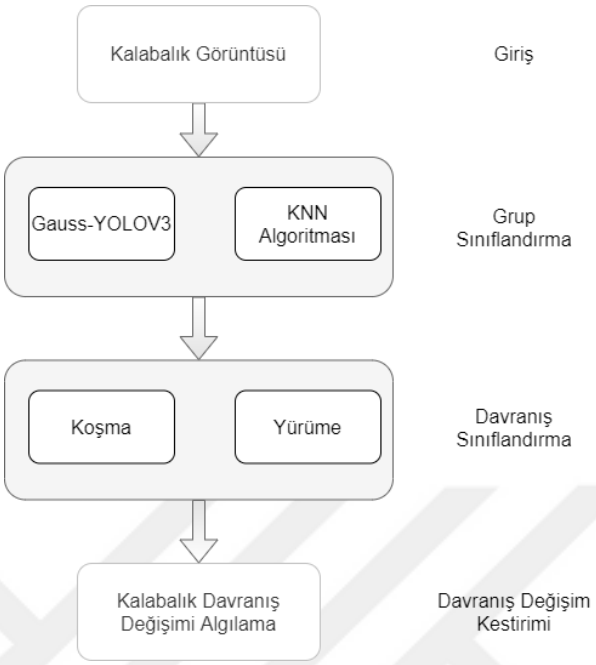
Videolar, ardışık görüntülerin dizi olarak ifade edilmesidir. Bu görüntülerde çeşitli nesneler ve insanlar yer almakta ve zamanla konum değiştirmektedir. Kalabalık izleme, sahneler içindeki insan gruplarının hareketini, hızını ve davranışını analiz eden bir kalabalık analizi alt alanıdır.

Çalışmanın bu aşamasında video kareleri arasındaki ilişkiden yola çıkarak kişi ve grupları belirlemek ve bunların davranış değişimlerini kestirmek hedeflendi. Sahnelerdeki kişiler, güncel derin öğrenme yöntemlerinden biri olan YOLOv3 kullanılarak takip edilmesi öngörüldü [152]. Görüntüde farklı gruplardan kişiler olduğunda grupların tespiti için KNN kümeleme algoritması kullanıldı [153]. Bu işlemlerden sonra yürüme ve koşma davranışları tahmin edildiği yeni bir yaklaşım önerildi [16]. Problem Şekil 2.17'deki hiyerarşik yapı ile ele alınmaktadır.

Yürüme ve koşma eylemleri şu şekilde tanımlanabilir:

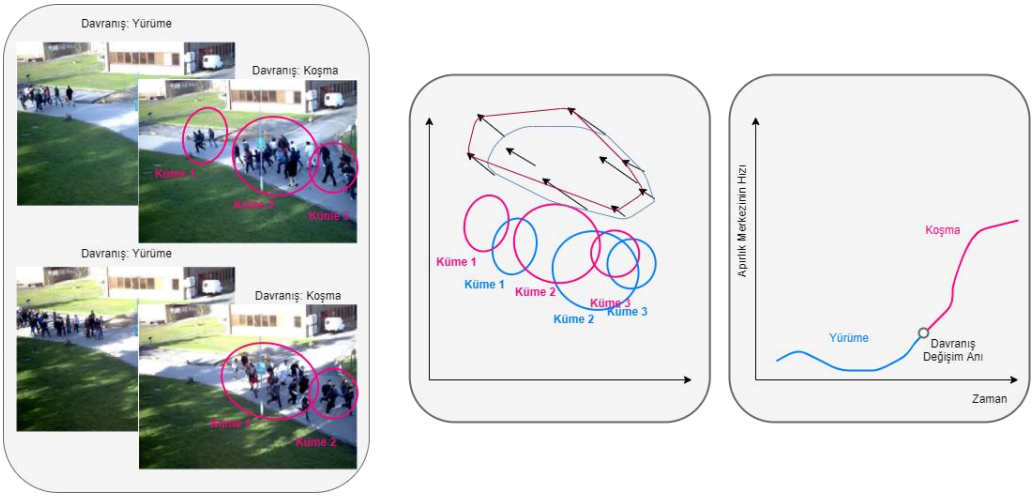
- Yürüme eylemi, sahne içindeki bir grup insanın ortalama hızdaki pozisyonundaki değişikliği göstermesidir.

- Koşma eylem, yürüme hareketine benzer, ancak sahneler arasında daha hızlı değişmesidir.




**Şekil 2.17** Problem ifadesi ve çözüm hiyerarşisi [16]

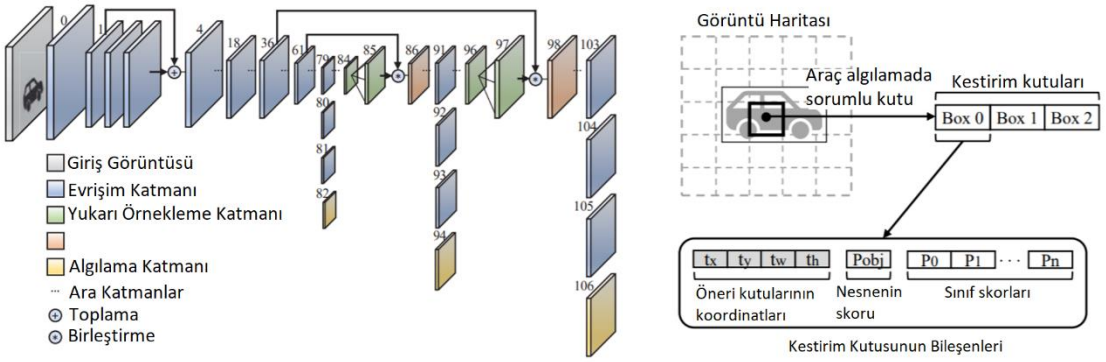
Yürüme ve koşma hareketlerini belirlemek için görüntüde tespit edilen her kümenin ağırlık merkezi hızları zaman serileri olarak hesaplanmaktadır. Şekil 2.18'de gösterilen hız değişim eğrisi kullanılarak görüntüdeki davranış değişim süreleri de belirlenmektedir.




**Şekil 2.18** Küme temelli kalabalık izleme yaklaşımı [16]

### 2.3.2.1 Gauss-YOLOv3

YOLO derin öğrenme mimarileri, nesneleri ve görüntüdeki konumlarını gerçek zamanlı olarak tespit etmek için tasarlandı. Öznitelik haritası tahminleri, belirtilen koordinatlara sahip bir sınırlayıcı kutuyla gösterilmektedir. YOLOv3, görüntüde bir nesne olup olmadığını hesaplayabilmektedir. Gauss YOLOv3, sınırlayıcı kutuların nesne kapsama alanında kesin bir iyileştirme sağlayan yeni bir surumudur. Şekil 2.19, Gauss YOLOv3 model mimarisini ve özellik haritası yapısını göstermektedir [152].



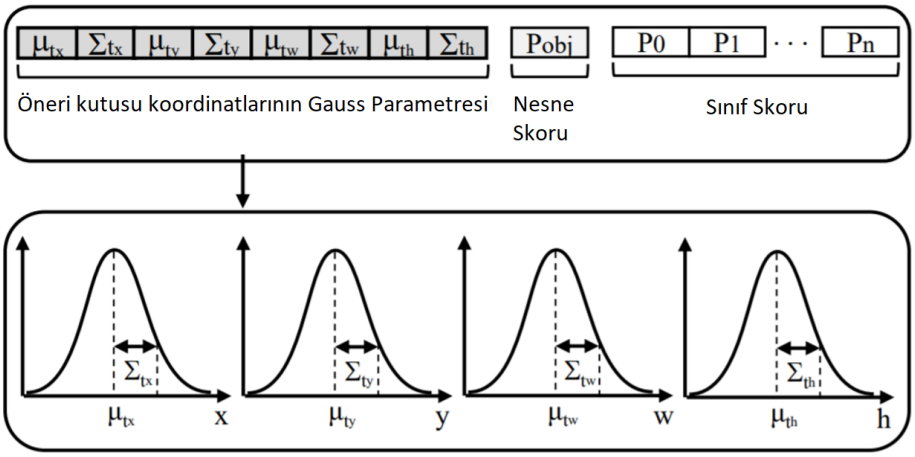

**Şekil 2.19** Gauss YOLOv3 derin öğrenme modeli mimarisi [152]

Bu modelde, sınırlayıcı kutu koordinatları regresyon kullanılarak hesaplanmaktadır. Bu şekilde konum belirleme yapmak mümkün olmaktadır. Ancak, her zaman belirsizlik sorunu vardır. Bu problem Gauss modeli kullanılarak aşılmaktadır. Bir x test girdisi için Gauss parametrelerinden oluşan y çıkışı denklem (2.21) gibi ifade edilebilir [152]:

$$p(y|x) = N(y; \mu(x), \Sigma(x)) \tag{2.21}$$

$\mu(x)$ ve $\Sigma(x)$ ayrı ayrı ortalama ve varyans fonksiyonlarıdır.

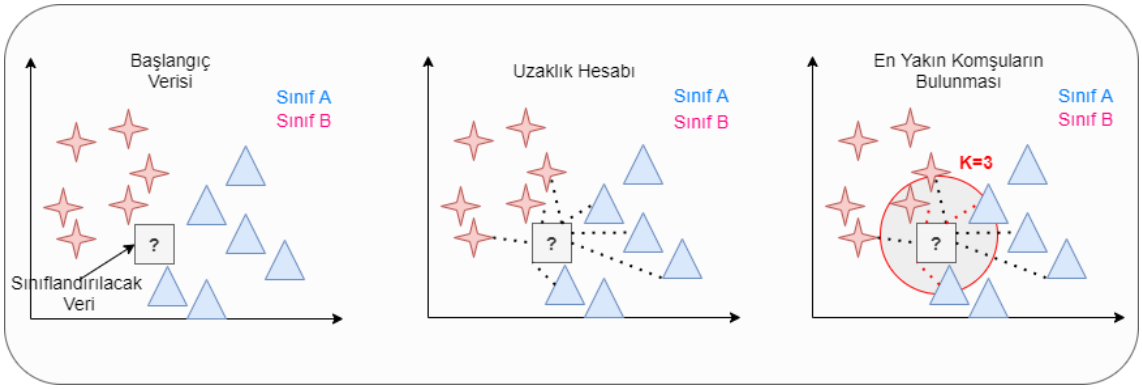
Sınırlayıcı kutunun belirsizliği, tahmin öznitelik haritasındaki sınırlayıcı kutu koordinatlarının her biri, Şekil 2.20'de gösterildiği gibi tahmin için ortalama (μ) ve varyans ($\Sigma$) olarak modellenmektedir.




**Şekil 2.20** Önerilen algoritmanın tahmin kutusundaki öğeler [152]

### 2.3.2.2 KNN Kümeleme

K-en yakın komşu (KNN) algoritması, etiketli noktaları ve mesafe bilgilerini kullanarak etiketlenmemiş noktaları sınıflandırmak için eski ama başarılı bir yaklaşımın kullanıldığı sağlam bir sınıflandırıcı modelidir. KNN, hızlı ve verimli bir algoritma olup gerçek zamanlı çalışmalarda tercih edilmektedir. Yeni etiketler atanırken Manhattan, Öklid ve Minkowski gibi farklı mesafe ölçütleri kullanılabilir ve algoritma komşu sayısında "k" olarak belirtilen bilgilere göre hesaplama yapmaktadır [153]. Çalışmanın bu aşamasında, sahnelerdeki kalabalık gruplarının tespiti ve birbirlerine göre hız ve yönlerinin incelenmesi önerilmektedir. Şekil 2.21'de temel KNN algoritmasının gösterilmektedir.




**Şekil 2.21** KNN algoritmasının temel gösterimi [16]

Kalabalık izleme yaklaşımları arasında, özellikle davranış değişikliği ve insan sayısını tahmin etme konusunda sıklıkla kullanılan PETS2009 veri kümesinde performans değerlendirmesi yapıldı. Bu veri kümesindeki 4 video dizisi için

performansı, güncel bir gerçek zamanlı derin öğrenme modeli olan YOLOv3'ün özel bir sürümü olan Gauss-YOLOv3 kullanılarak değerlendirilir. Etkili olduğu bilinen KNN kümeleme yöntemi, kalabalıktaki grupları belirlemek için de kullanılmaktadır.

Daha büyük ve daha kalabalık veri kümelerinde farklı kümeleme yaklaşımları ile yürüme ve koşmanın yanı sıra birleştirme ve bölme gibi davranışların tanınması üzerine çalışılabilir. Toplum ve insan güvenliği açısından kritik önem taşıyan kitle izleme araştırmalarının, bugün devam eden ve tüm dünyayı etkileyen Covid-19 salgınının kontrolünde kullanılması hedeflenebilir. Çalışmanın performans değerlendirmelerine tezin Sonuç ve Öneriler bölümünde yer verilmektedir.

### 2.3.3 Dikkat Temelli Kapsül Ağı ve İki Sütunlu CNN Kullanarak Kalabalık Yoğunluğu Tahmini

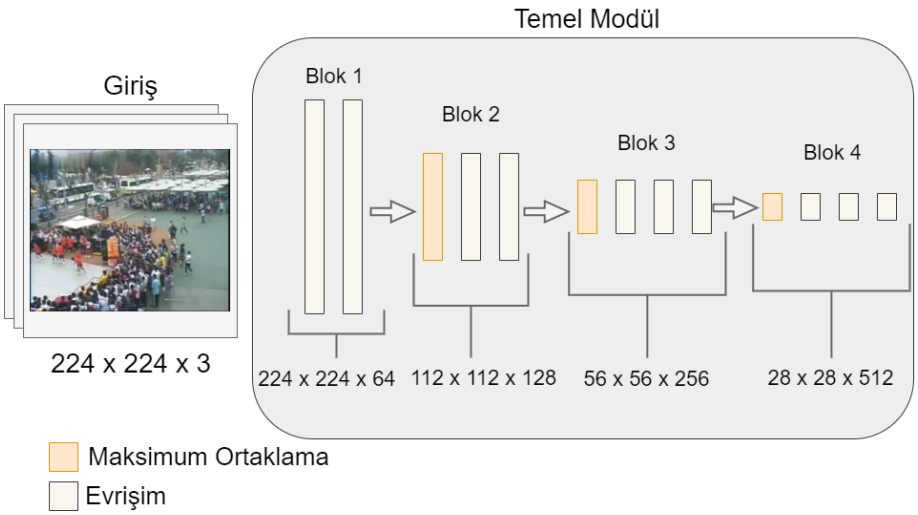
Kalabalık analizinde, sağlam bir yoğunluk haritası oluşturmak için yenilikçi bir yaklaşım önerilmektedir. CNN'leri ve CapsNet'i bir dikkat mekanizmasıyla birlikte kullanarak, kalabalık tahmini için daha verimli bir yoğunluk haritası elde edilmektedir. CNN'lerin kalabalık analizindeki başarısı kabul edilmektedir. Bununla birlikte, özellikle sürecin karmaşıklığı ve bu yaklaşımın henüz çok yeni olması nedeniyle bu alanda CapsNet uygulamasıyla ilgili literatürde sınırlı kaynaklar bulunmaktadır. Nguyen vd. [154], statik ve video görüntülerde saldırı türlerini belirlerken CNN'e ek olarak CapsNet kullandı. Algamdi vd. [155], açık hareket bilgisi olmadan videolardan kareler kullanarak insan eylemini tanımak için bir CapsNet mimarisi kullandı.

Önerilen yöntemde, temel model olarak önceden eğitilmiş VGG-16 modeli kullanılmaktadır. Bu model, VGG-16'nın 4. bloğundan sonra iki sütuna bölünmektedir. Böylece iki seviyeli bir özellik haritası elde edilmektedir. İkinci sütunda, iki farklı modül vardır, evrişim ve kapsül dikkat mekanizması modülleri sahnelerdeki varyasyon alanını genişletmeyi öğrenmektedir. Bu varyasyonlar, yoğunluk, seyreklik ve kalabalığın yönünü içermektedir. Bakış açısıyla değişmeyen bilgi, bir parça ile bütün arasındaki içsel uzamsal ilişkiyi kodlamayı öğrenmiş olan CapsNet dönüşüm matrislerinden kaynaklanmaktaıdr ve bu bilgi otomatik olarak yeni bakış açılarına genelleştirilebilir. Afin dönüşümlere karşı sağlam olan CapsNet kullanılarak, kalabalık yoğunluğu tahmini için daha kararlı sonuçların elde edilmesi

bekleniyordu. Bu amaç için önerilen mimari CNN'in nesne tanımadaki başarısından ve hızından yararlanırken, CapsNet'in dinamik yönlendirme algoritması, bir seviyedeki her bir kapsülün düşük seviyeli bazı aktif kapsüllere katılmasına ve diğerlerini göz ardı etmesine izin veren paralel bir dikkat mekanizması olarak kullanılmaktadır [145], [147]. Anlaşmalı yönlendirme algoritması, sahnedeki bireylerin segmentasyonunu desteklemek için bir kalabalığın şeklinin algılanmasını desteklemektedir. Piksel alanında daha yüksek seviyeli kararlar alma ihtiyacını ortadan kaldırmaktadır. Bu, önerilen modelin, insanlar çoğunlukla örtüşse bile kalabalık görüntüsündeki birden fazla kişiyi tanımasına izin vermesi sağlanmaktadır. CNN'leri ve CapsNet dikkat yaklaşımını birleştirerek, CNN modelinin örtüşme sebebiyle katlanarak artması beklenen verimsizliğinin önlenmesi başarılmaktadır. Son aşamada, temel model ve her iki dikkat modülünün birleştirilmesiyle elde edilen öznitelik haritaları kullanılarak bir yoğunluk haritası oluşturulmaktadır. Mekânsal ve zamansal bilgileri hesaplayabilen CapsNet dikkat mekanizmasının CNN'ler ile birlikte uygulanması ve orijinal amaçlarının ötesine geçilerek kalabalık analizi için kullanılması önerilmektedir.

#### 2.3.3.1 Temel Modül

İlk 4 evrişim, içinde blokların kullanıldığı ve tamamen bağlı katmanların bulunduğu önceden eğitilmiş bir VGG-16 ağını kullanılmaktadır [156]. VGG-16 ağındaki tüm evrişimli katmanların arkasına bir küme normalizasyonu (Batch Normalization) katmanı eklenmektedir. Temel model 10 evrişimli katmandan oluşmaktadır. İlk evrişim bloğu haricinde, diğer tüm katmanlarda maksimum ortaklama uygulanmaktadır. Giriş görüntüleri 224×224×3 olarak yeniden boyutlandırılmaktadır. Evrişim işlemleri sonucunda elde edilen 28×28×512 boyutundaki basit öznitelik haritası iki sütun halinde işlenmeye devam etmektedir. Doğrudan bir sütunun Birleştirme Modülü (BM)'ye iletilmektedir. Diğer sütun, evrişim dikkat modülünün girdisi olarak kullanılmaktadır. Önerilen model mimarisinin temel modülü Şekil 2.22'de gösterilmektedir.



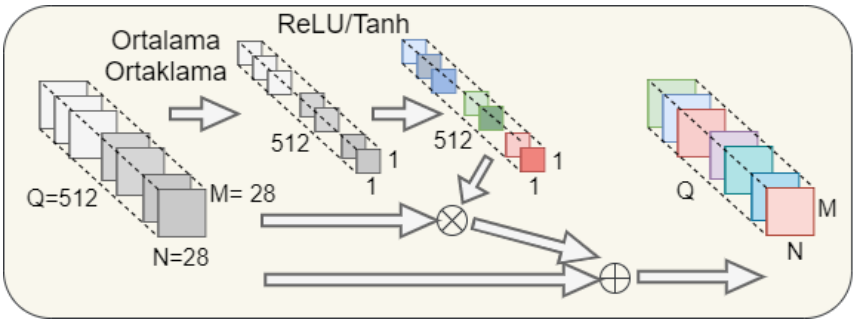

**Şekil 2.22** Önerilen modelin temel modülü [157]

### 2.3.3.2 Dikkat Modülleri

CNN'ler, yinelemeli sinir ağları ve dikkat modelleri, görüntü ve doğal dil işleme için başarılı uygulamalardır. Bu çalışmada, dikkat modülü hem CNN'lerde hem de CapsNet'te kullanılmaktadır. Böylelikle kalabalık analizinde ilk defa bu amaçla kullanılan CapsNet'in sağlamlığından ve mekânsal bilgilerinden yararlanma arzusu duyulmaktadır. Çalışmanın yenilikçi bakışı bu aşamada temellenmektedir.

- **Evrişim Dikkat Modülü**

Evrişim dikkat modülünün girişi daha önce temel model olarak kullanılan VGG16'nın 10. evrişim katmanından 28×28×512 olarak alınmaktadır. Öznitelik haritası; evrişimde ReLU ve düzlem bilgisini nokta bilgisine toplayan ve daha sonra sırasıyla ReLU ve tanh aktivasyon fonksiyonları kullanılan bir tamamen bağlı sinir ağında ortalama ortakalama işlemi ile elde edilmektedir. Son adımda, dikkat evrişimi ile çarpılmaktadır. Burada evrişim giriş boyut ayarlama denklemi $p_{sc}$ uzunluğu M, genişliği N ve kanal sayısı Q ile gösterilmektedir. Ortaklama işleminden sonra kullanılan ayarlama değeri $p_a$ denklem (2.22)'daki gibi hesaplanmaktadır. Boyutu 1×1×Q olarak elde edilmektedir. Burada amaç modelin doğrusal olmayan fonksiyonlar kullanarak öğrenme yeteneğini iyileştirmektir. $p_a$'ye biri ReLU ($p_1$) diğeri tanh ($p_2$) aktivasyon fonksiyonuna sahip iki tamamen bağlı katman uygulanmaktadır. Şekil 2.23'de önerilen evrişimli dikkat modülü gösterilmektedir.




**Şekil 2.23** Evrişim dikkat modülü [157]

Tamamen bağlı katmanlara karşılık gelen ağırlık matrisleri $W_1$ ve $W_2$, offset ise $b_1$ ve $b_2$ olarak gösterilmektedir. Bu iki tam bağlantı işleminin sonucunda elde edilen $p_2$'nin boyutu 1×1×Q olmaktadır ve denklem (2.23) ve (2.24)'de gösterilmektedir. Elde edilen $p_2$ ve girişten gelen $p_{sc}$ çarpımından $p_3$ denklemi (2.25) elde edilmektedir. Bir sonraki adımda $p_3$ ile $p_2$ toplanarak evrişim dikkat $p_{conv-att}$ denklemi (2.26) hesaplanmaktadır [166].

$$p_{aq_k} = \frac{1}{|MN|}\sum_{i=0}^{M}\sum_{j=0}^{N} q_{kij} \tag{2.22}$$

$$p_1 = ReLU(W_1 p_a + b_1) \tag{2.23}$$

$$p_2 = tanh(W_2 p_1 + b_2) \tag{2.24}$$

$$p_3 = p_{sc} * p_2 \tag{2.25}$$

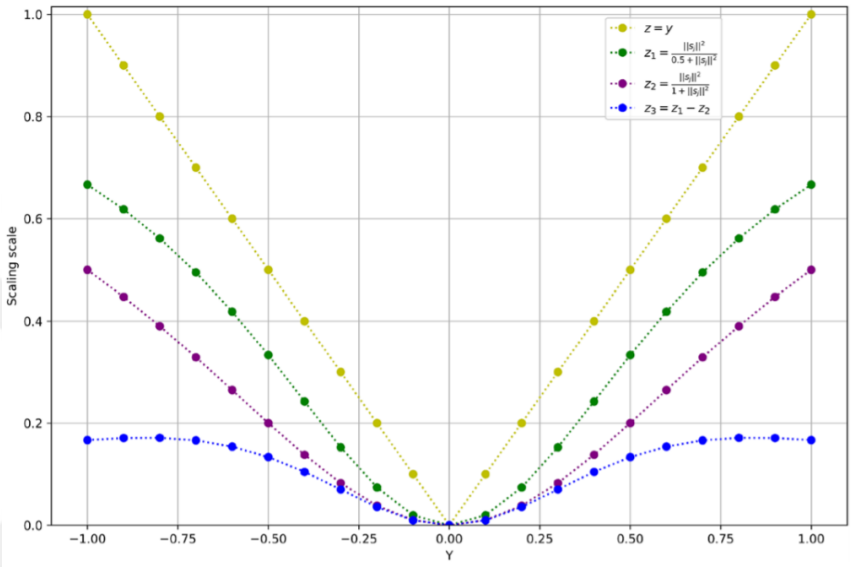
$$p_{conv-att} = p_{sc} + p_3 \tag{2.26}$$

- **Kapsül Dikkat Modülü**

Kapsül dikkat haritasını oluşturabilmek için öncül kapsülleri kullanarak yüksek seviyeli *DigitCaps* hesaplayabilmek gerekmektedir. CapsNet'lerin en önemli dezavantajı hesaplama gücü ihtiyacıdır. Bu problemin önüne geçmek için tam bağlantı katmanları bir boyutlu evrişim ve doğrusal aktivasyon fonksiyonu uygulanmaktadır [145]. Böylece öncül kapsül boyutu 100×(10×16) olarak elde edilmektedir. Öncül kapsüllere yeniden boyutlandırma işlemi uygulanarak vektörize hale getirilmektedir. Böylece ReLU ve tanh aktivasyon fonksiyonlarını kullanarak tam bağlantılı bir sinir ağına benzetilmektedir ve denklem (2.28) ve (2.29)'deki gibi hesaplanmaktadır.

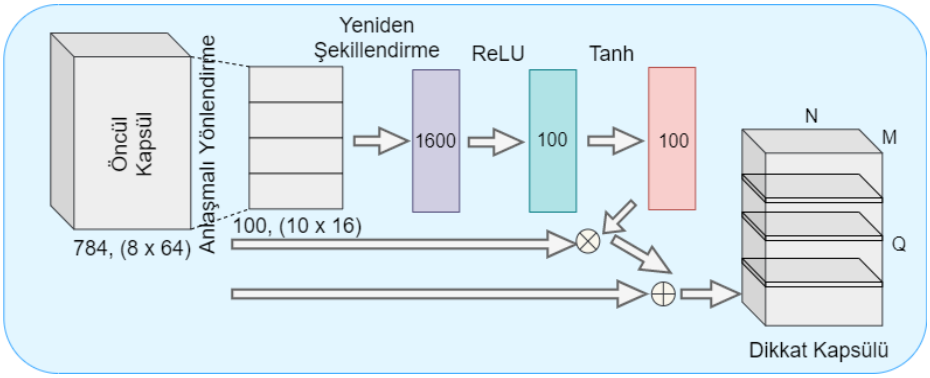
Sıkıştırma fonksiyonu için 1 değeri bu çalışmada 0,5 olarak ölçeklendirilmiştir. Bu yüzden sıkıştırma denklemi (2.27) gibi yeniden hesaplanmaktadır. Bu yaklaşım ile CapsNet'in doğruluğunun %0.5-1,59 aralığında yükseldiği Şekil 2.24'de gösterilmektedir [157].

$$v_j = Squash(s_j) = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \Longrightarrow \frac{\|s_j\|^2}{0.5 + \|s_j\|^2} \tag{2.27}$$




**Şekil 2.24** İki farklı ezme fonksiyonunun karşılaştırılması

DigitCaps adı verilen bir katman ile kod çözme yapısı tasarlanmaktadır. DigitCaps, ReLU ve tanh aktivasyon fonksiyonları tarafından kontrol edilen iki tam bağlantılı sinir hücresi katmanlarından meydana gelmektedir. Bu katmanda, eğitimde kullanılan görüntüler ile sigmoid katmanın çıktısı arasındaki Öklid mesafesi minimize edilmektedir. DigitCaps, eğitimde yeniden yapılandırma hedefi olarak doğru etiketi kullanmaya dayalı bir öğrenme stratejisine sahiptir. Şekil 2.25'de önerilen kapsül dikkat modülü gösterilmektedir.




**Şekil 2.25** Kapsül dikkat modülü [157]

Burada $p_{s1}$ ve $p_{s2}$ tam bağlantılı iki katmanın sonucu, $W_3$ ve $W_4$ bu katmanların ağırlık matrisleri, $b_3$ ve $b_4$ offset değerleridir. Bu modülün son adımında için $p_s$ ve $p_{s2}$ çarpımlarından $p_{s3}$ denklem (2.30)'daki gibi hesaplanmaktadır. Burada $p_s$ ile $p_{s3}$ toplanarak $p_{caps-att}$ adı verilen ve M×N×Q boyutlu kapsül dikkat haritası denklem (2.31)'de gösterildiği gibi elde edilmektedir [166].

$$p_{s1} = ReLU(W_3 p_{sr} + b_3) \tag{2.28}$$

$$p_{s2} = tanh(W_4 p_{s1} + b_4) \tag{2.29}$$

$$p_{s3} = p_s * p_{s2} \tag{2.30}$$

$$p_{caps-att} = p_s + p_{s3} \tag{2.31}$$

CapsNet ile klasik CNN'ler tarafından çözülemeyen; konum, yönelim, duruş ve açısal değerin değişmesi durumunda dâhi bir grup sinir hücresinden oluşan kapsüller aracılığıyla nesnenin başarıyla tanınabilmesi için nesneyi temsil eden kalınlık, ölçek, kaydırma vb. özelliklerin anlaşmalı yönlendirme ile öğrenilmesi önerilmektedir [146].

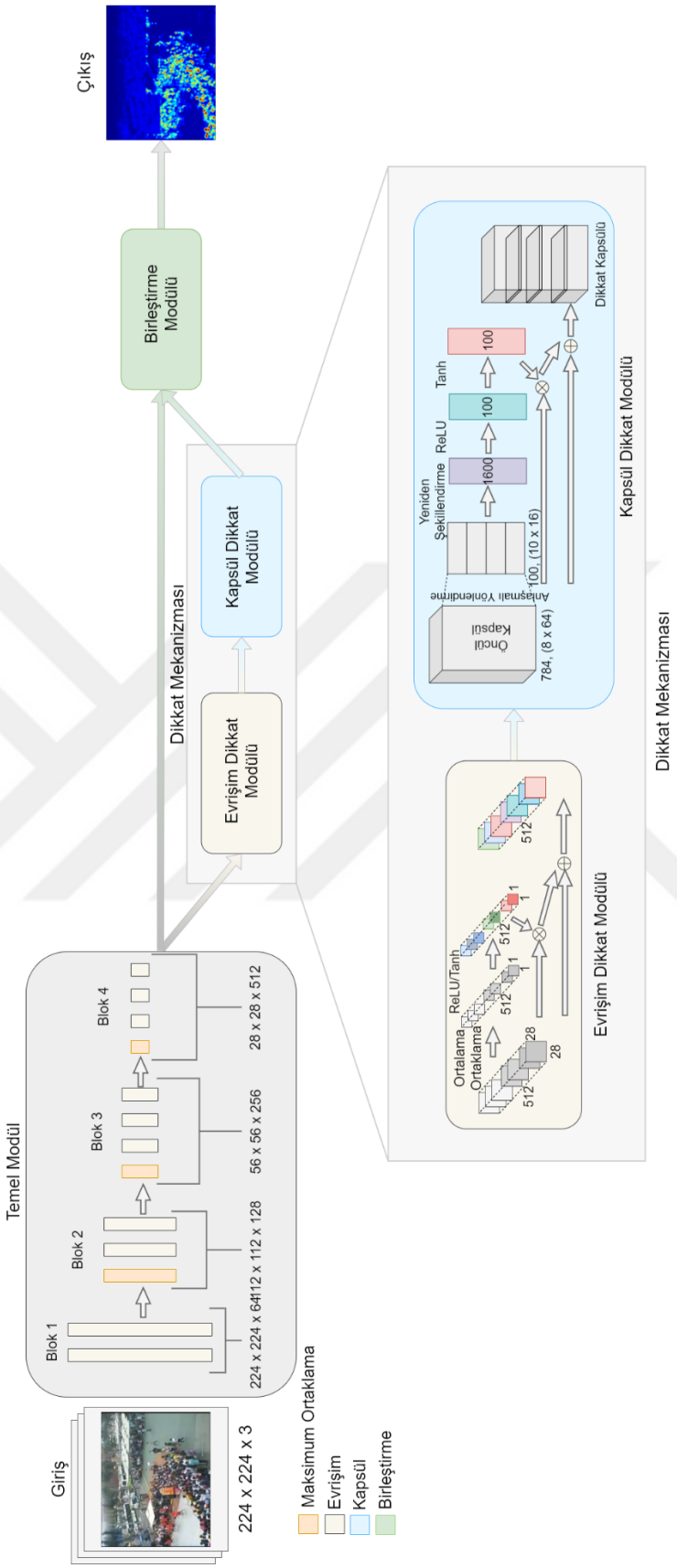
Buna ek olarak CapsNet ile daha kısıtlı eğitim kümeleriyle genellenebilir sonuçlar elde edebilmektedir. CNN'ler görütülerdeki belirsizlikle çok iyi başa çıkamamaktadır. CapsNet, kalabalık sahnelerde bile iyi performans göstermektedir. CNN'ler ortaklama katmanlarında görüntüye dair uzamsal çözünürlük azalmaktadır. Bu nedenle çıktılar, girdilerdeki küçük değişiklikleri algılayamamaktadır. CapsNet ile bu sorunun da üstesinden gelinmektedir. CNN'ler aynı nesneye ait parçaları bir bütün olarak algılamakta zorlanırken CapsNet,

parçalar arasındaki uzamsal ilişkiyi saklayabilmesi sayesinde bütüne yönelik daha etkin sınıflandırma ve bölütleme sonuçları vermektedir.

#### 2.3.3.3 Birleştirme Modülü

Çok seviyeli özellikler arasındaki uzamsal ve anlamsal bilgiyi iki farklı dikkat modülüyle elde ettikten sonra BM olarak iki dikkat haritası tek kanallı bir kalabalık yoğunluk haritası halinde birleştirilmektedir. Denklem (2.32)'deki gibi gösterilmektedir. Şekil 2.26'de önerilen dikkat temelli kapsül ağı ve iki sütunlu CNN modelin genel mimari temsili gösterilmektedir.

$$BM = [p_{conv-att}, p_{caps-att}] \tag{2.32}$$




**Şekil 2.26** Önerilen dikkat temelli kapsül ağı ve iki sütunlu CNN modelin genel mimari temsili [157]

## 2.4 Değerlendirme Metrikleri

Üzerinde çalışılan üç farklı yöntem için model performansı temelde Ortalama Mutlak Hata (MAE) ve Ortalama Karesel Hata (MSE) değerlendirme metrikleri tercih edilmektedir. Bunun en önemli sebebi ilgili literatür ile karşılaştırma yapabilmektir. MAE, model tarafından tahmin edilen sonuçların doğruluğunu yansıtmaktadır. MSE, modelin sağlamlığını ölçmek için kullanılmaktadır. Denklem (2.33) ve (2.34)'de $N$ test kümesindeki görüntü sayısı, $i$ görüntünün indeksidir, $y_i$ görüntü için hesaplanan tahmini sayıdır ve $y_i^{GT}$, görüntüdeki doğru etiket sayısıdır.

$$MAE = \frac{1}{N}\sum_{i=0}^{N}\left|y_i - y_i^{GT}\right| \tag{2.33}$$

$$MSE = \frac{1}{N}\sum_{i=0}^{N}\left|y_i - y_i^{GT}\right|^2 \tag{2.34}$$

ShanghaiTech Part A veri kümesi için Yapısal Benzerlik Endeksi (SSIM) karşılaştırması da gerçekleştirildi. Böylece yoğunluk haritasının kalitesi de ölçülmekte ve denklem (2.35)'daki gibi tanımlanmaktadır.

$$SSIM = \frac{1}{N}\sum_{i=0}^{N}\left(\frac{(2\mu P_{yi}.\mu G_{yi^{GT}} + y_1)}{\mu_{P_{yi}}^2 + \mu_{G_{yi}}^2 + y_1}\cdot\frac{(2\sigma P_{yi}.\sigma G_{yi^{GT}} + y_2)}{\sigma_{P_{yi}}^2 + \sigma_{G_{yi}}^2 + y_2}\right) \tag{2.35}$$

SSIM kaybı hesaplaması için $\mu P_{yi}$ ve $\mu G_{yi}$ ortalama, $\sigma P_{yi}$, $\sigma G_{yi}$ ve $\sigma P_{yi}.\sigma G_{yi^{GT}}$ standart sapma değerleri kullanılmaktadır.

PETS2009 veri kümesinde grup algılama ve davranış değişimi kestirimi için farklı metrikler de hesaplandı. Yoğunluk haritasını hesaplamak için Gauss çekirdeklerinin σ ve δ parametreleri denklem (2.36) 'da olduğu gibi hesaplanmaktadır.

$$Y = \delta(c - c_i) * G_\sigma, (Gaussian\ kernel\ \sigma = 4) \tag{2.36}$$

Bu yöntemde önceden eğitilmiş temel model eğitimi için, küme boyutu 16, momentum 0.9, ağırlık azalması L2 norm 5e-4, optimizasyon algoritması Adam ve öğrenme hızı 1e-4 seçilmektedir.

Veri kümesindeki 1, 2, 3 ve 5 dizileri kullanılmaktave bu sekanslarda elde edilen sonuçlar hem çerçevedeki kişi sayısı hem de doğru etiket değerleri ile davranış sınıfı, WCS (Dünya Koordinat Sistemi) sonuçları ve literatürdeki güncel çalışmalar açısından karşılaştırılmak üzere kullanılmaktadır. Bunun için ihtiyaç duyulan doğruluk değerlendirmesi denklem (2.37)'deki gibi hesaplanmaktadır.

$$1 - \ Kestirim - Doğru\ Etiket \Big/ Toplam\ Sahne\ Sayısı - Doğru\ Etiket \qquad (2.37)$$

Çalışmanın bu aşamasında elde edilen başarım karşılaştırmaları Sonuç ve Öneriler bölümünde gösterilmektedir.

## 2.5 Deney Ortamı

Çalışmanın öneri aşamasında NVIDIA GeForce GTX850M ekran kartına sahip kişisel dizüstü bilgisayar ile Matlab ortamında geliştirime yapılmıştır. Çalışmanın literatür incelemesi bölümünde de belirtilen optik akış, Kalman filtre gibi uygulamalar üzerinde bu donanım ortamında çalışılmıştır. Böyle bir çalışma ortamının yeterli olmayacağı derin öğrenme uygulamalarına geçiş aşamasında farkedilmiştir. Görüntüler üzerinde kalabalık analizi gerçekleştirmek için GPU (Grafik İşlem Birimi) ihtiyacının ortaya çıktığı tespit edildi ve NVIDIA Grant Program başvurusu gerekleştirilmiştir.

Bu çalışma; Pascal mimarisine, 1582 MHz bellek hızında ve 3840 CUDA çekirdeğine sahip bir Titan Xp GPU ile NVIDIA Grant Program tarafından desteklenmektedir. Bunun üzerine Intel i7-7700 3.60 GHz 8M 1151p özelliklerine sahip 32 GB RAM'e sahip bir masaüstü bilgisayar kurulmuştur. Linux Ubuntu 16.4 ve çalışmanın son aşamalarında ise Ubuntunun KDE kullanıcı arayüzü olan Kubuntu 20.10 üzerinde deney ortamı kurulmuştur. Programalma dili olarak Python'un 3.7 sürümü kütüphanelerin ve donanımların uyumluluğu sağlanacak şekilde kullanılmıştır. Derin öğrenme kütüphanesi olarak TensorFlow 2.0 ve 2.1 sürümleri kullanılarak model mimarileri kodlanmıştır.

Deneyler başta CNN olmak üzere derin öğrenme modellerine odaklanmaktadır. Gerekli görülen noktalarda yapay öğrenme, istatistiksel ve değerlendirme metriklerine ilişkin diğer matematiksel çıkarımlara da aynı deney ortamında

hesaplanabilecek şekilde organize edilmektedir. Tüm deneyler aynı deney ortamında gerçekleştirilerek literatürdeki diğer çalışmalarla karşılaştırma yapılmıştır. Bunun için deneylere özgü parametre optimizasyonları ve seçimleri de gerçekleştirilmektedir. Bununla ilgili detaylar tez çalışmasının Sonuç ve Öneriler bölümünde yer verilmektedir.

# 3
# SONUÇ VE ÖNERİLER

## 3.1 Sonuçlar

Kamuya açık alanlarda pek çok planlama ve denetleme gerektirdiği bilinmektedir. Bunun için tez kapsamında kalabalık alanlardan toplanmış olan farklı veri kümeleri kullanılarak kalabalık yoğunluğunun kestirilmesi sağlanmıştır. Kapsamlı ve detaylı bir literatür incelemesinin ardından YZ temelli yaklaşımlar üzerine odaklanılmıştır. Kalabalık görüntülerde insan sayısını tahmin etmek için kapsül ağlarının uzamsal bilgi öğrenme yeteneğinden faydalanma fikri ile yapılan araştırmada teoriye uygun pratik sonuçlarelde edilmiştir.

Bu tez çalışması CapsNet ve CNN mimarilerini bir dikkat mekanizması olarak kalabalık analizi için öneren ilk çalışmadır. Bu özgün ve etkin yaklaşımın uygulanması ve bu aşamaya gelirken yapılan deneylerin sonuçları son teknoloji diğer çalışmalarla karşılaştırmalı olarak tezin bu bölümünde sunulmaktadır. Bunun yanı sıra önerilen yaklaşım için ablasyon deneyler sonuçları ile her bir modülün sonuca etkisi incelenmektedir. Tezin amacı, hipotez ve elde edilen sonuçlardan yola çıkarak tartışma ve öneriler sunulmaktadır.

### 3.1.1 Yüksel Yoğunluklu Kalabalık için Derin ve Geniş CNN Modeli Sonuçları

Tez kapsamında yapılan ve 2019 yılında yayınlanan bir çalışmada, yoğun kalabalık sahnelerdeki insan sayısını tahmin etmek için bir derin sinir ağı modeli önerilmiştir. Önerilen ağa DWCNN adı verilmiştir [15]. Shanghai Tech Part A [82] veri kümesi, kalabalık yoğunun fazla olması sebebiyle çalışmanın ilk adımında kullanılmak üzere seçilmiştir. Sahne sayısı 482, her bir görüntüdeki ortalama kişi sayısı 501, maksimum kişi sayısı 3139 ve veri kümesinde yer alan toplam kişi sayısı 241677'dir. Konser, stadyum, siyasi ve dini toplantılardan elde edilen görüntülerden meydana getirilmiş bu veri kümesindeki her bir sahne için kişi sayısı kestirimine odaklanılmıştır.

Çalışma için veri kümesindeki RGB görüntüler gri tonlama görüntüye dönüştürülmüştür. Veri kümesinin %75'i eğitim, %15'i geçerleme ve %10'u ise test için ayrılmıştır. Veri kümesindeki kalabalık yoğunluğunun anlaşılması için örnek görüntüler Şekil 3.1'de gösterilmektedir.



**Şekil 3.1** Shanghai Tech Part A veri kümesinden örnekler [82]

Önerilen iki sütunlu CNN modelinin biri sığ diğeri paralel evrişim katmanlarından meydana gelmektedir. Sığ model olarak VGG-16 modelinin ilk iki katmanı ImageNet veri kümesinde eğitilmiş ve böylece önerilen modelin başarımı artırılmıştır [159]. Böylece sığ katmanlardan elede edilen öznitelikler daha hızlı ve etkin şekilde öğrenilmektedir. Bu transfer öğrenme süreci, özellikle ilk katmanlarda kullanıldığında hesaplama karmaşıklığını azalttığı için modelin çalışma performansını ve süresini de iyileştirmektedir. Çalışma için seçilen hiper parametreler ve optimizasyon yöntemi Tablo 3.2'de gösterilmektedir.

**Tablo 3.1** Seçilen parametreler ve optimizasyon yöntemleri [15]

| **Parametreler ve optimizasyon yöntemi** | Stokastik gradyan iniş algoritması | Öğrenme oranı | Küme boyutu |
|---|---|---|---|
| | 0.9 | 0.0001 | 32 |

Değerlendirme metriği olarak, MAE ve MSE kullanılmasının en önemli sebebi literatürde yapılan önemli çalışmalarla karşılaştırma yapabilme imkânının olmasıdır. Literatürde bu konudaki çalışmalarda doğruluk hesabına başvurulmamaktadır. Tablo 3.3'de elde edilen sonuçlar literatürdeki güncel çalışmalarla karşılaştırılmıştır.

- MSE'nin 181,8'den 84'e,
- MAE'nin ise 277,7'den 92,4'e düştüğü görülmüştür.

**Tablo 3.2** ShanghaiTech Part A veri kümesindeki DWCNN yönteminin başarım karşılaştırılması [15]

| **Referans** | **MSE** | **MAE** |
|---|---|---|
| Idrees vd. [149] | 181.8 | 277.7 |
| Shang vd. [90] | 110.2 | 173.2 |
| Kumagai vd. [19] | 126.5 | 173.5 |
| Boominathan vd. [91] | 101.3 | 152.4 |
| **Önerilen Yöntem DWCNN [15]** | **84.0** | **92.4** |

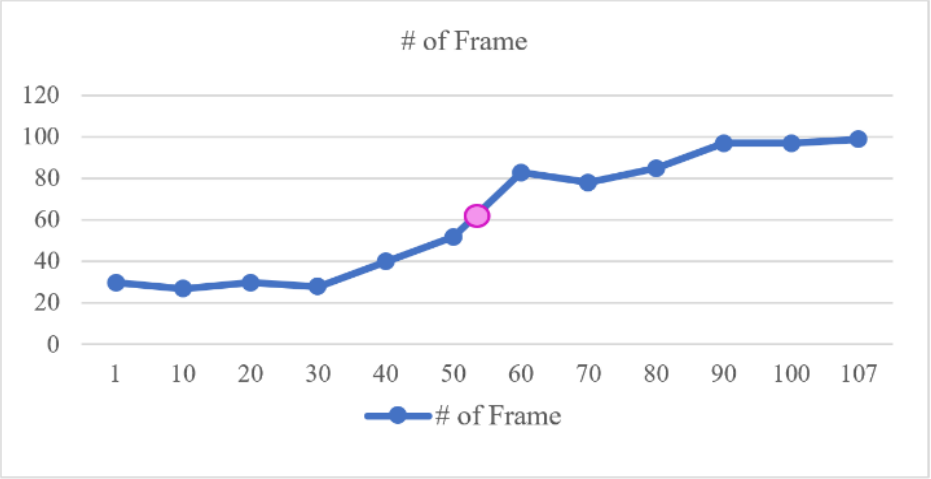
Önerilen DWCNN yöntemi özellikle güvenlik eksikliğinin önlenmesi ve etkinlik alanlarının doğru oranlarla dağıtımının sağlanması gibi düzenleme operasyonlarının yönetilmesi için birim çalışanlarına destek olarak kullanılabilecek bir destek sistemi olabilir.

### 3.1.2 Kalabalık Sayımı için Küme Temelli İzleme ve Konum Tahmini Sonuçları

PETS2009 veri kümesinin 4 video dizisi için performansı, güncel bir gerçek zamanlı derin öğrenme modeli olan YOLOv3'ün özel bir sürümünü olan Gauss-YOLOv3 kullanılarak değerlendirilir [152]. Verilerin uzaklık bilgilerinden faydalanarak sınıflandırma konusunda etkili olduğu bilinen KNN algoritması kalabalıktaki

grupları belirlemek için kullanılmaktadır. Çalışmanın bu aşamasında gerçekleştirilen deneyler kapsamında PETS2009 [150] veri kümesindeki kişi sayısı ve yürüme/koşma olaylarının belirlenmesi üzerine odaklanılmıştır [16]. Bu uygulamada işlem hızını artırmak için sabit genişlikte küme sayısı ve KNN algoritmasında iki sınıf ($k = 2$) seçilmiştir. Gauss-YOLOv3 için eşik ayarı 0 olarak ayarlanmıştır. Veri kümesindeki kalabalık içinde alt grupların oluştuğu ve zamanla değişen davranışların sergilendiği 1, 2, 3 ve 5 video sekansları kullanılmıştır. Bu sekanslarda elde edilen sonuçlar hem görüntüdeki kişi sayısı ve referans değerler arasındaki ilişkiye hem de Dünya Koordinat Sistemi (WCS) sonuçları ve literatürdeki güncel çalışmalar açısından karşılaştırılmıştır.

Sekans 1 için hesaplanan davranış değişim süresi, Şekil 3.2'de grafik olarak gösterilmektedir. Grafik; x ekseni sahne numarası, y ekseni ise her iki küme arasındaki Öklid mesafesinin ortalama değeri olarak ifade edilmektedir. Burada sekans 1'de yer alan 107 sahne içinden 52. sahne davranış değişiminin olduğu hesaplanmaktadır. Şekil 3.3'de, bir grup insanın yürüme davranışından koşma davranışına geçiş yaptığı ilgili video sahnesi olarak tespit edilen görüntüler gösterilmektedir. Tablo 3.4, referans değer ve önerilen yöntemle elde edilen her bir sekansdaki kişi sayısı sonuçlarını göstermektedir.




**Şekil 3.2** Sekans 1 için yürüme/koşma davranışı için elde edilen örnek çıktı zamanı serisi [16]



**Şekil 3.3** Yürüme davranışından koşma davranışına değişen bir grup insan [16]

**Tablo 3.3** Algılama sonuçları: yürüme / koşma davranışı [16]

| Video sekansları | Referans değer | Önerilen yöntem |
|---|---|---|
| Sekans 1 | 38 | 39 |
| Sekans 2 | 56 | 62 |
| Sekans 3 | 335 | 339 |
| Sekans 5 | 76 | 83 |

Sonuç olarak, önerilen kalabalık izleme yaklaşımı, diğer son teknoloji yöntemlerle karşılaştırıldığında kullanılan dört farklı video sekansı için de %83,2 ile%96,4 arasında değişen bir başarım göstermektedir. Tablo 3.5, elde edilen sonuçların doğruluk metriği açısından literatür ile karşılaştırmalı başarımı gösterilmektedir.

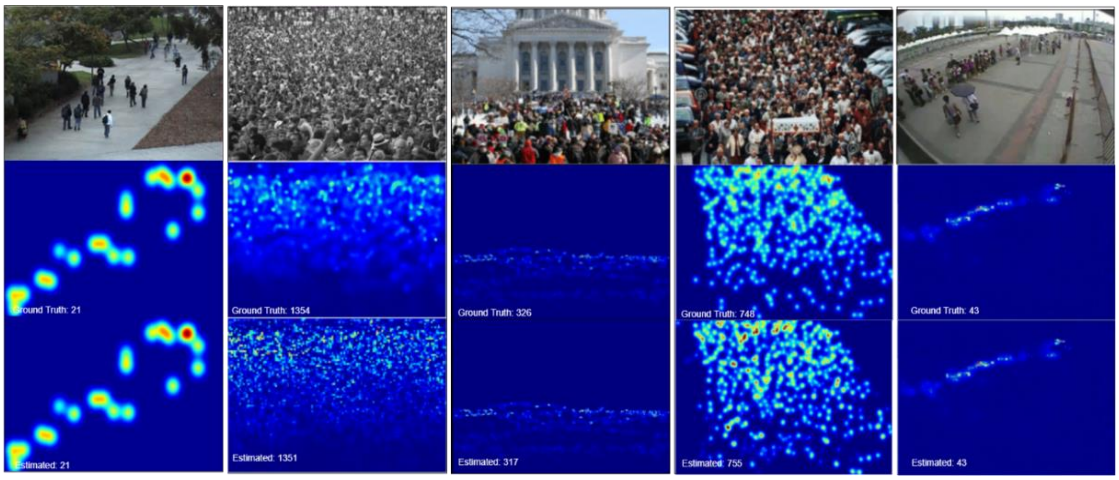
**Tablo 3.4** Dört farklı video sekansı için önerilen yöntemin doğruluk başarım karşılaştırması

| Referans | Sekans 1 için doğruluk başarımı | Sekans 2 için doğruluk başarımı | Sekans 3 için doğruluk başarımı | Sekans 5 için doğruluk başarımı |
|---|---|---|---|---|
| WCS | %85.5 | %63.8 | %81.0 | %87.0 |
| BRIASS | %79.7 | %50.0 | %69.0 | - |
| Yang vd. [160] | %95.7 | %82.8 | %95.2 | %90.7 |
| **Önerilen Yöntem [16]** | **%96.4** | **%83.2** | **%95.6** | **%93.8** |

Çalışmanın bu noktasında güncel derin öğrenme modeli Gauss-YOLOv3 ve KNN algoritması birlikte kullanılarak video sahnelerindeki kalabalık davranışı değişiklikleri tespit etmeye yönelik yeni bir bakış açısı sunulmuştur. Çalışamada güncel ve nispeten daha eski yöntemlerin birlikte kullanılması gücünden faydalanılarak davranış değişimini anları tespit edilebildiği gösterilmektedir. Aynı veri kümesinde davranış sınıflandırması ve davranış değişikliğinin tespiti üzerine yapılmış çalışmalarla karşılaştırılabilir bir başarının elde edildiği görülmektedir.

### 3.1.3 Dikkat Temelli Kapsül Ağı ve İki Sütunlu CNN Kullanarak Kalabalık Yoğunluğu Tahmini Sonuçları

Çalışmanın bu aşamasında nihai yöntem önerisi yapılmaktadır. Yöntem önerinin genelleştirilebilirliğinin sağlanması için kalabalık analizi üzerine hazırlanan veri kümelerinden farklı zorluklara sahip beşi kullanılmıştır. Bu veri kümeleri farklı boyutlara sahip olmalarından dolayı öncelikle hepsi önerilen yöntemin giriş parametresine uygun olacak şekilde yeniden boyutlandırılmıştır. Bu tezde farklı mekânlardan toplanan verilerin olduğu ve yoğunluk kestirimi ve kişi sayısına yönelinmiştir. Tablo 2.1'deki yedi veri kümesinden MALL ve CUHK veri kümeleri çalışmada kullanılmamıştır. Bunun nedeni MALL sadece iç mekân görüntülerinden oluşmaktadır ve CUHK kalabalık davranışıyla ilgilidir. Şekil 3.4’de UCF-QNRF, UCF_CC_50, UCSD, ShanghaiTech Part A ve WorldExpo'10 veri kümelerinin referans değer ve yoğunluk haritaları ile tahmin edilen kişi sayısı ve yoğunluk haritalarının karşılaştırılması gösterilmektedir.




**Şekil 3.4** Soldan sağa doğru sütun sırasıyla UCF-QNRF, UCF_CC_50, UCSD, ShangaiTech Part A ve WorldExpo'10 veri kümelerinde referans değer, kestirilen kalabalık yoğunluk haritası ve kişi sayısı sonuçları

Önerilen yaklaşımın performansını kapsamlı bir şekilde inceleyeilmek ve başarımını ispatlayabilmek için değerlendirme metrileri hesaplanmıştır. Sonuçlar son teknoloji sonuçlarla kapsamlı şekilde karşılaştırılmıştır. Elde edilen sonuçlardaki iyileştirme oranları aşağıda listelenmektedir.

- UCF-QNRF [147] veri kümesinde, önerilen yöntem MSE'de kendisine en yakın başarımı elde edilen Liu vd. [162] çalışmalarından %2.18'lik bir artış sağlamıştır.
- UCF_CC_50 [149] veri kümesinde, Idrees vd. [147]'e göre %33 daha başarılı MAE ve Liu vd. [162]'ye göre ise %1,68 daha iyi MSE sonucu elde edilmiştir.
- UCSD [32] veri kümesinde, Idrees vd. [102] ile karşılaştırıldığında MAE sonucu için %8,46 oranında ve Basalamah vd. [115] ile karşılaştırıldığında %7,03 MSE ile iyileştirilmiştir.
- ShanghaiTech Bölüm A [82] veri kümesinde, %4,69 MAE ve %3,94 MSE oranında iyileştirilme ile en yakın doğruluktaki Sindagi vd. [95]'in önüne geçilmiştir.
- WorldExpo'10 [81] veri kümesinde, Liu vd. [162]'nin elde ettiği en yakın sonuçtan %6,94 daha başarılıdır.

Bu beş veri kümesi için deneysel sonuçlar Tablo 3.6 – 3.11'de karşılaştırılmaktadır.

**Tablo 3.5** UCF-QNRF veri kümesinde karşılaştırma sonuçları

| **Yöntem** | **Kalabalık Sayısının Tahmin Hatası** | |
|---|---|---|
| | MAE | MSE |
| Sam vd. [83] | 252 | 514 |
| Idrees vd. [32] | 277 | 426 |
| Badrinarayanan vd. [161] | 228 | 445 |
| He vd. [90] | 190 | 277 |
| Huang vd. [162] | 163 | 226 |
| Idrees vd. [147] | 132 | 191 |
| Liu vd. [162] | 107 | 183 |
| **Önerilen Yöntem** | **109** | **179** |

**Tablo 3.6** UCF_CC_50 veri kümesinde karşılaştırma sonuçları

| **Yöntem** | **Kalabalık Sayısının Tahmin Hatası** | |
|---|---|---|
| | MAE | MSE |
| Kang vd. [96] | 406.2 | 404 |
| Zhang vd. [82] | 377.6 | 509.1 |
| Sam vd. [83] | 318.1 | 439.2 |
| Zhu vd. [86] | 302.3 | 411.6 |
| Shi vd. [85] | 288.4 | 404.7 |
| Li vd. [98] | 266.1 | 397.5 |
| Basalamah vd. [115] | 235.74 | 345.6 |
| Liu vd. [163] | 212.2 | 243.7 |
| Xu vd. [164] | 188.4 | 315.3 |
| **Önerilen Yöntem** | **184** | **239.6** |

**Tablo 3.7** UCSD veri kümesinde karşılaştırma sonuçları

| **Yöntem** | **Kalabalık Sayısının Tahmin Hatası** | |
|---|---|---|
| | **MAE** | **MSE** |
| Zhang vd. [82] | 1.60 | 3.31 |
| Sam vd. [83] | 1.62 | 2.10 |
| Kang vd. [96] | 1.17 | 2.15 |
| Li vd. [98] | 1.16 | 1.47 |
| Liu vd. [163] | 1.03 | 1.32 |
| Basalamah vd. [115] | 1.01 | 1.28 |
| Jiang vd. [102] | 0.97 | 1.30 |
| **Önerilen Yöntem** | **0.89** | **1.19** |

**Tablo 3.8** ShangaiTech Part A veri kümesinde karşılaştırma sonuçları

| **Yöntem** | **Kalabalık Sayısının Tahmin Hatası** | |
|---|---|---|
| | **MAE** | **MSE** |
| Zhang vd. [81] | 181.8 | 277.7 |
| Marsden vd. [104] | 126.5 | 173.5 |
| Zhang vd. [82] | 110.2 | 173.2 |
| Sam vd. [83] | 90.4 | 135.0 |
| Chen vd. [148] | 73.6 | 106.4 |
| Gao vd. [107] | 73.5 | 124.0 |
| Li vd. [98] | 68.2 | 115.0 |
| Shi vd. [99] | 66.3 | 106.4 |
| Jiang vd. [102] | 62.4 | 102.6 |
| Liu vd. [163] | 62.0 | 100.0 |
| Sindagi vd. [95] | 46.9 | 71.0 |
| **Önerilen Yöntem** | **44.7** | **68.2** |

**Tablo 3.9** ShangaiTech Part A veri kümesinde yoğunluk haritası kalitesi için karşılaştırma sonuçları

| **Yöntem** | **Yoğunluk Haritasının Kalitesi** |
|---|---|
| | **SSIM** |
| Zhang vd. [82] | 0.52 |
| Chen vd. [148] | 0.72 |
| Li vd [98] | 0.76 |
| **Önerilen Yöntem** | **0.43** |

**Tablo 3.10** WorldExpo'10 veri kümesinde karşılaştırma sonuçları

| **Yöntem** | **Kalabalık Sayısının Tahmin Hatası** |
|---|---|
| | MAE |
| Zhang vd. [81] | 12.9 |
| Zhang vd. [82] | 11.6 |
| Sam vd. [83] | 9.4 |
| Shi vd. [85] | 9.1 |
| Sindagi vd. [84] | 8.86 |
| Li vd. [98] | 8.6 |
| Shen vd. [110] | 7.5 |
| Basalamah vd. [115] | 7.42 |
| Liu vd. [163] | 7.2 |
| **Önerilen Yöntem** | **6.7** |

### 3.1.3.1 Yoğunluk Haritası Parametreleri Üzerine Ablasyon Deneyleri

Çalışmanın parametrelere seçimlerinin iyileştirilmesi için yoğunluk haritası üzerinde ablosyon deneyleri gerçekleştirilmiştir. Kalabalık sayma görevlerinde yoğunluk haritaları oluştururken Gauss çekirdek değeri $\sigma=4$ seçilmiştir. Bu deneyde, $\sigma$ değeri 16 seçildiğinde, sonuçlar tüm veri kümelerinde karşılaştırılmaktadır. Önerilen yöntemle seçilen $\sigma$ parametresi ile en iyi sonuca

ulaşıldığı gösterilmiştir. Kullanılan beş veri kümesi için Gauss çekirdek ablasyon deneyinin sonuçları Tablo 3.12'de karşılaştırılmıştır.

**Tablo 3.11** Yoğunluk haritası üretimi için Gauss çekirdek ablasyon değerinin sonuçları

| Veri kümeleri | Gauss çekirdek değeri | MAE | MSE |
|---|---|---|---|
| UCF-QNRF [147] | σ =4 | 109 | 179 |
| | σ =16 | 112.2 | 187.1 |
| UCF_CC_50 [149] | σ =4 | 184 | 239.6 |
| | σ =16 | 187.2 | 242 |
| UCSD [32] | σ =4 | 0.89 | 1.19 |
| | σ =16 | 0.93 | 1.32 |
| ShangaiTech Part A [82] | σ =4 | 44.7 | 68.2 |
| | σ =16 | 47.2 | 71.0 |
| WorldExpo'10 [81] | σ =4 | 6.7 | - |
| | σ =16 | 7.2 | - |

### 3.1.3.2 Önerilen Yöntemin Mimarisinde Ablasyon Deneyleri

Önerilen model çeşitli modüllerden oluşmaktadır. Özellikle temel modül ve CNN modülün kalabalık analizinde kullanımında literatür özetinde de görüldüğü gibi pek çok örnek vardır. Çnerilen modeled CNN ve beraberinde CapsNet modülleri dikkat mekanizması olarak kullanılmaktadır Çelışma temelinde CapsNet'in bu şekilde kullanılmasının pozitif etkilerini deneysel sonuçlarla ispatlamayı amaçlamaktadır. Bu yüzden deneyler; kullanılan veri kümeleri üzerinde önerilen mimarideki modüllerin birer birer çıkarılması, yani yalnızca CNN dikkat modülü ve yalnızca CapsNet dikkat modülü ile tekrarlanmıştır. Özetle bu ablosyan deneyi ile CapsNet ve CNN dikkat modüllerinin birlikte kullanılması ile elde edilen uzamsal zöniteliklerin kalabalık analizinde kişi sayısı belirlemek için kullanılabileceğini göstermektedir.

Sonuçlar Tablo 3.13'de karşılaştırılmıştır. Bu model mimarisinin kitle analizinde, insan sayısını son zamanlarda yapılan birçok çalışmaya göre daha sağlam ve yoğunluk haritası kalitesiyle başarılı bir şekilde tahmin ettiğini göstermektedir.

**Tablo 3.12** UCF-QNRF, UCF_CC_50, UCSD, ShangaiTech Part A ve WorldExpo'10 veri kümelerindeki ablasyon modeli yapılarının sonuçları

| | | **Modüller** | | |
|---|---|---|---|---|
| **Veri kümeleri** | **Yöntem** | Temel Modül + Evrişim Dikkat Modülü + Birleştirme Modülü | Temel Modül + Kapsül Dikkat Modülü + Birleştirme Modülü | **Önerilen Yöntem** |
| UCF-QNRF [147] | MAE | 143 | 110 | **109** |
| | MSE | 198 | 182 | **179** |
| UCF_CC_50 [149] | MAE | 225.6 | 206 | **184** |
| | MSE | 287 | 244.3 | **239.6** |
| UCSD [32] | MAE | 1.23 | 1.03 | **0.89** |
| | MSE | 1.55 | 1.20 | **1.19** |
| ShangaiTech Part A [82] | MAE | 65.6 | 59.6 | **44.7** |
| | MSE | 102.8 | 98.1 | **68.2** |
| WorldExpo'10 [81] | MAE | 8.8 | 7.6 | **6.7** |

## 3.2 Tartışma ve Öneriler

Kalabalık analizi konusunun pek çok alt alanı olduğu bilinmektedir. Bunlardan biri de bu tez çalışması kapsamında incelenen yoğunluk analizidir.

Probleme özgü yaklaşımlara ihtiyaç olduğu literatür incelemesi ile anlaşılmaktadır. Bu bağlamda çalışma üç temel aşamadan geçmiştir. Amaçlanan ve hipotez olarak öne sürülen sağlam, etkin ve genelleştirilebilir bir yaklaşım ile kalabalık içindeki kişi sayısı kamu güvenliği ve faydası için doğru şekilde tespit edebilmektir. Bunun için kalabalık görüntüsü video olsun veya olmasın uzamsal bilgiden faydalanarak çok daha keskin sonuçlar elde edilebileceği sunulan deneysel sonuçlarla ispatlanmaktadır.

CaspNet yaklaşımının teorik olarak uzamsal bilgileri elde ettiği bilimektedir. Ancak bu çalışmada CapsNet'in özel bir dikkat mekanizması olarak tasarlanması ve CNN ile birlikte iki sütunlu bir yapıda önermesinin başarılı olabileceği deneysel olarak gösterilmektedir. Teorik öngörülerin pratik olarak da veri kümeleri ve deney ortamındaki testler ile gösterilmektedir. Bu sayede teorik bir yeniliğin, pratik hayatta da karşılık bulması için yol açılmaktadır. Elbette bu çalışmanın da kendine has kısıtları tespit edilmiştir. Hesaplama karmaşıklığından doğan donanım ihtiyacı ve dolayısıyla testlerin gerçekleştirilmesi için ihtiyaç duyulan grafik işlem kartıdır. Donanım ve süre bakımından iyileştirmelere açık bir çalışmadır. Bu yüzden alandaki bilim insanlarının üzerinde geliştirme yapabileceği yeni bir derin öğrenme mimarisi sunulmaktadır.

Çalışmanın bilimsel katkıları şöyle özetlenebilir:

- Daha etkin öznitelik çıkarımı gerçekleştirmesi,
- Poz ve konum gibi bilgilerin genelleştirilebilmesi,
- Daha az veri ile perspektif sorununun önüne geçilmesi,
- Çok daha az katman ile çok derin ağların benzer tahmin performansına ulaşılabilmesi,
- Önerilen yaklaşım ile konum ve yön bilgilerinin de gelecekteki çalışmalarda kalabalık davranış analizi için kullanılabileceği gösterilmiştir.

### 3.2.1 Gelecek Çalışmalar

Önerilen yöntemin kalabalık görüntülerinde sıklıkla karşılaşıldığı gibi örtüşen nesnelerin bulunduğu görüntülerden elde edebildiği konum ve oryantasyon öznitelikleri sayesinde iyi birer bölütleme aracı olarak da kullanılabileceği öngörülmektedir.

Bu nedenle, kalabalık analizinin yanı sıra hareket bilgisine sahip farklı bilgisayarlı görü problemleri için de uygulanabileceği kanısına varılmaktadır.

# KAYNAKÇA

[1] H. Y. Swathi, G. Shivakumar ve H.S. Mohana, "Crowd behavior analysis: A survey," in Proc. Int. Conf. on Recent Advances in Electronics and Communication Technology (ICRAECT), March 2017, pp. 169-178.

[2] A. C. Sankaranarayanan, A. Veeraraghavan, ve R. Chellappa, "Object detection, tracking and recognition for multiple smart cameras," in Proc. IEEE, vol. 96, no. 10, 2008, pp. 1606–1624.

[3] S. Khan ve M. Shah, "Tracking multiple occluding people by localizing on multiple scene planes," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 3, pp. 505–519, Mar. 2009.

[4] R. Eshel ve Y. Moses, "Homography based multiple camera detection and tracking of people in a dense crowd," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2008.

[5] F. Fleuret, J. Berclaz, R. Lengagne, ve P. Fua, "Multicamera people tracking with a probabilistic occupancy map," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 2, pp. 267–282, Feb. 2008.

[6] N. Anjum ve A. Cavallaro, "Trajectory association and fusion across partially overlapping cameras," in Proc. IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS), Sep. 2009, pp. 201–206.

[7] Y. Xu, X. Liu, Y. Liu, ve S.-C. Zhu, "Multi-view people tracking via hierarchical trajectory composition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 4256–4265.

[8] P. Dollar, C. Wojek, B. Schiele, ve P. Perona, "Pedestrian detection: A benchmark," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 304–311.

[9] L. D. Pizzo, P. Foggia, A. Greco, G. Percannella, ve M. Vento, "Counting people by RGB or depth overhead cameras," Pattern Recognit. Lett., vol. 81, pp. 41–50, Oct. 2016.

[10] C. Ertler, H. Possegger, M. Opitz, ve H. Bischof, "Pedestrian detection in RGB-D images from an elevated viewpoint," in Proc. 22nd Comput. Vis. Winter Workshop, 2017, pp. 1–9.

[11] I. Ahmed ve A. Adnan, "A robust algorithm for detecting people in overhead views," Cluster Comput., vol. 21, no. 1, pp. 633–654, Mar. 2018.

[12] P. Baque, F. Fleuret, ve P. Fua, "Deep occlusion reasoning for multicamera multi-target detection," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 271–279.

[13] J. Wetzel, A. Laubenheimer, M. Heizmann, "Joint probabilistic people detection in overlapping depth images," IEEE Access, vol. 8, pp. 28349 - 28359, 2020.

[14] T. Chavdarova ve F. Fleuret, "Deep multi-camera people detection," in Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA), Dec. 2017, pp. 848–853.

[15] M. A. Kızrak ve B. Bolat, "Deep and wide convolutional neural network model for highly dense crowd," in Proc. Innovations in Intelligent Systems and Applications Conference (ASYU), Nov. 2019.

[16] M. A. Kızrak ve B. Bolat, "Cluster-based monitoring and location estimation for crowd counting," in Proc. Intelligent Decision Science, Jan. 2020, pp. 240-253.

[17] S. Ali, M. Shah, "Floor fields for tracking in high density crowd scenes," 10th European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science, Volume 5303, 2008, pp 1-14, 2008.

[18] Y. Mao, J. Tong ve W. Xiang, "Estimation of crowd density using multi-local features and regression," in Proc. of the 8th World Congress on Intelligent Control and Automotion, 2010, pp 6295-6300.

[19] W. Ma, L. Huang ve C. Liu, "Crowd density analysis using co-occurrence texture features," in 5th International Conf. on Computer Sciences and Convergence Information Technology (ICCIT'10), 2010, pp 170-175.

[20] J. Guo, X. Wu, T. Cao, S. Yu ve Y. Xu, "Crowd density estimation via markov random field (MRF)," in Proc. of 8th World Congress on Intelligent Control and Automation, 2010, pp 258-263.

[21] W. Li, X. Wu, K. Matsumoto ve H. Zhao, "A new approach of crowd density estimation," in Proc. IEEE Region 10 Conference TENCON, 2010, pp 200-203.

[22] W. Li, X. Wu, K. Matsumoto ve H. Zhao, "Crowd density estimation: an ımproved approach," in Proc. IEEE 10th International Conference on Signal Processing (ICSP'10), 2010, pp 1213-1216.

[23] W. Ge ve R. T. Collins, "Crowd density analysis with marked point processes," in IEEE Signal Processing Magazine, vol. 27, pp 107-123, 2010.

[24] G. Kim, K. Eom, M. Kim ve J. Jung, "Automated measurement of crowd density based on edge detection and optical flow," in IEEE 2nd International Conference on Industrial Mechatronics and Automation, vol. 2, pp 553-556, 2010.

[25] W. Hsu, K. Lin ve C. Tsai, "Crowd density estimation based on frequency analysis," in Proc. 7th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2011, pp 348-351.

[26] G. Xiong, X. Wu, J. Cheng, Y. Chen, Y. Ou ve Y. Liu, "Crowd density estimation based on ımage potential energy model," in Proc. of the IEEE International Conference on Robotics and Biometrics (ROBIO),2011, pp 538-543.

[27] H. Yu, Z. He, Y. Liu ve L. Zhang, "A crowd flow estimation method based on dynamic texture and GRNN," in Proc. 7th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2012, pp 79-84.

[28] H. Yang ve H. Zhao, "A novel method for crowd density estimations," in Proc. IET International Conference on Information Science and Control Engineering (ICISCE), 2012, pp 1-4.

[29] V. B. Subburaman, A. Descamps ve C. Carincotte, "Counting people in the crowd using a generic head detector," in Proc. IEEE 9th International Conference on Advenced Video and Signal-Based Surveillance, 2012, pp 470-475.

[30] A. Chan ve N. Vasconcelos, “Counting people with low-level features and bayesion regression,” in IEEE Trans. on Image Processing, vol 21, No. 4, pp. 2160-2177, 2012.

[31] A. B. Chan ve N. Vasconcelos, “Modeling, clustering and segmenting video with mixtures of dynamic textures,” in IEEE Trans. on Pattern Analysis and Machine Intelligence, vol 30 (5), pp 909–926, 2008.

[32] A. B. Chan, Z. J. Liang ve N. Vasconcelos, “Privacy preserving crowd monitoring: counting people without people models or tracking,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR’08), 2008, pp 1–7.

[33] Z. Wang, H. Liu, Y. Qian ve T. Xu, “Crowd density estimation based on local binary pattern co-occurence matrix,” in Proc. IEEE International Conference on Multimedia and Expo Workshops (ICMEW), 2012, pp 372-377.

[34] H. Fradi ve J. Dugelay, “People counting system in crowded scenes based on feature regression,” in Proc. of the 20th European Signal Processing Conference (EUSIPCO), 2012, pp 136-140.

[35] H. Fradi ve J. Dugelay, “Crowd density map estimation based on features tracks,” in Proc. IEEE 15th International Workshop on Multimedia Signal Processing, 2013, pp 40-45.

[36] F. Tehranipour, R. Shishegar, S. Tehrenipour ve S. Seterehdan, “Attention control using fuzzy ınference system in monitoring cctv based on crowd density estimation,” in Proc. IEEE 8th Iranian Conference on Machine Vision and Image Processing (MVIP), 2013, pp 204-209.

[37] H. Fradi, X. Zhao ve J. Dugelay, “Crowd density analysis using subspace learning on local binary pattern,” in Proc. IEEE International Conference on Multimedia and Expo Workshops (ICMEW), 2013, pp 1-6.

[38] A. S. Rao, J. Gubbi, S. Marusic, P. Stanley ve M. Palaniswami, “Crowd density estimation based on optical flow and hierarchical clustering,” in Proc. IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2013, pp 494-499.

[39] Y. Yuan, J. Zhao ve C. Qui, “Estimating crowd density in an rf-based dynamic environment,” in IEEE Sensors Journal, vol. 13, no. 10, pp 3837-3845, 2013.

[40] P. Karpagavalli ve A. V. Ramprasad, “Estimating the density of the people and counting the number of people in a crowd environment for human safety,” in Proc. IEEE International Conference on Communication and Signal Processing, 2013, pp. 663-667.

[41] Z. Wu, H. Zheng ve J. Wang, “Pedestrian counting based on crowd density estimation and Lucas-Kanade optical flow,” in Proc. IEEE 7th International Conference on Image and Graphics (ICIG), 2013, pp 471-476.

[42] K. Ping, P. Bo, Z. Wenying ve L. Shuai, “Research on central ıssues of crowd density estimation,” in Proc. 10th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2013, pp 143-145.

[43] Y. Yuan, “Crowd monitoring using mobile phones,” in IEEE 6th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), vol. 1, pp 261-264, 2014.

[44] M. Khansari, H. R. Rabiee, M. Asadi ve M. Ghanbari, “Object tracking in crowded video scenes based on the undecimated wavelet features and texture analysis,” Hindawi Publishing Corparation, EURASIP Journal on Advances in Signal Processing, pp18, 2008.

[45] M. Rodriguez, I. Laptev, J. Sivic ve J. Audibert, “Density-aware person detection and tracking in crowds,” in Proc. IEEE Internatinal Conference on Computer Vision, 2011, pp 2423-2430.

[46] M. Rodriguez, J. Sivic, I. Laptev ve J. Audibert, “Data-driven crowd analysis in videos,” in Proc. IEEE International Conference on Computer Vision, 2011, pp 1235-1242.

[47] D. Conte, P. Foggia, G. Percannella, F. Tufano ve M. Vento, “A method for counting moving people in video surveillance videos,” EURASIP Journal on Advances in Signal Processing, vol. (1), pp. 231-240, 2010.

[48] G. Antonini ve J-P. Thiran, “Counting pedestrians in video sequences using trajectory clustering,” in IEEE Trans. on Circuits and Systems for Video Technology, vol. 16(8), pp 1008–1020, 2006.

[49] E. L. Andrade, S. Blunsden ve R. B. Fisher, “Modelling crowd scenes for event detection,” in Proc. 18th International Conference on Pattern Recognition (ICPR’06), Volume 1, pp 175–178, 2006.

[50] X. Wang, X. Ma ve W. E. L. Grimson, “Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models,” in IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 31, no. 3, pp 539-555, 2009.

[51] C. C. Loy, T. Xiang ve S. Gong, “Multi-camera activity correlation analysis,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR’09), 2009, pp 1988-1995.

[52] R. Mehran, A. Oyama ve M. Shah, “Abnormal crowd behavior detection using social force model,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR’09), 2009, pp 935-942.

[53] T. Hospedales, S. Gong ve T. Xiang, “A markov clustering topic model for mining behaviour in video,” in Proc. IEEE 12th International Conference on Computer Vision, 2009, pp 1165–1172.

[54] S. Ali ve M. Shah, “A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR’07), 2007, pp 1–6.

[55] J. Liu, B. Kuipers ve S. Savarese,” Recognizing human actions by attributes,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR’11), 2011, pp 3337-3344.

[56] B. Zhou, X. Tang ve X. Wang, “Coherent filtering: detecting coherent motions from crowd clutters,” Conf.. Computer Vision (ECCV’12), 2012, pp 857-871.

[57] B. Zhou, X. Tang ve X. Wang, “Measuring crowd collectiveness,” in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR’13), 2013, pp 3049-3056.

[58] J. Shao, C. C. Loy ve X. Wang, “Scene-ındependent group profiling in crowd,” in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR’14), 2014, pp 2219-2226.

[59] K. Kang ve X. Wang, “Fully convolutional neural networks for crowd segmentation,” ArXiv preprint, arXiv:1411.4464, 2014.

[60] S. Yi, X. Wang, C. Lu ve J. Jia, “L0 regularized stationary time estimation for crowd group analysis,” in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR’14), 2014, pp 2211-2218.

[61] F. Zhu, X. Wang ve N. Yu, “Crowd tracking with dynamic evolution of group structures,” in European Conference on Computer Vision, Springer, 2014, pp 139-154.

[62] B. Zhou, X. Tang ve X. Wang, “Learning collective crowd behaviors with dynamic pedestrian-agents,” International Journal of Computer Vision, vol. 111(1), pp 50-68, 2015.

[63] S. Yi, H. Li ve X. Wang, “Understanding pedestrian behaviors from stationary crowd groups,” in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp 3488-3496.

[64] J. Shao, K. Kang, C. C. Loy ve X. Wang, “Deeply learned attributes for crowded sscene understanding,” in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR’15), 2015, pp 4657-4666.

[65] B. Zhou, X. Wang ve X. Tang, "Understanding collective crowd behaviors: learning a mixture model of dynamic pedestrian-agents,” in Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR’12), 2012, pp 2871-2878.

[66] N. Kumar, A. C. Berg, P. N. Belhumeur ve S. K. Nayar, “Attribute and simile classifiers for face verification,” in Proc. IEEE 12th International Conference on Computer Vision, 2009, pp 365–372.

[67] A. Farhadi, I. Endres, D. Hoiem ve D. Forsyth, “Describing objects by their attributes,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR’09), 2009, pp 1778–1785.

[68] C. H. Lampert, H. Nickisch ve S. Harmeling, “Learning to detect unseen object classes by between-class attribute transfer,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09), 2009, pp 951–958.

[69] T. L. Berg, A. C. Berg ve J. Shih, “Automatic attribute discovery and characterization from noisy web data.” in European Conference on Computer Vision, Springer, 2010, pp 663–676.

[70] Y. Fu, T. Hospedales, T. Xiang ve S. Gong, “Attribute learning for understanding unstructured social activity,” Computer Vision (ECCV’12), 2012, pp 530-543.

[71] G. Patterson ve J. Hays, “Sun attribute database: discovering, annotating, and recognizing scene attributes,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR’12), pp 2751-2758, 2012.

[72] A. Oliva ve A. Torralba, “Modeling the shape of the scene: a holistic representation of the spatial envelope,” International Journal of Computer Vision, vol. 42(3), pp 145–175, 2001.

[73] F-F. Li, I. Asha, K. Christof ve P. Pietro, “What do we perceive in a glance of a real-world scene?,” Journal of Vision, vol. 7(1), pp10-10, 2007.

[74] D. Parikh ve K. Grauman, “Interactively building a discriminative vocabulary of nameable attributes,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR’11), 2011, pp 1681-1688.

[75] F. Jiang, Y. Wu ve A. K. Katsaggelos, “Detecting contextual anomalies of crowd motion in surveillance video,” in Proc. 16th IEEE International Conference on Image Processing (ICIP’09), 2009, pp 1117-1120.

[76] L. Kratz ve K. Nishino, “Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR’09), 2009, pp 1446-1453.

[77] V. Mahadevan, W. Li, V. Bhalodia ve N. Vasconcelos, “Anomaly detection in crowded scenes,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR’10), 2010, pp 1975-1981.

[78] V. Reddy, C. Sanderson ve B. C. Lovell, “Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and textures,” in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2011, pp 55-61.

[79] V. Saligrama ve Z. Chen, “Video anomaly detection based on local statistical aggregates,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR’12), 2012, pp 2112-2119.

[80] L. Cao ve K. Huang, “Video-based crowd density estimation and prediction system for wide-area surveillance,” Future Video Technology, China Communications, vol. 10, no.5, pp. 79-88, 2013.

[81] C. Zhang, H. Li, X. Wang, ve X. Yang, “Cross-scene crowd counting via deep convolutional neural networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 833–841.

[82] Y. Zhang, D. Zhou, S. Chen, S. Gao, ve Y. Ma, “Single-image crowd counting via multi-column convolutional neural network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 589–597.

[83] D. B. Sam, S. Surya, ve R. V. Babu, “Switching convolutional neural network for crowd counting,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1, Jul. 2017, pp. 4031–4039.

[84] V. A. Sindagi ve V. M. Patel, “Generating high-quality crowd density maps using contextual pyramid CNNs,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 1861–1870.

[85] Z. Shi, L. Zhang, Y. Liu, X. Cao, Y. Ye, M.-M. Cheng, ve G. Zheng, "Crowd counting with deep negative correlation learning," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 5382–5390.

[86] L. Zhu, C. Li, Z. Yang, K. Yuan, ve S. Wang, "Crowd density estimation based on classification activation map and patch density level," in Neural Computing & Applications. Berlin, Germany: Springer, 2019, pp. 5105–5116.

[87] C. Wang, H. Zhang, L. Yang, S. Liu, ve X. Cao, "Deep people counting in extremely dense crowds," in ACM International Conference on Multimedia, pp. 1299–1302, 2015.

[88] M. Fu, P. Xu, X. Li, Q. Liu, M. Ye, ve C. Zhu, "Fast crowd density estimation with convolutional neural networks," Engineering Applications of Artificial Intelligence, vol. 43, pp. 81–88, 2015.

[89] K. He, X. Zhang, S. Ren, ve J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016.

[90] C. Shang, H. Ai, ve B. Bai, "End-to-end crowd counting via joint learning local and global count," in IEEE Int. Conf. on Image Process., pp. 1215–1219, 2016.

[91] L. Boominathan, S. S. S. Kruthiventi, ve R. V. Babu, "Crowdnet: A deep convolutional network for dense crowd counting," in ACM International Conference on Multimedia, 2016, pp. 640–644.

[92] C. Arteta, V. Lempitsky, ve A. Zisserman, "Counting in the wild," in European Conference on Computer Vision, 2016, pp. 483–498.

[93] T. N. Mundhenk, G. Konjevod, W. A. Sakla, ve K. Boakye, "A large contextual dataset for classification, detection and counting of cars with deep learning," in European Conference on Computer Vision, 2016, pp. 785–800.

[94] Z. Zhao, H. Li, R. Zhao, ve X. Wang, "Crossing-line crowd counting with two-phase deep neural networks," in European Conference on Computer Vision, 2016, pp. 712–726.

[95] V. A. Sindagi ve V. M. Patel, "Cnn-based cascaded multi-task learning of high-level prior and density estimation for crowd counting," in IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance, Italy, 2017.

[96] D. Kang, Z. Ma, ve A. B. Chan, "Beyond counting: Comparisons of density maps for crowd analysis tasks - counting, detection, and tracking," in IEEE Trans. on Circuits and Systems for Video Technology, vol. 29, no. 5, pp. 99, May 2019,

[97] N. Ilyas, A. Ahmad, K. Kim, "CASA-Crowd: a context-aware scale aggregation cnn-based crowd counting technique," in IEEE Access, vol. 7, pp. 182050 - 182059, 2019.

[98] Y. Li, X. Zhang, ve D. Chen, "CSRNet: dilated convolutional neural networks for understanding the highly congested scenes," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 1091–1100.

[99] M. Shi, Z. Yang, C. Xu, ve Q. Chen, "Revisiting perspective information for efficient crowd counting," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 7279–7288.

[100] J. Wan, W. Luo, B. Wu, A. B. Chan, ve W. Liu, "Residual regression with semantic prior for crowd counting," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2019, pp. 4036–4045.

[101] F. Dai, H. Liu, Y. Ma, J. Cao, Q. Zhao, ve Y. Zhang, "Dense scale network for crowd counting," 2019, arXiv:1906.09707. [Online

[102] X. Jiang, L. Zhang, P. Lv, Y. Guo, R. Zhu, Y. Li, Y. Pang, X. Li, B. Zhou, ve M. Xu, "Learning multi-level density maps for crowd counting," in IEEE Trans. Neural Netw. Learn. Syst., vol. 31, no. 8, pp. 2705–2715, Aug. 2020.

[103] D. Oñoro-Rubio ve R. J. López-Sastre, "Towards perspective-free object counting with deep learning," in Proc. Eur. Conf. Comput. Vis. (ECCV), Springer, Amsterdam, Netherlands, 2016, pp. 615–629.

[104] M. Marsden, K. McGuinness, S. Little, ve N. E. O'Connor, "Resnetcrowd: A residual deep learning architecture for crowd counting, violent behaviour detection and crowd density level classification," in IEEE International Conference on Advanced Video and Signal-Based Surveillance, p. 7, 2017.

[105] B. Yang, J. Cao, N. Wang, Y. Zhang, ve L. Zou, "Counting challenging crowds robustly using a multi-column multi-task convolutional neural network," in Signal Process., Image Commun., vol. 64, pp. 118–129, May 2018.

[106] Y. Tian, Y. Lei, J. Zhang, ve J. Z. Wang, "PaDNet: Pan-density crowd counting," in IEEE Trans. on Image Proces., vol. 29, pp 2714 – 2727, 2019.

[107] J. Gao, Q. Wang, ve X. Li, "PCC Net: Perspective crowd counting via spatial convolutional network," in IEEE Trans. Circuits Syst. Video Technol., vol. 30, no. 10, pp. 3486–3498, Oct. 2020.

[108] D. B. Sam, S. V. Peri, A. Kamath, ve R. V. Babu, "Locate, size and count: Accurately resolving people in dense crowds via detection," 2019, arXiv:1906.07538. [Online

[109] D. Guo, K. Li, Z.-J. Zha, ve M. Wang, "DADNet: dilated-attention deformable ConvNet for crowd counting," in Proc. 27th ACM Int. Conf. Multimedia, Oct. 2019, pp. 1823–1832.

[110] Z. Shen, Y. Xu, B. Ni, M. Wang, J. Hu, ve X. Yang, "Crowd counting via adversarial cross-scale consistency pursuit," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 5245–5254.

[111] J. Liu, C. Gao, D. Meng, ve A. G. Hauptmann, "Decidenet: Counting varying density crowds through attention guided detection and density estimation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2018, pp. 5197–5206.

[112] M. Saqib, S. D. Khan, N. Sharma, ve M. Blumenstein, "Person head detection in multiple scales using deep convolutional neural networks," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2018, pp. 1–7.

[113] M. Saqib, S. D. Khan, N. Sharma, ve M. Blumenstein, "Crowd counting in low-resolution crowded scenes using region-based deep convolutional neural networks," in IEEE Access, vol. 7, pp. 35317–35329, 2019.

[114] M. Shami, S. Maqbool, H. Sajid, Y. Ayaz, ve S.-C. S. Cheung, "People counting in dense crowd images using sparse head detections," in IEEE Trans. Circuits Syst. Video Technol., vol. 29, no. 9, pp. 2627- 2636, 2019.

[115] S. Basalamah, S. D. Khan, H. Ullah, "Scale Driven Convolutional Neural Network Model for People Counting and Localization in Crowd Scenes," in IEEE Access, vol. 7, pp. 71576 - 71584, 2019.

[116] J. Hu, L. Shen, ve G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7132–7141.

[117] S. Woo, J. Park, J.-Y. Lee, ve I. S. Kweon, "CBAM: Convolutional block attention module," in Proc. Eur. Conf. Comput. Vis. (ECCV), Munich, Germany, Sep. 2018, pp. 3–19.

[118] J. Park, S. Woo, J.-Y. Lee, ve I. So Kweon, "BAM: Bottleneck attention module," 2018, arXiv:1807.06514. [Online

[119] J. U. Kim ve Y. Man Ro, "Attentive layer separation for object classification and object localization in object detection," in Proc. IEEE Int. Conf. Image Process. (ICIP), Sep. 2019, pp. 3995–3999.

[120] M. Shaikh, V. A. Kollerathu, ve G. Krishnamurthi, "Recurrent attention mechanism networks for enhanced classification of biomedical images," in Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI), Apr. 2019, pp. 1260–1264.

[121] Z. Lan, Q. Huang, F. Chen, ve Y. Meng, "Aerial image semantic segmentation using spatial and channel attention," in Proc. IEEE 4th Int. Conf. Image, Vis. Comput. (ICIVC), Jul. 2019, pp. 316–320.

[122] H. Ling, J. Wu, L. Wu, J. Huang, J. Chen, ve P. Li, "Self-residual attention network for deep face recognition," in IEEE Access, vol. 7, pp. 55159–55168, 2019.

[123] X. Wang, R. Girshick, A. Gupta, ve K. He, "Non-local neural networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7794–7803.

[124] X. Li, W. Wang, X. Hu, ve J. Yang, "Selective kernel networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 510–519.

[125] Y. Cao, J. Xu, S. Lin, F. Wei, ve H. Hu, "GCNet: Non-local networks meet squeeze-excitation networks and beyond," in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW), Oct. 2019, pp. 1971–1980.

[126] L. Zhu, Z. Zhao, C. Lu, Y. Lin, Y. Peng, ve T. Yao, "Dual path multi-scale fusion networks with attention for crowd counting," 2019, arXiv:1902.01115. [Online

[127] Z. Huo, B. Lu, A. Mi, F. Luo, Y. Qiao, "Learning multi-level features to improve crowd counting," in IEEE Access, vol. 8, pp. 211391 - 2114006, 2020.

[128] L. Cosmides ve J. Tooby, "Evolutionary psychology: A primer," 1997, Erişim tarihi: Eylül 2019, https://www.cep.ucsb.edu/primer.html

[129] Veri kalitesi ve yapay zekâ - temel hakları korumak için önyargı ve hataların azaltılması, "Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights," Erişim tarihi: Mart 2021,

https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-data-quality-and-ai_en.pdf

[130] Yapay zekâ sistemi ve veri yaşam döngüsü, (The AI system and data life cycle), OECD, Erişim tarihi: Mart 2021, https://www.oecd-ilibrary.org/sites/8b303b6f-en/index.html?itemId=/content/component/8b303b6f-en

[131] A. M. Turing, "Computing machinery and intelligence," Mind, vol. 59, pp. 43-460, 1950.

[132] A. Newell, J. C. Shaw ve H. A. Simon, "Empirical explorations of the logic theory machine: a case study in heuristic," IRE-AIEE-ACM, 1957, Erişim tarihi: Eylül 2019, http://digitalcollections.library.cmu.edu/awweb/awarchive?type=file&item=354740

[133] M. A. Kızrak, "Şu kara kutuyu açalım: yapay sinir ağları," Erişim tarihi: Eylül 2019, https://medium.com/@ayyucekizrak/%C5%9Fu-kara-kutuyu-a%C3%A7alim-yapay-sinir-a%C4%9Flar%C4%B1-7b65c6a5264a

[134] F. F. Li, J. Johnson ve S. Young, "CS231n: Convolutional neural networks for visual recognition," Stanford Üniversitesi, 2017, Erişim tarihi: Eylül 2019, http://cs231n.github.io/neural-networks-1/

[135] E. Alpaydın, E., "Yapay öğrenme," İstanbul: Boğaziçi Üniversitesi Yayınevi, 2018.

[136] I. Goodfellow, Y. Bengio ve A. Courville, "Derin öğrenme," (Çeviri: Yarman Vural, F., Cinbiş, R. G., Kalkan, S.), Ankara: Buzdağı Yayınevi, 2018.

[137] A. Amidi ve S. Amidi, "CS229: Makine öğrenmesi," (Çeviri: M. A. Kızrak, B. Buluz ve Y. Kömeçoğlu), Erişim tarihi: Ekim 2019, https://stanford.edu/~shervine/l/tr/teaching/cs-229/cheatsheet-supervised-learning

[138] D. E. Rumelhart, G. E. Hinton ve R. J. Williams, "Learning representations by back-propagating errors," Nature, vol. 323, pp. 533–536, 1986.

[139] G. E. Hinton, S. Osindero ve Y. Teh, "A fast learning algorithm for deep belief nets," Neural Computation, vol.18, pp. 1527-1554, 2006.

[140] D. H. Hubel ve T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," The Journal of Physicology, vol. 160(1),pp. 106-154, 1962.

[141] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position", Biological Cybernetics, vol. 36(4), pp. 193–202, 1980.

[142] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard ve L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural Computation, vol. 1(4), pp. 541-551, 1989.

[143] A. Krizhevsky, I. Sutskever ve G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, vol. 25(1), pp. 1097-1105, 2012.

[144] F. Chollet, “Python ile derin öğrenme,” (Çeviri: Aksoy, B.), Ankara: Buzdağı Yayınevi, 2019.

[145] S. Sabour, N. Frosst, and G. E. Hinton, “Dynamic routing between capsules,” in Proc. 31st Conf. on Neural Information Proces. Systems (NIPS), 2017, pp. 3859–3869.

[146] A. Géron, “Introduction to capsules”, Erişim tarihi: Şubat 2018, https://www.oreilly.com/content/introducing-capsule-networks/

[147] H. Idrees, M. Tayyab, K. Athrey, D. Zhang, S. Al-Maadeed, N. Rajpoot, “Composition loss for counting, density map estimation and localization in dense crowds,” in Proc. of the European Conference on Computer Vision (ECCV), Munich, Germanypp, 8–14 September, 2018, pp. 532–546.

[148] K. Chen, S. Gong, T. Xiang, C. C. Loy, “Cumulative attribute space for age and crowd density estimation,” in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June, 2013, pp. 2467–2474.

[149] H. Idrees, I. Saleemi, C. Seibert, M. Shah, “Multi-source multi-scale counting in extremely dense crowd images,” in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Portland, OR, USA, 23-28 June 2013, pp. 2547–2554.

[150] J. Ferryman ve A. Shahrokni, “PETS2009: dataset and challenge,” in Proc., Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, (2009).

[151] V D Sindagi, V M Patel, “Survey of recent advances in cnn-based single ımage crowd counting and density estimation,” Pattern Recognition Letters, Elsevier, 2017.

[152] J. Choi, D. Chun, H. Kim ve H-J. Lee, “GaussianYOLOv3: an accurate and fast object detector using localization uncertainty for autonomous driving,” International Conference on Computer Vision (ICCV), 2019, Erişim tarihi: Mayıs 2020, https://arxiv.org/abs/1904.04620

[153] P. Cunningham ve S. J. Delany, “K-nearest neighbour classifiers,” Technical Report UCD-CSI-vol. 4 Mart 27, 2007.

[154] H. H. Nguyen, J. Yamagishi ve I, Echizen, “Capsule-forensics: using capsule networks to detect forged images and videos,” in Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), pp. 2301–2307, 2019.

[155] A. M. Algamdi, V. Sanchez, and C. T. Li, “Learning temporal information from spatial information using CapsNets for human action recognition,” in IEEE Int. Conf. Acoust. Speech Signal Process, pp 3867–3871, 2019.

[156] S. Liu, W. Deng, “Very deep convolutional neural network based image classification using small training sample size,” in 3rd IAPR Asian Conference on Pattern Recognition (ACPR), doi:10.1109/ACPR.2015.7486599, 2015.

[157] M. A. Kızrak ve B. Bolat, “Crowd density estimation by using attention based capsule network and multi column CNN”, IEEE Access, vol. , pp. -, 2021.

[158] S. Kumagai, K. Hotta, T. Kurita, "Mixture of counting CNNs: Adaptive integration of CNNs specialized to specific appearance for crowd counting", ArXiv preprint, arXiv:1703.09393, 2017.

[159] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition", in Proc.3rd International Conference on Learning Representations, ICLR 2015, pp. 1-14, 2015

[160] M. Yang, L. Rashidi, A. S. Rao, S. Rajasegarar, M. Ganji, M. Palaniswami ve C. Leckie, "Cluster-based crowd movement behavior detection", in Proc. of the Digital Image Computing: Techniques and Applications (DICTA), Australia, pp. 1–8, 2018.

[161] V. Badrinarayanan, A. Kendall, R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation", in IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481-2495, Jan. 2017.

[162] G. Huang, Z. Liu, L. V. Maaten, K. Q. Weinberger, "Densely connected convolutional networks", in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., USA, 21–26 July 2017, pp. 2261 – 2269.

[163] W. Liu, M. Salzmann, P. Fua, "Context-aware crowd counting", in IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), CA, USA, 2019.

[164] C. Xu, K. Qiu, J. Fu, S. Bai, Y. Xu, X. Bai, "Learn to scale: generating multipolar normalized density maps for crowd counting", in Proc. IEEE Int. Conf. Comput. Vis, Seoul, Korea, 2019, pp. 8382–8390.

# TEZDEN ÜRETİLMİŞ YAYINLAR

## Konferans Bildirileri

1. M. A. Kızrak ve B. Bolat, "A Novel Approach for People Counting and Tracking from Crowd Video," 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA), DOI: 10.1109/INISTA.2017.8001170

2. M. A. Kızrak ve B.Bolat, "Deep and Wide Convolutional Neural Network Model for Highly Dense Crowd," 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), DOI: 10.1109/ASYU48272.2019.8946395

3. M. A. Kızrak ve B. Bolat, "Cluster-Based Monitoring and Location Estimation for Crowd Counting," in Proc. Intelligent Decision Science, Jan. 2020, pp. 240-253.

## Dergi

1. M. A. Kızrak ve B. Bolat, "Derin Öğrenme ile Kalabalık Analizi Üzerine Detaylı Bir Araştırma," Bilişim Teknolojileri Dergisi, Cilt 11, Sayı 3, Yıl 2018, Sayfalar 263 – 286, DOI: 10.17671/gazibtd.419205

2. M. A. Kızrak ve B. Bolat, "Crowd Density Estimation by Using Attention Based Capsule Network and Multi Column CNN", IEEE Access, vol. 9, pp. 75435 - 75445, DOI: 10.1109/ACCESS.2021.3081529, 2021.