



**GÖZETİM SİSTEMLERİ İÇİN NORMAL OLMAYAN DAVRANIŞLARIN
ALGILANMASI**

Elvan DUMAN

**DOKTORA TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

OCAK 2020

Elvan DUMAN tarafından hazırlanan “GÖZETİM SİSTEMLERİ İÇİN NORMAL OLMAYAN DAVRANIŞLARIN ALGILANMASI” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Bilgisayar Mühendisliği Ana Bilim Dalında DOKTORA TEZİ olarak kabul edilmiştir.

Danışman: Prof. Dr. O. Ayhan ERDEM

Bilgisayar Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Başkan: Prof. Dr. İlyas ÇANKAYA

Elektrik Elektronik Mühendisliği Ana Bilim Dalı, Ankara Yıldırım
Beyazıt Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Üye: Doç. Dr. Necaattin BARIŞÇI

Bilgisayar Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Üye: Doç. Dr. Hasan Şakir BİLGE

Elektrik Elektronik Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Üye: Doç. Dr. Halil Murat ÜNVER

Bilgisayar Mühendisliği Ana Bilim Dalı, Kırıkkale Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Tez Savunma Tarihi: 24/01/2020

Jüri tarafından kabul edilen bu tezin Doktora Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Prof. Dr. Sena YAŞYERLİ
Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Elvan DUMAN

24/01/2020

GÖZETİM SİSTEMLERİ İÇİN NORMAL OLMAYAN DAVRANIŞLARIN ALGILANMASI

(Doktora Tezi)

Elvan DUMAN

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Ocak 2020

ÖZET

Günümüzde havaalanları, hastaneler, şehir merkezleri gibi kamuya açık alanlar gözetim sistemleri tarafından görüntülenmektedir. Gözetim sistemlerinin yaygın kullanımı ise güvenlik kaygılarını azaltırken, bununla birlikte gerçek zamanlı olarak insanlar tarafından incelenemeyecek boyutlarda video verisi oluşturmaktadır. Bu sebeple videolarda yer alan eylemlerin otomatik olarak analiz edilmesi güvenlik kamera sistemlerinin standartlarını yükseltmektedir. Tez çalışmasında, videolarda gözetimsiz bir şekilde normal olmayan eylem tespiti yapabilmek için Evrişimsel otokodlayıcı (Convolutional Autoencoder) ve Evrişimsel uzun-kısa süreli bellek (Convolutional Long Short-Term Memory) kullanan bir çerçeve (OF-ConvAE-LSTM) geliştirilmiştir. Derin öğrenme modelinin yanı sıra çerçevede optik akış yöntemi video ön plan nesnelerinin hareket ve hız bilgilerinin elde edilmesi için uygulanmıştır. Deneyler iyi bilinen üç veri kümesi olan Avenue, UCSD Ped1 ve UCSD Ped2 üzerinde gerçekleştirilmiştir. Deneysel sonuçlar, önerilen çerçevenin düzenli hareket değişimlerinin karmaşık dağılımını yüksek doğrulukla modellediğini göstermiştir. Ayrıca, önerilen yöntemin gözetimsiz ve yarı gözetimli derin öğrenme modellerine dayanan son teknoloji yaklaşımlardan eğri altındaki alan (EAA) metriğine göre daha iyi performans gösterdiği gözlemlenmiştir.

Bilim Kodu : 92431

Anahtar Kelimeler : Normal olmayan davranış tespiti, Evrişimsel otokodlayıcı, Evrişimsel uzun kısa süreli bellek, Optik akış

Sayfa Adedi : 80

Danışman : Prof. Dr. O. Ayhan ERDEM

ABNORMAL HUMAN BEHAVIOR RECOGNITION FOR SURVEILLANCE
SYSTEMS

(Ph. D. Thesis)

Elvan DUMAN

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

January 2020

ABSTRACT

Today, public areas, such as airports, hospitals, city centers are monitored by surveillance systems. The widespread use of surveillance systems reduces security concerns while creating an amount of video data that cannot be examined by people in real-time. Therefore, the concept of automatic understanding of video activities has raised the standards of security camera systems. In this thesis, a framework (OF-ConvAE-LSTM) is proposed to detect anomalies using Convolutional Autoencoder and Convolutional Long Short-Term Memory in an unsupervised manner. Besides the deep learning model, the feature extraction stage based on dense optical flow is applied in the framework to obtain the velocity and direction information of foreground objects. The experiments were carried out on three popular public datasets consisting of Avenue, UCSD Ped1, and UCSD Ped2. The experimental results have shown that the proposed framework models the complex distribution of the pattern of regular motion changes with high accuracy. Besides, this method was observed to outperform in terms of Area Under The Curve (AUC) values against state-of-the-art approaches based on unsupervised and semi-supervised deep learning models.

Science Code : 92431

Key Words : Abnormal behavior detection, Convolutional autoencoder,
Convolutional long short-term memory, Optical flow

Page Number : 80

Supervisor : Prof. Dr. O. Ayhan ERDEM

TEŞEKKÜR

Doktora çalışmalarım süresince, her türlü destekte bulunan, kritik noktalarda bana doğru yolu gösterip ufkumu açan ve elde ettiğim başarılarda büyük pay sahibi olan danışman hocam Prof. Dr. O. Ayhan ERDEM'e; ihtiyacım olduğu her an bana bilgi ve fikirlerini esirgemeyen başta tez izleme komitemde yer alan değerli hocalarım Doç. Dr. Necaattin BARIŞCI ve Doç. Dr. Hasan Şakir BİLGE hocalarıma; yoğun çalışmalarım boyunca bana sabır ve anlayış gösteren, maddi manevi her zaman yanımda olarak yükümü hafifleten ve her daim bana moral ve motivasyon kaynağı olan sevgili eşim Ebru AYDOĞAN DUMAN'a, canım kızım Beril'e ve her zaman yanımda olduklarını hissettiren ve bu günlere gelmemde büyük emeği olan başta annem Ayfer DUMAN olmak üzere aileme gönülden sonsuz teşekkür ederim.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	x
SİMGELER VE KISALTMALAR.....	xii
1. GİRİŞ.....	1
2. VİDEOLARDA İNSAN EYLEM ANALİZİ YAKLAŞIMLARI.....	7
2.1. Tek Katmanlı Yaklaşımlar	7
2.1.1. Uzay-zaman hacmi yaklaşımları	7
2.1.2. Ardışık yaklaşımlar	11
2.2. Hiyerarşik Yaklaşımlar	13
2.2.1. İstatistiksel yaklaşımlar	13
2.2.2. Dizilimsel yaklaşımlar	15
2.2.3. Tanımlama temelli yaklaşımlar.....	15
2.3. Videolarda Normal Olmayan Eylem Tespit Yaklaşımları	16
2.3.1. Davranış gösterimi	16
2.3.2. Normal olmayan eylem tespit yöntemleri	19
2.3.3. Derin öğrenme yöntemleri	21
2.4. Ulusal Tez Çalışmaları.....	22
3. MATERYAL VE YÖNTEMLER.....	29
3.1. Derin Öğrenme.....	29

	Sayfa
3.1.1. İleri beslemeli sinir ağları.....	31
3.1.2. Evrişimsel sinir ağları	34
3.1.3. Özyineli sinir ağları.....	37
3.2. Uzun Kısa Süreli Bellek.....	38
3.3. Evrişimsel Uzun Kısa Süreli Bellek.....	40
3.4. Otokodlayıcılar.....	40
3.5. Evrişimsel Otokodlayıcılar.....	42
3.6. Optik Akış	43
3.6.1. Farneback yöntemi	46
4. NORMAL OLMAYAN DAVRANIŞ ALGILANMASI	49
4.1. Önişleme Aşaması.....	50
4.2. Derin Öğrenme Modeli	50
4.3. Düzenlilik Skoru	51
4.4. Deneysel Sonuçlar.....	52
4.4.1. Veri kümeleri	52
4.4.2. Uygulama detayları	57
4.4.3. Performans metrikleri.....	57
4.4.4. Deneysel sonuçlar	58
5. SONUÇ VE ÖNERİLER	65
KAYNAKLAR	67
ÖZGEÇMİŞ	79

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. Davranış sunum yöntemleri ve özellikleri	17
Çizelge 2.2. Normal olmayan eylem algılamada kullanılan sınıflama yöntemleri.....	19
Çizelge 2.3. Normal olmayan eylem tespiti için kullanılan yöntemler.....	21
Çizelge 2.4. Ulusal tez çalışmaları ve detaylı incelemesi	27
Çizelge 4.1. Önerilen yöntemin çerçeve seviyesinde performans karşılaştırması.....	61
Çizelge 4.2. Önerilen modelin çerçeve seviyesindeki performans sonuçları	62
Çizelge 4.3. Önerilen modelin eylem seviyesinde performans karşılaştırması	62
Çizelge 4.4. Önileme ve test aşamaları çalışma zaman analizi	63

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Yumruklama eylemine ait XYT hacim çıkarım örneği	8
Şekil 2.2. Hareket hikâye resmi örnek gösterimi	9
Şekil 2.3. a) Yürüme eylemi eklem yörüngelerinin sunum örneği b) Sunum örneğinin uzay-zamanda gösterimi	10
Şekil 2.4. Doğrusal olmayan oluş hızlarına göre ayak açma dizisinin eşleme örneği	12
Şekil 2.5. Kol açma eylemi için SMM örneği	12
Şekil 2.6. Yumruklama eylemi için sıralı SMM örneği	14
Şekil 2.7. Normal olmayan eylem tespiti için sistem aşamaları	16
Şekil 2.8. İlgi noktaları ve yerel alan gösterimleri için örnek durumlar	18
Şekil 3.1. Bir gizli katmana sahip ileri beslemeli sinir ağı örneği	31
Şekil 3.2. İki gizli katmana sahip ileri beslemeli sinir ağı örneği.....	32
Şekil 3.3. LeNet evrışimsel sinir ağı mimarisi.....	34
Şekil 3.4. Evrişim katmanında filtre hesaplama örneği	36
Şekil 3.5. Evrişimsel sinir ağında 2x2 boyutunda filtre ile maksimum örneklendirme..	36
Şekil 3.6. (a) Geleneksel sinir ağı mimarisi örneği (b) Seyretilme uygulanmış sinir ağı örneği	37
Şekil 3.7. Özyineli sinir ağı mimarisi	38
Şekil 3.8. UKSB biriminin mimarisi.	39
Şekil 3.9. Geleneksel otokodlayıcı mimarisi	41
Şekil 3.10. Evrişimsel otokodlayıcı genel mimarisi	43
Şekil 3.11. Hareket eden noktanın şiddet değişmezliği gösterimi	44
Şekil 3.12. Hız bileşenlerinin gösterilmesi	44
Şekil 3.13. İki ardışık video çerçevesinin komşuluk polinomları (a,b)	46
Şekil 3.14. Vektör alanı ve optik akış renklendirme şeması.....	47

Şekil	Sayfa
Şekil 4.1. Önerilen yöntem mimarisi: Önışleme ve eğitim aşamaları	49
Şekil 4.2. Önerilen yöntemin detaylı evrişimsel ağ mimarisi.....	51
Şekil 4.3. Avenue veri kümesi düzensiz eylem örnekleri	54
Şekil 4.4. UCSD Ped1 veri kümesi düzensiz eylem örnekleri.....	55
Şekil 4.5. UCSD Ped2 veri kümesi düzensiz eylem örnekleri.....	56
Şekil 4.6. Model girişinin örnek gösterimi: Sekiz ardışık çerçeve ve karşılık gelen optik akış haritaları.	59
Şekil 4.7. Video 4 için düzenlilik skor örneği	60

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Kısaltmalar

Açıklamalar

AÖM	Aşırı Öğrenme Makinesi (Extreme Learning Machine)
DAA	Doğrusal Ayrımcılık Analizi (Linear Discriminant Analysis)
DBA	Dinamik Bayes Ağları (Dynamic Bayesian Networks)
DDK	Dinamik Doku Karışımı (Mixture of Dynamic Texture)
DGKM	Derin Gauss Karışım Modeli (Deep Gaussian Mixture Model)
DVM	Destek Vektör Makinesi (Support Vector Machine)
DZB	Dinamik Zaman Bükme (Dynamic Time Warping)
EAA	Eğri Altındaki Alan (Area Under the Curve)
EAK	En Büyük Artçıl Kestirimi (Maximum a Posteriori Probability)
EOK	En Büyük olabilirlik kestirimi (Maximum Likelihood Estimation)
EOK	Evrişimsel Otokodlayıcı (Convolutional Autoencoder)
ESA	Evrişimsel Sinir Ağı (Convolutional Neural Network)
GGO	Gürültü Giderici Otokodlayıcı (Denoising Autoencoder)
GKM	Gauss Karışım Modeli (Gaussian Mixture Model)
GSEM	Gri Seviyeli Eş-oluşum Matrisleri (Gray Level Co-occurrence Matrix)
HEG	Hareket Enerji Görüntüsü (Motion Energy Image)
HGG	Hareket Geçmiş Görüntüsü (Motion History Image)
İBD	İçerikten Bağımsız Dil (Context-free grammar)
KEYK	K En Yakın Komşu (K-Nearest Neighbor)
KÇ	Kelime Çantası (Bag of Words)
MRA	Markov Rastgele Alanı (Markov Random Field)
OA	Optik Akış (Optical Flow)
OAİ	Optik Akış Histogramı (Histogram of Optical Flow)
OAGM	Optik Akış Gauss Modeli (Gaussian Model of Optical Flow)

Kısaltmalar**Açıklamalar**

OK	Otokodlayıcı (Autoencoder)
ÖBÖD	Ölçek Bağımsız Özellik Dönüşümü (Scale-invariant feature transform)
ÖSA	Özyineli Sinir Ağı (Recurrent Neural Network)
PSO	Parçacık Sürü Optimizasyonu (Particle Swarm Optimization)
RDB	Rektifiye Doğrusal Birim (Rectified Linear Unit)
SA	Sinir Ağları (Neural Networks)
SİBD	Stokastik İçerikten Bağımsız Dil (Stochastic Context-Free Grammar)
SMM	Saklı Markov Modeli (Hidden Markov Model)
TBA	Temel Bileşen Analizi (Principal Component Analysis)
UKSB	Uzun Kısa Süreli Bellek (Long Short Term Memory)
UZİN	Uzay-zaman İlgi Noktaları (Space-time Interest Points)
UZH	Uzaysal-Zaman Hacim (Spatio-Temporal Volume)
UZY	Uzay Zaman Yörüngeleri (Space-time Trajectories)
YGH	Yönlü Gradyan Histogramı (Histogram of Oriented Gradients)

1. GİRİŞ

Bilgi, insan hayatı için her zaman en önemli unsurlardan birisi olmuştur. İnsanoğlu tarih boyunca yeni bilgiler edinmek, kaydetmek, saklamak ve istenildiğinde hızlı ulaşabilmek için yöntem ve teknikler geliştirmiştir. Bunun yanında teknolojinin gelişimi bilginin değerini artırırken; bilgi edinme, arşivleme ve bilgiye ulaşma yöntemlerinde kökten değişiklikler getirmiştir. Özellikle bilgisayar sistemleri ve internetin yaygınlaşması ile veri saklama işlemi ucuzlamış; veriye erişim süresi kısalmıştır. Günümüzde, yaşamımıza dair tüm eylemler sayısal ortama taşınmış, kayıt altına alınmaya başlanmış ve bunun sonucu olarak hızla büyüyen devasa veri yığınları oluşmuştur. Araştırmalara göre birkaç yılda üretilen veri miktarı, daha önce üretilen tüm verinin iki katına çıkmaktadır. Bu durum sahip olduğumuz verinin üstel şekilde arttığını göstermektedir [1]. Verinin bu denli hızlı artışı; işlenmesi, gizli ve değerli bilgilerin çıkarılması, özetlenmesi ve sınıflandırılması problemlerini öne çıkarmış ve verinin işlenmesini karmaşık hale getirmiştir.

Video gözetim sistemleri, kameraların görüntü sinyallerini bir merkeze ileten sistemlerdir. Bu tür sistemler sinyalleri taşıırken herkese açık sistemde değil, belirli protokollere bağlı güvenli iletim sağlamaktadır. Belirli alanın görüntüsü güvenlik amacıyla merkezi birimlere ve monitörlere ulaştırılmaktadır. Teknolojinin hızlı gelişimi ve Türkiye'nin teknolojik gelişmeleri yakından takip ederek günlük hayata yansıtması sebebiyle Türkiye'de kameralı izleme sistemlerinin kullanımı büyük artış göstermiştir. Gerek kamuya açık alanlarda ve devlet kurumlarında, gerekse özel teşebbüslerde varlıkların korunması amacıyla kameralı izleme sistemleri etkin şekilde kullanılmaktadır. Özellikle son yıllarda Türkiye'de güvenlik teşkilatları mobese izleme sistemini tüm illerde kritik bulunan noktalarda hızla yaygınlaştırmayı amaç edinmiştir.

İlk yıllarında video gözetim sistemleri anlık izleme ve olay gerçekleşikten sonra kamera kayıtlarının incelenmesi amacıyla kullanırken, günümüzde olayların meydana gelmesini önleme ve olay gerçekleştiğinde tespit edip hızlı müdahale edebilme gibi amaçlarla da kullanılmaya başlanmıştır. Bu amaçları gerçekleştirebilmek için kritik öneme sahip alanların kamera kayıtları yetkili personel tarafından sürekli izlenmekte ve çeşitli yardımcı teknolojilerle desteklenmektedir. Ancak kamera kullanımının hızlı artışı, kamera kayıtlarının personel tarafından 7/24 anlık olarak incelenmesini imkânsız hale getirmiştir.

Yaygınlaşmaya başladığı ilk yıllarda sadece kritik alanların takibinde kullanılan güvenlik kamera sistemlerinin sayısı günümüzdeki milyonlarla ifade edilmekte ve neredeyse tüm halka açık alanları kayıt altına almaktadır. Kullanımındaki bu hızlı gelişim, video kayıt verisinin devasa boyutlara ulaşmasına neden olmaktadır. Bu sebeple, elde edilen veriyi saklama ve inceleme için gereken zaman konusunda yetersizlikler yaşanmaktadır. Hem kamera kayıtlarının daha sağlıklı izlenmesi hem de kameralar üzerinde değerlendirme yaparak insan gücünden ve zamandan tasarruf edebilmek için akıllı gözetim sistemlerine gereksinim duyulmaktadır. Bu yolla kamera kayıtlarının gerçek zamanlı ve otomatik incelenmesine imkân sağlanarak, önemli bilgilerin otomatik olarak çıkartılması mümkün olmaktadır.

Günümüzde, güvenlik kamera sistemlerinin kayıtları büyük çoğunlukla olay meydana geldikten sonra analiz ve ipuçlarını elde etmek için kullanılmaktadır. Görüntü kayıtlarının yapısı gereği yüksek kapasiteli bellek alanları gerektirmesinin yanı sıra görüntülerin çok büyük bir kısmı normal durumları içerdiğinden değersizdir. Tutulan kayıtların büyük bir kısmı saklama maliyeti ve yetersizliğinden dolayı kısa bir süre saklanmakta ve süresi dolan kayıtlar incelenmeden silinmektedir. Günümüzde gözetim sistemlerinde daha çok hareketli cisim tespiti, plaka tanıma, araç tipi tanıma, çizgi ihlali gibi belirli kuralı olan ve yoruma ihtiyaç duyulmayan durumlara çözüm getiren makine destekli uygulamalar yaygındır. Yüz tanıma, davranış tespiti, duygu analizi gibi yoruma açık, kuralları mantıksal ve matematiksel olarak ortaya açık şekilde çıkarılamayan problemler için uygulamalar oldukça yetersizdir. Tez çalışması kapsamında yapılan araştırmada normal olmayan eylem tespitinde kullanılan yöntem ve uygulamaların kısıtlılıklarının çokluğu nedeniyle hayata geçen örneklerin sayısının çok az olduğu tespit edilmiştir. Bu alandaki başlıca problemleri gerçek hayattaki çevre koşulları yerine kontrollü bir mekân kullanma zorunluluğu, kamera açısı ve uzaklığına göre yöntemin çalışmaması veya başarımın ciddi şekilde düşmesi, gölge ve örtüşme, yağış gibi problemleri çözememe ve eğitimde kullanılmamış senaryolarla karşılaşıldığında başarısız olunması olarak sıralayabiliriz. Ayrıca normal olmayan eylem sayısının çokluğu ve tanımının gözlem yapılan alana ve amaca göre değişiklik göstermesi de en önemli problemler arasında yer almaktadır. Sözü geçen bu kısıtlılıklar video kayıtlarında eylem analizi konusunda yapılan çalışmaların gerçek hayatta uygulanabilirliğini büyük ölçüde düşürmektedir.

Video anlamlandırma çalışmalarının uygulama alanındaki artan genişlik ile beraber, bilginin kaynağını oluşturan en önemli varlığın insan olduğu düşünülmektedir. Bu durum, insan eylemlerinin tanınmasını ve can ve mal güvenliğinin gözetim sistemleri ile takip edilmesini bir gereksinim haline getirmiştir. Artan veri miktarı ve kullanılan kameranın çok büyük sayılara gelmesi, video görüntülerinden anlam çıkarımının otomatikleşmesini zorunlu hale getiren diğer etmenlerdir. Tez çalışmasının bu yönüyle hem literatüre hem de Türkiye’de özellikle son yıllarda ihtiyaç duyulan güvenlik alanında yapılan çalışmalara katkı sağladığı düşünülmektedir.

Tez çalışma kapsamında yapılan literatür araştırması sonucunda elde edilen bilgiler ışığında, normal olmayan eylem tespiti üzerine yapılan çalışmaların çözmeye çalıştıkları problemler aşağıdaki gibi sıralanabilir:

- Normal olmayan eylem tanımının belirsiz olması. Çoğunlukla problem temelli normal olmayan eylem tanımı yapılmaktadır.
- Görsel davranışların karmaşıklığı ve kısıtlı mekânda pek çok çeşitliliğin bulunması.
- Gürültülü veri, düşük seviye özellik seçimi ve sunumu gibi faktörler sınıflandırıcı performansını etkilemesi.
- Video kalitesi, gölgeler, görüntü kapanması, ışıklandırma, hareketli kamera ve karmaşık arka plan gibi başarıya etki eden birçok etken bulunması
- Bazı normal olmayan eylemlerin insanlar tarafından normal davranışmış izlenimi içerisinde gerçekleştirilmesi veya aynı normal davranışın birçok farklı yapılaşının bulunması.

Normal olmayan eylemler üzerine yapılacak çalışmada karar verilmesi gereken önemli hususlar şu şekilde sıralanabilir:

- Normal olmayan eylem tanımı ve tipleri
- Öğrenme Çerçevesi: gözetimli, gözetimsiz ve yarı gözetimli
- Hareketli Hedefin Yoğunluğu: kalabalık bir sahne mi yoksa kalabalık olmayan bir sahne üzerinde mi çalışılacağı

- Kullanılacak veri kümesi
- Kamera sahne uzaklığı
- Kamera sahne açısı

Bilgisayarlı görü alanında video gözetim sistemleriyle davranış analizi, duygu analizi, yüz tanıma gibi kritik öneme sahip karmaşık problemlerin çözümleri günlük hayata yansması henüz yavaş yavaş olmaktadır. Güvenlik alanında yapılan çoğu görü çalışması hareketli cisim tespiti, plaka tanıma, araç tipi tanıma, çizgi ihlali gibi belirli kuralları olan yoruma ihtiyaç duyulmayan durumlar için kullanılmaktadır. İnsan aklının karmaşık yapısının ürünü olan çözümlerin bilgisayarlar tarafından da tespit edilmesi ve günlük hayatta uygulanabilirliği robotik ve insan-makine etkileşimi alanında büyük öneme sahiptir.

Tezin geri kalan kısmı ise şu şekilde oluşturulmuştur.

Bölüm 2, videolarda insan eylemlerinin analizinde kullanılan yöntemler hakkında detaylı bilgi sunmaktadır. Literatürde insan eylemlerini algılama için geliştirilen yaklaşımlar ve özelde normal olmayan eylemler için geliştirilmiş yaklaşımlar ayrı ayrı ele alınmış ve mümkün olduğunca kategorik ve anlamsal yakınlıklarına göre incelenmiştir. Mevcut yöntem ve çerçevelerin avantajlı ve dezavantajlı yönleri karşılaştırmalı olarak ortaya konulmuştur.

Bölüm 3, tez kapsamında uygulanan yöntem ve tekniklerin temelini ortaya koymaktadır. Yöntemlerin çalışma prensipleri ve hangi amaçla kullanıldıkları belirtilmiş sonrasında mimarileri açıklanmıştır. Önerilen yöntemin temelinde kullanılan optik akış özellik çıkarımı, otokodlayıcılar ve uzun kısa süreli bellek hakkında detaylı bilgi verilmiştir.

Bölüm 4, verilen temel bilgileri ve literatür ışığında önerilen yöntemin detaylarını sunmaktadır. Kullanılan özellik çıkarım yöntemi, geliştirilen derin öğrenme çerçevesi ve normal olmayan eylem tespitinde kullanılan skor yöntemi hakkında bilgiler verilmiştir. Ayrıca, çalışmanın diğer yöntemlerle karşılaştırılabilmesini mümkün kılan iyi bilinen veri kümelerini, uygulama detaylarını, değerlendirmede kullanılan performans metriklerini ve yöntemin birçok açıdan değerlendirmesini içermektedir.

Bölüm 5’de tez çalışmasında elde edilen bulgular ve ortaya koyulan etkinlik sonuçları incelenmiş, modern yöntemler ile karşılaştırılmıştır. Ayrıca, bu bölümde geleceğe yönelik araştırma önerileri sunulmuştur.





2. VİDEOLARDA İNSAN EYLEM ANALİZİ YAKLAŞIMLARI

Bu bölümde, videolarda insan eylem analizi için kullanılan yaklaşımlar anlamsal birliktelik gözetilerek incelenmiştir.

2.1. Tek Katmanlı Yaklaşımlar

Tek katmanlı yaklaşımlar insan eylemlerini doğrudan video verisinden algılamayı amaçlamışlardır. Bu tür yaklaşımlar bir eylemi resim dizi sınıfı olarak ele alır ve bilinmeyen resim dizisinden eylemi algılar. Tek katmanlı yöntemlerin birçoğu, tüm mümkün alt sıraları sınıflayan kayan pencereler tekniğini kullanır. Bu teknik, belli bir eylem şablonu algılanmak istenilen davranış eğitim kümesinde var ise etkili yöntemdir. Doğası gereği tek katmanlı yaklaşımlar yürüme, zıplama, el sallama gibi kısa ve ardışık hareketleri algılama amacıyla kullanılmaktadır [2].

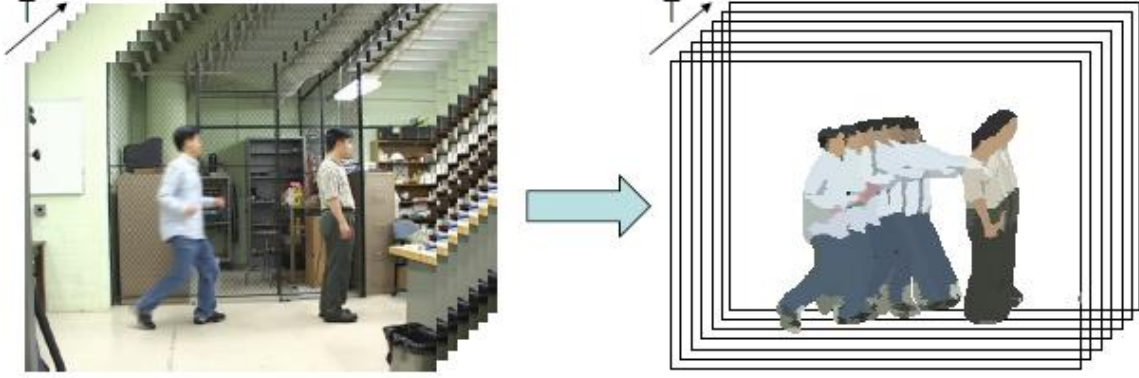
Tek katmanlı yaklaşımlar insan eylemlerini modelleme yöntemine göre ikiye ayrılırlar. Bunlar uzay-zaman hacmi (space-time volume) yaklaşımları ve ardışık yaklaşımlardır. Ardışık yaklaşımlar bir giriş videosunu gözlemler dizisi olarak ele alırken uzay-zaman hacmi yaklaşımları ise $3B$ (XYT) olarak ele alır. Uzay-zaman yaklaşımları kendi içinde $3B$ uzay-zaman hacmini, uzay-zaman hacmi, yörüngeler ve yerel ilgi noktası tanımlayıcıları özellikleri kullanımına göre üç ayrı kategoride incelemektedir.

2.1.1. Uzay-zaman hacmi yaklaşımları

Resim, üç boyutlu gerçek dünya sahnesinin iki boyutlu veriye formüle edilmesidir ve insan veya nesnelerin şekil ve görünüş gibi uzaysal özelliklerini içerir. Video ise bu iki boyutlu resimlerin zaman dizisi olarak sıralanmasıdır. Bu yüzden bir eylemin gerçekleştiği video verisi, $2B$ XY resimlerin zaman t boyunca sıralanmasıyla kendine özgü $3B$ uzay-zaman hacmi olarak oluşturulabilir [3].

Geleneksel bir uzay-zaman hacim yaklaşımı eğitim verisine dayanarak her bir eylemi temsil etmek için bir $3B$ uzay-zaman hacim modeli oluşturur [4]. Yeni bir video ele alındığında videoya ait $3B$ uzay-zaman hacmi oluşturulur ve var olan eylem modelleriyle

karşılaştırılarak benzerlik oranı en yüksek olan sınıfa dâhil edilir. Şekil 2.1’de yumruklama eylemine ait 3B hacim örneği gösterilmektedir.



Şekil 2.1. Yumruklama eylemine ait XYT hacim çıkarım örneği [5]

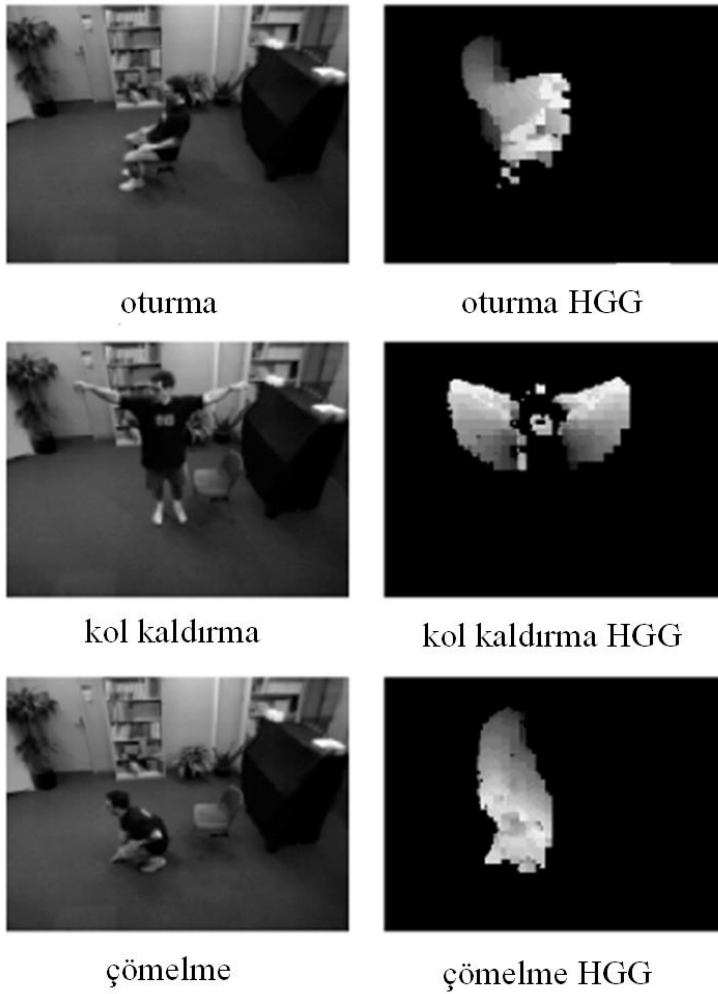
Uzay-zaman yaklaşımları ile oluşturulan şablonların karşılaştırılmasında çeşitli algoritmalar kullanılmaktadır. Bu alanda şablon karşılaştırma ve komşu temelli karşılaştırma yöntemleri sıklıkla uygulanmaktadır. Komşu temelli karşılaştırma yönteminde bir eylemi tanımlamada örnek hacim kümesi kullanılır ve test edilen giriş tüm örneklerle karşılaştırılır. Ayrıca bir eylemin oluşma olasılığını modelleyen istatistiksel modelleme algoritmaları da geliştirilmiştir [6].

Geleneksel uzay-zaman yaklaşımlarına ek olarak çeşitli varyasyonlar da önerilmiştir. Sistem bir eylemi bir hacim yerine bir yörünge olarak da ele alabilmektedir. Eğer sistem insanların eklem bölgeleri gibi özelliklerini takip edebiliyorsa, insan hareketleri bu yörünge kümelerinden daha açık şekilde temsil edilebilmektedir. Bir başka yöntem ise özellik kümesinin hacim veya yörünge bilgisinden çıkarılmasıdır. Bu sunum yöntemlerine göre uzay-zaman yaklaşımları uzay-zaman hacimleri, uzay-zaman yörüngeleri ve uzay-zaman yerel özellikleri olarak üç kategoride incelenebilmektedir [6].

Uzay-zaman hacimleri

Uzay-zaman hacmi (UZH, space-time volume) yaklaşımları iki hacim arasındaki benzerlik hesabına dayanmaktadır. Sistem iki hacimdeki benzerliği hesaplayabilmelidir ve benzerliğin en doğru şekilde hesaplanması için çeşitli hacim sunum ve algılama yöntemleri geliştirilmiştir. Bazı yaklaşımlar, tüm resimleri zaman boyunca birleştirmektense insan

siluetinin arka plandan ayrılarak tek olarak incelenmesini önermektedir. Bobick ve Davis çalışmalarında [7] şablon karşılaştırma algoritması kullanarak gerçek zamanlı algılama sistemi geliştirmişlerdir. Her bir eylemin sunumu için 3 boyutlu uzay-zaman hacmi yerine iki boyutlu hareket enerji görüntüsü (HEG, motion-energy image) ve sayısal değerli hareket geçmişi görüntüsüne (HGG, motion history image) dönüşüm yapılmaktadır. Bu yöntemle oturma, el sallama, eğilme gibi düşük seviye hareketler geleneksel şablon eşleme tekniğiyle algılanabilmektedir. Şekil 2.2’de üç farklı eyleme ait HGG örnekleri verilmiştir.

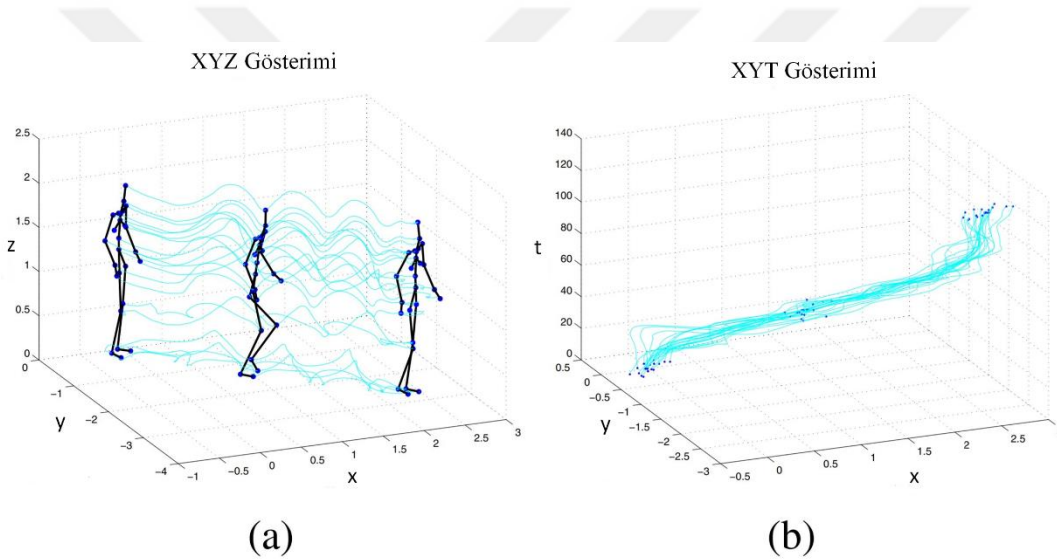


Şekil 2.2. Hareket hikâye resmi örnek gösterimi [7]

Uzay-zaman hacmi yaklaşımlarının temel kısıtlılığı, sahnede birden çok insan olduğunda hareketleri tanımadaki zorluğudur. Bu problemle başa çıkmak için kayan pencereler metodu uygulanır ancak yöntem yüksek hesaplama ihtiyacı gerektirir.

Uzay-zaman yörüngeleri

Uzay-zaman yörüngeleri (UZY, Space-time trajectories) yöntemi ile bir eylem uzay-zaman yörünge kümesi olarak çevrilir. Bu yaklaşımda bir insan iki boyutlu XY veya eklem pozisyonlarını da içeren üç boyutlu XYZ ile temsil edilir. İnsan vücut kısımları tahmin yöntemleri kullanılarak her bir çerçeveden insan eklem noktaları elde edilir. Yılmaz ve Shah çalışmalarında [8] hem sunum hem de sınıflamada uzay zaman yörüngelerini doğrudan kullanmıştır. Bir eylemi 13 ayrı yörünge kümesinden oluşturarak $XYZT$ uzayında temsil etmişlerdir. Şekil 2.3’de yürüme eylemine ait yörünge ve uzay-zamanda gösterim örneği verilmiştir.



Şekil 2.3. a) Yürüme eylemi eklem yörüngelerinin sunum örneği b) Sunum örneğinin uzay-zamanda gösterimi [9]

Yörünge temelli yaklaşımların temel avantajı insan hareketlerinin detaylı incelenmesine olanak sağlamasıdır. Dahası bu yöntemlerin birçoğu insan davranışlarını analiz ederken kamera açısından bağımsızdırlar. Fakat insan vücut parçalarının tahmin edilmesi gibi güçlü düşük seviye işlemlere ihtiyaç duymaktadır. 3B vücut parça tahmini ve izlemesi problemi hala verimli bir şekilde çözülmüş değildir.

Uzay-zaman yerel özellikleri

Bu yaklaşımlar insan eylemlerini algılamak için 3B uzay-zaman hacimlerinden elde edilen yerel özellikleri kullanır. Burada yerel özellikler 3B video kayıtlarındaki önemli noktalar ve çevresinin karakteristik özelliklerini taşımaktadırlar. Yerel özellikler daha kararlı ve daha

bilgi verici özelliğe sahiptirler. Bu yerel özellikler yoğunluklarına göre seyrek ve yoğun sunum sınıflarına ayrılmaktadır.

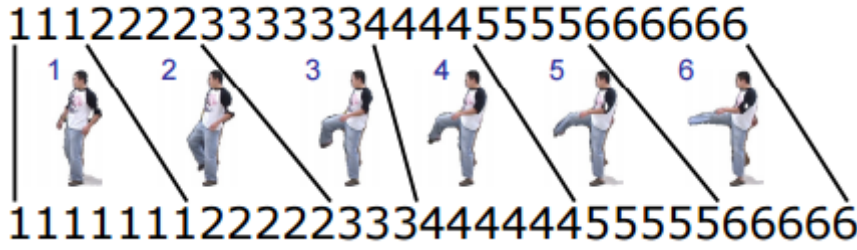
2.1.2. Ardışık yaklaşımlar

Ardışık yaklaşımlar, sıralı özellikler üzerinde inceleme yaparak insan davranışlarını algılamayı amaçlayan tek katmanlı yaklaşımlardır. Bu yaklaşımlarda giriş videosu özellik dizisi olarak ele alınır ve bir eylem kendine özgü bir sıralanış olarak tanımlıdır. Ardışık yaklaşımlarda öncelikle resim dizisi, yörüngelerin açısı gibi bir insanın durumunu tanımlayan özellik vektör dizisine dönüşüm yapılır. Bir eylemi oluşturan sıralı özellikler belirlendikten sonra analiz edilen video içerisinde bu sıraya benzeyen bir diziliş algılandığında davranışın meydana geldiği sonucuna varılır.

Ardışık yöntemler, örnek temelli algılama ve model temelli durum algılama yaklaşımları olmak üzere ikiye ayrılır. Örnek temelli yaklaşımda bir eylemi doğrudan eğitim verisinden oluşan insan hareketleri kümesinden oluşurken model temelli yaklaşımda bir eyleme karşılık eğitim sonrası model oluşturulur.

Örnek temelli yaklaşımlar

Örnek temelli yaklaşımlarda, insan eylemleri şablon dizisine çevrilerek temsil edilir. Yeni bir video analiz edileceğinde bu videodan elde edilen sıralı özellik vektörüyle eylemi temsil eden sıralı özellik şablonu karşılaştırılır. Eğer benzerlik yeteri kadar yakın çıkarsa eylem incelenen videoda gerçekleşmiştir kararına varılır. Dinamik zaman bükme (DZB, Dynamic time warping) algoritması iki sıralı dizinin karşılaştırılmasında yaygın olarak kullanılmaktadır [10]. Bu algoritma ile iki dizi arasındaki doğrusal olmayan karşılaştırma problemi, optimum polinom zamanda çözülebilmektedir. Şekil 2.4’de farklı oluş hızına göre iki serinin mantıksal eşlemesi gösterilmiştir. Şekilde her bir farklı numara özel bir durumu temsil etmektedir.

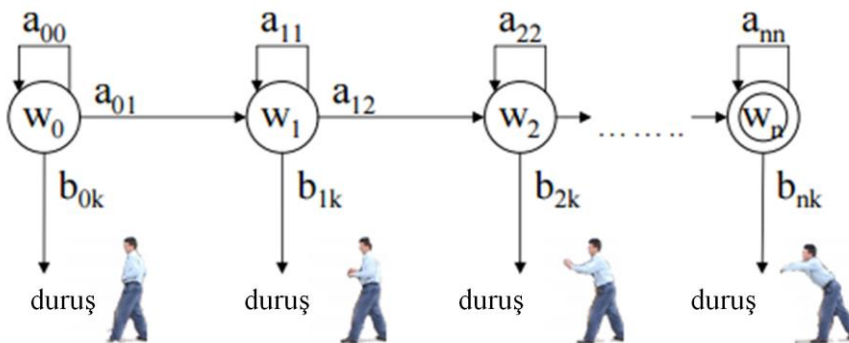


Şekil 2.4. Doğrusal olmayan oluş hızlarına göre ayak açma dizisinin eşleme örneği [5]

Model temelli durum yaklaşımları

Model temelli durum yaklaşımlarında insan davranışı birçok alt eylemden meydana gelen bir model olarak temsil edilir. Model istatistiksel olarak eğitilir ve sıralı özellik vektörü olarak ortaya çıkar. Genel olarak her bir eylem için bir istatistiksel model oluşturulur.

Saklı Markov Modelleri (SMM, Hidden Markov Models) ve dinamik bayes ağları (DBA , Dynamic Bayesian Networks) model temelli durum yaklaşımlarında yaygın kullanılmaktadır. İki yöntemde de insan eylemi gizli durumlar kümesi ile temsil edilmektedir. Bir insan her bir zaman çerçevesinde sadece bir durumda olabilir ve her bir durum, özellik vektörü gibi yeni bir gözlem üretir. Bir sonraki çerçevede, durumlar arası geçiş olasılığına bakılarak karar verilir. Geçiş ve gözlem istatistikleri model için çıkarılır ve değerlendirme probleminin çözümüyle davranışlar ortaya çıkarılır. Değerlendirme ise verilen dizi olasılıklarının hesaplanması problemidir. Model temelli durum yönteminde, hesaplanan olasılık belirlenen olasılıktan yüksek ise değerlendirilen olayın videoda gerçekleştiği sonucuna varılır. Şekil 2.5’de sıralı bir SMM örneği gösterilmiştir. Şekilde durum w_j için en yüksek gözlem olasılığı b_{jk} olan duruş gösterilmektedir.



Şekil 2.5. Kol açma eylemi için SMM örneği [5]

Ardışık yaklaşımlar, uzay-zaman yaklaşımlarının aksine özellikler arasındaki sıralı ilişkilerle ilgilenmektedir. Böylece periyodik olmayan eylemler gibi daha karmaşık davranışların algılanmasında başarı sağlanabilmektedir.

Örnek temelli yaklaşımlar, sistem tarafından çoklu şablon dizisi sağlayabildiği için model temelli durum yaklaşımlarına kıyasla esneklik sağlamaktadır. Ayrıca örnek temelli yaklaşımlarda yaygın kullanılan DZB algoritması sayesinde, gerçekleşme hızını da dikkate alan doğrusal olmayan karşılaştırma yöntemi kullanılır. Bunun yanı sıra, örnek temelli yaklaşımlar daha az eğitim verisiyle çalışabilirler. Durum temelli yaklaşımlarda ise olasılıksal analizler yapılmaktadır ve geniş bir eğitim kümesine ihtiyaç duyması en büyük dezavantajdır.

2.2. Hiyerarşik Yaklaşımlar

Hiyerarşik yaklaşımların arkasında yatan ana fikir, temel eylemlerin tanınmasından yola çıkarak yüksek seviyeli eylemlerin algılanmasıdır. Örneğin kavga etmek gibi yüksek seviyeli bir eylem, tekme ve yumruk atmak gibi düşük seviyeli alt eylemlerin birkaç kez tekrarlanması tanımlanabilir. Hiyerarşik sunum, mantıksal açıdan anlaşılabilir ve hesaplama açısından işlenebilir olmasının yanında alt olayları üst olaylarda yeniden kullanmasıyla algılama sürecindeki fazlalıkları azaltır. Genel olarak, yüksek seviye eylemler atomik (ilkel) seviye eylemlerin hiyerarşik bir şekilde modellenmesiyle temsil edilir. Çoğu hiyerarşik yaklaşımda atomik eylemler daha önce bahsedilen tek katmanlı yaklaşımlar ile algılanmaktadır.

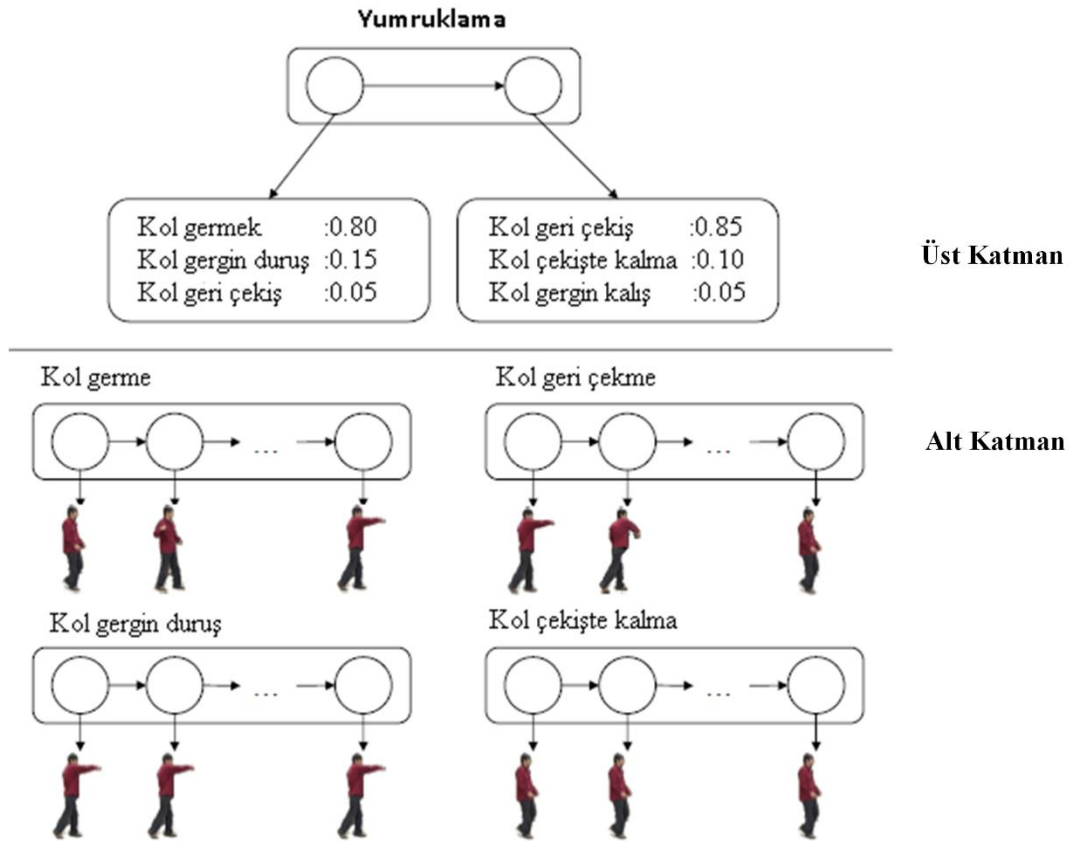
Hiyerarşik yaklaşımların tek katmanlı yaklaşımlar gibi hiyerarşik olmayan yöntemlere göre en büyük avantajı, karmaşık yapıdaki yüksek seviye eylemleri algılamadaki başarısıdır. Ayrıca, hiyerarşik yaklaşımlar insan-insan, insan-nesne ve insan-grup gibi anlamsal seviyedeki analizler için de uygundur. Bu başarısını ise daha az eğitim verisine ihtiyaç duymasından ve öncü bilgileri davranış sunumuna dahil edebilmesinden almaktadır.

2.2.1. İstatistiksel yaklaşımlar

İstatistiksel yaklaşımlar durum temelli istatistiksel modelleri kullanırlar. Genellikle insan eylemlerinin algılanmasında SMM ve DBA gibi iki katmanlı durum temelli dizi yapısı

kullanılır. Alt katmanda tek katmanlı sıralı yaklaşımlara dönüştürülen özellik vektörlerinden atomik hareketler çıkarılır. Bir başka deyişle özellik vektör dizisi, atomik hareket dizisine çevrilir. İkinci katmanda ise atomik hareket dizisi olarak temsil edilen eylem ile giriş videosunun benzerlikleri hesaplanır. Bu hesaplama için genellikle En büyük olabilirlik kestirimi (EOK, maximum likelihood estimation) veya en büyük artçıl kestirimi (EAK, maximum a posteriori probability) sınıflandırıcıları tercih edilir.

Şekil 2.6’da yumruklama eylemini algılamak üzere tasarlanmış bir istatistiksel sıralı model örneği verilmiştir. Model iki katmandan oluşmaktadır. Alt katmanda Saklı Markov Modeli kullanılarak düşük seviye atomik eylemler algılanmakta, üst katmanda ise alt katmanda çıkarılan dizi yardımıyla yumruk atma eylemi algılanmaya çalışılmaktadır.



Şekil 2.6. Yumruklama eylemi için sıralı SMM örneği [5]

İstatistiksel yaklaşımlar yeterli eğitim kümesi olduğunda gürültülü girişlerde bile iyi sonuçlar verebilmektedir. Ana kısıtlılığı ise eş zamanlı alt eylemle olduğu durumlardaki tanımlama ve algılama zorluğudur.

2.2.2. Dizilimsel yaklaşımlar

Sıralı istatistiksel yaklaşımlar gibi, dizilimsel yaklaşımlar da karmaşık davranışların algılanmasında öncelikle atomik seviye eylemlerin tespitine ihtiyaç duymaktadır. İnsan eylemleri, her bir sembolün atomik seviyede bir hareketi temsil ettiği bir sembol dizisi ile ifade edilir. Dizilimsel yaklaşımlarda içerikten bağımsız dil (İBD, context-free grammar) ve stokastik içerikten bağımsız dil (SİBD, stochastic context-free grammar) yüksek seviyeli eylemlerin algılanmasında kullanılır.

Dizilimsel yaklaşımlar tüm gözlemlerin kural tabanında gösterilmiş olduğunu varsaymaktadır. Bu sistemlerde kullanıcının tüm olası olaylar için kural tabanı oluşturması gerekir. Bu yüzden sistem daha önce gözlemlenmeyen bir olay tipi ile karşılaştığında zorluk çeker. Bu kısıtlılığın üstesinden gelmek için Kitani ve arkadaşları gözlemlerden otomatik öğrenme gerçekleştiren gramer temelli bir algoritma geliştirmiştir [11]. Fakat bu yaklaşımda, eşzamanlı meydana gelen atomik hareketlerin algılanması problemi yaşamaktadır.

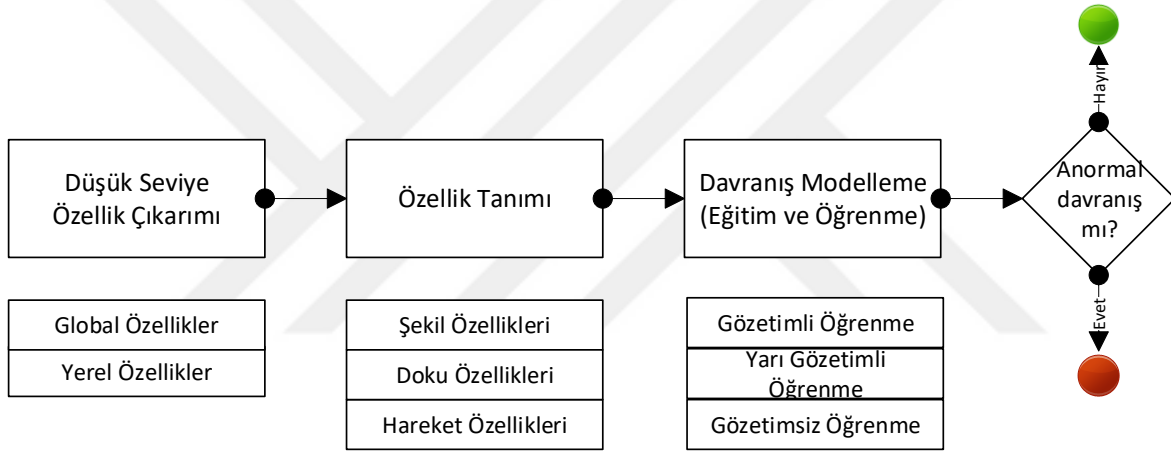
2.2.3. Tanımlama temelli yaklaşımlar

Tanımlama temelli yaklaşımlar da yüksek seviye insan hareketlerini, meydana geldiği atomik yapıda hareketlerin zamansal, uzaysal ve mantıksal ilişkilerini tanımlayarak temsil etmektedir. Tanımlama temelli yaklaşımlar insan eylemlerini belirli ilişkileri karşılayan alt olayların oluşumu ile modeller. Tüm tanımlama temelli yaklaşımlar hiyerarşiktir. Ayrıca bu yaklaşımlar eşzamanlı yapılar içeren eylemleri tanımlamada da başarılıdır.

Sıralı yaklaşımlar alt hareketlere bölünebilen yüksek seviye insan eylemlerini algılamada yeteneklidirler. İnsan bilgisini sisteme ekleyebildiği için doğası gereği daha az eğitim verisine ihtiyaç duyar [12]. İstatistiksel ve dizilimsel yaklaşımlar olasılıksal bir yapı sunarlar ancak bu yaklaşımlar eşzamanlı oluşan alt olaylardan meydana gelmiş eylemleri algılamada kısıtlılık yaşamaktadırlar. Tanımlama temelli yaklaşımların ise hem sıralı oluşan olaylarda hem de eş zamanlı gerçekleşen alt olaylarda karmaşık hareketleri algılama yeteneği vardır. Fakat alt seviye eylemlerin algılanmasında yaşanan başarısızlık yöntemin başarımını büyük ölçüde etkilemektedir.

2.3. Videolarda Normal Olmayan Eylem Tespit Yaklaşımları

İnsan davranışı algılamanın nesne bölütleme, özellik çıkarımı, özellik sunumu ve davranış sınıflandırması aşamalarında çalışmalar sürmektedir. Akıllı gözetim sistemlerinde normal olmayan eylem tespiti yapan sistemler genellikle iki video işleme aşamasından oluşmaktadır. Birinci aşamada sahnedeki ilgi alanlarını elde etmek için düşük seviye özellikler çıkartılmakta ve daha sonra düşük seviye özelliklerden yararlanılarak ilgi alanlarını belirlemek için temel hareketler tanımlanmaktadır. İkinci aşamada ise elde edilen bilgiler kullanılarak bir eylemin normal veya normal olmadığı hakkında karar verilmektedir. Şekil 2.7.'de normal olmayan eylem tespiti için oluşturulacak sistemin temel aşamaları verilmiştir.



Şekil 2.7. Normal olmayan eylem tespiti için sistem aşamaları

2.3.1. Davranış gösterimi

Davranış gösterimi, davranış analizinin düşük seviyeli bir işlemidir. Video içindeki hedef nesnenin ilgili özelliklerinin elde edilmesi sürecidir. Düşük seviye özellikler sayesinde ilgi alanları algılanır ve daha sonra hedef nesne alanının tanımlanması yapılır. Davranış gösteriminde yaşanan en büyük zorluk arka plandaki ve nesnedeki değişimler gibi durumlara karşı sağlam ve uygun özelliklerin bulunmasıdır. Çizelge 2.1'de davranış gösteriminde yaygın kullanılan yöntemler ve özellikleri listelenmiştir.

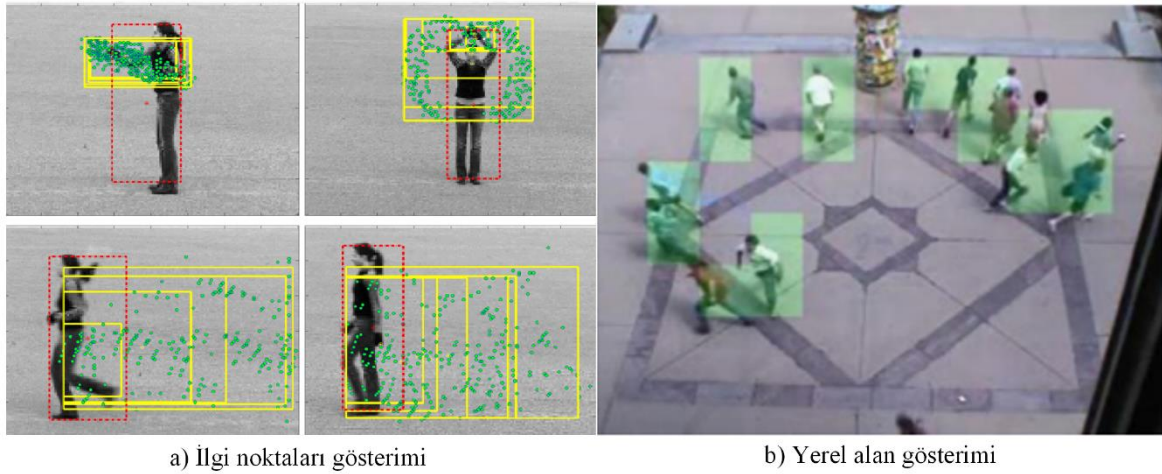
Çizelge 2.1. Davranış sunum yöntemleri ve özellikleri

ÖZELLİK TİPİ	TANIMLAMA	YORUMLAR	REFERANS
Optik akış temelli özellikler (histogram, hızlanma gibi) hareket bilgisi	Optik akıştan istatistiksel özelliklerin çıkarımıyla hareketin karakterize edilmesi	Global özelliklerle çalışır. Hesaplama maliyeti yüksektir.	[13-28]
İlgi Noktaları Uzay-zaman ilgi noktaları (UZİN, Space-time interest points) Ölçek bağımsız özellik dönüşümü (ÖBÖD, Scale-invariant feature transform)	Uzay ve zaman çalışma alanında belirgin noktalar algılama yöntemidir. Nadir eylemlere karşılık gelen önemli hareket bilgisinin sunumu yapılır.	Yerel özelliklerle çalışır. Hesaplama maliyeti düşüktür. Gürültüye karşı duyarlıdır. Kalabalık sahneler için uygun değildir.	[29-32]
UZH, küp, blob gibi	Üçüncü boyut yani zamansal boyutun ardışık çerçevelerden elde edilmesi yöntemidir.	Yerel Özelliklerle çalışır. Bellek maliyeti yüksektir. Seçilecek parametrelere duyarlıdır (ardışık çerçevelerin sayısı, küpün boyutu vb.)	[31,33-39]
Şekil (Siluet, YGH, dikdörtgen ve sınır kutuları (bounding box))	Ardışık çerçevelerde hareketli nesnelerin şekillerinin tanımlanması yöntemidir. Normal olmayan davranışlar şekil değişiminden ortaya çıkarılır.	Düşme algılama gibi şekil değişikliği içeren eylemlerin tespitinde sıklıkla kullanılır. Kalabalık sahneler için uygun değildir.	[40-42]
Doku: Moments, Gri seviyeli eş-oluşum Matrisleri (GSEM, Gray Level Co-occurrence Matrix), DDK, dalgacıklar (wavelets) gibi.	Sınırları çıkarılmış veya bir kutu içerisinde belirlenmiş her bir hareketli nesne için yerel şablon çıkarımı yapar	Kalabalık sahneler için uygundur.	[27,41,43,44]
Nesne takibi ve yörünge çıkarımı	Yörüngeler kullanılarak hareketli nesnelerin takip edilmesi yöntemidir.	Bir kişinin takip edilmesine uygundur. Kalabalık sahneler için uygun değildir. Arka plan durağan olan sahnelerde verimlidir. Hareketli arka plan veya kamera problem yaratır.	[45-60]

Davranış sunumu ortaya konurken, analiz edilen nesne şekil ve doku gibi farklı açılardan ele alınmış olabilmektedir. Örneğin Li ve arkadaşları çalışmalarında [43], kalabalık sahnelerde normal olmayan davranış tespiti için dinamik doku karışımı (DDK, mixtures of dynamic texture) modeli temelli bir yöntem önermişlerdir. Çalışmalarda gri seviyeli eş dizilim matrisi (GSEM, Gray level co-occurrence matrix) [41,61] ve uzay-zaman özellikleri [62] kullanılmıştır. Bunların yanında, hedef nesnelerin hareket bilgisi ise o nesnenin davranışı hakkında bilgi veren en önemli kaynakların başında gelmektedir. Zaman boyutunda hareketli

nesnelerin tanımlanması için çoklu özellikler kullanılmaktadır. Bu özellikler yerel ve global özellikler olmak üzere ikiye ayrılmaktadır.

Yerel özellikler çerçevedeki ön tanımlı alanlar içinde algılanırlar. Bu alanlar ilgi noktaları veya yerel bir alan olarak temsil edilirler. Yerel özellik olarak, ilgi noktaları üzerine Debapratim ve Soharab bir literatür çalışması [63] ortaya koymuş, çalışmada ilgi noktaları yöntem aşamalarını, kullanım çeşitlerini ve yöntemlerini açıklamıştır. Çalışmalar [64-66] ölçek bağımsız hareket özellik dönüşümü kullanarak yerel hareketleri tanımlamışlardır. Zhao ve arkadaşları [39] yerel uzay-zamansal tanımlayıcısını YGH ve optik akış histogramı (OAH, Histogram of optical flow) yöntemlerini birlikte kullanarak tanımlamıştır. Li ve arkadaşları [33] uzay zamansal video küpleri analizi yöntemini yerel normal olmayan durum tespitinde kullanmıştır. Uzay zamansal blob çıkarımı [34] istatistiksel yöntemi de yerel normal olmayan eylem algılamasında kullanılmıştır.



Şekil 2.8. İlgi noktaları ve yerel alan gösterimleri için örnek durumlar

Global özellikler ise tüm çerçeve içindeki hareketleri tanımlamak için kullanılır. Optik akış en sık karşılaşılan global özelliklerden birisidir [28]. Optik akış yöntemi, hızlanma ve histogram olmak üzere iki özellikle birlikte kullanılabilir [18]. Şiddet olaylarının tespitinde Optik Akış Gauss Modeli (OAGM, Gaussian Model of Optical Flow) de kullanılmıştır [67]. Optik akış yönteminin ilgi noktaları gösterimi [68] kullanımı ve yerel alan gösterimi [33] sırasıyla Şekil 2.8’de gösterilmiştir. Nesne takibi için hareket bilgisi arı kolonisi [58], parçacık sürü optimizasyonu (PSO, Particle Swarm Optimization) [69], kalman filtresi [70], HGG [59] gibi yöntemlerle elde edilebilir. Hareketli nesnelerin yörünge bilgisi de normal olmayan davranışların tespitinde sıklıkla kullanılır [45,60,71,72].

2.3.2. Normal olmayan eylem tespit yöntemleri

Alt seviye aşamada sahnede yer alan hareketli nesnelerin algılanması ve tanımlanması işlemleri yapılmaktadır. Ancak bu aşamada hareketli nesnenin hangi eylemde bulunduğu veya eylemin normal olup olmadığı konusunda herhangi bir çıkarım yapılmamaktadır. Bundan dolayı, normal olmayan eylemlerin tespit edilerek sınıflandırılmasında kullanılan algoritma ve yöntemler normal olmayan eylem tespitinde büyük önem arz etmektedir. Normal olmayan eylem sınıflandırmasında yöntemleri gözetimli, gözetimsiz ve yarı gözetimli olarak üç gruba ayrılmaktadır. Çizelge 2.2’de normal olmayan eylem algılamada davranış sınıflamasında kullanılan öğrenme yöntemleri, yöntemlerin avantajları ve kısıtlılıkları detaylı bir şekilde verilmiştir.

Çizelge 2.2. Normal olmayan eylem algılamada kullanılan sınıflama yöntemleri

Öğrenme Yöntemi	Tanımlama	Avantajlar	Kısıtlılıklar
Gözetimli	-Etiketli veri kullanılarak normal ve normal olmayan eylemler için model oluşturma	-Veri kümesinde olan davranışların tespitinde etkilidir. -Anlaşılır ve kolay uygulanabilir.	-Tespit edilecek eylem sayısı belirlidir. -Eğitim verisinin içeriğine bağlıdır. -Bilinmeyen eylemler tespit edilemez -Gerçek hayat uygulamaları için çok uygun değildir.
Yarı-gözetimli (kural tabanlı)	-Normal şablonlar kullanılarak kural oluşturma -Kurallara uymayan yeni bir eylem normal olmayan eylem olarak kabul edilir.	-Anlaşılır ve kolay uygulanabilir. -Anlamlıdır. -İnsan muhakemesine yakındır.	-Yüksek bellek ve hesaplama karmaşıklığı gerektirebilir.
Yarı-gözetimli (model tabanlı)	-Normal davranışları temsil eden bir model oluşturur. -Modele uymayan yeni bir davranış normal olmayan davranış olarak kabul edilir.	-Anlaşılır ve kolay uygulanabilir. -Ele alınacak durum hızlı bir şekilde sınıflandırılır.	-Çoklu parametrelere hassastır. -Normal veri sistem tarafından normal olmayan eylem olarak algılanabilir.
Gözetimsiz	Etiketsiz veriden elde edilen istatistiksel özelliklerden öğrenme gerçekleştirir.	-Hızlı ve kolay çalışma -Ön bilgi ihtiyacı yoktur.	-Sonuç yorumlaması için vakit ihtiyacı vardır -Normal olmayan eylemler, normal eylemlere göre nadir meydana gelir varsayımı temellidir.

Gözetimli öğrenme normal ve normal olmayan davranışları etiketli veri yardımıyla modellemektedir. Bu yöntemler genellikle kavga tespiti [30], aylak gezen tespiti [73], düşme tespiti[74] gibi eğitim kümesinde yer alan belirli davranışların algılanmasında kullanılır. Literatürde hem normal olmayan davranış tespitinde hem de diğer bilim dalları ve problemlerinde uygulanan birçok gözetimli öğrenme yöntemi bulunmaktadır. Normal

olmayan eylemlerin tespitinde en yaygın kullanılan yöntemlerden bir tanesi kelime çantası (KÇ, Bag of Words) [75,76] yöntemidir. KÇ yönteminde her bir çerçeve yerel resim nitelikleri ve yörünge gibi özelliklerin histogramı olarak temsil edilmektedir. Öncelikle bir kelime sözlüğü (dictionary of words) oluşturulur ve arkasından video sözlüğündeki her bir sözcüğün frekansı kullanılarak histogram hesaplanır. KÇ çoğunlukla saldırı [30] ve kalabalık anormallik problemleri [13] için destek vektör makinası (DVM, Support Vector Machine) sınıflandırıcısı ile birlikte kullanılmaktadır.

Yarı gözetimli yöntemler eğitim için genellikle normal eylem barındıran videolarına ihtiyaç duymaktadır. Bu yöntemler kural tabanlı ve model tabanlı olmak üzere iki farklı yaklaşımda sınıflandırılabilir. Kural tabanlı yaklaşımlar normal eylem şablonlarını kullanarak kural tanımlamaya çalışmaktadır. Kurala uymayan test verisi normal olmayan durum olarak algılanmaktadır. Zihang ve arkadaşları yarı gözetimli öğrenme yöntemlerinde[77], sözlük öğrenmesi ve seyrek sunum yöntemini kullanmışlardır. Lu ve Jia yaptıkları çalışmada [78] seyrek kodlama yaklaşımını kullanarak normal olmayan eylemlerin tespitinde kural tabanlı yeni bir yöntem önermişlerdir. Çalışmada önerilen yöntem, saniyede 150 çerçeve işleyebilme başarısı göstermesine rağmen eşik değeri seçiminin başarısında çok büyük farklılıklar göstermesi yönüyle kısıtlılık göstermektedir. Nguyen ve arkadaşları [42] şekil özellikleri kullanarak oluşturdukları kural temelli yöntemi ile düşme algılama sistemi önermişlerdir. Tani ve arkadaşları [79] video gözetim sistemleri için normal olmayan davranış tespitinde ontoloji temelli yaklaşım kullanarak kural oluşturan bir yöntem önermişlerdir. Bir başka çalışmada ses, görüntü ve sensör gibi çoklu kaynaklardan alınan bilgiler kullanılan bulanık mantık temelli bir adaptif kural tabanlı yöntem ise sızma tespiti için geliştirilmiştir [80]. Ayrıca bulanık kümeleme tabanlı algoritmalar da önerilmiştir [81]. Bulanık karar verme süreciyle çok katmanlı sinir ağlarını birlikte öneren davranış analiz sistemleri de bulunmaktadır [82].

Yarı gözetimli öğrenmede Markov rastgele alanı (MRA, Markov Random Field), Gauss karışım modeli (GKM, Gaussian Mixture Model) ve SMM en çok kullanılan model tabanlı yöntemlerdir. Nallaivarothayan ve arkadaşları [18] yaptıkları çalışmada normal olmayan durum tespiti için GMM temelli bir MRA yöntemi önermiştir. Literatürde Gauss modelini öneren birçok çalışma mevcuttur [22]. Rojas ve Tozzi [83] yaptıkları çalışmada normal davranışlar optik akış ve arka plan çıkarımı yöntemlerinden alınan veriden GKM ile kodlanmıştır ve kalabalık sahneler için önerilen modern yöntemlerle hem çalışma zamanı

hem de doğruluk üzerinde deneysel sonuçlar verilmiştir. Ayrıca Feng ve arkadaşları [84] olay şablonlarını modellemek için derin GMM önermişler ve bu derin modeli diğer modern yöntemlerle karşılaştırıp etkinliğini deneysel sonuçlarla ortaya koymuşlardır. Kai-Wen ve arkadaşları çalışmalarında [85], Gauss regresyon işlemi (GRİ, Gaussian Process Regression) kullanan sıralı bir normal olmayan eylem algılama çerçevesi önermiştir.

Gözetimsiz öğrenmede etiketsiz veriden çıkarılan istatistiksel özellikleri kullanarak normal ve normal olmayan davranışların çıkarılması amaçlanmaktadır. Alvar ve arkadaşları [86] normal olmayan eylemlerin tespiti için baskın kümeler (dominant sets) kullanan bir gözetimsiz öğrenme çerçevesi önermişlerdir. Literatürde normal olmayan eylem tespitinde kullanılan yöntemler Şekil 2.3’de verilmiştir

Çizelge 2.3. Normal olmayan eylem tespiti için kullanılan yöntemler

NORMAL OLMAYAN DAVRANIŞ TESPİTİ İÇİN YÖNTEM VE ALGORİTMALAR	REFERANS
KÇ yaklaşımı	[13,30,33,87-89]
DVM	[13,26,40,67,90,91]
Sinir ağları (SA)	[23,82,92]
Özyineli sinir ağı (ÖSA, Recurrent Neural Network)	[93]
Evrişimsel sinir ağları (ESA, Convolutional neural network)	[94-97]
GKM	[17,18,98,99]
Derin Gauss Karışım Modeli (DGKM, Deep gaussian mixture model)	[84]
K en yakın komşu	[16]
SMM	[100-102]
Bulanık Sistem	[103]
Seyrek Sunum	[20,33,104]

2.3.3. Derin Öğrenme Yöntemleri

Son yıllarda derin öğrenme yöntemleri, çoğu zaman herhangi bir ön işleme ve özellik çıkarımı yapmaksızın başarılı sonuçlar almasıyla büyük bir popülerliğe ulaşmıştır. İlk önceleri evrişimsel sinir ağları kenar algılama [105], nesne algılama [106], yüz tanıma [107] gibi ileri seviye problemlere çok önemli çözümler getirmiştir. İlerleyen aşamalarda yapay sinir ağı modelleri [108,109] videolarda normal olmayan eylemlerin tespitinde de dikkat çekici sonuçlar almıştır. Evrişimsel sinir ağlarını geleneksel makine öğrenmesi yöntemlerinden ayıran en önemli noktalardan biri herhangi bir ön işlem olmaksızın video içerisindeki önemli özellikleri otomatik çıkarmasıdır. Ancak eğitim aşamasında etiketli veri ile çalışmaya ihtiyaç duymaları, bazı verinin etiketlenmesinin güç olması ve uzmanlık gerektirmesi gibi sebepler evrişimsel sinir ağlarının hayata geçirilmesi ve kullanımını zorlaştırmaktadır.

Araştırmacılar son yıllarda ESA gibi gözetimli öğrenme gerektiren modellerin performansı ile gözetimsiz öğrenmenin rahatlığını bir arada uygulayabilmenin yollarını araştırmışlardır. Bu kapsamda Evrişimsel otokodlayıcı [110], değişimsel otokodlayıcı [111], çekişmeli üretici ağ [112], evrişimsel uzun kısa süreli bellek [113] ve bahsedilen modellerin varyasyonları gibi etiketli veriye ihtiyaç duymadan eğitim yapabilen derin öğrenme modelleri önerilmiştir.

Araştırmacılar gözetimli veya yarı gözetimli üretici yapay sinir ağları kullanarak videolarda normal olmayan eylem algılanması konusunda da çözümler sunmuşlardır. Evrişimsel sinir ağının uzaysal ve zamansal özelliklerin elde edilerek normal olmayan eylem tanımda kullanılması, ilk olarak Zhou ve arkadaşları [108] tarafından geliştirilmiştir. Fakat ESA modelleri video ve zamansal özellikleri öğrenmek için ortaya çıkarılmadığından araştırmacılar evrişimsel otokodlayıcıyı, ESA modellerine alternatif olarak sunmuştur. Hasan ve arkadaşları [110] tam bağlı otokodlayıcı ve evrişimsel ileri beslemeli kodlayıcı önermiştir. Bu yaklaşım hem otokodlayıcının elde ettiği önemli özelliklerden hem de hareket özelliğinin elde edilmesinden yararlanmaktadır. Uzaysal ve zamansal özelliklerin daha iyi ortaya çıkarılıp modeller tarafından öğrenilebilmesi için Medel ve Savakis[114] evrişimsel uzun kısa süreli bellek ağı önermiştir.

Literatür incelemesi sonucunda, gözetimsiz derin öğrenme modellerinin performanslarının ve kullanımlarının son yıllarda hızla arttığı gözlemlenmiştir. Video işlemenin uzaysal özelliklerin yanında zamansal özellikleri de barındırması ve zamanla değişen durumların videolarda meydana gelen olaylar hakkında önemli ipuçları taşıması, görüntülerde başarılı olan derin öğrenme modellerinin zamansal özellikleri nasıl elde edeceği konusunda çalışmalara neden olmuştur. Özellikle evrişimsel uzun kısa süreli bellek kullanarak ardışık video serilerinden anlam çıkarılması öne çıkmış ve etkili sonuçlar elde edilmiştir.

2.4. Ulusal Tez Çalışmaları

Bu bölümde video üzerinde insan eylem analizi alanında ülkemizde yapılan tez çalışmaları ele alınmıştır. Çalışmalarda genel olarak daha önceden belirlenmiş insan eylemlerinin video içerisinde tanınması problemi üzerinde durulmuştur. Araştırmalar sonucunda video kayıtları üzerinde normal olmayan eylem tespitinde bulunan bir çalışmaya rastlanmamıştır. Aşağıda,

çalışmaların genel çerçeveleri, kullanılan yöntemler ve deneysel sonuçları hakkında bilgi verilmiştir. Ayrıca bu çalışmaların detaylı bilgileri Çizelge 2.4 ile sunulmuştur.

Nurullahoğlu [115] tarafından yapılan tez çalışmasında video görüntülerinde insan hareketlerinin tespiti için iyi bilinen makine öğrenmesi yaklaşımları uygulanmıştır. Öncelikle, farklı çevre koşulları ve kişilere ait koşma, hızlı koşma, el çırpma ve boks yapma eylemleri için kısa video görüntülerinden oluşan bir veri tabanı kullanılmıştır. Bu eylemleri tanımlamak için hareket geçmişi görüntülerinden 2B veya 3B şekilleri tanımlamada yaygın olarak kullanılan ve iyi bilinen bir özellik kümesi olan Hu momentleri hesaplanmıştır. Temel bileşen analizi (TBA, Principal Component Analysis) ve doğrusal ayırmacılık analizi (DAA, Linear Discriminant Analysis) özellik çıkarımı yöntemleri ile elde edilen özellikler, K en yakın komşu ve DVM yaklaşımları ile hareket tespitinde kullanılmıştır. Hu momentleri ile hesaplanan özellik setinden TBA ile özellik çıkarımı ve K en yakın komşu (KEY, K-Nearest Neighbor) sınıflandırıcı ile test sonucunda boks yapma ve el çırpma sınıflarında %95 ile en yüksek başarıyı sağladığı belirtilmiştir.

Baysal [116] tarafından yapılan tez çalışmasında video görüntülerinde insan eylemlerini tanımak için çizgiye dayalı bir poz temsilinden faydalanılmaktadır. Böylece el, kol ve eklem hareketleri daha iyi tanımlanmış ve insan figürünü oluşturan çizgiler arasındaki geometrik ilişkilerin yakalanması sağlanmıştır. İki poz arasında çizgi-çiftlerini eşleştirip benzerliklerini hesaplayarak literatüre katkı sunulmuştur. Poz temsili yaklaşımlarının gücünü ortaya koymak için Weizmann ve KTH veri setleri üzerinde deneysel çalışmalar yapılmıştır. Sıralı poz ve çizgi-akış histogramlarının birlikte kullanımı ile eylemleri birbirinden ayırt etmede etkin sonuçlar ortaya konulmuştur. Ayrıca önerilen yaklaşımın çoklu kamera sistemlerinde uygulanabilirliği de IXMAS veri kümesi üzerinde gösterilmiştir.

Kurt [117] tarafından yapılan tez çalışmasında video kayıtlarındaki insan hareketlerinin algılanmasında anahtar kareye dayalı bir poz sunumu kullanılmaktadır. İnsan figürünün oluşturduğu duruşun, video çerçevelerinde sürmekte olan hareket ve eylemi tanımlamak için çok güçlü ipuçları verdiği ifade edilmiştir. Çalışmada anahtar kare, bir eylemi diğer eylemlerden ayıran önemli bilgidir. Verilen iki kare üzerindeki duruşlar arasında bir benzerlik değeri hesaplanmakta daha sonra bir kümeleme algoritması kullanılarak, her eylemin benzer çerçeveleri belirli bir sayıda kümede toplanmakta ve bu kümelerin ağırlık merkezleri şekil eşleme metoduyla “Anahtar Kare” olarak tanımlanmaktadır. Weismann ve

KTH veri kümeleri ile yaptıkları deneysel çalışmalar ile insan hareketlerini temsil etme ve tanımlamada etkin bir yöntem oluşturdıklarını ortaya konulmuştur.

Güçlü [118] tarafından yapılan tez çalışmasında video görüntülerinin incelenmesiyle meydana gelen hareketin ne olduğuna tespit eden yeni bir sistem önerilmiştir. Bu yolla, sistem otomatik olarak hareketleri tespit edebilmekte ve ortam anlamlandırmaya katkı sağlayan bir mimari ortaya konulmuştur. Çalışmada hareket bilgisiyle anlık durağan görüntü bilgisi birlikte kullanılarak hareketler, zamansal bir örüntü olarak tanımlanmıştır. Zamansal bir örüntü çıkarılmasında, ilk olarak arka plan çıkarımı ile hareket bilgisi içermeyen ve anlamsız olan alanlar göz ardı edilmiş ve beraberinde bu bilgi hareketin gücünü ifade eden bir hale dönüştürülmüştür. Elde edilen optik akış güç şekli yönlü gradyanlar histogramı yardımıyla betimleyici bir hale dönüştürülmüştür. Hareket, durgun görüntü ve zaman bilgisinin birlikte kullanılması ile insan eylemlerinin etkili biçimde ifade edilebileceği ortaya konulmuştur. Tez çalışması kapsamında önerilen yapı kabiliyetleriyle değişen koşullar altında başarıyla çalışabilmektedir. Bu otonom hareket tanıyabilen sistem sayesinde videoların anlamlandırılabilmesi mümkün kılınmış ve gelecek çalışmalara öncü olmuştur.

Pehlivan [119] tarafından yapılan tez çalışmasında, çoklu kamera kullanımıyla insan eylemlerinin tanınmasına yönelik çalışmalar ortaya konulmuştur. Tez kapsamında iki farklı senaryo üzerinde çalışılmış ve iki farklı yöntem önerilmiştir. Birinci yöntem, kalibre edilmiş kameralardan elde edilen hacimleri eşleştiren bir model, ikinci yöntem ise görüntü çerçevelerini eşleştiren esnek bir model olarak tanımlanmıştır. Kullanılan iki farklı yöntemde elde edilen sonuçlar, tek kamera görüntüleriyle elde edilen sonuçlarla, farklılıklarının ve performanslarının ortaya konulabilmesi için karşılaştırılmıştır. Kaydedilen kamera görüntüleri, hacim yeniden yapılandırma yöntemi ile birleştirilmekte ve elde edilen hacimler hareket duruşları eşleniği olarak kabul edilmektedir. Birbiriyle örtüşen mevcut görüntüler yeterli sayıda olmadığı durumlar için ise farklı sayıda kamera ve öznelikle çalışabilen bir hareket tanıma sistemi düşünülmüştür. Böylece kameraların kalibre edilmesine de gerek duyulmadan çalışabilecek bir sistem önerilmiştir.

Keçeli [120] tarafından yapılan tez çalışmasında Microsoft Kinect RGBD algılayıcıdan alınan derinlik verisi kullanarak, insan eylemlerini algılamaya çalışan çeşitli yöntemler ortaya koymaktadır. Bir eklem iskelet yaklaşımından elde edilen açı ve yer değiştirme bilgileri hareket algılama amaçlı kullanılmaktadır. Daha sonra hareketleri zamansal örüntü

olarak ele alan saklı Markov modelleri ve zaman serilerini temel alan yöntemler üzerine durulmaktadır. Elde edilen özniteliklerden oluşturulmuş bir sözlük yardımıyla gözlem dizileri haline getirilen hareketler, saklı Markov modelleri yardımıyla tanınmaya çalışılmıştır. Hareketler ayrıca zaman serileri olarak ele alınmış ve serilerden elde edilen öznitelikler üzerinde boyut indirgemesi yaparak sınıflama yapan bir yöntem önerilmiştir. Daha sonra, iskelet modelinden elde edilen nitelikler yanında ham derinlik bilgisi de kullanılarak doğru tanıma oranı yükseltilmeye çalışılmıştır. Tez çalışmasında sonuç olarak yapılan çalışmalardan yola çıkılarak düşük gecikmeli hareket tanımaya yönelik bir yaklaşım ortaya konulmuştur. Geliştirilen yöntemler akademik çalışmalarda sıklıkla kullanılan ve yaygın olarak bilinen MSRC-12 ve MSR-Action 3D veri kümeleri üzerinde ve kendi oluşturdukları HUN-3D veri kümesi üzerinde test edilmiştir. Geliştirilen yöntemler veri kümelerinden bağımsız olarak kolay hesaplanabilir öznitelikler ile başarılı sonuçlar üretmiştir.

Agahian [121] tarafından yapılan tez çalışmasında, bir eylemi önceden tanımlanmış pozlarla ifade ederek eylemi pozların histogramı ile kodlamaya dayalı bir eylem tanıma yöntemi önerilmiştir. Eğitim pozlarından elde edilen tanımlayıcılar, k-means ve GKM ile kümelenecek anahtar pozlar oluşturulmuştur. Eylemleri temsil eden öznitelik vektörünün uzunluğu sabit ve görüntüyü oluşturan çerçeve sayından bağımsızdır. Eylemlerin vektör temsillerinin oluşturulmasında Fisher vektör kodlama ve öznitelik seçiminde PCA uygulanmıştır. Sonrasında eylemlerin sınıflandırılması için ise Aşırı Öğrenme Makinesi (AÖM, Extreme Learning Machine) kullanılmıştır. AÖM, farklı uygulama alanlarında öğrenme hızı yüksek ve tutarlı doğruluk sağlayan tek katmanlı ileri beslemeli sinir ağıdır. UTKinect, CAD-60, UTD-MHAD, MSR Action 3D ve MSRC-12 veri setleri üzerinde pek çok farklı eyleme göre analizler yapılmıştır. 3B iskelet verisini kapsayan bu beş veri kümesi üzerinde yapılan testlerde ilk üç veri kümesinde literatürde mevcut yöntemlerden daha başarılı sonuçlar üretilmiştir. Diğer iki veri kümesinde ise literatür ile kıyaslanabilir sonuçlar ortaya konulmuştur.

Yalçın [122] tarafından yapılan tez çalışmasında derin öğrenme ile insan hareketi tanıma üzerine odaklanılmıştır. İskeletsel ve KYM(RGB) verisi kullanılarak, 2 ve 3 boyutlu evrişimsel sinir ağı modelleri ve LSTM gibi derin öğrenme teknikleri ile insan eylemleri sınıflandırılmıştır. İskelet tabanlı insan eylemlerinin tanınmasında eklem verisi kullanılarak otomatik öznitelik çıkarımı yapılmaktadır. Bu amaçla NTU RGB+D veri kümesi

kullanılmıştır. Ayrıca etkileşimli eylemlerin sınıflandırmasında MSR Daily Activity 3D veri kümesinde ile eğitim yapıp çalışma kapsamında oluşturulan veri kümesi üzerinde test işlemi yapılmıştır. Bu aşamada DVM, UKSB ve ESA modelleri yeni kişi eylemlerinin tespiti için uygulanmış ve başarılı sonuçlar sağlamıştır. RGB tabanlı eylem tanımada ise 3B ESA mimarisi, UKSB ağı kullanılan hibrit model ve optik akış ile 3B ESA ağı modeli ile karşılaştırmalı deneyler yapılmıştır. Dağınık arka plan, yüksek kamera hareketleri gibi zorlayıcı özellikleri bulunan veri kümelerinden olan UFC101 ve HMDBM51 performans analizi için kullanılmıştır. Optik akış özniteliklerinin 3B ESA öznitelikleri ile birleştirilmesi sonucunda yapılan sınıflandırmanın başarı değerinin arttığı deneysel çalışmalar ile ortaya konmuştur.

Erkut [123] tarafından yapılan tez çalışmasında insan eylemlerin sınıflandırılmasında ESA yöntemi kullanılmıştır. Mevcut yaklaşımlardan farklı olarak hazır veri tabanları yerine çeşitli arama motorları kullanılarak beş farklı eyleme ait görsel veri toplanmış ve veri kümesi oluşturulmuştur. Ayakta, koşma, oturma, zıplama ve yatma eylemleri üzerine resimler ile yapılan sınıflandırma sonucunda %97 başarı oranı ortaya konmuştur. Video testlerinde yapılan deneylerde ise %92 doğruluk ile eylemler sınıflandırılmıştır.

Çizelge 2.4. Ulusal tez çalışmaları ve detaylı incelemesi

REF. NO	YIL	EYLEM TİPİ	ÖZELLİK ÇIKARIMI	YÖNTEM	TANIMLAMA
[115]	2008	Koşma, tempolu koşu, el çırpma, boks yapma	Hu momentleri, TBA, DAA	DVM, KEYK	Geçmiş hareket imgelerinden Hu moment hesapları ve farklı özellik çıkarımları ile hareket tanıma.
[116]	2011	Her veri kümesi için 10 ve üzerinde sayıda yürüme, oturma, atma vb. farklı eylem	Çizgi tabanlı poz çıkarımı, çizgi-akış histogramları	KEYK ile hareket tanıma	Çizgiye dayalı bir poz temsilinden faydalanan video görüntülerinde insan eylemlerini tanıma.
[117]	2011	Her veri kümesi için 10 ve üzerinde sayıda yürüme, oturma, atma vb. farklı eylem	Çizgi tabanlı poz çıkarımı, pozların şekil bağlam vektörleri	Farklı küme uzaklığı ölçüm yöntemleri	Anahtar kare tabanlı poz temsili ile video görüntülerinde insan hareketleri tanıma.
[118]	2012	Eğilme, koşma, yürüme, zıplama vb. 10 farklı eylem	Optik akış, Yönlü gradyanlar histogramları	KEYK ile hareket sınıflandırma	Gerçek zamanlı video görüntülerinden insan hareketlerini tanıma ve yorumlama.
[124]	2014	Eğilme, koşma, yürüme, zıplama vb. 20 farklı eylem	Eklem açıları histogramları	DVM ve rasgele orman ile sınıflandırma, SMM ile hareket tanıma	Derinlik bilgisi tabanlı insan hareketleri tanıma.
[121]	2018	Her veri kümesi için 10 ve üzerinde sayıda yürüme, oturma, atma vb. farklı eylem	Fisher vektör kodlama, TBA	KEYK ve GKM ile anahtar pozları elde etme, DVM ve ELM ile sınıflandırma.	İnsan eylemlerinin tespiti için 3B iskelet verisini kullanan bir poz çantası yöntemi uygulanmıştır.
[122]	2018	Farklı veri setleri için farklı sayılarda eylem	Optik akış	DVM, ESA, UKSB	İnsan eylemlerinin tanımlanmasında farklı derin öğrenme teknikleri ile sınıflandırma yapılmıştır.
[123]	2019	Ayakta, koşuyor, zıplıyor, oturuyor, yatıyor	-	ESA	Video görüntülerinde derin öğrenme yöntemleri ile farklı eylemlerin tespiti ve sınıflandırılması yapılmıştır.



3. MATERYAL VE YÖNTEMLER

Bu bölüm, tez kapsamında önerilen yöntem ve tekniklerin temelini ortaya koymaktadır. Yöntemlerin çalışma prensipleri, hangi amaçla kullanıldıkları ve mimarileri açıklanmıştır.

3.1. Derin Öğrenme

Yapay zekâ gelişiminin başlarında, matematiksel olarak ifade edilebilen, insanlar tarafından çözümünün zor olduğu fakat bilgisayarların işlem yapma kapasiteleri yardımıyla çözebildiği problemlere yoğunlaşılmıştır. Yüz tanıma, nesne tanıma, ses tanıma gibi insanın tam olarak nasıl çözdüğü bilinmeyen ve matematiksel olarak ifade edilemeyen sezgisel problemlerin çözümü ise yapay zekânın ilk yıllarında çözülmesi zor problemlerdi. Derin öğrenme, problemlerin matematiksel olarak modellenmesine gerek olmaksızın edindiği bilgiler arası ilişkilerden problemi anlama ve ilişkiyi kurabilme yeteneği sayesinde sezgisel problemlere kolay çözümler getirmiştir.

Geleneksel makine öğrenmesi yaklaşımlarının kısıtlılıklarından birisi, doğal veriyi ham haliyle işlemede yaşadığı zorluklardır. On yıllardır, alan uzmanları ve mühendislik yaklaşımlarıyla ham veri, genellikle bir sınıflandırıcı olan makine öğrenmesi sisteminin girişine uygun özellik vektörleri sunumuna dönüştürülmeye çalışılmıştır [125]. Makine öğrenmesi yöntemlerinin performansı, girdi olarak kullanılan bilgilerin temsil gücüne yüksek oranda bağlıdır. Örneğin, ele alınan konuşma ses kaydının bir kadına mı yoksa erkeğe mi ait olduğu probleminin çözümünde ses aralığı bilgisinin elde edilmesi problemin çözümünde önemli bir özellik olarak kullanılabilir. Fakat insanların bazı problemleri deneyim ve sezgileri ile çözebilmesi bu gibi problemler için ayırt edici özelliğin ne olduğuna karar verebilmesini mümkün kılmamaktadır. Örneğin, bir görüntüdeki yüzleri tanıma probleminde, yüz tanımadaki önemli özellikleri seçmek oldukça zordur. İmge üzerinde yüzlerin bulunması, yüz üzerindeki nesnelerin tespit edilmesi, bulunan nesnelerin şekilleri, büyüklükleri birbirlerine olan uzaklıkları ve oranları gibi birçok özelliğin çıkarılması gerekebilir. Ancak ortamın ışığı, insanların uzaklığı, açısı ve benzeri durumlardan dolayı özelliklerin anlamlı bilgiyi verebilmesi de zordur. Oysaki insanoğlu çoğu zaman yüzleri resimde gözükmeyen veya belirgin olmayacak kadar uzakta duran insanları bile sezgi ve deneyimlerle tanıyabilmektedir. Derin öğrenmenin sunduğu gösterim öğrenme

(representation learning), ham verinin algılama ve sınıflandırma için anlamlı bir sunumunu otomatik olarak elde edebilmektedir.

Gösterim öğrenme yöntemleri, uzmanların deneyimleri ve uzun yıllar çalışmalarının sonucunda elde ettiği anlamlı bilgileri, günler bazen saatler içerisinde özellik kümeleri olarak çıkarabilmektedir. Gösterim öğrenme yöntemlerinden en bilineni otokodlayıcılardır. Otokodlayıcılar çok boyutlu giriş verisini önce çözücü fonksiyonu ile saklı uzaya sıkıştıran sonra da saklı uzaydaki sıkıştırılmış veriyi yeniden inşa etmeye çalışan bir sinir ağı modelidir. Otokodlayıcı algoritmaları bu yöntem ile giriş verisinin önemli değişkenlerini ayırmayı amaçlamaktadır.

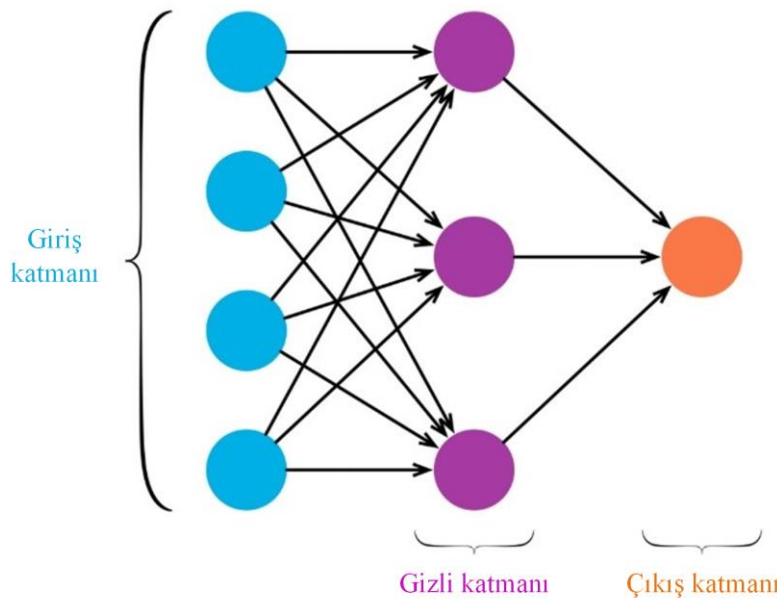
Derin öğrenmenin altyapısı sinir ağlarından oluşmaktadır. 1940'lı yılların sonlarına doğru insan beynindeki sinir hücrelerinin çalışma prensibinden yola çıkarak başlayan çalışmalar günümüzde birçok problemin çözümünde en yaygın kullanılan yöntemlerden birisi olmuştur [126]. Gelişim yıllarında her ne kadar büyük beklentiler oluştursa da belirli kısıtlılıklardan dolayı zaman içerisinde problem çözümünde kullanılmasında gerileme yaşanmıştır. En önemli kilometre taşlarından birisi, 1985 yılında Rumelhart ve arkadaşlarının geri yayılım yönteminin gizli katmanlar yardımıyla sinir ağlarında kullanılmasının önerilmesi olmuştur [127]. Fakat o dönemin şartlarında ihtiyaç duyulan veri işleme kapasitesinin karşılanamaması yöntemin kullanışsız kalmasına sebep olmuştur. Ancak günümüzde, geçmişte işlenmesi gereken verinin çok üstünde veriyi işleyebilecek bilgisayar gücü, bellek kapasitesi ve özellikle sinir ağları yapısına uygun çok çekirdekli grafik işlemcilerinin devreye girmesiyle 2006 yılında Hinton ve arkadaşlarının öncü çalışmalarıyla birlikte günümüzde adlandırdığımız “Derin sinir ağları” terimi ortaya çıkmıştır [128]. Bu öncü çalışmanın ardından, yapay sinir ağlarında aşırı öğrenmeyi veya başka bir deyişle genelleştirme hatalarını düşürme çalışmaları [129,130], amaç fonksiyon optimizasyonu [131], aktivasyon fonksiyonu iyileştirme [132] ve yığın normalleştirme [133] gibi derin sinir ağlarının performansını ve kullanılabilirliğini artırıcı birçok çalışma yapılmış ve bu çalışmalar günümüzdeki derin öğrenme mimarilerinin başarısına ön ayak olunmuştur.

Derin öğrenme mimarileri içinde LeCun ve arkadaşlarının [134] önermiş olduğu evrimsel sinir ağları görüntü ve video işleme problemlerinde başarılı sonuçların elde edilmesini sağlamıştır. Rumelhart ve arkadaşlarının [135] çalışması temel alınarak ortaya çıkarılan uzun süreli ve ardışık veriyi modellemede başarılı özyineli sinir ağları (ÖSA, Recurrent Neural

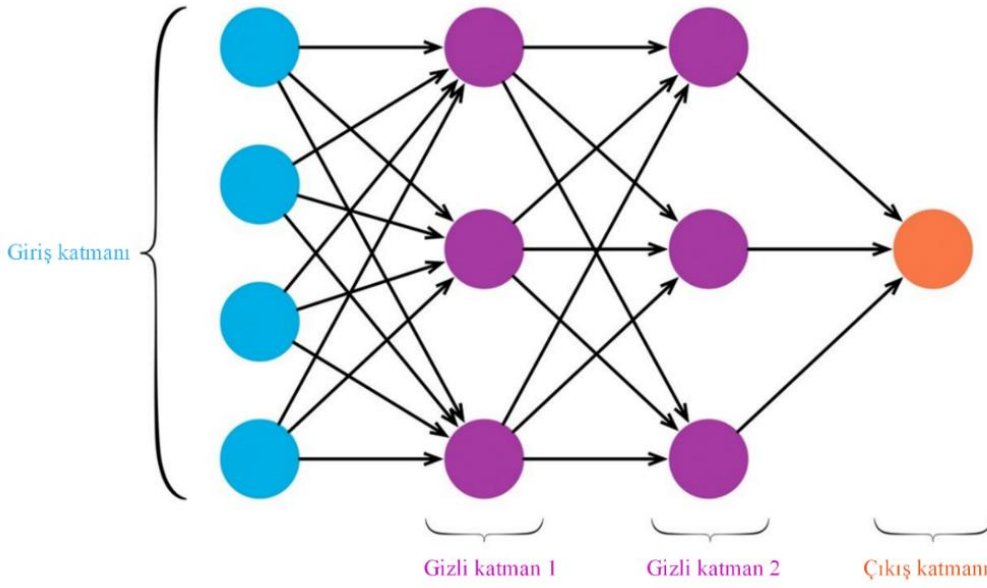
Network) ve ÖSA'ların kısıtlılıklarını gidermek amacıyla Hochreiter ve Schmidhuber [136] tarafından önerilen uzun-kısa süreli bellek (UKSB, Long short-term memory) ile de farklı problemlerin çözümünde başarılı bir şekilde uygulamıştır.

3.1.1. İleri beslemeli sinir ağları

Derin öğrenme yapılarını incelemeden önce yapay sinir ağları temellerini ve gerekli hesaplamaları açıklamak doğru olacaktır. Sinir ağlarında bir x_i girdisi ve bu girdiye karşılık üretilen y_i çıktısı bulunmaktadır. Verilen girdi ve çıktıların türleri, boyutları ve nitelikleri probleme göre değişiklik göstermektedir. Burada girdi görüntüye ait pikseller olabileceği gibi ses sinyaline ait bilgiler de olabilir. Bu tip ağlar, doğrusal veya doğrusal olmayan ilişkiye sahip problemleri x girdisi ve y çıktısı arasındaki ilişkiyi gözetimli öğrenme yöntemiyle ileri beslemeli modelleyerek çözmeyi amaçlamaktadır. Doğrusal olmayan problemlere çözüm bulabilmelerinin nedeni, çok katmanlı algılayıcı (Multi Layer Perceptron) ağlarının aksine sigmoid fonksiyonu gibi doğrusal olmayan bir aktivasyon fonksiyonu kullanmalarındır. İleri beslemeli sinir ağları giriş ve çıkış katmanlarının yanı sıra birden çok gizli katmana sahip olabilirler. Beyin yapısından esinlenerek bu birimlere nöron adı verilmiştir. Nöronlar arası ilişkiler giriş katmanından çıkış katmanına doğrudur. Gizli katman bulunuyor ise, bir gizli katmanın çıktısı diğer bir gizli katmanın girdisi olabilir. Bir ve iki gizli katmana sahip sinir ağı örnekleri sırasıyla Şekil 3.1 ve Şekil 3.2’de verilmiştir.



Şekil 3.1. Bir gizli katmana sahip ileri beslemeli sinir ağı örneği



Şekil 3.2. İki gizli katmana sahip ileri beslemeli sinir ağı örneği

Gizli katman sayısının birden çok olması derin bir ağı olduğu anlamına gelmektedir. Burada gizli katman sayısının artması problemin çözümünde kolaylık sağlamasına rağmen, ek bir hesaplama maliyeti getireceği anlamına da gelmektedir. Ayrıca bir katmandaki tüm nöronların istisnasız şekilde bir sonraki katman nöronlarının tamamına bağlanması durumu, ağı tam bağlı özellikte olduğu anlamına gelir ve bu yapıdaki mimariler tam bağlı ağı olarak adlandırılırlar.

Ağın girdisine karşılık elde edilen çıktı ile daha önceden bilinen ve gerçekte olması gereken çıktı arasındaki fark hatayı verir. Eğitim süresince, bulunan hatayı en aza indirmek için katmanların ağırlıkları güncellenir. Kullanılan fonksiyona göre her bir nöronun girdisi ile ağırlıklar işleme alınır ve çıkışlar elde edilir. Kuramsal olarak, elde edilen çıktı ile gerçekte olması gereken çıktı birbirine ne kadar benzer ise ağı o kadar iyi eğitilmiş denilmektedir. Ancak aşırı öğrenme ve genelleme gibi bu durumun bozulduğu problemler de bulunmaktadır.

Giriş katmanında herhangi bir matematiksel işlem yapılmaksızın ağı eğitime başlanır. Belirlenen aktivasyon fonksiyonu ile nöronların çıktıları elde edilir. Örneğin ileri beslemeli bir sinir ağında bir gizli katmana ait ağırlıklar girdi vektörü ile çarpılır.

$$z = Wx + b \quad (3.1)$$

Denklem 3.1’de W örnekteki gizli katmana ait ağırlığı, x nörona ait giriş vektörünü ve b (bias) yanlılık vektörünü ifade etmektedir. Elde edilen z ise toplam vektörünü temsil etmektedir ve z vektörüne daha sonra g aktivasyon fonksiyonu uygulanarak a çıkışı hesaplanır.

$$a = g(z) \quad (3.2)$$

Burada kullanılan aktivasyon fonksiyonu, doğrusal bir çözümü bulunmayan veri üzerinde başarısız olan adım fonksiyonu dışındaki veri problemleri üzerinde başarılı olan hiperbolik tanjant ve sigmoid fonksiyonlarından biri olabilmektedir. Girdi olarak kullanılan vektörlerin her birine ayrı ayrı uygulanan lojistik fonksiyonu sıfır ile bir arasında çıktı değerleri üretirken, tanjant fonksiyonu eksi bir ile bir arasında çıktılar üretir. İlk olarak 2010 yılında Nair ve Hinton tarafından ortaya çıkarılan ve yüksek popülerlik kazanan bir diğer aktivasyon fonksiyonu Rektifiye Doğrusal Birim (RDB, Rectified Linear Unit) [132]’dur. Aktivasyon fonksiyonu elde edilen tüm z toplam vektörleri için uygulanır ve gizli katmanın çıkışı a vektörü ilgili gizli katmanın çıkışı olarak bulunur. Son gizli katmanın çıkışı, çıkış katmanına girdi olarak ayarlanır ve gizli katman ile çıkış katmanı arasındaki hesaplamalar yapılır.

$$z^n = W^n x + b \quad (3.3)$$

$$y = f(z^n) \quad (3.4)$$

Çıkış vektöründe ise softmax fonksiyonu uygulanır ve Denklem 3.4’de f ile gösterilmiştir. Bu fonksiyonla girdiye karşılık olan çıktı elde edilmiş olur. Her bir girdiye karşılık toplamaları 1 olacak şekilde 0 ile 1 arasında olasılık tahmini ortaya koyulur ve bu sayede sinir ağının tahmini oluşur.

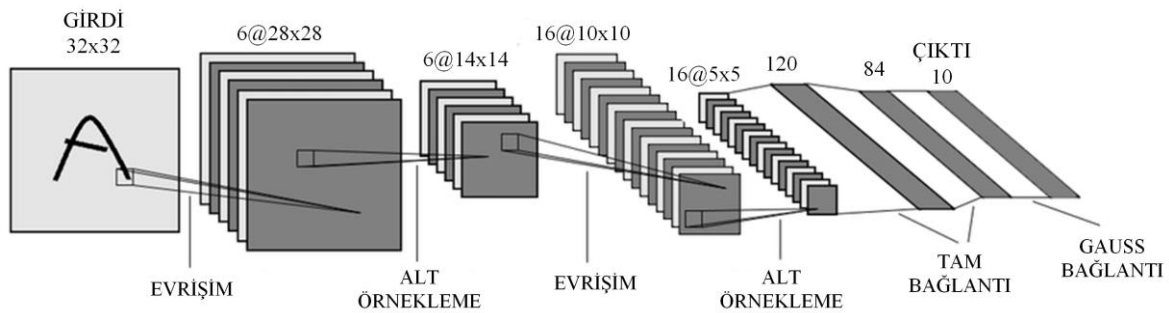
$$P(y = k|x) = \frac{e^{f_k(x)}}{\sum_{j=1}^K e^{f_j(x)}} \quad (3.5)$$

Denklem 3.5’de k . sınıf için $f_k(x)$ sonlu olasılıkları ifade eder ve k adet sınıf için k adet çıkış bulunduran bir sinir ağı bulunur. Softmax fonksiyonu çok kategorili fakat tek etiketli problemlerin çözümünde oldukça yaygın kullanılmaktadır. Ayrıca, sigmoid fonksiyonu bir girdinin sadece tek bir çıktıya gitmesini değil, her bir çıkış için ayrı ayrı sonsal olasılıklarını hesaplayabilmektedir.

Elde edilen çıktı ile gerçek çıktıların karşılaştırılmasında ise bir hata fonksiyonuna ihtiyaç vardır. Bu hata fonksiyonu sayesinde sinir ağının performansı ölçülüp geri besleme sağlanmaktadır. Hata fonksiyonları çok çeşitli olabileceği gibi en yalın olanı ve yaygın kullanılanı kare hata fonksiyonudur. Kare hata dışında çeşitli entropi fonksiyonları da kullanılabilir.

3.1.2. Evrişimsel sinir ağları

Evrişimsel sinir ağları Hubel ve Wiesel'in canlı görme merkezinden ilhamla geliştirmiş olduğu çok katmanlı algılayıcıdır [137]. Canlıların özellikle eşek gibi hayvanların görme merkezlerinin incelenmesi sonucunda, görme merkezindeki hücrelerin görsel alan üzerinde yerel maskeler gibi hareket ettiği ve görsel bölgeye ait belirgin özelliklerin ortaya çıkarılmasını gerçekleştirdiği gözlemlenmiştir. Görüntüler üzerinde çalışmaya imkân sağlayan evrişimsel sinir ağları da benzer şekilde görüntü üzerindeki önemli bilgilerin çıkartılmasına odaklanmaktadır. Derin öğrenme mimarileri arasında en temel ve bilinen model olan evrişimsel sinir ağları birçok tam bağlı evrişim katmanları ve örnekleme (pooling) katmanlarından meydana gelmektedir. Evrişimsel sinir ağının çıkış katmanı ise sınıflandırma için kullanılmaktadır. Şekil 3.3'de Lecun ve arkadaşları tarafından önerilen ilk evrişimsel sinir ağı modeli LeNet [134] gösterilmiştir.



Şekil 3.3. LeNet evrişimsel sinir ağı mimarisi

Şekil 3.3.'de verilen LeNet mimarisinde art arda eklenmiş katmanlar aracılığıyla giriş verisinin önemli niteliklerini çıkartılmakta ve bu bilgi en son katmanda sınıflandırma için kullanılmaktadır. Gözetimli öğrenme ile çalışan bu yöntemde evrişimsel sinir ağı tarafından belirlenen sınıf etiketleriyle gerçekte olması gereken etiketler arasında bir hata meydana gelmektedir. Bu noktada geriye yayılım algoritması kullanılarak hatanın optimize edilmesi için hesaplanan ağırlıkların güncellenmesi gerçekleştirilir. Eğitimin başarısında, kullanılan

sinir ağı mimarisinin yanı sıra eğitimde kullanılan verinin örnek uzayını iyi ve dengeli temsil etmesinin de önemi büyüktür. Mimari her ne kadar probleme uygun ve etkili olsa da dengesiz bir eğitim verisi öğrenmenin başarısız olmasına sebep olabilmektedir. Gözetimli öğrenme için kullanılan evrişimsel sinir ağları görüntü ve video işleme alanlarında yaygın olarak uygulansa da ses işleme, doğal dil işleme ve finans gibi çeşitli alanlarda etkili uygulamalar geliştirilmiştir.

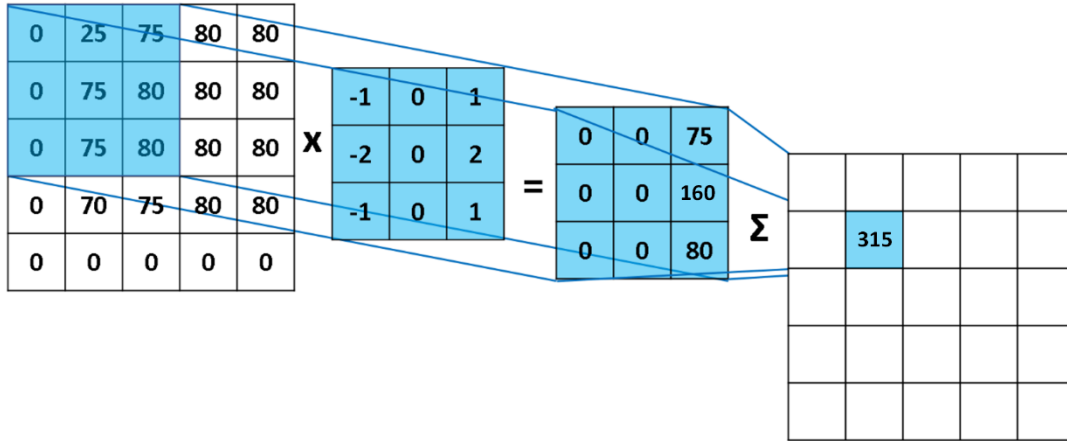
Giriş katmanı

Eğitim ve test verisinin doğrudan ağa sunulduğu katmandır. Eğitim için sunulan giriş verisinin yeterli miktarda olması ve iyi bir örneklem sunması ağın performansını, öğrenme ve sınıflandırma başarısını önemli derecede artırır. Bunun yanında, veri boyutunun büyük olması ve katman açısından zengin bir mimari ile çalışmak büyük işlemci ve geçici bellek birimi ihtiyacı doğurabilmektedir.

Evrişim katmanı

Evrişimsel sinir ağlarının temel yapı taşı olan evrişim katmanı, giriş verisindeki görüntü gibi özelliklerin algılanması için kullanılmaktadır. Derin öğrenme modelinde, giriş görüntüleri belirlenen büyüklükte kareler halinde ele alınır. Filtreler bu kareler üzerinde adım uzunluğu (stride) kadar kaydırılarak hesaplamalar yapılır. Evrişim işlemi matrislerin iç çarpımı (dot product) işlemidir. Denklem 3.6'da bir iç çarpım örneği verilmiştir. Şekil 3.4.'de giriş verisi üzerinde filtre uygulaması ve oluşan özellik haritasının bir hücresinin örneği gösterilmiştir.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} w & x \\ y & z \end{bmatrix} = aw + bx + cy + dz \quad (3.6)$$

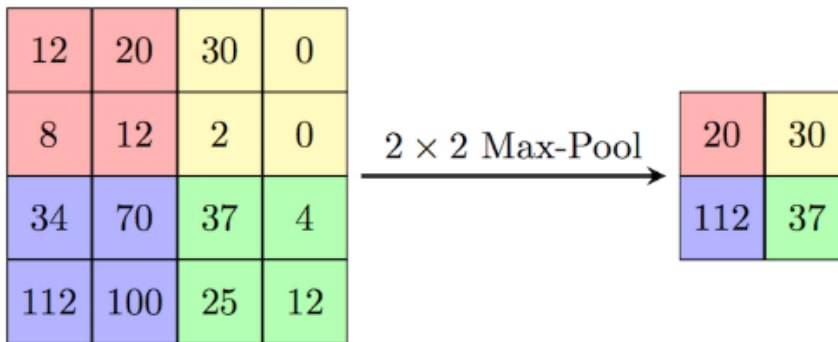


Şekil 3.4. Evrişim katmanında filtre hesaplama örneği

Evrişim katmanındaki her bir girdi için özellik haritaları oluşturulur ve eğitim verisinin işlenmesi süresince güncellenerek öğrenme gerçekleşir. Katman çıktılarında oluşabilecek negatif değerlerin sifıra eşitlenmesi için RDB aktivasyon fonksiyonu uygulanır.

Örnekleme Katmanı

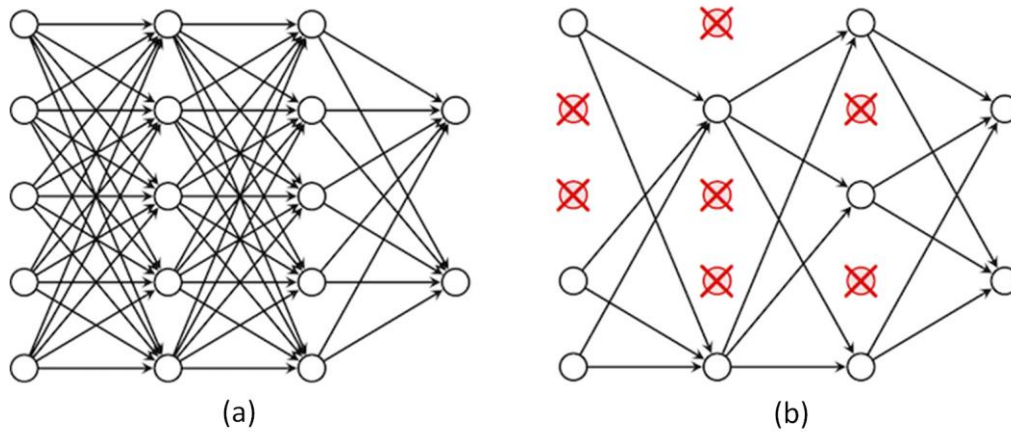
Örnekleme katmanı (pooling layer) evrişimsel katmanların arasına yerleştirilir ve verinin boyutlarının indirgenmesini sağlar. Böylece eğitimin daha hızlı ve sağlıklı yapılmasına imkân sunar. Şekil 3.5.'de yaygın olarak kullanılan maksimum örnekleme (max-pooling) yöntemi için bir örnek verilmiştir. 2 x 2 boyutunda bir filtre seçilen örnekte, her bir kare için maksimum değer seçilerek matris küçültülür. Bu sayede verinin en az kayıpla daha etkili ve hızlı bir eğitime tabi olması amaçlanır.



Şekil 3.5. Evrişimsel sinir ağında 2x2 boyutunda filtre ile maksimum örneklendirme

Seyreltme Katmanı

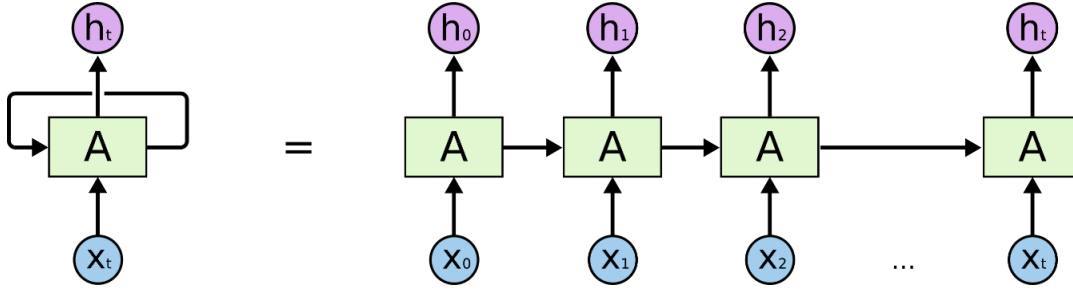
Ağın eğitimi yapılırken eğitimde kullanılan girişleri ezberlemesi problemi doğabilmektedir. Seyreltme katmanı (Dropout Layer), ezberleme probleminden kaçınabilmek için 2014 yılında Srivasta ve arkadaşları tarafından önerilmiş ve ezberleme problemiyle önemli ölçüde başa çıkabildiği ve eğitimin kalitesini arttırdığı kanıtlanmıştır [129]. Şekil 3.6’da seyreltme yöntemiyle (a)’da verilmiş olan sinir ağının bazı nöronları kaldırılarak (b) ile verilen yeni bir ağ yapısı elde edilmiştir.



Şekil 3.6. (a) Geleneksel sinir ağı mimarisi örneği (b) Seyreltme uygulanmış sinir ağı örneği

3.1.3. Özyineli sinir ağları

Geleneksel ileri beslemeli sinir ağları her bir girdiyi birbirinden bağımsız olarak ele alır. Fakat bu varsayım doğal dil işleme, zaman serisi problemleri ve video gibi zamana göre ilişki barındıran alanlarda doğru olmayabilir. Özyineli sinir ağları (ÖSA, recurrent neural networks), ağ girişinin özelliklerinin çıkarılmasının yanında zaman ve dil dizilimi gibi kavramlara dayalı ardışık girişlerin arasındaki ilişkilerin ortaya çıkarılması için tasarlanmıştır. ÖSA dil modelleme [120], konuşma algılama [138], tıbbi tahminleme [139] ve finansal tahminleme [140] gibi ardışık verinin anlamlandırılmasına dayalı birçok probleme çözüm sunmuştur. Katmanlar arasındaki çıkışların bir sonraki katmanın girişine bağlanması ile ardışık ilişkilerin modellenmesini mümkün kılmaktadır. Bir katmanda belirli bir zamanın verisi işlenirken kendinden önce gelen veri bilgisinin tutulması gerekir. Bu amaçla, özyineli sinir ağı girdi serisindeki son eleman tarafından belirlenen gizli durum parametresi bulunur. Özyineli sinir ağı mimarisi Şekil 3.7’de verilmiştir.



Şekil 3.7. Özyineli sinir ağı mimarisi

Şekil 3.7’deki sinir ağı mimarisinde gösterildiği üzere zaman adımı t ’deki seri elemanı işlenirken, zaman adımı $t - 1$ ‘e ait gizli durumu kullanılmaktadır. Özyineli sinir ağı çıkışının matematiksel gösterimi şu şekildedir:

$$s_t = \sigma(Ux_t + Ws_{t-1} + b_s) \quad (3.7)$$

$$h_t = \sigma(Vs_t + b_h) \quad (3.8)$$

Denklem 3.7’de x_t , t anındaki giriş bilgisi, s_{t-1} ise $t - 1$ anına ait yani bir önceki girişin gizli durum bilgisidir. Mevcut durum ve mevcut ağ çıkışı bu değerler üzerinden hesaplanmaktadır. U ileri beslemeli sinir ağlarındakine benzer olarak mevcut girişin çıkıştaki önemini belirleyen ağırlıklardır. W ve V ise ek ağırlıklar olup geçmiş verinin çıkıştaki ağırlığını belirlerler.

3.2. Uzun Kısa Süreli Bellek

Geleneksel ileri beslemeli sinir ağında, tüm giriş ve çıkışların birbirinden bağımsız olduğu varsayılır. Fakat ardışık işlemler yapılırken girişlerdeki zamansal değişim de önem arz etmektedir. Örneğin, bir kelime tahmin modelinde daha önceki kelime girişlerinden de bilgi çıkarımı yapılması gerekir. Bu amaçla kullanılan ÖSA benzer şekilde ileri beslemeli olarak çalışır. Ancak ÖSA’nın çıkışı sadece model girişinden değil tüm giriş geçmişinden etkilenir. Teoride ÖSA modelleri keyfi ardışık uzunluktaki veri için kullanılabilir. Fakat pratikte kaybolan gradyanlar yüzünden sadece birkaç adım geriye bakabilmektedir.

Gradyan kaybolması problemiyle başa çıkabilmek için, uzun kısa süreli bellek (UKSB) modeli önerilmiştir. Bu model unutmama kapısı adında bir tekrarlı kapı içermektedir. UKSB

modeli geri yayılım hatalarının yok olmasını veya gradyan patlamasını önlemektedir. Bu sayede UKSB modeli, hareket bilgisi gibi yüksek seviye bilgilerin elde edilebilmesi için uzun ardışık veriyle çalışabilmektedir. Tipik bir UKSB biriminin mimarisi Şekil 3.8 ile gösterilmiştir. Şekilde mavi hat daha iyi kararlar alınması için iç durumun geriye bakmasına olanak veren gözetleme deliği olarak adlandırılır. Geleneksel UKSB biriminin matematiksel hesaplamaları Denklem 3.9-3.14 ile özetlenebilir.

$$f_t = \sigma(W_f \otimes [h_t - 1, x_t] + b_f) \quad (3.9)$$

$$f_t = \sigma(W_i \otimes [h_t - 1, x_t] + b_i) \quad (3.10)$$

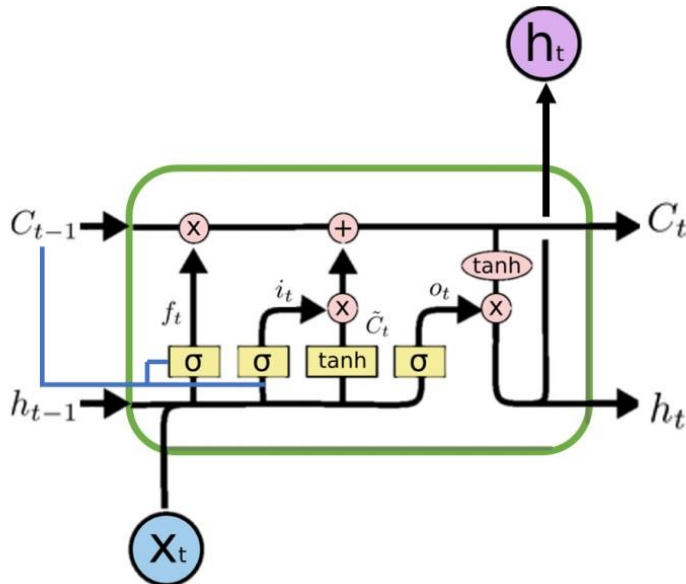
$$\hat{C}_t = \tanh(W_c \otimes [h_t - 1, x_t] + b_c) \quad (3.11)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \hat{C}_t \quad (3.12)$$

$$o_t = \sigma(W_o \otimes [h_t - 1, x_t] + b_o) \quad (3.13)$$

$$h_t = o_t \otimes \tanh(C_t) \quad (3.14)$$

UKSB biriminin matematiksel aşamalarındaki Denklem 3.9 unutmama katmanını temsil eder. Denklem 3.10 ve 3.11 yeni bilginin eklendiği yerdir. Denklem 3.12 eski ve yeni bilginin birleştirildiği aşamadır. Denklem 3.13 ve 3.14 ise bir sonraki adım için UKSB birimi tarafından öğrenilen çıktıyı temsil eder. x_t değişkeni giriş vektörünü, h_t gizli durumu ve C_t ise zaman t için hücre durumunu verir. W eğitilebilir ağırlık matrisi, b yanlılık değeri ve $\otimes$ ise Hamadart işlemidir.



Şekil 3.8. UKSB biriminin mimarisi.

3.3. Evrişimsel Uzun Kısa Süreli Bellek

Evrişimsel uzun kısa süreli bellek (EUKSB, convolutional long short term memory) UKSB'nin bir çeşidi olmakla birlikte ilk olarak Shi ve arkadaşları [141] tarafından önerilmiştir. Geleneksel UKSB'ye göre evrişimsel uzun kısa süreli bellek yönteminde, matris işlemleri yerini evrişim işlemlerine bırakmaktadır. Girişten gizli katmana ve gizli katmandan çıkış katmanına doğru evrişim kullanmak, EUKSB'in daha az ağırlığa ihtiyacı duymasını ve uzaysal özellik haritalamada daha iyi sonuçlar vermesini sağlamaktadır. EUKSB mimarisinin matematiksel aşamaları Denklem 3.15-3.20 ile özetlenebilir.

$$f_t = \sigma(W_f \otimes [h_t - 1, x_t, C_{t-1}] + b_f) \quad (3.15)$$

$$f_t = \sigma(W_i \otimes [h_t - 1, x_t, C_{t-1}] + b_i) \quad (3.16)$$

$$\hat{C}_t = \tanh(W_c \otimes [h_t - 1, x_t] + b_c) \quad (3.17)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \hat{C}_t \quad (3.18)$$

$$o_t = \sigma(W_o \otimes [h_t - 1, x_t, C_{t-1}] + b_o) \quad (3.19)$$

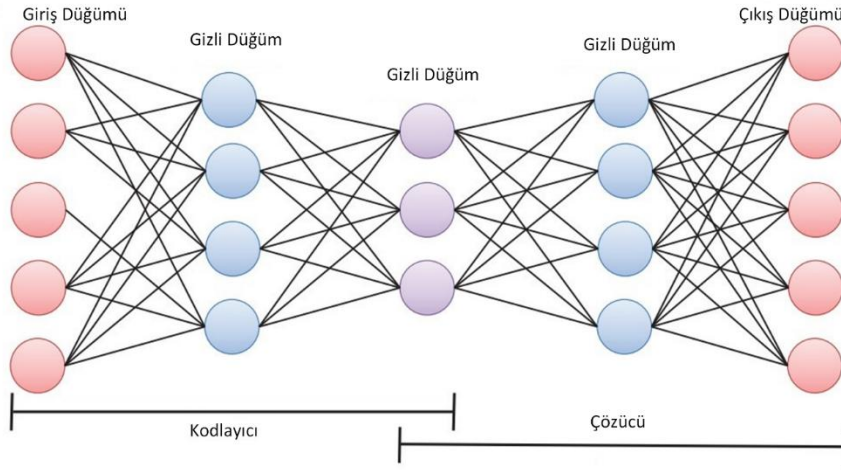
$$h_t = o_t \otimes \tanh(C_t) \quad (3.20)$$

EUKSB denklemleri tipik UKSB denklemlerine benzemektedir. Fakat her bir bağlantı için ağırlık kümesi evrişim filtreleriyle değiştirilir ve giriş resimlerle beslenir. Bu durum EUKSB modelinin resimlerle daha iyi çalışmasını sağlamaktadır.

3.4. Otokodlayıcılar

Otomatik kodlayıcı ilk olarak Rumelhart ve arkadaşları [135] tarafından önerilmiştir. Etiketsiz veri kümelerinden öğrenebilen otokodlayıcı yöntem, gizli katmanlı gözetimsiz bir yapay sinir ağıdır. Otokodlayıcıların altında yatan temel prensip model çıkışında, giriş verisini tekrar oluşturmaya çalışmaktır. Otokodlayıcılar $x \in \mathbb{R}^d$ girişini haritalama fonksiyonu $h = f_\Theta = \sigma(Wx + b)$ ve $\Theta = \{W, b\}$ parametrelerini kullanarak $h \in \mathbb{R}^{d'}$ gizli katmanına haritalar. Girişi tekrar oluşturmak için ise $f: y = f_{\Theta'}(h) = \sigma(b'W'h + b')$ ters haritalama fonksiyonu ve $\Theta' = \{W', b'\}$ parametrelerini kullanılır. Giriş katmanından gizli

katmana doğru öğrenilen W , kodlayıcı aşamasını oluşturur. Gizli katmandan çıkış katmanına doğru öğrenilen W' ise, çözücünün yapısını oluşturur. Çözücü parametreleri doğrudan kodlayıcı parametreleriyle ilgilidir ve $W' = W^T$ eşitliğindeki W 'nin transpozuna eşittir [142].



Şekil 3.9. Geleneksel otokodlayıcı mimarisi

Otokodlayıcılar giriş verisi için etikete ihtiyaç duymazlar yani gözetimsiz bir öğrenme gerçekleştirirler. Her bir giriş x_i ve girişlere karşılık gelen her bir çıkış y_i için yeniden oluşturma hatası e 'yi minimize etmek için bir geri yayılım algoritmasına ihtiyaç duyulmaktadır. Geri yayılım algoritması ile yeniden oluşturma hatası e kullanılarak kodlayıcı parametreleri W ve çözücü parametreleri W' güncellenir ve Denklem 3.21'deki gibi gösterilir.

$$e(x, y) = \frac{1}{2N} \sum_{i=1}^N \|x_i - y_i\|_2^2 \quad (3.21)$$

Literatürde otokodlayıcıların farklı türleri bulunmaktadır. Gürültü giderici otokodlayıcı [143] en popüler olanlarından birisidir. Gürültü giderici otokodlayıcı (GGO), gürültü barındıran bir giriş için mümkün olduğunca gürültüsü temizlenerek yeniden oluşturulmuş bir halini çıktı olarak üretir. Deneyisel çalışmalarda önce deney ortamında Gauss ve Speckle gibi üzerinde çalışılan gürültü tipleri ile görüntüler gürültülü hale getirilir. Daha sonra gürültüsü giderilerek yeniden oluşturulan görüntü ile orijinal görüntü karşılaştırılarak etkinlik ölçülmektedir. Otokodlayıcıların bir başka uygulama alanı ise görüntülerin sıkıştırılmasıdır. Bu yöntemlerde, bir giriş görüntüsü daha küçük boyutlu gizli bir katmanda

sıkıştırıldığı haliyle temsil edilmekte ve bu gizli sunum ters evrişimle tekrar görüntüsüne yakın bir giriş elde etmeye çalışmaktadır [143]. Otokodlayıcıya ait genel mimari Şekil 3.9'da örneklendirilmiştir.

3.5. Evrişimsel Otokodlayıcılar

Geleneksel otokodlayıcı ve gürültü giderici otokodlayıcı yöntemlerinin en büyük kısıtlılığı video ve resim gibi iki ve üç boyutlu görüntüleri temsil edememesidir [142]. Bu dezavantajı ağ parametrelerinde fazlalığa ve resimler üzerindeki yerel bilgilerin kaybolmasına neden olur. Konu özelinde örneğin; video çerçeveleri üzerindeki normal olmayan eylemler, çerçeveler üzerinde yerel bulunur ve geleneksel otokodlayıcılar tarafından kaybedilirler. Bu problemle başa çıkabilmek için Masci ve arkadaşları [142] evrişimsel otokodlayıcıyı önermişlerdir. Evrişimsel otokodlayıcılarda geleneksel otokodlayıcılardan farklı olarak ağırlıklar girişin tüm konumları için paylaşılmaktadır. Bu yolla evrişimsel sinir ağlarında olduğu gibi uzaysal bilginin saklanması mümkün olmaktadır. Evrişimsel otokodlayıcılar için kayıp fonksiyonu Denklem 3.22'de verilmiştir.

$$e(x, y, W) = \frac{1}{2N} \sum_{i=1}^N \|x_i - y_i\|_2^2 + \lambda \|W\|_2^2 \quad (3.22)$$

Denklem 3.22'deki λ eğitim süreci boyunca $\|W\|_2^2$ düzenleme terimi için bir düzenleme parametresidir. Evrişimsel sinir ağlarında olduğu gibi evrişimsel otokodlayıcılarda evrişimsel örnekleme, ters evrişimsel örnekleme ve dağıtım katmanları bulundurabilmektedir.

Evrişimsel katman bir filtrenin bilgisini, skalar bir değere özetler ve bir filtrenin sabit alıcı alanındaki çoklu giriş aktivasyonlarını, özellik haritasındaki tek bir aktivasyon çıkışına bağlar. Bir x girişi için k 'nıncı özellik haritasının gizli katman haritalaması Denklem 3.23'deki gibi tanımlanır.

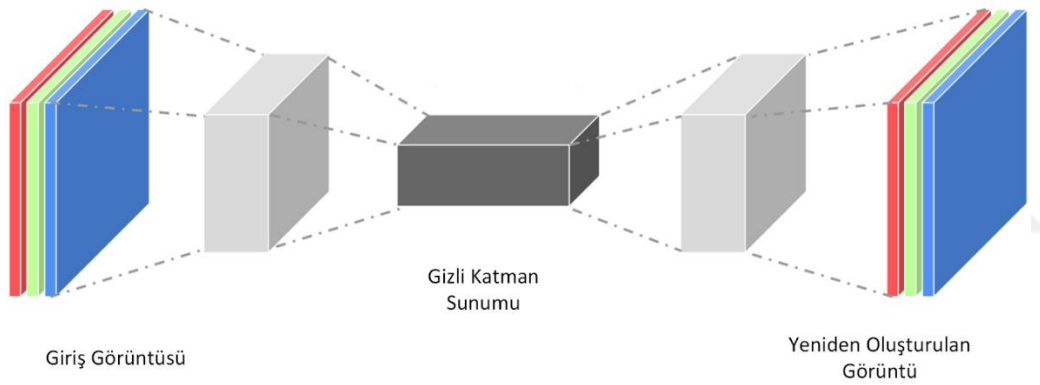
$$h_k = \sigma(x * W_k + b_k) \quad (3.23)$$

Denklem 3.23'de b yanlılığı, σ aktivasyon fonksiyonunu ve $*$ sembolü 2 boyutlu evrişimi temsil eder. Yeniden üretim ise Denklem 3.24 ile elde edilir.

$$y = \sigma\left(\sum_{k \in H} h_k * \tilde{W}_k + c\right) \quad (3.24)$$

Denklem 3.24’de c yanlılık, H gizli özellik harita grubunu, $\tilde{W}$ ise W boyutların ağırlıkları üzerinde çevirme işlemini (flip operation) ifade eder.

Ters evrişim katmanı, evrişim katmanının ters işlemini yapar. Ters evrişim katmanlarından öğrenilen filtreler giriş şeklinin yeniden oluşturulmasında temel görevi görmektedir [110]. Evrişimsel ve ters evrişimsel katmanlar düzenli bir şekilde birleştirilerek evrişimsel otokodlayıcı mimarisi oluşturulur. Evrişimsel ilk katmanlardaki filtreler düşük seviyeli özellikleri çıkarırken, son katmanlar hareket ve görünüm gibi yüksek seviyeli özellikleri çıkartabilir. Evrişimsel otokodlayıcıya ait genel bir mimari örneği Şekil 3.10’da verilmiştir.



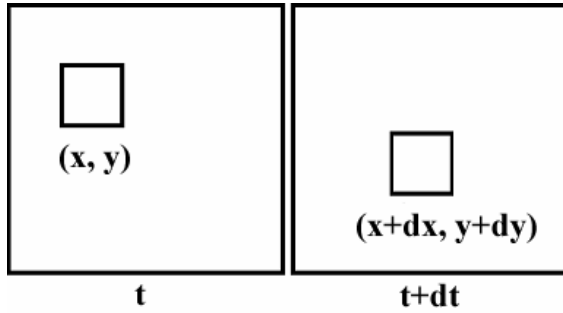
Şekil 3.10. Evrişimsel otokodlayıcı genel mimarisi

3.6. Optik Akış

Görüntü içindeki bir nesnenin şiddet bilgisinden faydalanılarak görüntü üzerindeki yeri tespit edilip sonraki çerçevelerdeki konumundan faydalanılarak hız tespiti yapılabilir. İlk olarak James Jerome Gibson [144] tarafından önerilen optik akış yöntemi olmak üzere literatürde görüntü üzerindeki her bir pikselin hızını bulabilmek için çeşitli yaklaşımlar önerilmiştir. Hız bilgisini elde etmek için ardışık görüntü dizileri arasında her şiddet noktasının hareket yönünü bulmak gerekmektedir. Görüntüdeki takip edilen nesne hareket ettiği anda meydana gelen harekete “Optik Akış” denilmektedir. Burada optik akış hız ve yön bilgisidir.

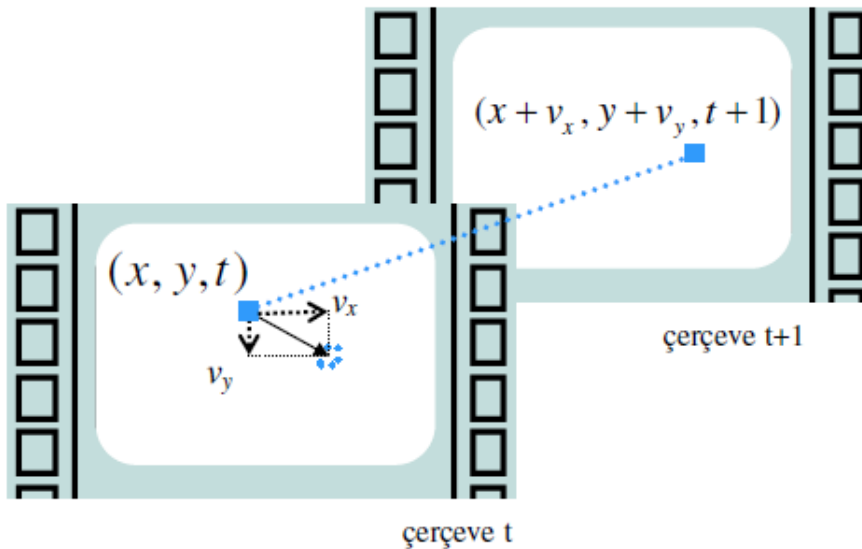
Hız bilgisinin sadece iki boyutlu şiddet bilgisinden tahmin edilmesi mümkün değildir. Nesne hareketinin görsel sensör tarafından algılanması sonucu iki boyutlu düzlem üzerinde meydana gelen görünür hız, optik akış olarak adlandırılır ve kullanılan yöntemler ile optik akış alanı, hız alanına yakınlştırılır. Bir görüntüde bulunan $r = (x, y)$ noktasının parlaklığı $I(x, y, t)$ olarak ifade edilmektedir. Görüntüde belirli bir noktanın parlaklığı zamanla sabittir ve sonraki çerçevede de aynı kalacağı varsayılmaktadır. Bu varsayım ile problemin çözümü Denklem 3.25'deki eşitliğe düşürülebilmektedir.

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (3.25)$$



Şekil 3.11. Hareket eden noktanın şiddet değişmezliği gösterimi

Bir noktanın şiddeti değişirse hareket tespiti hatalı yapılmaktadır. Optik akış yöntemlerinde şiddetin değişmeyeceği varsayılmaktadır. Bir noktanın şiddet değişmezliği gösterimi Şekil 3.11'de verilmiştir. Bir noktanın çerçeveler boyunca yer değiştirmesi ile hız bileşenleri bulunur. Şekil 3.12'de hız bileşenlerinin gösterimi sunulmuştur.



Şekil 3.12. Hız bileşenlerinin gösterilmesi

Parlaklığın uzay-zamansal düzlemde değişmediği kabulü ile, $r = (x, y)$ pikseli etrafındaki (dx, dy, dt) küçük değişimi için bir noktadaki parlaklığın Taylor açılımı Denklem 3.26 ile gösterilebilir:

$$I(x + dx, y + dy, t + dt) = I(x, y, t) + dx \frac{\partial I}{\partial x} + dy \frac{\partial I}{\partial y} + dt \frac{\partial I}{\partial t} + \text{H.O.T} \quad (3.26)$$

Lucas ve Kanade[145] ve Horn ve Schunck [146] optik akışları en yaygın bilinen yöntemlerdir. Lucas ve Kanade optik akış yöntemi normal hız hesabını mümkün kılarken Horn ve Schunck yöntemi mümkün kılmamaktadır.

Görüntü dizilerinin işlenmesindeki temel problem optik akışın veya görüntü hızının ölçülme zorluğudur. Bu bölümde anlatılan yöntemlerin genel amacı, iki boyutlu hareket alanına yaklaşık bir hesaplama yapmaktır.

Optik akış hesaplaması için birçok teknik ileri sürülmekle birlikte optik akış teknikleri aşağıda bahsedilen üç gruptan birine bağlı olarak sınıflandırılabilir [147]:

- Fark Teknikleri: uzay-zamansal şiddet türevlerinden görüntü hızını hesaplar[146].
- Frekans-temelli Teknikler: hız ayarlı filtre çıkışındaki enerji/faz bilgisini kullanırlar.
- Eşleştirme Teknikleri: az sayıda görüntüden (genellikle iki veya üç görüntü dizisinden) değişik görüntü özelliklerini eşleştirerek görüntü yer değişimlerini hesaplar.

Bu üç grup yaklaşım arasında uygulama ve performans farklılıkları olmasına karşın genelde birçok bakımdan birbirlerine denk olarak düşünülürler. Genel yapısı bakımından bu teknikler üç işleme aşamasına dayanarak incelenebilir:

- Görüntü üzerinde ilgilenilen işaret yapısını elde edebilmek ve işaret/gürültü oranını arttırmak için alçak-geçiren veya bant-geçiren filtre kullanarak ön-filtreleme veya yumuşatma,
- Hızın normal bileşenlerini hesaplamak için temel ölçümlerin hesaplanması; örneğin uzay-zamansal türevlerin elde edilmesi,
- İki boyutlu akış alanını üretebilmek için ön filtreleme ve temel hesaplamaların entegrasyonunun yapılması.

3.6.1. Farneback Yöntemi

Farneback yöntemi [148], polinom açılım dönüşümü kullanarak iki video çerçevesindeki hareket komşuluğunu tahmin eden bir yöntemdir. Polinom açılımı bir polinom ile her bir pikselin komşularının tahmin edilmesi mantığına dayanır. Bu polinomlar ikinci dereceden polinom olup yerel sinyal modelini verir ve lokal koordinat sisteminde şu şekilde ifade edilir.

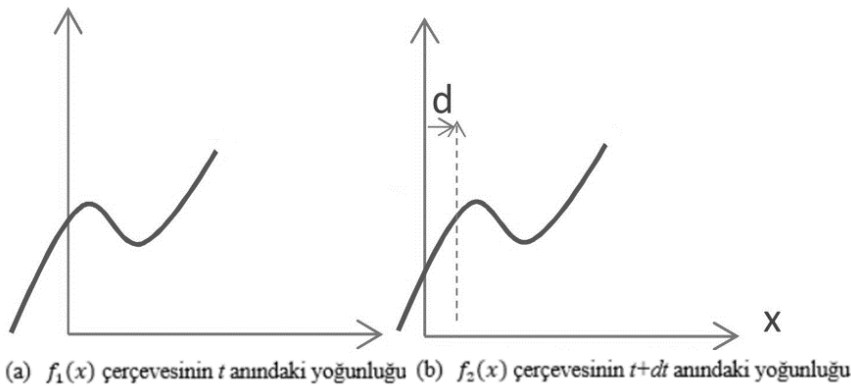
$$f(x) \sim x^T A x + b^T x + c \quad (3.27)$$

Denklem 3.27’de A simetrik bir matrisi, b bir vektörü ve c bir sabiti temsil eder. Katsayılar, komşu sinyal değerlerine uygun olarak ağırlıklı en küçük karelerden tahmin edilir. İlk kare için polinom Denklem 3.28’de ve bir sonraki çerçeve değişim ile Denklem 3.29’da verilmiştir.

$$f_1(x) \sim x^T A_1 x + b_1^T x + c_1 \quad (3.28)$$

$$f_2(x) = f_1(x - d) = (x - d)^T A (x - d) + b^T (x - d) + c_1 = x^T A_2 x + b_2^T x + c_2 \quad (3.29)$$

Yer değiştirme d komşu iki video çerçevesinin polinomları ile tahmin edilir. Çerçeve $f_1(x)$ t zamanda, $f_2(x)$ ise $(t + dt)$ zamanda elde edilmiştir. x ’in tek boyutlu basit bir gösterimi Şekil 3.13’de gösterilmiştir ve d yer değişimini göstermektedir.



Şekil 3.13. İki ardışık video çerçevesinin komşuluk polinomları (a,b)

İkinci dereceden polinomların katsayıları eşitlendiğinde:

$$A_2 = A_1 \quad (3.30)$$

$$b_2 = b_1 - 2A_1d \quad (3.31)$$

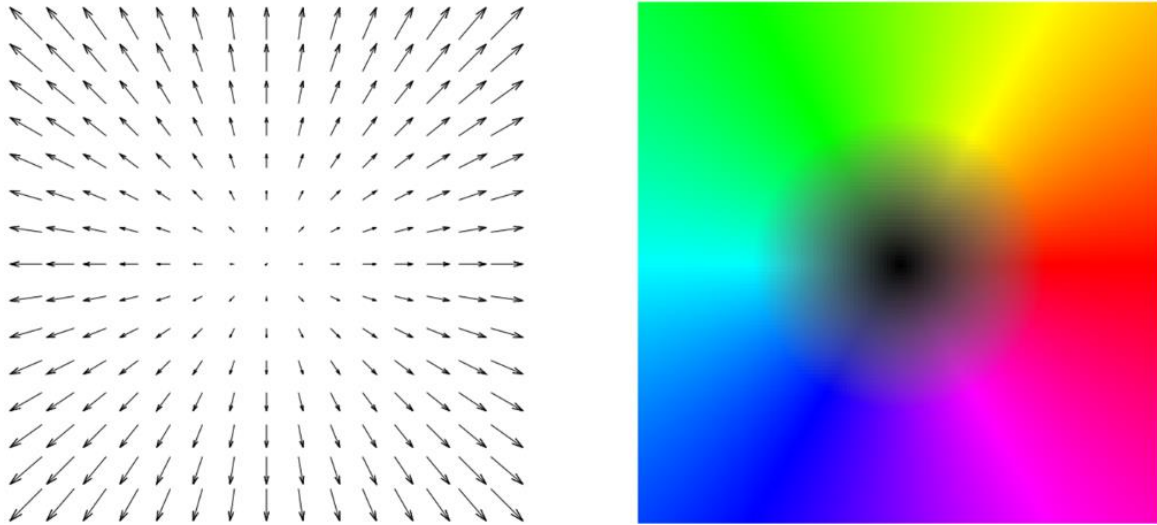
$$c_2 = d^T A_1 d - b_2^T x + c_2 \quad (3.32)$$

Denklem 3.31’de en azından A_1 tekil olmadığında yer değışimi (d) çözülebilir.

$$2A_1d = -(b_2 - b_1) \quad (3.33)$$

$$d = -\frac{1}{2}A_1^{-1}(b_2 - b_1) \quad (3.34)$$

Elde edilen yer değıştirme vektörü, yer değışimine karşılık gelen hareket vektörünü verir. Optik akış yönteminin son aşamasında ise yer değıştirme vektörü renk kodlarıyla kodlanır. Kullanılan renk kod şeması Şekil 3.14’de verilmiştir.



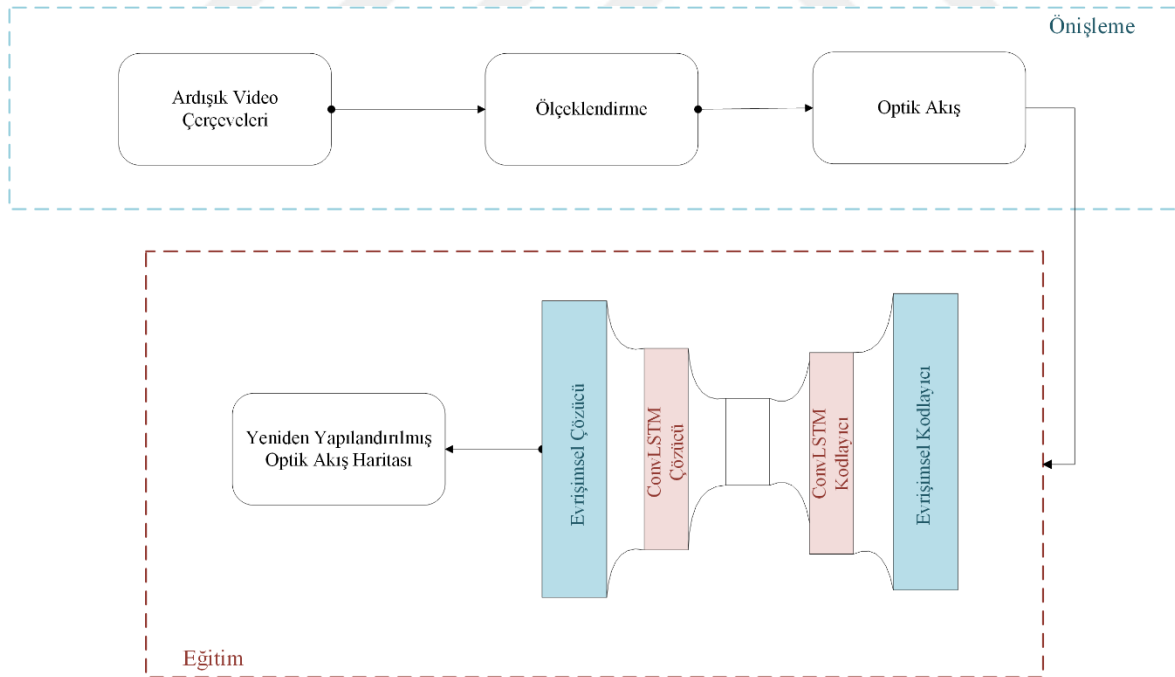
Şekil 3.14. Vektör alanı ve optik akış renklendirme şeması

Şekil 3.14’de verilen renkler optik akış vektörlerinin yönlerini temsil eder. Renklerin yoğunluğu ise vektörün yön büyüklüklerini ortaya çıkarır. Koyu renkler nispeten düşük akış büyüklüklerini temsil ederken, mavi kırmızı, yeşil ve diğer renkler belli bir yönü ve renk yoğunluğu akış büyüklüğünü vermektedir.



4. NORMAL OLMAYAN DAVRANIŞ ALGILANMASI

Tez çalışmasında, videolardaki normal eylemlerin şablonunu öğrenmek için bir evrişimsel otokodlayıcı model önerilmiştir. Önerilen yöntemin altında yatan fikir normal olmayan eylemleri barındıran video çerçevelerinin, düzenli eylemlerin meydana geldiği çerçevelere göre oldukça farklı hareket şablonu oluşturacaklarıdır. Model girişi için her biri ardışık 8 çerçeveden elde edilen optik akış haritaları kullanılmıştır. Model eğitiminde sadece düzenli eylemleri içeren eğitim video çerçeveleri kullanılmıştır. Bu sayede model, eğitim bittikten sonra sadece düzenli hareket şablonunun karmaşık dağılımını öğrenir. Model çerçevelerin oluşturduğu optik akış haritasına karşılık gelen yeniden oluşturulmuş optik akış haritası üretir. Eğer giriş olarak verilen eylemler düzenli hareketler içeriyorsa, yöntem düzenli hareket şablonlarını modelleyebildiği için çıkış olarak girişe yakın optik akış haritası üretmesi beklenir ve düşük yeniden oluşturma hatası oluşur. Ancak düzensiz hareket içeren çerçeveler giriş olarak verilirse, yeniden oluşturulan optik akış haritalarının giriş haritalarından oldukça farklı çıkması ve yüksek yeniden oluşturma hatası oluşması beklenir.



Şekil 4.1. Önerilen yöntem mimarisi: Ön işleme ve eğitim aşamaları

Önerilen yöntemin eğitimi temelde 2 aşamadan oluşmaktadır. İlk aşama, çerçevelerin çıkartılması, sekizli seriler haline getirilmesi, ölçeklendirilmesi ve optik akış haritalarının çıkarılmasını içeren ön işlem aşamasıdır. İkinci aşama ise ön işleme aşamasında elde edilen

girişleri kabul eden bir derin öğrenme sürecidir. Bu aşamada normal eylem şablonu öğrenilmeye çalışılmaktadır. Ön işleme ve derin öğrenme modelini içeren yöntem mimarisi Şekil 4.1’de verilmiştir.

4.1. Ön işleme Aşaması

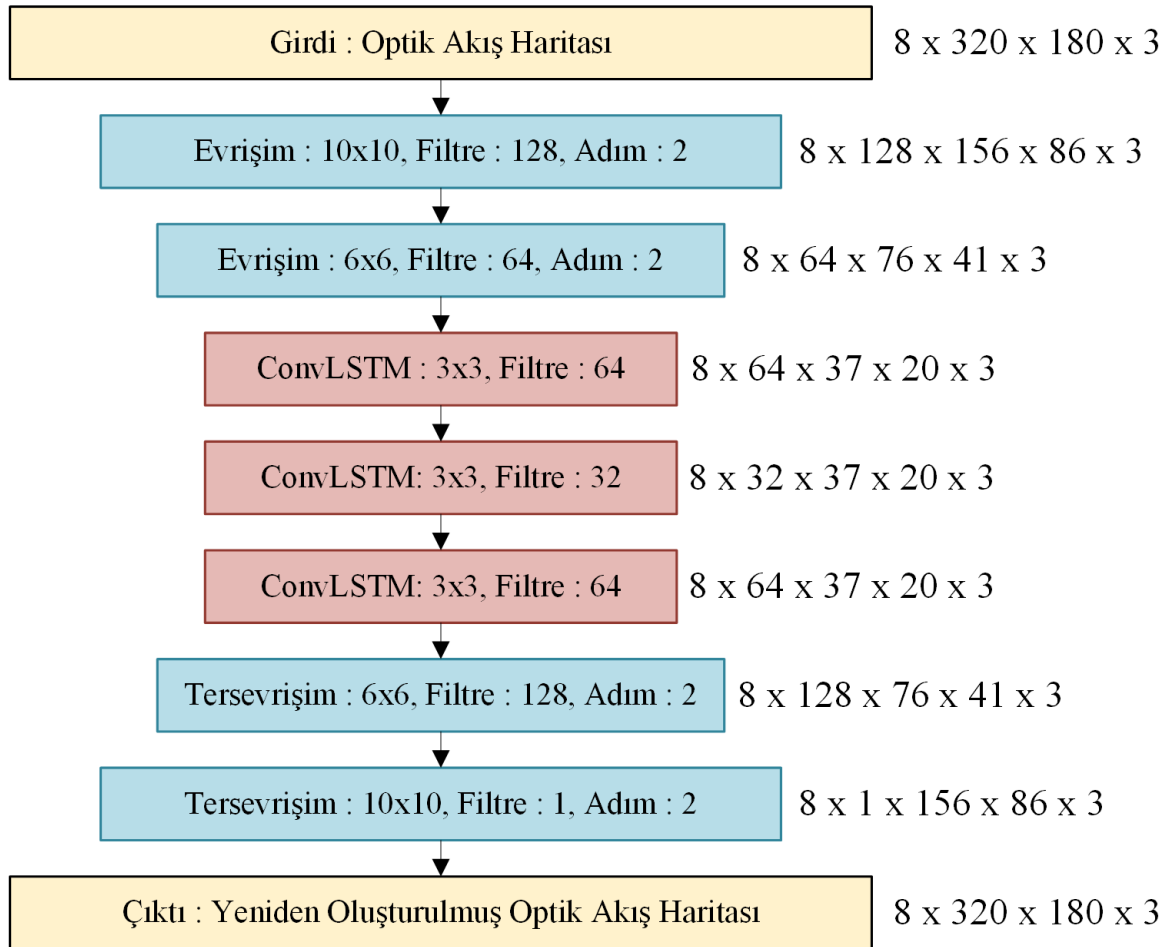
Ön işleme aşamasında video çerçeveleri hem model girişine hem de gerçek zamanlı uygulamalarda kullanıma uygun hale getirilir. Model mimarisindeki ön işleme aşaması Şekil 4.1 üzerinde gösterilmiştir. Ön işleme aşaması yeniden ölçeklendirme ve ardışık sekizli çerçeveler için optik akış haritalarının elde edilmesi olarak iki bölümden oluşur. Günümüzde kabul gören ve yaygın olan 16 birim genişlik ve 9 birim yükseklik en-boy oranına sahip 16:9 video formatıdır [120]. Ön işleme aşamasının ilk bölümünde, veri kümesi videolarından elde edilen çerçeveler 16:9 en-boy oranını koruyacak şekilde 320 x 180 boyutlarında yeniden ölçeklendirilmiştir. Bu yolla hem çerçevelerin boyutu gerçek zamanlı uygulamalara daha uygun olması için küçültülmüş hem de en-boy oranı korunmuştur.

Önerilen derin öğrenme modeli, ardışık sekiz video çerçevesinin optik akış yer değiştirme haritalarını giriş olarak kabul etmektedir. Bu nedenle ön işleme aşamasının ikinci bölümünde, Farneback optik akış yöntemi [148], yoğun optik akış haritalarını elde etmek için uygulanmıştır. Bu sayede sahnedeki hareketli nesnelerin yön ve hız bilgileri elde edilmektedir. Farneback yöntemi, gözlemlenen pikselin komşularının kuadratik polinom bazında modellenmesine dayanan bir yoğun optik akış algoritması sağlamıştır. Burada kullanılan yoğun terimi optik akışın video çerçevelerindeki her bir piksel için hesaplanacağı anlamını taşır. Her piksel için optik akışın hesaplanmasından doğan maliyet, doğrusal olduğu için Farneback algoritması ile kolayca çözülebilmektedir.

4.2. Derin Öğrenme Modeli

Önerilen yöntem mimarisinin eğitim aşaması kapalı olarak Şekil 4.1’de verilmiştir. Evrişimsel derin öğrenme modelinin içyapısı ise Şekil 4.2’de verilmiştir. Şekil 4.1’de sağ tarafta verilen bilgiler ilgili katmanın çıktı boyutlarıdır. Yöntem uzaysal kodlayıcısı, uzaysal çözücüsü ve zamansal hareket korelasyonu için evrişimsel LSTM katmanlarını içermektedir. Bu yolla, eğitim aşamasında normal davranışların hareket şablonunu çıkaran gözetimsiz üretici ağ modelinin oluşturulması amaçlanmıştır. Bilindiği üzere geleneksel otokodlayıcı

modelleri resimlerin uzaysal özelliklerini korumakta yetersizdir [149]. Bu sebeple, Masci ve arkadaşları [142] yeniden oluşturma hatasını minimize etmek için en uygun filtreyi öğrenen evrişimsel otokodlayıcı önermişlerdir. Evrişimsel sinir ağları genellikle gözetimli eğitim sunar ve sınıflandırma problemleri gibi ayrıştırıcı problemleri çözer. Bunun aksine, evrişimsel otokodlayıcılar gözetimsiz öğrenme ile eğitim yapma imkanı sağlarken video çerçevelerinin uzaysal ilişkilerini de korur.



Şekil 4.2. Önerilen yöntemin detaylı evrişimsel ağ mimarisi.

4.3. Düzenlilik Skoru

Test aşamasında performans değerlendirmesi için modelin girişi olan optik akış haritaları ile model tarafından girişe karşılık yeniden üretilen optik akış haritaları karşılaştırılır. Karşılaştırmada çok ölçekli yapısal benzerlik (ÇÖYP, multi-scale structural similarity) metriği kullanılır. Modern yöntemler genellikle giriş görüntüsü ile yeniden oluşturulan çıkış görüntülerinin kare yoğunluk farklarını kullanan piksel temelli ortalama kare hatası (mean

squared error) metriğini kullanır. Ancak ÇÖYP yapısal benzerlik, kontrast benzerliği ve parlaklık benzerliği olmak üzere üç tamamlayıcı bileşeni kullanarak benzerlik ölçümü yapar. Wang ve arkadaşları tarafından geliştirilmiş ve matematiksel notasyonu çalışmalarında [150] ayrıntılı olarak verilmiştir. Tez çalışmasında ise düzensizlik skoru, hesaplamada bir optik akış hacmi x ve x 'e karşılık model tarafından çıkış olarak üretilen $f_w(x)$ kullanılarak şu şekilde hesaplanır:

$$a(x) = 1 - (\text{ÇÖYP}(x, f_w(x))) \quad (4.1)$$

Denklem 4.1'de f_w , model tarafından elde edilen ağırlıkları temsil eder. ÇÖYP metriği ile iki resmin benzerliği hesaplanmıştır. İki eş optik akış haritasının ÇÖYP skorunun bir olması beklenir ve elde edilen değerin birden çıkarılmasıyla anomali skoru sıfır bulunur. Birbirinden farklılaşan iki optik akış haritası için ise, ÇÖYP skoru sıfıra yaklaşırken anomali skoru bire yaklaşır. Son aşamada düzenlilik skorun, sıfır ile bir arasında olacak şekilde normalizasyonu yapılmıştır. Düzenlilik skoru hesaplaması ve normalizasyon işlemi denklem 4.2'de verilmiştir.

$$r(x) = 1 - \frac{a(x) - \min_t a(t)}{\max_t a(t) - \min_t a(t)} \quad (4.2)$$

4.4. Deneysel Sonuçlar

Bu bölüm tez çalışmasının etkinlik karşılaştırmasında kullanılan veri kümeleri, uygulama detayları, değerlendirmede kullanılan etkinlik metrikleri hakkında detaylı bilgiler sunmaktadır.

4.4.1. Veri kümeleri

Kapsamlı bir deneysel değerlendirme yapabilmek için iyi bilinen üç veri kümesi Avenue [151], UCSD normal olmayan eylem tespiti veri kümeleri Ped1 ve Ped2 [152] üzerinde çalışılmıştır. Veri kümeleri hem eğitim hem de test videoları içermektedir. Eğitim süreci yalnızca düzenli olaylar içeren eğitim videolarıyla gerçekleştirilmiştir. Her bir veri kümesi için model normal ve normal olmayan eylemleri içeren test videoları ile test edilmiştir.

Avenue veri kümesi tren istasyonu girişine yerleştirilen sabit bir kamera ile elde edilen 16 eğitim ve 21 test videosu içermektedir. Toplam çerçeve sayısı 30652 ve videoların uzunluğu bir dakika veya daha az sürededir. Eğitim videolarının ortalama uzunluğu 37.75 saniye, test videolarının ortalama uzunluğu ise 28.6 saniyedir. Eğitim videoları tren istasyonu önündeki yaya hareketlerinden oluşmaktadır. Test videoları ise yaya hareketlerinin yanında ters yöne yürüme, koşuşturma, çanta ve evrak fırlatma gibi çeşitli normal olmayan eylemleri içermektedir. Avenue veri kümesinde kalabalık topluluklar, bazı anlarda kamera sarsıntısı ve eğitim videolarında bulunmayan bazı normal eylemler gibi zorluklar bulundurmaktadır. Avenue veri kümesine ait normal olmayan eylem örnekleri Şekil 4.3’de verilmiştir.

UCSD veri kümesi ise çok yüksek bir yere yerleştirilmiş sabit bir kamerayla yaya hareketlerini içermektedir. Sahne yoğunluğu ise seyrekten, oldukça kalabalığa kadar çeşitlilik göstermektedir. Eğitim videoları sadece düzenli yaya hareketlerini barındırırken, test videoları yaya hareketleri yanı sıra bisiklet gibi yaya olmayan araçların geçtiği normal olmayan durumları da içermektedir. Hem eğitim hem de test videoları eşit uzunlukta olup her biri 200 çerçeveden oluşmaktadır. Ped1 veri kümesi 32 eğitim ve 36 test videosundan, ped2 veri kümesi ise 16 eğitim ve 12 test videosundan oluşmaktadır. UCSD Ped1 ve UCSD Ped2 veri kümelerine ait normal olmayan eylem örnekleri sırasıyla Şekil 4.4 ve Şekil 4.5.’de verilmiştir.



Şekil 4.3. Avenue veri kümesi düzensiz eylem örnekleri



Şekil 4.4. UCSD Ped1 veri kümesi düzensiz eylem örnekleri



Şekil 4.5. UCSD Ped2 veri kümesi düzensiz eylem örnekleri

4.4.2. Uygulama detayları

Önerilen modelin eğitim ve test işlemleri 64-bit Windows 10 Pro işletim sistemi kurulu Intel Core i7-7700K merkezi işletim birimi (CPU) ve 16 gigabayt DDR4 rastgele erişim belleği (RAM) bulunan kişisel bir bilgisayarda gerçekleştirilmiştir. GeForce marka GTX 1070 model 8 gb GDDR5 belleği ve 256 bit 1920 CUDA çekirdeği bulunan bir ekran kartı kullanılmıştır. Modelin uygulaması python genel amaçlı programlama dilinde CUDA kullanarak yapılmıştır. Python sürümü 3.6 ve CUDA sürümü 9.0 kullanılmıştır.

Model uygulamasında aktivasyon fonksiyonu olarak hiperbolik tanjant ve iyileştirici olarak Adam iyileştiricisi kullanılmıştır. Öğrenme hızı model ağırlıklarının güncelleme geçmişine göre otomatik olarak belirlenmektedir. Parti boyutu (batch size) 64 ve her bir eğitim 150 devir olarak seçilmiştir. Ancak yeniden oluşturma kaybı (reconstruction loss) son 20 devirde azaltılamazsa eğitime ara verilecek şekilde ayarlandı. Optik akış için ise pyr_scale, window size, iterations, poly_n ve poly_sigma parametreleri sırasıyla 0.5, 15, 3, 5 ve 1.1 olarak belirlenmiştir.

4.4.3. Performans metrikleri

Videolarda normal olmayan eylem tespit yöntemlerinin performansının değerlendirilebilmesinde alıcı işletim karakteristiği (AİK, Receiver Operating Characteristic) eğrisi ve eğri altında kalan alan (EAKA, area under the curve) metrikleri kullanılmıştır. Eğri altında kalan alan metriği ile önerilen yöntemin nicel performans değerlendirmesi mümkün olmuştur. Karşılaştırmalar için performans ölçütünün tanımları aşağıdaki gibidir:

- Doğru Pozitif (DP): Doğru tahmin edilen normal olmayan eylem sayısı
- Doğru Negatif (DN): Doğru tahmin edilen normal eylem sayısı
- Yanlış Pozitif – (YP): Yanlış tahmin edilen normal eylem sayısı
- Yanlış Negatif – (YN): Yanlış tahmin edilen normal olmayan eylem sayısı

Hassasiyet (Precision), pozitif olarak etiketlenen tüm örneklerin saflığının bir ölçüsüdür ve $DP/(DP+YP)$ formülüyle hesaplanır. Geri çağırma (recall) ise tüm pozitif olarak etiketlenmiş

örneklerin doğruluğudur ve $DP/(DP+YN)$ formülüyle hesaplanır. AİK, sahte pozitif oranını (SPO) ile gerçek pozitif oranını (GPO) arasındaki değişim ilişkisini yansıtan bir grafikdir. SPO, $YP/(YP+DN)$ formülüyle GPO ise $DP/(DP+YN)$ formülüyle hesaplanır. Burada AİK eğrisi, sınıflayıcı tarafından belirlenen eşik değerlerine karşılık modelin değerlendirmesini mümkün kılar. EAKA, AİK eğrisi altında kalan alan anlamına gelen ve AİK eğrisi değerlendirilirken kullanan yaygın bir ölçüttür. EAKA, AİK eğrisini tek bir sayıyla ifade ettiğinden karşılaştırma yapmayı oldukça kolaylaştırmıştır. EAKA sıfır ile bir arasında değerler almakta ve değer bire yaklaştıkça değerlendirilen yöntemin başarısı artmakta sıfıra yaklaştıkça azalmaktadır. İstenilen ideal seviye bir değeri olup herhangi yanlış bir değerlendirme yapılmamış anlamına gelir. Değerlendirme metrikleri hakkında daha detaylı bilgi edinebilmek için Power'ın değerlendirme ve değerlendirme metrikleri üzerine çalışmasını[153] incelemek faydalı olabilir.

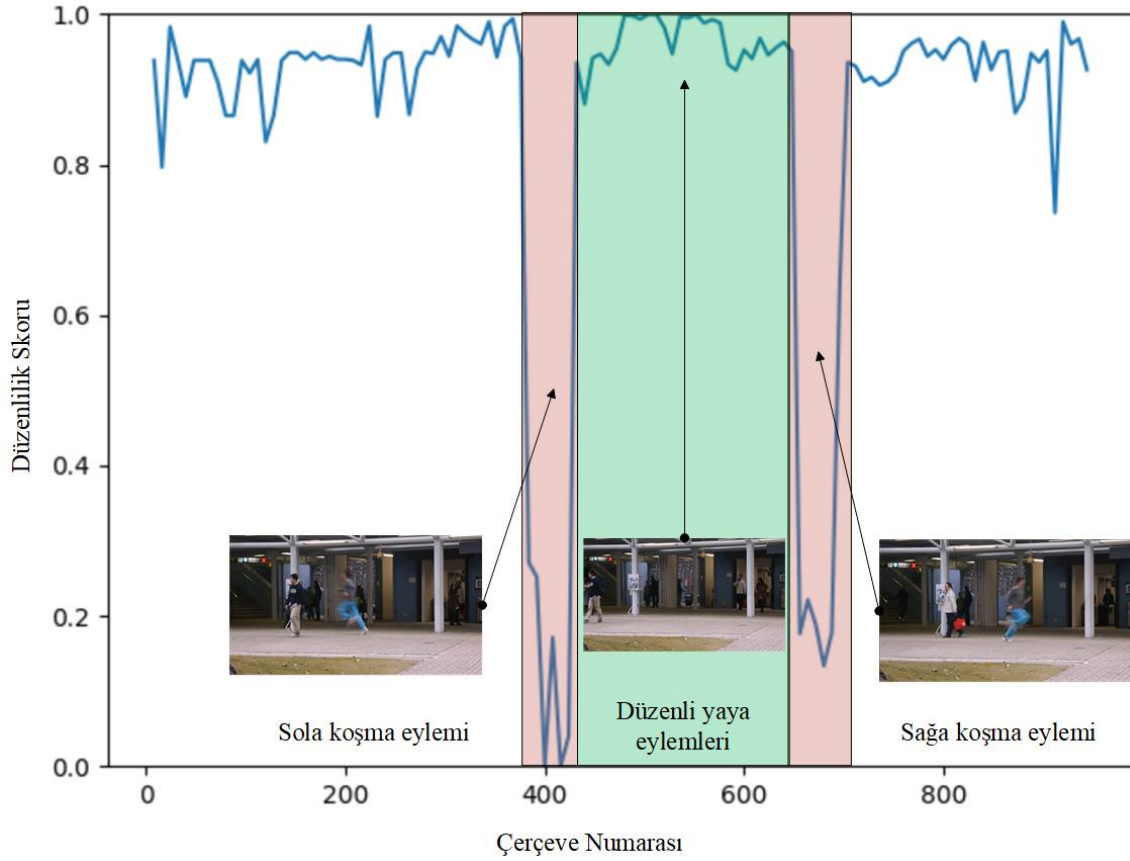
4.4.4. Deneysel sonuçlar

Önişleme safhasında video çerçeveleri çıkarılmış ve 320 x 180 oranında ölçeklendirilmiştir. Her bir çerçevenin optik akışları çıkarılırken giriş çerçeveleri sekizli gruplar haline model girişi olarak hazırlanmaktadır. Model girişi olan sekizli çerçeve ve karşılık gelen optik akış haritasına örnekler Şekil 4.6'de verilmiştir. Şekil 4.6 a1 ve Şekil 4.6 a2 örnek ardışık çerçeveyi Şekil 4.6 b1 ve Şekil 4.6 b2 ise çerçeveler karşılık gelen optik akış haritalarını göstermektedir.

Model eğitiminden sonra model girişi ve model çıkışının karşılaştırılması sonucunda elde edilen düzenlilik skoru, video çerçevelerinde normal olmayan bir eylemin gerçekleşip gerçekleşmediği hakkında bilgi verir. Normal olmayan bir eylem içeren video çerçevelerinin düşük düzenlilik skoru, normal eylemler içeren video çerçevelerinin ise yüksek düzenlilik skoru üretmesi beklenmektedir. Şekil 4.7'de Avenue test veri kümesinde bulunan 4. Video için düzenlilik skorları verilmiştir. Şekilde x ve y koordinatları sırasıyla çerçeve numarası ve düzenlilik puanını göstermektedir. Düzenlilik puanının sıfıra yaklaşması ilgili çerçevelerde düzensiz bir eylemin gerçekleştiğine, bire yaklaşması ise düzenli eylemlere işaret etmektedir.



Şekil 4.6. Model girişinin örnek gösterimi: Sekiz ardışık çerçeve ve karşılık gelen optik akış haritaları.



Şekil 4.7. Video 4 için düzenlilik skor örneği

Tez çalışmasında üç veri kümesi için çerçeve ve olay seviyesinde performans değerlendirmesi sağlanmıştır. Önerilen yöntemin etkisini ölçebilmek için otokodlayıcı temelli altı modern yaklaşımla karşılaştırılmıştır. Bu yöntemler; ConvAE [110], ST-AE [154], ConvLSTM-AE [155], Two-Stream R-ConvVAE[156], WCAE-LSTM[157] ve STAN[158] kısaltmalarıyla kendi makalelerinde isimlendirilmiştir. ConvAE [110] hem tam bağlı otokodlayıcı hem de gezinga temelli el yapımı uzaysal ve zamansal özelliklerden faydalanmıştır. ConvAE yönteminin temel kısıtlılığı gezinga temelli özellik çıkarım aşamasının başarımından büyük oranda etkilenmesidir. Özellikle kalabalık sahnelerde yakalama ve nesnelerin takibi zor olduğundan gezinga temelli yöntemlerin başarımı düşmektedir. ST-AE [154] evrişimsel otokodlayıcı ve evrişimsel uzun kısa süreli bellek temelli uzaysal-zamansal bir mimari önermektedir. Bu şekilde, yöntem hızlı eğitim ve test süresi ile videoların uzaysal ve zamansal özelliklerini yakalar. Bunun yanında özellik çıkarımı yöntemlerini kullanmadığından performansında düşüklük görülmektedir. ConvLSTM-AE [155] düzenli eylemleri modellemek için evrişimsel uzun kısa süreli bellek ve otokodlayıcıyı birlikte kullanan bir yapı sunmaktadır. Two-Stream R-ConvVAE [156], iki akışlı özyineli değişimsel kodlayıcıyı yarı gözetimli bir öğrenme yaklaşımıyla

modellemiştir. WCAE-LSTM[157] ise video çerçevelerinin uzaysal-zamansal bilgileri kodlamak için ağırlıklı evrişimsel kodlayıcı ve kısa süreli bellek yöntemi kullanılmıştır. Yöntem, ön plan nesnelerinin çıkarılması için ağırlıklı Öklid kaybı önermiştir. Bu yolla, yöntem arka plandan olumsuz etkilenmeden sadece hareketli nesnelere yoğunlaşmayı amaçlamıştır. Ancak sahneden elde edilen nesneler sadece nesnenin varlığı ve şekli hakkında bilgi verir. Oysa optik akış nesnelerin varlığı ve şekli dışında nesnelerin hızı ve yönleri hakkında da bilgi verir. STAN[158] ise ardışık video çerçevelerinin gerçek olup olmadığının tespitinden yola çıkan üretici çekişmeli ağ modeli önermektedir. Önerilen sinir ağı modeli düzenli insan davranış şablonunu uzaysal ve zamansal olarak uzaysal-zamansal üretici ve ayrıştırıcı ağdan oluşmaktadır. Yöntemin güçlü yanı yoğunluktan dolayı kapanmaların olduğu karmaşık sahnelerdeki başarısıdır. Bu yönüyle tez çalışması diğer çalışmalardan ayrışsa bile, optik akışın ve otokodlayıcının üstün yanlarını kullandığından dolayı önerilen yöntem tüm senaryolar genelinde daha başarılıdır.

Çizelge 4.1. Önerilen yöntemin çerçeve seviyesinde performans karşılaştırması

METHOD	AVENUE	PED1	PED2
ConvAE	%75.2	%81.0	%90.0
ST-AE	%80.3	%89.9	%87.4
ConvLSTM-AE	%77.0	%75.5	%88.1
Two-Stream R-ConvVAE	%79.6	%75.0	%91.7
WCAE-LSTM	%85.7	%85.1	%92.6
STAN	%87.2	%82.1	%96.5
Önerilen Model	%89.5	%92.4	%92.9

Önerilen modelin AUC metriğine göre çerçeve seviyesindeki karşılaştırması kullanılan üç veri kümesi için ayrı ayrı Çizelge 4.1’de verilmiştir. Deneysel sonuçlara göre, geliştirilen yöntem Avenue ve Ped1 veri kümelerinde karşılaştırılan derin öğrenme kullanan yöntemlere göre videolarda normal olmayan eylem tespitinde daha iyi sonuç vermişlerdir. Ped2 veri kümesinde ise STAN[158], geliştirilen yöntem de dahil karşılaştırılan tüm metotlardan daha iyi sonuç vermektedir. Önerilen yöntemin Aveue, Ped1 ve Ped2 için elde ettiği AUC değerleri sırasıyla %89.5, %92.4 ve %92.9’dur. Geliştirilen yöntemin Ped2 veri kümesinde ikinci sırada kalmasının sebebi ise geliştirilen yöntemin çok uzak mesafeden alınan görüntülerde, yakın ve orta uzaklıktaki görüntülere nazaran daha az başarılı olması gösterilebilir. Bunun dışında bir başka sebep de Ped2 veri kümesinde 16 eğitim ve 12 test videosu gibi çok kısıtlı eğitim ve test verisinin olması gösterilebilir.

Çizelge 4.2. Önerilen modelin çerçeve seviyesindeki performans sonuçları

VERİSETİ	TOPLAM ÇERÇEVE	DÜZENLİ ÇERÇEVE	DÜZENSİZ ÇERÇEVE	DP	YN	DN	YP
Avenue	15324	11504	3820	3738	82	11394	110
UCSD Ped1	7200	3195	4005	3919	86	3135	46
UCSD Ped2	2010	374	1636	1564	72	355	19

Toplam çerçeve sayısı, normal olmayan eylem içeren toplam çerçeve sayısı ve düzenli eylemleri içeren toplam çerçeve sayısı gibi veri kümesi hakkında detaylı bilgilerle, deneysel sonuçlara göre doğru pozitif, yanlış negatif, doğru negatif ve yanlış pozitif istatistikleri Çizelge 4.2’de verilmiştir. Avenue veri kümesindeki toplam çerçeve sayısı 15324 iken, normal olmayan eylem içeren çerçeve sayısı 3820 olup, toplam çerçeve sayısının yaklaşık %2,5’luk bölümünü oluşturmaktadır. Avenue veri seti için 3820 normal olmayan çerçeveden 3738 adeti doğru etiketlenirken, 82 adet normal olmayan eylem içeren çerçeve tespit edilememiştir. Yine aynı veri kümesinde 11504 düzenli eylemler içeren çerçeveden 11394 tanesi düzenli eylem etiketiyle işaretlenirken 110 tanesi düzenli olmasına rağmen normal olmayan eylem olarak işaretlenmiştir. UCSD Ped1 veri kümesinde toplam 7200 çerçeveden 3195 çerçeve düzenli eylemleri barındırırken 4005 çerçeve normal olmayan eylem bulundurmaktadır. Normal olmayan eylem içeren 4005 çerçeveden 3919 adeti doğru tespit edilirken 86 çerçeve normal olmayan eylem olarak yakalanamamıştır. 3195 düzenli eylem içeren çerçeveden ise 3135’i doğru etiketlenerek büyük oranda başarı elde edilmiştir. Düzenli eylemlerin yanlış etiketlendiği çerçeve sayısı yalnızca 46 olmuştur. UCSD Ped2 veri kümesinde toplam 2010 çerçeveden 374 çerçeve düzenli eylemleri barındırırken 1636 çerçeve normal olmayan eylem bulundurmaktadır. Bu yönüyle test verisi için büyük oranda normal olmayan eylem barındırdığı söylenebilmektedir. Normal olmayan eylem içeren 1636 çerçeveden 1564 adeti doğru tespit edilirken, 72 çerçeve normal olmayan eylem olarak yakalanamamıştır. 374 düzenli eylem içeren çerçeveden ise 355’i doğru 19 tanesi yanlış işaretlenmiştir.

Çizelge 4.3. Önerilen modelin eylem seviyesinde performans karşılaştırması

VERİSETİ	ANOMALİ	DOĞRU ALGILAMA/HATALI ALARM				
		ST-AE [154]	ConvAE [110]	TS-R-ConvVAE[156]	STAN[158]	ÖNERİLEN YÖNTEM
Avenue	47	43/8	45/4	34/6	N/A	45/2
UCSD Ped1	40	36/11	38/6	38/6	37/3	38/2
UCSD Ped2	12	12/3	12/1	12/0	12/0	12/2

Çerçeve seviyesinde performans analizinin yanında eylem seviyesinde performans analizi de yapılmıştır. Her bir veri kümesi için ayrıntılı karşılaştırmalı sonuçlar Çizelge 4.3’de

verilmiştir. Avenue, Ped1 ve Ped2 verisetlerinde sırasıyla 47, 40 ve 12 normal olmayan eylem bulunmaktadır. Normal olmayan eylemlerin belirlenmesinde veri setleri yayıncılarının paylaştığı gerçek referans değerler kullanılmıştır. Bu değerler ışığında, Avenue veri kümesi için ConvAE [110] ve önerilen yöntem 47 normal olmayan eylemden 45 tanesini doğru tahmin etmiştir. Ancak önerilen yöntem iki yanlış alarm verirken ConvAE [110] yöntemi 4 adet yanlış alarm vermiştir. Ped1 veri kümesi için 40 normal olmayan eylem için önerilen yöntem 38 algılama yanında 2 yanlış alarm verirken, en yakın sonuç 38 algılama ve 6 yanlış alarm ile Two-Stream R-ConvVAE[156] ve ConvAE [110] yöntemleri tarafından elde edilmiştir. Bunun yanında STAN[158] yöntemi, Two-Stream R-ConvVAE ve ConvAE [110] yönteminden bir eksik algılama yapmasına rağmen 3 yanlış alarm vermiştir. Bu yönüyle Two-Stream R-ConvVAE[156] ve ConvAE [110] yöntemlerinden yanlış alarm yönünden bir üstünlük sağladığı söylenebilir. Ped2 veri için toplam 12 normal olmayan eylemi karşılaştırma yapılan bütün yöntemler ve önerilen yöntem algılayabilirken hiç yanlış alarm vermeyen Two-Stream R-ConvVAE[156] ve STAN[158] yöntemleri en başarılı sonuçları almıştır. Önerilen yöntem ise düzenli eylemlerin olduğu iki sahneyi normal olmayan eylem olarak tanımlayarak iki yanlış alarm üretmiştir.

Çizelge 4.4. Önışleme ve test aşamaları çalışma zaman analizi

İŞLEMÇİ BİRİMİ	ÖNİŞLEME AŞAMASI	TEST AŞAMASI	TOPLAM ZAMAN
CPU	0.016 saniye	0.235 saniye	0,251 saniye. (~4fps)
GPU	0.016 saniye	0.008 saniye	0,024 saniye (~42.5 fps)

Çerçeve seviyesinde ve eylem seviyesinde performans karşılaştırmasının yanında, yöntemlerin gerçek zamanlı uygulamalarda da kullanılabilmesi için çalışma zamanı analizi de bir diğer önemli unsurdur. Önerilen yöntemin performans karşılaştırması Çizelge 4.4’de verilmiştir. Merkezi işlem birimi kullanarak yapılan testlerde önışleme 0.016 saniye ve test aşaması yaklaşık 0.235 saniye almaktadır. Bu sonuç da saniyede yaklaşık 4 çerçeve işleminin anlamına gelir ki gerçek zamanlı uygulamalar için kullanışsızdır. Bilindiği üzere günümüz videoları saniyede ortalama 24 çerçeve sunulmaktadır. Grafik işlemci desteğiyle önerilen yöntem test aşamasını çok hızlı bir hale getirerek yaklaşık saniyede 42-43 çerçeve işleyebilmektedir. Bu yönüyle merkezi işlemci biriminden yaklaşık 10 kat hızlı ve gerçek zamanlı uygulamalar için kullanılabilir.



5. SONUÇ VE ÖNERİLER

Video görüntülerinde normal olmayan davranışların tespit edilmesi konusu gerçek zamanlı uygulamalar için çalışma zamanı kısıtlılığı, kalabalık sahneler sebebiyle kapanma gibi çevresel etkiler ve normal olmayan eylem tanımının karmaşıklığı gibi zorluklar içermektedir. Tez çalışması kapsamında, bu problemlerle başa çıkmada deneysel sonuçlar ışığında başarılı sonuçlar ortaya koyan video görüntülerinde normal olmayan eylem tespiti yapabilen gözetimsiz, üretici ve model temelli bir çerçeve önerilmiştir. Yöntem optik akış haritalarını model çıkışında üreterek, model girişindeki optik akış haritalarıyla karşılaştırmak ve yeniden oluşturma hatası sunmaktadır. Evrişimsel otokodlayıcı ile evrişimsel UKSB mimarisini birleştirerek hem uzaysal hem de zamansal özelliklerin çıkarılması mümkün kılınmıştır. Modelin etkinliğinin deneysel sonuçlarla ölçülebilmesi için model giriş ve çıkışlarının karşılaştırılmasında yeni bir düzenlilik skoru sunulmuştur. Modelin düzenli olmayan eylemleri yakalamasındaki başarısının yanı sıra gerçek zamanlı uygulamalar konusundaki etkinliği test edilmiş ve aşağıdaki sonuçlara ulaşılmıştır:

- Önerilen modelin normal olmayan eylemleri tespit etmedeki başarısı Aveue, Ped1 ve Ped2 veri kümeleri üzerine test edilmiş ve sırasıyla %89.5, %92.4 ve %92.9 EAA değerleri elde edilmiştir. Etkinliğinin diğer modern yaklaşımlarla karşılaştırılabilmesi için EAA temelli sonuçlar birlikte Çizelge 4.1’de verilmiştir.
- Önerilen modelin gerçek zamanlı uygulamalar için uygunluğunun ölçülebilmesi için çalışma zamanı analizi yapılmış ve Çizelge 4.4’de verilen bilgiler ışığında, test aşamasında grafik işlemci desteğiyle saniyede yaklaşık 43 çerçeve işlenebileceği ortaya konulmuştur. Saniyede bu miktarda çerçeve işleyebilmek günümüz video kayıtlarının gerçek zamanlı işlenmesine olanak tanımaktadır. Ayrıca deneysel çalışmada grafik işlemci desteği yardımıyla yalnızca merkezi işlem birimiyle çalışmaktan 10 kat daha hızlı çerçeve işlendiği ortaya çıkarılmıştır.

Ayrıca normal olmayan eylemlerin tespitinde karşılaşılan zorluklarda aşağıdaki başarımlar elde edilmiştir:

- Normal olmayan eylemlerin tespitinde EAA metriği temelli yaklaşık %90 ortalama doğruluk oranı
- Gözetimsiz öğrenme yöntemi kullanılarak etiketsiz video verisinde çalışabilme

- Gerçek zamanlı uygulamalar için saniyede yaklaşık 43 çerçeve işleyebilme
- Optik akış yöntemiyle bireysel eylemlerin dışında toplu normal olmayan eylemlerin tespitinin yapılabilmesi
- Model normal eylem şablonu çıkardığından daha önce karşılaşılmamış normal olmayan eylemlerin ayırt edilebilmesi
- Yerel yerine sahneye ait özellikler kullanıldığından yoğun sahnelerde de başarı sağlanabilmesi

Bu tez çalışmasında normal olmayan eylemlerin algılanması için geliştirilen yöntemin iyileştirilmesi için yapılabilecek yeni çalışmalara dair değerlendirmeler aşağıda sunulmuştur.

- Yüksek hızda çalışan ve üstün başarı elde eden derin öğrenme temelli optik akış modellerinin model içerisinde kullanılması ile optik akış haritalarının daha yüksek doğrulukta çıkarılması
- Ön plan nesnelerin birbirlerine olan etkilerinin, yöntemin gerçek zamanlı uygulamalara uygunluğunu bozmadan incelenmesi böylece yön ve hız bilgisinin yanı sıra nesnelerin etkileşimi hakkında da modele bilgi sunulması.

KAYNAKLAR

1. Wang, Q., Kealy, A. and Zhai, S. (2019). Introduction for the Special Issue on Beyond the Hypes of Geospatial Big Data: Theories, Methods, Analytics, and Applications. *Mechanical Engineering Faculty Publications*. 112(2), 245-245
2. Vrigkas, M., Nikou, C. and Kakadiaris, I.A. (2015). A review of human activity recognition methods. *Frontiers in Robotics and Artificial Intelligence*, 2, 28.
3. Hu, T., Zhu, X., Guo, W., Wang, S. and Zhu, J. (2019). Human action recognition based on scene semantics. *Multimedia Tools and Applications*, 78(20), 28515-28536.
4. Raval, R., Prajapati, H. and Dabhi, V. (2019). Survey and analysis of human activity recognition in surveillance videos. *Intelligent Decision Technologies*, 13(2), 271-294.
5. Aggarwal, J.K. and Ryoo, M.S. (2011). Human activity analysis: A review. *ACM Computing Surveys (CSUR)*, 43(3), 1-43.
6. Nigam, S., Singh, R. and Misra, A. (2019). A Review of Computational Approaches for Human Behavior Detection. *Archives of Computational Methods in Engineering*, 26(4), 831-863.
7. Bobick, A.F. and Davis, J.W. (2001). The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis & Machine Intelligence* (3), 257-267.
8. Yilmaz, A. and Shah, M. (2005). *Actions sketch: A novel action representation*. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05).
9. Sheikh, Y., Sheikh, M. and Shah, M. (2005). *Exploring the space of a human action*. Tenth IEEE International Conference on Computer Vision (ICCV'05), 1.
10. Leightley, D., Li, B., McPhee, J.S., Yap, M.H. and Darby, J. (2014). *Exemplar-based human action recognition with template matching from a stream of motion capture*. International Conference Image Analysis and Recognition, 12-20 .
11. Kitani, K.M., Sato, Y. and Sugimoto, A. (2008). Recovering the basic structure of human activities from noisy video-based symbol strings. *International Journal of Pattern Recognition and Artificial Intelligence*, 22(08), 1621-1646.
12. Hongeng, S., Nevatia, R. and Bremond, F. (2004). Video-based event recognition: activity representation and probabilistic recognition methods. *Computer Vision and Image Understanding*, 96(2), 129-162.
13. Cho, S.H. and Kang, H.B. (2014). Abnormal behavior detection using hybrid agents in crowded scenes. *Pattern Recognition Letters*, 44, 64-70.

14. Mu, C.D., Xie, J.B., Yan, W., Liu, T. and Li, P.Q. (2015). A fast recognition algorithm for suspicious behavior in high definition videos. *Multimedia Systems*, 22(3), 275-285.
15. Cong, Y., Yuan, J.S. and Liu, J. (2013). Abnormal event detection in crowded scenes using sparse representation. *Pattern Recognition*, 46(7), 1851-1864.
16. Gnanavel, V.K. and Srinivasan, A. (2015). *Abnormal Event Detection in Crowded Video Scenes*. Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014, 2, 328, 441-448.
17. Gu, X.X., Cui, J.R. and Zhu, Q. (2014). Abnormal crowd behavior detection by using the particle entropy. *Optik*, 125(14), 3428-3433.
18. Nallaivarothayan, H., Fookes, C., Denman, S. and Sridharan, S. (2014). *An MRF based Abnormal Event Detection Approach using Motion and Appearance Features*. 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 343-348.
19. Huang, J.F. and Chen, S.L. (2014). *Detection of Violent Crowd Behavior Based on Statistical Characteristics of the Optical Flow*. 2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 565-569.
20. Huo, J., Gao, Y., Yang, W.Q. and Yin, H.J. (2014). Multi-Instance Dictionary Learning for Detecting Abnormal Events in Surveillance Videos. *International Journal of Neural Systems*, 24(3).
21. Leyva, R., Sanchez, V. and Li, C.T. (2014). *Video Anomaly Detection based on Wake Motion Descriptors and Perspective Grids*. 2014 IEEE International Workshop on Information Forensics and Security (WIFS'14), 209-214.
22. Li, N.N., Wu, X.Y., Guo, H.W., Xu, D., Ou, Y.S. and Chen, Y.L. (2015). Anomaly Detection in Video Surveillance via Gaussian Process. *International Journal of Pattern Recognition and Artificial Intelligence*, 29(6).
23. Rasheed, N., Khan, S.A. and Khalid, A. (2014). *Tracking and Abnormal Behavior Detection in Video Surveillance using Optical Flow and Neural Networks*. 2014 28th International Conference on Advanced Information Networking and Applications Workshops (Waina), 61-66.
24. Zhang, T., Jia, W.J., Yang, B.Q., Yang, J., He, X.J. and Zheng, Z.L. (2017). MoWLD: a robust motion image descriptor for violence detection. *Multimedia Tools and Applications*, 76(1), 1419-1438.
25. Wang, L.J. and Dong, M. (2014). *Detection of abnormal human behavior using a matrix approximation-based approach*. 2014 13th International Conference on Machine Learning and Applications (ICMLA), 324-329.
26. Wang, T. and Snoussi, H. (2014). Detection of Abnormal Visual Events via Global Optical Flow Orientation Histogram. *IEEE Transactions on Information Forensics and Security*, 9(6), 988-998.

27. Zhang, Y., Lu, H.C., Zhang, L.H. and Ruan, X. (2016). Combining motion and appearance cues for anomaly detection. *Pattern Recognition*, 51, 443-452.
28. Colque, R.V.H.M., Caetano, C., de Andrade, M.T.L. and Schwartz, W.R. (2017). Histograms of Optical Flow Orientation and Magnitude and Entropy to Detect Anomalous Events in Videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(3), 673-682.
29. Bellamine, I. and Tairi, H. (2015). Motion Detection using Color Structure-Texture Image Decomposition. *2015 Intelligent Systems and Computer Vision (ISCV)*, 1-8.
30. Nieves, E.B., Suarez, O.D., Garcia, G.B. and Sukthankar, R. (2011). *Violence Detection in Video Using Computer Vision Techniques*. Computer Analysis of Images and Patterns: 14th International Conference, 6855, 332-339.
31. Ben Mabrouk, A. and Zagrouba, E. (2017). Spatio-temporal feature using optical flow based distribution for violence detection. *Pattern Recognition Letters*, 92, 62-67.
32. Xu, L., Gong, C., Yang, J., Wu, Q. and Yao, L.X. (2014). *Violent Video Detection Based on Mosift Feature and Sparse Coding*. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 3538-3542.
33. Li, N.N., Wu, X.Y., Xu, D., Guo, H.W. and Feng, W. (2015). Spatio-temporal context analysis within video volumes for anomalous-event detection and localization. *Neurocomputing*, 155, 309-319.
34. Zhu, S.H., Hu, J.J. and Shi, Z. (2016). Local abnormal behavior detection based on optical flow and spatio-temporal gradient. *Multimedia Tools and Applications*, 75(15), 9445-9459.
35. Xu, D., Song, R., Wu, X.Y., Li, N.N., Feng, W. and Qian, H.H. (2014). Video anomaly detection based on a hierarchical activity discovery within spatio-temporal contexts. *Neurocomputing*, 143, 144-152.
36. Yogameena, B. and Priya, K.S. (2015). *Synoptic Video Based Human Crowd Behavior Analysis for Forensic Video Surveillance*. 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR), 69-74.
37. Mousavi, H., Galoogahi, H.K., Perina, A. and Murino, V. (2016). Detecting Abnormal Behavioral Patterns in Crowd Scenarios. *Toward Robotic Socially Believable Behaving Systems, Vol II: Modeling Social Signals*, 106, 185-205.
38. Raptis, M. and Soatto, S. (2010). Tracklet Descriptors for Action Modeling and Video Analysis. *Computer Vision-ECCV 2010, Pt I*, 6311, 577-590.
39. Zhao, Y., Qiao, Y., Yang, J. and Kasabov, N. (2015). Abnormal Activity Detection Using Spatio-Temporal Feature and Laplacian Sparse Representation. *Neural Information Processing, ICONIP 2015, Pt Iv*, 9492, 410-418.

40. Aslan, M., Sengur, A., Xiao, Y., Wang, H.B., Ince, M.C. and Ma, X. (2015). Shape feature encoding via Fisher Vector for efficient fall detection in depth-videos. *Applied Soft Computing*, 37, 1023-1028.
41. Miao, Y.Y. and Song, J.X. (2014). *Abnormal Event Detection Based on SVM in Video Surveillance*. Proceedings of 2014 IEEE Workshop on Advanced Research and Technology in Industry Applications (Wartia), 1379-1383.
42. Nguyen, V.D., Le, M.T., Do, A.D., Duong, H.H., Thai, T.D. and Tran, D.H. (2014). *An Efficient Camera-based Surveillance for Fall Detection of Elderly People*. Proceedings of the 2014 9th IEEE Conference on Industrial Electronics and Applications (ICIEA), 994.
43. Li, W.X., Mahadevan, V. and Vasconcelos, N. (2014). Anomaly Detection and Localization in Crowded Scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1), 18-32.
44. Mahadevan, V., Li, W.X., Bhalodia, V. and Vasconcelos, N. (2010). *Anomaly Detection in Crowded Scenes*. 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1975-1981.
45. Arroyo, R., Yebes, J.J., Bergasa, L.M., Daza, I.G. and Almazan, J. (2015). Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls. *Expert Systems with Applications*, 42(21), 7991-8005.
46. Li, C., Han, Z.J., Ye, Q.X. and Jiao, J.B. (2013). Visual abnormal behavior detection based on trajectory sparse reconstruction analysis. *Neurocomputing*, 119, 94-100.
47. Conci, N. and Lizzi, L. (2009). *Camera Placement Using Particle Swarm Optimization in Visual Surveillance Applications*. 2009 16th IEEE International Conference on Image Processing, 1-6, 3485.
48. Rai, H., Kolekar, M.H., Keshav, N. and Mukherjee, J.K. (2015). Trajectory Based Unusual Human Movement Identification for Video Surveillance System. *Progress in Systems Engineering*, 366, 789-794.
49. Ko, J.G. and Yoo, J.H. (2013). *Rectified Trajectory Analysis based Abnormal Loitering Detection for Video Surveillance*. 2013 First International Conference on Artificial Intelligence, Modelling and Simulation (AIMS 2013), 289-293.
50. Lai, D.X., Chang, Y.H. and Zhong, Z.H. (2009). Active Contour Tracking of Moving Objects Using Edge Flows and Ant Colony Optimization in Video Sequences. *Advances in Image and Video Technology, Proceedings*, 5414, 1104-1116.
51. Leach, M.J.V., Sparks, E.P. and Robertson, N.M. (2014). Contextual anomaly detection in crowded surveillance scenes. *Pattern Recognition Letters*, 44, 71-79.
52. Ngo, D.V., Do, N.T. and Nguyen, L.A.T. (2016). *Anomaly Detection in Video Surveillance: A Novel Approach Based on Sub-Trajectory*. 2016 International Conference on Electronics, Information, and Communications (ICEIC), 1-4.

53. Lee, S.C. and Nevatia, R. (2014). Hierarchical abnormal event detection by real time and semi-real time multi-tasking video surveillance system. *Machine Vision and Applications*, 25(1), 133-143.
54. Zhang, X.Q., Hu, W.M., Maybank, S., Li, X. and Zhu, M.L. (2008). Sequential particle swarm optimization for visual tracking. 2008 IEEE Conference on Computer Vision and Pattern Recognition, 1-12, 1317.
55. Zhang, Y.H., Lin, W.Y., Zhang, G.W., Luo, C.F., Jiang, D. and Yao, C.L. (2014). *A New Approach for Extracting and Summarizing Abnormal Activities in Surveillance Videos*. 2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), 1-5.
56. Li, X.X., Zhu, S.H. and Chen, L.L. (2016). Statistical background model-based target detection. *Pattern Analysis and Applications*, 19(3), 783-791.
57. Subudhi, B.N., Ghosh, S., Shiu, S.C.K. and Ghosh, A. (2016). Statistical feature bag based background subtraction for local change detection. *Information Sciences*, 366, 31-47.
58. Wang, F.S., Lin, B.W. and Li, X.C. (2017). *An Ant Particle Filter for Visual Tracking*. 2017 16th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2017), 417-422.
59. Albawendi, S., Appiah, K., Powell, H., & Lotfi, A. (2016). *Video based fall detection with enhanced motion history images*. Proceedings of the 9th ACM International Conference on Pervasive Technologies Related to Assistive Environments, 29.
60. Saini, R., Ahmed, A., Dogra, D.P. and Roy, P.P. (2017). *Surveillance scene segmentation based on trajectory classification using supervised learning*. Proceedings of International Conference on Computer Vision and Image Processing, 261-271.
61. Lloyd, K., Rosin, P.L., Marshall, D. and Moore, S.C. (2017). Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (GLCM)-based texture measures. *Machine Vision and Applications*, 28(3-4), 361-371.
62. Zhao, Y., Zhou, L., Fu, K.R. and Yang, J. (2016). *Abnormal Event Detection Using Spatio-Temporal Feature and Nonnegative Locality-Constrained Linear Coding*. 2016 IEEE International Conference on Image Processing (ICIP), 3354-3358.
63. Das Dawn, D. and Shaikh, S.H. (2016). A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector. *Visual Computer*, 32(3), 289-306.
64. Elabbessi, S., Abdellaoui, M. and Douik, A. (2016). *An improved selective spatio-temporal interest point detector*. 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 147-150.

65. Chakraborty, B., Holte, M.B., Moeslund, T.B. and Gonzalez, J. (2012). Selective spatio-temporal interest points. *Computer Vision and Image Understanding*, 116(3), 396-410.
66. Cahyadi, R. and Fadlil, J. (2015). *Video Anomaly Detection using Selective Spatio-Temporal Interest Points and Convolutional Sparse Coding*. 2015 IEEE /WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (Wi-Iat), Vol 3, 203-206.
67. Wang, T. and Snoussi, H. (2015). Detection of Abnormal Events via Optical Flow Feature Analysis. *Sensors*, 15(4), 7156-7171.
68. Bregonzio, M., Gong, S. and Xiang, T. (2009). *Recognising action as clouds of space-time interest points*. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 1948-1955.
69. Raghavendra, R., Del Bue, A., Cristani, M. and Murino, V. (2011). *Optimizing Interaction Force for Global Anomaly Detection in Crowded Scenes*. 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 136-143.
70. Rezaee, K., Haddadnia, J. and Delbari, A. (2015). Modeling abnormal walking of the elderly to predict risk of the falls using Kalman filter and motion estimation approach. *Computers & Electrical Engineering*, 46, 471-486.
71. Dotti, D., Popa, M. and Asteriadis, S. (2017). *Unsupervised Discovery of Normal and Abnormal Activity Patterns in Indoor and Outdoor Environments*. VISIGRAPP (5: VISAPP), 210-217.
72. Rai, H., Kolekar, M.H., Keshav, N. and Mukherjee, J. (2015). Trajectory Based Unusual Human Movement Identification for Video Surveillance System. *Progress in Systems Engineering*. Springer, pp. 789-794.
73. Gomez, A.H.F., Tomas, R.M., Tapia, S.A., Caballero, A.F., Ratte, S., Eras, A.G. and Gonzalez, P.L. (2015). Identification of Loitering Human Behaviour in Video Surveillance Environments. *Artificial Computation in Biology and Medicine, Pt I (IWINAC 2015)*, 9107, 516-525.
74. Stone, E.E. and Skubic, M. (2015). Fall Detection in Homes of Older Adults Using the Microsoft Kinect. *IEEE Journal of Biomedical and Health Informatics*, 19(1), 290-301.
75. Foggia, P., Percannella, G., Saggese, A. and Vento, M. (2013). *Recognizing Human Actions by a bag of visual words*. 2013 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2013), 2910-2915.
76. Hernandez-Garcia, R., Ramos-Cozar, J., Guil, N., Garcia-Reyes, E. and Sahli, H. (2018). Improving Bag-of-Visual-Words model using visual n-grams for human action classification. *Expert Systems with Applications*, 92, 182-191.

77. Zhang, T., Jia, W., Gong, C., Sun, J., & Song, X. (2017). Semi-supervised dictionary learning via local sparse constraints for violence detection. *Pattern recognition letters*, 107: 98-104.
78. Lu, C.W., Shi, J.P. and Jia, J.Y. (2013). *Abnormal Event Detection at 150 FPS in MATLAB*. 2013 IEEE International Conference on Computer Vision (ICCV), 2720-2727.
79. Tani, M.Y.K., Lablack, A., Ghomari, A. and Bilasco, I.M. (2015). Events Detection Using a Video-Surveillance Ontology and a Rule-Based Approach. *Computer Vision - ECCV 2014 Workshops, Pt Ii*, 8926, 299-308.
80. Castro, J.L., Delgado, M., Medina, J. and Ruiz-Lozano, M.D. (2011). Intelligent surveillance system with integration of heterogeneous information for intrusion detection. *Expert Systems with Applications*, 38(9), 11182-11192.
81. Chen, Z.Y., Tian, Y.H., Zeng, W. and Huang, T.J. (2015). *Detecting Abnormal Behaviors in Surveillance Videos Based on Fuzzy Clustering and Multiple Auto-Encoders*. 2015 IEEE International Conference on Multimedia & Expo (ICME), 1-6.
82. Acampora, G., Foggia, P., Saggese, A. and Vento, M. (2015). A hierarchical neuro-fuzzy architecture for human behavior analysis. *Information Sciences*, 310, 130-148.
83. Rojas, O.E. and Tozzi, C.L. (2016). Abnormal Crowd Behavior Detection Based on Gaussian Mixture Model. *Computer Vision - ECCV 2016 Workshops, Pt Ii*, 9914, 668-675.
84. Feng, Y.C., Yuan, Y. and Lu, X.Q. (2017). Learning deep event models for crowd anomaly detection. *Neurocomputing*, 219, 548-556.
85. Cheng, K.-W., Chen, Y.-T. and Fang, W.-H. (2015). *Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2909-2917 .
86. Alvar, M., Torsello, A., Sanchez-Miralles, A. and Armingol, J.M. (2014). Abnormal behavior detection using dominant sets. *Machine Vision and Applications*, 25(5), 1351-1368.
87. Mehran, R., Oyama, A. and Shah, M. (2009). *Abnormal Crowd Behavior Detection using Social Force Model*. CVPR: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Vols 1-4, 935.
88. Huang, B.S., Hsu, S.C. and Huang, C.L. (2016). *Abnormal Behavior Detection Using Conditional Random Fields*. Proceedings of 2016 International Conference on Audio, Language and Image Processing (ICALIP), 235-240.
89. Singh, D. and Mohan, C.K. (2017). Graph formulation of video activities for abnormal activity recognition. *Pattern Recognition*, 65, 265-272.

90. Mu, C.D., Xie, J.B., Yan, W., Liu, T. and Li, P.Q. (2016). A fast recognition algorithm for suspicious behavior in high definition videos. *Multimedia Systems*, 22(3), 275-285.
91. Wang, T., Chen, J., Zhou, Y. and Snoussi, H. (2013). Online Least Squares One-Class Support Vector Machines-Based Abnormal Visual Event Detection. *Sensors*, 13(12), 17130-17155.
92. Kumar, M. and Bhatnagar, C. (2017). Crowd Behavior Recognition Using Hybrid Tracking Model and Genetic algorithm Enabled Neural Network. *International Journal of Computational Intelligence Systems*, 10(1), 234-246.
93. Damla Arifoglu, A.B. (2017). *Activity Recognition and Abnormal Behaviour Detection with Recurrent Neural Networks*. The 14th International Conference on Mobile Systems and Pervasive Computing, 110, 86-93.
94. Ding, C.H., Fan, S.K., Zhu, M., Feng, W.G. and Jia, B.Z. (2014). Violence Detection in Video by Using 3D Convolutional Neural Networks. *Advances in Visual Computing (ISVC 2014), Pt II*, 8888, 551-558.
95. Ko, K.E. and Sim, K.B. (2018). Deep convolutional framework for abnormal behavior detection in a smart surveillance system. *Engineering Applications of Artificial Intelligence*, 67, 226-234.
96. Zhou, S.F., Shen, W., Zeng, D., Fang, M., Wei, Y.W. and Zhang, Z.J. (2016). Spatial-temporal convolutional neural networks for anomaly detection and localization in crowded scenes. *Signal Processing-Image Communication*, 47, 358-368.
97. Sabokrou, M., Fayyaz, M., Fathy, M. and Klette, R. (2017). Deep-Cascade: Cascading 3D Deep Neural Networks for Fast Anomaly Detection and Localization in Crowded Scenes. *IEEE Transactions on Image Processing*, 26(4), 1992-2004.
98. Paris, A., Arbaoui, S., Cislo, N., El-Amraoui, A. and Ramdani, N. (2015). *Using Hidden Semi-Markov Model for Learning Behavior in Smarthomes*. 2015 International Conference on Automation Science and Engineering (Case), 752-757.
99. Marsden, M., McGuinness, K., Little, S. and O'Connor, N.E. (2016). *Holistic Features for Real-Time Crowd Behaviour Anomaly Detection*. 2016 IEEE International Conference on Image Processing (ICIP), 918-922.
100. Fu, R.R., Wang, H. and Zhao, W.B. (2016). Dynamic driver fatigue detection using hidden Markov model in real driving condition. *Expert Systems with Applications*, 63, 397-411.
101. Gunduz, A.E., Temizel, T.T. and Temizel, A. (2014). *Pedestrian Zone Anomaly Detection by Non-Parametric Temporal Modelling*. 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (Avss), 131-135.
102. Wang, T., Qiao, M., Deng, Y., Zhou, Y., Wang, H., Lyu, Q., & Snoussi, H. (2018). Abnormal event detection based on analysis of movement information of video sequence. *International Journal for Light and Electron Optics*, 152, 50-60.

103. Batchuluun, G., Kim, J.H., Hong, H.G., Kang, J.K. and Park, K.R. (2017). Fuzzy system based human behavior recognition by combining behavior prediction and recognition. *Expert Systems with Applications*, 81, 108-133.
104. Yuan, Y., Feng, Y.C. and Lu, X.Q. (2018). Structured dictionary learning for abnormal event detection in crowded scenes. *Pattern Recognition*, 73, 99-110.
105. Xie, S. and Tu, Z. (2015). *Holistically-nested edge detection*. Proceedings of The IEEE International Conference On Computer Vision, 1395-1403.
106. Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012). *Imagenet classification with deep convolutional neural networks*. Advances in neural information processing systems.
107. Li, H., Lin, Z., Shen, X., Brandt, J. and Hua, G. (2015). *A convolutional neural network cascade for face detection*. Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition.
108. Zhou, S., Shen, W., Zeng, D., Fang, M., Wei, Y. and Zhang, Z. (2016). Spatial-temporal convolutional neural networks for anomaly detection and localization in crowded scenes. *Signal Processing: Image Communication*, 47, 358-368.
109. Ding, L., Fang, W., Luo, H., Love, P.E., Zhong, B. and Ouyang, X. (2018). A deep hybrid learning model to detect unsafe behavior: integrating convolution neural networks and long short-term memory. *Automation in Construction*, 86, 118-124.
110. Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A.K. and Davis, L.S. (2016). *Learning temporal regularity in video sequences*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 733-742.
111. Kingma, D.P. and Welling, M. (2014). *Stochastic gradient VB and the variational auto-encoder*. Second International Conference on Learning Representations, ICLR, 19.
112. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2014). *Generative adversarial nets*. Advances in neural information processing systems, 2672-2680 .
113. Sainath, T.N., Vinyals, O., Senior, A. and Sak, H. (2015). *Convolutional, long short-term memory, fully connected deep neural networks*. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4580-4584 .
114. Medel, J.R. and Savakis, A. (2016). Anomaly detection in video using predictive convolutional long short-term memory networks. *arXiv preprint arXiv:1612.00390*.
115. Nurullahoğlu, M. (2008). *Human Motion Analysis And Action Recognition*, Bahçeşehir Üniversitesi Institute of Science Computer Engineering Graduate Program, 62.
116. Baysal, S. (2011). *A Line Based Pose Representation For Human Action Recognition*, Bilkent University The Institute of Engineering and Science of Bilkent University.

117. Kurt, M.C. (2011). *A Key-Pose Based Representation for Human Action Recognition*, The Degree of Master of Science The Graduate School of Engineering and Science of Bilkent University.
118. Güçlü, A. (2012). *Gerçek Zamanlı Video Görüntülerinden İnsan Hareketlerinin Tanınması Ve Yorumlanması*, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü.
119. Pehlivan, S. (2012). *Multiple View Human Activity Recognition*, Doctor of Philosophy, Bilkent University The Graduate School of Engineering and Science.
120. Zeng, H., Li, L., Cao, Z. and Zhang, L. (2019). *Reliable and Efficient Image Cropping: A Grid Anchor based Approach*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
121. Agahian, S. (2018). *Uzay-Zaman Poz Çantası İle 3b İnsan Hareketlerinden Eylem Tanıma*, Doktora Tezi, Karadeniz Teknik Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı.
122. Yalçın, M. (2018). *Human Activity Recognition Using Deep Learning*, M.Sc. THESIS, Istanbul Technical University Department of Electronics and Communication Engineering.
123. Erkut, A. (2019). *İnsan Aktivitelerinin Derin Öğrenme Tekniği İle Tespit Edilmesi* Yüksek Lisans Tezi, Trakya Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı.
124. Keçeli, A.S. (2015). *Derinlik Bilgisi Kullanılarak İnsan Hareketlerinin Tanınması*, Doktora Tezi, Hacettepe Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı.
125. LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436.
126. McCulloch, W.S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133.
127. Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1985). Learning internal representations by error propagation. *California Univ San Diego La Jolla Inst for Cognitive Science*, 8506.
128. Hinton, G.E., Osindero, S. and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7), 1527-1554.
129. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929-1958.
130. Zaremba, W., Sutskever, I. and Vinyals, O. (2014). Recurrent neural network regularization. *arXiv preprint arXiv:1409.2329*.
131. Kingma, D.P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

132. Nair, V. and Hinton, G.E. (2010). *Rectified linear units improve restricted boltzmann machines*. Proceedings of the 27th International Conference on Machine Learning (ICML-10), 807-814.
133. Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.
134. LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
135. Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088), 533-536.
136. Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
137. Hubel, D.H. and Wiesel, T.N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1), 215-243.
138. Graves, A., Mohamed, A.-r. and Hinton, G. (2013). *Speech recognition with deep recurrent neural networks*. 2013 IEEE International Conference on Acoustics, Speech And Signal Processing, 6645-6649. .
139. Saha, S. and Raghava, G.P.S. (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins: Structure, Function, and Bioinformatics*, 65(1), 40-48.
140. Kamijo, K.-i. and Tanigawa, T. (1990). *Stock price pattern recognition-a recurrent neural network approach*. 1990 IJCNN International Joint Conference on Neural Networks, 215-221.
141. Xingjian, S., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K. and Woo, W.-c. (2015). *Convolutional LSTM network: A machine learning approach for precipitation nowcasting*. Advances in Neural Information Processing Systems, 802-810.
142. Masci, J., Meier, U., Cireşan, D. and Schmidhuber, J. (2011). *Stacked convolutional auto-encoders for hierarchical feature extraction*. International Conference on Artificial Neural Networks, 52-59.
143. Vincent, P., Larochelle, H., Bengio, Y. and Manzagol, P.-A. (2008). *Extracting and composing robust features with denoising autoencoders*. Proceedings of the 25th International Conference on Machine Learning, 1096-1103.
144. Gibson, J.J. (1950). The perception of the visual world.
145. Lucas, B.D. and Kanade, T. (1981). *An iterative image registration technique with an application to stereo vision*, Proceeding 7th Int Joint Conference on Artificial Intelligence (IJCAI), 674-679 .
146. Horn, B.K. and Schunck, B.G. (1981). Determining optical flow. *Artificial intelligence*, 17(1-3), 185-203.

147. Barron, J.L., Fleet, D.J. and Beauchemin, S.S. (1994). Performance of optical flow techniques. *International journal of computer vision*, 12(1), 43-77.
148. Farnebäck, G. (2003). *Two-frame motion estimation based on polynomial expansion*. Scandinavian Conference On Image Analysis, 363-370 .
149. Mei, S., Wang, Y. and Wen, G. (2018). Automatic fabric defect detection with a multi-scale convolutional denoising autoencoder network model. *Sensors*, 18(4), 1064.
150. Wang, Z., Simoncelli, E.P. and Bovik, A.C. (2003). *Multiscale structural similarity for image quality assessment*. The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2, 1398-1402 .
151. Lu, C., Shi, J. and Jia, J. (2013). *Abnormal event detection at 150 fps in matlab*. Proceedings of the IEEE International Conference on Computer Vision, 2720-2727 .
152. Mahadevan, V., Li, W., Bhalodia, V. and Vasconcelos, N. (2010). *Anomaly detection in crowded scenes*. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1975-1981.
153. Powers, D.M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, *Journal of Machine Learning Technologies*, 2, 37-63. .
154. Chong, Y.S. and Tay, Y.H. (2017). *Abnormal event detection in videos using spatiotemporal autoencoder*. International Symposium on Neural Networks, 189-196.
155. Luo, W., Liu, W. and Gao, S. (2017). *Remembering history with convolutional LSTM for anomaly detection*. 2017 IEEE International Conference on Multimedia and Expo (ICME), 439-444.
156. Yan, S., Smith, J.S., Lu, W. and Zhang, B. (2018). Abnormal Event Detection from Videos using a Two-stream Recurrent Variational Autoencoder. *IEEE Transactions on Cognitive and Developmental Systems*, 1-1.
157. Yang, B., Cao, J., Ni, R. and Zou, L. (2018). Anomaly detection in moving crowds through spatiotemporal autoencoding and additional attention. *Advances in Multimedia*, 2018.
158. Lee, S., Kim, H.G. and Ro, Y.M. (2018). *STAN: spatio-temporal adversarial networks for abnormal event detection*. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1323-1327.

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : DUMAN, Elvan
 Uyruğu : T.C.
 Doğum tarihi ve yeri : 24.07.1985, Kırıkkale
 Medeni hali : Evli
 Telefon : 0 (312) 202 85 83
 Faks : 0 (312) 202 89 47
 E-mail : elvanduman@gazi.edu.tr



Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Doktora	Gazi Üniversitesi / Bilgisayar Mühendisliği	Devam Ediyor
Yüksek lisans	Kırıkkale Üniversitesi / Bilgisayar Mühendisliği	2013
Lisans	Gazi Üniversitesi / Bilgisayar Mühendisliği	2018
Lisans	Muğla Üniversitesi – Elektronik ve Bilgisayar Eğitimi	2010
Lise	Kırıkkale Anadolu Ticaret Meslek Listesi	2005

İş Deneyimi

Yıl	Yer	Görev
2013-Halen	Gazi Üniversitesi	Araştırma Görevlisi
2011-2013	Kırıkkale Üniversitesi	Araştırma Görevlisi

Yabancı Dil

İngilizce

Yayınlar

1. Duman, E. and Erdem, O.A. (2019) Anomaly Detection in Videos Using Optical Flow and Convolutional Autoencoder. *IEEE Access*,7, 183914-183923.
2. Ünver, H.M., Kökver, Y., Duman, E. and Erdem, O.A. (2019). Statistical Edge Detection and Circular Hough Transform for Optic Disk Localization. *Applied Sciences*, 9(2), 350.
3. Duman, E. and Erdem, O.A. (2018). *A statistical edge detection framework for noisy images*. 2018 26th Signal Processing and Communications Applications Conference 1-4.
4. Duman, E., Kökver, Y., Ünver, H.M. and Erdem, O.A. (2017). *Automatic landmark detection through circular hough transform in cephalometric X-rays*. 2017 10th International Conference on Electrical and Electronics Engineering (ELECO), 583-587.
5. Duman, E. and Erdem, O.A. (2017). *A new image denoising method based on region growing segmentation*. 2017 25th Signal Processing and Communications Applications Conference (SIU), 1-4.
6. Duman, E. and Erdem, O.A. (2015). *Edge preserving image denoising method based on averaging*. 2015 23rd Signal Processing and Communications Applications Conference (SIU), 1857-1860.

Hobiler

Yüzme, doğa yürüyüşü, kamp



GAZİ GELECEKTİR..