

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

VİSİON TRANSFORMER VE RETENTİVE NETWORK
YÖNTEMLERİ İLE GÖRÜNTÜ ÜZERİNE
DUYGU ANALİZİ

VELİ DEMİR

YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

DANIŞMAN: DOÇ. DR. YAKUP GENÇ

TEMMUZ 2025

**T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

**VISION TRANSFORMER VE RETENTİVE NETWORK
YÖNTEMLERİ İLE GÖRÜNTÜ ÜZERİNE
DUYGU ANALİZİ**

VELİ DEMİR

**YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

DANIŞMAN: DOÇ. DR. YAKUP GENÇ

TEMMUZ 2025

T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL

**IMAGE-BASED SENTIMENT ANALYSIS WITH
VISION TRANSFORMER AND
RETENTIVE NETWORK METHODS**

VELİ DEMİR

**A THESIS OF MASTER OF SCIENCE
DEPARTMENT OF COMPUTER ENGINEERING**

ADVISOR: ASSOC. PROF. DR. YAKUP GENÇ

JULY 2025

YÜKSEK LİSANS JÜRİ ONAY FORMU

GTÜ Lisansüstü Eğitim Enstitüsü Yönetim Kurulunun 10/02/2025 tarih ve 2025/10 sayılı kararıyla oluşturulan jüri tarafından 27/05/2025 tarihinde tez savunma sınavı yapılan Veli Demir'in tez çalışması Bilgisayar Mühendisliği Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE

(TEZ DANIŞMANI) : Doç. Dr. Yakup GENÇ

ÜYE

: Dr. Öğr. Üyesi Salih SARP

ÜYE

: Dr. Öğr. Üyesi Tarkan Aydın

ONAY

Gebze Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü Yönetim Kurulunun

...../...../..... tarih ve/..... sayılı kararı.

İMZA/MÜHÜR

*Bu tez, bana olan sarsılmaz desteęi, sevgisi ve inancıyla en büyük güç
kaynaęım olan,
Aileme ithaf edilmiştir.*

ÖZET

Son yıllarda bilgisayarla görme ve veri odaklı modelleme alanındaki ilerlemeler sayesinde video duygu analizi alanında önemli gelişmeler yaşanmıştır. Video içeriğindeki duygusal ifadelerin anlaşılması, farklı alanlarda giderek daha fazla önem kazanmakta ve bu amaçla doğruluk ve verimliliği artırmak üzere yeni yöntemler geliştirilmektedir.

Bu çalışmada, Retentive Networkler ile Vision Transformerların entegrasyonu video verilerindeki duygu analizinde kullanılan bir çerçeve olarak incelenmiştir. Geleneksel görüntü tabanlı değerlendirmelere ek olarak, performans ve kaynak kullanımı arasında denge sağlamak amacıyla kısa video segmentlerinde kare örnekleme yöntemi uygulanmıştır (örneğin, saniyede 7 kare). Ardışık kareler arasındaki yüzler mekânsal tutarlılık kullanılarak takip edilmiş ve geçerli yüz dizileri üzerindeki sonuçların ortalaması alınarak duygu sınıflandırması gerçekleştirilmiştir. YOLO, SSD, Faster R-CNN, MTCNN, Haar Cascades ve RetinaFace gibi çeşitli yüz tespit modelleri test edilmiş olup, güvenilir performansı ve hafif hesaplama gereksinimi nedeniyle RetinaFace tercih edilmiştir.

Önerilen RetViT modeli, video tabanlı veri setinde 0.80 doğruluk oranı elde ederek, dinamik duygusal ifadeleri yakalamada etkinliğini göstermiştir. Karşılaştırmalı sonuçlar, önerilen modelin doğruluk, uyarlanabilirlik ve ölçeklenebilirlik açısından geleneksel modellere göre tutarlı gelişmeler sağladığını ortaya koymaktadır. Vision Transformerlar ve benzeri mimariler görüntü sınıflandırma görevlerinde yaygın olarak kullanılmış olsa da, bu çalışma onların video tabanlı duygu analizinde umut vadeden uygulamalarını vurgulamakta ve gerçek zamanlı uygulamalar gibi alanlarda gelecekteki araştırmalar için zemin hazırlamaktadır.

Anahtar Kelimeler: Video Duygu Analizi, Vision Transformer, Retentive Network, Yüz Algılama, Derin Öğrenme, Duygu Tanıma

ABSTRACT

The field of video sentiment analysis has seen considerable progress in recent years, supported by advances in computer vision and data-driven modeling. As understanding emotional expressions in video content gains importance across various domains, new methods are being developed to improve accuracy and efficiency.

This study explores the integration of Retentive Networks and Vision Transformers as a framework for analyzing emotions in video data. In addition to conventional image-based evaluations, an extended approach was applied to short video segments. Frames were sampled at reduced frequency (e.g., 7 frames per second) to balance performance and resource usage. Faces were tracked across consecutive frames using spatial consistency, and emotions were classified by averaging results across valid face sequences. Among several face detection models tested—including YOLO, SSD, Faster R-CNN, MTCNN, Haar Cascades, and RetinaFace—RetinaFace was selected for its reliable performance and lightweight computational demand.

Comparative results show that the proposed combination achieves consistent improvements over traditional models in terms of accuracy, adaptability, and scalability. While Vision Transformers and related architectures have been widely used in image classification tasks, this work highlights their promising application in video-based emotion analysis and sets the stage for future exploration, including potential use in real-time settings such as online meetings.

Keywords: Video Sentiment Analysis, Vision Transformers, Retentive Network, Face Detection, Deep Learning, Emotion Recognition

TEŞEKKÜR

Bu araştırma sürecinde bana yol gösteren, desteğini esirgemeyen ve rehberliğiyle çalışmanın yönünü belirlememde büyük katkı sağlayan danışmanım Doç. Dr. Yakup Genç'e en içten teşekkürlerimi sunarım.

Ayrıca, bu süreçte her zaman yanımda olan, sabırla destek veren ve moral kaynağım olan aileme en derin şükranlarımı sunuyorum.

İş yerimde bana zaman ve esneklik sağlayarak bu çalışmayı yürütmeme imkân veren müdürüm Sayın İbrahim Arslan'a da anlayışı ve desteği için teşekkür ederim.

Bununla birlikte, akademik bakış açımı geliştiren değerli tartışmaları ve deneyim paylaşımlarıyla katkı sağlayan meslektaşlarıma ve arkadaşlarıma da teşekkür ederim.

Son olarak, bu tezin hayata geçmesinde doğrudan ya da dolaylı katkısı olan herkese gönülden teşekkür ederim.

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	vi
ABSTRACT	vii
TEŞEKKÜR	viii
İÇİNDEKİLER	ix
SİMGELER ve KISALTMALAR DİZİNİ	x
ŞEKİLLER DİZİNİ	xi
TABLOLAR/ÇİZELGELER DİZİNİ	xii
1. GİRİŞ	1
1.1. Arka Plan ve Motivasyon	2
1.1.1. Ekonomik Kazanımlar	2
1.1.2. Ticari Başarı Potansiyeli	3
2. LİTERATÜR TARAMASI	5
2.1. Duygu Analizi Teknikleri ve Teknolojileri	5
2.1.1. Öğrenme Mimarileri	5
2.1.2. Mesafe ve Benzerlik Ölçütleri	6
2.1.3. Renk Dönüşüm Algoritmaları	8
2.1.4. Veri Artırma Teknikleri	9
2.1.5. Normalizasyon	9
2.1.6. Özellik Çıkarma Yöntemleri	9
2.2. Yüz Tespit Algoritmaları	11
2.3. Duygu Analizi Veri Kümeleri	13
2.3.1. FER-2013 Veri Kümesi	13
2.3.2. Video Tabanlı Veri Kümesi	14
3. YÖNTEM	16
3.1. Önerilen Sistem Mimarisi	16
3.1.1. Kamera Tabanlı Veri Toplama	17
3.1.2. Video Tabanlı Duygu Analizi	18
3.1.3. Sunucu Altyapısı	19
3.1.4. Kullanıcı Segmentasyonu ve Uygulama Bağlamları	19
3.2. Duygu Analizi için Optimal Teknoloji	20
4. DENEYLER VE SONUÇLAR	22
3.1. FER-2013 Veri Kümesi Üzerindeki Değerlendirme	22
3.2. Video Veri Kümesi Üzerindeki Değerlendirme	22
5. SONUÇLAR VE YORUMLAR	24
KAYNAKLAR	25
ÖZGEÇMİŞ	29
TEZ METNİNDEN TÜRETİLEN YAYINLAR/SUNUMLAR	30

SİMGELER ve KISALTMALAR DİZİNİ

ACC	: Accuracy
AI	: Artificial Intelligence
API	: Application Programming Interface
CNN	: Convolutional Neural Network
CNN-LSTM	: Convolutional Neural Network - Long Short-Term Memory
Dlib	: Dlib (Open-source Machine Learning Library)
DTW	: Dynamic Time Warping
FER	: Facial Expression Recognition
FPS	: Frames Per Second
GPU	: Graphics Processing Unit
GTÜ	: Gebze Teknik Üniversitesi
HOG	: Histogram of Oriented Gradients
ML	: Machine Learning
MSER	: Maximally Stable Extremal Regions
MTCNN	: Multi-task Cascaded Convolutional Networks
NLP	: Natural Language Processing
ResNet	: Residual Network
R-CNN	: Region-based Convolutional Neural Network
RN	: Retentive Network
SaaS	: Software as a Service
SIFT	: Scale-Invariant Feature Transform
SSD	: Single Shot MultiBox Detector
ViT	: Vision Transformer
YOLO	: You Only Look Once

ŞEKİLLER DİZİNİ

	<u>Sayfa</u>
Şekil 2.3.1: FER-2013 veri kümesinden 7 duyguya ait örnekler.	13
Şekil 2.3.2: Video veri kümesinden örnekler ve analiz sonuçları.	15
Şekil 3.1: Önerilen duygu analizi sistem mimarisi.	17



TABLÖLAR DİZİNİ

	<u>Sayfa</u>
Tablo 4.1: FER-2013 veri kümesi için sonuçların karşılaştırması.	22
Tablo 4.2: Video tabanlı kendi veri kümemiz için sonuçların karşılaştırması.	23



1. GİRİŞ

Duygu tanıma ya da duygu analizi, insan yüzündeki ifadelerin sınıflandırılması yöntemidir. Genel olarak bu sınıflandırmada yedi temel ifade kullanılır. Bazı uygulamalarda bu analiz yalnızca görüntülerle değil, metin ve ses verilerinin bir araya gelmesiyle yapılır. Bu çalışma ise görüntü işleme alanında başarı oranını artırmayı amaçlamaktadır. Retentive Network yapıları, Doğal Dil İşleme (NLP) alanında umut vadeden sonuçlar göstermiştir; bağlamı kapsamlı biçimde anlayabilme yetenekleri ve anlamlı ilişkiler kurabilmeleri, onları metin analizi için oldukça uygun hale getirir. Diğer yandan, Vision Transformer (ViT) modelleri, görüntü segmentleri üzerinde şaşırtıcı derecede başarılı sonuçlar vermektedir. Manhattan Distance, özellikle görüntü veya piksel tabanlı sistemlerde, ızgara yapısındaki iki nokta arasındaki mesafeyi hesaplamak için kullanılan bir tekniktir ve bu modeller bu sistemle ilginç biçimde bir araya getirilebilir.

Bu çalışmada, Retentive Network ve Vision Transformer birleşiminin video üzerinden duygu analizi için ne derece uygun olduğu tartışılmakta ve diğer modellere göre ne gibi avantajlar sunduğu analiz edilmektedir [2], [3]. Son dönem çalışmalarda, Retentive Network ile Vision Transformer modellerinin birleşimi, RNN, Transformer ve CNN gibi rekabetçi modellerin önüne geçmiştir. RNN ve CNN modelleri tek başlarına yalnızca yüz ifadeleri gibi fiziksel özellikleri dikkate alırken, bu bütünleşik yapı çevresel detayları ve bağlamı da göz önünde bulundurur. Üstelik klasik transformer modelleri verinin tamamı üzerindeki bağlam bilgisine odaklanırken, Retentive Network yapıları ilgili olan bağlamı önceliklendirerek gereksiz yükü azaltır.

Bu, Vision Transformer modellerini CNN tabanlı yöntemlere göre daha esnek hale getirir; zira CNN'ler genellikle sabit boyutlu girdilere ihtiyaç duyarken, Vision Transformer yapıları farklı çözünürlüklerdeki video verileriyle de çalışabilir. Sonuç olarak, Manhattan Distance destekli Retentive Network yapılarının Vision Transformer üzerinde çalışması video tabanlı duygu analizinde yeni bir paradigma oluşturma potansiyeline sahiptir. Büyük modellerin davranışlarının daha iyi anlaşılması ve teknolojilerin birleştirilerek yeni uygulamalara öncülük etmesiyle birlikte, bu modellerin evrimleşerek daha doğru ve etkili duygu analizleri yapması mümkün olacaktır. Bu teknoloji, video duygu analizinde standartları yukarıya

taşıyarak Yapay Zeka ve Makine Öğrenimi'nin dijital deneyimlerde daha fazla anlayış ve bağ kurabilmesine olanak tanıyacaktır. Bu nedenle, bu modeller üzerine yapılacak Ar-Ge yatırımlarının artırılması gerektiği açıktır.

1.1. Arka Plan ve Motivasyon

Canlı destek sistemlerine, çevrim içi müşteri görüşmelerine, toplantılara, mağaza kameralarına ve akıllı araçlara entegre edilebilecek bir müşteri memnuniyeti ve personel analiz paneli hayal edin. Bu çözüm; canlı destek görüşmeleri ve mağaza görüntülerini işleyerek müşterilerin ne kadar memnun olduğunu izleyebilecek, temsilci tahminleri sağlayabilecek ve akıllı temsilci yönlendirmeleri sunabilecektir. Duygular hakkında ön bilgi sağlayarak canlı destek sistemlerinin sürece daha doğru yaklaşmasına, müşterileri doğru temsilciye yönlendirmesine ve kriz anlarında öncelik tanınmasına olanak sağlayacaktır. Ayrıca bu sistem, çevrim içi görüşmeler ve toplantılarda katılımcıların ne kadar mutlu veya mutsuz olduklarını ve bu durumun ne kadar sürdüğünü analiz ederek öz değerlendirme ve personel analizini de mümkün kılacaktır.

Sistem, mağaza kamera sistemlerine entegre edilerek müşterilerin duygusal durumlarını analiz ederek öz değerlendirme ve çalışan analizi sağlayacaktır. Kameralar reklam panolarına yerleştirildiğinde, insanların reklama tepkileri ölçülebilecektir. İstenirse bir ekran yerleştirilerek, potansiyel müşteriden ekrana kısaca bakarak reklamla ilgili duygusunu belirtmesi istenebilir ve bu duygu anlık olarak gösterilebilir. Bu etkileşim, müşterinin teknolojiye dokunduğu hissini artıracak ve marka itibarı yükselmiş olacaktır. Ayrıca bu sistem, akıllı araçlarda yolcuların duygusal durumuna göre yardım hizmetlerinin çeşitlendirilmesini sağlayacaktır. Örneğin, “Moralin bozuk görünüyor, biraz hareketli müzik çalalım mı?” gibi proaktif yönlendirmeler kullanıcının teknolojiyle bağ kurmasını sağlayacaktır.

1.1.1 Ekonomik Kazanımlar

Gerçek zamanlı SaaS video ve fotoğraf duygu analizi teknolojisi şirketlerin kullanımına sunulabilecektir. Böylece diğer teknolojiler de kendi ürünlerinin yeteneklerini bu çözümle artırabilecektir. Teknolojinin her firma tarafından tekrar tekrar üretilmesine gerek kalmayacak, yeni fikirlerin hayata geçirilmesi

kolaylaşacaktır — tıpkı Google Maps'in taksi uygulamasında kullanılması gibi. Chatbot, çevrim içi görüşme, mağaza, robot veya akıllı araç uygulamalarında entegre edilebilir. Sistem, öz değerlendirme imkânı sağlayarak müşteri memnuniyetini artıracaktır. Müşterilerin ve çalışanların duygusal durumları çevrim içi görüşmeler ve kameralar yoluyla analiz edilebilecek, öfkeli ve mutsuz müşterilere kriz oluşmadan önce müdahale edilecek, mutlu müşteriler elde tutulabilecek ve mutsuzluk kaynaklı gelir kayıpları en aza indirilebilecektir.

Bu teknolojinin sağladığı ekonomik faydalar birçok sektörde verimliliği artırma potansiyeline sahiptir. Şirketler, müşteri geri bildirimlerini ve kullanıcı deneyimlerini bu gelişmiş analiz araçlarıyla daha hızlı ve etkili şekilde değerlendirebileceklerdir. Bu teknolojiler yalnızca müşteri ilişkilerini geliştirmekle kalmayıp ürün ve hizmet optimizasyonuna da katkı sağlayacaktır. Analizlerden elde edilen verilerle, pazarlama stratejileri daha isabetli şekilde belirlenecek ve müşteri memnuniyeti artırılabilecektir. Gerçek zamanlı duygu analizi teknolojileri, işletmelerin operasyonel süreçlerini dönüştürerek sektörel verimliliği artırabilecek ekonomik bir çarpan görevi görecektir.

Bu tür analiz araçları müşteri ilişkilerini iyileştirmenin ötesine geçerek ürün ve hizmetleri geliştirmeye katkı sağlar. Analizler aracılığıyla elde edilen veriler, işletmelerin pazarlama taktiklerini daha doğru anlamasını sağlar ve müşteri memnuniyetini yükseltir. Duygu analizine dayalı bu teknolojiler perakendeden sağlığa kadar birçok sektörde kullanılabilir. Örneğin bir perakende mağazası, müşteri etkileşimlerini analiz ederek daha uygun ürün yerleşimleri planlayabilir ya da yeni bir mağaza tasarlayabilir. Sağlık sektöründe ise, hastaların duygusal tepkileri izlenerek kişiselleştirilmiş tedavi süreçleri uygulanabilir. Bu tür yenilikçi teknolojilerin yaygın kullanımı, iş süreçlerinin kolaylaştırılmasına ve işletmelerin pazarda daha güçlü konumlanmasına olanak tanıyacaktır. Gelecekte, daha özel kontrol sistemleri, ekonomik süreç iyileştirmeleri ve veri otomasyonundaki fırsatlar sayesinde daha yenilikçi çözümler mümkün olacaktır.

1.1.2. Ticari Başarı Potansiyeli

Müşteri deneyimi ve çalışan içgörüsüne yönelik ciddi boşlukları doldurmayı hedefleyen bir ürün, özellikle görüntü ve video temelli duygu analizi başarısına dayandığında yüksek ticari potansiyele sahiptir. Şirketler için müşteri memnuniyeti ve

personel analizleri rekabet avantajı sağlama noktasında öncelikli alanlar hâline gelmiş ve bu tür çözümlere olan talep artmıştır. Paydaş mimarisi, özel ihtiyaçları olan bireysel ve kurumsal müşterilere hizmet edecek şekilde tasarlanmıştır.

Geniş müşteri hizmeti veri kümeleri içinde, şirketler duygu analizi çözümleriyle kullanıcı deneyimini artırabilir ve çalışan performansını izleyebilir. Farklı platformlara entegrasyon sayesinde bu çözümler birçok sektörde yaygın şekilde kullanılabilir. Sistem yalnızca tek bir platforma değil; canlı destek sistemleri, çevrim içi tüketici görüşmeleri, toplantılar, mağaza kameraları, reklam panoları ve akıllı araçlara kadar geniş bir kullanım yelpazesine sahiptir.

Analiz sonuçlarına dayanarak şirketler daha bilinçli kararlar alabilir ve iş süreçlerini, kullanıcı deneyimini geliştirecek adımlar atabilirler. Bu teknoloji, işletmelere müşteri memnuniyetine dair içgörü sunar ve personel performansını izleme imkânı verir. Veri analizi, müşteri ilişkilerini geliştirmek, hizmet kalitesini artırmak ve insan kaynakları yönetimini güçlendirmek için değerli bir araç hâline gelir. Bu unsurlar da projenin ticari başarı potansiyelini doğrudan artıran temel etmenlerdir.

2. LİTERATÜR TARAMASI

2.1. Duygu Analizi Teknikleri ve Teknolojileri

Duygu analizi, insan duygularını anlamak amacıyla yüz ifadelerini algılama ve analiz etme sürecidir. Bu, görüntü veya videolarda yüzleri doğru şekilde tanımlamayı ve anlamlı yüz özelliklerini çıkarmayı gerektirir. Bu tekniklerdeki ilerlemeler, farklı uygulamalarda duyguların güvenilir şekilde yorumlanmasını mümkün kılmıştır.

2.1.1. Öğrenme Mimarileri

Bu makalede, günümüzde bilgisayarla görme çalışmalarında yaygın olarak kullanılan bazı deep learning modellerini inceleyeceğiz. Bu yöntemlerin temel prensipleri, avantajları ve dezavantajlarını Convolutional Neural Networks (CNN), ResNet50, EfficientNet, MobileNet, Vision Transformers, Retention Networks ve VGG19 modellerini dikkate alarak ele alacağız:

- **Vision Transformers (ViTs) [2], [3]:** Vision Transformers, NLP alanında tanıtılan transformer mimarisi üzerine kurulu, görüntü tanıma için yeni bir yöntem ailesidir. ViTs, görüntüleri metindeki tokenlara benzer şekilde sabit boyutlu parçalara ayırır, bu parçaları vektörlere dönüştürür ve mekânsal bilgiyi korumak için pozisyonel kodlamalar ekler. Ardından, çok başlıklı self-attention ve feed-forward katmanlarından oluşan transformer encoder'lerden geçirilir. Son sınıflandırma aşamasında ise bir sınıflandırma başlığı kullanılır. ViTs, uzun menzilli ilişkileri yakalamada güçlüdür ve attention tabanlı yapısı nedeniyle esnektir; ancak genellikle büyük veri setlerine ihtiyaç duyar ve hesaplama açısından ağırdır.
- **Convolutional Neural Networks (CNNs) [12], [13]:** CNN'ler görüntü sınıflandırmada yaygın şekilde kullanılır. Görüntüden mekânsal özellikleri çıkarır, boyutları pooling katmanları ile azaltır ve dense katmanlar ile sınıflandırma yapar. CNN'ler daha küçük veri setlerinde verimli çalışır ve özellik çıkarma konusunda etkilidir; ancak model boyutları büyük olabilir ve hesaplama maliyeti yüksektir.

- **ResNet50 [11]:** Residual Network ailesinin bir üyesi olan ResNet50, derin ağlarda karşılaşılan gradyan azalması sorununu skip bağlantılarla çözer. Her residual blok, giriş ile convolution çıktısını toplar. Bu sayede daha derin modellerin eğitilmesine olanak sağlar ve gradyan akışını iyileştirir; fakat daha fazla hafıza gerektirir ve doğru ayarlanmazsa overfitting'e yol açabilir.
- **EfficientNet [14]:** EfficientNet, derinlik, genişlik ve çözünürlüğü dengeli şekilde ölçeklendiren compound scaling yaklaşımını önerir. Depthwise separable convolution ve squeeze-and-excitation modülleri kullanır. EfficientNet, daha az parametre ile yüksek doğruluk sağlar ancak dikkatli manuel ayar gerektirebilir ve eğitim süresi uzun olabilir.
- **MobileNet [15], [17]:** MobileNet, mobil ve gömülü cihazlar için tasarlanmıştır. Standart convolutions işlemlerini depthwise ve pointwise convolutions olarak ikiye ayırır. MobileNet hafif ve hızlıdır, gerçek zamanlı uygulamalar için idealdir; ancak daha büyük modellere göre doğrulukta biraz kayıp yaşanabilir.
- **VGG19 [14], [21]:** VGG19, 19 katmanlı derin bir CNN modelidir. Çok sayıda 3×3 convolution filtreleri ve son kısımda sınıflandırma için fully connected katmanlar kullanır. Uygulaması ve anlaşılması kolaydır ve iyi genelleme yapar; ancak model boyutu büyük ve hafıza tüketimi yüksektir.

Sonuç olarak, her modelin kendi güçlü ve zayıf yönleri vardır. Doğru model seçimi, yapılacak işe, mevcut veri setine ve hesaplama kaynaklarına bağlıdır. Bazı modeller doğruluk açısından daha üstünken, diğerleri gerçek zamanlı veya mobil uygulamalar için daha verimli olabilir.

2.1.2. Mesafe ve Benzerlik Ölçütleri

Manhattan Distance'ın Retentive Networks ve Vision Transformers ile birlikte kullanılması, model performansını artırmak için değerli bir yaklaşımdır. Bu bölümde, istatistiksel karşılaştırma amacıyla değerlendirilen bazı yöntemler; dayandıkları gerekçeler, avantajları ve dezavantajlarıyla birlikte ele alınmaktadır:

- **Manhattan Distance [29],** aynı zamanda L1 normu ya da taxicab distance olarak da adlandırılır; bir ızgara üzerindeki iki nokta arasındaki uzaklığı, koordinatlar arasındaki mutlak farkların toplamı olarak hesaplar. İsmi, şehir

bloklarının ızgara benzeri yapısından alır. Avantajları arasında, yalnızca basit toplama ve mutlak farklara dayanması nedeniyle yüksek boyutlu verilerde bile hızlı çalışması yer alır. Euclidean Distance'a kıyasla daha az hesaplama maliyeti gerektirir. Ayrıca, tek bir boyuttaki büyük farkların sonucu orantısız şekilde etkilemesini engellediği için uç değerlere karşı daha dayanıklıdır. Özellikle birçok özelliğin sıfır ya da düşük katkı sunduğu seyrek veri yapılarında etkili çalışır. Ayrıca, ızgara hareketleri ya da ayırık eylemler içeren senaryolarda kolay yorumlanabilir olması önemli bir avantajdır. Bununla birlikte, sürekli uzaylarda düzgünlük sağlamaması dezavantaj oluşturabilir. Ayrıca, Euclidean Distance kadar sezgisel geometrik yorum sunmaz ve tüm boyutların eşit katkı sunduğu varsayımına dayanır ki bu her veri türü için geçerli olmayabilir. Düzensiz yapıli ortamlarda en kısa yolu yansıtamayabilir.

- **Euclidean Distance [30]**, iki nokta arasındaki düz çizgi mesafesini hesaplar ve sıklıkla özellik uzaylarındaki gömülmelerin karşılaştırılmasında kullanılır. Basitliği ve doğrudan benzerlik ifadesi sunması nedeniyle yaygındır; ancak karekök işlemi nedeniyle daha fazla hesaplama maliyeti getirir ve uç değerlere karşı hassastır.
- **Cosine Similarity [31], [36], [37]**, sıfır olmayan iki vektör arasındaki açıya odaklanır ve yönelimi karşılaştırır; büyüklüğü dikkate almaz. Yüksek boyutlu verilerde iyi performans gösterir ve vektör uzunluklarını göz ardı eder. Bununla birlikte, gerçek mesafeyi veremez ve sıfır vektörlere karşı duyarlıdır.
- **Mahalanobis Distance [32]**, veri içindeki değişkenler arası korelasyonu dikkate alır ve ölçekten bağımsızdır. Bu sayede, bir nokta ile bir dağılım arasındaki mesafeyi ölçmek için uygundur. Veri bağımlılıklarını hesaba kattığı ve çok değişkenli uç değerleri belirlemede etkili olduğu için avantajlıdır. Ancak, kovaryans matrisinin hesaplanması oldukça maliyetlidir ve doğru sonuçlar için yüksek hassasiyet gerektirir.
- **Hamming Distance [33]**, bir dizinin diğerine dönüştürülmesi için gerekli minimum karakter değişikliklerini sayar ve çoğunlukla kategorik veya ikili verilerde kullanılır. Basitliği ve hata tespiti için uygunluğu avantajlarındandır; ancak sürekli veri üzerinde uygulanamaz ve yalnızca sınıflandırılmış verilere özgüdür.

- **Jaccard Similarity [34]**, sonlu örnek kümelerini karşılaştırır ve kesişim ile birleşim oranına göre benzerlik hesaplar. Küme karşılaştırmaları ve kümeleme algoritmaları için etkilidir. Ancak sürekli verilere uygulanamaz ve seyrek veriye karşı duyarlıdır.
- **Wasserstein Distance [35]**, ya da diğer adıyla Earth Mover's Distance, bir dağılımdan diğerine geçmek için gereken minimum maliyeti ölçer. Olasılık dağılımlarının karşılaştırılması için kullanılır. Farklılıkları daha iyi yansıtmaları ve uç değerlere karşı dayanıklı olması avantajları arasındadır. Bununla birlikte, hesaplama açısından yoğundur ve lineer programlama çözümleri gerektirir.
- **Dynamic Time Warping (DTW) [36]**, zaman içinde hızları değişebilen seriler arasındaki benzerliği ölçer. Zaman eksenini esneterek dizileri en uygun şekilde hizalamayı hedefler. Zaman serisi verilerinde esnek karşılaştırma sağlaması önemli bir avantajdır; ancak hesaplama maliyeti yüksektir ve gürültüye karşı hassastır.
- **Bhattacharyya Distance [37]**, iki olasılık dağılımı arasındaki yakınlığı ölçer ve özellikle özellik seçimi ile sınıflandırma uygulamalarında kullanılır. Olasılık dağılımlarını karşılaştırmakta etkili olmasına rağmen, bu dağılımların doğru biçimde değerlendirilmesini gerektirir ve boyut sayısı yüksek olduğunda boyut indirgeme yapılmadan uygulanması zordur.

Genel olarak, Manhattan Distance; hesaplama maliyeti, uç değerlere karşı dayanıklılık ve belirli veri türlerine olan uygunluğu açısından dengeli bir seçenektir. Retentive Networks ile Vision Transformers arasında bağ kurmak için uygun bir adaydır. Bununla birlikte, etkinliği kullanılacak veri setine ve görev tanımına göre değişebilir.

2.1.3. Renk Dönüşüm Algoritmaları

Renk dönüşüm algoritmaları, görüntüleri bir renk uzayından başka bir renk uzayına dönüştürmek için kullanılır; örneğin RGB'den grayscale ya da HSV'ye. Bu adım, bazı renk bileşenlerinin diğerlerinden daha anlamlı olabileceği bilgisayarla görme görevlerinde önemli bir rol oynar. Avantajları arasında, renk bilgisindeki fazlalığın azaltılması sayesinde, renkli pikseller dizisi gibi birçok benzer öğeye sahip görüntü verisinin sadeleştirilmesi yer alır. Bu sayede her bileşeni ayrı ayrı analiz etmek her zaman gerekli olmayabilir. Ayrıca, karanlık bir büyük görüntüdeki kenar gibi belirli özelliklere odaklanmayı kolaylaştırır ve bu da sonraki işlemler için fayda sağlar.

Ancak dezavantajları arasında, dönüşüm sırasında bazı renk bilgilerinin kaybolması yer alır ve bu durum analiz türüne göre önemli olabilir. Ayrıca, dönüşümün etkinliği büyük ölçüde yürütülen spesifik göreve bağlıdır.

2.1.4. Veri Artırma Teknikleri

Veri artırma, mevcut verilere gürültü, döndürme ve diğer dönüşümler gibi çeşitli varyasyonlar eklenerek veri kümesinin yapay olarak genişletilmesini sağlayan bir tekniktir. Makine öğrenmesi uygulamalarında modelin dayanıklılığını artırmak için yaygın olarak kullanılan bir yaklaşımdır. Avantajları arasında daha iyi genelleme yeteneği bulunur; zira model, çeşitlendirilmiş veri formları ile eğitildiğinde, daha önce görülmemiş veriler üzerindeki tahmin başarısı artar. Ayrıca, gerçek dünya koşullarındaki kamera gürültüsü veya nesne yönelimindeki değişiklikler gibi kusurlara karşı dayanıklılığı artırır. Dezavantajları ise, veri karmaşıklığının ve boyutunun artması nedeniyle eğitim süresinin uzamasıdır. Ayrıca, modele anlamlı olmayan özellikler yerine gürültü kaynaklı örüntüleri öğrenme riski nedeniyle overfitting oluşabilir.

2.1.5. Normalizasyon

Normalizasyon, verinin belirli bir küçük aralığa (örneğin $[0, 1]$) ölçeklenmesi sürecidir ve öğrenme algoritmalarının daha hızlı yakınsamasını sağlar. Bu işlemin avantajları arasında, farklı özelliklerin ortak bir ölçeğe getirilmesiyle tutarlılığın artırılması ve model eğitiminin iyileştirilmesi yer alır. Ayrıca, veri dağılımındaki çarpıklığın azaltılması yoluyla algoritmaların daha verimli çalışmasına katkı sunar. Bununla birlikte, normalizasyon dikkatli uygulanmazsa, önemli bağlamsal bilgilerin kaybolma riski vardır. Ayrıca, min-max scaling veya z-score gibi uygun normalizasyon tekniğinin seçimi de ayrı bir zorluk oluşturabilir.

2.1.6. Özellik Çıkarım Yöntemleri

Bu teknikler, ham verideki anlamlı örüntüleri veya ayırt edici özellikleri aramak ve böylece makine öğrenimi modellerini daha hızlı ve daha etkili hale getirmek amacıyla kullanılır. Avantajları arasında, yüksek boyutlu verinin daha düşük boyutlu, bilgi açısından zengin temsillere indirgenmesiyle etkin veri yönetimi sağlanması ve

yalnızca en önemli bilgilere odaklanılarak model doğruluğunun artırılması yer alır. Bununla birlikte, bu yöntemler genellikle yüksek hesaplama gücü ve zaman gerektirir. Ayrıca uygulanmaları karmaşık olabilir ve veri yapısının, çıkarılacak özelliklerin ve görev gereksinimlerinin iyi anlaşılmasını zorunlu kılar. En yaygın kullanılan özellik çıkarım yöntemleri aşağıda, temel ilkeleri ve performans nitelikleriyle birlikte özetlenmiştir:

- **Histogram of Oriented Gradients (HOG) [24]**, bilgisayarla görme alanında nesne tespiti için yaygın olarak kullanılan bir tanımlayıcıdır. Görüntünün yerel bölümlerindeki gradyan yönelimlerinin dağılımını analiz ederek nesneleri tanımlar. HOG, ışık değişimlerine ve gölgelere karşı görece dayanıklıdır; bu da tutarlı özellik çıkarımı sağlar. Özellikle, güvenlik kamerası videolarında insan gibi nesnelerin tespiti için etkilidir. Ancak, hesaplama açısından yoğun kaynak tüketir; bu da gerçek zamanlı uygulamalarda performans sınırlamalarına yol açabilir. Ayrıca, karmaşık ya da kalabalık arka planlarda etkinliği azalabilir.
- **Maximally Stable Extremal Regions (MSER) [27]**, görüntülerdeki blob yapılarını, farklı eşik seviyeleri boyunca kararlılığını koruyan bölgeleri bularak tespit eder. Farklı çevresel koşullarda dahi tutarlı özellikler sunarak yüksek tekrarlanabilirlik sağlar; bu da özellikle metin tespiti gibi yüksek doğruluk gerektiren görevlerde avantaj yaratır. Öte yandan, MSER gürültüye karşı duyarlıdır ve bazen alakasız bölgeleri de algılayabilir. Ayrıca, parametrik ayarlarının karmaşıklığı nedeniyle deneyimli kullanıcılar için bile uygulaması zor olabilir.
- **VLFeat [28]**, SIFT ve HOG gibi birçok bilgisayarla görme algoritmasını içeren açık kaynaklı bir kütüphanedir. Ekim 2023'e kadar olan veri kümesi desteğiyle geniş kapsamlı veri işleyebilir ve diğer kütüphane ve uygulamalarla kolayca entegre olarak hızlı prototipleme süreçlerine olanak tanır. Ancak, VLFeat'in sunduğu geniş özellik seti, uygulamanın ve veri boyutunun büyüklüğüne bağlı olarak yüksek hesaplama kaynakları gerektirebilir. Etkin şekilde kullanımı için bilgisayarla görme konularında sağlam bir bilgi birikimi gerekir.
- **Scale-Invariant Feature Transform (SIFT) [25], [26]**, görüntülerdeki yerel özellikleri bulur ve tanımlar. Üretilen keypoint'ler, ölçek, döndürme ve affine

dönüşümlerine karşı değişmezlik sağlar. Bu sayede farklı görüntüleme koşullarına uyarlanabilir. SIFT, çok çeşitli dönüşümlere karşı dayanıklılık ve zengin tanımlayıcılar sunarak görüntü eşleştirme doğruluğunu artırır; bu da nesne tanıma gibi görevlerde avantaj sağlar. Ancak geçmişte patent kısıtlamalarına tabi olması (ki bu kısıtlamalar sonlanmış olabilir) ve hesaplama açısından yüksek maliyetli olması, özellikle gerçek zamanlı uygulamalarda kullanılabilirliğini sınırlamıştır.

SIFT, özellikle ölçek ve yön değişmezliğinin kritik olduğu eşleştirme ve tanıma görevlerinde öncelikli tercih olurken; hızın daha önemli olduğu veya ışık koşullarının değişken olduğu durumlarda HOG tercih edilebilir. Yöntem seçimi, belirli görev ihtiyaçlarına ve mevcut hesaplama kaynaklarına bağlıdır.

2.2. Yüz Tespit Algoritmaları

YOLO [6], SSD, Faster R-CNN, MTCNN, Dlib, Mediapipe, Haar Cascades (OpenCV), RetinaFace ve YuNet gibi çeşitli yüz tespit kütüphaneleri mevcuttur. Bu çalışmada değerlendirilen yöntemler aşağıda listelenmiştir:

- **Haar Cascades (OpenCV) [9], [10]**, gerçek zamanlı uygulamalar için hızlıdır ve eğitim gerektirmediğinden doğrudan kullanılabilir. OpenCV gibi yaygın kütüphaneler içinde yerleşik olarak sunulmaktadır. Ancak, ışık ve bakış açısındaki değişimlere karşı hassastır, yanıltabilir ve karmaşık arka planlarda doğruluğu düşüktür.
- **MTCNN (Multi-task Cascaded CNN) [8]**, yüz tespiti ve göz, burun, ağız gibi yüz özelliklerinin belirlenmesinde iyi hassasiyet sunar. Farklı yüz pozları ve aydınlatma koşullarında başarılı çalışır. Ancak, hızı nispeten düşük olduğu için gerçek zamanlı uygulamalar için uygun olmayabilir ve özellikle düşük kaliteli videolarda yanlış tespitler yapabilir.
- **Dlib kütüphanesi [16]**, yüz hatlarını işaretleme konusunda oldukça hassas ve doğrudur; farklı ifadeler ve pozlarla çalışmak için uygundur. Ancak, gerçek zamanlı uygulamalarda daha fazla işlem gücüne ihtiyaç duyar ve düşük çözünürlüklü ya da karışık görüntülerde doğruluğu azalabilir.
- **YOLOv8 [6]**, nesneleri hızlı ve etkili biçimde gerçek zamanlı olarak tespit eder ve tek bir model ile hem yüz hem diğer nesne tespiti yapabilmesi açısından

esneklik sağlar. Yüz tespiti ve sınırlayıcı kutular (bounding box) için iyi doğruluk sunar. Bununla birlikte, YOLO gibi derin öğrenme tabanlı yöntemler genellikle GPU gibi güçlü donanım gerektirir. Belirli senaryolarda hassas ayarlamalar yapılabilmesi için büyük ve çeşitli veri setlerine ihtiyaç duyar. Ayrıca, çok karmaşık ya da kalabalık arka planlarda yanlış pozitifler veya kaçırılan tespitler meydana gelebilir.

- **SSD (Single Shot MultiBox Detector) [17]**, YOLO'ya benzer şekilde hızlı bir nesne tespiti algoritmasıdır ve gerçek zamanlı uygulamalar için uygundur. VGG ve MobileNet gibi önceden eğitilmiş ağlarla kolayca entegre edilebilir. Ancak, küçük nesnelerin tespitinde daha az başarılı olabilir ve özel uygulamalar için özelleştirilmiş eğitim gerekebilir.
- **Faster R-CNN [18], [19]**, region proposal networks (bölge öneri ağları) sayesinde daha yüksek doğruluk sağlar ve özellikle karmaşık senaryolarda etkilidir. Yüksek çözünürlüklü görüntülerde daha ayrıntılı sonuçlar sunar ve küçük nesnelerde iyi performans gösterir. Ancak, diğer yöntemlere kıyasla daha yavaştır; bu da gerçek zamanlı uygulamalar için uygunluğunu azaltır. Aynı zamanda daha fazla hesaplama kaynağına ihtiyaç duyar.
- **Mediapipe [21]**, hızlı ve verimli çalışması sayesinde özellikle mobil ve web tarayıcı tabanlı uygulamalarda uygundur. Android ve iOS platformlarında sorunsuz çalışabilir. Ancak çok karmaşık ortamlarda doğruluğu azalabilir ve özelleştirme seçenekleri sınırlıdır; çünkü genel kullanım senaryoları için optimize edilmiştir.
- **RetinaFace [7]**, yüz tespitinde yüksek doğruluk oranları sunar ve birden fazla yüzü aynı anda etkili şekilde tespit edebilir. Ancak, daha fazla işlem gücü gerektirir; bu da düşük güçlü cihazlarda kısıtlayıcı olabilir. Özelleştirilebilir olmakla birlikte, yapılandırması için ileri düzey bilgi gerektirir.
- **YuNet [22]**, katman ve hesaplama açısından iyi optimize edilmiş yapısı sayesinde hızlı çalışır. Hem mobil hem de masaüstü uygulamalarda kullanılabilir. Ancak karmaşık duygu analizi veya poz tespiti gibi görevlerde performansı sınırlı olabilir ve başarımı büyük ölçüde eğitimde kullanılan veri setinin çeşitliliğine bağlıdır.

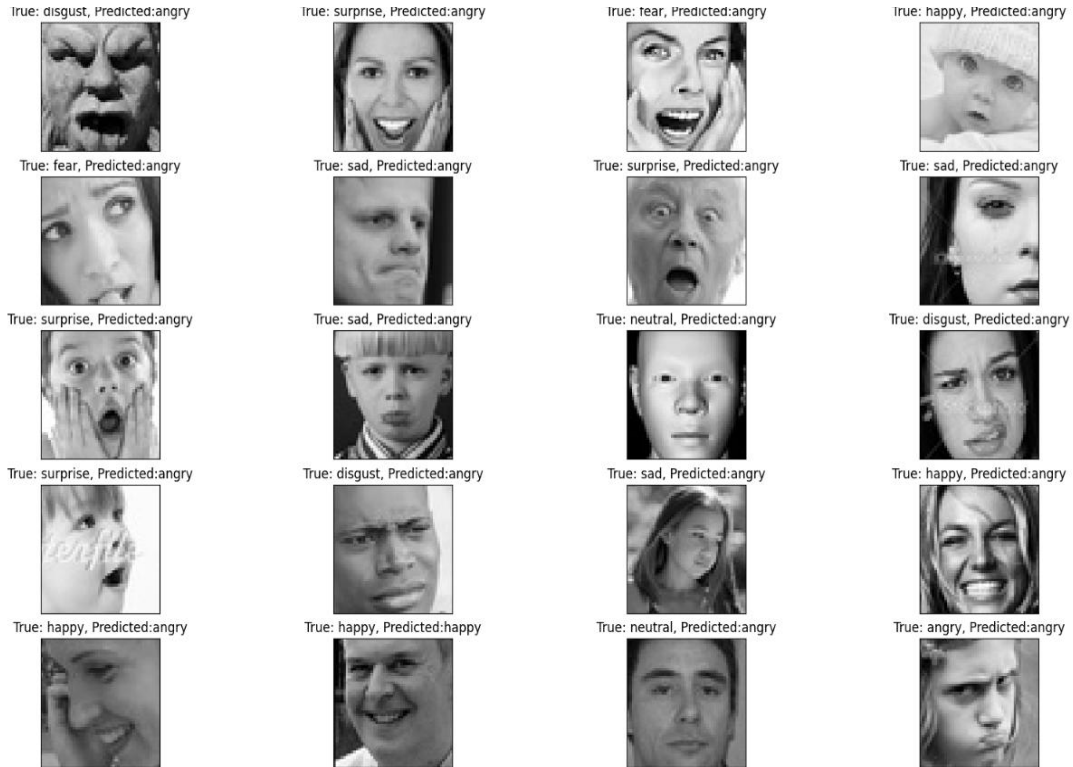
En iyi sonuçlar MTCNN, Haar Cascades (OpenCV), RetinaFace ve SSD yöntemleri ile elde edilmiştir. Daha az kaynak gerektirmesi nedeniyle RetinaFace yöntemiyle

devam edilmiştir. Bu modeller karşılaştırılırken resmi doğruluk testleri yapılmamış, değerlendirmeler manuel testler yoluyla gerçekleştirilmiştir.

2.3. Duygu Analizi Veri Kümeleri

Duygu analizi araştırmaları, modellerin etkili biçimde eğitilmesi ve değerlendirilmesi için özenle hazırlanmış veri setlerine büyük ölçüde bağımlıdır. Bu veri setleri; biçim, karmaşıklık düzeyi ve uygulama bağlamı açısından çeşitlilik göstermektedir. Alandaki temel veri setleri iki ana kategoride toplanabilir: statik görüntü veri setleri ve video-based veri setleri. Statik veri setleri, örneğin FER-2013, etiketlenmiş yüz görüntülerini kontrollü koşullarda sunarak birçok duygu analizi algoritması için karşılaştırmalı değerlendirme olanağı sağlar. Öte yandan, video-based veri setleri zaman içinde değişen yüz ifadelerini yakalayarak, gerçek dünya etkileşimlerini daha doğru yansıtan zengin zamansal bilgi içerir. Biz ise kafayı sağa sola sallamanın olumsuz olması gibi zamansal bağlamda elde edilebilecek sonuçları yok sayıp direk frame'ler ve ortalamaları üzerinden sonuç elde ediyoruz. Bu çalışmada hem halka açık veri seti olan FER-2013 hem de kendi hazırladığımız video-based veri seti analiz edilmiştir.

2.3.1. FER-2013 Veri Kümesi



Şekil 2.3.1: FER-2013 veri kümesinden 7 duyguya ait örnekler.

FER-2013 (Facial Expression Recognition 2013) [1], [5], insan yüz ifadelerini sınıflandırmak amacıyla oluşturulmuş popüler bir dataset'tir. Bilgisayarla görme ve makine öğrenimi alanlarında yüz ifadelerinin tanınması için model eğitimi ve değerlendirmesinde yaygın biçimde kullanılmaktadır. FER-2013 dataset'i, 48x48 piksel boyutunda toplam 35.887 gri tonlamalı görüntü içermektedir. Düşük çözünürlüklü ve standart biçimde sunulan bu görüntüler, yedi farklı yüz ifadesi sınıfını kapsamaktadır: angry, disgust, fear, happy, sad, surprise ve neutral.

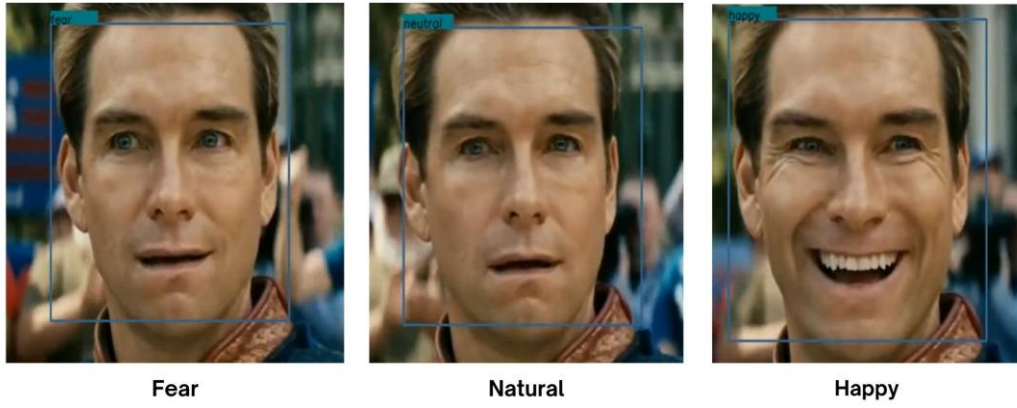
FER-2013 dataset'inin avantajları arasında, yedi duygusal durumu kapsayan geniş etiket kümesi, küçük boyut ve gri tonlama nedeniyle veri işleme süreçlerini kolaylaştıran standart yapısı ve açık erişimli olması sayesinde akademik çalışmalarda sıkça tercih edilmesi sayılabilir. Bununla birlikte, bazı dezavantajları da bulunmaktadır: düşük çözünürlük, ayrıntılı yüz ifadelerinin gerektiği uygulamalar için yetersiz olabilir; sınıf dağılımı dengesizdir (örneğin happy sınıfı, angry sınıfından daha fazla örneğe sahiptir); ayrıca dataset'in laboratuvar ortamında veya yapay şekilde oluşturulmuş görüntüler içermesi, gerçek dünya senaryolarını tam olarak yansıtamayabilir.

Sonuç olarak, FER-2013 dataset'i, yüz ifadesi tanıma alanında başlangıç araştırmaları ve model geliştirme çalışmaları için değerli bir kaynak oluşturmaktadır. Ancak, gerçek dünya uygulamaları için ek veri kaynaklarıyla desteklenmesi ve uygun ön işleme adımlarıyla tamamlanması gerekebilir.

2.3.2. Video Tabanlı Veri Kümesi

Bu çalışmada kullanılan video-based dataset, kayıtlı video dizilerinden çıkarılan çeşitli yüz ifadesi örneklerinden oluşmaktadır. Bu dataset, her biri temel yedi duygusal kategoriye temsil eden yaklaşık 10.000 örnek içeren toplam 70.000 renkli görüntüden oluşmaktadır. Standartlaştırılmış statik görüntü dataset'lerinden farklı olarak, bu koleksiyondaki görüntüler video karelerinden doğrudan yüz bölgesinin kırılması yoluyla elde edilmiştir. Bu nedenle, görüntü çözünürlükleri ve en-boy oranları değişkenlik göstermektedir. Örneğin, 27×39, 38×49, 51×74 gibi çeşitli boyutlarda görüntüler mevcuttur. Bu durum, gerçek dünya video kayıtlarında karşılaşılan doğal değişkenliği yansıtmaktadır.

Veri kümesi, FER-2013 veri kümesinde olduğu gibi aynı yedi farklı yüz ifadesi sınıfını kapsamaktadır: Angry, Disgust, Fear, Happy, Sad, Surprise ve Neutral. Bu kategoriler, duygu analizi alanında yaygın olarak kabul gören standartlarla uyumludur ve dinamik, gerçek yaşam koşullarında model eğitimi ve değerlendirmesi için kapsamlı bir temel sağlar. Dataset'teki görüntü boyutları, aydınlatma koşulları ve poz farklılıkları, duygu analizi algoritmalarının dayanıklılığını ve genellenebilirliğini değerlendirmek için hem zorlu hem de gerçekçi bir ortam sunar.



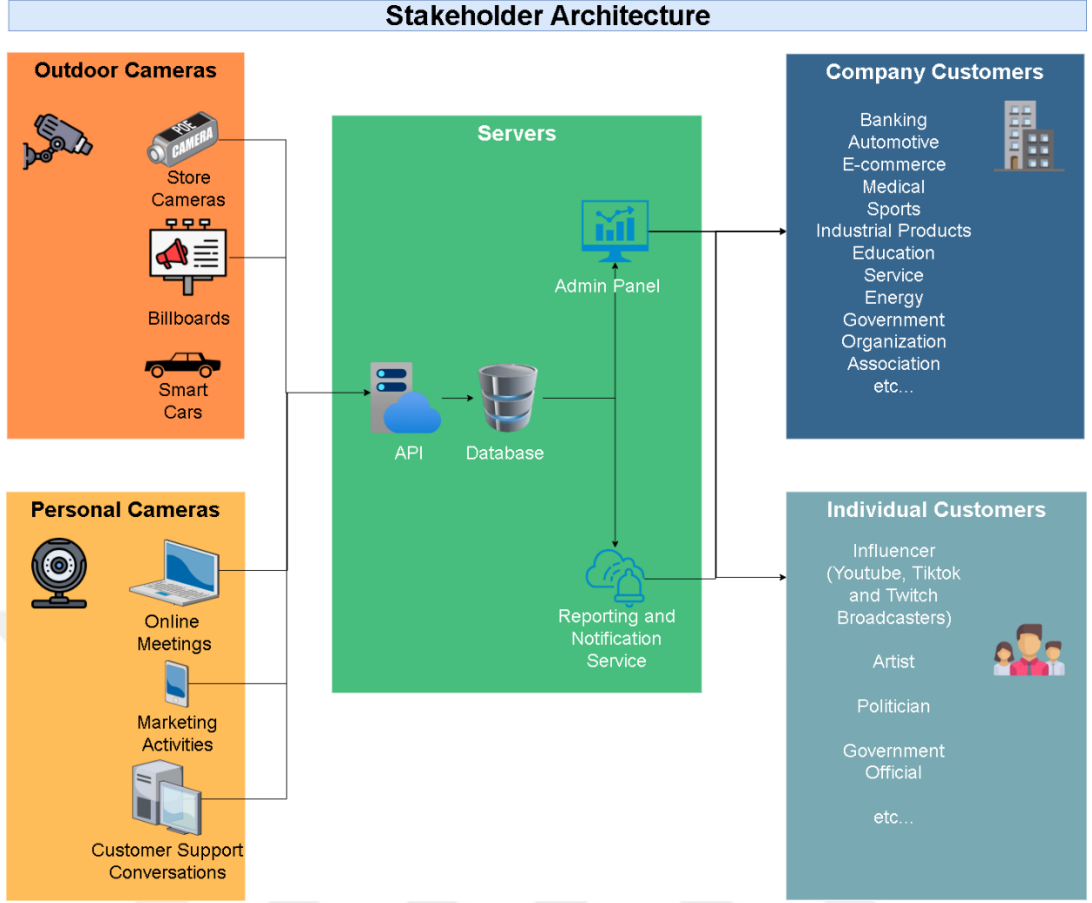
Şekil 2.3.2: Video veri kümesinden örnekler ve analiz sonuçları.

3. YÖNTEM

Bu bölümde, geliştirilen duygu analiz sisteminin teknik temelleri iki ana başlık altında ele alınmaktadır. İlk olarak, çok kaynaklı kamera verilerinin toplanması, işlenmesi ve son kullanıcıya servis edilmesini sağlayan yazılım platformunun mimarisi ve bileşenleri tanıtılmaktadır. Ardından, bu altyapı üzerinde çalışan yapay zekâ tabanlı duygu analizi modellerinin tasarımı, eğitimi ve uygulanmasına ilişkin yöntemler açıklanmaktadır. Bu iki temel başlık, sistemin genel başarımını doğrudan etkileyen yazılım mühendisliği ile makine öğrenmesi süreçlerini bütüncül biçimde değerlendirmeyi hedeflemektedir. öngörülmüştür.

3.1. Önerilen Sistem Mimarisi

Bu bölümde, önerilen sistemin genel tasarımı ve mimari bileşenleri sunulmaktadır. Yöntem, çoklu kamera kaynakları, sunucu altyapısı ve kullanıcı odaklı uygulamaların nasıl bir araya getirilerek sağlam bir duygu analizi platformu oluşturduğunu ortaya koymayı amaçlamaktadır. Sistem, farklı ortamlar ve kullanım senaryoları için verimli çalışacak şekilde tasarlanmıştır ve çeşitli son kullanıcı profillerine uyarlanabilir yapıdadır. Sistem; çeşitli kamera türleri aracılığıyla veri toplama, yapılandırılmış bir arka uç (backend) üzerinden veri işleme ve özel müşteri profillerine yönelik hizmetler aracılığıyla çıktının sunulması adımlarını içermektedir. Bu çok katmanlı sistem mimarisinin şematik genel görünümü, görüntü elde etmeden analitik hizmet sunumuna kadar olan temel işlem akışını gösteren **Şekil 3.1**'de sunulmuştur. Bu yapı, hem bireysel hem kurumsal müşterilerin çeşitli ihtiyaçlarını destekleyebilmek açısından kritik önemdedir.



Şekil 3.1: Önerilen duygu analizi sistem mimarisi.

3.1.1. Kamera Tabanlı Veri Toplama

Kamera kaynakları, farklı uygulama ortamlarında veri toplama ve analiz süreçlerinin temelini oluşturur. Genel olarak iki ana kategoriye ayrılır: kişisel kameralar ve dış ortam kameraları, ki bunlar farklı amaçlara hizmet eder. Dış ortam kameraları, mağaza kameraları, reklam panosu (billboard) kameraları ve akıllı araç kameralarını içerir. Mağaza kameraları, güvenlik amacıyla kullanılmasının yanı sıra, mağaza içi tüketici davranışlarını anlamak açısından da önemlidir. Billboard kameraları, yoldan geçen bireylerin reklamlarla ne düzeyde etkileşim kurduğunu algılayarak gerçek etkileşim düzeyini ölçebilir. Akıllı araç kameraları ise yol koşullarını değerlendirmek ve navigasyon sistemlerini iyileştirmek için kullanılır. Kişisel kameralar, çevrim içi toplantılarda, satış süreçlerinde, pazarlama kampanyalarında ve müşteri hizmeti etkileşimlerinde kullanılan çeşitli kamera türlerini içerir. Bu kameralar, kişiselleştirilmiş kullanıcı deneyimleri oluşturmak için zengin veri sağlar. Pazarlama alanında, müşteri tepkilerini ve geri bildirimlerini ölçerek kampanya etkinliğini

artırmaya yardımcı olur. Müşteri destek süreçlerinde ise, doğrudan müşteriyle etkileşim kurulmasına olanak sağlar ve sorunların çözümünde etkin rol oynar.

3.1.2. Video Tabanlı Duygu Analizi

Görüntü ve video tabanlı duygu analizinde esas odak noktası duygusal ifadelerin doğru şekilde tanınmasıdır. Ancak duygusal ifadeler oldukça karmaşıktır ve kültürler ve bireyler arasında değişkenlik gösterebilir. Bu nedenle, duyguların doğru şekilde sınıflandırılıp yorumlanmasına yönelik geliştirilen modellerin doğruluğu ve güvenilirliği konusunda belirsizlikler söz konusudur. Veri miktarı arttıkça kapsayıcılık artacak ve öğrenilmemiş verilerde başarı oranı yükselecektir. Ancak bu veri artışı aynı zamanda yapay zekânın eğitim süresini ve tepki süresini de artıracaktır. Veri miktarı ve başarı ihtiyacı arttıkça sistem gereksinimi de artmaktadır. Bu sorunun çözümü için yapay zekâ öğrenme modelinin ve sistem mimarisinin geliştirilmesi gerekmektedir.

Teknolojinin sunumunda mikroservis mimarileri denenecek ve sistem ihtiyacını yatay olarak bölerek birçok container'ın tek bir hizmet olarak çalıştırılması sağlanabilecektir. Verinin anlık olarak gönderilip alınabileceği bir teknoloji geliştirilmelidir. Örneğin, bu verinin TCP protokolüyle parçalı bir şekilde (örneğin 5 MB'lık dilimler halinde) REST servisi olarak gönderilmesi ve yanıt alınması, veya SignalR gibi bir yapıyla iki yönlü Web Socket veri paketleri ile gönderilmesi, yahut UDP protokolü ile teknolojiye canlı bağlanılıp bağlantı kopmadan analiz sonucunun alınması gibi tüm teknolojiler denenerek nihai çözüm geliştirilebilir.

Ayrıca birden fazla yüzün yer aldığı ve düşük kaliteli görüntülerde duygu analizinin yapılması hedeflenmektedir. Bu yüzlerin bulunması, video süresince bu işlemin tekrarlanmaması ve kişinin sahneden çıkıp geri döndüğünde aynı kişi olduğunun anlaşılması gibi problemler de çözülmesi gereken konulardır. Bu bağlamda, bir saniyelik video kesiti her ne kadar 24 kare içerse de, kaynak kullanımını azaltmak amacıyla yalnızca 7 kare üzerinden analiz gerçekleştirilmiştir. Her karedeki yüzler tespit edilmiş, ardışık karelerde yaklaşık olarak aynı ekran konumunda yer alan yüzler aynı kişi olarak kabul edilmiştir. Yüzün belirli bir piksel mesafesini aşarak yer değiştirmesi veya sahne geçişi yaşanması hâlinde ise takip işlemi sonlandırılmıştır. Teorik olarak bir yüz için saniyede 7 kare işlenebilir olsa da, en az 3 kareye ulaşamaması durumunda ilgili yüz analiz kapsamı dışında bırakılmıştır. Kamera

kapanması veya sahne geçişi gibi nedenlerle tam 7 karelik takip sağlanamasa da, 3 karelik veri analiz için yeterli kabul edilmiştir. Bu şekilde elde edilen her yüz için 3 ila 7 kare arasında duygu analizleri gerçekleştirilmiş; her bir duygu kategorisi için model çıktılarındaki olasılık değerlerinin ortalaması alınarak, en yüksek ortalama olasılığa sahip duygu, ilgili zaman aralığında kişinin duygusu olarak belirlenmiştir.

3.1.3. Sunucu Altyapısı

Sunucu altyapısı, kamera kaynaklarından toplanan verilerin işlenmesini ve dağıtılmasını sağlayan temel iskelet yapısını oluşturur. Bu yapı, aşağıda belirtilen birkaç temel bileşenden oluşur: Yönetim paneli (admin panel), sistem içindeki farklı bileşenlerin yapılandırılması ve kontrol edilmesi için bir merkez görevi görür. Yöneticiler, sistemin işleyişini izleme, parametreleri ayarlama ve genel mimarinin düzgün çalışmasını sağlama yetkisine sahiptir. API (Application Programming Interface), teknoloji yığını içindeki farklı bileşenler arasında veri akışını sağlayarak sistem içi iletişimi mümkün kılar. Ayrıca, üçüncü parti sistemlerle entegrasyonlara olanak tanıyarak mevcut altyapının esnekliğini ve gücünü artırır. Veritabanı, büyük miktarda veriyi güvenli bir şekilde saklar ve bu bilgilere erişimi kolaylaştırarak yönetilebilir kılar. Bu kaynak, hem kurumsal hem bireysel kullanıcılar için karar alma ve stratejik planlama süreçlerinde kritik öneme sahiptir. Raporlama ve bildirim hizmetleri, toplanan verilerden içgörü üretir ve paydaşlara zamanında bilgi sağlar. Raporlama araçları, veri eğilimlerinin yorumlanmasına yardımcı olurken, bildirim sistemleri önemli olaylar veya gerekli aksiyonlar hakkında kullanıcıları bilgilendirir.

3.1.4. Kullanıcı Segmentasyonu ve Uygulama Bağlamları

Müşteriler kurumsal ve bireysel olarak kategorize edilmektedir. Kurumsal müşteriler; kamu kuruluşları ve dernekler gibi sektörlerde faaliyet gösteren işletmeleri içermekte olup, mimari yapı bu müşterilerin operasyonel gereksinimlerine göre uyarlanmıştır. Örneğin, bazı müşteriler potansiyel risklerin ortadan kaldırılmasına odaklanabilirken, e-ticaret şirketleri müşteri deneyimindeki sürtünme noktalarının giderilmesine öncelik verebilir. Sistem, müşteri etkileşimi ve kişiselleştirilmiş pazarlama için temel bir çerçeve sunmaktadır. Bireysel müşteriler ise YouTube, TikTok ve Twitch yayıncıları gibi influencer'lar, sanatçılar, politikacılar ve kamu görevlilerinden oluşmaktadır.

Influencer'lar mimariyi içerik ve izleyici analizleri için, sanatçılar hayran etkileşimlerini takip etmek için kullanabilirken; politikacılar ve kamu görevlileri için bu yapı halkla ilişkiler ve iletişim stratejilerini destekleyici niteliktedir. Özetle, bu paydaş mimarisi; kamera kaynakları, güçlü sunucu altyapısı ve geniş müşteri tabanı ile oldukça karmaşık bir yapıya sahiptir. Bu unsurların kusursuz bir şekilde birleştirilmesi sayesinde işletmeler daha çevik biçimde operasyonel verimliliklerini artırabilir, hizmet tekliflerini yenileyebilir ve dijital dünyanın hızlı değişen taleplerine daha iyi yanıt verebilir.

3.2. Duygu Analizi için Optimal Teknoloji

Bu çalışmada, yüksek kaynak kullanımı ve bazı önemli özelliklerin kaybı riskleri nedeniyle SIFT ve HOG gibi Özellik Çıkarım Yöntemleri tercih edilmemiştir. Ancak, duygu analizi için bu yöntemler oldukça etkili olabilir. Görsel verilerdeki kritik desen ve özellikleri detaylı şekilde ortaya çıkararak yüz ifadelerindeki ince değişikliklerin tanınmasında önemli avantajlar sunarlar. Hesaplama açısından yoğun olmalarına rağmen, duygusal ipuçlarını belirginleştirmeleri sayesinde duygu analizi sistemlerinin doğruluğunu ve etkinliğini önemli ölçüde artırır.

Yöntemimiz, Retentive Networks ve Vision Transformers modellerinin Manhattan Distance yöntemiyle birleştirilmesini içermektedir. Retentive Network [2], [3], zaman serileri veya sıralı verilerle çalışırken uzun menzilli korelasyonları ve geçmiş bilgileri etkili biçimde hatırlama gibi avantajlara sahiptir; önemli zaman periyotlarını ya da öznitelikleri hedef alarak karar verme sürecinde etkili olabilir. Ancak karmaşık yapısı nedeniyle hesaplama açısından maliyetli olabilir ve bazı durumlarda küçük değişimlere aşırı duyarlı hale gelebilir. Vision Transformers (ViT) [2], [3] ise görüntünün farklı bölgeleri arasındaki uzun mesafeli korelasyonları öğrenmede etkili olan küresel attention mekanizması sunar; konvolüsyon yerine attention tabanlı yaklaşımlar sayesinde farklı veri türlerinde esnek kullanım sağlar. Bu modelin dezavantajları ise büyük veri ihtiyacı ve yüksek parametre sayısı nedeniyle eğitim ve ayarlama süreçlerinin zor olmasıdır.

Bu modelleri birleştirmede iki temel yaklaşım vardır. RMT modeli, RetNet'in sıralı verilerdeki unutma (decay) mekanizmasını görsel verilerin uzamsal düzenine uyarlayarak Manhattan Self-Attention ile görsel dikkat yapısına açık mekânsal

öncüller kazandırmaktadır [2]. Benzer şekilde, ViR (Vision Retention Networks) modeli, RetNet'ten esinlenen dual paralel ve ardışık formülasyonlarla hem eğitim hem de çıkarımda verimli görsel işlemeye olanak sağlar [3]. Bu modellerde geçen “zaman” veya “temporal” kavramı, klasik anlamıyla gerçek dünya zamanının akışı olarak değil, görüntü verisinin işleme sürecinde ortaya çıkan soyut bir sıralama veya ardışıklık olarak ele alınır. Görüntüler, Vision Transformer yaklaşımlarında olduğu gibi küçük parçalara (patch) bölünür ve bu parçalar model içinde bir dizi şeklinde sırayla işlenir. Bu işleme sürecinde her bir patch, modelin hesaplama adımlarında belirli bir pozisyonda yer alır; bu pozisyonlar, bir nevi “zamansal” veya ardışık adım gibi değerlendirilir. Temporal decay mekanizması ise, modelin bu ardışık pozisyonlar arasında bilgi aktarımını düzenlerken, önceki patch'lerden gelen bilgilerin etkisini aşamalı olarak azaltmasını sağlar. Böylece, model çok uzak geçmişteki (pozisyon olarak çok önceki) bilgilerin güncel kararlar üzerindeki etkisini azaltırken, daha yakın ve ilgili bilgileri önceliklendirebilir. Bu yaklaşım, hem hesaplama verimliliğini artırır hem de uzun menzilli ilişkilerin modellenmesinde daha dengeli ve seçici bir bilgi akışı sağlar.

RMT ve ViR ayrımları ise, temporal decay mekanizmasını uygulama biçimleri ve mimari tasarımlarında ortaya çıkar. RMT, bilgiyi daha esnek ve seçici bir şekilde saklayıp unutmaya odaklanırken, ViR bu süreci daha yalın ve hesaplama açısından verimli hale getirmeyi amaçlar. RMT'de retentive mekanizmalar detaylı bilgi tutma ve unutma stratejileri içerirken, ViR daha sade ve optimize edilmiş bir retention yapısı ile benzer işlevi daha hızlı ve hafif şekilde gerçekleştirir. RMT daha karmaşık ve güçlü bilgi yönetimi sağlarken, ViR performans ve verimlilik odaklı bir yaklaşım sunar.

RMT ve ViR modelleri, temel olarak statik görüntüler üzerinde çalışan derin öğrenme mimarileridir ve doğrudan video ya da gerçek zamanlı zaman serisi verisiyle çalışmazlar. Videodan duygu analizi çalışmamızda ViR tercih edilmiştir çünkü hesaplama açısından daha verimli ve hızlıdır, bu sayede yüksek hacimli video verisini daha pratik ve gerçek zamanlı işleyebilir. Yöntemimizde kafa sallama gibi önceki saniyeye bağlı veriler yok sayılmış, her bir kare bağımsız işlenmiştir. Sınıflandırma sonuçlarının ortalaması alınarak videodan duygu analizi ihtiyacı çözümlenmiştir.

4. DENEYLER VE SONUÇLAR

4.1. FER-2013 Veri Seti Üzerinde Değerlendirme

FER-2013 veri seti kullanılarak T4 GPU üzerinde eğitim ve test işlemleri gerçekleştirilmiştir. Eğitim süresinin uzun olmasına rağmen yüksek başarı sağlandığı söylenebilir. **Tablo 4.1**'de güncel yöntemler incelendiğinde, yüksek başarı ve düşük sistem gereksinimi sağlayan yöntemlerin K-Fold testleri ile belirlendiği görülmektedir. K-Fold çapraz doğrulama testi, eğitim ve test setlerinin örneğin 1/10 oranında ayrılarak 10 kez test edilmesi anlamına gelir. Bu sayede modele sorulan öğrenilmemiş verilerin ezberlenmesi ve 10 defa sorulduğunda yanlış çıkarımlar yapılması sorunları aşılmış olur. Elbette geliştirme sürecini hızlandırmak için 10 kez eğitim ve test yapmak doğru değildir. İlk eleme için yüksek başarının elde edildiği senaryoların 3 kez veya daha az test edilmesi yeterli olacaktır. Sonuçları paylaşırken 1/10 k-fold deneyleri 3 kez yapılmış ve ortalamalar alınmıştır.

Tablo 4.1: FER-2013 veri kümesi için sonuçların karşılaştırması.

Model	Eğitim Süresi	Test Süresi (3.589 görüntü)	Kayıp (Loss)	Doğruluk (Accuracy) 3 tekrar 1/10 k-fold
RetViT (Bizim)	01:01:08	00:00:18	0.75	0.73
ViT	01:01:08	00:00:18	0.75	0.73
VGG19	00:23:28	00:00:10	0.92	0.69
CNN	00:11:04	00:00:03	1.16	0.63
EfficientNet	00:06:02	00:00:05	1.10	0.61
ResNet 50	00:08:34	00:00:05	1.37	0.55

4.2. Video Veri Seti Üzerinde Değerlendirme

Video analizleri, görüntü analizlerinden farklı ihtiyaçlar doğurur. Örneğin, bir videoda insanların yüzleri bir kare içinde hareket eder, ancak her seferinde yüz alanını tekrar yakalamaya çalışmak veya her kareyi ayrı ayrı analiz etmek yüksek maliyetlere yol açar. Mevcut video ve yayın verilerinde bu ihtiyaçların ve kişinin aynı olup olmadığı sorununun araştırılması ve çözülmesi gerekir. Video verisiyle yapılan testlerde

MTCNN, Haar Cascades (OpenCV), RetinaFace ve SSD yöntemlerinden en iyi sonuçlar alınmıştır. Kaynak ihtiyacının daha az olması nedeniyle RetinaFace ile devam edilmiştir. Bu modellerin karşılaştırmasında resmi bir doğruluk testi yapılmamış, değerlendirme manuel testlerle gerçekleştirilmiştir. CNN, ResNet 50, EfficientNet, MobileNet, VGG19 ve RetViT yöntemleri karşılaştırıldığında RetViT en iyi sonucu vermiştir ancak eğitim ve test sürelerinin uzun olduğu görülmüştür. Vision Transformer ve Ret+Vit modelleri rakiplerine göre çok daha yüksek başarı sunmaktadır. Ayrıca Ret+Vit modeli tek başına Vision Transformer modeline kıyasla eğitim ve test süresi yarı yarıya azalmış, başarı oranı %77'den %80'e yükselmiştir.

Tablo 4.2: Video tabanlı kendi veri kümemiz için sonuçların karşılaştırması.

Model	Eğitim Süresi	Test Süresi (3.589 görüntü)	Doğruluk (Accuracy)
RetViT (Our)	01:32:06	00:00:45	0.80
VGG19	00:51:49	00:00:10	0.78
ViT	02:50:42	00:01:32	0.77
MobileNet	00:36:18	00:00:12	0.71
ResNet 50	01:32:22	00:00:05	0.65

5. SONUÇLAR VE YORUMLAR

Sonuç olarak, bu iki güçlü yöntemin kombinasyonu RetViT, özellikle karmaşık veri setlerinde duygusal ifadelerin sınıflandırılmasında yeni kapılar açmakta ve mevcut model performanslarını artırmaktadır. Video analizinde ise yüz tanıma tekniklerinin kaynak gereksinimleri ve doğruluğu göz önüne alınarak RetinaFace yöntemi tercih edilmiştir. Makalede yöntemlerin avantajları ve dezavantajları kısaca ele alınmıştır. Veri setimiz modellere dahil edildiğinde başarı oranı artmaya devam etmiş ve diğer modellere göre eğitim süresi fazla artmamıştır. Bu durumda Ret+Vit için veri seti büyüdükçe, eğitim ve test sürelerinin diğer modellere göre fazla artmayacağı, yalnız Vision Transformers’a kıyasla çok daha az zaman gerektireceği ve başarı oranını yükselteceği öngörülmektedir. Geliştirilen uygulama sayesinde, gerçek zamanlı kamera görüntülerinde tespit edilen yüzler üzerinde modellerin çıktıları etkin biçimde görüntülenebilmekte ve değerlendirilebilmektedir. Yüz ifadelerinin analizi problemi için hem görüntü (FER-2013) hem de video analizinde RetViT ve RetinaFace en iyi yöntemler olarak bulunmuştur.

Bununla birlikte, kaynak kullanımı, eğitim süresi ve çözülmesi gereken özgün problemlere bağlı olarak farklı yöntemlerin tercih edilmesi gerekebilir. Örneğin, kameradaki kişinin şirket üniforması giyip giymediğinin tespiti veya bir aracın yetkili sürücü tarafından kullanılıp kullanılmadığı gibi yeni uygulama senaryoları, farklı ve optimize edilmiş çözüm yaklaşımlarını zorunlu kılacaktır.

KAYNAKLAR

- [1] P.-L. Carrier and A. Courville, "FER2013 dataset," Kaggle, 2024. [Online]. Available: <https://www.kaggle.com/datasets/deadskull7/fer2013>. [Accessed: Jul. 17, 2025].
- [2] Q. Fan, H. Huang, M. Chen, H. Liu, and R. He, "RMT: Retentive networks meet vision transformers," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Seattle, WA, USA, 2024, pp. 5641–5651, doi: 10.1109/CVPR52733.2024.00539.
- [3] A. Hatamizadeh, M. Ranzinger, and J. Kautz, "VIR: Vision retention networks," arXiv, 2023. [Online]. Available: <https://arxiv.org/abs/2310.19731v1>. [Accessed: Jul. 17, 2025].
- [4] P. Shah, D. Ambekar, H. Bodat, and S. Kumari, "Multimodal sentiment analysis: Techniques, implementations and challenges across diverse modalities," in Proc. 11th Int. Conf. Comput. Sustainable Global Dev. (INDIACom), New Delhi, India, 2024, pp. 405–413, doi: 10.23919/INDIACom61295.2024.10498914.
- [5] M. C. Gursesli, S. Lombardi, M. Duradoni, L. Bocchi, A. Guazzini, and A. Lanata, "Facial emotion recognition (FER) through custom lightweight CNN model: Performance evaluation in public datasets," IEEE Access, vol. 12, pp. 45543–45559, 2024, doi: 10.1109/ACCESS.2024.3380847.
- [6] S. V. M. Dev Vanamoju, M. V. Vineetha, H. Tekchandani, P. Joshi, P. K. Shukla, and A. Khanna, "Facial emotion recognition using YOLO based deep learning classifier," in Proc. 1st Int. Conf. Electron., Commun. Signal Process. (ICECSP), New Delhi, India, 2024, pp. 1–5, doi: 10.1109/ICECSP61809.2024.10698173.
- [7] M. E. Wibowo, A. Ashari, A. Subiantoro, and W. Wahyono, "Human face detection and tracking using RetinaFace network for surveillance systems," in Proc. IECON 2021 – 47th Annu. Conf. IEEE Ind. Electron. Soc., Toronto, ON, Canada, 2021, pp. 1–5, doi: 10.1109/IECON48115.2021.9589577.
- [8] Y. Wang, L. Huang, J. Li, and T. Sun, "Research on face detection based on lightweight MTCNN," in Proc. 11th Joint Int. Conf. Inf. Technol. Artif. Intell. Conf. (ITAIC), Chongqing, China, 2023, pp. 345–348, doi: 10.1109/ITAIC58329.2023.10408812.
- [9] L. Cuimei, Q. Zhiliang, J. Nan, and W. Jianhua, "Human face detection algorithm via Haar cascade classifier combined with three additional classifiers," in Proc. 13th IEEE Int. Conf. Electron. Meas. Instrum. (ICEMI), Yangzhou, China, 2017, pp. 483–487, doi: 10.1109/ICEMI.2017.8265863.
- [10] X. Huang and X. Cao, "Face detection and tracking using Raspberry Pi based on Haar cascade classifier," in Proc. 37th Youth Acad. Annu. Conf. Chinese Assoc. Automat. (YAC), Beijing, China, 2022, pp. 505–509, doi: 10.1109/YAC57282.2022.10023612.

- [11] Z. Zhu and R. Jiao, "Real-time facial expression recognition research based on BlazeFace face detection and ResNet emotion classification," in Proc. 5th Int. Conf. Comput. Vis., Image Deep Learn. (CVIDL), Zhuhai, China, 2024, pp. 401–408, doi: 10.1109/CVIDL62147.2024.10603598.
- [12] P. Asha, A. Vipulendiran, A. Kumaravelu, J. Refonaa, S. L. Jany Shabu, and L. K. Joshila Grace, "Emotion detection by employing deep learning CNN model," in Proc. 2nd Int. Conf. Data Sci. Inf. Syst. (ICDSIS), Hassan, India, 2024, pp. 1–6, doi: 10.1109/ICDSIS61070.2024.10594468.
- [13] A. S. Negi, A. Arora, S. Bisht, S. Devliyal, B. V. Kumar, and G. Kaur, "Facial emotion detection using CNN & VGG16 model," in Proc. 9th Int. Conf. Convergence Technol. (I2CT), Pune, India, 2024, pp. 1–6, doi: 10.1109/I2CT61223.2024.10543618.
- [14] P. Utami, R. Hartanto, and I. Soesanti, "The EfficientNet performance for facial expressions recognition," in Proc. 5th Int. Semin. Res. Inf. Technol. Intell. Syst. (ISRITI), Yogyakarta, Indonesia, 2022, pp. 756–762, doi: 10.1109/ISRITI56927.2022.10053007.
- [15] A. Nouisser, R. Zouari, and M. Kherallah, "Enhanced MobileNet and transfer learning for facial emotion recognition," in Proc. Int. Arab Conf. Inf. Technol. (ACIT), Abu Dhabi, United Arab Emirates, 2022, pp. 1–5, doi: 10.1109/ACIT57182.2022.9994192.
- [16] A. V, A. S. Bharadwaj, C. C. Bagan, D. K, and S. G, "Eye-Move: An eye gaze typing application with OpenCV and Dlib library," in Proc. Int. Conf. Autom., Comput. Renew. Syst. (ICACRS), Pudukkottai, India, 2022, pp. 952–957, doi: 10.1109/ICACRS55517.2022.10029276.
- [17] P. Ranjana, K. Ramesh, S. J. S, and M. B, "Face mask detection using single shot multibox detector and MobileNet," in Proc. 2nd Int. Conf. Adv. Technol. Intell. Control, Environ., Comput. Commun. Eng. (ICATIECE), Bangalore, India, 2022, pp. 1–4, doi: 10.1109/ICATIECE56365.2022.10047292.
- [18] K. Jayanthi, D. Chitradevi, N. Saranya, and S. Anbukkarasi, "Group-scanning by face identification and real-time emotion detection using Faster R-CNN," in Proc. 12th Int. Conf. Adv. Comput. (ICoAC), Chennai, India, 2023, pp. 1–6, doi: 10.1109/ICoAC59537.2023.10249758.
- [19] W. Hua and Q. Tong, "Research on face expression detection based on improved Faster R-CNN," in Proc. IEEE Int. Conf. Artif. Intell. Comput. Appl. (ICAICA), Dalian, China, 2020, pp. 1189–1193, doi: 10.1109/ICAICA50127.2020.9182525.
- [20] N. Yamsani, M. B. Jabar, M. M. Adnan, A. H. A. Hussein, and S. Chakraborty, "Facial emotional recognition using Faster R-CNN with VGG16 feature extraction," in Proc. 3rd Int. Conf. Mobile Netw. Wireless Commun. (ICMNWC), Tumkur, India, 2023, pp. 1–6, doi: 10.1109/ICMNWC60182.2023.10435819.

- [21] D. Gandhi, K. Shah, and M. Chandane, "Dynamic sign language recognition and emotion detection using MediaPipe and deep learning," in Proc. 13th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT), Kharagpur, India, 2022, pp. 1–7, doi: 10.1109/ICCCNT54827.2022.9984592.
- [22] S. V. Vasantha, B. Kiranmai, M. A. Hussain, S. S. Hashmi, L. Nelson, and S. Hariharan, "Face and object detection algorithms for people counting applications," in Proc. 2nd Int. Conf. Autom., Comput. Renew. Syst. (ICACRS), Pudukkottai, India, 2023, pp.
- [23] Y. Sun, L. Dong, S. Huang, S. Ma, Y. Xia, J. Xue, J. Wang, and F. Wei, "Retentive Network: A successor to Transformer for large language models," arXiv preprint arXiv:2307.08621, Jul. 2023. [Online]. Available: <https://arxiv.org/abs/2307.08621>. [Accessed: Jul. 17, 2025].
- [24] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05), San Diego, CA, USA, Jun. 2005, pp. 886–893. doi:10.1109/CVPR.2005.177
- [25] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, Nov. 2004. doi:10.1023/B:VISI.0000029664.99615.94
- [26] D. G. Lowe, "Object Recognition from Local Scale-Invariant Features," in Proc. 7th IEEE Int. Conf. Computer Vision (ICCV), Kerkyra (Corfu), Greece, Sep. 1999, vol. 2, pp. 1150–1157.
- [27] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions," in Proc. British Machine Vision Conference (BMVC), Cardiff, Wales, UK, 2002, pp. 384–393.
- [28] A. Vedaldi and B. Fulkerson, "VLFeat – An Open and Portable Library of Computer Vision Algorithms," in Proc. 18th ACM Int. Conf. Multimedia (ACM MM '10), Firenze, Italy, Oct. 2010, pp. 1469–1472. doi:10.1145/1873951.1874249
- [29] P. C. Mahalanobis, "On the Generalized Distance in Statistics," Proc. Nat. Inst. Sci. India, vol. 2, no. 1, pp. 49–55, Jan. 1936.
- [30] Krislock, N., & Wolkowicz, H. (2012). Euclidean distance matrices and applications. In Handbook on semidefinite, conic and polynomial optimization (pp. 879-914). Springer, New York, NY.
- [31] R. W. Hamming, "Error Detecting and Error Correcting Codes," Bell Syst. Tech. J., vol. 29, no. 2, pp. 147–160, Apr. 1950.
- [32] P. Jaccard, "Étude comparative de la distribution florale dans une portion des Alpes et des Jura," Bull. Soc. Vaudoise Sci. Nat., vol. 37, pp. 547–579, 1901.

- [33] H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 26, no. 1, pp. 43–49, Feb. 1978. doi:10.1109/TASSP.1978.1163055
- [34] A. Bhattacharyya, "On a Measure of Divergence between Two Statistical Populations Defined by Their Probability Distributions," *Bull. Calcutta Math. Soc.*, vol. 35, pp. 99–109, Mar. 1943.
- [35] Y. Rubner, C. Tomasi, and L. J. Guibas, "The Earth Mover's Distance as a Metric for Image Retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, Nov. 2000. doi:10.1023/A:1026543900054
- [36] Y. Zhang, Y. Liu, and X. Zhang, "Emotion-Anchored Contrastive Learning Framework for Emotion Recognition in Conversation," *arXiv preprint arXiv:2403.20289*, 2024.
- [37] S. Qin, H. Shao, Z. Wang, K. Shi, C. Gao, and J. Zhang, "Efficient cosine similarity-based image correlation algorithm for object detection and localization," in *Proc. SPIE 12317, Optoelectronic Imaging and Multimedia Technology IX*, 2023, pp. 1231711.

ÖZGEÇMİŞ

Veli Demir, 2017 yılında Kocaeli Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü'nden mezun olmuştur. Sekiz yılı aşkın profesyonel yazılım mühendisliği deneyimine sahip olan Veli Demir, halen İstanbul, Türkiye'de Concentrix firmasında Yazılım Yöneticisi olarak görev yapmakta; yazılım geliştirme ekiplerini yönetmekte ve proje teslim süreçlerini koordine etmektedir. Ağustos 2022'den bu yana Gebze Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü'nde Makine Öğrenmesi ve Doğal Dil İşleme alanlarına odaklanarak Yapay Zekâ yüksek lisans programını sürdürmektedir. Kurumsal görevlerinin yanı sıra, müşteri memnuniyeti ve rekabet analizi için yapay zekâ tabanlı dijital ve sosyal medya veri analiz çözümleri geliştiren AISenseUs adlı şirketin kurucusudur. Uzmanlık alanları arasında .NET ve React gibi teknolojilerle full-stack geliştirme, yapay zekâ destekli yazılım çözümleri ve ekip liderliği yer almaktadır. Kariyeri boyunca ekip yönetimi, teknik yeniliklerin hayata geçirilmesi ve yapay zekâ uygulamalarının baştan sona geliştirilmesi konularında güçlü yetkinlikler sergilemiştir.

TEZ METNİNDEN TÜRETİLEN YAYINLAR/SUNUMLAR

[1] Demir V., Genç Y., (2025), “Combination of Retentive Networks and Visions Transformers for Facial Emotion Recognition in Image and Video”, International Conference on Interdisciplinary Innovations: Arts, Education,Business and Technology (ICAEBT-2025), Baku, Azerbaijan, 19-20 August.

