

KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

ALTIN ORANA DAYALI İYİLEŞTİRİLMİŞ KAPSÜL AĞLARININ
FÜZYONU İLE DERİN SAHTE VIDEO TESPİTİ

YÜKSEK LİSANS TEZİ

Samet DİNÇER

AĞUSTOS 2024

TRABZON



KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

ALTIN ORANA DAYALI İYİLEŞTİRİLMİŞ KAPSÜL AĞLARININ
FÜZYONU İLE DERİN SAHTE VIDEO TESPİTİ

Samet DİNÇER

Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünce
"BİLGİSAYAR YÜKSEK MÜHENDİSİ "
Unvanı Verilmesi İçin Kabul Edilen Tezdir.

Tezin Enstitüye Verildiği Tarih : 16 / 07 / 2024

Tezin Savunma Tarihi : 05 / 08 / 2024

Tez Danışmanı : Prof. Dr. Güzin ULUTAŞ

Trabzon 2024

ÖNSÖZ

“Altın Orana Dayalı İyileştirilmiş Kapsül Ağlarının Füzyonu ile Derin Sahte Video Tespiti” isimli bu tez Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, Yüksek Lisans Programı’nda hazırlanmıştır.

Tez çalışmam boyunca hiçbir desteğini esirgemeyen ve bilim insanı olma yolundaki ilk adımlarımı atarken kıymetli bilgileriyle ve tecrübesiyle yolumu aydınlatan danışman hocam Sayın Prof. Dr. Güzin ULUTAŞ’a teşekkürü bir borç bilirim.

Tez çalışması sürecinde bana desteklerini esirgemeyen tüm arkadaşlarıma teşekkürlerimi sunarım.

Son olarak, doğduğum günden beri elimi hiç bırakmayan, hayatım boyunca aldığım tüm kararlarda bilgi ve tecrübelerinden faydalandığım ve her kararında arkamda olan, maddi ve manevi desteklerini hiçbir zaman esirgemeyen ve bugünlere gelmemde en büyük paya sahip olan annem Gönül DİNÇER, babam Hasan DİNÇER ve kardeşim Ekrem Sami DİNÇER’e teşekkür eder, minnettarlığımı sunarım.

Bu tez çalışmasının bundan sonraki çalışmalara katkı sağlamasını temenni ederim.

Samet DİNÇER

Trabzon 2024

TEZ ETİK BEYANNAMESİ

Yüksek Lisans Tezi olarak sunduğum “Altın Orana Dayalı İyileştirilmiş Kapsül Ağlarının Füzyonu İle Derin Sahte Video Tespiti” başlıklı bu çalışmayı baştan sona kadar danışmanım Prof. Dr. Güzin ULUTAŞ’ın sorumluluğunda tamamladığımı, verileri/örnekleri kendim topladığımı, deneyleri/analizleri ilgili laboratuvarlarda yaptığımı/yaptırdığımı, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecinde bilimsel araştırma ve etik kurallara uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim. 05/08/2024

Samet DİNÇER

İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ.....	III
TEZ ETİK BEYANNAMESİ.....	IV
İÇİNDEKİLER.....	V
ÖZET	VIII
ŞEKİLLER DİZİNİ	X
TABLolar DİZİNİ.....	XII
SEMBOLLER VE KISALTMALAR DİZİNİ	XIII
1. GENEL BİLGİLER	1
1.1. Giriş.....	1
1.2. Sahte Yüz Görüntüsü Üretim Türleri	4
1.2.1. Tüm Yüz Sentezi.....	4
1.2.2. Kimlik Değiştirme.....	5
1.2.3. Yüz Geçişi.....	6
1.2.4. Özellik Manipülasyonu	7
1.2.5. İfade Değişimi.....	7
1.3. Derin Sahte Video Üretimi.....	8
1.4. Derin Sahte Video Tespit Yöntemleri	11
1.4.1. Kamera Parmak İzi Tabanlı Yöntemler	12
1.4.2. Görsel Olgulara Dayalı Yöntemler	13
1.4.3. Zamansal Tutarlılık Tabanlı Yöntemler.....	14
1.4.4. Biyolojik Sinyallere Dayalı Yöntemler.....	16
1.4.5. Genel Ağ Tabanlı Yöntemler.....	18
1.5. Derin Öğrenme Mimarileri.....	21
1.5.1. VGG19.....	23
1.5.2. Efficient-Net B0.....	25
1.5.3. Efficient-Net B4.....	26
1.5.4. Kapsül Ağları	27
1.6. Yüz Tespit Algoritmaları.....	31
1.6.1. MTCNN	32
1.6.2. S3FD	32

2. YAPILAN ÇALIŞMALAR.....	34
2.1. Yüzdeki Altın Oran Bilgisi Kullanılarak Çerçeve Seçimi	35
2.1.1. Yüz Tespiti ve Hizalama.....	37
2.1.2. En Düz Bakan Yüzlerin Belirlenmesi	38
2.1.3. Yüzdeki Altın Oran Bilgisinin Kullanımı	41
2.2. Öğrenme Aktarımı Yoluyla Özellik Çıkarıcıların Kullanılması	42
2.2.1. VGG19 Ağının Özellik Çıkarıcı Olarak Kullanımı	43
2.2.2. Efficient-Net B0 Ağının Özellik Çıkarıcı Olarak Kullanımı	44
2.2.3. Efficient-Net B4 Ağının Özellik Çıkarıcı Olarak Kullanımı	46
2.3. Kapsül Ağ Tabanlı Sınıflandırma Ağları Kullanılarak Sınıflandırma Yapılması.....	47
2.3.1. Capsule Forensics Mimarisinin Kullanımı	48
2.3.2. ArCapsNet Mimarisinin Kullanımı.....	49
2.4. Önerilen Ağ Mimarilerinin Eğitimi.....	50
2.5. Ağ Mimarilerinin Çıktılarının Füzyonunun Yapılması	53
2.6. Kapsül Ağlarının Ses Sahteciliği Tespitinde Kullanımı.....	54
3. BULGULAR VE TARTIŞMA	57
3.1. Kullanılan Veri Kümeleri.....	57
3.2. Kullanılan Performans Değerlendirme Ölçütleri	59
3.3. Uygulanan Yöntemlerin Değerlendirilmesi ve Elde Edilen DeneySEL Sonuçlar.....	61
3.3.1. Capsule Forensics Ağ Kullanımının Analizi ve Elde Edilen DeneySEL Sonuçlar.....	62
3.3.2. ArCapsNet Ağ Kullanımının Analizi ve Elde Edilen DeneySEL Sonuçlar	65
3.3.3. Farklı Özellik Çıkarıcıların DeneySEL Sonuçlar Üzerindeki Etkisinin Değerlendirilmesi	69
3.3.4. Füzyon Yapılmasının Sonuçlar Üzerindeki Etkisinin Değerlendirilmesi.....	70
3.3.5. Farklı Füzyon Yöntemlerinin Etkilerinin Değerlendirilmesi.....	76
3.3.6. Farklı Aktivasyon Fonksiyonlarının Kullanımının DeneySEL Sonuçlar Üzerindeki Etkisinin Değerlendirilmesi	77
3.3.7. Veri Artırma Tekniklerinin DeneySEL Sonuçlar Üzerindeki Etkisinin Değerlendirilmesi	79
3.4. Önerilen Yöntemin Literatürdeki Diğer Yöntemlerle Karşılaştırılması.....	81
3.5. Elde Edilen DeneySEL Sonuçların İstatistiksel Yöntemler Kullanılarak Değerlendirilmesi	85

3.6. Ses Birleřtirme Sahtecilięi Tespitinde Elde Edilen Sonular	87
3.7. Kopyala-Yapıřtır Ses Sahtecilięi Tespitinde Elde Edilen Sonular.....	88
4. SONULAR.....	91
5. NERİLER.....	93
6. KAYNAKLAR	95
ZGEMİř	



Yüksek Lisans Tezi

ÖZET

ALTIN ORANA DAYALI İYİLEŞTİRİLMİŞ KAPSÜL AĞLARININ FÜZYONU İLE DERİN SAHTE VIDEO TESPİTİ

Samet DİNÇER

Karadeniz Teknik Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı
Danışman: Prof. Dr. Güzin ULUTAŞ
2024, 100 Sayfa

Derin sahte videolar, gelişmiş yapay zekâ teknikleri kullanılarak üretilen sahte dijital içeriklerdir ve son yıllarda büyük bir tehdit haline gelmiştir. Bu videoların tespiti, toplumsal güvenliği ve kişisel mahremiyeti koruma açısından kritik bir önem taşımaktadır. Derin sahte video tespiti sürecinde, yüzdeki altın oran bilgisi kullanılarak çerçeve seçimi gerçekleştirilmiştir. Altın oran, yüzün estetik ve simetrik görünmesini sağlayan oranlar kullanılarak hesaplanır ve en uygun çerçeveler bu kritere göre seçilir. Bu yöntem, tespit sürecinin genel performansını artırmakta ve işlem süresini optimize etmektedir. Çalışmada, derin sahte videoların tespiti için özellik çıkarıcıların kullanımı ve kapsül ağı tabanlı sınıflandırma ağlarının kullanımı ve bunların füzyonu aşamaları uygulanmıştır. İlk olarak, VGG19, Efficient-Net B0 ve Efficient-Net B4 gibi derin ağ modelleri, video karelerinden anlamlı özelliklerin çıkarılmasında kullanılmıştır. Özellik çıkarma işlemi sonrası, bu özellikler kapsül ağı tabanlı sınıflandırıcı ağlara giriş olarak verilmiştir. Capsule Forensics ve ArCapsNet gibi kapsül ağı mimarileri kullanılmıştır. Deneysel sonuçlar, önerilen yöntemin mevcut derin sahte video tespit yöntemlerine kıyasla daha yüksek doğruluk ve güvenilirlik sağladığını göstermektedir. Çalışma, derin sahte video tespitinde yeni ve etkili bir yaklaşım sunarak, bu alandaki literatüre önemli bir katkı sağlamaktadır.

Anahtar Kelimeler: Derin sahte video, altın oran, kapsül ağları

Master Thesis

SUMMARY

DEEP FAKE VIDEO DETECTION BY FUSION OF OPTIMIZED CAPSULE NETWORKS BASED ON GOLDEN RATIO

Samet DİNÇER

Karadeniz Technical University
The Graduate School of Natural and Applied Sciences
Computer Engineering Graduate Program
Supervisor: Prof. Dr. Güzin ULUTAŞ
2024, 100 Pages

Deepfake videos are fake digital content produced using advanced artificial intelligence techniques and have become a major threat in recent years. Detecting these videos is critical for protecting public safety and personal privacy. In the deepfake video detection process, frame selection is performed using the golden ratio information on the face. The golden ratio is calculated using ratios that ensure the face appears aesthetically pleasing and symmetrical, and the most suitable frames are selected according to this criterion. This method improves the overall performance of the detection process and optimizes processing time. In this study, the stages of using feature extractors, capsule network-based classification networks, and their fusion were applied for deepfake video detection. First, deep network models such as VGG19, Efficient-Net B0, and Efficient-Net B4 were used to extract meaningful features from video frames. After the feature extraction process, these features were input into capsule network-based classifier networks. Capsule network architectures such as Capsule Forensics and ArCapsNet were used. Experimental results show that the proposed method provides higher accuracy and reliability compared to existing deepfake video detection methods. The study offers a new and effective approach to deepfake video detection, making a significant contribution to the literature in this field.

Key Words: Deepfake video, golden rate, capsule networks

ŞEKİLLER DİZİNİ

	<u>Sayfa No</u>
Şekil 1. Yıllara göre yayın sayıları.	3
Şekil 2. ÇÜA tabanlı tekniklerle üretilmiş sahte yüz örnekleri.	5
Şekil 3. Kimlik değiştirme türüne örnekler.....	5
Şekil 4. Kişi 1 ve Kişi 2'nin görüntüleri ile yeni yüzün oluşturulması.....	6
Şekil 5. Yüzdeki özelliklerin değişimine örnekler.....	7
Şekil 6. İfade değişimi yöntemi ile üretilen sahte yüzler.....	8
Şekil 7. OK çalışma prensibi	10
Şekil 8. ÇÜA çalışma prensibi.....	10
Şekil 9. Derin sahte video tespit yöntemleri	12
Şekil 10. Oluşturulan iç yüzün mevcut arka plan ile oluşturduğu uyumsuzluklar	13
Şekil 11. Genel derin öğrenme mimarisi	22
Şekil 12. VGG19 ağı mimarisi.....	24
Şekil 13. Efficient-Net B0 ağı mimarisi	26
Şekil 14. Efficient-Net B4 ağı mimarisi	27
Şekil 15. Capsule Forensics model mimarisi	29
Şekil 16. ArCapsNet model mimarisi	30
Şekil 17. Önerilen yöntemin temel aşamaları	35
Şekil 18. Solda S3FD ile tespit edilen yeşil dikdörtgen kutular içine alınmış yüzler, sağda MTCNN ile tespit edilen mavi dikdörtgen kutu içine alınmış yüz.....	37
Şekil 19. Ön işleme adımında yapılan işlemler. MTCNN ile yüzün tespit edilmesi ve 256x256 piksel boyutunda görüntü kırılması. Ardından FAN ile yüzün hizalanması	38
Şekil 20. Solda yan tarafa baktığı için tamamı görünmeyen yüz, sağda tamamı görünen yüz	39
Şekil 21. MTCNN ile göz bebekleri ve yüz kenar sınırlarının tespit edilmesi.....	40
Şekil 22. Yüzdeki altın oran hesabında kullanılan uzaklıklar (Kaya vd., 2019)	41
Şekil 23. Celeb-DF veri kümesinden sahte videolar için seçilen en düşük altın orana sahip görüntü kareleri	42
Şekil 24. VGG19 ağıнын özellik çıkarıcı olarak kullanılan katmanları.....	43
Şekil 25. Efficient-Net B0 ağıнын özellik çıkarıcı olarak kullanılan katmanları.....	45
Şekil 26. Efficient-Net B4 ağıнын özellik çıkarıcı olarak kullanılan katmanları.....	46

Şekil 27.	Eğitim ve doğrulama aşamalarında kayıp ve doğruluk grafikleri	52
Şekil 28.	Celeb-DF veri kümesinde en iyi performans gösteren üç modele ait ROC eğrisi	68
Şekil 29.	Celeb-DF veri kümesinde en iyi performans gösteren üç modele ait ROC eğrisi	68
Şekil 30.	Celeb-DF veri kümesinde en iyi performans gösteren üç ikili füzyon modeline ait ROC eğrisi	73
Şekil 31.	Celeb-DF veri kümesinde en iyi performans gösteren üç ikili füzyon modeline ait ROC eğrisi	73
Şekil 32.	Önerilen yöntemin Celeb-DF veri kümesinde literatürdeki diğer çalışmalarla karşılaştırması.....	81
Şekil 33.	Önerilen yöntemin DFDC-P veri kümesinde literatürdeki diğer çalışmalarla karşılaştırması.....	82

TABLolar DİZİNİ

	<u>Sayfa No</u>
Tablo 1. Veri kümelerinin eğitim, doğrulama ve test bölümlerine ayrılışı.....	58
Tablo 2. Capsule Forensics modelinin Celeb-DF veri kümesi üzerinde elde ettiği performans sonuçları	63
Tablo 3. Capsule Forensics modelinin DFDC-P veri kümesi üzerinde elde ettiği performans sonuçları	64
Tablo 4. ArCapsNet modellerinin Celeb-DF veri kümesi üzerinde elde ettiği performans sonuçları	66
Tablo 5. ArCapsNet modellerinin DFDC-P veri kümesi üzerinde elde ettiği performans sonuçları	67
Tablo 6. Celeb-DF veri kümesi için en iyi performans gösteren modellerin füzyon sonuçları.....	71
Tablo 7. DFDC-P veri kümesi için en iyi performans gösteren modellerin füzyon sonuçları.....	72
Tablo 8. Celeb-DF veri kümesi için 3 model füzyonu sonuçları.....	74
Tablo 9. DFDC-P veri kümesi için 3 model füzyonu sonuçları.....	75
Tablo 10. Celeb-DF veri kümesi için en iyi performans gösteren modellerin farklı füzyon stratejilerinin AUC ölçütü sonuçları	76
Tablo 11. DFDC-P veri kümesi için en iyi performans gösteren modellerin farklı füzyon stratejilerinin AUC ölçütü sonuçları.....	77
Tablo 12. Celeb-DF Veri Kümesindeki En İyi Performans Gösteren Modelin Farklı Aktivasyon Fonksiyonlarıyla Sonuçları.....	78
Tablo 13. DFDC-P Veri Kümesindeki En İyi Performans Gösteren Modelin Farklı Aktivasyon Fonksiyonlarıyla Sonuçları.....	79
Tablo 14. Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4 modeli üzerinde farklı artırma veri kümeleriyle eğitim yapıldığında Celeb-DF veri kümesi üzerindeki sonuçlar..	80
Tablo 15. Celeb-DF veri kümesinde karşılaştırmalı F1 skoru	83
Tablo 16. Celeb-DF veri kümesinde karşılaştırmalı karmaşıklık analizi.....	84
Tablo 17. Wilcoxon işaretli sıralar testi	86
Tablo 18. Friedman testi	86
Tablo 19. Ses birleştirme veri kümesi üzerinde elde edilen performans sonuçları.....	88
Tablo 20. Kopyala-yapıştır veri kümesi üzerinde elde edilen performans sonuçları.....	89

SEMBOLLER VE KISALTMALAR DİZİNİ

OK	: Otomatik Kodlayıcılar
ÇÜA	: Çekişmeli Üretici Ağlar
FTD	: Foto Tepki Düzensizliği
RNN	: Tekrarlayan Sinir Ağları
LSTM	: Uzun Kısa Süreli Bellek
PPG	: Fotopletismogram
DSA	: Derin Sinir Ağları
YSA	: Yapay Sinir Ağları
CNN	: Konvolüsyonel Sinir Ağları

1. GENEL BİLGİLER

Bu bölümde, derin sahte video üretimi ve tespiti konusundaki temel kavramlar ve yöntemler ele alınacaktır. İlk olarak, sahte yüz görüntüsü üretim türleri ve bu türlerin derin sahte videoların üretimindeki rolü üzerinde durulacaktır. Tüm yüz sentezi, kimlik değiştirme, yüz geçişi, özellik manipülasyonu ve ifade değişimi gibi çeşitli yöntemler detaylandırılacaktır. Bu yöntemlerin her biri, sahte yüz görüntülerinin nasıl oluşturulduğunu ve hangi teknolojilerin kullanıldığını açıklayacaktır. Ardından, derin sahte videoların üretim süreci incelenecektir. Bu süreçte, oto kodlayıcılar (OK) ve çekişmeli üretici ağlar (ÇÜA) gibi ileri düzey derin öğrenme tekniklerinin kullanımı ve bu tekniklerin sahte içerik üretimindeki etkinliği üzerinde durulacaktır. Derin sahte videoların üretimi, gerçekçi ve yanıltıcı içeriklerin nasıl oluşturulduğunu anlamak için önemli bir temel sağlayacaktır.

Sonrasında, derin sahte video tespit yöntemleri ele alınacaktır. Kamera parmak izi tabanlı yöntemler, görsel olgulara dayalı yöntemler, zamansal tutarlılık tabanlı yöntemler, biyolojik sinyallere dayalı yöntemler ve genel ağ tabanlı yöntemler gibi çeşitli tespit yaklaşımları detaylandırılacaktır. Bu yöntemler, sahte içeriklerin tespitinde kullanılan teknikleri ve bu tekniklerin etkinliğini inceleyecektir.

Bunun yanı sıra, bu tez çalışması kapsamında kullanılan derin öğrenme ağ mimarileri hakkında detaylı bir bakış sunulacaktır. Tez çalışmasında özellik çıkarıcı ve sınıflandırıcı olarak kullanılan derin ağ mimarileri hakkında ayrı ayrı bilgi verilecektir.

Son olarak, tez çalışması kapsamında videolarda yer alan yüzlerin tespiti için kullanılan yüz tespit algoritmaları detaylı şekilde incelenecektir.

1.1. Giriş

İnternetin hayatımızda büyük yer kapladığı ve kullanımının hızla yaygınlaştığı son yıllarda, çoklu ortam dosyalarının (fotoğraflar, videolar ve ses kayıtları) birçok alanda kullanılabilir hale gelmesi, bu tür içeriklerin doğrulanmasının önemini artırmıştır. Günümüzde dijital içerikler, sosyal medya, haber platformları, eğlence sektörü ve kişisel iletişim gibi pek çok alanda yoğun bir şekilde kullanılmaktadır. Bu nedenle, çoklu ortam

dosyalarının doğrulanması ve bu içeriklerin gerçekliğinin tespiti giderek daha kritik bir hale gelmektedir.

Çoklu ortam dosyalarının doğrulanması için çeşitli yaklaşımlar geliştirilmiştir. Bu yaklaşımlar arasında sayısal damgalama (digital watermarking) ve adli bilim analizi (forensic analysis) öne çıkmaktadır. Sayısal damgalama, dijital içeriklere gizli bilgilerin eklenmesi yoluyla içeriklerin doğruluğunun sağlanmasını amaçlamaktadır. Bu yöntem, ek bilgiyi görsele veya videoya saklayarak doğrulama yapmayı gerektirdiği için bazı zorluklar taşımaktadır. Örneğin, sayısal damgalama yöntemleri, içeriklerin orijinal haline müdahale edebilir ve bu da bazı uygulamalar için istenmeyen bir durum olabilir. Öte yandan, adli bilim analizi yaklaşımları, dijital içeriklerin gerçekliğinin tespiti konusunda daha etkin çözümler sunmaktadır. Bu yaklaşımlar, içeriklerin doğal özelliklerini ve dijital izlerini inceleyerek sahte olup olmadıklarını belirlemeye çalışır. Adli analiz yöntemleri, içeriklerin oluşturulma sürecindeki izleri ve anormallikleri tespit ederek sahte içeriklerin ortaya çıkarılmasında önemli bir rol oynar.

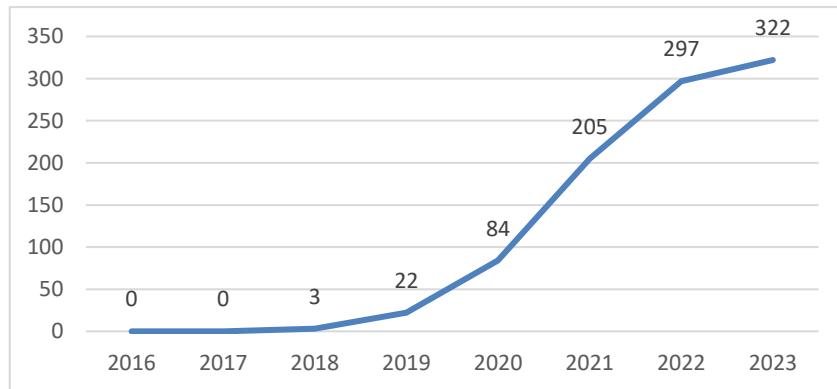
Son yıllarda, özellikle derin öğrenme algoritmalarının gelişimi ile birlikte derin sahte (deepfake) video, ses ve görüntü üretimi büyük bir hız ve gerçekçilik kazanmıştır. Derin öğrenme, büyük veri kümeleri üzerinde eğitim alan yapay sinir ağlarının kullanılmasıyla gerçekleşen bir makine öğrenmesi türüdür. Bu algoritmalar, karmaşık verilerden anlamlı desenler ve ilişkiler çıkararak tahminler ve sınıflandırmalar yapabilir.

Derin sahte içerikler, OK veya ÇÜA gibi ileri düzey derin öğrenme teknikleri kullanılarak üretilmektedir (Kingma & Welling, 2013) (Goodfellow vd., 2020). Oto kodlayıcılar, veri sıkıştırma ve yeniden yapılandırma işlemlerini öğrenen sinir ağlarıdır. Çekişmeli üretici ağlar ise iki ağın (üretici ve ayırt edici ağlar) karşılıklı olarak birbirlerini eğittiği bir yöntemdir. Bu yöntemlerle hedef kişinin yüzü, kaynak videodaki kişinin yüzüyle değiştirilebilmekte ve böylece son derece gerçekçi sahte içerikler oluşturulabilmektedir. Derin öğrenme yöntemleri kullanılarak sahte içerik üretilmesi nedeniyle bu yöntem, “derin öğrenme” ve “sahte” kelimelerinin birleşimi olan “derin sahte” ifadesi ile adlandırılmaktadır.

Derin sahte videoların üretilebilmesi, film endüstrisi ve sanal gerçeklik gibi alanlarda fayda sağlayan çözümler geliştirilmesini sağlamaktadır. Örneğin, film endüstrisinde, oyuncuların yüzlerinin dijital olarak değiştirilmesi, özel efektlerin daha gerçekçi hale getirilmesi ve vefat etmiş ünlü kişiliklerin yeniden canlandırılması gibi uygulamalarda kullanılmaktadır. Sanal gerçeklik alanında ise kullanıcıların sanal ortamlarda gerçekçi

avatarlar oluşturabilmeleri sağlanmaktadır. Ancak, yüz manipülasyonu yapılmış videolar ciddi toplumsal sorunlara da neden olabilmektedir. Örneğin, 2017 yılında deepfakes isimli bir Reddit forumu kullanıcısı, bazı ünlü isimlerin yüzlerini pornografik videolardaki yüzlerle değiştirerek yayınlamıştır. Bu olay, derin sahte teknolojisinin kötüye kullanılma potansiyelini gözler önüne sermiştir. 2019 yılında DeepNude adlı bir akıllı telefon uygulaması, kişilerin fotoğraflarını kıyafetleri çıkarılmış hale getirerek büyük bir tepki çekmiştir. Aynı yıl, ZAO isimli IOS uygulaması ve FaceApp isimli Android uygulaması ile kullanıcılar, yükledikleri fotoğraflardaki yüzleri otomatik olarak başka bir yüzle değiştirme, saç ve sakal ekleme, cinsiyet ve yaş değiştirme gibi çeşitli değişiklikler yapabilmektedir. 2022 yılında Ukrayna-Rusya arasında sürmekte olan savaşla ilgili olarak Ukrayna Devlet Başkanı Zelenskiy'in ülkesinin teslim olduğunu açıkladığı sahte videosunun paylaşılması, derin sahte teknolojisinin siyasi manipülasyon ve kamuoyunu yanıltma potansiyelini göstermiştir. Derin sahte videolar, kişisel mahremiyete zarar vermenin yanı sıra siyasi manipülasyon ve kamuoyunu yönlendirme gibi ciddi zararlar oluşturabilecek bir tehdit haline gelmiştir. Sahte görüntü ve video üretimi, açık kaynak yazılımlar ve çeşitli mobil uygulamalar sayesinde teknik bilgisi olsun ya da olmasın herkesin kolaylıkla yapabileceği bir noktaya gelmiştir. Üretilen sahte görüntü ve videolar oldukça gerçekçi görünmekte ve insanların bu içeriklerin sahte olduğunu anlaması zor olmaktadır.

Bu nedenle, derin sahte içeriklerin tespiti konusundaki çalışmalar son yıllarda büyük bir artış ve ihtiyaç göstermektedir. Şekil 1'de webofscience'da listelenen yayınlarda 'deepfake' terimini konusunda içerenlerin yıllara göre sayıları yer almaktadır. Şekilden görülebildiği gibi derin sahte alanında yayın sayısı yıldan yıla artmaktadır.



Şekil 1. Yıllara göre yayın sayıları.

Araştırmacılar, derin sahte içeriklerin tespiti için yeni yöntemler ve algoritmalar geliştirmekte ve bu tehditlere karşı daha güvenli dijital ortamlar oluşturmayı hedeflemektedir. Derin öğrenme tabanlı tespit yöntemleri, sahte içeriklerin tespitinde büyük bir potansiyele sahip olup, bu alanda yapılan çalışmaların sayısı ve önemi giderek artmaktadır. Derin sahte videoların ve videolarda yer alan yüzlerin üretimi konusuna sonraki başlıklarda detaylı bir şekilde değinilecektir.

1.2. Sahte Yüz Görüntüsü Üretim Türleri

Derin sahte videoların üretiminde kullanılan teknikler, farklı sahte yüz görüntüsü üretim türlerine dayanmaktadır. Bu bölümde, derin sahte videolarda yer alan yüzlerin üretiminde hangi türlerin kullanıldığı detaylı bir şekilde ele alınacaktır. Sahte yüz görüntüleri üretim türlerini beş başlıkta toplayabiliriz: tüm yüz sentezi, kimlik değiştirme, yüz geçişi, özellik manipülasyonu ve ifade değişimi (Rehaan vd., 2024) (Tolosana vd., 2020) (Malik vd., 2022).

1.2.1. Tüm Yüz Sentezi

Tüm yüz sentezi, var olmayan bir yüzün tamamen sıfırdan üretilmesi sürecidir. Genellikle ÇÜA tabanlı Style-GAN, Style-GAN2 ve StyleGAN2-Ada gibi modeller kullanılmaktadır (Karras vd., 2019) (Karras, Laine, vd., 2020) (Karras, Aittala, vd., 2020). ÇÜA'lar, birbirine karşı çalışan iki yapay sinir ağı olan üretici ve ayırt edici ağlardan oluşur. Üretici ağ, rastgele gürültüden gerçekçi görünen yüz görüntüleri üretmeye çalışırken, ayırt edici ağ bu görüntülerin sahte mi yoksa gerçek mi olduğunu ayırt etmeye çalışır. Bu karşılıklı öğrenme süreci, çok gerçekçi yüz görüntüleri oluşturulmasını sağlar. Üretilen görüntüler oldukça gerçekçidir ve insan gözüyle sahte olduğu anlaşılamayacak kalitededir. Ancak, bu yöntemde üretilen sahte görüntüler, gerçek görüntülerde yer alan cihaz parmak izine benzer bir ÇÜA parmak izi içermektedir. Tüm yüz sentezi yöntemi, video oyunları ve 3 boyutlu modelleme gibi çeşitli sektörlerde faydalı kullanımlar sağlarken, gerçekçi profil fotoğraflarıyla sahte online hesapların oluşturulması gibi sorunlara da neden olabilmektedir. Tüm yüz sentezi yönteminde üretilmiş örnek sahte yüz görüntüleri Şekil 2'de verilmiştir.

Sahte



Şekil 2. ÇÜA tabanlı tekniklerle üretilmiş sahte yüz örnekleri.

Şekilde yer alan yüz görüntüleri <https://thispersondoesnotexist.com> adresli web sitesinden alınmıştır. Örneklerden görülebildiği gibi üretilmiş sahte yüzler oldukça gerçekçidir ve sahte olduğunu anlayabilmek neredeyse imkansızdır.

1.2.2. Kimlik Değiştirme

Kimlik değiştirme, bir videodaki kişinin yüzünün başka bir kişinin yüzüyle değiştirilmesidir. Bu teknik, son zamanlarda büyük toplumsal endişelere neden olan örnekleri nedeniyle en popüler yüz manipülasyon alanlarından biridir. Bu yöntemle üretilen görüntülere örnekler Şekil 3'te verilmektedir.



Şekil 3. Kimlik değiştirme türüne örnekler.

Şekilde gösterildiği gibi kaynak kişinin yüz görüntüsü, hedef kişinin yüzüyle değiştirilir ve sahte yüz üretilir. Bilgisayar grafikleri tabanlı ve derin öğrenme tabanlı olmak üzere iki farklı yaklaşım kullanılır. Bilgisayar grafikleri tabanlı teknikler, genellikle karmaşık 3D modelleme ve animasyon araçları kullanarak yüz değişimini gerçekleştirir. Derin öğrenme tabanlı teknikler ise, ÇÜA ve diğer derin sinir ağı modelleri kullanarak yüz değişimini daha otomatik ve hızlı bir şekilde gerçekleştirir. Bu tekniğin kullanımı, özellikle film endüstrisinde faydalı olarak kullanılmaktadır; oyuncuların yüzlerinin başka karakterlerle değiştirilmesi veya dijital ikizlerin yaratılması gibi. Ancak, kimlik değiştirme yöntemi kötü amaçlı olarak sahte videoların üretiminde de kullanılmaktadır.

1.2.3. Yüz Geçişi

Yüz geçişi, iki veya daha fazla kişinin yüz görüntüsündeki biyometrik özellikler kullanılarak yeni bir yapay biyometrik yüz oluşturulması işlemidir. Bu yöntem, kaynak yüzlerden alınan özelliklerin bir kombinasyonunu içerdiğinden, yüz tanıma sistemleri için ciddi bir tehdit oluşturmaktadır. Kimlik değişiminden farklı olarak yüz geçişinde, video seviyesinde sahte içerik oluşturmak yerine, görüntü düzeyinde sahte örnekler oluşturulur. Bu yöntemde üretilen sahte görüntüler, yüz tanıma sistemlerinin yanıltılması için kullanılabilir ve bu da güvenlik sistemlerinde büyük zorluklar yaratabilir. Bu yöntemde üretilen sahte görüntüye bir örnek Şekil 4'te verilmektedir.



Şekil 4. Kişi 1 ve Kişi 2'nin görüntüleri ile yeni yüzün oluşturulması.

Şekilden görülebildiği gibi kişi 1 ve kişi 2'ye ait biyometrik özellikler ile yeni bir yüz oluşturulmakta ve bu yeni yüz kendisini oluşturan diğer iki yüzün özelliklerini taşımaktadır.

1.2.4. Özellik Manipülasyonu

Özellik manipülasyonu, saç veya cilt rengi, cinsiyet, yaş, gözlük kullanımı gibi özelliklerin değiştirilmesi tekniğidir. ÇÜA tabanlı teknikler kullanılarak üretilen bu manipülasyonlar, piyasada son kullanıcının kullanımına sunulan birçok ücretli ve ücretsiz uygulama tarafından desteklenmektedir. Uygulamalar genellikle ÇÜA tabanlı StarGAN temellidir (Choi vd., 2018). Bu uygulamalar sayesinde kullanıcılar, herhangi bir teknik bilgiye ihtiyaç duymadan, tek tuşla üzerinde oynama yapılmış görüntüler oluşturabilirler. Örneğin, FaceApp gibi mobil uygulamalar, kullanıcıların yüzlerinde çeşitli değişiklikler yapmalarına olanak tanır. Bu tür uygulamalar, kozmetik sektöründe ürünleri kullanmadan elde edilecek sonuçları görme gibi faydalı kullanımlar sağlayabilirken, aynı zamanda kötü niyetli kullanımlara da zemin hazırlayabilir. Şekil 5'te bu manipülasyon türü için örnekler görülmektedir.



Şekil 5. Yüzdeki özelliklerin değişimine örnekler.

Verilen örneklerdeki sahte görüntüler FaceApp uygulaması ile oluşturulmuştur. Kozmetik sektöründe ürünü kullanmadan elde edilecek sonucu görebilme gibi kullanımlar sağlayabileceği için faydalı olmaktadır. Bunun yanında kötü niyetli kullanımları da olabilmektedir.

1.2.5. İfade Değişimi

İfade değişimi, bir video veya görüntüdeki kişinin yüz ifadesinin değiştirilmesi işlemidir. Bu teknikte, kaynak videodaki kişinin yüz ifadeleri, hedef videodaki kişinin ifadelerine yansıtılır. Şekil 6'da ifade değişimi yöntemi için örnek görseller verilmektedir.



Şekil 6. İfade değişimi yöntemi ile üretilen sahte yüzler

Şekilde görüldüğü gibi kaynak videoda yer alan yüze hedef videoda yer alan yüzün ifadesi yerleştirilmiştir. ÇÜA tabanlı teknikler kullanılarak görüntü düzeyinde sahte içerik üretimi ve Face2Face, NeuralTextures gibi yöntemler kullanılarak video düzeyinde sahte içerik üretimi gerçekleştirilmektedir (He vd., 2019) (Liu vd., 2019) (Thies, Zollhöfer, Stamminger, vd., 2019) (Thies, Zollhöfer, & Nießner, 2019). İfade değişimi, kişinin gerçekte söylemediği bir sözü söylüyor gibi gösterilmesini mümkün kılarak, özellikle siyasi manipülasyon ve yanlış bilgilendirme gibi alanlarda ciddi sorunlara yol açabilir.

1.3. Derin Sahte Video Üretimi

Derin sahte videolar, gelişen yapay zeka teknolojileriyle üretilen sahte dijital içerikler arasında önemli bir yer tutmaktadır. Bu videolar, gerçek görüntülerin veya videoların manipüle edilmesiyle oluşturulan yanıltıcı içeriklerdir. Derin sahte videoların üretimi, özellikle OK ve ÇÜA gibi yapay zeka algoritmalarının kullanımıyla mümkün hale gelmiştir. Bu videolar, gerçek görüntülerin ve seslerin manipüle edilerek oluşturulmuş sanal sahneleri içerir ve gerçeklik hissi yaratmak için dikkatlice hazırlanır. Derin sahte videoların üretimi, son yıllarda artan hesaplama gücü ve gelişen ileri derin öğrenme algoritmalarıyla kolaylaşmıştır.

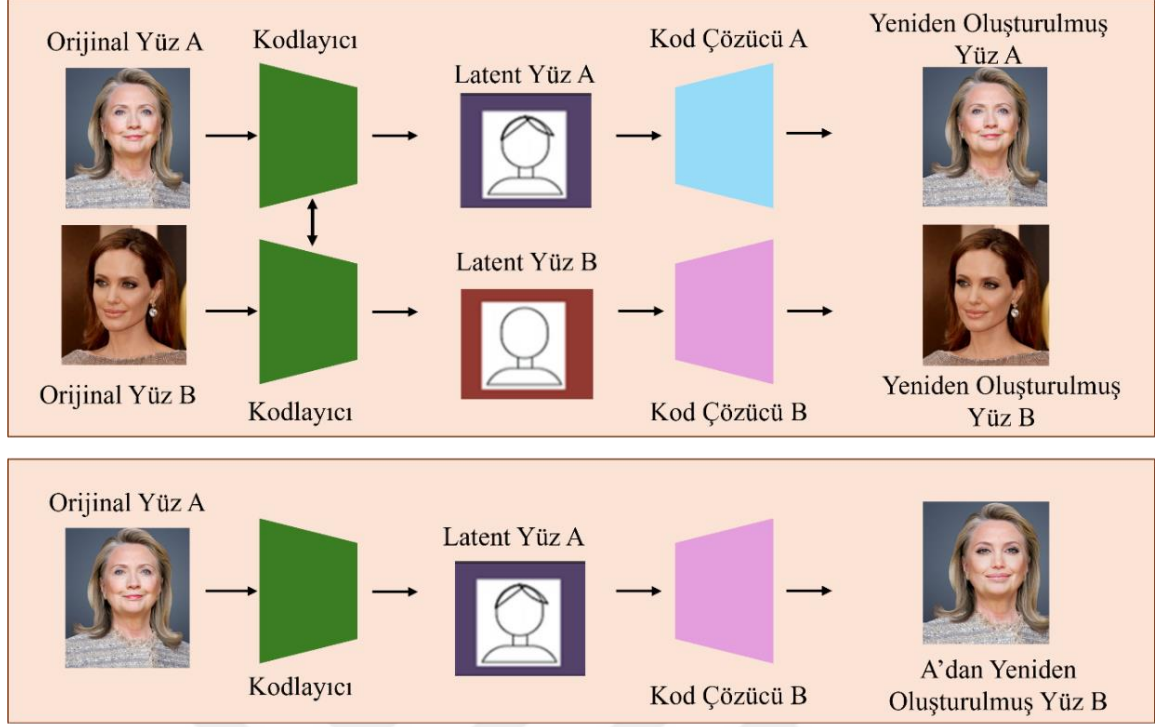
Oto kodlayıcılar (OK), veri sıkıştırma ve kodlama için kullanılan bir tür yapay sinir ağıdır. OK'ler, kodlayıcı (encoder) ve kod çözücü (decoder) olarak adlandırılan iki ana bileşenden oluşur. Şekil 7'de gösterilen OK çalışma prensibi, kodlayıcı kısmında giriş görüntüsünün özelliklerini içeren sıkıştırılmış bir veri oluştururken, kod çözücü kısmında

bu sıkıştırılmış veriden tekrar giriş görüntüsüne benzer bir görüntü üretilebilmesini sağlamaktadır.

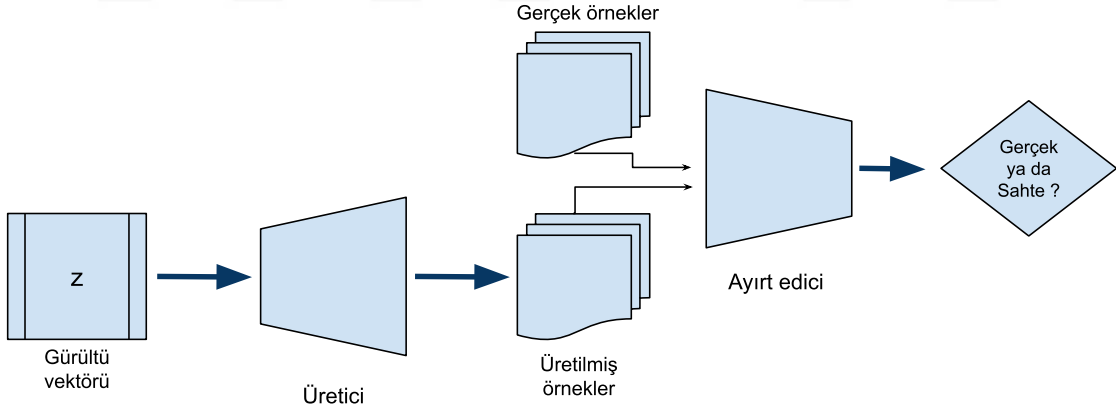
OK'lerin çalışma prensibi şu şekilde özetlenebilir: Kodlayıcı, giriş görüntüsünü alır ve bu görüntüyü daha düşük boyutlu, sıkıştırılmış bir temsile dönüştürür. Bu sıkıştırılmış temsil, genellikle "latent uzay" olarak adlandırılan bir gizli uzaya kodlanır. Latent uzay, görüntüdeki temel özellikleri ve desenleri içerir. Daha sonra, kod çözücü, bu latent temsili alır ve orijinal görüntüyü yeniden oluşturur. Derin sahte video üretiminde, oto kodlayıcılar, hedef kişinin yüz özelliklerini öğrenmek için eğitilir. Bu süreçte, hedef kişinin yüz ifadeleri, hareketleri ve diğer ayırt edici özellikleri latent uzayda kodlanır. Daha sonra, bu kodlanmış bilgiler, başka bir yüz üzerinde yeniden oluşturularak sahte bir görüntü elde edilir. Örneğin, bir kişinin yüzü, OK ile kodlanabilir ve daha sonra bu kodlama üzerinde çeşitli değişiklikler yapılarak sahte bir görüntü oluşturulabilir. Derin sahte video üretiminde oto kodlayıcılar, genellikle hedef kişinin yüz özelliklerini ayrıntılı bir şekilde öğrenmek için geniş veri kümeleri üzerinde eğitilir. Bu veri kümeleri, çeşitli açılardan ve farklı ışık koşullarında çekilmiş yüz görüntülerini içerir. Kodlayıcı, bu görüntülerin özelliklerini öğrenerek, yüzün temel yapı taşlarını latent uzaya kodlar. Kod çözücü ise, bu kodlanmış bilgileri kullanarak, gerçekçi ve inandırıcı yüz görüntüleri oluşturur. Bu süreç, OK'lerin derin sahte video üretiminde etkin bir şekilde kullanılmasını sağlar.

ÇÜA, birbirine karşıt iki ağdan oluşan bir derin öğrenme modelidir: bir üretici (generator) ve bir ayırt edici (discriminator). ÇÜA'ların çalışma prensibi Şekil 8'de gösterilmiştir. Üretici ağ, gerçekçi sahte veriler üretmeye çalışırken, ayırt edici ağ bu verilerin gerçek mi yoksa sahte mi olduğunu saptamaya çalışır. Eğitim süreci boyunca, üretici ağ giderek daha inandırıcı sahte videolar üretmeyi öğrenirken, ayırt edici ağ da sahte ve gerçek veriyi ayırt etme konusunda daha iyi hale gelir.

ÇÜA'ların eğitim süreci iteratif bir yapıya sahiptir. Başlangıçta, üretici ağ rastgele gürültü (noise) kullanarak veri üretir. Bu üretilen veriler, ayırt edici ağa sunulur ve ayırt edici ağ, bu verilerin gerçek mi yoksa sahte mi olduğunu belirlemeye çalışır. Ayırt edici ağın çıktısına göre, üretici ağın parametreleri güncellenir ve daha gerçekçi veriler üretmesi sağlanır. Benzer şekilde, ayırt edici ağ da sahte verileri daha iyi ayırt edebilmek için kendini geliştirir. Bu süreç, her iki ağın da performansını sürekli olarak artırır ve sonuçta son derece gerçekçi sahte videolar üretilir.



Şekil 7. OK çalışma prensibi



Şekil 8. ÇÜA çalışma prensibi

ÇÜA'ların üretim süreci, temel olarak bir oyun teorisi yaklaşımıyla çalışır. Üretici ağ, başlangıçta rastgele gürültü kullanarak görüntüler üretir ve bu görüntüler ayırt edici ağına sunulur. Ayırt edici ağ, bu görüntülerin gerçek olup olmadığını belirlemeye çalışır ve üretici ağın geri bildirimi ile kendini geliştirir. Bu geri bildirim döngüsü, her iki ağın da daha iyi hale gelmesine olanak tanır. Üretici ağ, gerçekçi görüntüler üretmeyi öğrenirken, ayırt edici ağ, sahte görüntüleri daha iyi tespit etmeyi öğrenir.

ÇÜA'lar, özellikle insan yüzleri ve ifadeleri konusunda başarılı sonuçlar verebilir. Üretici ağ, hedef kişinin yüz ifadelerini, mimiklerini ve diğer karakteristik özelliklerini taklit ederek gerçekçi sahte videolar üretir. Ayırt edici ağ ise, bu videoların gerçekliğini değerlendirir ve üretici ağın daha inandırıcı videolar üretmesine yardımcı olur. Bu süreç, derin sahte videoların gerçekçiliğini artırmaktadır.

Bu teknolojilerin yardımıyla üretilen derin sahte videolar hem pozitif hem de negatif yönde etkileyici olabilir. Eğitim, film yapımı, sanat ve terapi gibi alanlarda yenilikçi ve faydalı kullanımları mevcutken, yanlış bilgilendirme, siyasi manipülasyon ve kişisel hakların ihlali gibi ciddi toplumsal riskleri de beraberinde getirmektedir. Derin sahte içeriklerin üretimi ve yayılması, toplumsal güveni sarsma ve bireylerin mahremiyetini ihlal etme potansiyeline sahiptir. Bu nedenle, bu teknolojilerin etik kullanımı ve derin sahte içeriklerin tespiti konusundaki çalışmalar büyük önem taşımaktadır.

1.4. Derin Sahte Video Tespit Yöntemleri

Derin sahte videoların tespiti, günümüzde dijital medya güvenliğinin en önemli meselelerinden biri haline gelmiştir. Yapay zekâ ve makine öğrenimi teknolojilerinde yaşanan büyük gelişmelerle birlikte üretilen sahte videoların kalitesi artmış, gerçekle sahteyi ayırt etmek giderek daha zor bir hale gelmiş ve bu da derin sahte içeriklerin tespiti için yeni ve etkili yöntemlerin geliştirilmesine olan ihtiyacı artırmıştır. Bu bağlamda, araştırmacılar, sahte videoları belirleyebilmek için çeşitli tespit yöntemleri üzerinde çalışmalar yapmaktadırlar. Derin sahte tespit yöntemleri Şekil 9'da gösterildiği gibi kabaca beş ana kategori altında incelenebilir: Kamera parmak izi tabanlı yöntemler, Görsel olgulara dayalı yöntemler, Zamansal tutarlılık tabanlı yöntemler, Biyolojik sinyallere dayalı yöntemler ve Genel ağ tabanlı yöntemler. Her bir tespit yöntemi, derin sahte içerikleri saptamada farklı bir yaklaşım sunar ve bu yöntemler, sahteciliği belirlemede kritik roller oynar. Bu başlık altında, söz konusu beş ana kategoriye detaylı bir bakış sunulacak ve her bir metodun nasıl çalıştığına, avantajlarına ve sınırlılıklarına dair bilgiler verilecektir.



Şekil 9. Derin sahte video tespit yöntemleri

1.4.1. Kamera Parmak İzi Tabanlı Yöntemler

Kamera parmak izi tabanlı tespit yöntemleri, derin sahte video tespiti alanında kullanılan ilk dönem yöntemlerden biridir. Bu yöntemler, kamera sensörlerinin her birinden kaynaklanan benzersiz "Foto Tepki Düzensizliği" (FTD) adı verilen izleri kullanır. FTD, kamera sensörlerinin imalat sürecindeki hafif kusurlar nedeniyle her bir pikselin ışığa duyarlılığının farklı olması sonucunda ortaya çıkar. Bu benzersiz özellikler, videonun hangi cihaz tarafından çekildiğini belirlemek için kullanılabilir. Bu ayrıca sahte içeriklerin tespit edilmesinde de belirleyici özellik olarak kullanılabilir.

Bu yöntemde, genellikle her karedeki yüz bölgeleri belirlenir ve bu bölgelerden FTD izleri çıkarılır. Sonra, bu izler orijinal kaynakla karşılaştırılır. Eğer video düzenlenmişse, FTD izlerindeki tutarsızlıklar bu düzenlemeleri ortaya çıkarabilir. Araştırmalar, bu yöntemin küçük veri kümelerinde oldukça başarılı sonuçlar verdiğini göstermektedir. Ancak, özellikle sahte içeriklerde FTD izlerinin kaldırılması gibi yöntemler geliştirildikçe, bu tür tespit tekniklerinin etkinliği azalabilir. Kamera parmak izi tabanlı yöntemler, aynı zamanda dijital forensik alanında da sıkça kullanılmakta olup, sahteciliğin yanı sıra, içeriğin hangi cihazla çekildiğini ve dolayısıyla kaynağını belirleme konusunda da bilgi sağlar.

Koopman vd. (2018), FTD kullanarak derin sahte videoları tespit etme yöntemi önerdi. Her görüntü karesindeki yüz bölgesini 8 gruba ayırdılar ve her grubun ortalama FTD değerini hesapladılar. Deneysel sonuçlar, gerçek video ile derin sahte video arasında ortalama normalleştirilmiş korelasyon katsayısında önemli bir fark gösterdi. Lugstein vd.

(2021) de FTD kullanan bir diğer araştırmacıdır ve derin sahte video tespitinde başarılı sonuç elde etmişlerdir. Frank vd. (2020), FTD'nin küçük veri kümeleri için derin sahte video tespitinde etkili olduğuna inanıyor. Başlarda olumlu sonuçlar alınmasına rağmen, parmak izinin kaldırılabilirdiği yeni sahte video örnekleri nedeniyle, kamera parmak izine dayalı yöntemler etkinliklerini yitirmiştir.

1.4.2. Görsel Olgulara Dayalı Yöntemler

Görsel olgulara dayalı tespit yöntemleri, derin sahte videoların tespitinde kullanılan yaklaşımlardan bir tanesidir. Bu yöntemler, video veya görüntülerde yapay olarak oluşturulan içeriklerin meydana getirdiği görsel kusurları ve anormallikleri tespit etmeye odaklanır. Özellikle derin öğrenme tekniklerinin gelişmesiyle birlikte, bu tür kusurları belirleyebilmek daha da mümkün hale gelmiştir. Derin sahte videolar oluşturulurken, arka plan ile yüz arasında tutarsızlıklar meydana gelebilir. Bu tutarsızlıklara örnekler Şekil 10'da gösterilmektedir. Görsel olgulara dayalı tespit yöntemleri, bu tutarsızlıkları, özellikle yüz ve arka plan arasındaki birleşme noktalarında, cilt tonu farklılıkları, aydınlatma hataları ve sınır çizgilerindeki düzensizlikler gibi görsel ipuçlarını analiz ederek belirler.



Şekil 10. Oluşturulan iç yüzün mevcut arka plan ile oluşturduğu uyumsuzluklar

Yüzler ve arka plan arasındaki tutarsızlık gözlemlerine dayanarak Li & Lyu (2018) yeni bir derin öğrenme tabanlı yöntem önerdiler. Sahte videoları tespit için harmanlama işlemiyle üretilen yüz çarpıtma olguları kullanıldı. Sentetik yüzler hedef yüzlerin üzerine yerleştirilirken iç yüz ve arka plan arasında renk ve çözünürlük farkları olabilmektedir. Bu farklılıktan yararlanmayı hedefledikleri çalışmada VGG16, ResNet50, ResNet101 ve ResNet152 gibi dört farklı CNN modelini eğiterek tespit yapmaya çalışmışlardır. Çeşitli ilk nesil veri kümeleri üzerinde kendilerinden önceki çalışmalara kıyasla başarılı sonuçlar elde etmişlerdir.

Bu yaklaşımdan farklı olarak Li vd. (2020), giriş görüntüsünün ön plan yüzü ve arka plan olarak ayrıştırılıp ayrıştırılamayacağını gözlemlemek için kullanılan yüz röntgeni adı verilen yeni bir görüntü temsili önerdiler. Manipüle edilen ön plan ile arka plan arasındaki birleşim sınırı, yüz X-ışını olarak tanımlandı. Görüntü birleştirmede bilinen birleştirme sınırını hedef aldı ve böylece çeşitli veri kümelerinde test edildiğinde başarılı performans gösterdi. Genelleştirmede yüz çarpıtma olguları yöntemine göre daha başarılı olduğu görüldü. Bununla birlikte, birleştirme sınırına aşırı odaklanıldığı için bu sınırı gizleyebilen örnekler için başarılı bir yöntem değildir.

Diğer bir yaklaşımda Yang vd. (2019), iki boyutlu yüz işaret noktalarından üç boyutlu baş pozisyonunu tahmin ederek derin sahte videoları tespit eden bir yaklaşım sundular. Baş pozisyonları arasındaki farklar gerçek ve sahte ayrımı yapmak için kullanılan bir SVM sınıflandırıcıyı eğitmek için özellik vektörü olarak kullanılır. Bu yöntem derin sahte tespitinde iyi performans elde edebilmektedir. Bununla beraber görüntü kalitesi düşük, bulanık videolarda yüz işaret noktaları tam olarak doğru konumlandırılmadığı için bu koşullarda performansı düşmektedir.

Görsel olgulara dayalı yöntemler, derin sahte videoların tespitinde önemli bir rol oynamaktadır. Yüz ve arka plan arasındaki tutarsızlıklar, cilt tonu farklılıkları, aydınlatma hataları ve sınır çizgilerindeki düzensizlikler gibi görsel ipuçları, sahte videoları gerçek videolardan ayırt etmek için kullanılabilir. Bu yöntemler, derin öğrenme algoritmaları ile desteklenerek sahtecilik tespitinde kullanılmaktadır.

1.4.3. Zamansal Tutarlılık Tabanlı Yöntemler

Zamansal tutarlılık tabanlı tespit yöntemleri, video içeriklerindeki zaman dizileri boyunca oluşan düzensizlikleri ve anormallikleri yakalayarak derin sahte videoları tespit etmeye odaklanmaktadır. Bu yöntemler, videoların doğal akışındaki kesintileri ve uyumsuzlukları belirleyerek manipüle edilmiş içerikleri ayırt eder. Videolar, bir dizi görüntü karesinden oluştuğu için, her bir karenin bir önceki ve bir sonraki kare ile olan zamansal bağlantısının korunması beklenir. Ancak, derin sahte videoların üretimi esnasında bu kareler arasında tutarsızlıklar meydana gelebilmektedir. Bu tespit yönteminde genellikle, video kareleri arasındaki hareket vektörlerini analiz eden optik akış tekniklerinden yararlanır. Optik akış, bir kareden diğerine nesnelerin nasıl hareket ettiğini

gösteren vektör alanlarıdır. Derin öğrenme modelleri, bu hareket desenlerinin tutarlılığını öğrenir ve anormal hareketleri saptayarak sahte içerikleri belirleyebilir.

Görüntü kareleri arasındaki zamansal sürekliliği göz önünde bulunduran Güera & Delp (2018) derin sahte video tespitinde tekrarlayan sinir ağlarının (RNN) kullanımını önermektedir. Sahte videolardaki yüz görüntüleri kare kare oluşturulur. OK'lar görüntü oluştururken önceki görüntüleri dikkate almaz, bu da önceki ve sonraki kareler arasında bazı alanlarda anormalliğe yol açabilir. Bitişik görüntü kareleri arasındaki sürekliliği kontrol etmek için uçtan uca eğitilebilir tekrarlayan derin sahte video tespit sistemi önerildi. Önerilen sistem görüntü karesi dizilerini işlemek için bir evrimsel uzun kısa süreli bellek (LSTM) yapısından oluşmaktadır. Bu yapıda iki temel bileşen kullanılır. Görüntü karesi özelliği çıkarıcı olarak CNN ve zamansal dizi analizi olarak LSTM kullanılır. Önceden eğitilmiş InceptionV3 ağı her bir görüntü karesinin derin çıktısını almak için uyarlanmıştır. Havuzlama katmanları tarafından çıkarılan 2048 boyutlu özellik vektörleri LSTM girişi olarak verilir. Sahtecilik olasılıklarını hesaplamak için tam bağlantılı katman ve softmax katmanı eklenir. Kendi oluşturdukları veri kümesi üzerinde yapılan deneyler başarılı sonuçlar alınabileceğini göstermektedir.

Zamansal tutarlılığa dayalı tespit yöntemi etkinliğini gösterdikten sonra bu yöntem temelinde birçok çalışma geliştirilmiştir. Guera'dan esinlenerek Sabir vd. (2019) sahte videoları tespit etmek için geçici ağ ve yüz ön işleme tekniklerini birleştirmeyi önermiştir. Bu çalışma yüksek kaliteli videolarda başarı elde etse de sıkıştırma yapılmış videolarda ve düşük kaliteli videolarda iyi performans gösterememektedir. Bu sorunu çözmek için Montserrat vd. (2020) otomatik ağırlıklandırma mekanizmasına sahip bir CNN-RNN ağı önermektedir. Tespit algoritmasının sağlamlığını iyileştirmek için video düzeyinde tespit sırasında yüz görüntülerinin ağırlığı artırılır. Deneyler, CNN ve RNN'nin birleştirilmesinin bazı veri kümelerinde yüksek doğruluk sağladığını göstermiştir. Algoritmaların sağlamlığının yanında genelleme yeteneği de gereklidir. Zhao vd. (2020) bitişik kareler arasındaki yüz ifadelerinin farklılıklarını yakalamak için optik akış kullandı ancak güçlü bir sağlamlık veya genelleme elde edilemedi. Bu sorunu çözmek için Wu vd. (2020) hem düşük seviyeli olgulardan hem de zamansal tutarsızlıklardan yararlanan SSTNet adlı yeni bir manipülasyon tespit sistemi önerdi. Masi vd. (2020) tespit yönteminin geliştirilmesini geliştirmek için iki dallı ağ önermektedir. Bir dal ağı orijinal görüntü özelliği bilgilerini çıkarmak için kullanılır ve diğer dal yüz içeriğini bastırmak için kullanılır. Çeşitli veri kümeleri üzerinde çapraz test yapıldığında iyi performans elde

edilebilmektedir. Zhang vd. (2021), video eylem tanıma için bir dönüştürücü tabanlı model olan Video Transformer (VidTr)'ı tanıttı. İlk olarak uyguladıkları video dönüştürücüsünün yüksek bellek kullanımına sebep olduğunu tespit ettiler. VidTr, bu sorunu çözmek için mekânsal ve zamansal dikkati ayrı ayrı gerçekleştiren ayrılabilir dikkati tanıttı. Çeşitli veri kümelerinde daha düşük hesaplama ile iyi performans elde etti. Zhao vd. (2023), dönüştürücü mimarisinde, mekânsal ve zamansal bilgileri ayrı ayrı ele alan ayrıştırılmış mekânsal–zamansal dikkat mekanizması kullanan bir yöntem olan ISTVT'yi önerdi. Ayrıca, zamansal dikkati çerçeve tutarsızlıklarına odaklamak için bir kendinden çıkarma mekanizması tanıttı. Çeşitli veri kümelerinde yapılan deneysel sonuçlar, ISTVT'nin iyi tespit performansına sahip olduğunu göstermektedir.

1.4.4. Biyolojik Sinyallere Dayalı Yöntemler

Biyolojik sinyallere dayalı yöntemler, derin sahte video tespitinde kullanılan yenilikçi ve etkili yaklaşımlardan biridir. Bu yöntemler, insanların doğal biyolojik tepkilerini ve sinyallerini analiz ederek gerçek ve sahte içerikler arasında ayırım yapmayı amaçlar. Derin sahte videoların üretiminde kullanılan derin öğrenme teknolojileri oldukça gerçekçi görünen görseller üretme kapasitesine sahip olsa da ince biyolojik detayları taklit etmede yeterli başarıya sahip değiller ve bu durum tespit metodunun temelini oluşturur. İnsan yüzü ve davranışları üzerine kurulu bu yöntemler, yüz ifadeleri, göz kırpma hızları, dudak hareketleri ve hatta kalp atış hızı gibi unsurları inceler. Gerçek insan yüzleri ve onların ifadeleri, doğal ve tutarlı bir biyolojik ritme sahiptir. Örneğin, bir kişi konuşurken veya bir duygu ifade ederken, yüz kasları, göz hareketleri ve dudak senkronizasyonu belirli bir doğallık ve akıcılık gösterir. Ancak, derin sahte teknolojisi ile üretilen videolarda bu tür doğal ritimler ve senkronizasyonlar eksik olabilir ya da yapay görünebilir.

Bu tespit yöntemlerinden biri göz kırpma oranları ve sıklığı gibi basit ama güçlü biyolojik göstergelere dayanmaktadır. İnsanlar belirli bir ritimde göz kırparlar, ancak yapay olarak üretilmiş videolarda bu göz kırpma sıklığı doğal olması gerekenden farklı bir şekilde düşük ya da yüksek olabilir. Yapay zekâ destekli analiz araçları, bu tür düzensizlikleri tespit ederek gerçek olmayan içerikleri ayırt edebilir.

Jung vd. (2020) tarafından önerilen bir göz kırpma anomali tespit yöntemi, gerçek ve sahte videolar arasında ayırım yapmak üzere geliştirildi. Göz kırpmayı algılamak için Fast-HyperFace ve EAR tekniklerini birleştirdiler (Ranjan vd., 2017) (Soukupova & Cech,

2016). Cinsiyet, yaş, zaman gibi faktörlere bağlı olarak göz kırpma dalgalanmaları izlenerek bir bütünlük doğrulama yöntemi kullanılır ve gerçek-sahte ayrımı yapılmaya çalışılır. Belirli koşullarda derin sahte tespit performansını artırsa da çeşitli hastalıklar nedeniyle göz kırpma düzeni değişen kişiler bu yöntemde tespit sorunlarına neden olabilmektedir. Göz kırpma frekansına dayalı yöntemler, derin sahte üretiminde kapalı göz örneği yeterince bulunmadan üretilen sahte örneklerde başarılı sonuçlar alabiliyordu. Ancak, eğitim setine yeterince kapalı göz görüntüsü eklendiği sürece, sahtecilik algoritmaları makul göz kırpma sıklığına sahip videoları kolayca üretebilir. Anormal göz kırpma frekansına aşırı dikkat nedeniyle, bu yöntem, göz kırpma frekansı sorunu çözüldükten sonra mevcut derin sahte tespit görevleri için geçerliliğini yitirmektedir.

Bir diğer önemli biyolojik sinyal ise fotopletismogram (PPG) sinyalleridir. PPG sinyalleri, yüzden yansıyan ışık aracılığıyla kalp atış hızı gibi kardiyovasküler değişiklikleri tespit etmekte kullanılır. Bu yöntem, derin sahte videolarda yer alan manipüle edilmiş yüzlerdeki bu tür biyolojik sinyallerin eksikliğini veya anormalliğini algılayabilir. Ciftci vd. (2020) yüzdeki biyolojik sinyalleri (kalp atış hızı gibi) rPPG tekniği ile hesaplayarak sahte yüzleri tespit eden bir yaklaşım geliştirdiler. Çıkarılan biyolojik özellikler SVM ve CNN modellerini eğitmek için kullanıldı. Sonuçlar, düşük çözünürlüklü veya düşük kaliteli videolar için bile yüksek algılama doğruluğuna sahip olduğunu göstermiştir. J. Wu vd. (2024), rPPG sinyallerini kullanarak derin sahteleri tespit etmek için iki aşamalı bir ağ önermektedir. Birden fazla yüz bölgesinden bilgi toplayan çok ölçekli bir mekânsal-zamansal temsil kullanarak yüz videolarından rPPG sinyallerini elde eder. Maske güdümlü yerel dikkat modülü, rPPG haritalarındaki değiştirilmiş bölgeleri vurgularken, zamansal bir dönüştürücü bitişik rPPG haritalarının özelliklerini etkileşimli bir şekilde işleyerek uzun mesafeli bağımlılıkları yakalar. Çeşitli veri kümeleri üzerinde yapılan deneyler, önerilen yöntemin diğer rPPG tabanlı yöntemlerden daha üstün performans sergilediğini göstermektedir. Qi vd. (2020) kalp atış ritmini kullanarak derin sahte video tespiti yapan DeepRhythm isimli modeli duyurdular. Kalp ritim sinyalini görebilmek için MMSTR isimli yöntemi oluşturdular. Bu yöntemin çıkışını temel alarak çift mekânsal-zamansal bir dikkat ağı kullanarak tespit yaptılar. Benzer şekilde DeepFakesOn-Phys ten rengindeki değişiklikler yoluyla kalp atış hızını tahmin ederek tespit yapmaya çalışmaktadır (Hernandez-Ortega vd., 2020). Oldukça başarılı sonuçlar elde etmişlerdir.

Biyolojik sinyallere dayalı tespit yöntemleri, derin sahte videoların biyolojik ve fizyolojik tutarsızlıklarını analiz ederek tespit edilmesine olanak sağlar. Bu yöntemleri kullanan çalışmalar başarılı sonuçlar elde etmektedir. Bununla beraber derin sahte üretim yöntemlerinin kalp atış hızını taklit edebilir yeteneğe erişmesi durumunda geçerliliğini yitiren bir yöntem olacağı açıktır. Ayrıca biyolojik sinyallerden elde edilen bilgiler video kalitesinden ciddi şekilde etkilenmekte ve bu nedenle biyolojik sinyallere dayalı yaklaşımlar için bazı dezavantajlar ve sınırlı uygulama aralığı vardır.

1.4.5. Genel Ağ Tabanlı Yöntemler

Genel ağ tabanlı yöntemler, derin sahte içeriklerin tespitinde kullanılan önemli ve etkili bir teknolojidir. Bu yöntemler, genellikle yapay sinir ağları ve özellikle derin öğrenme modellerini kullanarak sahte ve gerçek içerikleri sınıflandırır. Derin öğrenme, video ve görüntü analizinde karmaşık özellikleri tanıma konusunda yüksek başarı gösterdiğinden, derin sahte tespitinde de yaygın olarak tercih edilmektedir. Bu yöntemlerin avantajları arasında, yüksek doğruluk oranları ve otomatik özellik çıkarma yetenekleri yer alır. Ancak, büyük miktarda eğitim verisi gerektirmeleri ve modellerin karmaşıklığı nedeniyle yüksek hesaplama gücüne ihtiyaç duyulması gibi zorlukları da beraberinde getirmektedir. Bununla birlikte, bu yöntemler, derin sahte içeriklerin yayılmasını önlemede kritik bir rol oynamaktadır ve dijital içerik doğrulama sistemlerinin geliştirilmesinde önemli bir yere sahiptir.

Genel ağ tabanlı yöntemler, sahte içerik tespiti için geliştirilen sistemlerde sıklıkla tercih edilen ve güvenilir sonuçlar sunan bir yaklaşımdır. Bu yöntemlerin erken dönem çalışmalarında öğrenme aktarımı stratejisi sıklıkla görülmektedir. Öğrenme aktarımı stratejisinde, önceden eğitilmiş olan bir derin ağ mimarisi mevcut problem için kullanılırken ağın önceden öğrenmiş olduğu özellikler yeni görevde de kullanılır. Örneğin Zhou vd. (2017), yüz yapılarını analiz etmek için GoogleNet ve steganaliz özelliklerini analiz etmek için yama tabanlı bir üçlü ağ kullanan iki akışlı bir ağ önermektedir. Bir dal evrişimli bir sinir ağı kullanır ve diğeri sahte öznitelikleri çıkarır ve sınıflandırır. Daha sonra iki dalın sınıflandırma puanları bir araya getirilerek sınıflandırma kararı verilir. Rossler vd. (2019) tespit yapmak için Xception ağını kullandılar. Diğer ağlara göre daha başarılı sonuçlar elde ettiler. DFDC veri kümesinin duyurulduğu çalışmada Dolhansky vd. (2019) benzer tespit yöntemleri kullandılar. Kıyaslama yapabilmeyi sağlayan temel

performans değerleri oluşturmak için mevcut iki model test edildi: Küçük bir DNN (altı evrişim katmanından ve tam bağlantılı bir katmandan oluşur) ve mevcut bir XceptionNet. İlk sonuçlar, en iyi yöntemin (XceptionNet) %93 kesinlik sağladığını gösterdi. Bonettini vd. (2020) farklı eğitilmiş CNN modelleri gurubu üzerinde çalışmış ve CNN'ler gurubunun derin sahte tespitinde umut verici sonuçlar elde edebileceğini göstermiştir. Bununla beraber bu tür ağ tabanlı yöntemler aşırı uyum problemine meyillidirler. Bu nedenle ön işleme yoluyla gerçek ve sahte videolar arasındaki içsel farklılıklardan yararlanılmaya çalışılır. Optik akış hesaplama gibi bazı ön işleme yöntemlerinin, ağ tabanlı yöntemlerde görüntü kareleri arası farklılıklardan yararlanmak için yararlı olduğu kanıtlanmıştır (Amerini vd., 2019).

Öğrenme aktarımı yöntemiyle kullanılan derin ağ modellerinin performansı derin sahte tespit görevlerinde sınırlı kalmaktadır. Yeni ve çok sayıda örnek içeren veri kümelerinin duyurulmasıyla birlikte özel tasarlanmış derin ağ modellerinin oluşturulması ve eğitilmesi konusundaki çalışmaların önü açılmıştır. Nguyen vd. (2019), algılama ağlarının performansını artırmak için bir kapsül ağı tanıttı. Yüz görüntüleri ilk önce önceden eğitilmiş VGG-19 ağına giriş olarak verilir. Çıkarılan özellikler daha sonra birkaç birincil kapsül ve iki çıkış kapsülü içeren kapsül ağına girilir. Birincil kapsüller tarafından çıkarılan özellikler arasındaki uyum, dinamik bir yönlendirme algoritması tarafından dinamik olarak hesaplanır ve sonuçlar son olarak uygun çıkış kapsülüne yönlendirilir. Çıkarılan gizli özelliklerin görselleştirilmesi, kapsül ağları ve dinamik yönlendirme algoritmasının kombinasyonunun manipülasyonları tespit etmek için etkili olduğunu göstermiştir. Afchar vd. (2018) Meso-4 ve MesoInception-4 ağlarını önermektedir. Derin sahte tespiti görüntü mezoskopik özellikleri yardımıyla gerçekleştirilir. Yöntem kendi oluşturdukları derin sahte veri kümesi üzerinde eğitilmiş ve test edilerek 98.4% doğruluk elde edilmiştir. Wodajo & Atnafu (2021) CNN ve Görüntü Dönüştürücüsü'nü birleştiren bir yöntem önermiş ve DFDC veri kümesi üzerinde rekabetçi bir sonuç elde etmiştir. Roy vd. (2022) 3D ResNet ve dikkat mekanizmalarının kombinasyonlarını kullanmıştır. Cao vd. (2022) genelleme kabiliyetini ve sağlamlığı artırmak için yeniden yapılandırma-sınıflandırma öğrenmesine dayalı bir yöntem önermektedir. Wang & Deng (2021), Sahtecilik Dikkat Haritaları ve Şüpheli Sahtecilik Silme olmak üzere iki ana bileşeni olan Temsili Sahtecilik Madenciliği adlı bir yöntem sunmaktadır. Deneyler, TSM'nin çeşitli manipülasyon tekniklerine karşı tespit performansını artırdığını göstermektedir. Daichi vd. (2021) derin sahte videoları tespit etmek için TD3DCNN isimli bir ağ önermektedir. Sabit

uzunluktaki çerçeve hacimlerini bir 3DCNN'ye besleyerek ve çerçeveleri rastgele örneklemek için zamansal bir bırakma işlemi ekleyerek video çerçeveleri arasındaki olası tutarsızlıklardan yararlanır. Sun vd. (2022) genel yüz sahteciliği tespiti için DCL adlı bir yöntem önermektedir. DCL, ayırt edici özellikleri öğrenmek için farklı seviyelerde kontrastlı öğrenme gerçekleştirir. Çeşitli veri kümeleri üzerinde yapılan deneyler, yöntemin görülmemiş örneklerde genelleme konusunda iyi sonuçlar verdiğini göstermektedir. D. Liu vd. (2023) kalıntı birleşik öğrenme kullanarak genelleştirilmiş yüz sahteciliği tespiti için FedForgery adlı yeni bir yöntem önermektedir. Sahtecilik ipuçlarını yakalamak için gerçek ve sahte yüzlerden sağlam kalıntı özellik haritaları öğrenmek için varyasyonel bir otomatik kodlayıcı tasarlar. Asha vd. (2024) videolardan mekânsal ve zamansal özellikler çıkarmak için VGG16 ve optik akış kullanan bir görsel dikkat ağı sunmaktadır. Mevcut veri tabanlarındaki gerçek/sahte örnekleri ve farklı teknikler kullanılarak oluşturulan yeni derin sahte örneklerini birleştiren özel birçok modlu derin sahte veri kümesi oluşturulmuştur. Bu yaklaşım, özel veri kümesinde %97,6 doğruluk oranına ulaşmıştır. Heidari vd. (2024) veri gizliliğini korurken sahte videoların tespitini iyileştirmek için Blok zinciri-Etkin Birleşik Derin Sahte Tespiti adı verilen bir mimari sunmaktadır. Çeşitli kaynaklardan gelen videoları standartlaştırmak için bir veri normalleştirme prosedürü uygular. Derin öğrenme modelleri daha sonra derin sahteciliği gösteren kalıpları saptamak için kullanılır. Birleştirilmiş öğrenme yaklaşımı, özel verileri paylaşmadan yerel model güncellemelerinden küresel bir modeli eğitir. Deneysel sonuçlara göre, önerdikleri sistem çeşitli veri kümeleri üzerinde derin sahtecilikleri güvenilir bir şekilde belirleyebilmektedir. Nguyen vd. (2024) yüksek kaliteli derin sahte tespiti için LAA-Net adlı bir yöntem sunmaktadır. LAA-Net, yapaylığa eğilimli hassas noktalara odaklanmak için çok görevli bir öğrenme mimarisi içinde açık bir dikkat mekanizması önermektedir. Bu noktaları modellemek için bir ısı haritası dalı ve öz tutarlılık dalı olmak üzere iki dallı bir yapı kullanılır. Ek olarak, çok ölçekli özellikleri artıklık olmadan toplamak için E-FPN adı verilen bir Geliştirilmiş Özellik Piramidi Ağı tanıtılmış ve yerleştirilmiş özelliklerin modellenmesine olanak sağlanmıştır. Birden fazla karşılaştırma ölçütü üzerinde yapılan deneyler, LAA-Net'in iyi sonuçlar verdiğini ve Celeb-DF veri kümesinde %95,40 AUC elde edildiğini göstermektedir. Kaddar vd. (2024) Evrimsel Sinir Ağını Görüntü Dönüştürücüsü ile birleştiren bir yöntem önermektedir. Çeşitli veri kümeleri üzerinde rekabetçi sonuçlar elde etmişlerdir.

Bu yöntemler, derin sahte içeriklerin tespitinde yüksek doğruluk ve güvenilirlik sağlar. Öğrenme aktarımı ve özel tasarlanmış ağlar gibi farklı stratejiler kullanılarak geliştirilen bu yöntemler, sahte içeriklerin belirlenmesinde önemli bir rol oynamaktadır. Gelecekte, bu yöntemlerin daha da geliştirilerek daha yüksek doğrulukta ve güvenilirlikte sahte içerik tespit sistemleri oluşturulması hedeflenmektedir.

1.5. Derin Öğrenme Mimarileri

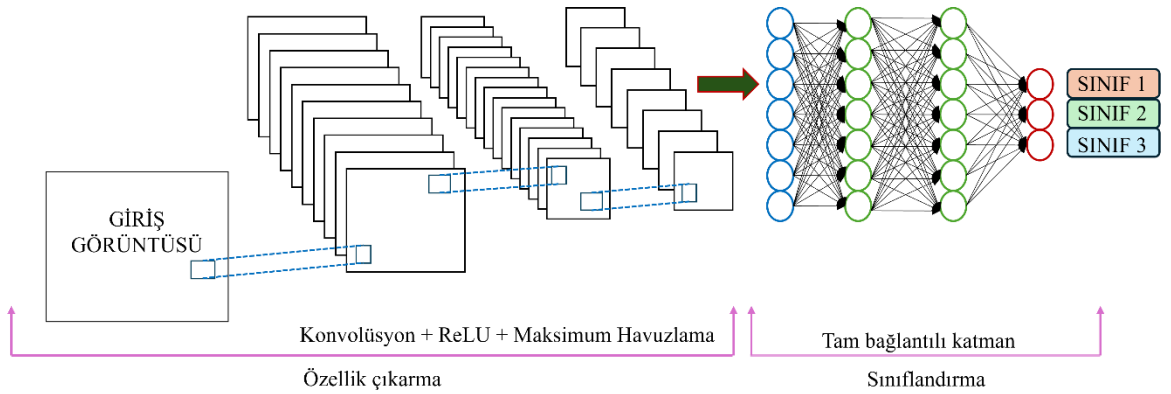
Derin öğrenme, makine öğrenmesinin bir alt dalıdır ve karmaşık veri analizleri için çok katmanlı yapay sinir ağlarını kullanır. Derin ağ mimarileri, verilerin özelliklerini otomatik olarak çıkararak daha yüksek doğrulukta sonuçlar elde etmeye yardımcı olur. Bu mimariler, özellikle görüntü işleme, ses tanıma ve doğal dil işleme gibi alanlarda başarıyla kullanılmaktadır.

Derin sinir ağları (DSA), yapay sinir ağlarından (YSA) birkaç önemli açıdan ayrılır. Öncelikle, DSA'lar daha fazla katmana sahiptir ve bu da onların daha karmaşık veri ilişkilerini öğrenmelerine olanak tanır. YSA'lar genellikle daha az katmanlıdır ve bu nedenle basit görevlerde etkili olsalar da karmaşık verileri analiz etmekte yetersiz kalabilirler. İkinci olarak, DSA'lar verilerden özellikleri otomatik olarak çıkarabilirken, YSA'lar genellikle önceden belirlenmiş özelliklerle çalışır. Bu, DSA'ların büyük veri kümeleri üzerinde çalışırken daha esnek ve güçlü olmasını sağlar, çünkü manuel olarak özellik seçimine ihtiyaç duymazlar. Son olarak, DSA'lar yüksek hesaplama gücü gerektiren işlemleri gerçekleştirebilmek için genellikle GPU'lar gibi güçlü donanımlara ihtiyaç duyar. Bu da onları, büyük ölçekli veri analizleri ve derin öğrenme modelleri için ideal kılar. Bu üç temel farklılık, DSA'ların karmaşık veri işleme ve analiz görevlerinde YSA'lara kıyasla daha üstün performans göstermesini sağlar.

DSA, birbirine bağlı yapay nöronlardan oluşan çok katmanlı yapılardır. Her katman, önceki katmandan gelen verileri işler ve bir sonraki katmana iletir. Bu süreç, ağın derinliğine bağlı olarak birçok kez tekrarlanır. DSA, verileri işleme ve öğrenme süreçlerinde birkaç önemli aşamadan geçer. Bu aşamalar, ileri yayılım (forward propagation) ve geri yayılım (backward propagation) olarak iki ana başlık altında toplanabilir. İleri yayılımda, girdiler katmanlar boyunca ileriye doğru aktarılır. Her nöron, aldığı verileri belirli ağırlıklar ve aktivasyon fonksiyonları ile işler. Bu aşama, verilerin ağın son katmanına kadar taşınmasını ve ilk tahminlerin yapılmasını sağlar. Geri yayılım

aşamasında ise, hesaplanan hata geriye doğru yayılır ve bu hata, ağırlıklarının güncellenmesi için kullanılır. Bu işlem, ağı öğrenme sürecini sağlar ve modelin doğruluğunu artırmak için gerçekleştirilir.

DSA, genel mimarisi Şekil 11’de gösterildiği gibi kabaca özellik çıkarıcı ve sınıflandırıcı bölümlerden oluşur. Özellik çıkarma, verinin önemli kısımlarını belirleyerek yüksek boyutlu verilerin daha düşük boyutlu fakat anlamlı temsillerini elde etme sürecidir. Derin ağlarda bu işlem, ağı öğrenme süreci boyunca otomatik olarak gerçekleştirilir. Özellik çıkarımının en büyük avantajı, manuel olarak özelliklerin çıkarılmasına gerek kalmamasıdır. Ayrıca, öğrenilen özellikler genellikle farklı veri kümelerine de uygulanabilir, bu da genel geçerliliği sağlar. Özellikle görüntü işleme alanında başarılı olan konvolüsyonel sinir ağları (CNN), verinin yerel özelliklerini öğrenmek için konvolüsyon (evrişim) katmanlarını kullanır. Konvolüsyon katmanları, filtreler (kerneller) olarak adlandırılan küçük matrislerle çalışır. Bu filtreler, giriş verisi üzerinde kaydırılarak (stride) çarpım işlemleri yapılır ve belirli özelliklerin çıkarılmasını sağlar. Aktivasyon fonksiyonları, genellikle ReLU (Rectified Linear Unit) gibi doğrusal olmayan fonksiyonlar kullanılır ve bu fonksiyonlar, doğrusal olmayan ilişkileri öğrenmeye yardımcı olur. Havuzlama (pooling) katmanları ise boyut küçültme işlemleri yapar; örneğin, max-pooling belirli bir bölgedeki maksimum değeri seçer. Bu işlemler, hesaplama maliyetini düşürür ve aşırı öğrenmeyi (overfitting) önler.



Şekil 11. Genel derin öğrenme mimarisi

Derin ağların çıktıları, son katmanda elde edilen ve genellikle sınıflandırma, regresyon gibi görevleri yerine getiren sonuçlardır. Bu çıktılar, genellikle softmax gibi bir aktivasyon fonksiyonundan geçirilir ve nihai tahminler yapılır. Çıktı aşamaları, tam

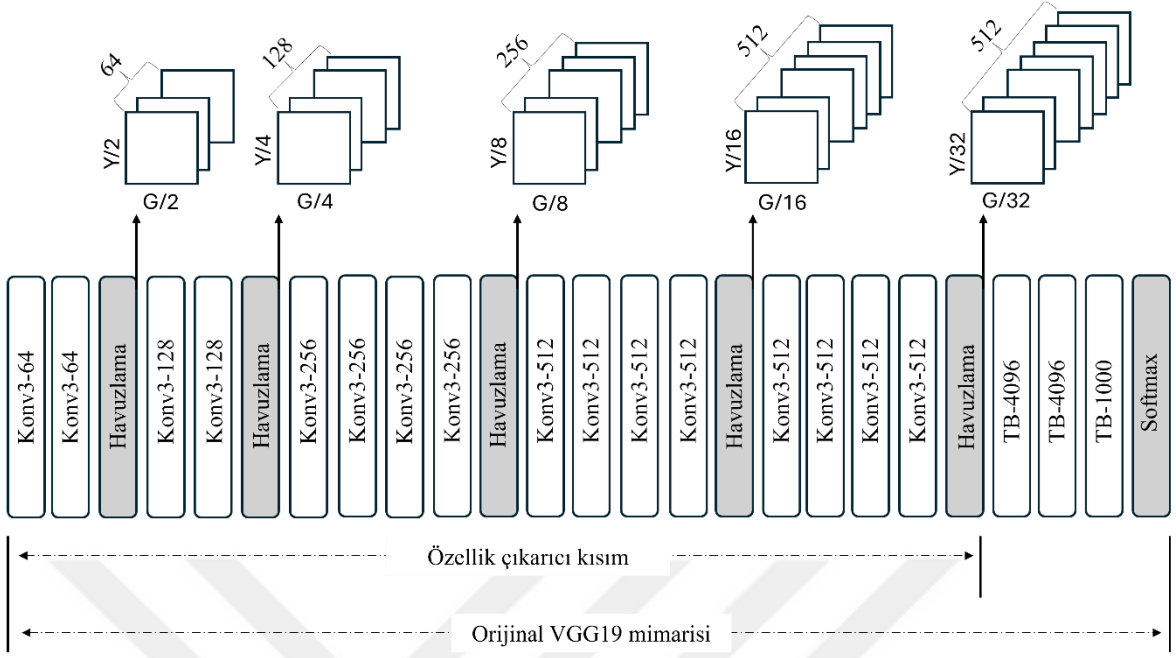
bağlantılı (fully connected) katmanlardan geçerek başlar. Bu katmanlar, özellik haritalarını düzleştirerek tek boyutlu bir vektöre dönüştürür. Sonuçlar, softmax veya sigmoid gibi aktivasyon fonksiyonlarından geçirilerek olasılık dağılımı elde edilir ve nihai tahminler yapılır. En yüksek olasılığa sahip sınıf, nihai sonuç olarak belirlenir.

Sonuç olarak, derin ağ mimarileri, özellik çıkarma ve konvolüsyon aşamaları sayesinde verilerin karmaşık yapısını öğrenir ve yüksek doğrulukta tahminler yapar. YSA'lardan daha fazla katman ve karmaşıklık içerirler, bu da onların geniş çaplı ve karmaşık veri analizlerinde başarılı olmasını sağlar. Özellik çıkarma ve konvolüsyon aşamaları, verilerin anlamlı temsillerini elde etmeye yardımcı olur ve nihai tahminlerin doğruluğunu artırır. Bu nedenle, derin ağ mimarileri, günümüzün veri yoğun dünyasında önemli bir rol oynamaktadır ve birçok alanda yenilikçi çözümler sunmaktadır.

Alt bölümlerde bu tez çalışması kapsamında kullanılan DSA modelleri hakkında detaylı bilgiler verilecektir. Bu modeller VGG19, Efficient-Net B0, Efficient-Net B4, Capsule Forensics ve ArCapsNet'tir.

1.5.1. VGG19

VGG19, Oxford Üniversitesi'ndeki Görsel Geometri Grubu (VGG) tarafından önerilen etkili bir CNN mimarisidir (Simonyan & Zisserman, 2014). Derin öğrenme alanında sıkça kullanılan bir mimaridir ve özellikle görüntü sınıflandırma problemlerinde yüksek performans gösterir. VGG19 ağı Şekil 12'de gösterildiği gibi toplamda 19 katmandan oluşmaktadır ve adında yer alan 19 sayısı buradan gelmektedir. Temelinde evrişimli katmanlar, havuzlama katmanları, tam bağlantılı katmanlar ve softmax çıkışı gibi çeşitli bileşenler bulunmaktadır. Katmanlar mimari içindeki işlemlerine göre başlıklar altında gruplandırılabilir.



Şekil 12. VGG19 ağı mimarisi

Evrişimli Katmanlar (Convolutional Layers): VGG19 ağı, 16 adet evrişimli katmana sahiptir. Mimari içerisinde bulunan bu katmanlar giriş görüntüsündeki özellikleri çıkarmada ve daha karmaşık özellikleri öğrenmede kullanılır. 16 katmanın her birinde 3x3 boyutunda çekirdekler kullanılır. Çekirdekler, görüntüdeki yerel bölgeleri analiz ederek özellik haritalarını çıkarır. VGG19 ağında evrişimli katmanların her birinden sonra uygulanan aktivasyon fonksiyonu genelde ReLU olarak kullanılır.

Havuzlama Katmanları (Pooling Layers): VGG19 ağı, toplamda 5 maksimum havuzlama katmanına sahiptir. Evrişimli katmanlardan elde edilen özellik haritalarının boyutlarını ve hesaplama maliyetini düşürmek için kullanılır. Ağda bulunan maksimum havuzlama katmanların her birinde 2x2 boyutunda fitreler ve adım 2 (stride 2) kullanılır. Böylelikle her havuzlama katmanından sonra özellik haritalarının boyutunun yarıya düşürülmesi sağlanır.

Tam Bağlantılı Katmanlar (Fully Connected Layers): VGG19 ağı, son kısmında 3 adet tam bağlantılı katmana sahiptir. Tam bağlantılı katmanlar, önceki katmanlarda çıkarına özelliklere göre sınıflandırma işlemini yapar. Tam bağlantılı katmanlardan ilk ikisi 4096 nörona sahiptir ve bu katmanlar genellikle ReLU aktivasyon fonksiyonunu kullanır. Son katman ise sınıflandıracak sınıfların sayısı kadar nörona sahiptir ve softmax aktivasyon fonksiyonunu kullanarak olasılık dağılımını elde eder.

Softmax Çıkışı (Softmax Output): VGG19 ağıının son katmanı olarak softmax katmanı yer alır. Bu katman, her bir sınıf için olasılık değeri üretir ve bu değerler toplamda 1 olacak şekilde normalize edilerek olasılığı en yüksek sınıf, ağıın tahmini olarak belirlenir.

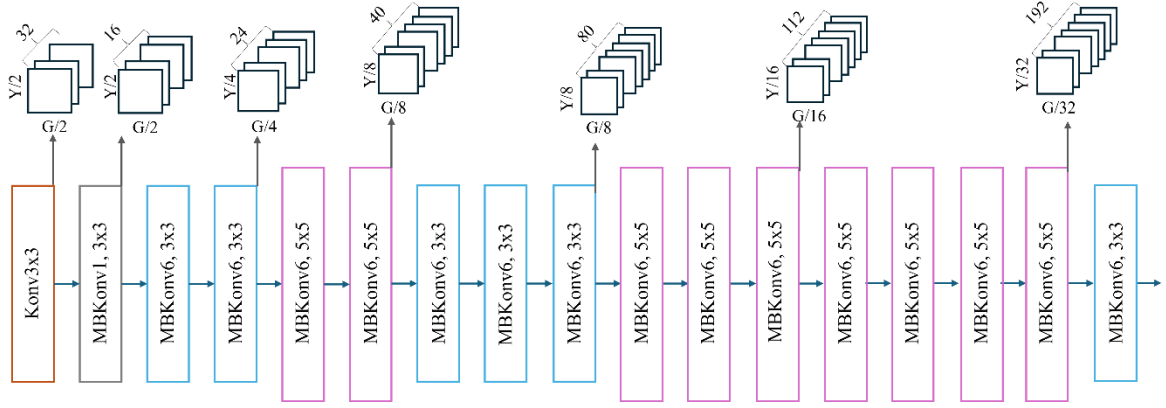
VGG19 ağı, Bölüm A (Özellik Çıkarma Parçası) ve Bölüm B (Sınıflandırma Parçası) olmak üzere temelde iki ana bölümde incelenir.

Özellik Çıkarma Bölümü (Feature Extraction Part): Bu bölüm, ağa verilen giriş görüntüsünden anlamlı özellikler çıkarmak için evrişimli katmanlarını ve havuzlama katmanlarını içermektedir. İlk katmanlar kenarlar ve dokular gibi düşük seviye özellikleri öğrenirken, derinde bulunan katmanlar ise nesne parçaları gibi daha karmaşık özellikleri öğrenir.

Sınıflandırma Bölümü (Classification Part): Sınıflandırma bölümü, öğrenilen özellikleri bir sınıflandırma kararına dönüştüren tam bağlantılı katmanlarını ve softmax katmanını içerir ve bu bölümde elde edilen özelliklere göre tanıma işlemi gerçekleştirilir.

1.5.2. Efficient-Net B0

Efficient-Net, Google AI'daki araştırmacılar tarafından tanıtılan bir CNN mimarisidir (Tan & Le, 2019). Efficient-Net'teki "Efficient" terimi, önceki mimarilere kıyasla önemli ölçüde daha az parametre ve hesaplama gereksinimi kullanırken dikkate değer doğruluk derecesi elde etme özelliğini ifade eder. Bu verimlilik, optimum performans elde etmek için model derinliğini, genişliğini ve çözünürlüğünü dengeleyen bileşik ölçekleme adı verilen bir teknikle elde edilir. Efficient-Net B0, Efficient-Net ailesinin temel modeli olarak sunulmaktadır. Ağın içerdiği katmanlar Şekil 13'te gösterilmektedir. Ağ derinliğini, genişliğini ve çözünürlüğünü dengeli bir şekilde ölçeklendiren bileşik ölçeklendirme yöntemini kullanır. Bu yaklaşım, model karmaşıklığı ve performans arasında doğru dengeyi bulmasını sağlayarak verimli ancak güçlü özellik çıkarma yetenekleriyle sonuçlanır.



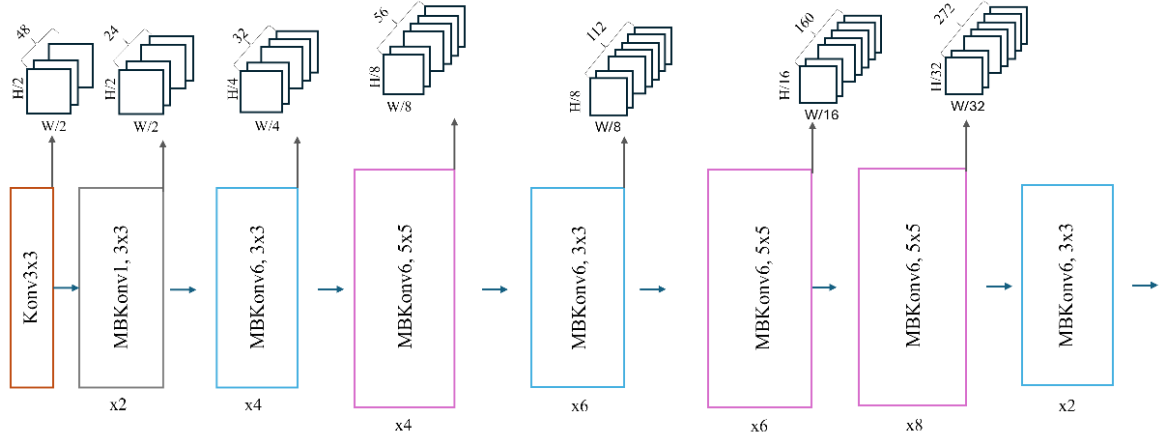
Şekil 13. Efficient-Net B0 ağı mimarisi

Efficient-Net B0 mimarisi, önceden eğitilmiş bir modelin yeni bir problem çözmek için kullanılmasını ifade eden ve transfer öğrenme, derin öğrenmede yaygın olarak kullanılan bir tekniktir. Bu model, derinlik, genişlik ve çözünürlük boyutlarını eşit şekilde ölçeklendirmek için bileşik katsayı kullanan bir CNN mimarisi ve ölçekleme yöntemi sunar.

Bileşik ölçeklendirme yaklaşımı, genişlik, derinlik ve çözünürlük boyutlarını eşit şekilde ölçeklendirerek model performansını artırır. Genellikle katman sayısının artırılması bir doyum noktasına kadar iyidir, fakat katman sayısı doyum noktasını aştığı zaman algoritmanın etkinliği azalacak ve performans eğrisindeki artış yerini sabit bir çizgiye bırakacaktır. Bir CNN'nin derinliği, genişliği ve çözünürlüğü üç ölçeklendirme boyutudur. Ağın derinliği, katman sayısını ifade eder ve daha derin ağlar, daha karmaşık özellikleri öğrenebilir. Genişlik, her katmandaki kanal sayısını ifade eder ve daha fazla kanal, daha fazla özellik haritası çıkarabilir. Çözünürlük ise modele verilen görüntülerin boyutlarını ifade eder ve daha yüksek çözünürlük, daha ayrıntılı özellikler sunar.

1.5.3. Efficient-Net B4

Efficient-Net, bileşik ölçeklendirme uygulayarak değiştirilmiş bir sinir ağı mimarisi modelidir. Efficient-Net B4, aynı Efficient-Net ailesine aittir ancak Efficient-Net B0'a kıyasla daha büyük ve daha güçlü bir modeli temsil eder. Modelin içerdiği katmanlar Şekil 14 üzerinde gösterilmektedir.



Şekil 14. Efficient-Net B4 ağı mimarisi

Bu model ağı genişliğini, derinliğini ve çözünürlüğünü dengeleyerek doğruluğu ve verimliliği artırmayı amaçlar (Tan & Le, 2019). Aynı bileşik ölçeklendirme ilkelerini takip eder, ancak daha fazla derinlik, genişlik ve çözünürlük ile giriş verilerindeki daha karmaşık desenleri ve ayrıntıları yakalamasını sağlar. Efficient-Net B4, hem hesaplama verimliliği hem de doğruluk açısından yüksek performans sunabilen bir modeldir. Bu modelin ölçeklenebilir yapısı, farklı görevler ve veri kümeleri için kolayca uyarlanabilir olmasını sağlamaktadır. Efficient-Net B4, ağ genişliğini, derinliğini ve çözünürlüğünü eşit şekilde ölçeklendirmek için 4'lük bir bileşik katsayı kullanır. Model, ImageNet-1k veri kümesinde 380x380 çözünürlükte eğitilmiştir ve görüntü sınıflandırma görevlerinde kullanılabilir. Efficient-Net B4, nesne tanıma, anlamsal segmentasyon ve resim sınıflandırma gibi çalışmalarda üstün performansı ve verimliliği ile tercih edilen bir yöntemdir.

1.5.4. Kapsül Ağları

Kapsül ağları, Geoffrey Hinton ve ekibi tarafından tanıtılan ve geleneksel konvolüsyonel sinir ağlarının (CNN) sınırlamalarını aşmayı amaçlayan yenilikçi bir derin öğrenme mimarisidir (Sabour vd., 2017). Kapsül ağları, özellikle nesnelerin ve özelliklerin hiyerarşik ilişkilerini daha iyi modelleyerek, görüntü tanıma ve sınıflandırma görevlerinde daha yüksek doğruluk ve dayanıklılık sağlar. Geleneksel CNN'ler, görüntülerdeki nesneleri tanıırken bazen konum, ölçek ve döndürme gibi varyasyonları anlamakta zorluk çekebilir. CNN'ler, her ne kadar güçlü özellik çıkarıcılar olsalar da nesnelerin uzaysal ilişkilerini tam anlamıyla modelleyemezler. Bu durum, özellikle karmaşık ve değişken görüntülerde sınıflandırma hatalarına yol açabilir. Kapsül ağları ise, her biri bir nesnenin veya

özelliklerin belirli bir yönünü temsil eden küçük sinir ağı grupları olan kapsüller kullanarak bu sorunu çözmeyi amaçlar. Her kapsül, bir nesnenin varlığı ve özellikleri hakkında bilgi taşıyan bir vektör çıkışı üretir. Bu vektör, nesnenin pozisyonu, ölçeği, yönelimi gibi bilgilerle birlikte, nesnenin mevcut olup olmadığına dair bir olasılık değeri içerir. Kapsül ağlarının temel bileşenlerinden biri olan dinamik yönlendirme algoritması, kapsüller arasındaki bilgi akışını kontrol eder. Bu algoritma, düşük seviyeli kapsüllerin çıktılarının, yüksek seviyeli kapsüllere nasıl yönlendirileceğini belirler. Böylece, nesnelerin parçaları arasındaki ilişkiler daha iyi anlaşılır. Dinamik yönlendirme, kapsüllerin bir araya gelerek daha anlamlı yapılar oluşturmasını sağlar.

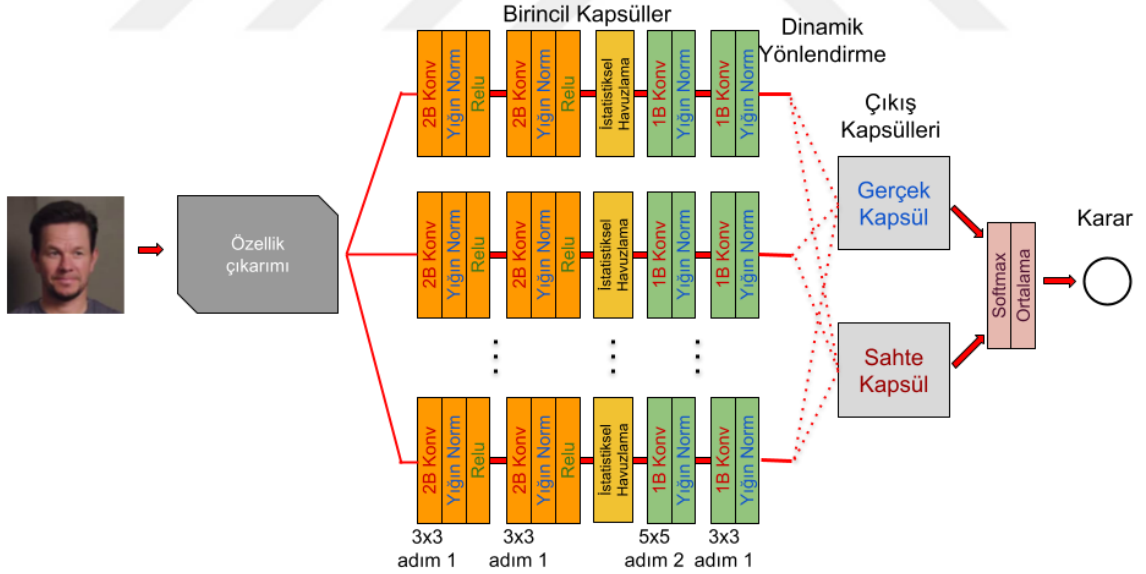
Kapsül ağları, katmanlar arasındaki ilişkiyi ve bilgi akışını yeniden tanımlayarak çalışır. Temel olarak, kapsül ağları iki ana katmandan oluşur. Bunlar; Birincil Kapsül Katmanı (Primary Capsule Layer) ve Sınıf Kapsül Katmanı (Class Capsule Layer) bölümleridir. Birincil kapsül katmanı geleneksel CNN'lerdeki konvolüsyon katmanları ile benzer şekilde çalışır ve girdiden temel özellikleri çıkarır. Ancak, burada her bir kapsül, belirli bir özellik haritasının aktivasyonlarını temsil eden bir vektör üretir. Örneğin, 32 farklı özelliği çıkaran bir konvolüsyonel katman, her biri 8 boyutlu 32 farklı kapsül oluşturabilir. Bu kapsüller, görüntüdeki basit özellikleri temsil eder. Sınıf kapsül katmanı birincil kapsül katmanından gelen çıktıları alır ve daha karmaşık yapılar oluşturur. Her bir sınıf kapsül, daha yüksek seviyede bir nesne veya özelliği temsil eder. Dinamik yönlendirme algoritması, bu katmanda aktif olarak kullanılarak birincil kapsüllerin çıktılarını sınıf kapsüllere yönlendirir. Yüksek seviyeli kapsüller, daha düşük seviyeli kapsüllerin çıktılarını alarak daha karmaşık ve ilişkili yapılar oluşturur.

Kapsül ağlarının temel bileşenlerinden biri olan dinamik yönlendirme algoritması, kapsüller arasındaki bilgi akışını kontrol eder. Her bir düşük seviyeli kapsül, yüksek seviyeli kapsüllere olan bağlanma olasılıklarını öğrenir. Bu olasılıklar, eğitim sürecinde sürekli olarak güncellenir ve en iyi bağlantılar seçilerek bilgi akışı optimize edilir. Böylece, her bir kapsül, en anlamlı ve doğru bilgiye sahip olacak şekilde organize edilir. Dinamik yönlendirme süreci, kapsül çıktılarının yüksek seviyeli kapsüllere nasıl iletileceğini belirler. Bu süreçte, düşük seviyeli kapsüllerin çıktıları, bir ağırlık matrisi aracılığıyla yüksek seviyeli kapsüllere iletilir. Ağırlık matrisleri, kapsüller arasındaki bağlantıların gücünü ve doğruluğunu temsil eder. Yönlendirme süreci, birkaç iterasyon boyunca tekrarlanarak en iyi bağlantılar ve bilgi akışı optimize edilir.

Kapsül ağlarının bir diğer önemli özelliği, kapsül aktivasyonu olarak adlandırılan ve kapsüllerin çıktılarının etkinleştirilme derecesini belirleyen fonksiyondur. Kapsül aktivasyonu, bir kapsülün belirli bir nesne veya özelliği ne kadar doğru temsil ettiğini gösterir. Bu mekanizma, kapsüllerin doğru nesne veya özellikleri temsil etme olasılığını artırır ve yanlış pozitifleri azaltır. Kapsül aktivasyonu, genellikle bir squashing fonksiyonu ile gerçekleştirilir. Bu fonksiyon, kapsül çıktılarını normalize eder ve uzunluğu 0 ile 1 arasında olacak şekilde sıkıştırır. Bu sayede, daha güçlü kapsüller daha belirgin hale gelirken, zayıf ve anlamsız kapsüller baskılanır.

Bu tez çalışmasında, kapsül ağlarının iki farklı modifikasyonu olan Capsule Forensics ve ArCapsNet kullanılmıştır. Bu modifikasyonlar, derin sahte video tespitinde kullanılan özel kapsül ağlarıdır ve çeşitli iyileştirmelerle performansları artırılmıştır.

Capsule Forensics, orijinal kapsül ağı mimarisinin uyarlanmış bir versiyonudur ve özellikle derin sahte içeriklerin tespiti için geliştirilmiştir (Nguyen vd., 2019). Bu model, birden fazla birincil kapsül ve iki çıkış kapsülünden (gerçek ve sahte) oluşur. Capsule Forensics ağının katman mimarisi Şekil 15'te gösterilmektedir.

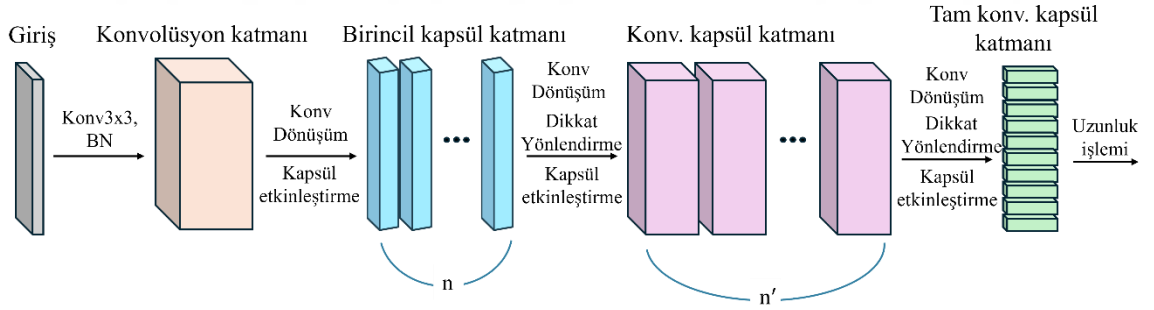


Şekil 15. Capsule Forensics model mimarisi

Her birincil kapsül, 2D konvolüsyonel bir parça, istatistiksel bir havuzlama katmanı ve 1D konvolüsyonel bir parçadan oluşur. Bu yapı, modelin karmaşık görsel özellikleri daha derinlemesine analiz etmesini sağlar. Birincil kapsüllerin çıktıları, dinamik

yönlendirme algoritması kullanılarak çıkış kapsüllerine yönlendirilir. Dinamik yönlendirme, kapsüller arasındaki bilgi akışını optimize ederek, her bir kapsülün doğru özellikleri ve ilişkileri öğrenmesini sağlar. Çıkış kapsülleri, belirli bir görüntünün veya videonun gerçek mi yoksa sahte mi olduğunu belirlemek için aktivasyon değerleri üretir. Bu aktivasyonlara dayanarak, nihai sonuç hesaplanır ve içerik sınıflandırılır. Capsule Forensics, geleneksel yöntemlere kıyasla daha yüksek doğruluk ve güvenilirlik sunar. Bu model, sahte içeriklerin belirlenmesinde daha etkili olmasını sağlayan, özellikle kapsüller arasındaki hiyerarşik ilişkileri ve uzaysal düzenlemeleri daha iyi anlama yeteneğine sahiptir. Sonuç olarak, Capsule Forensics, derin sahte tespitinde yenilikçi ve güçlü bir araç olarak öne çıkar.

Attention Routing Capsule-Net (ArCapsNet), orijinal kapsül ağından modifiye edilmiş bir başka versiyondur (J. Choi vd., 2019). Bu model, orijinal kapsül ağındaki dinamik yönlendirme ve squash aktivasyon fonksiyonlarını dikkat yönlendirme ve kapsül aktivasyonu ile değiştirmiştir. ArCapsNet, birincil kapsüller katmanı, konvolüsyonel kapsüller katmanı ve tam konvolüsyonel kapsüller katmanından oluşur. ArCapsNet modelinin mimari yapısı Şekil 16’da verilmektedir.



Şekil 16. ArCapsNet model mimarisi

Her bir katmanın çıktısına 1x1 konvolüsyon ve tanh aktivasyonu uygulanan bir kapsül aktivasyonuna sahiptir. Dikkat yönlendirme mekanizması, kapsüllerin özellikleri daha seçici bir şekilde işlemesine olanak tanır ve böylece modelin doğruluğunu ve verimliliğini artırır. Kapsül aktivasyonu, kapsüllerin etkinleştirilme derecesini belirler ve bu, modelin daha doğru ve güvenilir sınıflandırmalar yapmasını sağlar. Birincil kapsüller katmanı, görüntüden temel özellikleri çıkaran ve bu özellikleri kapsül aktivasyonu ile işleyen ilk katmandır. Bu katman, görüntüdeki düşük seviyeli özellikleri (örneğin, kenarlar, köşeler) yakalar ve bunları daha karmaşık özelliklere dönüştürmek için işler. Birincil

kapsüller katmanı, kapsüller arasındaki ilişkileri ve mekânsal düzenlemeleri daha iyi anlamak için dikkat yönlendirme mekanizmasını kullanır. Konvolüsyonel kapsüller katmanı, birincil kapsüllerden gelen özellikleri daha derin ve geniş bir şekilde işler. Bu katman, daha karmaşık ve yüksek seviyeli özelliklerin çıkarılmasını sağlar. Konvolüsyonel kapsüller, geleneksel konvolüsyonel sinir ağlarındaki filtreler gibi çalışır, ancak kapsüller arasındaki ilişkileri koruyarak ve bu ilişkileri dikkate alarak daha zengin bir özellik kümesi oluşturur. Konvolüsyonel kapsüller katmanı, birden fazla tekrar eden katmandan oluşabilir, bu da modelin esnekliğini ve adaptasyon yeteneğini artırır. Tam konvolüsyonel kapsüller katmanı, modelin son katmanıdır ve 1×1 konvolüsyon ve tanh aktivasyonu kullanarak özellikleri işler. Bu katman, konvolüsyonel kapsüllerden gelen çıktıları son bir işleme tabi tutarak, nihai sınıflandırma için hazır hale getirir. Tanh aktivasyonu, özelliklerin sıkıştırılmasını ve daha dengeli bir dağılıma sahip olmasını sağlar. ArCapsNet, özellikle nesne tanıma ve sınıflandırma görevlerinde başarılı bir performans gösterir. Bu mimari, değişken koşullar altında ve çeşitli perspektiflerden görüntülenen nesneleri tanıma konusunda yüksek başarı oranına sahiptir. Dikkat yönlendirme ve kapsül aktivasyonu kombinasyonu, modelin sahte ve gerçek içerikleri daha hassas bir şekilde ayırt etmesini sağlar. Sonuç olarak, ArCapsNet, kapsül ağlarının gelişmiş bir versiyonu olarak, derin sahte tespiti ve diğer görsel tanıma görevlerinde yenilikçi ve güçlü bir araçtır.

1.6. Yüz Tespit Algoritmaları

Bu tez çalışmasında derin sahte videolar üzerinde çalışılırken videolarda yer alan yüzlerin tespit edilmesi gerekmektedir. Bu amaçla çeşitli yüz tespit algoritmaları kullanılarak videolar yer alan yüzlerin tespiti gerçekleştirilmiştir. Yüz tespit algoritmaları, bilgisayarla görme ve yapay zekâ uygulamalarında kritik bir rol oynar. Bu algoritmalar, dijital görüntülerdeki ve videolardaki yüzleri algılayarak, yüz tanıma, yüz ifadesi analizi ve sahte içerik tespiti gibi çeşitli uygulamalara olanak tanır. Gelişmiş yüz tespit algoritmaları, yüksek doğruluk ve hız ile çalışarak, karmaşık ve değişken ortam koşullarında bile güvenilir sonuçlar sağlar. Bu tez çalışmasında, derin sahte videolarda yer alan yüzlerin tespiti için MTCNN ve S3FD gibi önde gelen yüz tespit algoritmaları kullanılmıştır. Bu bölümde, bu algoritmaların teknik detayları ve çalışma prensipleri hakkında bilgi verilecektir.

1.6.1. MTCNN

MTCNN (Multi-task Cascaded Convolutional Networks), yüz tespiti için yaygın olarak kullanılmakta olan güçlü bir algoritmadır (K. Zhang vd., 2016). MTCNN, yüz tespitini, yüz özellik noktalarının (gözler, burun, ağız) konumunu ve yüz hizalamasını P-Net, R-Net ve O-Net olarak adlandırılan üç aşamadan oluşan bir yöntemle gerçekleştirir.

İlk aşamada, tahmini yüz bölgelerini belirlemek için P-Net isimli öneri ağı kullanılır. P-Net, girdi görüntüsünü kaydırmalı bir pencere kullanarak tarar ve potansiyel yüz bölgelerini belirlemek için konvolüsyonel sinir ağı kullanır. Bu ağ yüz bölgelerini içeren tahminler üretir ve bu bölgeleri daha hassas bir şekilde değerlendirecek olan sonraki aşamalara iletir.

İkinci aşamada, P-Net tarafından önerilen bölgeler, daha kesin yüz sınırı kutuları ve yüz özellikleri belirlemek üzere R-Net isimli ağ tarafından değerlendirilir. R-Net, daha derin ve karmaşık bir konvolüsyonel ağ kullanarak, yüz bölgelerinin sınırlarını ve özellik noktalarını daha doğru bir şekilde belirler.

Son aşamada, O-Net yüzün kesin konumunu ve yönelimini belirlemek üzere yüz özellik noktalarını konumlandırır ve yüzü hizalar. O-Net, en detaylı ve yüksek çözünürlüklü katmandır ve yüz bölgelerinin küçük ayrıntılarına kadar incelenmesini sağlar.

MTCNN, çeşitli yüz ifadeleri, pozisyonları ve aydınlatma koşulları altında yüksek doğrulukla çalışabilen bir yüz tespit yöntemidir. Bu nedenle oldukça geniş bir uygulama alanına sahiptir ve birçok çalışmada tercih edilmektedir. MTCNN'in başarısı, çoklu görev öğrenme (multi-task learning) yaklaşımına dayanmaktadır ve bu sayede hem yüz tespiti hem de yüz hizalaması işlemleri aynı ağ tarafından öğrenilmek ve gerçekleştirilmektedir.

1.6.2. S3FD

S3FD (Single Shot Scale-invariant Face Detector), görüntülerde yer alan yüzlerin tespiti için geliştirilmiş başka bir etkili algoritmadır (S. Zhang vd., 2017). Derin öğrenme tabanlı olan bu yöntem, farklı büyüklüklerdeki yüzleri tek bir adımda tespit edebilmektedir. Bu model yüz tespitinde yüksek doğruluk ve hız sağlamak amacıyla tasarlanmıştır.

S3FD, farklı ölçekte yüz büyüklüklerini ve konumlarını işleyebilme yeteneği ile öne çıkar. Algoritma, farklı büyüklüklerdeki yüzleri ve bu yüzlerin konumlarını dikkate alarak,

çoklu çözünürlük düzeylerinde yüz tespiti yapabilen ölçeklenebilir bir yapıya sahiptir. S3FD'nin çalışma prensipleri şu şekildedir:

- Ölçeklenebilir Algılama: S3FD, çeşitli ölçeklerdeki yüzleri algılayabilmek için çoklu çözünürlük düzeylerinde çalışmaktadır. Bu, küçük, orta ve büyük yüzlerin tamamının aynı model tarafından yüksek doğrulukla tespit edilebilmesini imkân verir. Model, farklı çözünürlüklerdeki özellik haritalarını kullanarak yüz tespiti gerçekleştirmektedir.
- Anchors ve Feature Pyramid Network (FPN): S3FD, çeşitli ölçeklerde yüz tespiti yapabilmek için Anchor adı verilen sabit boyuta sahip tahmin kutuları kullanır. Farklı çözünürlük düzeylerindeki özellik haritaları üzerinde Anchor'lar uygulanarak, her bir çözünürlük düzeyinde yüz tespiti yapılır. Bu yaklaşım hem küçük hem de büyük yüzlerin aynı doğrulukla tespit edilmesini sağlamaktadır.
- Özellik Birleştirme: S3FD, farklı çözünürlük seviyelerindeki özellikleri birleştirerek her bir çözünürlük seviyesindeki tahminleri daha doğru hale getirmektedir. Bu, düşük çözünürlükteki özelliklerin yüksek çözünürlükteki özelliklerle birleştirilerek daha iyi bir genel tahmin yapılabilmesini sağlamaktadır.

Sahip olduğu özellikleri sayesinde S3FD, özellikle gerçek zamanlı yüz tespiti uygulamalarında ve değişken çevresel koşullar altında yüksek performans sunabilmektedir ve tercih edilen bir yöntem olmaktadır.

Giriş bölümünde, derin sahte video üretimi ve tespiti konusundaki temel kavramlar ve yöntemler ele alınmıştır. Sahte yüz görüntüsü üretim türleri, derin sahte video üretim teknikleri, tespit yöntemleri, derin öğrenme mimarileri ve yüz tespit algoritmalarının detaylı bir şekilde incelenmesi gerçekleştirilmiştir. Bir sonraki bölümde, yapılan çalışmalar başlığı altında, derin sahte video tespiti için bu tez çalışması kapsamında geliştirilen yöntemler ve bu yöntemlerin uygulanması detaylandırılacaktır. Ardından, bulgular ve tartışma bölümünde, yapılan çalışmalar sonucunda elde edilen sonuçlar analiz edilerek bu sonuçların literatürdeki yeri tartışılacaktır. Sonuçlar bölümünde, çalışmanın genel bulguları özetlenerek, bu bulguların ne anlama geldiği ve ne tür katkılar sağladığı ele alınacaktır. Son olarak, öneriler bölümünde, gelecekte yapılacak çalışmalar için yol gösterici öneriler ve bu alanda karşılaşılan zorlukların nasıl aşılabileceğine dair stratejiler sunulacaktır.

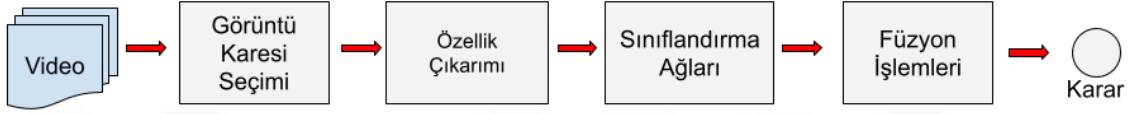
2. YAPILAN ÇALIŞMALAR

Derin sahte videoların neden olduğu kişisel ve toplumsal problemler nedeniyle videoların gerçek mi yoksa sahte mi olduğunun belirlenmesi ihtiyacı ortaya çıkmıştır. Bu amaçla pek çok sahte video tespit çalışması gerçekleştirilmektedir. Literatürdeki çalışmalar incelendiğinde bu çalışmaların bazı eksiklikleri olduğu görülmektedir. Belirlenen bu eksiklikler ve bunlara literatüre sağlanan katkılar aşağıda sıralanmaktadır;

- Derin sahte videoların yer aldığı genel kullanıma açık veri kümeleri içerdikleri örneklerin görsel kalitesine göre birinci ve ikinci nesil veri kümeleri olarak ayrılmaktadırlar. Birinci nesil veri kümelerinde yer alan örnek videoların görsel kalitesi düşük olduğundan dolayı ikinci nesil veri kümelerine kıyasla sahte videoların tespit edilmesi daha kolaydır. Bu nedenle birinci nesil veri kümeleri üzerinde tespit sonucu paylaşılan çalışmaların aksine bu çalışmada ikinci nesil veri kümeleri üzerinde çalışılmaktadır.
- Derin sahte videoların tespiti çalışmalarında tespit başarısı kıyasında değerlendirme ölçütü olarak genellikle Doğruluk ve AUC kullanılmaktadır. Çalışmaların çoğunluğunda bu iki ölçütten sadece birinde başarılı sonuç elde edilmektedir. Oysaki her iki ölçütte de yüksek başarılı elde edilebilir olması yöntemin sağlamlığını göstermesi açısından önem arz etmektedir. Bu duruma paralel olarak çalışmamızda her iki ölçüt için de önceki çalışmalardan daha iyi sonuç elde edebilen bir yöntem geliştirmek amaçlanmıştır.
- Derin sahte içerikleri otomatik olarak tespit etmeye yönelik geliştirilen pek çok yöntem bulunmaktadır, ancak bu yöntemlerin yüksek hesaplama maliyeti ve uzun işlem süreleri ile ilgili sorunları bulunmaktadır. Tespit maliyeti ve işlem süresinin azaltılmasına ihtiyaç bulunmaktadır. Videolarda yer alan görüntü karelerinin hangilerinin derin sahte tespit sistemlerinde kullanılacağına belirlenmesi gerekir. Tamamının kullanılması hesaplama maliyeti ve işlem süresinin artmasına neden olacaktır. Bu nedenle bu çalışmada yüzde altın oran bilgisi kullanılarak hangi görüntü karelerinin derin sahte tespiti için kullanılacağı belirlenmektedir. Bu yöntemle doğrulama performansını artırmak ve tespit süresini azaltmak için sahtecilik kanıtı

içerebilecek en uygun görüntü karesini seçmeye çalışılır. Deneysel çalışmalar, bu yöntemin diğer çalışmalara kıyasla daha az hesaplama maliyeti getirdiğini göstermektedir.

Bu çalışmada önerilen yöntem, derin sahte videoları tespit etmeyi amaçlayan dört aşamalı bir süreci içermektedir. Önerilen yöntemin blok tasarımı, dört ana aşamasını gösteren Şekil 17'de gösterilmektedir.



Şekil 17. Önerilen yöntemin temel aşamaları

İlk aşama, yüzdeki altın oran bilgisi kullanılarak videolardan sahtecilik tespitini yakalamayı sağlayacak en uygun tek görüntü karesinin seçilmesidir. Bu aşamada her video için en düşük altın orana sahip görüntü karesi seçilir. Bir sonraki aşamada, transfer öğrenme yönteminin bir uygulaması olarak VGG19, Efficient-Net-B0 ve Efficient-Net-B4 gibi önceden eğitilmiş ağ modelleri belirli aşamalarına kadar kullanılarak özellik çıkarımı gerçekleştirilir. Sonraki aşamada Capsule Forensics ve ArCapsNet gibi kapsül ağı tabanlı sınıflandırma ağları kullanılır ve modeller en iyi sonucu elde etmek için farklı parametrelerle eğitilerek sınıflandırma yapılır. Son olarak, en iyi performans gösteren modeller füzyon aşamasında birleştirilerek sınıflandırma başarımı daha da artırılır. Alt başlıklarda yöntemle ait aşamalar derinlemesine incelenecektir.

2.1. Yüzdeki Altın Oran Bilgisi Kullanılarak Çerçeve Seçimi

Derin sahte video tespiti sürecinde ilk aşama olarak videolardan görüntü karelerinin çıkarımı gerçekleştirilmelidir. Videolar çok sayıda görüntü karesinden oluşmaktadır. Tüm görüntü kareleri kullanılarak videonun sahte ya da gerçek olduğuna karar verilmesi, tespit için gereken hesaplama maliyetini ve işlem süresini oldukça fazla artıracaktır. Ayrıca sahte videolar için tüm görüntü karelerinde yer alan yüzler aynı kalitede olmayacağından dolayı her görüntü karesini eşit değerde tutmak uygun olmayacaktır. Sahtecilik tespiti için en uygun görüntü karelerinin belirlenmesi ve buna göre seçim yapılması büyük önem taşır.

Böylelikle sahtecilik kanıtını en fazla içerebilecek görüntü karesinin belirlenmesi ile tespit yapılması, sistemin genel performansını artıracaktır.

Bu amaçla yüzdeki altın oran bilgisi kullanılarak videolar için hangi görüntü karesinin seçileceği belirlenmektedir. Altın oran, matematiksel olarak yaklaşık 1.6180339887 değerine sahip olan ve estetik açıdan hoş kabul edilen bir orandır. Yüzdeki altın oran, gözler, burun ve ağız gibi yüz özellikleri arasındaki mesafelerin bu orana yakın olmasını ifade eder. İnsan yüzündeki bu oranlar, yüzün doğal ve simetrik görünmesini sağlar. Altın oran, antik çağlardan beri sanat ve mimaride kullanılan bir orandır ve doğada da sıkça gözlemlenir. Yüzdeki altın oranı elde etmek için bazı temel ölçümler yapılır:

- Gözler arasındaki mesafe: İki göz arasındaki mesafe, yüz genişliğine oranlandığında altın oranı verir.
- Burun ve ağız arasındaki mesafe: Burun tabanı ile ağızın üst kısmı arasındaki mesafe, yüz yüksekliğine oranlandığında altın oranı verir.
- Yüz genişliği ve yüksekliği: Yüz genişliği, yüz yüksekliğine oranlandığında altın oranı verir.

Bu ölçümler, yüzün simetrik ve estetik olarak hoş görünmesini sağlayan temel oranlardır.

Videolardan görüntü karelerinin çıkarımı sırasında, altın oran kriterlerine göre en düşük altın orana sahip yüzün bulunduğu kareler seçilir. Bu süreç, aşağıdaki adımları içerir:

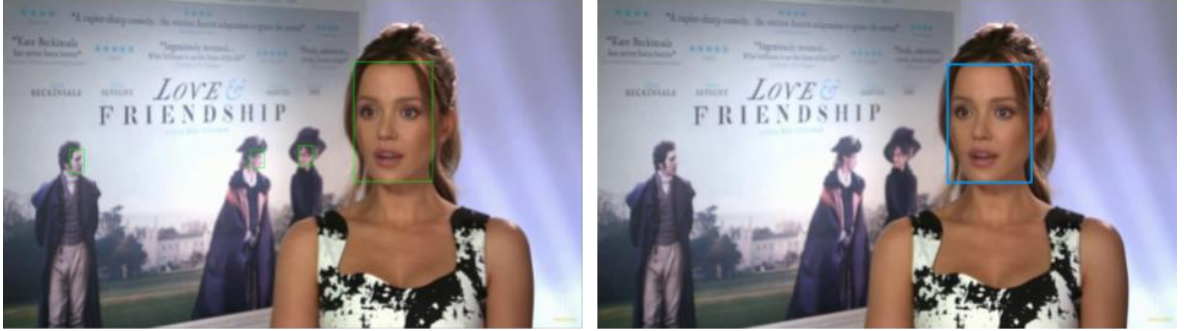
- Yüz Tespiti ve Yüzlerin Hizalanması: İlk olarak, videodaki tüm görüntü karelerinde yüz tespiti yapılır. Yüz tespiti için MTCNN veya S3FD gibi güçlü algoritmalar kullanılır. Tespit edilen yüzler, yüz hizalama algoritmaları ile hizalanır. Yüz hizalama, yüz özelliklerinin (gözler, burun, ağız) belirli bir standart pozisyonda olmasını sağlar.
- En Düz Bakan Yüzlerin Belirlenmesi: Hizalanan yüzler arasında en düz bakan yüzler belirlenir. Bu, yüzün en az eğimli ve en net görüldüğü kareleri seçmek anlamına gelir. En düz bakan yüzlerin belirlenmesi sonraki aşamada hesaplanacak olan altın oran değerinin doğruluğu üzerinde büyük öneme sahiptir.
- Altın Oran ile Yüz Seçimi: En düz bakan yüzler arasından, altın oran değeri en düşük olan seçilir. Bu yüzler, gözler, burun ve ağız arasındaki mesafelerin altın orana en uzak olduğu yüzlerdir. Bu süreç, yüz özellikleri arasındaki

mesafelerin hesaplanması ve altın orana ne kadar uygun olduğunun belirlenmesi ile gerçekleştirilir.

Bu aşamada yüz tespiti ve yüzlerin hizalanması, en düz bakan yüzlerin belirlenmesi ve altın oran ile yüz seçimi adımları uygulanarak her video için tek bir görüntü karesi belirlenmektedir. Bu yöntem, sahtecilik kanıtını en fazla içerebilecek görüntü karesinin belirlenmesini sağlayarak, tespit sürecinin genel performansını artırır ve işlem süresini optimize eder. Alt başlıklarda her bir aşama hakkında detaylar verilecektir.

2.1.1. Yüz Tespiti ve Hizalama

Yüz tespiti ve hizalama adımı, videolarda yer alan yüzlerin doğru bir şekilde tespit edilmesi ve hizalanması sağlanır. FaceSwap açık kaynak yazılımı içerisinde yüz tespiti ve yüzlerin hizalanması için MTCNN ve S3FD yüz tespit araçları ve FAN yüz hizalama aracı hazır olarak yer almaktadır (Bulat & Tzimiropoulos, 2017). Bu araçlar kullanılarak video karelerinden yüzler otomatik olarak çıkarılır ve hizalanır. Şekil 18’de MTCNN ve S3FD araçları ile tespit edilen yüzlerle ait örnek görseller yer almaktadır.



Şekil 18. Solda S3FD ile tespit edilen yeşil dikdörtgen kutular içine alınmış yüzler, sağda MTCNN ile tespit edilen mavi dikdörtgen kutu içine alınmış yüz

Şekil 18’de de görüldüğü gibi S3FD aracı MTCNN tarafından tespit edilemeyen küçük ölçekli yüzleri de tespit edebilmektedir. S3FD’nin küçük boyutlu yüzleri tespit edebilmedeki başarısı ile arka planda yer alan reklam panosu üzerindeki yüzleri de tespit edebildiği görülmektedir. Bu durum her ne kadar S3FD’nin başarısını kanıtlasa da derin sahte tespiti için gerekli olan asıl yüzün dışında diğer yüzleri de belirlemesi ve bunları çıkarması hatalı olacaktır. Bu nedenle ön işleme adımı MTCNN kullanımı tercih

edilmiştir. Örnek görselin Şekil 19’da verildiği yüz tespiti ve hizalama adımları şu şekilde özetlenebilir:

- Videodan yüzlerin çıkarılması: MTCNN kullanılarak yüzler tespit edilir ve 256x256 piksel boyutunda kırpılır.
- Yüz hizalama: Yüzler, FAN yüz hizalama aracı kullanılarak hizalanır. Bu, tüm yüzlerin doğru bir şekilde merkezlenmesini ve doğru oryantasyona getirilmesini sağlar.



Şekil 19. Ön işleme adımında yapılan işlemler. MTCNN ile yüzün tespit edilmesi ve 256x256 piksel boyutunda görüntü kırılması. Ardından FAN ile yüzün hizalanması

2.1.2. En Düz Bakan Yüzlerin Belirlenmesi

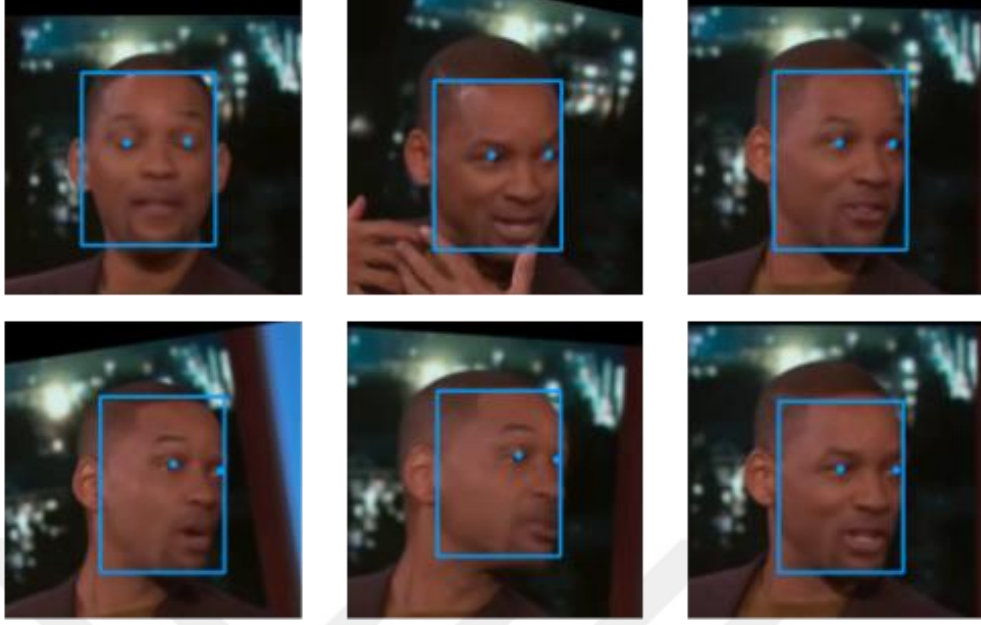
Videolardaki yüzlerin tespiti yapıp 256x256 piksel boyutlarında görüntü kareleri çıkarıldıktan sonra ikinci adım olan en düz bakan yüzlerin belirlenmesi adımına geçilir. Bir sonraki adımda yüzdeki altın oran bilgisinin hesaplanması yapılacaktır. Videolarda yer alan yüzler bazı durumlarda kameraya direkt bakmak yerine yan taraflara bakan durumda olabilir. Bu nedenle yüzün tamamı görünmeyebilir. Yüzün bir kısmının görünmediği ve tamamının görüldüğü örnek görseller Şekil 20’de gösterilmektedir.



Şekil 20. Solda yan tarafa baktığı için tamamı görünmeyen yüz, sağda tamamı görünen yüz

Yüzdeki altın oran hesabında yüzün tamamının görünür olması altın oran hesabının doğru yapılabilmesi için gerekliliktir. Bu nedenle videoda yer alan görüntü karelerindeki yüzlerin en fazla görünür olanlarının belirlenmesi ve bunlar üzerinde altın oran hesabının yapılması daha doğru olacaktır. Bu amaçla videodaki en düz bakan yüzlerin yer aldığı görüntü kareleri göz bebeklerinden görüntü karesi seçimi adımı belirlenir. Burada yapılan işlem adımları şu şekildedir:

- Göz bebekleri ve yüz kenar sınırlarının konumlarının belirlenmesi: MTCNN kullanılarak göz bebekleri ve yüz kenar sınırlarının konumları tespit edilir. Göz bebekleri ve yüz kenar sınırlarının konumlarının tespitine ilişkin örnekler Şekil 21’de gösterilmektedir.
- Göz bebeklerinin yüz kenarlarına olan mesafelerinin hesaplanması: Sol göz bebeğinin sol yüz kenarına ve sağ göz bebeğinin sağ yüz kenarına olan piksel mesafeleri hesaplanır.
- Oran hesaplaması: Daha küçük olan mesafenin, daha büyük olan mesafeye oranı hesaplanır.
- En uygun karelerin seçimi: En yüksek orana sahip k adet görüntü karesi seçilir.



Şekil 21. MTCNN ile göz bebekleri ve yüz kenar sınırlarının tespit edilmesi

Sol göz bebeğinin yüzün sol sınır kenarına olan piksel uzaklığı mesafesi dL_i ve sağ göz bebeğinin yüzün sağ sınır kenarına olan piksel uzaklığı mesafesi de dR_i olmak üzere

$$R_i = \min(dL_i, dR_i) / \max(dL_i, dR_i) \quad (1)$$

denklemlerle yüzün düzlük oranı belirlenmektedir. Burada i görüntü karesi numarasını, $\min$ ve $\max$ sırasıyla en küçük ve en büyük değeri geri döndüren fonksiyonları temsil etmektedir. Videoda yer alan tüm görüntü kareleri için R oranı belirlendikten sonra en yüksek orana sahip k sayıda görüntü karesi sonraki adımda kullanılmak üzere aşağıdaki gibi seçilir.

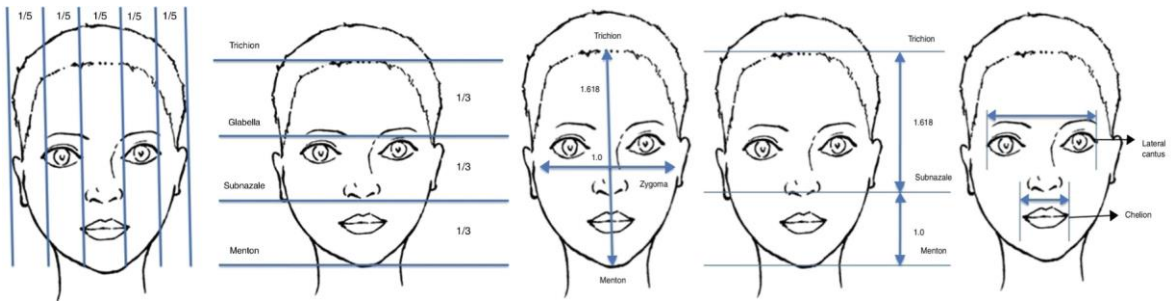
$$S = \{i_1, i_2, i_3, \dots, i_k\} \text{ ve } R_{i_1} \geq R_{i_2} \geq R_{i_3} \geq \dots \geq R_{i_k} \quad (2)$$

S seçilen k sayıdaki görüntü karelerini temsil ederken $i_1, i_2, i_3, \dots, i_k$ seçilen görüntü karelerinin numarasını temsil etmektedir. Bu çalışmada k sayısı 20 olarak belirlenmiştir.

2.1.3. Yüzdeki Altın Oran Bilgisinin Kullanımı

En düz bakan yüzü içeren 20 adet görüntü karesi belirlendikten sonra bu görüntü kareleri üzerindeki yüzlerde altın oranın hesaplandığı adıma geçilir. Altın oran yüz seçimi adımı, her videodan en uygun tek bir görüntü karesini seçmek için yüzün altın oran bilgileri kullanılır. Altın oran, yüzün yüksekliği ve genişliği gibi çeşitli yüz özellikleri kullanılarak hesaplanır. Yüzdeki altın oran hesaplamaları Şekil 22’de gösterilen şu temel ölçümlere dayanarak hesaplanmaktadır:

- **Gözler Arasındaki Mesafe:** Gözlerin birbirine yakın olan iç köşeleri arasındaki mesafe (d_1), gözlerin birbirine uzak dış köşeleri arasındaki mesafe (d_2) ile oranlanır. Altın oran, bu iki mesafenin oranı olarak tanımlanır: $d_2 / d_1 \approx 1.618$.
- **Yüzün Yüksekliği ve Genişliği:** Yüzün toplam yüksekliği (y_1) ve genişliği (g_1) arasındaki oran da altın oranı verir: $y_1 / g_1 \approx 1.618$.
- **Ağız ve Burun Arasındaki Mesafe:** Burun tabanı ile ağız arasındaki mesafe (d_3), burun ucu ile çene arasındaki mesafe (d_4) ile oranlanır: $d_4 / d_3 \approx 1.618$.
- **Gözler ve Ağız Arasındaki Mesafe:** Gözlerin orta noktası ile ağzın orta noktası arasında bulunan mesafe (d_5), gözlerin orta noktası ile burun arasındaki mesafe (d_6) ile oranlanır: $d_6 / d_5 \approx 1.618$.



Şekil 22. Yüzdeki altın oran hesabında kullanılan uzaklıklar (Kaya vd., 2019)

Bu temel ölçümler, yüzün çeşitli noktalarından alınan mesafelerle hesaplanarak altın oranın belirlenmesini sağlamaktadır. Açık kaynaklı bir yazılım olan GoldenFace yazılımı, bu ölçümleri otomatik olarak yapar ve her görüntü karesinde yer alan yüz için altın oran değeri hesaplayarak yüzün altın orana yüzdesel olarak ne kadar yakın olduğunu belirler.

(Aksoylu, 2021). Her video için altın oran değeri, en düz yüzlerin önceki adımda seçildiği 20 görüntü karesi üzerinde hesaplanır. En düşük altın oran değerine sahip yüzü içeren görüntü karesi video için seçilmektedir. Celeb-DF veri kümesinde yer alan derin sahte videoların bazıları için seçilen en düşük altın orana sahip görüntü karesi örnekleri Şekil 23'te gösterilmektedir.



Şekil 23. Celeb-DF veri kümesinden sahte videolar için seçilen en düşük altın orana sahip görüntü kareleri

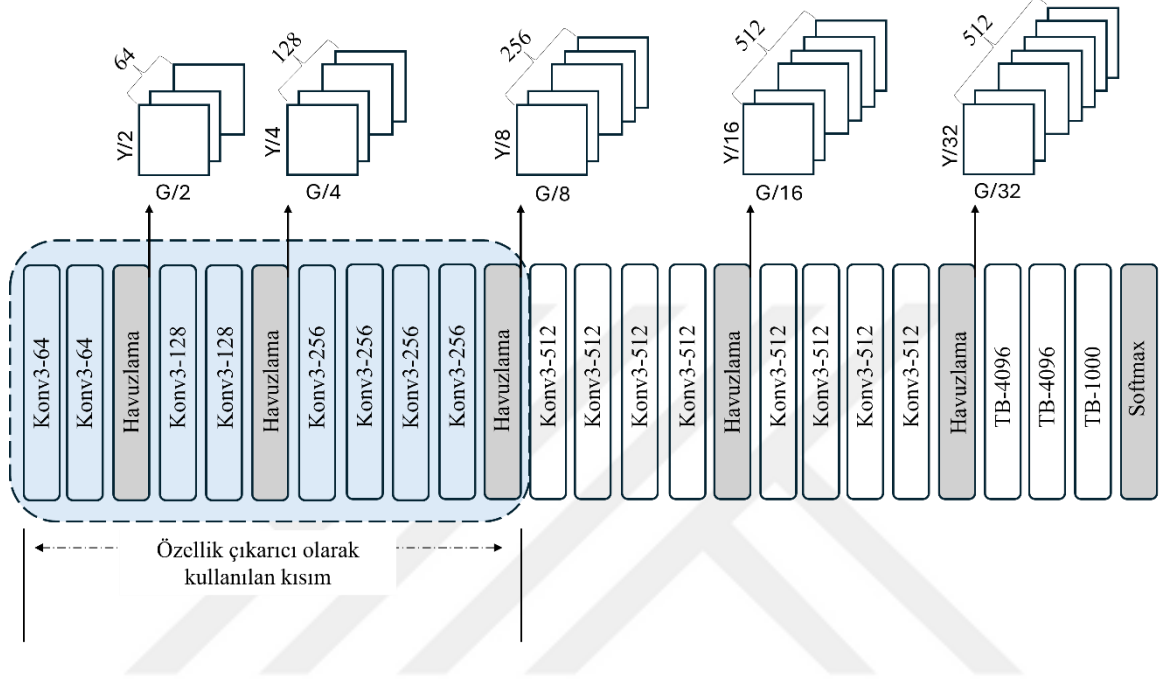
Şekilde verilen örneklerden görüldüğü gibi sahte videolarda yer alan en düşük altın oran değerine sahip yüzlerin sahte olduğunu anlamak çıplak gözle bile kolay olmaktadır. Bu durum sınıflandırma modellerinin tespit yeteneğinin de artmasını sağlayacaktır.

2.2. Öğrenme Aktarımı Yoluyla Özellik Çıkarıcıların Kullanılması

Öğrenme aktarımı, önceden büyük veriler üzerinde eğitilmiş modelleri yeni ve benzer görevler için tekrar eğitim yapmadan kullanma stratejisidir. Bu yöntem, derin öğrenme modellerinin daha hızlı ve etkili bir şekilde eğitilmesinde yardımcı olur. Özellikle büyük ve karmaşık veri kümeleri üzerinde eğitilmiş modellerden yararlanarak, küçük veri kümeleri üzerinde de yüksek performans elde edilmesi sağlanabilir. Bu bölümün alt başlıklarında, öğrenme aktarımı yoluyla VGG19, Efficient-Net B0 ve Efficient-Net B4 ağlarının özellik çıkarıcı olarak kullanılması açıklanacaktır.

2.2.1. VGG19 Ağının Özellik Çıkarıcı Olarak Kullanımı

VGG19 ağı kendi başına bir sınıflandırma modelidir. Bu çalışma kapsamında özellik çıkarıcı olarak kullanıldığından, yapısında yer alan tüm katmanlar kullanılmamaktadır. Şekil 24’te VGG19 ağının özellik çıkarıcı olarak kullanılan katmanları gösterilmektedir.



Şekil 24. VGG19 ağının özellik çıkarıcı olarak kullanılan katmanları

VGG19 ağı, 19 derin katmandan (16 konvolüsyonel ve 3 tam bağlantılı katman) oluşur. Bu çalışmada, modelin 256 kanallı çıkışın elde edildiği 3. havuzlama katmanına kadar olan kısmı kullanılmaktadır. Bu nedenle, giriş görselinden anlamlı özelliklerin çıkarılmasında VGG19 ağının ilk 8 konvolüsyonel katmanı ve 3 havuzlama (pooling) katmanı kullanılmaktadır.

VGG19 ağı, temel olarak 3x3 boyutunda filtreler kullanan konvolüsyonel katmanlardan oluşur. Bu filtreler, giriş görüntüsünden özellikleri çıkararak daha düşük boyutlu özellik haritaları oluşturur. İlk katmanlar, kenarları ve basit desenleri öğrenirken, derin katmanlar daha karmaşık ve yüksek seviyeli özellikleri öğrenir. Aşağıda, VGG19 ağının ilk 8 konvolüsyonel katmanının detayları verilmiştir:

- Conv3-64 × 2: 64 filtre, 3x3 çekirdek boyutu, padding=1, stride=1.
- Max Pooling: 2x2 çekirdek boyutu, stride=2.
- Conv3-128 × 2: 128 filtre, 3x3 çekirdek boyutu, padding=1, stride=1.
- Max Pooling: 2x2 çekirdek boyutu, stride=2.

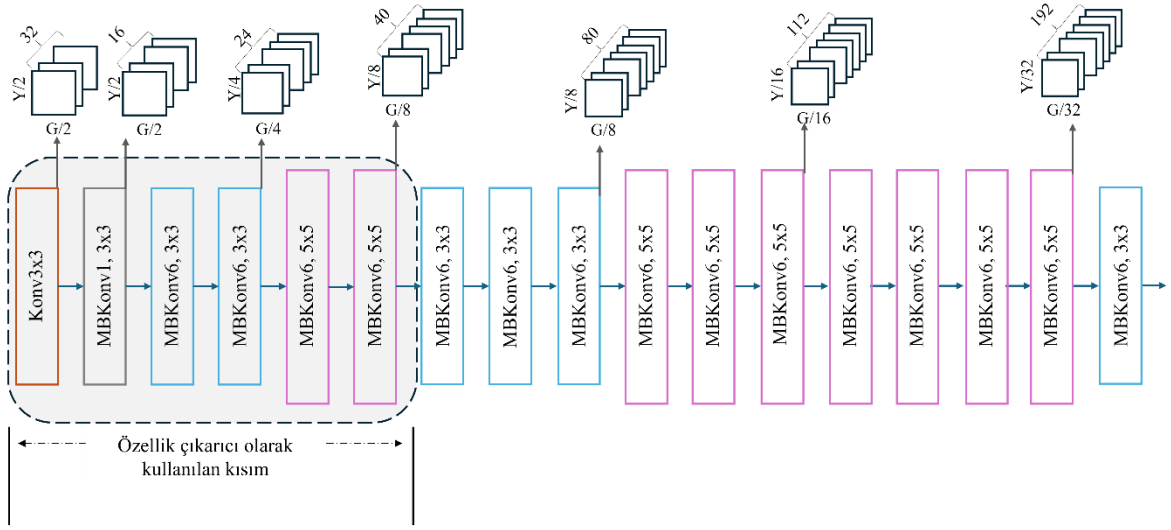
- Conv3-256 $\times$ 4: 256 filtre, 3x3 çekirdek boyutu, padding=1, stride=1.
- Max Pooling: 2x2 çekirdek boyutu, stride=2.

Tüm konvolüsyonel katmanlarda ReLU aktivasyon fonksiyonu kullanılır. ReLU, negatif değerleri sıfırlar ve pozitif değerleri olduğu gibi geçirir. Bu, ağı doğrusal olmayan ilişkileri öğrenmesini sağlar ve modeli daha hızlı ve etkili hale getirir. Kullanılan max-pooling katmanları, özellik haritalarının boyutunu düşürmek için kullanılır. Bu katmanlar, 2x2 çekirdek boyutu ve stride=2 kullanarak özellik haritalarının her bir bölgesindeki maksimum değeri seçer. Bu işlem, modelin hesaplama maliyetini düşürür ve overfitting'i önlemeye yardımcı olur. İlk üç havuzlama katmanından sonra elde edilen özellik haritaları, toplamda 256 kanal içermektedir. Bu 256 kanallı çıkış, VGG19 ağına ilk 8 konvolüsyonel katmanı ve 3 havuzlama katmanından süzülerek elde edilir. Bu özellik haritaları, giriş görselindeki anlamlı desenleri ve yapıları temsil eder. Elde edilen bu özellik haritaları, daha sonraki sınıflandırma aşamasında kullanılacak olan modele giriş olarak verilmektedir. Bu özellikler, derin sahte video tespiti gibi karmaşık görevlerde önemli bir rol oynar. Özellik haritaları, sınıflandırıcı modellere giriş olarak verilerek bu modellerin daha yüksek doğrulukta tahminler yapmasını sağlar.

2.2.2. Efficient-Net B0 Ağına Özellik Çıkarıcı Olarak Kullanımı

Efficient-Net ağı mimarisi ailesinin güçlü özellik çıkarıcı yanından faydalanmak üzere bu modelin B0 kodlu temel sürümünden kullanılmaktadır. Efficient-Net B0 modeli, verimli hesaplama ve yüksek doğruluk sağlamak üzere tasarlanmış olup çeşitli görüntü işleme görevlerinde üstün performans göstermektedir. Efficient-Net B0 modelinin özellik çıkarıcı olarak kullanılan kısmına ait katman mimarisi Şekil 25'te gösterilmektedir.

Efficient-Net B0 modelini özellik çıkarıcı olarak kullanırken, 40 kanallı çıkışı üreten 4. blok katmanına kadar olan kısım alınmaktadır. Bu katmanlar, ham görüntü verisinden anlamlı ve kullanışlı özelliklerin çıkarılması için optimize edilmiştir. Efficient-Net B0'ın bu bölümü, görsel verilerden hem düşük seviyeli hem de daha karmaşık olan özellikleri etkili bir şekilde çıkarmaktadır.



Şekil 25. Efficient-Net B0 ağının özellik çıkarıcı olarak kullanılan katmanları

Efficient-Net B0, temelinde MobilNetV2 tabanlı genişletilmiş konvolüsyonlar kullanır. Bu konvolüsyonlar, daha düşük hesaplama maliyeti ile daha yüksek performans sağlamak üzere tasarlanmıştır. Genişletilmiş konvolüsyonlar, giriş özellik haritalarını genişleterek kanal sayısını artırır ve ardından daraltarak önemli özellikleri çıkarır. Aktivasyon fonksiyonu olarak Swish kullanılmaktadır. Swish, doğrusal olmayan ve ReLU'dan daha üstün performans gösteren bir aktivasyon fonksiyonudur. Swish fonksiyonu, $f(x) = x/(1 + e^{-x})$ formülüne sahiptir ve ağın öğrenme kapasitesini artırarak daha karmaşık desenleri öğrenmesini kolaylaştırmaktadır. Efficient-Net B0, katmanlarını hafif ve verimli olacak şekilde tasarlamıştır. Özellik çıkarıcı olarak kullanılan ilk dört blok katmanına ait yapının detayları aşağıdaki gibidir.

- Blok 1: 32 filtre, 3x3 konvolüsyon, stride=2.
- Blok 2: 16 filtre, genişletme faktörü=1, derinlikli ayırık konvolüsyon, stride=1.
- Blok 3: 24 filtre, genişletme faktörü=6, derinlikli ayırık konvolüsyon, stride=2.
- Blok 4: 40 filtre, genişletme faktörü=6, derinlikli ayırık konvolüsyon, stride=2.

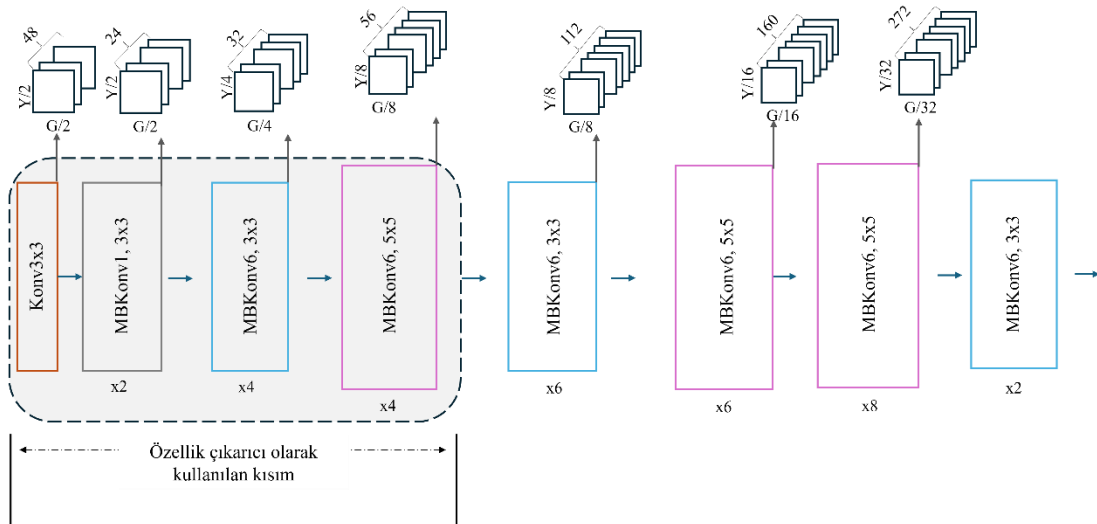
Efficient-Net B0, konvolüsyonel katmanların çıktısını daha anlamlı hale getirmek için havuzlama (pooling) katmanları kullanır. Bu katmanlar, özellik haritalarının boyutunu azaltarak hem hesaplama maliyetini düşürür hem de modelin genelleme yapabilme

kapasitesini artırır. 4. blokta elde edilen 40 kanallı çıkış, daha sonraki işleme adımlarında kullanılmak üzere elde edilmiştir. İlk dört blok, giriş görselinden çeşitli ölçeklerde ve karmaşıklıklarda özellikler çıkarır. Bu özellikler, ham görüntü verisinden anlamlı ve kullanışlı özelliklerin çıkarılmasını sağlar. Bu özellikler daha sonra derin sahte video tespiti için ileri işleme adımlarına giriş olarak verilecektir.

Özetle, Efficient-Net B0 modelinin özellik çıkarıcı olarak kullanımı, modelin ilk dört blok katmanını içerir. Bu katmanlar, giriş görselinden anlamlı ve kullanışlı özellikler çıkararak daha sonraki işleme adımlarında kullanılacak olan özellik haritalarını üretir. Bu yaklaşım, derin sahte video tespiti gibi görevlerde modelin performansını artırmak için etkili bir yöntemdir.

2.2.3. Efficient-Net B4 Ağının Özellik Çıkarıcı Olarak Kullanımı

Efficient-Net B4, görsel verilerden zengin ve detaylı özellikler çıkaran gelişmiş bir derin öğrenme modelidir. Bu model, yüz tanıma, nesne tespiti ve sahtecilik tespiti gibi görevlerde üstün performans gösterir. Efficient-Net B4'ün özellik çıkarıcı katmanları, büyük veri kümeleri üzerinde önceden eğitilmiş haliyle öğrenme aktarımı stratejisi kullanılarak daha küçük ve benzer veri kümeleri üzerinde yeniden kullanılabilir. Modelin katman mimarisi ve kanal sayılarına ilişkin bilgiler Şekil 26'da gösterilmektedir.



Şekil 26. Efficient-Net B4 ağının özellik çıkarıcı olarak kullanılan katmanları

Bu çalışmada, Efficient-Net B4 modeli özellik çıkarıcı olarak kullanılırken 4. blok katmanına kadar olan kısım alınmaktadır. Bu katmanlar, ham görüntü verisinden anlamlı ve kullanışlı özelliklerin çıkarılmasını sağlamaktadır. Efficient-Net B4'ün bu bölümü, görsel verilerden hem düşük seviyeli hem de daha karmaşık olan özellikleri etkili bir şekilde çıkarır. Elde edilen özellik verisi 56 kanallıdır. Bu özellik verileri, sonraki aşamada kullanılacak olan sınıflandırıcı ağılara giriş olarak verilmektedir.

Efficient-Net B4 modelinde, B0 modelinde olduğu gibi temelinde MobilNetV2 tabanlı genişletilmiş konvolüsyonlar kullanır ve aktivasyon fonksiyonu olarak Swish kullanılır. Efficient-Net B4 modelinin kullanılan katmanları şu şekildedir:

- Blok 1: 32 filtre, 3x3 konvolüsyon, stride=2.
- Blok 2: 24 filtre, genişletme faktörü=1, derinlikli ayırık konvolüsyon, stride=1.
- Blok 3: 40 filtre, genişletme faktörü=6, derinlikli ayırık konvolüsyon, stride=2.
- Blok 4: 56 filtre, genişletme faktörü=6, derinlikli ayırık konvolüsyon, stride=2.

Özet olarak, Efficient-Net B4 modelinin özellik çıkarıcı olarak kullanımı, modelin ilk dört blok katmanını içerir. Bu katmanlar, giriş görselinden anlamlı ve kullanışlı özellikler çıkararak, daha sonraki işleme adımlarında kullanılacak olan yüksek kaliteli özellik haritaları üretir. Bu yaklaşım, derin sahte video tespiti gibi görevlerde modelin performansını artıran bir etki göstermektedir.

2.3. Kapsül Ağı Tabanlı Sınıflandırma Ağları Kullanılarak Sınıflandırma Yapılması

Kapsül ağı tabanlı sınıflandırma ağları, geleneksel yapay sinir ağlarının ötesine geçerek daha yüksek doğruluk ve güvenilirlik sunmayı amaçlayan yenilikçi modeller olarak öne çıkmaktadır. Bu modeller, verinin içsel ilişkilerini ve hiyerarşik yapısını daha iyi modelleyerek, özellikle karmaşık sınıflandırma görevlerinde başarılı performans göstermektedirler. Bu bölümde, Capsule Forensics ve ArCapsNet mimarilerinin sınıflandırma görevlerinde nasıl kullanıldığına ilişkin açıklamalar verilecektir.

2.3.1. Capsule Forensics Mimarisinin Kullanımı

Capsule Forensics, orijinal kapsül ağı mimarisinden uyarlanmış bir versiyondur ve derin sahte video tespiti için yeniden düzenlenmiştir. Bu mimari, birincil kapsüller ve iki çıkış kapsülünden (gerçek ve sahte) oluşur. Her birincil kapsül, 2B konvolüsyonel bir parça, istatistiksel bir havuzlama katmanı ve 1B konvolüsyonel bir parçadan oluşur. Birincil kapsüllerin çıktıları, dinamik yönlendirme algoritması kullanılarak çıkış kapsüllerine yönlendirilir. Çıkış kapsüllerinin aktivasyonlarına dayanarak nihai sonuç hesaplanır. Capsule Forensics ağına temel bileşenleri şunlardır:

- **Birincil Kapsüller (Primary Capsules):** Bu kapsüller, 2B konvolüsyonel katman, istatistiksel havuzlama katmanı ve 1B konvolüsyonel katmandan oluşur. Bu katmanlar, giriş verisinden temel özellikleri çıkarır ve bu özellikleri yüksek boyutlu kapsüller şeklinde temsil eder.
 - **2B Konvolüsyonel Katman:** Bu katman, giriş verisinden ilk özellik haritalarını çıkarır. 32 filtre, 3x3 çekirdek boyutu ve stride=1 kullanılır.
 - **İstatistiksel Havuzlama Katmanı:** Bu katman, ağı giriş görüntü boyutundan bağımsız olmasına yardımcı olur. Bu, mevcut olan ağı yeniden tasarlamak zorunda kalmadan, farklı giriş boyutlarına sahip farklı problemlere uygulanabileceği anlamına gelir. İstatistiksel havuzlama katmanında, her filtrenin ortalaması ve varyansı hesaplanır.
 - **1B Konvolüsyonel Katman:** Bu katman, havuzlama katmanından gelen özellikleri daha yüksek boyutlu kapsüller şeklinde temsil eder. 128 filtre, 1x1 çekirdek boyutu ve stride=1 kullanılır.
- **Dinamik Yönlendirme Algoritması (Dynamic Routing Algorithm):** Bu algoritma, birincil kapsüllerden gelen çıktıları çıkış kapsüllerine yönlendirir. Her bir kapsül, hangi çıkış kapsülüne bağlanacağını ve ne kadar katkıda bulunacağını belirler. Bu süreç, iteratif olarak gerçekleştirilir ve her iterasyon, kapsüller arasındaki bağlanma olasılıklarını günceller.
- **Çıkış Kapsülleri (Output Capsules):** Bu kapsüller, gerçek ve sahte sınıfları temsil eder. Her çıkış kapsülü, bir nesnenin veya özelliğin belirli bir yönünü temsil eden yüksek boyutlu vektörler üretir. Bu vektörler, nesnenin varlığı ve

özellikleri hakkında bilgi taşır. Çıkış kapsüllerinin aktivasyonları, nihai sınıflandırma sonucunu belirler.

Capsule Forensics ağına veri girişi yapılmadan önce özellik çıkarımı aşaması yer almaktadır. Orijinalinde özellik çıkarıcı olarak VGG19'un kullanıldığı mimaride bu çalışmada ek olarak Efficient-Net B0 ve B4 modelleri de özellik çıkarıcı olarak kullanılmıştır. Önceden eğitilmiş modellerin tekrar eğitilmeden özellik çıkarıcı olarak kullanılması, eğitim aşamasında sınıflandırıcı ağıın örnek verileri aşırı öğrenmesi problemini önleyecektir. Capsule Forensics mimarisinin eğitimi aşamasında kullanılan parametrelere ve eğitim detaylarına “Önerilen Ağ Mimarilerinin Eğitimi” başlığında değinilecektir.

2.3.2. ArCapsNet Mimarisinin Kullanımı

ArCapsNet, orijinal kapsül ağından uyarlanmış bir başka versiyondur. Bu çalışmada derin sahte tespiti için kullanılmaktadır. Bu mimari, dinamik yönlendirme ve squash aktivasyon fonksiyonlarını, dikkat yönlendirme ve kapsül aktivasyonu ile değiştirir. ArCapsNet, birincil kapsüller katmanı, konvolüsyonel kapsüller katmanı ve tam konvolüsyonel kapsüller katmanından oluşur. Her katmanın çıktısına 1x1 konvolüsyon ve tanh aktivasyonu uygulanır. ArCapsNet mimarisinin temel bileşenleri şunlardır:

- Birincil Kapsüller Katmanı (Primary Capsule Layer): Bu katman, giriş görüntüsünden temel özellikleri çıkarır ve bu özellikleri kapsüller şeklinde temsil eder. 64 filtre, 3x3 çekirdek boyutu ve stride=1 kullanılır.
- Konvolüsyonel Kapsüller Katmanı (Convolutional Capsule Layer): Bu katman, birincil kapsüllerden gelen özellikleri daha karmaşık yapılar şeklinde temsil eder. Bu süreçte, konvolüsyonel kapsüller, giriş özelliklerini daha yüksek seviyede özellikler şeklinde çıkarır.
- Tam Konvolüsyonel Kapsüller Katmanı (Fully Convolutional Capsule Layer): Bu katman, giriş görüntüsünden en karmaşık özellikleri çıkarır ve bu özellikleri yüksek boyutlu kapsüller şeklinde temsil eder.

Orijinal ArCapsNet mimarisinde, ham görüntü verisinden özellik çıkaran tek katmanlı bir CNN bulunurken, bu çalışmada bu kısım VGG19, Efficient-Net B0 ve

Efficient-Net B4 gibi daha büyük ve güçlü özellik çıkarıcı mimarilerle değiştirilmiştir. Bu değişiklik, modelin performansını ve doğruluğunu önemli ölçüde artırmaktadır. Önceden eğitilmiş hallerinin kullanıldığı bu özellik çıkarıcı mimariler, ArCapsNet mimarisinin eğitim aşamasında aşırı öğrenme problemi yaşamasını önler. Bu mimariler, büyük ve çeşitli veri kümeleri üzerinde eğitildiğinden, geniş bir genelleme yeteneğine sahiptir ve derin sahte tespiti görevinde performansı artırıcı etki gösterirler. Eğitim aşaması ile ilgili detaylı bilgiler “Önerilen Ağ Mimarilerinin Eğitimi” başlığında verilecektir.

2.4. Önerilen Ağ Mimarilerinin Eğitimi

Bu bölümde, Capsule Forensics ve ArCapsNet mimarileri kullanılarak yapılacak olan sınıflandırmada ağların eğitimi ile ilgili detaylar paylaşılacaktır. Her iki ağın eğitimi ile ilgili bilgiler, parametre ve yapılandırma bilgileriyle birlikte açıklanacaktır. Capsule Forensics ve ArCapsNet mimarileri tanıtıldıkları orijinal parametreleri ile benzer şekilde eğitilmektedir. Bu tez çalışmasının kapsamında yürütülen tüm deneyler 64 bit Windows 11 işletim sistemi, Intel i5 13600K işlemci, 32GB DDR5 Ram ve Nvidia GeForce RTX 3080 10GB ekran kartına sahip bir sistem üzerinde gerçekleştirilmiştir. Tüm modeller Python programlama dili ve Pytorch kütüphanesi kullanılarak kodlanmış ve gerçekleştirilmiştir.

Capsule Forensics model mimarisi orijinalinde özellik çıkarıcı olarak VGG19 ağının bir kısmını kullanmaktadır. Özellik çıkarıcı ağının çıkış verisinin kanal sayısı 256’dır. Dolayısıyla kapsül ağı kısmının girişi 256 kanallı veriyi alır. Bu çalışma kapsamında orijinal mimariden farklı olarak VGG19 dışında Efficient-Net B0 ve B4 modelleri de özellik çıkarıcı olarak kullanıldığından dolayı bu modellerin çıkış verileri kapsül ağının giriş kanalı sayısı ile uyumamaktadır. Efficient-Net B0 modeli 40 kanallı çıkışa sahip iken B4 modeli 56 kanallı çıkış verisine sahiptir. Bu nedenle özellik çıkarıcı olarak Efficient-Net modelleri kullanılırken kapsül ağının girişine uygun kanal sayısına sahip veri üretebilmek için özellik çıkarıcı kısmın peşine 1x1 CNN bölümü eklenmektedir. Bu CNN bölümü veriyi 256 kanallı hale getirmeye yarar ve kapsül ağının işleyebileceği formata getirmiş olur.

Capsule Forensics ağının yapılandırma detayları şu şekildedir:

- Özellik çıkarıcı bölüm: VGG19 (256 kanallı çıkış) veya Efficient-Net B0 (40 kanallı çıkış) veya Efficient-Net B4 (56 kanallı çıkış).

- 1x1 CNN bölümü: Efficient-Net B0 ve B4 modelleri kullanıldığında veriyi 256 kanala getirmek için kullanılır.
- Birincil kapsüller: 2D konvolüsyon, istatistiksel havuzlama ve 1D konvolüsyon parçalarından oluşur.
- Çıkış kapsülleri: Gerçek ve sahte sınıfları temsil eder.

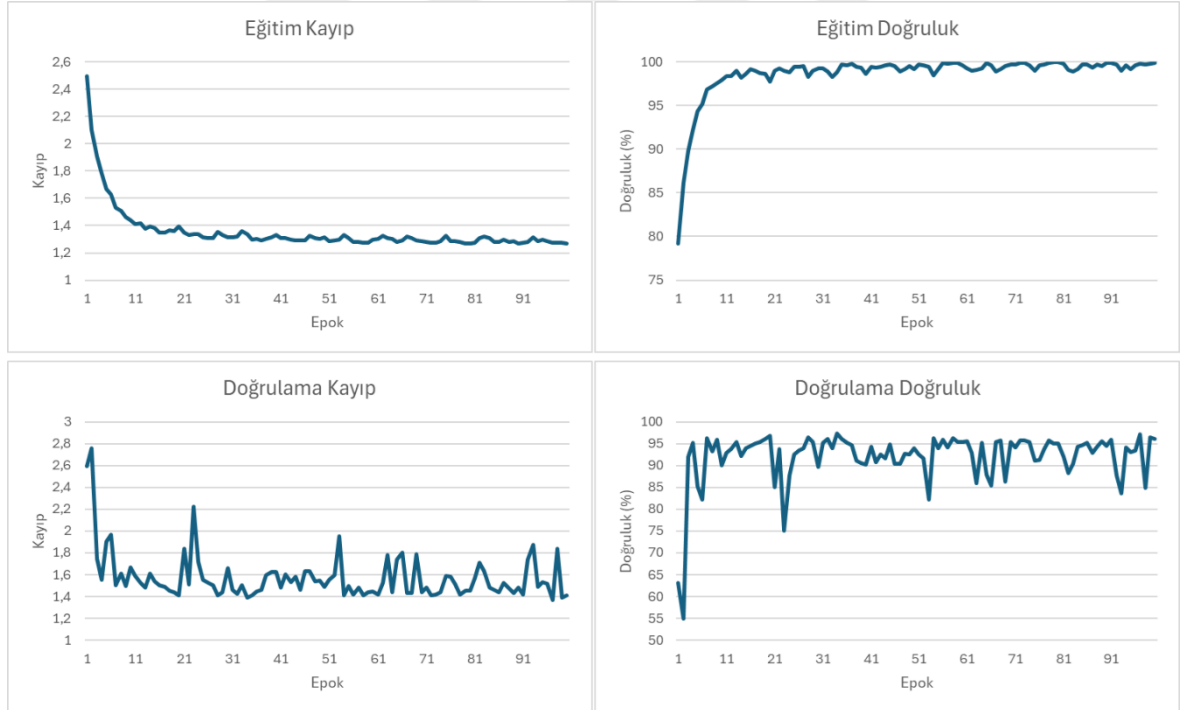
Capsule Forensics ağının eğitiminde yakınsama hızı ve hassasiyet arasında bir denge kurmak sağlayacak şekilde öğrenme oranı 5×10^{-4} olarak belirlenmiştir. $\beta_1 = 0.9$ değeriyle adam optimize edici eğitim sürecinde kullanılmaktadır. Ayrıca, eğitim sırasında sinir ağı birimlerinin küçük bir bölümünü rastgele olarak devre dışı bırakarak aşırı uyumu önlemekte fayda sağlayan %5'lik bir bırakma oranı uygulanmıştır. Bu oran, ağın belirli nöronlarını rastgele devre dışı bırakarak, modelin eğitim verilerine aşırı uyum sağlamasını önler. Kayıp hesabında çapraz entropi kullanılmaktadır. Eğitimde kullanılan veri kümelerinde yer alan örnekler sahte ve gerçek sınıfları için dengeli sayıda değiller. Veri kümelerindeki sınıf dengesizlikleri dikkate alınarak, kayıp fonksiyonunda sınıf ağırlıkları ayarlanmıştır. Bu daha az örneğe sahip sınıfların ağırlıklarının artırılması ve daha fazla örneğe sahip sınıfların ağırlıklarının azaltılması anlamına gelir.

ArCapsNet model mimarisinin orijinal halinde kapsül kısmı öncesinde giriş görüntüsünü hazırlamak ve özellik çıkarmak için bir CNN katmanı kullanılmaktadır. Bu CNN katmanı 64 kanallı çıkış üretmektedir ve kapsül kısmı bu 64 kanallı veriyi girdi olarak almaktadır. Bu katman yerine bu çalışmada VGG19 ve Efficient-Net B0, B4 modelleri özellik çıkarıcı olarak kullanılmaktadır. VGG19 modelinin özellik çıkarıcı olarak kullanılan kısmının çıkış verisinin kanal sayısı 256, Efficient-B0 modelinin özellik çıkarıcı kısmının çıkış verisinin kanal sayısı 40 ve Efficient-Net B4 modelinin özellik çıkarıcı olarak kullanılan kısmının çıkış verisinin kanal sayısı 56'dır. ArCapsNet model mimarisinin kapsül kısmının giriş olarak aldığı veri 64 kanallı olduğundan özellik çıkarıcı kısmından gelecek olan verilerin kanal sayısını 64'e eşitlemek gerekecektir. Bu nedenle özellik çıkarıcı kısmın peşine 1x1 CNN bölümü eklenmektedir. Bu bölüm özellik çıkarıcı kısmından gelen verileri 64 kanala sahip olacak şekilde değiştirerek kapsül kısmına iletir. ArCapsNet model mimarisinde bulunan konvolüsyonel kapsül katmanları tekrarlanabilir yapıdadır. Bu katmanların sayıları 1 ile 3 arasında değişmektedir. ArCapsNet modelinin eğitiminde öğrenme oranı 1×10^{-3} olarak ayarlanmış ve adam optimize edici $\beta_1 = 0.9$ ile kullanılmıştır.

ArCapsNet modelinin yapılandırma detayları şu şekildedir:

- Özellik çıkarıcı: VGG19 (256 kanallı çıkış) veya Efficient-Net B0 (40 kanallı çıkış) veya Efficient-Net B4 (56 kanallı çıkış)
- 1x1 CNN bölümü: Özellik çıkarıcı modellerinden gelen veriyi 64 kanallı hale getirmek için kullanılır.
- Birincil kapsüller katmanı: Görüntüden temel özellikleri çıkaran ve kapsül aktivasyonunu uygulayan katman.
- Konvolüsyonel kapsüller katmanı: Daha derin ve geniş özellikler çıkaran katmanlar. Bu katmanlar tekrarlanabilir yapıdadır ve 1 ile 3 arasında değişir.
- Tam konvolüsyonel kapsüller katmanı: 1x1 konvolüsyon ve tanh aktivasyonu kullanılarak çıktıları işleyen katmanlar.

Eğitim aşamalarında tüm modeller için mini-paket boyutu 32 olarak ayarlanmıştır. Şekil 27’de Celeb-DF veri kümesinde en iyi sonucu elde eden modele ait eğitim ve doğrulama aşamalarına ait kayıp ve doğruluk grafikleri görülmektedir.



Şekil 27. Eğitim ve doğrulama aşamalarında kayıp ve doğruluk grafikleri

Şekil 27’de görülebildiği gibi 100 epok düzeyi civarında sonuçlarda iyileşme durmaktadır. Bu nedenle tüm modeller 100 epok boyunca eğitilmiştir.

2.5. Ağ Mimarilerinin Çıkışlarının Füzyonunun Yapılması

Derin sahte tespiti gibi karmaşık görevlerde, tek bir modelin doğruluğu ve güvenilirliği genellikle yeterli olmayabilir. Bu nedenle, farklı ağ mimarilerinin çıkışlarının birleştirilmesi (füzyon) daha sağlam ve güvenilir sonuçlar elde etmek için kullanılan etkili bir yöntemdir. Bu bölümde, Capsule Forensics ve ArCapsNet modellerinden elde edilen çıkışların nasıl birleştirildiği ve bu füzyonun derin sahte tespitindeki performansı nasıl artırdığı detaylı bir şekilde açıklanmaktadır.

Füzyon stratejileri, farklı modellerden elde edilen bilgileri birleştirerek daha doğru ve sağlam kararlar almayı amaçlar. Bu stratejiler genellikle iki ana kategoriye ayrılır: özellik seviyesi füzyonu ve skor seviyesi füzyonu. Özellik seviyesi füzyonu, modellerin ara katmanlarından elde edilen özelliklerin birleştirilmesiyle gerçekleştirilir. Bu yaklaşım, daha zengin ve kapsamlı bir özellik seti elde edilmesini sağlar, ancak bu yöntem yüksek hesaplama maliyetlerine neden olabilir. Skor seviyesi füzyonu her modelin bağımsız olarak verdiği karar skorlarının birleştirilmesiyle daha güvenilir sonuçlar elde edilmesini sağlar. Bu çalışmada, skor seviyesi füzyon stratejisi benimsenmiştir. Skor seviyesinde yapılan füzyon işlemleri üç farklı şekilde gerçekleştirilebilmektedir.

- Minimum Füzyon: Korumacı bir yaklaşımdır ve sadece sınıflandırıcılar arasında fikir birliği olduğunda karar alınır. Sınıflandırıcılar sınıflandırma kararı verirken en düşük skoru veren sınıflandırıcının kararı kabul edilir.
- Maksimum Füzyon: Agresif bir yaklaşımdır. Sınıflandırıcılar sınıflandırma kararı verilirken en yüksek skoru veren sınıflandırıcının kararı kabul edilir.
- Ortalama Füzyon: Tüm sınıflandırıcıların ortak görüşünü dikkate alarak dengeli bir yaklaşım sunar ve sınıflandırıcıların kendi içindeki hatalara karşı daha dayanıklı olmasını sağlar. Tüm sınıflandırıcıların sınıflandırma skorlarının ortalaması alınarak nihai skor hesaplanır.

Füzyon stratejisinin seçimi, ele alınan problemin doğasına, bireysel sınıflandırıcıların özelliklerine ve karar verme sürecinde korumacı veya agresif bir yaklaşım arasındaki dengeye bağlıdır. Örneğin, ikili sınıflandırma probleminde yanlış pozitif veya yanlış negatiflerin farklı maliyetlere sahip olması durumunda, belirli bir hata türünü en aza indiren bir füzyon yöntemi tercih edilebilir. Diğer yandan, çok sınıflı bir problemde genel doğruluk veya farklı sınıf bazlı performans ölçütleri arasında bir denge önceliklendirilebilir. Bu çalışmada, çeşitli model füzyon stratejileri incelenmiş ve ortalama

füzyon stratejisinin, derin sahtekarlık tespiti görevimizde en iyi performansı sağladığı tespit edilmiştir. Bu yöntem, her bir modelin tahminlerini tek bir birleşik skora dönüştürerek dengeli ve sağlam bir sonuç elde edilmesini sağlar. Ortalama füzyon, basit fakat etkili bir yöntemdir. Bu yöntemde, N adet modelden elde edilen bireysel sınıflandırma skorları $(S_1, S_2, \dots, S_N)$ şu şekilde birleştirilir:

$$F = \frac{1}{N} \sum_{i=1}^N S_i \quad (3)$$

Bu denklemde, tüm bireysel skorlar toplanır ve toplam skor sayısına (N) bölünerek ortalama füzyon skoru (F) elde edilir. Ortalama füzyon, modellerin tahminlerinin harmanlanmasını sağlayarak dengeli bir sonuç ortaya çıkarır.

Bu tez çalışması kapsamında modellerin ikili ve üçlü füzyonunun yapıldığı farklı senaryolar uygulanmıştır. Her iki füzyon senaryosu ile ilgili kapsamlı deneysel sonuçlar ve bilgiler sonraki bölümlerde paylaşılmaktadır.

2.6. Kapsül Ağlarının Ses Sahteciliği Tespitinde Kullanımı

Derin sahte videoların tespitinde bu tez çalışması kapsamında kullanılan kapsül ağ modelleri olan Capsule Forensics ve ArCapsNet sınıflandırma ağları ayrıca ses sahteciliği tespitinde de kullanılmıştır. Günümüzde ses üretimi ve manipülasyonu, ticari veya ücretsiz gelişmiş ses düzenleme yazılımları kullanılarak kolaylıkla yapılabilir hale gelmiştir. Ses sahteciliğinde uygulanan yöntemlerden olan ses birleştirme ve kopyala-yapıştır sahtecilikleri için tespit yöntemleri uygulanmıştır.

Ses birleştirme sahteciliği, iki farklı ses kaydının birleştirilerek orijinal ses kaydının içeriğinin değiştirilmesi işlemidir. Bu tür manipülasyonlar, ses kayıtlarının bütünlüğünü tehlikeye atmaktadır. Örneğin, "Olay günü Ali'yi gördüm" cümlesini içeren bir ses kaydı, aynı konuşmacının başka bir zaman diliminde kaydedilmiş olan "Ali ve Ayşe'yi görmedim" cümlesi ile birleştirilebilir. Bu birleştirme sırasında ikinci kayıttaki "ve Ayşe" kısmı alınarak ilk kayda eklenir ve "Olay günü Ali ve Ayşe'yi gördüm" şeklinde sahte bir ses kaydı oluşturulur. Bu sahtecilik işlemi orijinal ses kayıtlarının anlamsal içeriğini kısmen veya tamamen çarpıtabilmektedir.

Bu çalışma kapsamında, ses birleştirme sahteciliği tespiti için kullanılan veri kümesi, TIMIT veri kümesinde yer alan ses kayıtlarından oluşturulmuştur. TIMIT veri kümesi, 438

erkek ve 192 kadın konuşmacıdan toplam 6300 ses kaydı içermektedir. Her ses kaydı yaklaşık olarak 2 ila 6 saniye uzunluğundadır. Bu çalışmada kullanılan 3 saniyelik sahte ses veri kümesi, TIMIT veri kümesindeki orijinal ses dosyalarının 1 ve 2 saniyelik parçalara bölünmesiyle oluşturulmuştur. 3 saniyelik sahte sesler, 2 saniyelik konuşma kayıtlarının ortasına 1 saniyelik ses kayıtlarının eklenmesiyle üretilmiştir. Veri kümesi, eğitim ve test setlerine ayrılmıştır. Toplamda, 3112 gerçek ve 6000 sahte ses eğitim için, 1145 gerçek ve 1000 sahte ses ise test için kullanılmıştır.

Tespit için kullanılan yöntem iki aşamadan oluşmaktadır. İlk aşama ön işleme aşamasıdır. Bu aşamada giriş olarak verilen ses dosyası bir kokleagram görüntüsüne dönüştürülür. Kokleagram, insan kokleasının frekans seçiciliğine dayalı frekans bileşenlerini temsil eder ve bu bileşenler gammatone filtresi kullanılarak elde edilir (Patterson vd., 1992). Gammatone filtreleme işlemi kullanılarak her frekans kanalı için pencerelemiş sinyalin enerjisi toplanır ve bir spektrogram görüntüsü benzeri temsil çıktısı üretilir. İkinci aşama sınıflandırma işleminin yapıldığı bölümdür. Bu aşamada kokleagram görüntülerinden özellikler Efficient-Net B4 ile çıkarılır ve ArCapsNet bu özelliklerle eğitilir. Capsule Forensics ağı orijinalinde olduğu haliyle VGG19 özellik çıkarıcı olarak kullanılarak eğitilir. Eğitim sonucunda, test için verilen ses dosyaları sahte veya gerçek olarak sınıflandırılır.

Kopyala-yapıştır sahteciliğinde, bir ses kaydının belirli bir bölümü kopyalanarak, aynı kaydın başka bir bölümüne yapıştırılır. Kopyala-yapıştır sahteciliği, özellikle adli bilişim, medya güvenliği ve ses kayıtlarının doğruluğunun kritik öneme sahip olduğu diğer alanlarda ciddi bir tehdit oluşturur. Bu sahtecilik türü, genellikle bir konuşmadaki belirli kelimelerin veya cümlelerin yerini değiştirmek, yeni anlamlar yaratmak veya orijinal mesajı saptırmak amacıyla kullanılır. Örneğin, bir kişinin “Ben Ahmet’i olay günü gördüm” ifadesi, “Ben Mehmet’i olay günü gördüm” şeklinde manipüle edilebilir. Bu işlem, kaydın anlamını veya içeriğini değiştirmek amacıyla yapılır ve genellikle orijinal kaydın tonal, zamansal veya frekans özelliklerini bozmadan gerçekleştirildiği için tespiti zor olabilir.

Ses kopyala-yapıştır sahteciliğinin tespiti için özel olarak oluşturulmuş bir veri kümesi kullanılmıştır. Bu veri kümesi, Arabic Speech Corpus temel alınarak geliştirilmiştir. Veri kümesinin oluşturulmasındaki temel amaç, ses manipülasyonlarının tespitinde kullanılabilecek güvenilir bir test ortamı sağlamaktır.

Veri kümesi, orijinal ve sahte olmak üzere iki farklı türde ses dosyasından oluşmaktadır. Orijinal ses dosyaları, Arapça konuşmalardan oluşan doğal kayıtları içermektedir. Bu orijinal ses kayıtları, yaklaşık 2 ila 6 saniye uzunluğunda olan 6300 kayıt içermektedir. Kayıtlar, 438 erkek ve 192 kadın konuşmacıdan alınmıştır, bu da veri kümesinin geniş bir demografik yelpazeyi kapsamasını sağlamaktadır. Sahte ses dosyaları ise, orijinal kayıtların belirli bölümlerinin kopyalanıp aynı kayıt içinde başka bir yere yapıştırılmasıyla oluşturulmuştur. Bu işlem, kayıtların anlamını bozmak ve manipüle etmek amacıyla yapılmıştır. Örneğin, bir kayıttan alınan 1 saniyelik bir bölüm, başka bir kaydın ortasına yapıştırılarak 3 saniyelik sahte bir kayıt oluşturulmuştur. Bu şekilde, toplamda 6000 sahte ses dosyası üretilmiştir.

Veri kümesi, eğitim ve test setleri olarak ikiye ayrılmıştır. Eğitim seti, modelin öğrenmesi için kullanılan 3112 orijinal ve 6000 sahte ses dosyasından oluşmaktadır. Test seti ise modelin performansını değerlendirmek için ayrılmış olup, 1145 orijinal ve 1000 sahte ses dosyasını içermektedir. Ayrıca, eğitim setinin bir kısmı doğrulama işlemleri için ayrılarak modelin aşırı öğrenme yapması engellenmeye çalışılmıştır.

Tespit için kullanılan yöntem iki aşamadan oluşmaktadır. İlk aşama ön işleme aşamasıdır. Bu aşamada giriş olarak verilen ses dosyası bir mel spektrogram görüntüsüne dönüştürülür. Mel spektrogram, insan işitme duyusuna uygun olarak frekans bileşenlerini temsil eder ve ses sinyalinin zaman içindeki değişimini görsel olarak sunar. Bu işlem, kısa zamanlı Fourier dönüşümü (STFT) kullanılarak gerçekleştirilmiştir. İkinci aşamada sınıflandırma işlemi gerçekleştirilir. Sınıflandırma için öncelikle mel spektrogramlardan özelliklerin çıkarılması için Efficient-Net B0 mimarisi kullanılır. Daha sonra bu özellikler Capsule Forensics ağına giriş olarak verilerek ağ eğitilir.

Her iki ses sahteciliği yöntemi için de elde edilen test sonuçları, bulgular ve tartışma bölümünde ilgili alt başlıklarda paylaşılmaktadır.

3. BULGULAR VE TARTIŞMA

Bu bölümde, derin sahte tespiti için geliştirilen yöntemlerin performans değerlendirmeleri, kullanılan veri kümeleri ve performans ölçütleri ile birlikte detaylandırılacaktır. Çeşitli ağ mimarilerinin ve özellik çıkarıcıların etkileri analiz edilecek, elde edilen deneysel sonuçlar kapsamlı bir şekilde incelenecektir. Ayrıca, farklı füzyon stratejilerinin performansa olan katkıları ve farklı aktivasyon fonksiyonlarının kullanımının deneysel sonuçlar üzerindeki etkileri de değerlendirilecektir. Son olarak, önerilen yöntemlerin literatürdeki diğer yöntemlerle karşılaştırılması ve elde edilen sonuçların istatistiksel yöntemler kullanılarak değerlendirilmesi gerçekleştirilecektir. Bu sayede, geliştirilen modellerin etkinliği ve geçerliliği hakkında daha kapsamlı bir bakış elde edilmesi hedeflenmektedir.

3.1. Kullanılan Veri Kümeleri

Derin sahte veri kümeleri genel olarak birinci nesil ve ikinci nesil veri kümeleri olarak ikiye ayrılır. İlk nesil veri kümeleri, düşük kaliteli sahte videolar içerirken, ikinci nesil veri kümeleri daha yüksek kaliteli ve daha zor tespit edilebilir sahte videolar sunar. Birinci nesil veri kümeleri, erken dönem derin sahte videoları içerir ve genellikle düşük kaliteli sahtecilik örneklerinden oluşur. Bu veri kümeleri, sahte videoların tespiti için daha kolaydır çünkü kullanılan sahtecilik teknikleri daha az gelişmiş ve daha az gerçekçidir. Bu veri kümelerinde var olan sahte video örneklerinde yer alan yüzlerin sahte oldukları çıplak gözle kolayca fark edilebilmektedir. Bu veri kümelerinin bazı özellikleri şunlardır:

- Düşük Kalite: Sahte videoların düşük çözünürlükte ve düşük kalitede olması.
- Kolay Tespit: Kullanılan sahtecilik tekniklerinin daha az karmaşık ve daha belirgin olması, sahte videoların tespit edilmesini kolaylaştırır.
- Sınırlı Çeşitlilik: Farklı sahtecilik teknikleri ve senaryoların sınırlı olması.

Birinci nesil veri kümelerinin sınırlamaları nedeniyle, bu çalışmada ikinci nesil veri kümeleri kullanılmıştır.

İkinci nesil veri kümeleri, daha gelişmiş sahtecilik teknikleri kullanılarak oluşturulmuş yüksek kaliteli sahte videolar içerir. Bu veri kümeleri, sahte videoların tespitini daha zor hale getiren karmaşık ve gerçekçi sahtecilik örneklerini barındırır. Bu

çalışmada derin sahte tespiti için ikinci nesil veri kümelerinden olan DFDC-P ve Celeb-DF kullanılmıştır. Bu veri kümeleri, geniş bir çeşitlilikte gerçek ve sahte videolar içermekte ve derin sahte tespit modellerinin eğitimi ve test edilmesi için ideal ortamı sağlamaktadır.

DFDC-P (Dolhansky vd., 2019) veri kümesi, Facebook tarafından başlatılan “DeepFake Detection Challenge” adlı yarışma kapsamında yayınlanan bir veri kümesidir. Bu veri kümesi, sahte video tespiti teknolojilerinin geliştirilmesini teşvik etmek ve bu alandaki araştırmaları ilerletmek amacıyla oluşturulmuştur. DFDC-P veri kümesi, geniş bir yelpazede gerçek ve sahte videolar içerir ve farklı sahtecilik tekniklerini kapsar. Bu veri kümesinde farklı yaş gruplarından, etnik kökenlerden ve cinsiyetlerden bireyleri içeren videolar yer alır. Veri kümesinde, yüz değiştirme ve yüz yeniden hedefleme gibi çeşitli sahtecilik teknikleri kullanılarak oluşturulmuş sahte videolar bulunur. Bu, modellerin farklı sahtecilik yöntemlerine karşı dayanıklılığını test etmeyi sağlar. Veri kümesinde 1131 gerçek video ve 4113 derin sahte video olmak üzere toplam 5244 video yer almaktadır. Test için ayrılan videolar veri kümesinin oluşturucuları tarafından test aşamalarında kullanılmasını önerdikleri videolardan oluşmaktadır ve bu videolar eğitim aşamalarında yer almamaktadır. Test için ayrılan videolar çıkarıldıktan sonra geri kalan videolar yaklaşık olarak %80 ve %20 olacak şekilde eğitim ve doğrulama bölümlerine ayrılmaktadır. Tablo 1’de veri kümesinde yer alan videoların eğitim, doğrulama ve test bölümlerine kaç adet ile ayrıldığı gösterilmektedir. Tablodan da görülebileceği gibi gerçek videoların 685 adeti eğitim, 170 tanesi doğrulama ve 276 tanesi test bölümleri için ayrılmıştır. Sahte videoların 2899 tanesi eğitim, 710 tanesi doğrulama ve 504 adeti test bölümlerine atanmıştır.

Tablo 1. Veri kümelerinin eğitim, doğrulama ve test bölümlerine ayrılışı

Veri Kümesi	Sınıf	Eğitim	Doğrulama	Test
Celeb-DF	Gerçek	610	102	178
	Sahte	4299	1000	340
DFDC-P	Gerçek	685	170	276
	Sahte	2899	710	504

Celeb-DF veri kümesi, derin sahte tespiti çalışmalarında yaygın olarak kullanılan bir başka önemli veri kümesidir (Li vd., 2020). Bu veri kümesi, ünlü kişilerin videolarından

oluşturulmuş sahte videoları içeren bir derin sahte veri kümesidir. Veri kümesi, yüksek kaliteli ve gerçekçi sahte videolar üretmeyi amaçlayan en son derin sahte algoritmaları kullanılarak oluşturulmuştur. Celeb-DF, hem sahte hem de gerçek videoları içeren dengeli bir veri kümesidir, bu sayede derin sahte tespiti için etkili modellerin eğitilmesine olanak tanır. Veri kümesi, 59 farklı ünlü kişiye ait toplamda 890 adet gerçek video ve bu gerçek videolardan üretilmiş 5639 adet derin sahte video içermektedir. Gerçek videolar, çeşitli röportajlar, konuşmalar ve televizyon programlarından alınmış farklı ünlü kişilerin yüzlerini içermektedir. Sahte videolar ise bu gerçek videolardan, yüz değiştirme algoritmaları kullanılarak üretilmiştir. Veri kümesi, yüksek çözünürlüklü videolardan oluşmakta olup, görsel açıdan yüksek kalite sunmaktadır. Derin sahte teknolojilerinin en son yenilikleri kullanılarak üretildiğinden, sahte videolar yüksek gerçekçilikte olup tespit edilmesi zor özellikler taşır. Literatürdeki çalışmalardan Khalil vd. (2021)'de olduğu gibi aynı stratejiyi kullandık ve veri kümesindeki ilk 13 ünlünün videolarını doğrulama için ve geri kalan 46 ünlünün videolarını eğitim için ayırdık. Tablo 1'de görülebileceği gibi, toplamda 610 gerçek ve 4299 sahte video eğitim için, 102 gerçek ve 1000 sahte video doğrulama için ve 178 gerçek ve 340 sahte video test için ayrılmıştır. Bu veri kümesindeki test kümesi ilgili yazarlar tarafından önceden tanımlanmıştır ve bu çalışmada da aynı şekilde kullanılmıştır. Test setinde yer alan videolar eğitim ve doğrulama aşamalarında kullanılmamıştır.

3.2. Kullanılan Performans Değerlendirme Ölçütleri

Derin sahte tespiti çalışmalarında, geliştirilen modellerin başarısını ve etkinliğini ölçmek için çeşitli performans değerlendirme ölçütleri kullanılır. Bu ölçütler, modellerin ne kadar doğru, hassas ve etkili olduğunu belirlemek için kullanılır. Bu çalışmada kullanılan başlıca performans değerlendirme ölçütleri doğruluk, kesinlik, duyarlılık, F1 skoru ve ROC-AUC eğrisidir. Bu ölçütler, modellerin gerçek ve sahte videoları ayırt etme yeteneğini kapsamlı bir şekilde değerlendirmek için kritik öneme sahiptir.

Doğruluk, modelin doğru sınıflandırdığı örneklerin sayısının toplam örnek sayısına oranıdır. Bu ölçüt, modelin genel performansını değerlendirmek için yaygın olarak kullanılmaktadır. Doğruluk aşağıdaki formülle hesaplanır,

$$\text{Doğruluk} = \frac{DP + DN}{DP + YN + DN + YP} \times 100 \quad (4)$$

burada DP, DN, YP, YN terimleri sırasıyla doğru pozitif, doğru negatif, yanlış pozitif ve yanlış negatif değerlerini ifade etmektedir. Pozitif sahte sınıfı, negatif ise gerçek sınıfı temsil etmektedir. Doğruluk, modelin hem gerçek hem de sahte videoları ne kadar doğru bir şekilde sınıflandırdığını gösterir. Ancak, dengesiz veri kümelerinde yanıltıcı olabilmektedir çünkü sayıca çok olan sınıfa ait örnekleri doğru sınıflandırarak yüksek doğruluk elde edilebilir.

Kesinlik, modelin sahte olarak sınıflandırdığı videoların ne kadarının gerçekten sahte olduğunu ölçer. Bu ölçüt, modelin yanlış pozitif oranını düşük tutma yeteneğini belirler. Kesinlik şu formülle hesaplanır:

$$Kesinlik = \frac{DP}{DP + YP} \times 100 \quad (5)$$

Kesinlik, özellikle sahte videoların tespitinde yanlış pozitiflerin azaltılmasının önemli olduğu durumlarda kritik bir ölçüttür.

Duyarlılık, modelin tüm gerçek sahte videoları ne kadar iyi tespit edebildiğini ölçer. Aynı zamanda doğru pozitif oranı (DPO) olarak da bilinir. Bu ölçüt, modelin yanlış negatif oranını düşük tutma yeteneğini değerlendirir. Duyarlılık şu formülle hesaplanır:

$$Duyarlılık = DPO = \frac{DP}{DP + YN} \times 100 \quad (6)$$

F1 skoru, kesinlik ve duyarlılık arasındaki dengeyi ölçer. Bu ölçüt, her iki ölçütü de dikkate alarak modelin genel performansını daha kapsamlı bir şekilde değerlendirir. F1 skoru şu formülle hesaplanır:

$$F1 = 2 \times \frac{Kesinlik \times Duyarlılık}{Kesinlik + Duyarlılık} \times 100 \quad (7)$$

ROC eğrisi, modelin farklı eşik değerlerinde doğru pozitif oranı (DPO) ve yanlış pozitif oranı (YPO) arasındaki ilişkiyi gösterir. AUC ise bu eğrinin altındaki alanı temsil eder ve modelin ayırt edici yeteneğini değerlendirir ve aşağıdaki eşitlikteki gibi hesaplanır:

$$AUC = \int_0^1 DPO(YPO) dYPO \times 100 \quad (8)$$

ve burada yer alan YPO aşağıdaki eşitlikte olduğu gibi hesaplanmaktadır,

$$YPO = \frac{YP}{DP + YN} \quad (9)$$

ROC-AUC, şu şekilde yorumlanır:

- $AUC = 1.0$: Mükemmel ayırım
- $0.9 \leq AUC < 1.0$: Çok iyi ayırım
- $0.8 \leq AUC < 0.9$: İyi ayırım
- $0.7 \leq AUC < 0.8$: Orta ayırım
- $0.6 \leq AUC < 0.7$: Kötü ayırım
- $0.5 \leq AUC < 0.6$: Rastgele tahmin

Yüksek AUC değeri, modelin hem pozitif hem de negatif sınıfları doğru bir şekilde ayırt edebilme yeteneğinin yüksek olduğunu gösterir.

Bu çalışmada, geliştirilen modellerin performansı yukarıda belirtilen ölçütler kullanılarak kapsamlı bir şekilde değerlendirilmiştir. Doğruluk, kesinlik, duyarlılık, F1 skoru ve ROC-AUC eğrisi, modellerin başarısını ölçmek için kullanılmıştır. Elde edilen sonuçlar, modellerin güçlü ve zayıf yönlerini ortaya koymaktadır. Bu ölçütlerin birlikte kullanılması, modellerin genel etkinliğinin ve gerçek dünya uygulamalarında ne kadar başarılı olabileceğinin daha iyi anlaşılmasına olanak tanımaktadır.

3.3. Uygulanan Yöntemlerin Değerlendirilmesi ve Elde Edilen DeneySEL Sonuçlar

Bu bölümde, derin sahte tespiti için uygulanan çeşitli yöntemlerin detaylı bir analizi ve bu yöntemlerle elde edilen deneySEL sonuçlar sunulmaktadır. Derin sahte tespiti alanında, kullanılan modellerin etkinliğini ve güvenilirliğini değerlendirmenin temel yolu, kapsamlı deneyler yaparak elde edilen performans ölçütlerini analiz etmektir. Bu analizler, modelin sahte ve gerçek videoları ayırt etmedeki başarısını, çeşitli veri kümelerinde gösterdiği performansı ve genel doğruluğunu ortaya koyar.

Çalışmada kullanılan temel modeller, Capsule Forensics ve ArCapsNet'tir. Bu modellerin her biri, sahtecilik tespiti konusunda yenilikçi yaklaşımlar sunar ve performansları çeşitli ölçütlerle değerlendirilmiştir. Ayrıca, farklı özellik çıkarıcı mimariler (VGG19, Efficient-Net B0 ve B4) ve füzyon stratejileri de kullanılarak modellerin

performansı iyileştirilmiştir. Tüm deneyler Celeb-DF ve DFDC-P veri kümeleri üzerinde gerçekleştirilmiştir.

3.3.1. Capsule Forensics Ağı Kullanımının Analizi ve Elde Edilen Deneysel Sonuçlar

Capsule Forensics ağı, derin sahte tespiti için optimize edilmiş ve orijinal kapsül ağı mimarisine dayanan bir modeldir. Bu model, dinamik yönlendirme ve squash aktivasyon fonksiyonları kullanarak, sahte videoların tespitinde yüksek doğruluk ve güvenilirlik sağlamaktadır. Bu bölümde, Capsule Forensics ağının performansına detaylı bir bakış sunulacak ve elde edilen deneysel performans sonuçları paylaşılacaktır.

Capsule Forensics modelinin temel özelliklerinden biri, görüntülerin çeşitli düzeylerdeki hiyerarşik özelliklerini kapsüller aracılığıyla modellemesidir. Bu özellik, özellikle karmaşık sahtecilik örüntülerinin tespitinde önemli avantajlar sağlar. Model, birincil kapsüller katmanı, konvolüsyonel kapsüller katmanı ve çıkış kapsülleri olmak üzere üç ana bileşenden oluşur. Dinamik yönlendirme algoritması, kapsüller arasındaki bilgi akışını optimize eder ve sahtecilik belirtilerini daha doğru bir şekilde yakalar. Giriş görüntüsünü işlemek için öncelikle özellik çıkarıcı bir ağ kullanır. Bu çalışmada, VGG19, Efficient-Net B0 ve Efficient-Net B4 gibi çeşitli özellik çıkarıcılar kullanılmaktadır. Özellik çıkarıcı katmanlar, giriş görüntülerindeki düşük seviyeli özellikleri çıkararak kapsül ağına girdi oluşturur.

Tablo 2’de Capsule Forensics modellerinin Celeb-DF veri kümesi üzerinde elde etmiş olduğu performans sonuçları yer almaktadır. Bu tablo, modellerin VGG19, Efficient-Net B0 ve Efficient-Net B4 özellik çıkarıcıları ile kullanılarak elde edilen AUC, Doğruluk, Kesinlik, Duyarlılık ve F1 skorlarını göstermektedir. Modellerden Capsule Forensics orijinal haliyle oluşturulan modeli ifade etmektedir. Capsule Forensics (Ağırlıklı Kayıp) olarak ifadelendirilen modelde diğerinden farklı olarak kayıp fonksiyonunda veri kümesindeki dengesizliği göze alarak sınıflar için ağırlıklandırma yapılmıştır.

Tablo 2 incelendiğinde, Capsule Forensics modeli için VGG19 özellik çıkarıcı ile elde edilen sonuçlar, modelin yüksek doğruluk ve kesinlik sağladığını göstermektedir. Ancak, duyarlılık %74.16 olarak nispeten düşük kalmıştır. Efficient-Net B0 ile elde edilen sonuçlar, yüksek AUC ve kesinlik değerlerine rağmen doğruluk ve F1 skorunun VGG19’a göre daha düşük olduğunu göstermektedir. Efficient-Net B4 ile elde edilen sonuçlar,

modelin yüksek doğruluk, kesinlik ve F1 skoru ile en iyi performansı sağladığını göstermektedir. Capsule Forensics modeli için Efficient-Net B4 modelinin özellik çıkarıcı olarak kullanılmasının en iyi performansı sağladığı görülmektedir.

Tablo 2. Capsule Forensics modelinin Celeb-DF veri kümesi üzerinde elde ettiği performans sonuçları

Model	Özellik Çıkarıcı	AUC (%)	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1 (%)
Capsule Forensics	VGG19	94.49	90.93	99.25	74.16	84.89
	Efficient-Net B0	97.93	89.58	96.97	71.91	82.58
	Efficient-Net B4	97.58	92.66	98.61	79.78	88.20
Capsule Forensics (Ağırlıklı Kayıp)	VGG19	97.54	92.28	97.92	79.21	87.58
	Efficient-Net B0	96.18	91.12	95.21	78.09	85.80
	Efficient-Net B4	98.61	93.05	97.97	81.46	88.96

Ağırlıklı kayıp fonksiyonu kullanılarak yapılan deneylerde, modelin performansında belirgin bir iyileşme gözlemlenmiştir. Özellikle, duyarlılık ve F1 skoru önemli ölçüde artmıştır. Ağırlıklı kayıp fonksiyonunun kullanımı kesinlik ölçütü dışındaki diğer tüm ölçütlerde performans artışı sağlamıştır. Efficient-Net B4 özellik çıkarıcı olarak kullanıldığında, en yüksek performans değerlerine ulaşılmıştır.

Tablo 3'te Capsule Forensics modellerinin DFDC-P veri kümesi üzerinde elde etmiş olduğu performans değerleri gösterilmektedir. Bu tablo, modellerin VGG19, Efficient-Net B0 ve Efficient-Net B4 özellik çıkarıcıları ile kullanılarak elde edilen AUC, Doğruluk, Kesinlik, Duyarlılık ve F1 skorlarını göstermektedir. Ayrıca, ağırlıklı kayıp fonksiyonu kullanılarak yapılan deneylerin sonuçları da tabloya dahil edilmiştir.

Tablo 3. Capsule Forensics modelinin DFDC-P veri kümesi üzerinde elde ettiği performans sonuçları

Model	Özellik Çıkarıcı	AUC (%)	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1 (%)
Capsule Forensics	VGG19	79.68	73.68	60.47	75.36	67.10
	Efficient-Net B0	81.57	72.52	58.68	77.17	66.67
	Efficient-Net B4	81.89	73.81	60.52	76.09	67.42
Capsule Forensics (Ağırlıklı Kayıp)	VGG19	82.92	74.06	61.47	72.83	66.67
	Efficient-Net B0	79.86	69.81	55.28	79.71	65.28
	Efficient-Net B4	80.55	72.13	58.47	75.00	65.71

VGG19 özellik çıkarıcı olarak kullanıldığında elde edilen sonuçlar, modelin doğruluk ve duyarlılık açısından iyi düzeyde performans gösterdiğini ancak kesinlik ölçütünde daha düşük başarı yakaladığını göstermektedir. F1 skoru, modelin genel performansını ve sahte video tespitindeki etkinliğini yansıtmaktadır. Efficient-Net B0 ile elde edilen sonuçlara bakıldığında, AUC ve duyarlılık açısından iyileşmeler görülmektedir. Ancak kesinlik ve F1 skorunda düşüş bulunmaktadır. Bu, modelin bazı sahte videoları kaçırabileceğini ve yanlış pozitif oranlarının daha yüksek olabileceğini işaret etmektedir. Efficient-Net B4 tarafına bakıldığında, modelin genel performansını daha da iyileştirmiştir. Tüm ölçütler için en iyi sonuçların alındığı görülmektedir.

Ağırlıklı kayıp fonksiyonunun modele eklenmesi ile yapılan deneylerde, Celeb-DF veri kümesi üzerindeki belirgin performans artışı sağlamasına rağmen DFDC-P veri kümesinde büyük bir etki yapamadığı görülmektedir. Bunun nedeni olarak Celeb-DF veri kümesindeki sınıf dengesizliğinin DFDC-P veri kümesinde olana göre çok daha fazla olması gösterilebilir.

Capsule Forensics ağırları, her iki veri kümesinde de yüksek performans sergilemiştir. Celeb-DF veri kümesi üzerinde, model yüksek AUC, doğruluk, kesinlik ve F1 skorları elde

etmiştir. Efficient-Net B4 özellik çıkarıcı ile kullanıldığında en iyi sonuçlar elde edilmiştir. Ağırlıklı kayıp fonksiyonunun eklenmesi özellikle Celeb-DF veri kümesinde performansı önemli ölçüde artırmaktadır. DFDC-P veri kümesinde, modeller makul performans göstermiş, ancak Celeb-DF veri kümesine göre biraz daha düşük değerler elde edilmiştir. Bu durum literatürde yer alan çalışmaların elde ettiği sonuçlara benzerdir ve DFDC-P veri kümesinin daha zorlu ve çeşitli olmasından kaynaklanabilir. Efficient-Net B4 özellik çıkarıcı ile kullanıldığında, modeller en yüksek performans değerlerine ulaşmıştır.

3.3.2. ArCapsNet Ağı Kullanımının Analizi ve Elde Edilen Deneysel Sonuçlar

ArCapsNet, orijinal kapsül ağından uyarlanmış ve derin sahte tespiti için bu çalışma kapsamında optimize edilmiş bir modeldir. Bu model, dinamik yönlendirme ve squash aktivasyon fonksiyonlarını, dikkat yönlendirme ve kapsül aktivasyonu ile değiştirerek sahtecilik tespitinde daha yüksek doğruluk ve verimlilik sağlamayı amaçlar. ArCapsNet, birincil kapsüller katmanı, konvolüsyonel kapsüller katmanı ve tam konvolüsyonel kapsüller katmanından oluşur. Her katmanın çıktısına 1x1 konvolüsyon ve tanh aktivasyonu uygulanarak, özelliklerin daha odaklı ve seçici bir şekilde işlenmesi sağlanır. Bu bölümde, ArCapsNet modelinin Celeb-DF ve DFDC-P veri kümeleri üzerinde elde ettiği performans sonuçları detaylı olarak incelenecektir.

ArCapsNet'in tekrarlanabilir yapıdaki konvolüsyonel kapsüller katmanı nedeniyle bu çalışmada 1'den 3'e kadar farklı tekrarlama seviyesiyle 3 farklı model oluşturulmuştur. ArCapsNet, üç farklı katman derinliğinde (Katman1, Katman2, Katman3) ve üç farklı özellik çıkarıcı (VGG19, Efficient-Net B0, Efficient-Net B4) kullanılarak test edilmiştir. Tablo 4'te farklı katman sayıları ve özellik çıkarıcılar ile Celeb-DF veri kümesi üzerinde elde edilen performans gösterilmektedir.

Katman1 seviyesinde, ArCapsNet modeli, VGG19 ve Efficient-Net B4 ile beraber kullanıldığında yüksek performans göstermiştir. VGG19 ile %96.67 kesinlik ve %77.85 F1 skoru elde edilmiştir. Efficient-Net B4 ile %96.88 kesinlik ve %81.05 F1 skoru dikkat çekmektedir. Katman2 seviyesinde, Efficient-Net B4 ile en yüksek performansı göstermiştir. %97.04 kesinlik, %73.60 duyarlılık ve %83.71 F1 skoru elde edilmiştir. Bu sonuçlar, daha derin katmanların model performansını artırdığını göstermektedir. Katman3 seviyesinde, VGG19 ve Efficient-Net B4 yine yüksek performans sergilemiştir. VGG19 ile %94.89 kesinlik ve %82.54 F1 skoru elde edilmiştir. Efficient-Net B4 ile %96.43 kesinlik

ve %74.48 F1 skoru elde edilmiştir. Efficient-Net B0, bu seviyede daha düşük bir performans göstermiştir. ArCapsNet modelinin çeşitli katman derinliklerinde ve farklı özellik çıkarıcılarla elde ettiği sonuçlar, modelin esneklik ve adaptasyon yeteneğini göstermektedir. VGG19, her seviyede yüksek performans sunarak derin ve zengin özellik çıkarımı sağlamaktadır. Efficient-Net B4 ise, özellikle Katman2 seviyesinde yüksek performans göstererek, yüksek doğruluk ve F1 skorları sağlamaktadır. Sonuçlar, daha derin katmanların (Katman2 ve Katman3) modelin performansını artırdığını göstermektedir. Bu, modelin daha karmaşık ve detaylı özellikleri daha iyi öğrenmesini sağlamaktadır.

Tablo 4. ArCapsNet modellerinin Celeb-DF veri kümesi üzerinde elde ettiği performans sonuçları

Model	Özellik Çıkarıcı	AUC (%)	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1 (%)
ArCapsNet (Katman1)	VGG19	89.88	87.26	96.67	65.17	77.85
	Efficient-Net B0	92.89	84.94	93.86	60.11	73.29
	Efficient-Net B4	91.40	88.80	96.88	69.66	81.05
ArCapsNet (Katman2)	VGG19	93.38	88.80	95.45	70.79	81.29
	Efficient-Net B0	94.57	85.91	89.47	66.85	76.53
	Efficient-Net B4	96.41	90.15	97.04	73.60	83.71
ArCapsNet (Katman3)	VGG19	94.02	89.38	94.89	73.03	82.54
	Efficient-Net B0	88.77	83.40	97.92	52.81	68.61
	Efficient-Net B4	90.86	85.71	96.43	60.67	74.48

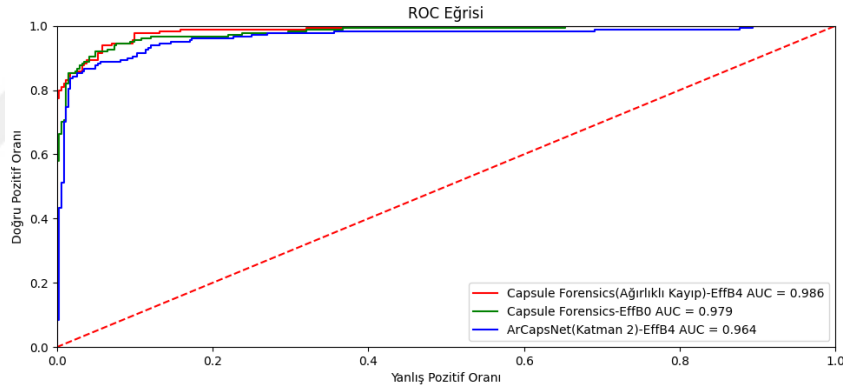
Efficient-Net B0, daha düşük hesaplama maliyeti ile belirli seviyelerde etkili olmuş ancak daha derin katmanlarda performansında düşüş olduğu gözlemlenmiştir. ArCapsNet Celeb-DF veri kümesi üzerinde her ne kadar başarılı sonuçlar elde etse de Capsule Forensics ile elde edilen sonuçların gerisinde kalmıştır. Tablo 5'te ArCapsNet modellerinin DFDC-P veri kümesi üzerindeki performans sonuçları gösterilmektedir.

Tablo 5. ArCapsNet modellerinin DFDC-P veri kümesi üzerinde elde ettiği performans sonuçları

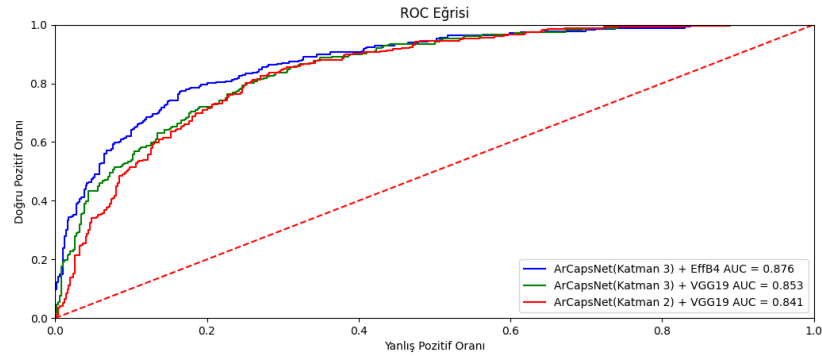
Model	Özellik Çıkarıcı	AUC (%)	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1 (%)
ArCapsNet (Katman1)	VGG19	82.63	76.39	66.43	68.12	67.26
	Efficient-Net B0	82.57	75.48	64.05	71.01	67.35
	Efficient-Net B4	82.94	76.90	65.40	74.64	69.71
ArCapsNet (Katman2)	VGG19	84.12	77.16	67.62	68.84	68.22
	Efficient-Net B0	82.57	75.61	63.55	73.91	68.34
	Efficient-Net B4	84.03	77.42	67.60	70.29	68.92
ArCapsNet (Katman3)	VGG19	85.28	77.55	68.75	67.75	68.25
	Efficient-Net B0	81.64	75.74	64.86	69.57	67.13
	Efficient-Net B4	87.56	80.52	71.93	74.28	73.08

ArCapsNet'in Katman1 yapılandırması VGG19 ve Efficient-Net B4 özellik çıkarıcıları ile birlikte kullanıldığında benzer performans göstermiştir. Efficient-Net B4, biraz daha yüksek doğruluk ve F1 skoru ile öne çıkmaktadır. %82.94 AUC ve %76.90 Doğruluk ile Katman1 yapılandırmasında en yüksek başarı Efficient-Net B4 ile elde

edilmiştir. Katman2 yapılandırmasında, ArCapsNet'in performansı genel olarak artmıştır. VGG19 ve Efficient-Net B4 ile yapılan deneyler, bu katmanda en yüksek performansı göstermiştir. Efficient-Net B4, daha yüksek duyarlılık ve F1 skoru ile öne çıkmaktadır. Katman3 yapılandırmasında, ArCapsNet'in performansı en yüksek seviyeye ulaşmıştır. Efficient-Net B4 ile yapılan deneyler, bu katmanda en yüksek AUC (%87.56), doğruluk (%80.52) ve F1 skorunu (%73.08) sağlamıştır. Bu sonuçlar, Efficient-Net B4'ün özellik çıkarıcı olarak kullanımının ArCapsNet'in performansını önemli ölçüde artırdığını göstermektedir. ArCapsNet'in farklı katman sayıları ve özellik çıkarıcılar ile performansı değerlendirildiğinde genel olarak en yüksek performansın Efficient-Net B4 ile Katman3 yapılandırılmasında elde edildiği görülmektedir. Bu model, %87.56 AUC, %80.52 doğruluk ve %73.08 F1 skoru ile DFDC-P veri kümesinde ArCapsNet modelleri içinde en iyi başarıyı sağlamıştır. VGG19 ve Efficient-Net B0 da iyi performans göstermiştir ancak Efficient-Net B4, özellikle yüksek duyarlılık ve kesinlik oranları ile öne çıkmaktadır.



Şekil 28. Celeb-DF veri kümesinde en iyi performans gösteren üç modele ait ROC eğrisi



Şekil 29. Celeb-DF veri kümesinde en iyi performans gösteren üç modele ait ROC eğrisi

Capsule Forensics ve ArCapsNet modellerinin Celeb-DF ve DFDC-P veri kümeleri üzerinde elde ettiği performans sonuçlarından en iyi üçüne ait ROC eğrileri çizilmiştir. Şekil 28’de Celeb-DF ve Şekil 29’da DFDC-P veri kümeleri için ROC eğrilerine ait grafikler görülmektedir.

3.3.3. Farklı Özellik Çıkarıcıların Deneysel Sonuçlar Üzerindeki Etkisinin Değerlendirilmesi

Derin sahte tespiti çalışmalarında, kullanılan özellik çıkarıcı mimarilerin model performansı üzerinde belirleyici bir etkisi vardır. Önceki bölümlerde Celeb-DF ve DFDC-P veri kümeleri üzerinde verilen performans sonuçlarından da görülebildiği üzere farklı özellik çıkarıcıların kullanımı performans üzerinde büyük farklılıklar ortaya koymuştur. Bu bölümde, farklı özellik çıkarıcıların (VGG19, Efficient-Net B0 ve Efficient-Net B4) Capsule Forensics ve ArCapsNet modelleri üzerindeki etkileri Celeb-DF ve DFDC-P veri kümeleri kullanılarak elde edilen deneysel sonuçlarla değerlendirilmiştir.

Capsule Forensics modelleri incelendiğinde, Celeb-DF veri kümesinde Efficient-Net B4 özellik çıkarıcısı en yüksek performansı sağlamış, AUC ve doğruluk değerlerinde diğer özellik çıkarıcılardan daha yüksek sonuçlar elde edilmiştir. DFDC-P veri kümesinde sonuçlar incelendiğinde farklı özellik çıkarıcıların birbirlerine karşı belirli ölçütlerde üstünlüğü varken genele bakıldığında belirgin bir üstünlüğün olmadığı görülmektedir.

ArCapsNet modeli farklı özellik çıkarıcılar ile test edildiğinde Efficient-Net B4 ile en yüksek performansı göstermiştir. Celeb-DF veri kümesinde, Efficient-Net B4 özellik çıkarıcısı ile Katman2 yapılandırmasında %96.41 AUC ve %90.15 doğruluk ile ArCapsNet modelleri içinde en yüksek sonuçlar elde edilmiştir. DFDC-P veri kümesinde ise aynı özellik çıkarıcı ile Katman3 yapılandırmasında %87.56 AUC ve %80.52 doğruluk elde edilmiştir. Bu sonuçlar Efficient-Net B4’ün ArCapsNet modeli için en etkili özellik çıkarıcı olduğunu göstermektedir.

Genel olarak farklı özellik çıkarıcıların kullanımı hem Capsule Forensics hem de ArCapsNet modellerinin performansını önemli ölçüde etkilemektedir. Efficient-Net B4, her iki modelde de en yüksek performans değerlerine ulaşılmasını sağlarken, VGG19 ve Efficient-Net B0 da iyi performans göstermiştir.

3.3.4. Füzyon Yapılmasının Sonuçlar Üzerindeki Etkisinin Değerlendirilmesi

Bu bölümde derin sahte videoların tespiti için geliştirilen farklı modellerin füzyonunun sonuçlar üzerindeki etkisi değerlendirilecektir. Füzyon stratejileri çeşitli modellerin sınıflandırma skorlarını birleştirerek performansın iyileştirilmesini amaçlar. Farklı modellerin sınıflandırma skorlarının birleştirilmesiyle elde edilen sonuçlar genel performansı önemli ölçüde artırabilir. Bu değerlendirme Celeb-DF ve DFDC-P veri kümelerinde gerçekleştirilen deneysel sonuçlar üzerinden yapılmıştır.

Füzyon yapılacak modellerin seçimi her bir modelin bağımsız performansına dayanmaktadır. Füzyon aşamalarında iki modelin füzyonu ve üç modelin füzyonu olarak iki farklı füzyon senaryosu uygulanmıştır. İkili füzyonda Capsule Forensics, Capsule Forensics (Ağırlıklı Kayıp), ArCapsNet (Katman1), ArCapsNet (Katman2) ve ArCapsNet (Katman3) modelleri için her bir modelde en iyi doğruluk sonuçlarının elde edildiği iki özellik çıkarıcı ile olan yapılandırma modeli birleştirilmektedir. Ayrıca Capsule Forensics modelleri içinde en iyi sonucu veren yapılandırma ile ArCapsNet modelleri içinde her katman için en iyi sonucu veren yapılandırma birleştirilmiştir. Bu senaryo Celeb-DF ve DFDC-P veri kümelerinin her ikisi için ayrı ayrı uygulanmıştır. Bu senaryoya göre gerçekleştirilen füzyon işlemleri sonrası oluşan sonuç tabloları Celeb-DF veri kümesi için

Tablo 6’da ve DFDC-P veri kümesi için

Tablo 7’de gösterilmektedir.

Bu sonuçlar, farklı özellik çıkarıcılar ile eğitilmiş Capsule Forensics ve ArCapsNet modellerinin füzyonunun model performansını önemli ölçüde artırdığını göstermektedir. Celeb-DF veri kümesi için, Capsule Forensics (Ağırlıklı Kayıp) + VGG19 ve Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4 kombinasyonu en yüksek performansı sağlayan birleşim olmuştur. Bu kombinasyon %99.14 AUC, %93.63 doğruluk, %99.32 kesinlik, %82.02 duyarlılık ve %89.85 F1 skoru değerleri ile Celeb-DF veri kümesinde en yüksek performans sağlayan model birleşimi olmuştur. DFDC-P veri kümesinde de benzer bir şekilde farklı modellerin füzyonunun performansı artırdığı görülmektedir. En yüksek performans, ArCapsNet (Katman3) + VGG19 ve ArCapsNet (Katman3) + Efficient-Net B4 kombinasyonunda elde edilmiştir. Bu kombinasyon %89.08 AUC, %82.84 doğruluk, %76.00 kesinlik, %75.72 duyarlılık, %75.86 F1 skoru ile DFDC-P veri kümesi üzerinde en yüksek başarı sağlayan füzyon kombinasyonu olarak öne çıkmaktadır.

Tablo 6. Celeb-DF veri kümesi için en iyi performans gösteren modellerin füzyon sonuçları

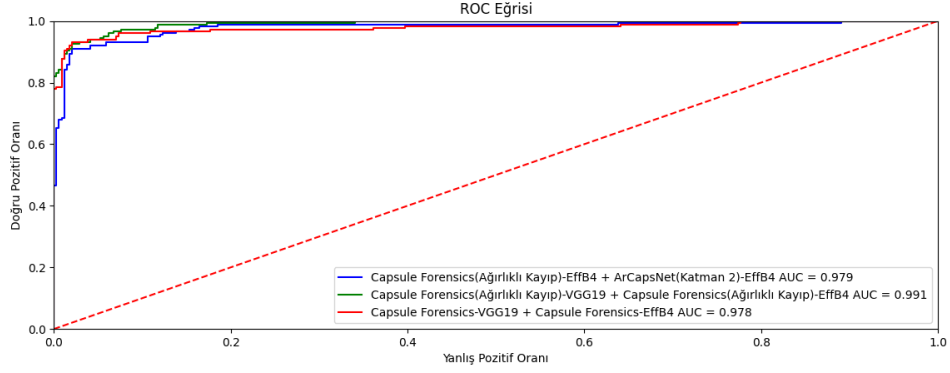
Model 1	Model 2	AUC (%)	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1 (%)
Capsule Forensics (Ağırlıklı kayıp) + VGG19	Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4	99.14	93.63	99.32	82.02	89.85
Capsule Forensics + VGG19	Capsule Forensics + Efficient-Net B4	97.83	91.51	100.00	75.28	85.90
ArCapsNet (Katman1) + VGG19	ArCapsNet (Katman1) + Efficient-Net B4	94.33	87.64	99.14	64.61	78.23
ArCapsNet (Katman2) + VGG19	ArCapsNet (Katman2) + Efficient-Net B4	97.79	89.96	97.73	72.47	83.23
ArCapsNet (Katman3) + VGG19	ArCapsNet (Katman3) + Efficient-Net B4	96.33	88.80	96.88	69.66	81.05
Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4	ArCapsNet (Katman1) + Efficient-Net B4	95.55	89.38	97.67	70.79	82.08
Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4	ArCapsNet (Katman2) + Efficient-Net B4	97.88	91.51	97.18	77.53	86.25
Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4	ArCapsNet (Katman3) + VGG19	97.44	89.38	94.89	73.03	82.54

Tablo 7. DFDC-P veri kümesi için en iyi performans gösteren modellerin füzyon sonuçları

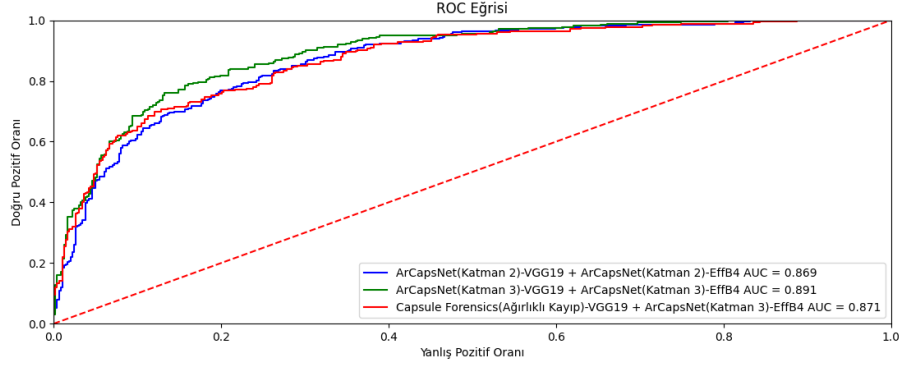
Model 1	Model 2	AUC (%)	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1 (%)
Capsule Forensics + VGG19	Capsule Forensics + Efficient-Net B4	83.50	73.29	59.89	75.72	66.88
Capsule Forensics (Ağırlıklı Kayıp) + VGG19	Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4	83.40	75.35	62.54	76.81	68.94
ArCapsNet (Katman1) + VGG19	ArCapsNet (Katman1) + Efficient-Net B4	86.58	79.87	69.87	76.45	73.01
ArCapsNet (Katman2) + VGG19	ArCapsNet (Katman2) + Efficient-Net B4	86.86	79.10	70.21	71.74	70.97
ArCapsNet (Katman3) + VGG19	ArCapsNet (Katman3) + Efficient-Net B4	89.08	82.84	76.00	75.72	75.86
Capsule Forensics (Ağırlıklı Kayıp) + VGG19	ArCapsNet (Katman1) + Efficient-Net B4	83.97	76.65	64.62	76.09	69.88
Capsule Forensics (Ağırlıklı Kayıp) + VGG19	ArCapsNet (Katman2) + Efficient-Net B4	84.46	77.29	66.45	73.19	69.66
Capsule Forensics (Ağırlıklı Kayıp) + VGG19	ArCapsNet (Katman3) + Efficient-Net B4	87.14	78.84	68.67	74.64	71.53

İkili füzyon sonuçları incelendiğinde her iki veri kümesi için de tek modellere göre daha yüksek ölçüt değerlerine ulaşıldığı görülmektedir. Bu durum füzyon yapılmasının başarıyı artırdığını açıkça göstermektedir.

İkili füzyon modellerinin Celeb-DF ve DFDC-P veri kümeleri üzerinde elde ettiği performans sonuçlarından en iyi üçüne ait ROC eğrileri çizilmiştir. Şekil 30'da Celeb-DF ve Şekil 31'de DFDC-P veri kümeleri için ROC eğrilerine ait grafikler görülmektedir.



Şekil 30. Celeb-DF veri kümesinde en iyi performans gösteren üç ikili füzyon modeline ait ROC eğrisi



Şekil 31. Celeb-DF veri kümesinde en iyi performans gösteren üç ikili füzyon modeline ait ROC eğrisi

Üçlü model füzyon stratejilerinde üç farklı modelin sınıflandırma skorlarının birleştirilmesiyle karar verilir. Tablo 8 Celeb-DF veri kümesi için üçlü model füzyon sonuçlarını gösterirken Tablo 9 DFDC-P veri kümesi için üçlü model füzyonunun sonuçlarını göstermektedir.

Üçlü füzyon kombinasyonları incelendiğinde Celeb-DF veri kümesinde en başarılı sonuçları elde eden kombinasyonun Capsule Forensics (Ağırlıklı Kayıp) + VGG19, Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B0 ve Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4 modellerinin birleşimi ile oluştuğu görülmektedir. Bu kombinasyon %99.09 AUC, %92.86 doğruluk, %97.32 kesinlik, %81.46 duyarlılık ve %88.69 F1 skoru sonuçlarına ulaşmaktadır. DFDC-P veri kümesinde en iyi sonuç veren üçlü füzyon kombinasyonu ArCapsNet (Katman3) + VGG19, ArCapsNet (Katman3) + Efficient-Net B4 ve Capsule Forensics + Efficient-Net B4 olmaktadır. Bu kombinasyon

%88.67 AUC, %82.32 doğruluk, %74.22 kesinlik, %77.17 duyarlılık ve %75.67 F1 skoru değerlerini elde etmektedir.

Tablo 8. Celeb-DF veri kümesi için 3 model füzyonu sonuçları

Model 1	Model 2	Model 3	AUC (%)	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1 (%)
Capsule Forensics + VGG19	Capsule Forensics + Efficient-Net B0	Capsule Forensics + Efficient-Net B4	99.10	90.54	99.24	73.03	84.14
Capsule Forensics (Ağırlıklı Kayıp) + VGG19	Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B0	Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4	99.09	92.86	97.32	81.46	88.69
ArCapsNet (Katman1) + VGG19	ArCapsNet (Katman1) + Efficient-Net B0	ArCapsNet (Katman1) + Efficient-Net B4	95.35	87.07	97.44	64.04	77.29
ArCapsNet (Katman2) + VGG19	ArCapsNet (Katman2) + Efficient-Net B0	ArCapsNet (Katman2) + Efficient-Net B4	97.06	89.38	96.24	71.91	82.32
ArCapsNet (Katman3) + VGG19	ArCapsNet (Katman3) + Efficient-Net B0	ArCapsNet (Katman3) + Efficient-Net B4	96.60	87.26	99.12	63.48	77.40
Capsule Forensics (Ağırlıklı Kayıp) + VGG19	Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4	ArCapsNet (Katman1) + Efficient-Net B4	97.77	90.93	99.25	74.16	84.89
Capsule Forensics (Ağırlıklı Kayıp) + VGG19	Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4	ArCapsNet (Katman2) + Efficient-Net B4	98.34	92.28	97.92	79.21	87.58
Capsule Forensics (Ağırlıklı Kayıp) + VGG19	Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4	ArCapsNet (Katman3) + VGG19	97.96	90.15	97.74	73.03	83.60

Tablo 9. DFDC-P veri kümesi için 3 model füzyonu sonuçları

Model 1	Model 2	Model 3	AUC (%)	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1 (%)
Capsule Forensics + VGG19	Capsule Forensics + Efficient-Net B0	Capsule Forensics + Efficient-Net B4	83.43	74.32	60.66	79.35	68.76
Capsule Forensics (Ağırlıklı Kayıp) + VGG19	Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B0	Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4	83.41	72.39	58.33	78.62	66.98
ArCapsNet (Katman1) + VGG19	ArCapsNet (Katman1) + Efficient-Net B0	ArCapsNet (Katman1) + Efficient-Net B4	86.20	78.58	67.86	75.72	71.58
ArCapsNet (Katman2) + VGG19	ArCapsNet (Katman2) + Efficient-Net B0	ArCapsNet (Katman2) + Efficient-Net B4	86.59	79.61	69.93	75.00	72.38
ArCapsNet (Katman3) + VGG19	ArCapsNet (Katman3) + Efficient-Net B0	ArCapsNet (Katman3) + Efficient-Net B4	88.18	81.29	73.14	75.00	74.06
ArCapsNet (Katman3) + VGG19	ArCapsNet (Katman3) + Efficient-Net B4	Capsule Forensics + Efficient-Net B4	88.67	82.32	74.22	77.17	75.67
ArCapsNet (Katman3) + VGG19	ArCapsNet (Katman3) + Efficient-Net B4	Capsule Forensics (Ağırlıklı Kayıp) + VGG19	88.81	81.68	73.59	75.72	74.64

Füzyon, farklı modellerin güçlü yönlerini birleştirerek genel performansı artırmış ve sahtecilik tespitinde daha yüksek doğruluk ve güvenilirlik sağlamıştır. Tekli modellerin performansları ile karşılaştırıldığında, ikili ve üçlü füzyon stratejilerinin performansı önemli ölçüde artırdığı görülmektedir. Bununla birlikte ikili füzyon kombinasyonları üçlü füzyon kombinasyonlarına kıyasla daha başarılı sonuçlar elde etmektedir.

Sonuç olarak, farklı modellerin füzyonu, derin sahte tespiti modellerinin performansını önemli ölçüde artırmaktadır. İkili füzyon stratejileri yüksek performanslı sonuçlar elde edilmesinde etkili olmuştur.

3.3.5. Farklı Füzyon Yöntemlerinin Etkilerinin Değerlendirilmesi

Bu bölümde, Celeb-DF ve DFDC-P veri kümelerinde uygulanan farklı füzyon stratejilerinin etkileri AUC ölçütü temel alınarak analiz edilecektir. Celeb-DF veri kümesi üzerinde yapılan deneylerde farklı füzyon stratejilerinin minimum, maksimum ve ortalama AUC değerleri Tablo 10'da gösterilmektedir.

Benzer şekilde DFDC-P veri kümesi üzerinde yapılan deneylerde farklı füzyon stratejilerinin minimum, maksimum ve ortalama AUC değerleri de Tablo 11'de gösterilmektedir.

Tablo 10. Celeb-DF veri kümesi için en iyi performans gösteren modellerin farklı füzyon stratejilerinin AUC ölçütü sonuçları

Model 1	Model 2	Min. (%)	Maks. (%)	Ortalama (%)
Capsule Forensics (Ağırlıklı Kayıp) + VGG19	Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4	98.41	97.68	99.14
Capsule Forensics + VGG19	Capsule Forensics + Efficient-Net B4	97.73	94.83	98.55
Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4	ArCapsNet (Katman2) + Efficient-Net B4	98.52	96.44	97.88

Tablo 11. DFDC-P veri kümesi için en iyi performans gösteren modellerin farklı füzyon stratejilerinin AUC ölçütü sonuçları

Model 1	Model 2	Min. (%)	Maks. (%)	Ortalama (%)
ArCapsNet (Katman1) + VGG19	ArCapsNet (Katman1) + Efficient-Net B4	81.09	83.81	86.58
ArCapsNet (Katman2) + VGG19	ArCapsNet (Katman2) + Efficient-Net B4	83.32	85.07	86.86
ArCapsNet (Katman3) + VGG19	ArCapsNet (Katman3) + Efficient-Net B4	84.10	88.56	89.08

Her iki tablodan da görüldüğü üzere füzyon stratejileri içinde en başarılı sonuçların elde edildiği durum ortalama füzyon stratejisinde bulunmaktadır. Bu durum daha önceki bölümde verilen füzyon kombinasyonlarında tercih edilen ortalama füzyon stratejisinin seçiminin doğruluğunu onaylamaktadır.

3.3.6. Farklı Aktivasyon Fonksiyonlarının Kullanımının Deneysel Sonuçlar Üzerindeki Etkisinin Değerlendirilmesi

Aktivasyon fonksiyonları, sinir ağlarının davranışlarını ve öğrenme dinamiklerini şekillendirmede önemli bir rol oynamaktadır. Çalışmanın bu bölümünde Capsule Forensics ve ArCapsNet modelleri yapısında yer alan ReLu aktivasyon fonksiyonu yerine farklı aktivasyon fonksiyonlarının kullanımının sonuçlar üzerindeki etkisi değerlendirilecektir. Celeb-DF veri kümesinde tekil modeller içinde en iyi sonucu veren Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4 yapılandırmasında Capsule Forensics kısmı içinde yer alan ReLu fonksiyonları farklı aktivasyon fonksiyonları ile değiştirilerek yeniden eğitimler yapılmıştır. DFDC-P veri kümesinde ise tekil modeller içinde en iyi sonucu veren ArCapsNet (Katman3) + Efficient-Net B4 yapılandırmasında ArCapsNet kısmı içinde yer alan ReLu fonksiyonları farklı aktivasyon fonksiyonları ile değiştirilerek yeniden eğitimler yapılmıştır. Modellerin farklı aktivasyon fonksiyonları ile yeniden eğitilerek yapılan yeni testlerin sonuçlarını içeren tablolar Celeb-DF veri kümesi için Tablo 12’de verilirken DFDC-P veri kümesi için Tablo 13’te sunulmaktadır.

Hardsigmoid aktivasyon fonksiyonu, her iki veri kümesinde de performans düşüşüne neden olmuştur. Celeb-DF'de test doğruluğu %9.07, AUC ise %10.02 azalmıştır. DFDC-P'de ise test doğruluğunda %2.71 ve AUC'de %6.41'lik bir düşüş gözlenmiştir.

Lelelu (Leaky ReLU), Celeb-DF'de doğrulukta %0.97'lik bir azalma ve AUC'de %4.67'lik bir artış göstermiştir. DFDC-P'de ise doğrulukta %4.13 ve AUC'de %3.39'luk bir düşüşe neden olmuştur.

Mish aktivasyon fonksiyonu, Celeb-DF'de test doğruluğunda %0.58'lik bir artış ve AUC'de %0.40'lık bir artış sağlamıştır. DFDC-P'de ise doğrulukta %0.65 ve AUC'de %1.99'luk bir düşüş göstermiştir.

PReLU, Celeb-DF'de test doğruluğunda %0.97 ve AUC'de %0.93'lük bir artış sağlamıştır. Ancak, DFDC-P'de test doğruluğunda %1.94 ve AUC'de %2.26'lık bir düşüş göstermiştir.

Swish aktivasyon fonksiyonu, Celeb-DF'de test doğruluğu sabit kalırken, AUC'de %0.56'lık bir düşüş gözlenmiştir. DFDC-P'de ise test doğruluğunda %1.81 ve AUC'de %2.56'lık bir düşüş olmuştur.

Tanh aktivasyon fonksiyonu her iki veri kümesinde de belirgin bir performans düşüşü sergilemiştir. Celeb-DF'de test doğruluğu %9.85 ve AUC %11.63 oranında azalmıştır. DFDC-P'de ise doğrulukta %1.94 ve AUC'de %2.53'lük bir düşüş gözlenmiştir.

Tablo 12. Celeb-DF Veri Kümesindeki En İyi Performans Gösteren Modelin Farklı Aktivasyon Fonksiyonlarıyla Sonuçları

Aktivasyon Fonksiyonu	AUC (%)	Doğruluk (%)
ReLu	98.61	93.05
Hardsigmoid	88.59	83.98
Lelelu	97.72	92.08
Mish	99.01	93.63
PReLU	97.49	94.02
Swish	98.05	93.05
Tanh	86.98	83.20

Tablo 13. DFDC-P Veri Kumesindeki En İyi Performans Gösteren Modelin Farklı Aktivasyon Fonksiyonlarıyla Sonuçları

Aktivasyon Fonksiyonu	AUC (%)	Doğruluk (%)
ReLu	87.56	80.52
Hardsigmoid	81.15	77.81
Lelelu	84.17	76.39
Mish	85.57	79.87
PReLU	85.30	78.58
Swish	85.00	78.71
Tanh	85.03	78.58

Bu kapsamlı çalışma aktivasyon fonksiyonlarının model performansı üzerindeki etkilerini net bir şekilde ortaya koymaktadır. ReLU, genellikle en iyi performansı sağlarken, Mish ve PReLU da bazı veri kümelerinde başarılı sonuçlar göstermiştir. Hardsigmoid ve Tanh gibi fonksiyonlar ise her iki veri kümesinde de performansı düşürme eğiliminde olmuştur. Bu bulgular, derin öğrenme modellerinin performansını optimize etmek için doğru aktivasyon fonksiyonunun seçilmesinin önemini vurgulamaktadır. Çalışmamız, farklı aktivasyon fonksiyonlarının, veri kümesinin özelliklerine ve modele bağlı olarak nasıl değişken sonuçlar verebileceğini göstermektedir. Bu nedenle, model geliştirme sürecinde aktivasyon fonksiyonlarının dikkatli bir şekilde değerlendirilmesi ve deneysel olarak test edilmesi gerekmektedir.

3.3.7. Veri Artırma Tekniklerinin Deneysel Sonuçlar Üzerindeki Etkisinin Değerlendirilmesi

Bu bölümde veri artırma tekniklerinin deneysel sonuçlar üzerindeki etkisini inceleyerek kullanılan farklı veri artırma yöntemlerinin model performansına olan etkilerini değerlendiriyoruz. Veri artırma, derin öğrenme modellerinin genelleme yeteneklerini ve sağlamlığını artırmada kullanılan yöntemlerden biridir. Bu çalışmada, çeşitli veri artırma teknikleri kullanılarak modelin performansı analiz edilmiştir.

Deneylerde kullanılan veri artırma teknikleri arasında rastgele döndürme, çevirme ve yeniden boyutlandırıp kırma bulunmaktadır. Bu teknikler, eğitim veri kümesine çeşitlilik kazandırarak modelin farklı açılardan nesneleri tanıma yeteneğini geliştirir.

Veri artırma tekniklerinin model performansı üzerindeki etkisini değerlendirmek için farklı artırma stratejileri ile eğitim yapılarak sonuçlar karşılaştırılmıştır. Deneyler, Celeb-DF veri kümesi üzerinde Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4 modeli kullanılarak gerçekleştirilmiştir. Tablo 14 farklı veri artırma stratejileri ile eğitilen modelin Celeb-DF veri kümesi üzerinde elde ettiği sonuçları sunmaktadır. Bu senaryolar, artırma yapılmamış temel durum, tüm görüntülerin dört kez artırıldığı (Artırma 1) ve sadece gerçek görüntülerin artırıldığı (Artırma 2) durumlardan oluşmaktadır.

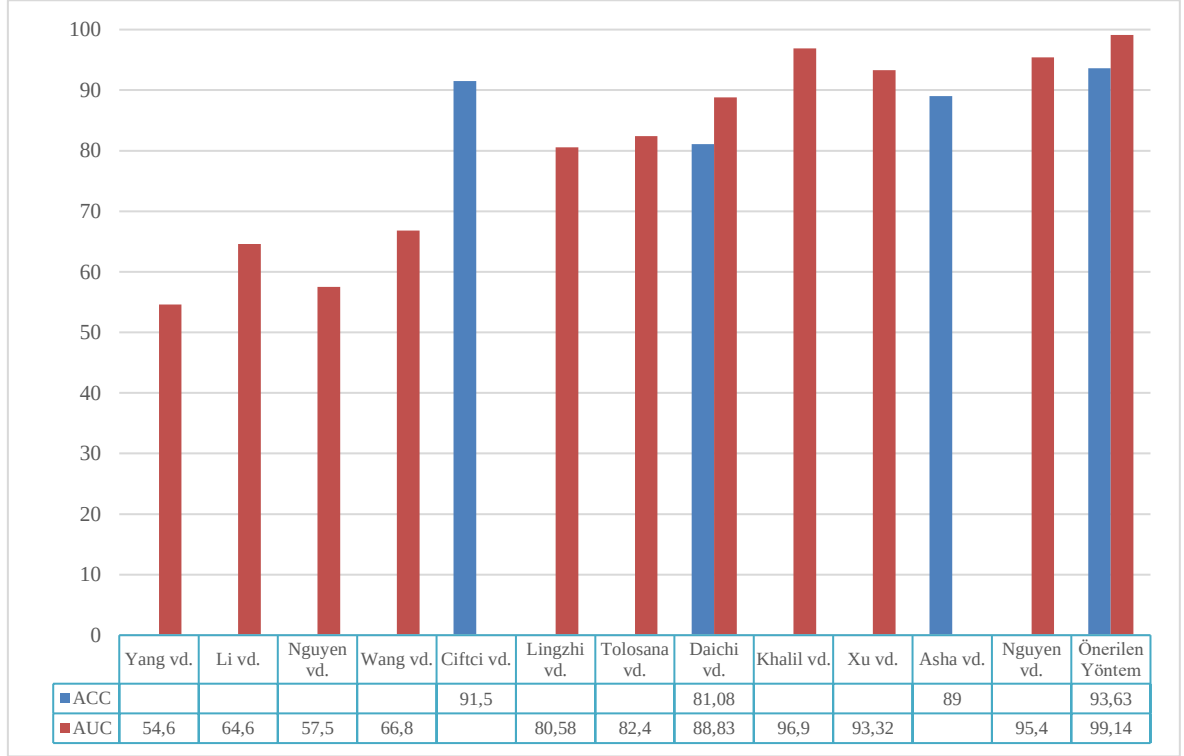
Tablo 14. Capsule Forensics (Ağırlıklı Kayıp) + Efficient-Net B4 modeli üzerinde farklı artırma veri kümeleriyle eğitim yapıldığında Celeb-DF veri kümesi üzerindeki sonuçlar.

Veri kümesi	AUC (%)	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1 (%)
Artırma Yok	98.61	93.05	97.97	81.46	88.96
Artırma 1	94.95	90.54	98.50	73.60	84.24
Artırma 2	98.75	93.63	98.01	83.15	89.97

Artırma yapılmamış senaryo, yüksek performans ölçütleri ile temel bir karşılaştırma noktası oluştururken, tüm görüntülerin artırıldığı senaryoda performans ölçütlerinde hafif bir düşüş gözlenmiştir. Bu durum, zaten dengesiz sınıf dağılımına sahip olan Celeb-DF veri kümesinin, veri artırma ile tüm örneklerinin dört kez artırılmasının sınıf dengesizliğini daha da artırdığından dolayı aşırı artırmanın model performansını olumsuz etkileyebileceğini göstermektedir. Ancak, sadece gerçek görüntülerin artırıldığı ve veri kümesinin dengelendiği senaryoda, performans ölçütlerinde belirgin bir iyileşme görülmüştür ve en iyi sonuçlar elde edilmiştir. Bu, dengeli artırma stratejilerinin modelin genelleme yeteneğini artırarak daha iyi performans sağlamasına olanak tanıdığını göstermektedir. Sonuç olarak, veri artırma tekniklerinin dikkatli ve duruma uygun kullanımı, derin öğrenme modellerinin performansını önemli ölçüde etkileyebilir. Bu deneysel sonuçlar, veri artırma stratejilerinin model eğitimi sırasında dikkatle seçilmesi gerektiğini ve doğru şekilde uygulanmasının model performansını artırdığını göstermektedir.

3.4. Önerilen Yöntemin Literatürdeki Diğer Yöntemlerle Karşılaştırılması

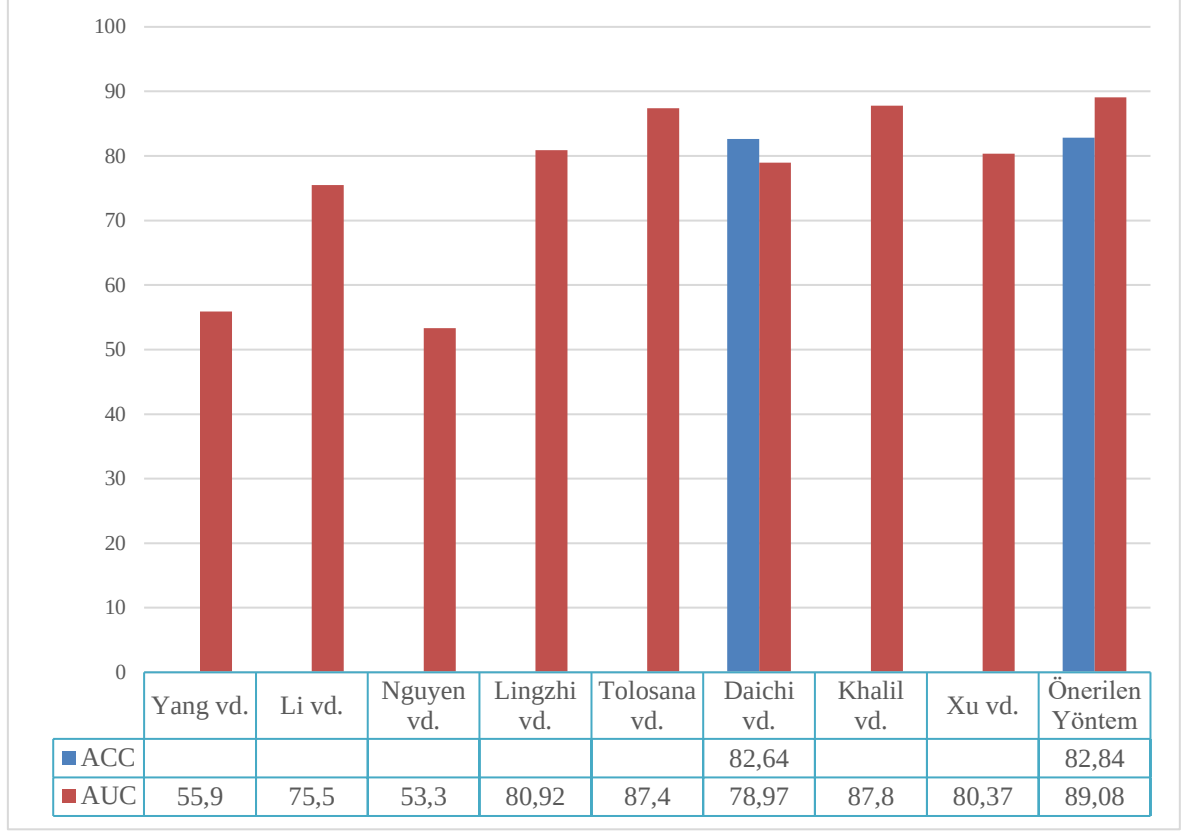
Bu bölümde önerilen yöntemle Celeb-DF ve DFDC-P veri kümeleri üzerinde elde edilen performans değerlerinin literatürdeki diğer çalışmalar ile kıyaslaması yapılacaktır. Önerilen yöntem, her iki veri kümesinde de yüksek performans sergilemiştir ve literatürdeki diğer yöntemlerle yapılan karşılaştırmalar, yöntemin üstünlüğünü açıkça ortaya koymaktadır. Önerilen modelin Şekil 32 ve Şekil 33’da gösterilen karşılaştırmaları doğrultusunda, doğruluk ve AUC ölçütlerinde karşılaştırma yapılan tüm çalışmalardan daha yüksek sonuçlar elde edildiği görülmektedir.



Şekil 32. Önerilen yöntemin Celeb-DF veri kümesinde literatürdeki diğer çalışmalarla karşılaştırması

Celeb-DF veri kümesinde önerilen yöntem, %93.63 doğruluk ve %99.14 AUC değerleri ile diğer çalışmalara göre belirgin bir üstünlük sağlamıştır. Ciftci vd. (2020) %91.5 doğruluk değeri ile öne çıkarken, AUC sonucu raporlanmamıştır. Khalil vd. (2021) %96.9 AUC ile en yüksek değerlerden birine sahiptir. Nguyen vd. (2024) %95.4 AUC değeri ile yüksek performans göstermiştir. Daichi vd. (2021) %81.08 doğruluk ve %88.83

AUC değerleri elde etmiştir. Önerilen yöntem, hem doğruluk hem de AUC ölçütlerinde diğer çalışmaları aşarak üstün performans göstermektedir.



Şekil 33. Önerilen yöntemin DFDC-P veri kümesinde literatürdeki diğer çalışmalarla karşılaştırması

DFDC-P veri kümesindeki yöntemler arasında önerilen yöntem, 82.84% doğruluk ve 89.08% AUC değerleri ile en yüksek performansı elde etmektedir. Daichi vd. (2021) %82.64 doğruluk ve %78.97 AUC değerleri elde etmiştir. Tolosana vd. (2021) %87.4 AUC ile yüksek bir değer elde etmiş ancak doğruluk değeri raporlanmamıştır. Khalil vd. (2021) %87.8 AUC ile yüksek bir performans göstermiştir. Lingzhi vd. (2020) %80.92 AUC değerini elde etmiştir. Diğer yöntemlerle karşılaştırıldığında önerilen yöntem, özellikle AUC ölçütünde diğer tüm yöntemleri geride bırakmaktadır.

Önerilen yöntemin performans sonuçlarının AUC ve doğruluk ölçütlerinde diğer çalışmaları yapılan karşılaştırmasının dışında F1 skoru ölçütü için de karşılaştırma yapılmıştır. Bu karşılaştırma Tablo 15'te gösterilmektedir.

Tablo 15. Celeb-DF veri kümesinde karşılaştırmalı F1 skoru

Yöntem	Doğruluk (%)	F1 (%)
Önerilen Yöntem	93.63	89.85
Daichi vd. (2021)	81.08	82.42
HCIT (Kaddar vd., 2024)	76.03	73.96
FedForgery (D. Liu vd., 2023)	75.74	72.87
DCL (Sun vd., 2022)	75.58	72.70
RECCE (Cao vd., 2022)	72.48	71.87
RFM (Wang & Deng, 2021)	70.72	69.94
CViT (Wodajo & Atnafu, 2021)	73.85	68.91
3D ResNet (Roy vd., 2022)	68.17	66.71
DSP-FWA (Li vd., 2020)	64.60	61.93
Face X-ray (Li vd., 2020)	54.87	50.79

Celeb-DF veri kümesinde, önerilen yöntem %89.85 F1 skoru ile en yüksek değere ulaşmıştır. Diğer yöntemlerin F1 skorlarına bakıldığında, Daichi vd. (2021) %82.42 F1 skoru ile en yakın performansı sergilemiştir. RECCE (Cao vd., 2022) %71.87 F1 skoru, CViT (Wodajo & Atnafu, 2021) %68.91 F1 skoru ile kısmen başarılı sonuç yakalamışlardır. Önerilen yöntem, F1 skorunda diğer yöntemleri önemli ölçüde geride bırakarak üstün performans göstermektedir.

Önerilen yöntemin literatürdeki diğer yöntemlerle model karmaşıklığı ve performans ölçütleri üzerinden karşılaştırması da gerçekleştirilmiştir. Model karmaşıklığı, parametre sayısı ve FLOPs gibi hesaplama maliyetleri ile değerlendirilirken, performans ölçütü AUC ve video başına kullanılan görüntü karesi sayısı gibi ölçütlerle incelenmiştir. Tablo 16'da çeşitli derin öğrenme modellerinin karmaşıklık ve performans açısından karşılaştırılması sunulmuştur. Bu tabloda, Celeb-DF veri kümesinde elde edilen AUC değerleri, modellerin parametre sayıları, FLOPs değerleri ve video başına kullanılan kare sayıları verilmiştir.

Tablo 16. Celeb-DF veri kümesinde karşılaştırmalı karmaşıklık analizi

Model	Celeb-DF AUC (%)	Parametre Sayısı	FLOPs (G)	Video Başına Kare Sayısı
Capsule (Nguyen vd., 2019)	57.50	3.90M	11.15	İlk 100
VidTr (Zhang vd., 2021)	83.30	93.00M	117.00	Videonun tüm kareleri
ISTVT (Zhao vd., 2023)	84.10	—	455.8	Videonun tüm kareleri
Caps. Foren.+VGG19	97.54	3.90M	8.14	1
Caps. Foren.+EffB0	98.10	1.65M	0.89	1
Caps. Foren.+EffB4	98.61	1.86M	1.08	1
ArCapsNet (Katman1) + VGG19	89.88	205.54M	7.35	1
ArCapsNet (Katman1) + EffB0	82.89	203.26M	0.09	1
ArCapsNet (Katman1) + EffB4	91.40	203.47M	0.28	1
ArCapsNet (Katman2) + VGG19	93.38	204.91M	7.35	1
ArCapsNet (Katman2) + EffB0	94.57	202.63M	0.09	1
ArCapsNet (Katman2) + EffB4	96.41	202.84M	0.28	1
ArCapsNet (Katman3) + VGG19	94.02	205.07M	7.35	1
ArCapsNet (Katman3) + EffB0	88.77	202.79M	0.09	1
ArCapsNet (Katman3) + EffB4	91.00	203.00M	0.28	1

Karmaşıklık analizi, model performansı ile hesaplama verimliliği arasındaki dengeyi anlamayı sağlayan önemli bir bakış sağlar. Orijinal Capsule Forensics modeli

(tabloda Capsule olarak tanımlanmıştır), 57.50% AUC değeri ile düşük bir performans sergilemektedir ve 3.90 milyon parametre ile 11.15 gigaflop hesaplama yüküne sahiptir. VidTr yöntemi, 83.30% AUC ile değerine sahipken 93 milyon parametre ve 117 gigaflop gibi yüksek bir hesaplama karmaşıklığına sahiptir. ISTVT yöntemi ise 84.10% AUC elde etmekte olup, parametre sayısı hakkında bilgi verilmemekle birlikte 455.8 gigaflop gibi yüksek bir hesaplama yüküne sahiptir. Öte yandan bu çalışma kapsamında uygulanan yöntemlerin AUC ölçütünde başarısı kıyaslama yapılan diğer yöntemlerden daha yüksek ve de FLOPs değerleri çok daha düşüktür. Capsule Forensics+VGG19 modeli, performans ve karmaşıklık arasında denge kurarak 97.54% AUC değeri elde etmektedir ve 3.90 milyon parametre ile 8.14 gigaflop hesaplama yüküne sahiptir. Capsule Forensics modeli özellik çıkarıcı olarak Efficient-Net mimarisini kullandığında, verimlilik-performans dengesini daha iyi bir şekilde göstermektedir. Özellikle, Capsule Forensics+EffB4 modeli, sadece 1.86 milyon parametre ve 1.08 gigaflop ile 98.61% AUC değerine ulaşarak, verimli mimarilerin karmaşıklığı azaltırken yüksek performans sağlayabileceğini göstermektedir.

ArCapsNet modelleri ise, farklı katman yapılandırmalarında (Katman1, Katman2 ve Katman 3) VGG19 veya Efficient-Net özellik çıkarıcılarında, 82.89% ile 98.61% arasında değişen AUC değerlerine ulaşmaktadır. ArCapsNet'in farklı katman seviyelerinde parametre sayıları ve FLOP değerleri nispeten sabit kalmakta olup karşılaştırma yapılan diğer çalışmalardan daha düşük FLOPs hesaplama maliyeti getirmektedir.

Önerilen modellerin bir diğer dikkat çekici özelliği, eğitimdeki verimliliğidir. Diğer çalışmalar her video için rastgele seçilmiş birçok görüntü karesi kullanma stratejisini benimserken, önerilen modeller her video için sadece tek bir görüntü karesi kullanmaktadır. Bu yaklaşım, eğitim sürecini basitleştirmenin yanı sıra hesaplama maliyetini de büyük oranda düşürmektedir.

3.5. Elde Edilen DeneySEL Sonuçların İstatistiksel Yöntemler Kullanılarak Değerlendirilmesi

Bu bölümde, füzyon işlemleri yapılarak sonuçların geliştirilmesi çalışmalarının ne derece başarılı olduğu istatistiksel analizler yaparak incelenecektir. Wilcoxon işaretli sıralar testi ve Friedman testi kullanılarak elde edilen tekil modellerin ikili füzyon ve üçlü füzyon sonuçlarından anlamlı bir farkı olup olmadığı tespit edilecektir. Bu testler, elde edilen sonuçların anlamlılık düzeylerini belirlemek için kullanılmıştır.

Wilcoxon işaretli sıralar testi, iki bağlı örneklem grubunun arasındaki farkı değerlendirmek için kullanılan parametrik olmayan bir testtir. Bu test, özellikle normal dağılım varsayımı sağlanmadığında tercih edilir. Celeb-DF ve DFDC-P veri kümeleri için tekli modeller, ikili füzyon modelleri ve üçlü füzyon modelleri içinde her biri için en iyi doğruluk değerini sağlayan durumlar seçilmiştir. Ardından Wilcoxon testi, tekli model performansının çift ve üçlü model füzyonlarından performans farkını değerlendirmek için uygulanmıştır. Tablo 17 Wilcoxon testi sonuçlarını göstermektedir.

Tablo 17. Wilcoxon işaretli sıralar testi

Durum	Test İstatistiği	P-değeri	Sonuç
Tekil model – Çift Füzyon	0	0.001953	Anlamlı fark
Tekil model – Üçlü Füzyon	7	0.066316	Anlamlı fark yok

Wilcoxon testinde elde edilen P-değeri 0.05'ten küçük olursa karşılaştırılan modeller arasında anlamlı fark olduğu, diğer durumda anlamlı bir fark olmadığı sonucu çıkmaktadır. Tablo 17'deki sonuçlardan da görülebileceği gibi tekli model ile ikili füzyon modeli arasında fark varken üçlü füzyon modeli ile arasında anlamlı fark yoktur.

Friedman testi, üç veya daha fazla bağlı örneklem grubunun arasındaki farkı değerlendirmek için kullanılan parametrik olmayan bir testtir. Friedman testi, tekli model, ikili füzyon ve üçlü füzyon sonuçları arasındaki performans farklarını değerlendirmek için uygulanmıştır. Tablo 18'de verilen test sonucunda görüldüğü gibi, Friedman testi sonuçlarına göre farklı model füzyonları arasında anlamlı farklar bulunmaktadır.

Tablo 18. Friedman testi

Test	Test İstatistiği	P-değeri	Sonuç
Friedman testi	13.282	0.001305	Anlamlı farklar mevcut

Wilcoxon işaretli sıralar testi ve Friedman testi sonuçları, model füzyonlarının performansları arasındaki farkların istatistiksel olarak anlamlı olduğunu göstermektedir. Tekli model ile çiftli model füzyonları arasındaki farkın anlamlı olması, çiftli model füzyonlarının performans açısından üstün olduğunu göstermektedir. Ancak, tekli model ile

üçlü model füzyonları arasındaki farkın anlamlı olmaması, üçlü model füzyonlarının ek bir avantaj sağlamadığını göstermektedir.

3.6. Ses Birleştirme Sahteciliği Tespitinde Elde Edilen Sonuçlar

Bu bölümde TIMIT veri kümesinden üretilmiş veri kümesi üzerinde ses birleştirme sahteciliği tespiti için kullanılan ArCapsNet ve Capsule Forensics modelleri ile elde edilen sonuçlar detaylı olarak değerlendirilmiştir. Her iki modelin performansı, doğruluk, kesinlik, duyarlılık ve F1 skoru gibi çeşitli metrikler kullanılarak ölçülmüştür.

ArCapsNet modeli katman 3 seviyesinde eğitilmiştir. Adam optimize edici $\beta 1 = 0.9$, $\beta 2 = 0.999$, değerleriyle kullanılmıştır ve öğrenme oranı 1×10^{-3} olarak seçilmiştir. Mini-paket boyutu 16 olarak ayarlanmıştır. 10 Kat çapraz doğrulama ve çapraz doğrulama olmayan iki farklı model eğitilmiştir. Girdi görüntü boyutları 334x217 olarak belirlenmiştir. Sonuçlar, çapraz doğrulama kullanmanın performansı artırdığını göstermektedir. Çapraz doğrulamalı ArCapsNet modeli ile Doğruluk, Kesinlik, Duyarlılık ve F1 metriklerinde sırasıyla %98.18, %98.59, %97.99 ve %98.29 değerleri elde edilmiştir.

Capsule Forensics modeli eğitiminde birincil kapsül katmanı sayısı 10 ve 30 olarak iki farklı seviyede seçilmiştir. Adam optimize edici $\beta 1 = 0.9$, $\beta 2 = 0.999$, olarak kullanılmıştır. Öğrenme oranı 5×10^{-4} olarak belirlenmiştir. Mini-paket boyutu 16 ve bırakma oranı %5 olarak seçilmiştir. Girdi görüntü boyutları 256x256 piksel boyutuna yeniden ölçeklendirilmiştir. Farklı parametrelere sahip dört farklı Capsule Forensics modeli eğitilmiştir.

- CapsuleForensics-1 : Giriş görüntüleri 256x256 piksel olarak yeniden ölçeklendirilmiştir. Birincil kapsül katmanı sayısı 10'dur ve çapraz doğrulama kullanılmamaktadır.
- CapsuleForensics-2 : Giriş görüntüleri üzerinde yeniden ölçeklendirme yapılmamıştır ve görüntü boyutları 334x217 pikseldir. Birincil kapsül katmanı sayısı 10'dur ve çapraz doğrulama kullanılmamaktadır.
- CapsuleForensics-3 : Giriş görüntüleri üzerinde yeniden ölçeklendirme yapılmamıştır ve görüntü boyutları 334x217 pikseldir. Birincil kapsül katmanı sayısı 30'dur ve çapraz doğrulama kullanılmamaktadır.

- CapsuleForensics-4 : Giriş görüntüleri üzerinde yeniden ölçeklendirme yapılmamıştır ve görüntü boyutları 334x217 pikseldir. Birincil kapsül katmanı sayısı 30'dur ve 10 kat çapraz doğrulama kullanılmaktadır.

ArCapsNet ve Capsule Forensics modelleri ile yapılan test sonuçları Tablo 19'da paylaşılmıştır.

Tablo 19. Ses birleştirme veri kümesi üzerinde elde edilen performans sonuçları

Model	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1 Skoru (%)
ArCapsNet	97.44	98.75	96.42	97.57
ArCapsNet (10 Kat)	98.18	98.59	97.99	98.29
Capsule Forensics 1	80.44	75.95	93.80	83.94
Capsule Forensics 2	87.23	92.91	82.36	87.31
Capsule Forensics 3	89.46	93.39	86.38	89.75
Capsule Forensics 4	90.72	91.35	91.27	91.31

Tablodaki sonuçlardan görüldüğü üzere ArCapsNet modeli daha başarılı sonuçlar elde edebilmiştir. Çapraz doğrulama kullanımının sonuçları iyileştirdiği de görülmektedir. Elde edilen sonuçlar hem ArCapsNet hem de Capsule Forensics modellerinin sahtecilik tespitindeki başarılarının yalnızca video kısmında kalmadığını, ses birleştirme sahteciliğinin tespitinde de başarılı sonuçlar elde edebildiklerini göstermektedir.

3.7. Kopyala-Yapıştır Ses Sahteciliği Tespitinde Elde Edilen Sonuçlar

Bu bölümde Arabic Speech Corpus veri kümesinden yararlanılarak oluşturulan kopyala-yapıştır ses sahteciliği veri kümesi üzerinde elde edilen performans sonuçları değerlendirilmiştir. Arabic Speech Corpus veri kümesinde yer alan orijinal ses dosyaları, belirli segmentler halinde bölünmüş ve bu segmentler, aynı ses dosyasında farklı pozisyonlara yapıştırılarak sahte sesler üretilmiştir.

Burada ses sahteciliği tespiti için kullanılan yöntemde öncelikle veri kümesinde yer alan ses dosyaları mel spektrogram görüntülerine dönüştürülmektedir. Ardından bu görüntülerden Efficient-Net B0 ile özellik çıkarımı gerçekleştirilmiştir ve bu özellikler

Capsule Forensics ağına giriş olarak verilmiştir. Bu ağ, ses dosyalarını "orijinal" veya "sahte" olarak etiketlemiştir.

Eğitim aşamasında, Capsule Forensics ağının birincil kapsül katmanı sayısı 10 olarak belirlenmiş ve modelin eğitimi toplamda 100 epoch boyunca sürdürülmüştür ve en iyi performansı gösteren model seçilerek testlerde kullanılmıştır. Adam optimize edici $\beta_1 = 0.9$, $\beta_2 = 0.999$, değerleriyle kullanılmıştır ve öğrenme oranı 5×10^{-4} olarak seçilmiştir. Mini-paket boyutu 12 olarak ayarlanmıştır ve aşırı öğrenmeyi engellemek amacıyla %5 oranında bırakma (dropout) uygulanmıştır. Modelin girişine verilen Mel spektrogram görüntüleri 434x343 piksel boyutunda olup, bu görüntüler üzerinde eğitim gerçekleştirilmiştir. Modelin performansı, doğruluk, kesinlik ve duyarlılık gibi ölçütlerde değerlendirilmiştir. Uygulanan yöntemle %97.59 doğruluk, %94.24 kesinlik ve %98.50 duyarlılık gibi yüksek performans değerleri elde edilmiştir. Ayrıca veri kümesinde yer alan ses dosyaları çeşitli ek işleme yöntemlerine tabi tutularak tespit için daha zor hale getirilmiştir ve bu şekilde de test işlemleri gerçekleştirilmiştir. Tablo 20’de ek işleme yöntemleri için elde edilen performans sonuçları yer almaktadır.

Tablo 20. Kopyala-yapıştır veri kümesi üzerinde elde edilen performans sonuçları

Saldırı Türü	Doğruluk (%)	Kesinlik (%)	Duyarlılık(%)
Ek Saldırı Yok	93.94	94.44	94.44
Median Filtreleme	93.94	94.44	94.44
64kbps Sıkıştırma	92.42	91.89	94.44
20db Gürültü	96.97	100.00	94.44
30db Gürültü	96.97	100.00	94.44

Burada yer alan ek saldırı türleri :

- **Ek Saldırı Yok:** Bu durumda, herhangi bir ek işlem uygulanmamış ses dosyaları üzerinde yapılan testlerde, model %93.94 doğruluk, %94.44 kesinlik ve %94.44 duyarlılık oranlarına ulaşmıştır. Bu sonuçlar, modelin temel performansını yansıtmaktadır.

- **Median Filtreleme:** Median filtreleme uygulanan sahte ses dosyaları üzerinde yapılan testlerde, modelin doğruluk ve kesinlik oranları saldırı olmayan durumla aynı seviyede kalmıştır. Bu durum, median filtrelemenin modelin performansını olumsuz etkilemediğini göstermektedir.
- **64kbps Sıkıştırma:** Sıkıştırma işlemi uygulanan sahte ses dosyaları üzerinde yapılan testlerde, modelin doğruluk oranı %92.42'ye düşmüş, kesinlik oranı ise %91.89 olarak ölçülmüştür. Bu sonuçlar, sıkıştırma işleminin modelin performansını bir miktar azalttığını göstermektedir, ancak model hala oldukça yüksek bir performans sergilemektedir.
- **20db Gürültü:** 20 desibel gürültü eklenmiş sahte ses dosyaları üzerinde yapılan testlerde, model %96.97 doğruluk ve %100 kesinlik oranlarına ulaşmıştır. Bu durum, modelin gürültü ekleme gibi saldırılara karşı oldukça dayanıklı olduğunu göstermektedir.
- **30db Gürültü:** 30 desibel gürültü eklenmiş sahte ses dosyaları üzerinde yapılan testlerde de benzer şekilde %96.97 doğruluk ve %100 kesinlik oranları elde edilmiştir. Bu da modelin daha yüksek seviyede gürültüye karşı bile performansını koruduğunu göstermektedir.

Capsule Forensics modelinin kopyala-yapıştır ses sahteciliğinde elde ettiği bu başarılı sonuçlar, yöntemin video sahteciliği tespitindeki başarısının yanında ses sahteciliğinde de başarılı sonuçlar verebildiğini göstermektedir.

4. SONUÇLAR

Bu tez çalışmasında, derin sahte video tespiti alanında uygulanan mevcut yöntemler ve geliştirilen yeni yaklaşımlar detaylı bir şekilde analiz edilmiştir. Derin sahte videolar, hızla gelişen yapay zekâ ve derin öğrenme teknolojileri sayesinde giderek daha gerçekçi ve tespit edilmesi zor hale gelmektedir. Bu nedenle sahte video içeriklerinin tespitine yönelik etkili ve güvenilir yöntemlerin geliştirilmesi büyük bir önem taşımaktadır. Çalışmamız, bu bağlamda önemli katkılar sunmakta ve mevcut literatürdeki yöntemlerin ötesinde bir performans sergilemektedir.

Çalışmada literatürde var olan derin sahte video tespit yöntemleri, kamera parmak izi tabanlı, görsel olgulara dayalı, zamansal tutarlılık tabanlı, biyolojik sinyallere dayalı ve genel ağ tabanlı yöntemler olarak beş ana kategori altında detaylıca incelenmiştir. Bu çalışmada kapsamında önerilen yöntem, derin sahte videoların tespiti için dört aşamalı bir süreçten oluşmaktadır. Bu aşamalar, görüntü karesi seçimi, özellik çıkarımı, sınıflandırma ve füzyon işlemidir.

DeneySEL sonuçlar, sadece doğruluk oranları açısından değil, aynı zamanda AUC, F1 skoru gibi önemli performans ölçütlerinde de üstün bir başarı göstermiştir. Özellikle, önerilen füzyon stratejisinin, derin sahte videoların karmaşıklığına rağmen etkili olduğu ve farklı sahtecilik türlerine karşı güçlü bir genel performans sergilediği görülmüştür. Bu durum, yöntemin sadece belirli bir veri kümesine bağlı kalmadığını, aynı zamanda genellenebilir bir çözüm sunduğunu göstermektedir.

Çalışmada ayrıca, farklı füzyon stratejileri ve aktivasyon fonksiyonlarının performans üzerindeki etkileri detaylı bir şekilde incelenmiştir. Ortalama füzyon stratejisinin özellikle karmaşık sahtecilik senaryolarında daha yüksek doğruluk sağladığı, farklı aktivasyon fonksiyonlarının ise belirli koşullarda performansı optimize edebildiği belirlenmiştir. Bu analizler, önerilen yöntemin esnek ve çeşitli uygulama alanlarına uyarlanabilir olduğunu göstermektedir.

Önerilen yöntemin literatürdeki diğer yöntemlerle karşılaştırılması, bu çalışmanın değerini daha da artırmaktadır. Önerilen yöntem karşılaştırmalı sonuçların paylaşıldığı bölümlerde de gösterildiği gibi var olan yöntemlere kıyasla performansı artırırken aynı zamanda hesaplama maliyetinde de büyük düşüş getirmektedir. Özellikle, Celeb-DF ve DFDC-P gibi yaygın olarak kullanılan veri kümeleri üzerinde yapılan deneylerde, bu

yöntemin üstün performansı kanıtlanmıştır. Bu veri kümelerinde elde edilen sonuçlar, önerilen yöntemin mevcut en iyi performans gösteren yöntemlerle karşılaştırıldığında dahi daha yüksek doğruluk ve güvenilirlik sunduğunu ortaya koymuştur. Kapsül ağlarının derin sahte video tespitiindeki etkili kullanımı, bu çalışmanın önemli bir yenilikçi yönünü oluşturmaktadır.

Bu tez çalışması, kapsül ağlarının yalnızca derin sahte video tespiti değil, aynı zamanda ses sahteciliği gibi diğer sahtecilik türlerinde de kullanılabileceğini göstermektedir. Bu, önerilen yöntemin çok yönlü yapısını ve farklı sahtecilik türlerine uygulanabilirliğini ortaya koymaktadır. Özellikle, ses ve görüntü sahteciliğinin bir arada kullanıldığı sahtecilik senaryolarında, bu yöntemin etkili sonuçlar sunma potansiyeline sahip olduğu değerlendirilmektedir.

Sonuç olarak, bu çalışma, derin sahte video tespiti alanında önemli bir adım atmış ve yeni bir yöntem sunarak literatüre değerli bir katkı sağlamıştır. Gelecekte, bu yöntemin daha büyük veri kümeleri üzerinde test edilmesi ve optimize edilmesi, genelleme yeteneğinin artırılması açısından önemli olacaktır.

5. ÖNERİLER

Derin sahte videoların üretimi yeni teknolojilerin gelişmesiyle birlikte daha kolay ve daha gerçekçi sonuçlar üretebilen noktaya yaklaşmaktadır. Üretilen sahte videolar gerçekten ayrımının çıplak gözle tespiti giderek daha zor hale gelmektedir. Sahte videolar toplumsal sorunlara neden olabilecek farklı kullanım senaryolarıyla dolaşıma sunulabilmektedir. Bu nedenle derin sahte videoların tespit edilebilmesi büyük bir ihtiyaç olarak kendini göstermektedir ve bu amaçla gerçekleştirilen çalışmalar artış göstermektedir.

Derin sahte videoların hızlı bir şekilde tespit edilmesi, özellikle sosyal medya platformları ve haber ajansları için büyük önem taşımaktadır. Gerçek zamanlı tespit sistemlerinin geliştirilmesi, sahte içeriklerin yayılmasını engellemede kritik bir rol oynayacaktır. Bu bağlamda daha hızlı ve verimli algoritmaların geliştirilmesi ve bu algoritmaların düşük hesaplama maliyeti ile çalışabilmesi sağlanmalıdır. Videolar çok sayıda görüntü karesinin arka arkaya sıralanması ile oluştuğundan, videonun sahte veya gerçek olduğunun kararı verilirken her bir görüntü karesi için ayrı ayrı karar verilip daha sonra kararlar sayılıp gerçek mi yoksa sahte mi kararı verilmesi durumunu her görüntü karesinin sahtecilikle ilgili kanıtları aynı oranda taşıdığı varsayımına dayanmaktadır. Oysaki sahte bir videoda her görüntü karesi aynı derecede sahtecilik kanıtı içermeyebilir. Daha fazla sahtecilik kanıtı içerebilecek görüntü karelerinin seçilip bu kareler üzerinde karar verilmesi tespit performansını artırmada büyük fayda sağlayabilecektir.

Bu çalışmada kullanılan Celeb-DF ve DFDC-P veri kümelerinin ötesine geçilerek, daha geniş ve çeşitli yeni veri kümeleri üzerinde testler yapılması önerilmektedir. Farklı dillerden, kültürel çeşitlilikten ve çeşitli senaryolardan gelen videoların kullanılması, yöntemin genelleme kabiliyetini artırabilir ve evrensel bir çözüm sunabilir.

Bu çalışma kapsamında geliştirilen yöntem, video sahteciliği üzerine odaklanmış olsa da, ses sahteciliği ve metin sahteciliği gibi diğer sahtecilik türlerine de adapte edilebilir. Multimodal sahtecilik tespiti alanında yapılacak araştırmalar, yöntemin etkinliğini daha da artırabilir ve karmaşık sahtecilik senaryolarında güçlü sonuçlar elde edilmesini sağlayabilir.

Sahtecilik tespitinde uygulanan derin öğrenme tabanlı yöntemlerin karar verme süreçlerinin daha iyi anlaşılması, tespit sonuçlarının güvenilirliğini artıracaktır. Model

içgörülerinin ve karar süreçlerinin açıklanabilirliği üzerine çalışmalar yapılmalıdır. Bu, özellikle yasal ve etik değerlendirmelerde modelin neden belirli bir kararı verdiğinin anlaşılmasını sağlayacaktır.



6. KAYNAKLAR

- Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). Mesonet: a compact facial video forgery detection network. *2018 IEEE international workshop on information forensics and security (WIFS)*, 1-7.
- Aksoylu. (2021). Aksoylu/GoldenFace: An image processing library about calculating face golden ratio, facial cosine similarity and more. İçinde *GitHub*. <https://github.com/Aksoylu/GoldenFace>
- Amerini, I., Galteri, L., Caldelli, R., & Del Bimbo, A. (2019). Deepfake video detection through optical flow based CNN. *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*. <https://doi.org/10.1109/ICCVW.2019.00152>
- Asha, S., Vinod, P., & Menon, V. (2024). A defensive attention mechanism to detect deepfake content across multiple modalities. *Multimedia Systems*, 30. <https://doi.org/10.1007/s00530-023-01248-x>
- Bonettini, N., Bondi, L., Cannas, E. D., Bestagini, P., Mandelli, S., & Tubaro, S. (2020). Video face manipulation detection through ensemble of CNNs. *Proceedings - International Conference on Pattern Recognition*. <https://doi.org/10.1109/ICPR48806.2021.9412711>
- Bulat, A., & Tzimiropoulos, G. (2017). How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks). *Proceedings of the IEEE international conference on computer vision*, 1021-1030.
- Cao, J., Ma, C., Yao, T., Chen, S., Ding, S., & Yang, X. (2022). End-to-End Reconstruction-Classification Learning for Face Forgery Detection. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4103-4112. <https://doi.org/10.1109/CVPR52688.2022.00408>
- Choi, J., Seo, H., Im, S., & Kang, M. (2019). Attention routing between capsules. *Proceedings of the IEEE/CVF international conference on computer vision workshops*, 0.
- Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). StarGAN: Unified Generative Adversarial Networks for Multi-domain Image-to-Image Translation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. <https://doi.org/10.1109/CVPR.2018.00916>
- Ciftci, U. A., Demir, I., & Yin, L. (2020). Fakecatcher: Detection of synthetic portrait videos using biological signals. *IEEE transactions on pattern analysis and machine intelligence*.

- Daichi, Z., Li, C., Lin, F., Zeng, D., & Ge, S. (2021). Detecting Deepfake Videos with Temporal Dropout 3DCNN. *IJCAI*, 1288-1294.
- Dolhansky, B., Howes, R., Pflaum, B., Baram, N., & Ferrer, C. C. (2019). The deepfake detection challenge (dfdc) preview dataset. *arXiv preprint arXiv:1910.08854*.
- Frank, J., Eisenhofer, T., Schönherr, L., Fischer, A., Kolossa, D., & Holz, T. (2020). Leveraging frequency analysis for deep fake image recognition. *International conference on machine learning*, 3247-3258.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139-144.
- Güera, D., & Delp, E. J. (2018). Deepfake video detection using recurrent neural networks. *2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS)*, 1-6.
- He, Z., Zuo, W., Kan, M., Shan, S., & Chen, X. (2019). AttGAN: Facial Attribute Editing by only Changing What You Want. *IEEE Transactions on Image Processing*, 28(11). <https://doi.org/10.1109/TIP.2019.2916751>
- Heidari, A., Navimipour, N. J., Dag, H., Talebi, S., & Unal, M. (2024). A Novel Blockchain-Based Deepfake Detection Method Using Federated and Deep Learning Models. *Cognitive Computation*, 1-19.
- Hernandez-Ortega, J., Tolosana, R., Fierrez, J., & Morales, A. (2020). Deepfakeson-phys: Deepfakes detection based on heart rate estimation. *arXiv preprint arXiv:2010.00400*.
- Jung, T., Kim, S., & Kim, K. (2020). Deepvision: Deepfakes detection using human eye blinking pattern. *IEEE Access*, 8, 83144-83154.
- Kaddar, B., Fezza, S. A., Akhtar, Z., Hamidouche, W., Hadid, A., & Serra-Sagristà, J. (2024). Deepfake Detection Using Spatiotemporal Transformer. *ACM Transactions on Multimedia Computing, Communications and Applications*.
- Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., & Aila, T. (2020). Training generative adversarial networks with limited data. *Advances in Neural Information Processing Systems, 2020-December*.
- Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019-June. <https://doi.org/10.1109/CVPR.2019.00453>
- Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and improving the image quality of stylegan. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. <https://doi.org/10.1109/CVPR42600.2020.00813>

- Kaya, K. S., Türk, B., Cankaya, M., Seyhun, N., & Coşkun, B. U. (2019). Assessment of facial analysis measurements by golden proportion. *Brazilian journal of otorhinolaryngology*, 85, 494-501.
- Khalil, S. S., Youssef, S. M., & Saleh, S. N. (2021). iCaps-Dfake: An integrated capsule-based model for deepfake image and video detection. *Future Internet*, 13(4), 93.
- Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.
- Koopman, M., Rodriguez, A. M., & Geradts, Z. (2018). Detection of deepfake video manipulation. *The 20th Irish machine vision and image processing conference (IMVIP)*, 133-136.
- Li, L., Bao, J., Zhang, T., Yang, H., Chen, D., Wen, F., & Guo, B. (2020). Face x-ray for more general face forgery detection. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5001-5010.
- Li, Y., & Lyu, S. (2018). Exposing deepfake videos by detecting face warping artifacts. *arXiv preprint arXiv:1811.00656*.
- Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-df: A large-scale challenging dataset for deepfake forensics. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3207-3216.
- Liu, D., Dang, Z., Peng, C., Zheng, Y., Li, S., Wang, N., & Gao, X. (2023). FedForgery: generalized face forgery detection with residual federated learning. *IEEE Transactions on Information Forensics and Security*.
- Liu, M., Ding, Y., Xia, M., Liu, X., Ding, E., Zuo, W., & Wen, S. (2019). STGAN: A unified selective transfer network for arbitrary image attribute editing. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019-June*. <https://doi.org/10.1109/CVPR.2019.00379>
- Lugstein, F., Baier, S., Bachinger, G., & Uhl, A. (2021). PRNU-based deepfake detection. *Proceedings of the 2021 ACM workshop on information hiding and multimedia security*, 7-12.
- Malik, A., Kuribayashi, M., Abdullahi, S. M., & Khan, A. N. (2022). DeepFake Detection for Human Face Images and Videos: A Survey. İçinde *IEEE Access* (C. 10). <https://doi.org/10.1109/ACCESS.2022.3151186>
- Masi, I., Killekar, A., Mascarenhas, R. M., Gurudatt, S. P., & AbdAlmageed, W. (2020). Two-branch recurrent network for isolating deepfakes in videos. *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII* 16, 667-684.
- Montserrat, D. M., Hao, H., Yarlagadda, S. K., Baireddy, S., Shao, R., Horváth, J., Bartusiak, E., Yang, J., Guera, D., Zhu, F., & others. (2020). Deepfakes detection

- with automatic face weighting. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 668-669.
- Nguyen, D., Mejri, N., Singh, I. P., Kuleshova, P., Astrid, M., Kacem, A., Ghorbel, E., & Aouada, D. (2024). LAA-Net: Localized Artifact Attention Network for High-Quality Deepfakes Detection. *arXiv preprint arXiv:2401.13856*.
- Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Use of a capsule network to detect fake images and videos. *arXiv preprint arXiv:1910.12467*.
- Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., & Allerhand, M. (1992). Complex Sounds and Auditory Images. *İçinde Auditory Physiology and Perception*. <https://doi.org/10.1016/b978-0-08-041847-6.50054-x>
- Qi, H., Guo, Q., Juefei-Xu, F., Xie, X., Ma, L., Feng, W., Liu, Y., & Zhao, J. (2020). DeepRhythm: Exposing deepfakes with attentional visual heartbeat rhythms. *Proceedings of the 28th ACM international conference on multimedia*, 4318-4327.
- Ranjan, R., Patel, V. M., & Chellappa, R. (2017). Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE transactions on pattern analysis and machine intelligence*, 41(1), 121-135.
- Rehaan, M., Kaur, N., & Kingra, S. (2024). Face manipulated deepfake generation and recognition approaches: a survey. *İçinde Smart Science* (C. 12, Sayı 1). <https://doi.org/10.1080/23080477.2023.2268380>
- Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). Faceforensics++: Learning to detect manipulated facial images. *Proceedings of the IEEE/CVF international conference on computer vision*, 1-11.
- Roy, R., Joshi, I., Das, A., & Dantcheva, A. (2022). 3D CNN Architectures and Attention Mechanisms for Deepfake Detection. *İçinde Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks* (ss. 213-234). Springer International Publishing. https://doi.org/10.1007/978-3-030-87664-7_10
- Sabir, E., Cheng, J., Jaiswal, A., AbdAlmageed, W., Masi, I., & Natarajan, P. (2019). Recurrent convolutional strategies for face manipulation detection in videos. *Interfaces (GUI)*, 3(1), 80-87.
- Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. *Advances in neural information processing systems*, 30.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Soukupova, T., & Cech, J. (2016). Eye blink detection using facial landmarks. *21st computer vision winter workshop, Rimske Toplice, Slovenia*, 2.

- Sun, K., Yao, T., Chen, S., Ding, S., Li, J., & Ji, R. (2022). Dual contrastive learning for general face forgery detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36, 2316-2324.
- Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *International conference on machine learning*, 6105-6114.
- Thies, J., Zollhöfer, M., & Nießner, M. (2019). Deferred neural rendering: Image Synthesis using Neural Textures. *ACM Transactions on Graphics*, 38(4). <https://doi.org/10.1145/3306346.3323035>
- Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2019). Face2Face: Real-time face capture and reenactment of RGB videos. *Communications of the ACM*, 62(1). <https://doi.org/10.1145/3292039>
- Tolosana, R., Romero-Tapiador, S., Fierrez, J., & Vera-Rodriguez, R. (2021). Deepfakes evolution: Analysis of facial regions and fake detection performance. *international conference on pattern recognition*, 442-456.
- Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A Survey of face manipulation and fake detection. *Information Fusion*, 64. <https://doi.org/10.1016/j.inffus.2020.06.014>
- Wang, C., & Deng, W. (2021). Representative forgery mining for fake face detection. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14923-14932.
- Wodajo, D., & Atnafu, S. (2021). *Deepfake Video Detection Using Convolutional Vision Transformer*.
- Wu, J., Zhu, Y., Jiang, X., Liu, Y., & Lin, J. (2024). Local attention and long-distance interaction of rPPG for deepfake detection. *The Visual Computer*, 40(2), 1083-1094.
- Wu, X., Xie, Z., Gao, Y., & Xiao, Y. (2020). Sstnet: Detecting manipulated faces through spatial, steganalysis and temporal features. *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2952-2956.
- Yang, X., Li, Y., & Lyu, S. (2019). Exposing deep fakes using inconsistent head poses. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 8261-8265.
- Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters*, 23(10), 1499-1503.
- Zhang, S., Zhu, X., Lei, Z., Shi, H., Wang, X., & Li, S. Z. (2017). S3fd: Single shot scale-invariant face detector. *Proceedings of the IEEE international conference on computer vision*, 192-201.

- Zhang, Y., Li, X., Liu, C., Shuai, B., Zhu, Y., Brattoli, B., Chen, H., Marsic, I., & Tighe, J. (2021). *VidTr: Video Transformer Without Convolutions*.
- Zhao, C., Wang, C., Hu, G., Chen, H., Liu, C., & Tang, J. (2023). ISTVT: Interpretable Spatial-Temporal Video Transformer for Deepfake Detection. *IEEE Transactions on Information Forensics and Security*, 18, 1335-1348. <https://doi.org/10.1109/TIFS.2023.3239223>
- Zhao, Y., Ge, W., Li, W., Wang, R., Zhao, L., & Ming, J. (2020). Capturing the persistence of facial expression features for deepfake video detection. *Information and Communications Security: 21st International Conference, ICICS 2019, Beijing, China, December 15–17, 2019, Revised Selected Papers 21*, 630-645.
- Zhou, P., Han, X., Morariu, V. I., & Davis, L. S. (2017). Two-stream neural networks for tampered face detection. *2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW)*, 1831-1839.

ÖZGEÇMİŞ

Samet DİNÇER, İlköğrenimini İskenderpaşa İlköğretim Okulu'nda bitirdikten sonra lise eğitimini 2011 yılında Kanuni Anadolu Lisesi'nde tamamladı. 2011-2012 öğretim yılında Karadeniz Teknik Üniversitesi Bilgisayar Mühendisliği bölümünde lisans eğitimine başladı. 2017 yılında bu bölümden mezun oldu. 2021 yılında Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim dalında tezli yüksek lisans programına başladı. 2022 yılından itibaren Karadeniz Teknik Üniversitesi Bilgisayar Mühendisliği Bölümü'nde araştırma görevlisi olarak görev yapmaktadır. İyi derece İngilizce bilmektedir.

Dincer, S., Ulutas, G., Ustubioglu, B., Tahaoglu, G., & Sklavos, N. (2024). Golden ratio based deep fake video detection system with fusion of capsule networks. Computers and Electrical Engineering, 117, 109234.