İSTANBUL TOPKAPI ÜNİVERSİTESİ

Ders/Dönem:
FET312 - Derin Öğrenme / 2025-2026 Güz Dönemi

Proje Başlığı:
Derin Evrişimli Sinir Ağları ile Deepfake Video Tespiti için Temel Model (Baseline Model) Geliştirilmesi ve Analizi

Ekip Adı: Detectify

Ekip Üyeleri:

- Khaiitmurod Khabibullayev 22040101116
- Abdumajid Abdulkhaev 22040101002
- Muhammed Ali Cüre 23040101006

GitHub Repo Bağlantısı:

https://github.com/murat-khabibullayev/Deepfake-detection-project/tree/main

Problem Tanımı & Motivasyon

İş/Bilimsel Soru:

Günümüzde "Deepfake" teknolojisi, üretken yapay zeka modellerinin gelişmesiyle birlikte gerçeği manipüle etme konusunda ciddi bir tehdit oluşturmaktadır. Bu proje, manipüle edilmiş videoların (fake) gerçek videolardan (real) ayırt edilmesini sağlayan otomatik bir tespit sistemi geliştirmeyi amaçlamaktadır. Temel problemimiz, insan gözüyle ayırt edilmesi zorlaşan görsel manipülasyonların, piksel düzeyindeki tutarsızlıklar kullanılarak tespit edilip edilemeyeceğidir.

Görev Türü:

Projemiz, **İkili Sınıflandırma (Binary Classification)** problemidir. Girdi olarak alınan video kareleri analiz edilerek çıktının "Gerçek (Real)" veya "Sahte (Fake)" olduğu belirlenir.

Hedef Değişkenler:

- **Hedef Değişkenler:**
  Projemiz bir ikili sınıflandırma (binary classification) problemi üzerine kuruludur. Model çıktısı, videonun manipüle edilip edilmediğini gösteren olasılıksal bir değer veya sınıf etiketidir:
- **Sınıf 0 (Negative Class - Real):** Orijinal, herhangi bir dijital manipülasyona uğramamış video dizileri.
- **Sınıf 1 (Positive Class - Fake):** Deepfakes, Face2Face, FaceSwap veya NeuralTextures yöntemleri kullanılarak yüz bölgesi değiştirilmiş veya manipüle edilmiş videolar.
  Modelin çıktısı
- $P(y=1 \mid x)$ olasılığını verecek, 0.5 eşik değeri (threshold) kullanılarak nihai karar verilecektir.
- **Başarı Kriterleri:**
  Bu proje iki aşamalı (Vize ve Final) olarak tasarlandığından, her aşama için farklı kriterler belirlenmiştir:
- **Vize Aşaması (Baseline):** Sıfırdan eğitilen sığ (shallow) CNN modelleri için hedefimiz, rassal tahminin (random guessing = %50) üzerine çıkarak modelin manipülasyon izlerini öğrenmeye başladığını kanıtlamaktır. Hedefimiz eğitim setinde **%60-%65** arası bir doğruluk yakalamak ve F1 skorunun 0.50'nin üzerinde olmasıdır.
- **Final Aşaması (Advanced):** Transfer learning kullanılacak modellerle hedefimiz, literatürdeki standartlara yaklaşarak **%85 üzeri doğruluk** ve **0.80 üzeri AUC (Alan Altındaki Alan)** değerine ulaşmaktır.

Proje Yönetimi

Önemli Noktalar ve Zaman Çizelgesi:

- **1. Hafta (20-27 Ekim):** Proje konusunun belirlenmesi ve FaceForensics++ veri setinin seçilmesi.
- **2. Hafta (3-10 Kasım):** Veri setinin incelenmesi, indirilmesi ve Google Drive üzerinde yapılandırılması. Veri ön işleme stratejilerinin belirlenmesi.
- **3-4. Hafta (11-24 Kasım):** Her grup üyesi için farklı temel CNN mimarilerinin tasarlanması, kodlanması ve 500 videoluk alt küme üzerinde eğitilmesi.
- **5. Hafta (25-30 Kasım):** Eğitilen temel modellerin performans analizlerinin yapılması, metriklerin hesaplanması ve vize raporunun hazırlanması.

Roller ve Sorumluluklar:

Ekibimiz, farklı temel model yaklaşımlarını karşılaştırmak amacıyla aşağıdaki şekilde iş bölümü

yapmıştır:

- **Khaiitmurod Khabibullayev:** Veri işleme boru hattını (pipeline) kurmuş; 3 evrişim katmanlı, Dropout destekli "SimpleCNN" temel modelini geliştirmiş ve analiz etmiştir. Final aşamasında **InceptionV3** mimarisi üzerinde çalışmayı planlamaktadır.
- **Muhammed Ali Cüre:** 4 evrişim katmanlı, Batch Normalization ve LeakyReLU aktivasyon fonksiyonlarını içeren "BatchNormCNN" modelini tasarlamış ve geliştirmiştir. Final aşamasında **ResNet50** mimarisi üzerinde çalışmayı planlamaktadır.
- **Abdumajid Abdulkhaev:** Girdi boyutunu optimize ederek (128x128) ve Rescaling katmanı kullanarak özelleştirilmiş bir CNN mimarisi tasarlamış ve test etmiştir. Final aşamasında hafif ve hızlı **MobileNetV2** mimarisi üzerinde çalışmayı planlamaktadır.

Çıktılar:
Proje sonunda oluşturulacak çıktılar:

- **https://github.com/murat-khabibullayev/Deepfake-detection-project/tree/main**

İlgili Çalışmalar (Mini Literatür Taraması)

Bu projede temel aldığımız ve problemimizi şekillendiren birkaç önemli çalışma bulunmaktadır:

1. **FaceForensics++: Learning to Detect Forged Facial Images:** Bu makale, projemizde kullandığımız FaceForensics++ veri setini tanıtan temel çalışmadır. Araştırmacılar, bu geniş ölçekli veri setini oluşturmuş ve XceptionNet gibi derin öğrenme modellerinin deepfake tespitinde ne kadar başarılı olabileceğini göstermişlerdir. Bizim çalışmamız, onların kullandığı karmaşık modeller yerine, daha basit, sıfırdan tasarlanmış modellerin sınırlarını test etmesiyle bu çalışmadan ayrılmaktadır.
2. **MesoNet: a Compact Facial Video Forgery Detection Network:** Bu çalışma, deepfake tespiti için özel olarak tasarlanmış, ResNet veya Inception gibi devasa mimarilere göre daha hafif ve kompakt bir CNN mimarisi olan MesoNet'i önermektedir. Bizim yaklaşımımız da benzer şekilde basit bir model tasarlamayı içerse de, biz belirli bir mimariyi kopyalamak yerine, temel CNN bloklarını kullanarak kendi denemelerimizi yapmaktayız.

Projemiz, bu büyük ve karmaşık veri seti üzerinde, literatürdeki devasa modeller yerine, bir temel seviye (benchmark) oluşturmak amacıyla çok derin olmayan, özel tasarlanmış CNN'lerin performansını analiz ederek literatürdeki bir boşluğu doldurmaktadır.

Veri Açıklaması ve Yönetimi

Veri Kümesi Açıklaması:

- **Veri Kümesi:** FaceForensics++ (C23 Sıkıştırma Seviyesi).
- **Kaynak:** Kaggle (xdxd003/ff-c23) ve TU Munich.
- **Bağlantı:** https://www.kaggle.com/datasets/xdxd003/ff-c23
- **Lisans:** Veri seti, akademik ve ticari olmayan araştırma amaçlı kullanıma açıktır.

Veri seti .mp4 formatında videolardan oluşmaktadır. Proje kapsamında:

- **Eğitim Seti:** 250 Gerçek, 250 Sahte (Her manipülasyon tekniğinden 50'şer adet) olmak üzere toplam 500 video.

- **Test Seti:** 25 Gerçek, 25 Sahte (Deepfakes) olmak üzere 50 video.
- Videolardan saniyede belirli aralıklarla kareler çekilmiş ve yüzler face_recognition kütüphanesi ile tespit edilip 224x224 piksel boyutuna indirgenmiştir.

Etik, Gizlilik, Önyargı:
Veri seti, internetteki halka açık videolardan oluşturulmuştur ve kişilerin yüzlerini içermektedir. Bu veriler, deepfake teknolojisinin yarattığı dezenformasyon tehdidiyle mücadele etmek gibi etik bir amaç doğrultusunda, akademik araştırma için kullanılmaktadır. Projemiz de aynı etik çerçeve içinde yürütülmektedir.

Yöntemler ve Mimari

Genel Veri İşleme Hattı (Pipeline):
Tüm grup üyeleri ortak bir veri hazırlama süreci izlemiştir. Videolardan belirli aralıklarla kareler (frames) çıkarılmış, face_recognition kütüphanesi ile yüz tespiti yapılmış ve yüzler kırpılmıştır. Ancak modellerin tasarımına göre görüntülerin yeniden boyutlandırılması (resize) aşamasında farklılaşılmıştır.

Veri Ön İşleme:

1. Seçilen 500 video, saniyede belirli sayıda kareye bölündü.
2. Her kareden face_recognition kütüphanesi kullanılarak insan yüzü tespit edildi.
3. Tespit edilen yüzler, tüm modeller için standart olarak 224x224 piksel boyutunda kırpılarak kaydedildi.
4. Bu yüz resimleri, real ve fake olarak iki ayrı klasörde toplandı.

1.Grup üyesi yaklaşımı (Khaiitmurod - Simple CNN):

- **Model Mimarisi:** 3 konvolüsyon katmanlı (Conv2d + ReLU + MaxPool) temel bir yapı kullanılmıştır.
- **Girdi Boyutu:** 224x224 piksel.
- **Öne Çıkan Özellik:** Tam bağlantı (Fully Connected) katmanında aşırı öğrenmeyi (overfitting) engellemek için **%50 Dropout** uygulanmıştır. Aktivasyon fonksiyonu olarak standart ReLU tercih edilmiştir.

2.Grup üyesi yaklaşımı (Muhammed Ali - BatchNorm CNN):

- **Model Mimarisi:** Daha derin bir özellik çıkarımı yapabilmek için **4 konvolüsyon katmanı** tasarlanmıştır.
- **Girdi Boyutu:** 224x224 piksel.
- **Öne Çıkan Özellik:** Her konvolüsyon katmanından sonra eğitimin daha stabil olması ve hızlı yakınsaması için **Batch Normalization (Toplu Normalizasyon)** katmanı eklenmiştir. Ayrıca "ölü nöron" problemini engellemek adına ReLU yerine **LeakyReLU** aktivasyon fonksiyonu kullanılmıştır.

3.Grup üyesi yaklaşımı (Abdulmecit - Rescaled Custom CNN):

- **Model Mimarisi:** 3 bloklu (Conv2D + MaxPooling) bir yapı kurulmuştur ancak filtre sayıları ve yoğun katman (Dense) boyutları diğer modellerden farklıdır.
- **Girdi Boyutu:** Diğer modellerden farklı olarak işlem yükünü azaltmak amacıyla görüntüler **128x128** boyutuna indirilmiştir.
- **Öne Çıkan Özellik:** Modelin giriş katmanına, piksel değerlerini [0-1] aralığına normalize eden yerleşik bir **Rescaling** katmanı eklenmiştir. Sınıflandırıcı kısmında daha kompakt (128 nöronlu) bir yapı kullanılarak modelin daha az parametre ile öğrenmesi hedeflenmiştir.

DENEY TASARIMI VE SONUÇLAR

Tüm modeller, FaceForensics++ veri setinden ayrılan 50 videoluk (25 Real, 25 Fake) "Hold-out Test Seti" üzerinde değerlendirilmiştir. Videoların sınıflandırılması, videodan çıkarılan karelerin çoğunluk oyu (Majority Voting) yöntemine göre yapılmıştır.

Performans Karşılaştırma Tablosu:

Model Geliştiricisi	Model Mimarisi	Doğruluk (Accuracy)	Hassasiyet (Precision - Fake)	Duyarlılık (Recall - Fake)	F1 Skoru
Khaiitmurod	**Simple CNN**	**%58.00**	**0.83**	**0.20**	**0.32**
Muhammed Ali	**BatchNorm CNN**	**%62.00**	**0.80**	**0.32**	**0.46**
Abdulmecit	**Rescaled CNN**	**%50.00**	**0.50**	**1.00**	**0.67**

Sonuçların Tartışılması ve Yorumlanması:

1. **En İyi Başarı (Muhammet - BatchNorm CNN):**
   Muhammet'in geliştirdiği model, **%62 doğruluk** oranı ile grup içindeki en iyi performansı göstermiştir. Bunun temel nedeni, **Batch Normalization** katmanlarının kullanılmasıdır. Batch Norm, ağın içindeki ağırlık değişimlerini dengeleyerek modelin veri setindeki gürültüye karşı daha dirençli olmasını sağlamış ve SimpleCNN'e göre daha iyi genelleme yapmıştır.
2. **Yüksek Precision - Düşük Recall (Simple CNN - Senin Modelin):**
   Basit CNN modeli %83 gibi yüksek bir **Precision** (Hassasiyet) değerine ulaşmıştır. Bu, modelin bir videoya "Fake" dediğinde, onun gerçekten Fake olma olasılığının çok yüksek olduğunu gösterir. Ancak **Recall** (0.20) değerinin düşük olması, modelin "muhafazakar" davrandığını, yani emin olamadığı birçok Fake videoyu Real olarak etiketlediğini göstermektedir.
3. **Model Sapması / Bias Problemi (Abdulmecit - Rescaled CNN):**
   Abdulmecit'in sonuçları incelendiğinde (Confusion Matrix: Real=0/25 yanlış, Fake=25/25 doğru), modelin **"Trivial Solution"** (Basit Çözüm) noktasına yakınsadığı görülmüştür. Model, ayırt edici özellikleri öğrenmekte zorlanmış ve eğitim sırasında Loss fonksiyonunu minimize etmek için tüm girdileri "Fake" olarak tahmin etme eğilimi göstermiştir (Mode Collapse).
   Olası Nedenleri:

- Giriş boyutunun 128x128'e düşürülmesi, Deepfake tespiti için kritik olan ince detayların (artifacts) kaybolmasına neden olmuş olabilir.
- Model ağırlıkları, yerel bir minimuma (local minima) sıkışmış olabilir.
- Bu durum, basit CNN mimarilerinin karmaşık Deepfake manipülasyonlarını (özellikle sıkıştırılmış videolarda) öğrenmekte yetersiz kalabileceğini, vize sonrası aşamada daha derin mimarilere (ResNet, EfficientNet) geçilmesinin zorunlu olduğunu kanıtlamaktadır.

Projenin Bir Sonraki Aşaması İçin Gelişmiş Mimari Planlaması:

Geliştirdiğimiz "Temel (Baseline) Modeller" ile yaptığımız deneyler, basit ve sığ CNN yapılarının; özellikle düşük çözünürlüklü yüzlerde ve sıkıştırma (compression) gürültüsünün olduğu videolarda "Deepfake" izlerini yakalamakta zorlandığını göstermiştir (Bkz. Deney Sonuçları). Bu nedenle, final projesinde modeli daha derinleştirmek ve önceden eğitilmiş (pre-trained) ağırlıkları kullanmak (Transfer Learning) zorunlu hale gelmiştir.

İleride her grup üyesi, literatürde başarısı kanıtlanmış farklı bir gelişmiş mimari üzerinde uzmanlaşacaktır:

1. **Khaiitmurod Khabibullayev: InceptionV3**
   - **Seçim Nedeni:** Inception modülleri sayesinde farklı boyutlardaki filtreleri (1x1, 3x3, 5x5) paralel olarak kullanarak görüntüdeki hem ince hem de kaba detayları aynı anda yakalayabilme yeteneğine sahiptir. Deepfake analizinde yüzdeki farklı ölçeklerdeki bozulmaları tespit etmek için bu mimari seçilmiştir.
2. **Abdumajid Abdulkhaev: ResNet50 (Residual Networks)**
   - **Seçim Nedeni:** Derin ağlarda karşılaşılan "Vanishing Gradient" (kaybolan gradyan) problemini "Skip Connections" (atlamalı bağlantılar) ile çözen ResNet, görsel özellikleri kaybetmeden derinlemesine analiz yapabilmektedir.
   - **Yaklaşım:** 50 katmanlı bu derin yapı sayesinde, yüzdeki çok ince manipülasyon izlerinin ve doku bozulmalarının (texture artifacts) tespit edilmesi hedeflenmektedir.
3. **Muhammed Ali Cüre: MobileNetV2**
   - **Seçim Nedeni:** Deepfake tespit sistemlerinin mobil cihazlarda veya gerçek zamanlı sistemlerde çalışabilmesi için "hafif" (lightweight) modellere ihtiyaç vardır.
   - **Yaklaşım:** MobileNet, parametre sayısı ve işlem yükü açısından diğer modellere göre çok daha verimlidir. Amacımız, Xception veya ResNet kadar yüksek bir doğruluğa, çok daha düşük hesaplama maliyetiyle ulaşıp ulaşamayacağımızı test etmektir.

Kullanılan Araçlar ve Frameworkler

- **Programlama Dili:** Python 3.x
- **Derin Öğrenme Kütüphanesi:** PyTorch
- **Veri İşleme ve Analiz:** Pandas, NumPy, Scikit-learn
- **Görüntü/Video İşleme:** OpenCV, face_recognition, Pillow
- **Geliştirme Ortamı:** Google Colaboratory (GPU ile)
- **GitHub Repo:** https://github.com/murat-khabibullayev/Deepfake-detection-project/tree/main

Kaynaklar:

A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, “Faceforensics++: Learning to detect forged facial images,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1–11.

D. Afchar, V. Nozick, J. Yamagishi, and I. Echigo, “Mesonet: a compact facial video forgery detection network,” in *2018 IEEE international workshop on information forensics and security (WIFS)*, 2018, pp. 1–7.