www.dergipark.gov.tr
ISSN:2148-3736
El-Cezerî Fen ve Mühendislik Dergisi
Cilt: 9, No: 4, 2022 (1282-1289)

El-Cezerî Journal of Science and Engineering
Vol: 9, No: 4, 2022 (1282-1289)

ECJSE

DOI : 10.31202/ecjse.1133527

Makale / Research Paper

Hafif Evrişimsel Sinir Ağları Kullanılarak Sahte Yüz Görüntülerinin Tespiti

Emre ŞAFAK[1,2a*], Necaattin BARIŞÇI[1b]

[1]Gazi Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü. Ankara/Türkiye
[2]HAVELSAN, ARGE Teknoloji ve İnovasyon Bölümü. Ankara/Türkiye, esafak@havelsan.com.tr

Received/Geliş: 23.06.2022 **Accepted/Kabul:** 18.11.2022

Öz: Sahte yüz bulunan görüntü ve video içerikleri en yaygın dijital manipülasyon türüdür. Genellikle eğlence amaçlı üretilen bu içerikler zararlı sonuçlar doğurabilir. Sahte yüz görüntüsü üretiminde makine öğrenmesi algoritmaları kullanılmaya başlanmıştır. Makine öğrenmesi algoritmaları ile gerçeğe oldukça yakın yüz manipülasyonları yapılabilmektedir. Bu nedenle gerçek ile sahte içeriklerin ayırt edilebilmesi oldukça zorlaşmıştır. Yüz manipülasyonları tüm yüz sentezi, kimlik değiştirme, nitelik manipülasyonu ve ifade değiştirme olmak üzere 4 temel gruba ayrılır. Tüm yüz sentezi ile çekişmeli üretici ağlar kullanılarak gerçekte olmayan yüzler üretilmektedir. Kimlik değiştirme video içerisindeki kişinin yüz görüntüsünün başka bir yüz ile değiştirilmesidir. Nitelik manipülasyonu yüzün cilt, cinsiyet, yaş, gözlük, saç rengi vb. özelliklerinin değiştirilmesidir. İfade değiştirme manipülasyon yöntemi kişinin yüz ifadesinin değiştirilmesidir. Yapılan çalışmada tüm yüz sentezi manipülasyon yöntemi ile üretilen sahte yüz görüntülerinin tespiti için hafif evrişimsel sinir ağları kullanılmıştır. Eğitim işlemi için MobileNet, MobileNetV2, EfficientNetB0 ve NASNetMobile algoritmaları kullanılmıştır. Kullanılan veri setinde FFHQ veri setindeki 70.000 gerçek görüntü ile FFHQ veri seti kullanılarak StyleGAN2 ile üretilen 70.000 sahte görüntü yer almaktadır. Eğitim işleminde modellerin ImageNet veri seti üzerinde eğitilmiş ağırlıkları transfer öğrenme ile tekrar kullanılmıştır. EfficientNetB0 algoritmasında %93,64 başarı oranı ile en yüksek doğruluk oranına ulaşılmıştır.

Anahtar Kelimeler: Sahte Yüz Tespiti, Yüz Manipülasyon Tespiti, Hafif Model, Derin Öğrenme, Transfer Öğrenme.

Detection of Fake Face Images Using Lightweight Convolutional Neural Networks

Abstract: Fake face images and videos are the most common type of digital manipulation. These content which are usually produced for entertainment purposes can have harmful consequences. Machine learning algorithms have started to be used in recent applications in fake face image processing. With machine learning algorithms, realistic facial manipulations can be made. Therefore, it has become very difficult to distinguish between real and fake content. Face manipulations are divided into 4 basic groups; entire face synthesis, face identity manipulation (deepfake), facial attribute manipulation, and facial expression manipulation. Faces that are not real are produced using entire face synthesis generative adversarial networks. In the identity change method the face image of the person in the video is replaced with another face. Attribute manipulation of face can be changed by skin, gender, age, glasses, hair color, etc. changing its properties. Expression manipulation method is to change the facial expression of the person. In this study, lightweight convolutional neural network algorithms were used to detect fake face images produced by entire face synthesis manipulation method. MobileNet, MobileNetV2, EfficientNetB0 and NASNetMobile algorithms were used for the training process. The dataset used includes 70,000 real images in the FFHQ dataset and 70,000 fake images produced with StyleGAN2 using the FFHQ dataset. In the training process, the weights of the models trained on the ImageNet dataset were reused with transfer learning. In the EfficientNetB0 algorithm, the highest accuracy rate was achieved with a accuracy of 93.64%.

Keywords: Fake Face Detection, Face Manipulation Detection, Lightweight Model, Deep Learning, Transfer

Bu makaleye atıf yapmak için
Şafak, E., Barışçı, N., "Hafif Evrişimsel Sinir Ağları Kullanılarak Sahte Yüz Görüntülerinin Tespiti", El-Cezerî Fen ve Mühendislik Dergisi, 2022, 9(4), 1282-1289.

How to cite this article
Şafak, E., Barışçı, N., "Detection of Fake Face Images Using Lightweight Convolutional Neural Networks", El-Cezerî Journal of Science and Engineering, 2022, 9(4), 1282-1289..
ORCID ID: [a]0000-0001-7579-3410, [b]0000-0002-8762-5091

Learning.

1. Giriş

Dijital manipülasyonlar sonucunda oluşturulan sahte görüntü ve video içeriklerinin sayısı giderek artmaktadır. Dijital manipülasyonlar fotoğraf ve videolarda başlangıçta insanları eğlendirmek amacıyla geleneksel yöntemler kullanılarak yapılmaktaydı. Günümüzde ise dijital manipülasyonlar için makine öğrenmesi algoritmaları kullanılmaya başlanmıştır. Makine öğrenmesi algoritmalarının kullanılmasıyla gerçek içerikler ile sahte içeriklerin ayırt edilebilmesi oldukça zorlaşmıştır [1]. Yüz manipülasyonları, dijital manipülasyonlar arasında en yaygın türdür. Yüz manipülasyonları ile kişiler hiç bulunmadıkları bir yerde gösterilebilir veya hiç yapmadıkları konuşmalar yaptırılabilir. Yüz manipülasyonları tüm yüz sentezi, kimlik değiştirme, nitelik manipülasyonu ve ifade değiştirme olmak üzere dört temel gruba ayrılır [2].

Kimlik değiştirme manipülasyon yönteminde videodaki bir kişinin yüzü ile başka bir kişinin yüzü değiştirilmektedir. Kimlik değiştirme yöntemi için genellikle klasik bilgisayar grafiği temelli (yüz değiştirme) ve derin öğrenme teknikleri (derin sahte) kullanılmaktadır. Kimlik değiştirme manipülasyon yöntemi ile hazırlanan veri setleri iki farklı nesile ayrılmıştır. Genel olarak 1.nesil sahte videolar; düşük kaliteli sentezlenmiş yüzleri, sentezlenen sahte maske ile orijinal yüz arasında farklı renk kontrastını, sahte maskenin görünen sınırlarını, orijinal videodan görünen yüz öğelerini, düşük poz varyantlarını ve sıralı çerçeveler arasındaki garip çıktıları içerir. Bu özelliklerin çoğu 2. nesilde geliştirilmiştir [3].

Nitelik manipülasyonu yönteminde yüzün cilt, cinsiyet, yaş, saç rengi vb. özelliklerinin düzenlenmesidir. FaceApp mobil uygulaması bu manipülasyon türünün örneğidir. Kullanıcıların fotoğraflarını hızlı bir şekilde düzenlemesini sağlayan bu uygulama kişilerin gerçek olmayan özellikleri eklemesine imkan vererek diğer kullanıcıların yanıltılmasına neden olmaktadır [2].

İfade değiştirme manipülasyonu kişinin yüz ifadesinin değiştirilmesi veya oynatılmasıdır. Bu yöntemle kullanıcı hiç bulunmadığı bir yerde gösterilerek istenilen konuşmalar yaptırılabilir. Bu yöntemin uygulanması için Face2Face ve NeuralTextures teknikleri kullanılmaktadır. Bu yöntemde kullanıcıları yanıltmaya yönelik zararlı içeriklerin oluşabilme potansiyeli yüksek olduğu için ciddi sonuçlar doğurabilir [2].

Tüm yüz sentezi manipülasyon yönteminde çekişmeli üretici ağlar kullanılarak var olmayan yüz görüntüleri üretilmektedir. Bu yöntem ile gelişmiş algoritmalar sayesinde yüksek çözünürlüklü gerçekçi yüz görüntüleri üretmek mümkündür. Bu manipülasyon yöntemi oyun ve 3B modellemede fayda sağlarken zararlı uygulamalar için de kullanılabilir. Sosyal ağlarda çok gerçekçi sahte profiller oluşturularak yanlış bilgi üretimine neden olunabilir [3]. Yapılan çalışmada çekişmeli üretici ağlar tarafından üretilen sahte yüz görüntülerinin tespit edilmesi sağlanmıştır.

Hong-Shuo Chen vd. [4] tarafından yapılan çalışmada DefakeHop adı verilen hafif ve yüksek performanslı bir derin sahte algılama yöntemi önerilmektedir. Sahte görüntüleri algılamak için derin öğrenme temelli olmayan DefakeHop adı verilen yeni bir sistem geliştirilmiştir. DefakeHop; PixelHop++, özellik damıtma ve topluluk sınıflandırması olmak üzere üç ana modülden oluşmaktadır. DefakeHop yüz görüntülerinin çeşitli bölümlerinden PixelHop++ birimlerini kullanarak özellikleri çıkarır. PixelHop++ düşük çözünürlüklü yüz görüntülerinden özellik öğrenme için kullanılır. PixelHop++ tarafından çıkarılan öznitelikler sınıflandırma için yeterli olmadığından özellik boyutunu düşürmek için özellik damıtma modülü önerilmektedir. Özellik damıtma modülü kullanılarak özellik boyutu önemli ölçüde azaltılır ve yalnızca önemli bilgiler tutulur. Son aşamada farklı bölgeler ve çerçeveler birleştirilerek sınıflandırma yapılır. DefakeHop Celeb-DF veri setinde %93,12 doğruluk oranına ulaşırken Celeb-DFv2 veri setinde %87,65 doğruluk oranına ulaşmıştır.

Ruben Tolosana vd. [5] tarafından yapılan çalışma 1. ve 2. nesil veri setlerinde sahte yüz görüntüsü tespiti ile ilgili yapılan çalışmalarla karşılaştırılmıştır. Önerilen yöntemde sahte yüz görüntülerini analiz etmek için iki farklı yaklaşım incelenmiştir. Bu yaklaşımlardan ilki tüm yüzün analizi ikincisi ise sadece belirli yüz bölgelerinin analiz edilmesidir. İkinci yaklaşımda yüz; burun, ağız, göz ve kalan kısım olmak üzere dört bölgeye ayrılmıştır. Her bölgenin segmentasyonu için açık kaynaklı OpenFace2 aracı kullanılmıştır. Sahte yüz görüntülerini tespit etmek için Xception ve Capsule ağı kullanılmıştır. Xception evrişimsel sinir ağı sahte görüntü tespiti çalışmalarında başarılı sonuçlar verdiği için tercih edilmiştir. Eğitim işleminde Imagenet üzerinde önceden eğitilmiş model kullanılmıştır. Xception modeli 20 adımda eğitildiğinde maksimum doğruluk oranına ulaşılmıştır. Capsule ağı geleneksel evrişimsel sinir ağları ile karşılaştırıldığında daha az parametre gerektirmektedir. Celeb-DF veri setinde yüzün tamamı analiz edildiğinde Xception ile %83,6 Capsule ağı ile %82,46 maksimum doğruluk oranına ulaşılmıştır. DFDC veri setinde yüzün tamamı analiz edildiğinde Xception ile %91,17, Capsule ağı ile %87,45 maksimum doğruluk oranına ulaşılmıştır.

Run Wang vd. [2] tarafından yapılan çalışmada sahte yüz görüntülerinin tespiti için derin öğrenme modelindeki katmanların nöron davranışlarının izlenmesine dayanan yeni bir evrişimsel sinir ağı modeli önerilmektedir. Nöron davranışlarının izlenmesine dayalı yöntem model saldırılarına karşı başarılı sonuçlar vermiştir. Katmanlarda nöron aktivasyonunun uygulanması sahte kalıpların daha iyi yakalanmasını sağlamaktadır. Dört sahte yüz üretme türü için uygulanması sonucu elde edilen sonuçlar mevcut çalışmalardan daha yüksek doğruluk oranına ulaşıldığını göstermektedir. Tüm yüz sentezi manipülasyon yönteminde FFHQ ve CelebA veri setleri kullanılarak StyleGAN2 ile üretilen sahte yüz görüntüsü veri setinde %91,9 doğruluk oranına ulaşılmıştır.

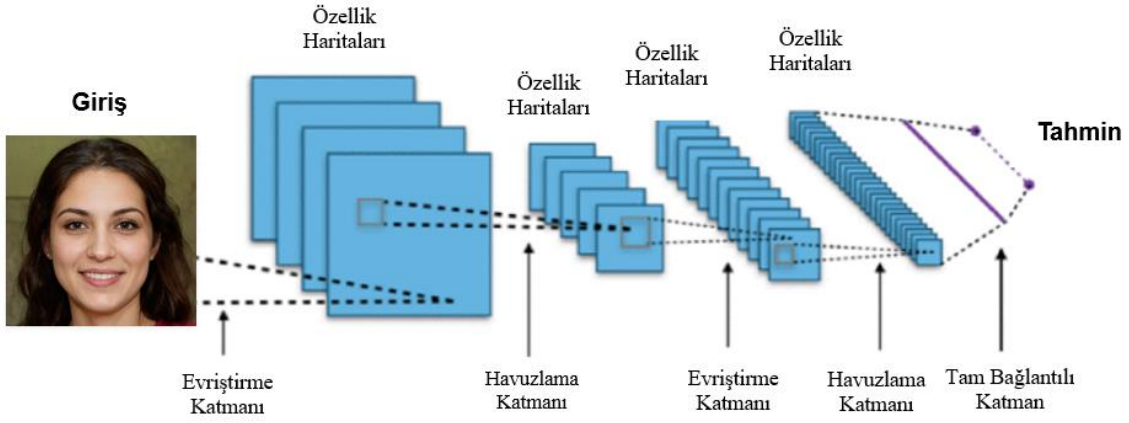
Literatürde tüm yüz sentezi manipülasyon yöntemi için hafif derin öğrenme temelli bir çalışma bulunmamaktadır. Tüm yüz sentezi manipülasyon tespiti için MobileNet, MobileNetV2, EfficientNetB0 ve NASNetMobile evrişimsel sinir ağları kullanılmıştır. Modeller ImageNet veri seti üzerinde ön eğitilmiş olarak transfer öğrenme ile tekrar kullanılmıştır. Bu makalede yapılan çalışmada kullanılan materyal ve metot, araştırma bulguları ve sonuçlar açıklanmıştır.

2. Materyal ve Metot

Sahte yüz görüntülerini tespit edebilmek için evrişimsel sinir ağları kullanılmıştır. Evrişimsel sinir ağı modelini eğitmek için 140.000 görüntüden oluşan veri seti kullanılmıştır. Kullanılan veri seti FFHQ veri setindeki 70.000 gerçek ve StyleGAN2 ile üretilmiş 70.000 sahte yüz görüntüsü içermektedir [6]. FFHQ veri seti çekişmeli üretici ağlar için hazırlanmış yaş ve farklı etnik köken çeşitliliğine sahip yüksek çözünürlüklü görüntülerden oluşan veri setidir [7]. StyleGAN2 veriye dayalı koşulsuz üretken modellemede oldukça başarılı sonuçlar vermektedir [8]. Bu nedenle StyleGAN2 ile oluşturulan sahte yüzler oldukça gerçekçi ve zorlayıcıdır. Anita Rácz vd. [9] tarafından yapılan çalışmada %80 eğitim - %20 test veri seti bölme oranında model performansının maksimuma ulaştığı görülmüştür. Bu nedenle veri setindeki görüntülerin %80'i eğitim ve geçerleme işlemi için kullanılırken %20'si test işlemi için kullanılmıştır [9].

Evrişimsel sinir ağları doğrudan verilerden öğrenen ve manuel özellik çıkarma ihtiyacını ortadan kaldıran derin öğrenmeye yönelik geliştirilmiş yapay sinir ağıdır. Evrişimsel sinir ağları bir girdi görüntüsünü alarak çeşitli bölgelere öğrenilebilir ağırlıklar atayarak öğrenme işlemini gerçekleştirir. Evrişimsel sinir ağları görüntü, zaman serileri, sinyal verileri ve ses tanıma/sınıflandırma konularında başarılı sonuçlar veren derin sinir ağıdır. Özellikle görüntü sınıflandırma alanında kendini kanıtlamıştır. Evrişimsel sinir ağları geliştiriciler tarafından yapılan özellik çıkarma ihtiyacını ortadan kaldırır. Özellikler doğrudan ağ tarafından öğrenilir [10]. Evrişimsel sinir ağı görüntünün farklı özelliklerini öğrenen birçok katmana sahiptir. Her eğitim görüntüsüne farklı

çözünürlükte filtreler uygulanır ve çıktı sonraki katmana giriş olarak alınır. Evrişimsel sinir ağları bir giriş katmanı ve bir çıkış katmanı arasında birçok gizli katmandan oluşur. En yaygın kullanılan katmanlar; evrişim, ReLU, havuzlama ve tam bağlantılı katmandır [11]. Girdi görüntüsünün ilk olarak evrişim katmanında evrişimsel filtreler kullanılarak belirli özellikleri çıkarılır. Evrişim işleminden sonra ReLU katmanında özellik haritasındaki negatif pikseller 0 olarak ayarlanarak ağa doğrusal olmama özelliği getirilir. Havuzlama katmanı girdideki parametre sayısını azaltarak ağın öğrenmesi gereken parametre sayısını düşürür. Son olarak tam bağlantılı katman önceki katmanlardan elde edilen özelliklere dayalı olarak sınıflandırma işlemini gerçekleştirir [12]. Sahte yüz görüntülerinin tespit edilebilmesini sağlayan evrişimsel sinir ağı mimarisinin çalışması Şekil 1'de sunulmuştur.



Şekil 1. Evrişimsel sinir ağının çalışması

Şekil 1'de görüldüğü gibi alınan görüntü evrişimsel sinir ağı katmanlarından geçirilerek özellik haritaları çıkarılır ve bu özellik haritalarına göre tahmin yapılır. Yapılan çalışmada MobileNet, MobileNetV2, NASNetMobile ve EfficientNetB0 evrişimsel sinir ağları kullanılmıştır.

2.1. MobileNet

MobileNet, mobil uygulamalarda kullanılmak üzere geliştirilmiş hesaplama açısından hafif ve verimli evrişimsel sinir ağıdır. MobileNet 28 katman ve 4,2 milyon parametreden oluşmaktadır. MobileNet, derinlemesine ayrılabilir evrişimler kullanır. Derinlemesine ayrılabilir evrişimler geleneksel evrişimler ile karşılaştırıldığında parametre sayısı önemli ölçüde azaltılmıştır. Derinlemesine ayrılabilir evrişim, derinlemesine ve noktasal evrişim olmak üzere iki katmandan oluşur. Derinlemesine evrişim katmanı her bir giriş kanalına tek bir filtre uygulamak için kullanılır. Derinlemesine evrişim katmanın çıktısının doğrusal bir kombinasyonunu hesaplamak için noktasal evrişim katmanı kullanılır [13]. MobileNet transfer öğrenmeyi çalıştırmak veya uygulamak için çok daha az hesaplama gücü gerektirir. MobileNet, tarayıcıların hesaplama, grafik işleme ve depolama konusunda sınırlamaları olduğundan, web tarayıcıları için de en uygunudur.

2.2. MobileNetV2

MobileNetV2, 53 katman ve 3,4 milyon parametreden oluşan evrişimsel sinir ağıdır. MobileNet mimarisi geliştirilerek artık bağlantılar ve darboğaz katmanı eklenmiştir. MobileNetV2 ağında derinlemesine ayrışabilir evrişim bloğu yerine bir darboğaz artık bloğu vardır. Darboğaz artık bloğu 3 katmandan oluşmaktadır. MobileNet evrişimsel sinir ağında bulunan 2 katmana ek olarak darboğaz katmanı yer almaktadır. MobileNet ağında bulunan noktasal evrişim kanal sayısını aynı tutar veya iki katına çıkarır. 1x1 evrişim içeren projeksiyon katmanı kanal sayısını azaltır. Darboğaz katmanı içerisinden geçen veri miktarını azalttığı için bu adı alır. Darboğaz artık bloğu

normalleştirme işlemi için artık bağlantılar içermektedir. MobileNetV2 modelleri MobileNet modellerine göre 2 kat daha az işlem gücü kullanırken çok daha hızlı çalışmaktadır [14].

2.3. NASNetMobile

NASNet, eğitilecek veri kümesi için evrişimsel sinir ağı mimarisinin eğitim işlemi sırasında oluşturulmasını sağlar. Temel olarak filtre boyutu, çıkış kanalları, katman sayısı vb. parametrelerin en iyi kombinasyonunu aramaktadır. Her arama işleminden sonraki ödül, veri seti üzerinde aranan mimarinin doğruluğuna göre hesaplanmaktadır. NASNet'te genel mimari önceden tanımlanmış olsa da hücreler önceden tanımlanmamıştır. NASNet ağında bulunan hücreler normal ve indirgeme hücresidir. Normal hücreler, aynı boyutta bir özellik haritası döndüren evrişimli hücrelerdir. Azaltma hücreleri, özellik haritası yüksekliği ve genişliğinin iki kat azaltıldığı bir özellik haritası döndüren evrişimli hücrelerdir [15]. NASNetMobile, NASNet ağının mobil cihazlar için parametre sayısının düşürüldüğü versiyonudur. NASNetMobile, 12 hücre ve 5,6 milyon parametre içermektedir [16].

2.4. EfficientNetB0

EfficientNet evrişimsel sinir ağı mimarisi 2019 yılında geliştirilmiştir. Model ölçeklendirilmesi sistematik olarak ele alınmıştır. EfficientNet bir bileşik katsayısını kullanarak ağ derinliği, genişliği ve çözünürlüğü dengelenmiş ve daha iyi performans elde edilmiştir. Bileşik ölçekleme yöntemi kullanılarak tüm derinlik, genişlik ve çözünürlük boyutları eşit olarak ölçeklenebilmektedir. Bileşik ölçekleme yönteminde giriş görüntüsü daha büyükse ağın alıcı alanını artırmak için daha fazla katmana ve daha küçük desenleri algılamak için daha fazla kanala ihtiyaç duyduğu tespit edilebilir. EfficientNetB0 ağı 5,3 milyon parametreden oluşmaktadır ve mobil/gömülü cihazlarda kullanımı uygundur. Temel EfficientNetB0, MobileNetV2 ağında bulunan darboğaz artık bloklarını temel almaktadır [17].

2.5. Transfer Öğrenme

Transfer öğrenme önceden eğitilmiş modelin tekrar kullanılabilmesidir. Transfer öğrenme ile yeni modeller eğitilirken eğitilmiş önceki modellerin ağırlıkları kullanılır [18]. Bu sayede her geliştirme bir sonraki çalışmanın temelini oluşturmaktadır. Transfer öğrenme bir görev için eğitilmiş modelin başka bir görevde tekrar kullanılarak hızlı gelişim ve yüksek performans elde edilmesini sağlayan optimizasyondur. Transfer öğrenme görüntü tanıma, ses tanıma ve doğal dil işleme alanlarında yaygın olarak kullanılmaktadır [19]. Yapılan çalışmada sahte yüz tespiti modelini eğitmek için ImageNet üzerinde ön eğitilmiş evrişimsel sinir ağı modelleri kullanılmıştır.

2.6. Tensorflow

Tensorflow, yapay zeka çalışmalarında kullanılan C++ programlama dili ile Google tarafından geliştirilmiş açık kaynak yazılım kütüphanesidir. Tensorflow Linux, macOS, Windows, Android ve iOS işletim sistemlerini desteklemektedir [20]. Tensorflow CPU, GPU ve TPU üzerinde çalışabilmektedir. Tensorflow, görüntü tanıma alanında iyi performans göstermesi ve akademik çalışmalarda yaygın olarak kullanılması nedeniyle yapılan çalışmada tercih edilmiştir.

3. Bulgular ve Tartışma

Yapılan çalışmada Python 3.6 ve Tensorflow kütüphanesi kullanılmıştır. Eğitim ve test işlemi için 70.000 sahte yüz ve 70.000 gerçek yüz görüntüsü içeren veri seti kullanılmıştır. Veri setinin %80'i eğitim %20'si test için kullanılmıştır. Transfer öğrenme yöntemi ile ImageNet üzerinde ön eğitilmiş

modeller kullanılmıştır. İşlem gücü düşük olan cihazlarda maksimum performansta çalışabilecek MobileNet, MobileNetV2, EfficientNetB0 ve NASNetMobile algoritmaları tercih edilmiştir. Literatürdeki mevcut çalışmalarda modeller eğitildikten sonra performansını değerlendirmek için doğruluk, kesinlik, duyarlılık ve f1 skor metrikleri kullanılmıştır. Doğruluk, modelin yaptığı doğru tahmin sayısının toplam girdi sayısına oranıdır. Doğruluk metriği ile modelin sahte ve gerçek yüz görüntülerini doğru etiketleme performansı ölçülmüştür. Kesinlik, modelin doğru tahmin ettiği sahte yüz görüntülerinin doğru veya yanlış sınıflandırılmış toplam sahte yüz tahmini sayısına oranıdır. Duyarlılık, modelin doğru tahmin ettiği sahte yüz görüntülerinin girdi olarak verilen toplam sahte yüz görüntüsüne oranıdır. F1 skor, kesinlik ve duyarlılık değerlerinin harmonik ortalaması alınarak hesaplanır. F1 skor tüm hata durumlarını içermesi sayesinde özellikle eşit dağılıma sahip olmayan veri kümelerinde model performansının daha iyi ölçülmesini sağlar [21].
Eğitilen modelin test işlemi sonucu elde edilen performans metrikleri ve diğer algoritmalar ile karşılaştırılması Tablo 1'de yer almaktadır.

Tablo 1. Sahte yüz görüntüleri tespiti modellerinin performans metrikleri

Algoritma	Doğruluk (%)	Kesinlik (%)	Duyarlılık (%)	F1 Skor (%)	Parametre Sayısı
MobileNet	87,83	93,66	87,25	90,34	4,200,000
MobileNetV2	91,12	92,53	91,87	92,19	3,500,000
EfficientNetB0	93,64	93,70	93,27	93,48	5,300,000
NASNetMobile	78,87	79,12	78,04	78,57	5,600,000

Tablo 1'de görüldüğü gibi EfficientNetB0 algoritması %93,64 doğruluk oranı ile diğer algoritmalarından daha yüksek doğruluk oranına ulaşmıştır.

Şekil 2. EfficientNetB0 modelinin sahte yüz görüntüleri üzerinde yaptığı tahminler

Şekil 2'de görüldüğü gibi model test veri setinden seçilen sahte yüz görüntüleri üzerinde yüksek doğrulukta tahmin yapmıştır. Literatürdeki çalışmalarla EfficientNetB0 modelinin karşılaştırılması Tablo 2'de yer almaktadır.

Tablo 2. Yapılan çalışmanın literatürdeki çalışmalarla karşılaştırılması

Çalışma	Teknik	Algoritma	Özellik	Veri Seti	Değerlendirme Metriği	Başarı Oranı
Yapılan Çalışma	Derin Öğreneme	EfficientNetB0	Sahte Yüz Tespiti	FFHQ	Doğruluk	%93,64
Run Wang vd. [2]	Derin Öğrenme	Nöron izlemesine dayalı yeni model	Sahte Yüz Tespiti	FFHQ ve CelebA	Doğruluk	%91,9
Hong-Shuo Chen vd. [4]	Derin Öğrenme	DefakeHop	Sahte Yüz Tespiti	Celeb-DF ve Celeb-DFv2	Doğruluk	%93,12 (Celeb-DF) - %87,65 (Celeb-DFv2)
Ruben Tolosana vd. [5]	Derin Öğrenme	Xception ve Capsule Ağı	Sahte Yüz Tespiti	Celeb-DF	Doğruluk	%83,6 (Xception) - %82,46 (Capsule)

Tablo 2'de görüldüğü gibi yapılan çalışmada kullanılan EfficientNetB0 evrişimsel sinir ağı ile önceki çalışmalardan daha yüksek doğruluk oranına ulaşılmıştır. Yapılan çalışmada yüksek çözünürlüklü ve çeşitliliğe sahip veri seti kullanılması, zorlayıcı ve gerçekçi sahte yüzlerin StyleGAN2 ile üretilmesi ve transfer öğrenme yöntemi ile önceden eğitilmiş modelin kullanılması sayesinde önceki çalışmalardan daha iyi sonuç vermiştir.

4. Sonuç ve Öneriler

Sahte yüz görüntüsü, yüz görüntülerinin farklı teknikler kullanılarak gerçeğe yakın olarak oluşturulmasıdır. Sahte yüz görüntüsü en yaygın dijital manipülasyon türüdür. Sahte yüz görüntüsü üretiminde makine öğrenmesi algoritmalarının kullanılması ile gerçek görüntülerden ayırt edilmesi giderek zorlaşmaktadır. Bu nedenle sahte yüz görüntüsü tespiti üzerine çalışmalar yapılmaya başlanmıştır. Yüz manipülasyonları tüm yüz sentezi, kimlik değiştirme, nitelik manipülasyonu ve ifade değiştirme olmak üzere 4 temel gruba ayrılır. Yapılan çalışmada tüm yüz sentezi ile üretilen sahte yüz görüntüleri MobileNet, MobileNetV2, EfficientNetB0 ve NASNetMobile evrişimsel sinir ağları kullanılarak tespit edilmiştir. Bu sayede geliştirilen model mobil cihazlar üzerinde gerçek zamanlı sahte yüz görüntüsü tespiti yapabilir. Eğitim işlemi için 70.000 gerçek ve 70.000 sahte görüntü içeren veri seti kullanılmıştır. Modeller ImageNet üzerinde ön eğitilmiş olarak kullanılmıştır. EfficientNetB0 evrişimsel sinir ağında %93,64 ile en yüksek doğruluk oranına ulaşılmıştır. Gelecekte evrişimsel sinir ağı modelleri revize edilerek doğruluk oranı artırılabilir. Ayrıca sonraki çalışmalarda 4 temel yüz manipülasyonunun tespit edilmesi sağlanabilir.

Yazar(lar)ın Katkıları

Her iki yazar da makalenin son halini okudu ve onayladı.

Çıkar Çatışması

Yazarlar, çıkar çatışması olmadığını beyan eder.

Kaynaklar

[1]. Pashine, S., Mandiya, S., Gupta, P. and Sheikh, R., "Deep Fake Detection : Survey of Facial Manipulation Detection Solutions", arXiv preprint arXiv:2106.126, 2021.

[2]. Wang, R., Juefei-Xu, F., Ma, L., Xie, X., Huang, Y., Wang, J. and Liu, Y., "FakeSpotter: A Simple yet Robust Baseline for Spotting AI-Synthesized Fake Faces", arXiv preprint arXiv:1909.06122, 2020.

[3]. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A. and Ortega-Garcia, "Deepfakes and beyond: A Survey of face manipulation and fake detection", Information Fusion, vol. 64, pp. 131-148, 2020.

[4]. Chen, H.-S., Rouhsedaghat, M., Ghani, H., Hu, S., You, S. and Kuo, C.-C. J., "DefakeHop: A Light-Weight High-Performance Deepfake Detector", arXiv preprint arXiv:2103.06929, 2021.

[5]. Tolosana, R., Romero-Tapiador, S., Fierrez, J. And Vera-Rodriguez, R., "DeepFakes Evolution: Analysis of Facial Regions and Fake Detection Performance", arXiv preprint arXiv:2004.07532, 2020.

[6]. StyleGAN2, [online] Available: https://github.com/NVlabs/stylegan2 [Accesed: 21.2.2022].

[7]. Flickr-Faces-HQ Dataset (FFHQ), [online] Available: https://github.com/NVlabs/ffhq-dataset [Accesed: 21.2.2022].

[8]. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J. and Aila, T., "Analyzing and Improving the Image Quality of StyleGAN", arXiv preprint arXiv:1912.04958, 2020.

[9]. Rácz, A., Bajusz, D. And Héberger, K., "Effect of Dataset Size and Train/Test Split Ratios in QSAR/QSPR Multiclass Classification", Molecules, vol. 26, no. 4, 2021.

[10]. Albawi, S., Mohammed, T. A. and Al-Zawi, S., "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), pp. 1-6, 2017.

[11]. Şafak, E. and Barışçı, N., "Age and Gender Prediction Using Convolutional Neural Networks", 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara-Türkiye, 19-21 Ekim, 2018.

[12]. Arı, A. and Hanbay, D., "Tumor detection in MR images of regional convolutional neural networks " , Journal of the Faculty of Engineering and Architecture of Gazi University, vol. 34, no. 3 1395-1408, 2019.

[13]. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications", arXiv preprint arXiv:1704.04861, 2017.

[14]. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L. C., "MobileNetV2: Inverted Residuals and Linear Bottlenecks", arXiv preprint arXiv:1801.04381, 2019.

[15]. Zoph, B., Vasudevan, V., Shlens, J. and Le, Q. V., "Learning Transferable Architectures for Scalable Image Recognition", arXiv preprint arXiv:1707.07012, 2018.

[16]. Saxen, F., Werner, P., Handrich, S., Othman, E., Dinges, L. and Al-Hamadi, A., "Face Attribute Detection with MobileNetV2 and NasNet-Mobile," 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), pp. 176-180, 2019.

[17]. Tan, M. and Le, Q.V., "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", arXiv preprint arXiv:1905.11946, 2019.

[18]. Bozinovski, S. and Fulgosi, A., "The influence of pattern similarity and transfer of learning upon training of a base perceptron B2", Proc. Symp. Informatica 3-121-5, Bled, 1976.

[19]. Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H. and He, Q., "A Comprehensive Survey on Transfer Learning", IEEE, vol. 109, no. 1, pp. 43-76, 2021.

[20]. Tensorflow, [online] Available: https://www.tensorflow.org/ [Accesed: 21.2.2022].

[21]. Güngör, S., Kaya, M. and Alhajj, R., "A Deep Learning Based Method for Detecting Covid-19 from Colorized CT Images", IDAP-2021 : 5th International Artificial Intelligence and Data Processing symposium, 391-399, 2021.