



Düzce Üniversitesi Bilim ve Teknoloji Dergisi

Araştırma Makalesi

Derin Öğrenme Teknikleri Kullanılarak Sahte Yüz Fotoğrafı ve Videosu Sentezi¹

 Mustafa Salih BAHAR^{a,*},  Ercan BULUŞ^b

^a Bilgisayar Mühendisliği Bölümü, Fen Bilimleri Enstitüsü, Yıldız Teknik Üniversitesi, İstanbul, TÜRKİYE

^b Bilgisayar Mühendisliği Bölümü, Fen Bilimleri Enstitüsü, Tekirdağ Namık Kemal Üniversitesi, Tekirdağ, TÜRKİYE

* Sorumlu yazarın e-posta adresi: mustafasalihbhr@gmail.com

DOI:10.29130/dubited.1017584

ÖZ

Çalışmadaki amaçla; bir kişinin gerçekte var olan kısa bir videosundan veya birkaç fotoğrafından bile, o kişinin yüz fotoğrafını içeren sahte videolar oluşturulabileceği kanıtlanmaktadır. Bu birkaç fotoğraf veya kısa bir video alınıp derin öğrenme teknikleriyle eğitilerek sahte fotoğraflar ve videolar oluşturulabilir. Sahte videolarda kişinin yüzüyle başka bir kişinin yüz değişimi uygulanabilir veya kişinin yüzüne yeniden canlandırma (hareketlendirme) yapılabilir. Yeniden canlandırmada ise kaynak bir kişinin yüzüne başka bir kişinin videosundaki yüz hareketleri uygulanabilir. Hatta StyleGAN gibi teknikler ile gerçek insan yüz fotoğraflarından oluşan bir fotoğraf kümesi kullanılarak var olmayan insan fotoğrafları bile üretilebilir. Yaygın olarak Derin sahtelik (Deepfake) teknolojisi olarak bilinen bu teknikler, bu çalışmada yüzde kullanılan çeşitleri ve yapıları ile birlikte ele alınmıştır. Bu teknikler, eski dönemlerde yaşamış bilim adamlarının, ünlülerin var olan fotoğraflarına yeniden canlandırma yapıp konuşturularak çocuklar için eğitim amaçlı kullanılabilir. Kuklacılıkta bu yöntem kullanılabilir. Oyuncuların (Aktörlerin-Aktrislerin) yerine sahnelerde, onların fotoğraflarıyla bilgisayarda yeniden canlandırma yapılabilir. Portreler canlandırılabilir (hareketlendirilebilir). Bu çalışmanın benzer araştırmalardan ayrıldığı nokta ise eğitim için kullanılacak olan verinin diğer çalışmalardaki verilerden daha az olması ve sahte video oluşturma çeşitlerinin, yapılarının birlikte ele alınmasıdır. Veri eğitimi için kullanılan materyal GPU ve veri seti olarak ise VoxCeleb veri seti, birkaç kısa video ve birkaç fotoğraftan oluşmaktadır. Kullanılan yöntem ise Çekişmeli üretici ağlar ve Otomatik kodlayıcılar gibi üretken ağlardır. Yapılan çalışma kullanılan video ve fotoğraflarda yüzün karşıya (öne) dönük veya hafif sağa ya da hafif sola dönük iken, yüz hareketinin belirli bir alanda sınırlı olduğunda ve yüzün yavaş hareket ettiğinde yapay zekayı daha iyi eğittiği ve bu eğitim verileri kullanılarak oluşturulan sahte videoların daha başarılı olduğunu göstermiştir.

Anahtar Kelimeler: Yüz değiştirme, Yüz canlandırma, Derinsahte video, Derin öğrenme.

Fake face photo and video synthesis using deep learning techniques

ABSTRACT

For the purpose of the study; It is proven that even from a short video or a few photos of a person that actually exists, fake videos can be created containing a photo of that person's face. Fake photos and videos can be created by taking these few photos or a short video and training them with deep learning techniques. In fake videos, face swapping of another person can be applied with the face of the person or face reenactment can be applied to the person's face. In re-enactment, facial movements of another person's video can be applied to the face of a source person. Even non-existent human photographs can be produced using techniques such as StyleGAN using a set of photographs of real human faces. These techniques, commonly known as deepfakes technology, are discussed in this study together with the types and structures used on the face. These techniques can be used for educational purposes for children by animating existing photographs of scientists and celebrities who lived in ancient times. This method can be used in puppetry. Instead of actors (Actors-Actresses), scenes can be animated with their photos on the computer. Portraits can be animated. The difference of this study from similar studies is that the data used for training is less than the data in other studies, and the types and structures of fake video creation are considered together. The material used for data training is the GPU and the dataset consists of VoxCeleb dataset,

¹ ICAIAME 2021 konferansında sunulmuş olup, tam metin olarak yayınlanmıştır.
Geliş: 01/11/2021, Düzeltme: 13/12/2021, Kabul: 19/12/2021

several short videos and several photos. The method used is generative networks such as Generative adversarial networks (GAN) and Auto-encoders. The study has shown that in the videos and photos used, when the face is turned forward or slightly to the right or slightly to the left, when the facial movement is limited in a certain area and the face moves slowly, it trains the artificial intelligence better. It has shown that the fake videos created using this training data are more successful.

Keywords: Face swapping, Face reenactment, Deepfake video, Deep learning.

I. GİRİŞ

Görüntü işleme bir şekilde elde edilmiş, ölçülmüş olan görüntü verilerinden yararlı bilgiler çıkarmak için kullanılan yöntemdir. Derin öğrenme ise bilgisayarın bir yapay sinir ağı ve birçok algoritma ile var olan verilerden yeni veriler elde etmesidir. Sahte fotoğraf ve video sentezi, derin öğrenme ve görüntü işleme alanlarının kesişiminde bulunan bir konudur. Yüz fotoğrafları içeren sahte fotoğraf ve video sentezi bir kişinin yüz özelliklerini kullanarak, o kişiye ait olmayan fotoğraf ve video oluşturmakla ilgilidir. Aynı zamanda hiç var olmayan bir kişi yüz özellikleri de oluşturularak fotoğraf ve video sentezlenebilir. Kaynak veriden alınan bu yüz özellikleriyle çeşitli işlemler yapılarak sahte fotoğraflar ve videolar oluşturulabilir.

Derin öğrenme, son yıllarda bilgisayar görüşü alanını güçlendirdiğinden, dijital görüntünün manipülasyonu, özellikle de insan portreleri görüntüsünün manipülasyonu, hızla gelişti ve çoğu durumda fotogerçekçi sonuç elde etti. Yüz değiştirme, kaynağın yüz hareketlerini ve ifade deformasyonlarını korurken hedefe bir kaynak yüzü aktararak sahte içerik oluşturmada göze çarpan bir görevdir. Yüz manipülasyon tekniklerinin arkasındaki temel motivasyon Çekişmeli Üretici Ağlar' dır [1]. GAN tabanlı yüz değiştirme yöntemleriyle sentezlenen çok sayıda sahte video Youtube ve diğer video web sitelerinde yayınlanmaktadır. Genel olarak sahte video oluşturma Deepfake (Derin sahtelik) olarak adlandırılrsa da birçok yöntem ve bu yöntemlerde kullanılan birçok yapı vardır.

Kullanılan yöntemler Çekişmeli Üretici Ağlar (Generative Adversarial Networks, GAN) ve Otomatik kodlayıcılar (Auto-encoders)' dır. Çekişmeli üretici ağlar iki ana birimden oluşur: Generator (Üretici) ve Discriminator (Ayırt Edici). Generator aldığı bir işaretten (gürültüden) sanal resimler oluşturmaktadır. Discriminator' a ise gerçek resimler ve sanal resimler bir arada verilip ayırt etmesi istenir. Discriminator, bu resimlere belirli değerler verir. Verdiği bu değerlere göre doğru ya da yanlış ayırt ettikleri bildirilir ve bu şekilde eğitilir. Generator ise oluşturduğu sanal resimlerle, tekrar bu yapıdan geçirilerek eğitilir. Bu iki yapı birbiri ile yarışarak, eğitilerek daha başarılı bir model ortaya çıkar. Otomatik kodlayıcı ise aldığı görüntü verilerinden gizli bir uzay vektörü oluşturur. Bu vektörü kullanarak daha sonra görüntünün temsili oluşturur. Otomatik kodlayıcı, Çekişmeli üretici ağdaki Generator' a benzetilebilir. Burada gerçek veri kullanılarak bir çıktı oluşturulduğu için Discriminator' a ihtiyaç yoktur. Çekişmeli Üretici Ağlar(ÇÜA)' da ise Generator alınan bir gürültüden (ses, video vb.) çıktı oluşturduğu için bunun doğruluğunu-yanlışlığını kontrol eden Discriminator' a ihtiyaç vardır.

Genel olarak üç çeşit sahte fotoğraf ve video sentezi vardır. Bunlar: Yüz Değiştirme (Face Swapping), Yüz Hareketlendirme, Canlandırma (Face Re-enactment), Yüz Oluşturma (Face Generation). Bu tekniklerde de kullanılan genel olarak iki çeşit yapı vardır. Bunlar: Otomatik Kodlayıcı (Auto- Encoder), Çekişmeli Üretici Ağ (GAN)' dır.

A. KONVOLÜSYONEL SİNİR AĞLARI

Konvolüsyonel Sinir Ağları [2](CNN) bir çeşit çok katmanlı algılayıcıdır (MLP). Görme merkezi hücreleri görselin tümünü içerecek biçimde alt alanlara bölünmüştür. Basit hücreler, kenara benzeyen özelliklerde odaklanırken, karmaşık hücreler geniş alıcılar yardımıyla, görselin bütününe odaklanır. Konvolüsyonel sinir ağları, bilgisayarlı görü alanında başarısı ispatlanmış binlerce değişik problem için tasarlanan yüzlerce modeline rastlanabilecek, derin öğrenmenin bir konusudur. Örneğin, robot ya da otonom araçların görü sistemlerinde; trafik işaretleri, nesne ve yüz tanıma vb. alanlarda faydalanılır. Bir

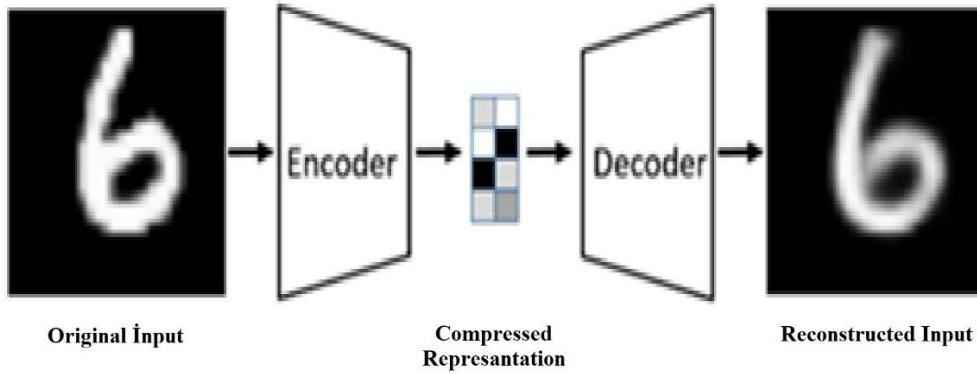
ileri beslemeli sinir ağı olan CNN, hayvanların görme merkezi ilham alınarak ortaya çıkmıştır. Burada yapılan matematiksel konvolüsyonel işlem, nöronun uyarı bölgesinden uyarılara verdiği cevaptır. CNN görüntüleri farklı katmanlarla işler. Evrişim katmanı (Convolutional layer): Özelliklerin saptanmasında yararlanır. Doğrusal olmayan katman (Non-linearity layer): Doğrusal olmayanlığın sisteme verilmesi amacıyla kullanılır. Ortaklama katmanı (Pooling layer): Ağırlıkların sayılarını düşürür ve uygunluk kontrolü yapar. Düzleştirme katmanı (Flattening layer): Temel sinir ağına veri hazırlar. Tam bağlı katman (Fully-connected layer): Sınıflama amacıyla yararlanan sinir ağıdır.

B. KULLANILAN YAPILAR

B. 1. Otomatik Kodlayıcı

Otomatik kodlayıcı [3], Encoder ve Decoder olmak üzere 2 yapıdan oluşur. Encoder yapısında veri alınır ve bu veriden gizli bir uzay vektörü oluşturulur. Decoder kısmında da bu gizli uzay vektörü kullanılarak giriş verisi yeniden yapılandırılarak bir çıkış verisi oluşturulur. Otomatik kodlayıcı, verilerin verimli bir şekilde nasıl sıkıştırılacağını ve kodlanacağını öğrenen, ardından veriyi indirgenmiş kodlanmış gösterimden orijinal girdiye mümkün olduğunca yakın bir temsile nasıl yeniden yapılandırılacağını öğrenen, denetimsiz bir yapay sinir ağıdır. Amacı giriş verisinden özellikleri çıkarıp bu özellikleri kullanarak yeniden inşa edilmiş çıkış verisi oluşturmaktır. Boyut azaltma, görüntü ve ses gürültü giderme, anormallik – aykırı değer algılama, resim boyama ve bilgi alma gibi konularda kullanılabilir. ÇÜA' ya göre oldukça küçük bir uzay vektörü oluşur ve eğitimi daha kolaydır. ÇÜA' lar gibi üretken modellerdir. Yapısında Konvolüsyonel Sinir Ağları (CNN) bulunabilir.

Şekil 1' deki gibi bir otomatik kodlayıcıda girdi görüntüsü sıkıştırılmış bir gösterime kodlanır ve sonra kodu çözülür.



Şekil 1. Bir otomatik kodlayıcı örneği [3].

Bir otomatik kodlayıcı, girdinin sıkıştırılmış bir temsilini geliştirmek için veri içindeki yapıyı keşfedebilen bir sinir ağı mimarisidir. Denetimsiz bir öğrenme tekniğidir. Genel otomatik kodlayıcı mimarisinin birçok farklı varyantı, sıkıştırılmış gösterimin orijinal veri girişinin anlamlı özelliklerini temsil etmesini sağlamak amacıyla mevcuttur; otomatik kodlayıcılarla çalışırken tipik olarak en büyük zorluk, modelinizin anlamlı ve genelleştirilebilir bir gizli alan temsilini gerçekten öğrenmesini sağlamaktır. Otomatik kodlayıcılar, eğitim sırasında verilerden keşfedilen özneliklere (yani, giriş özelliği vektörü arasındaki korelasyonlara) dayalı olarak verileri nasıl sıkıştıracağını öğrendikleri için, bu modeller tipik olarak yalnızca modelin eğitim sırasında gözlemlediği gözlem sınıfına benzer verileri yeniden yapılandırabilir.

Otomatik kodlayıcıda ağırlıkları güncellemek için kullanılan loss fonksiyonu Denklem 1'deki gibi encoder ve decoder kısmı için tanımlanan ve parametrelerine bağlıdır. Encoder ile temsil edilirken, decoder ile temsil edilir ve bunlar yalnızca sinir ağının ağırlıklarını ve bias değerlerini ifade eder.

$$L(\theta, \varphi) = \frac{1}{n} \sum_{i=1}^n (x^i - f_{\theta}(g_{\varphi}(x^i)))^2 \quad (1)$$

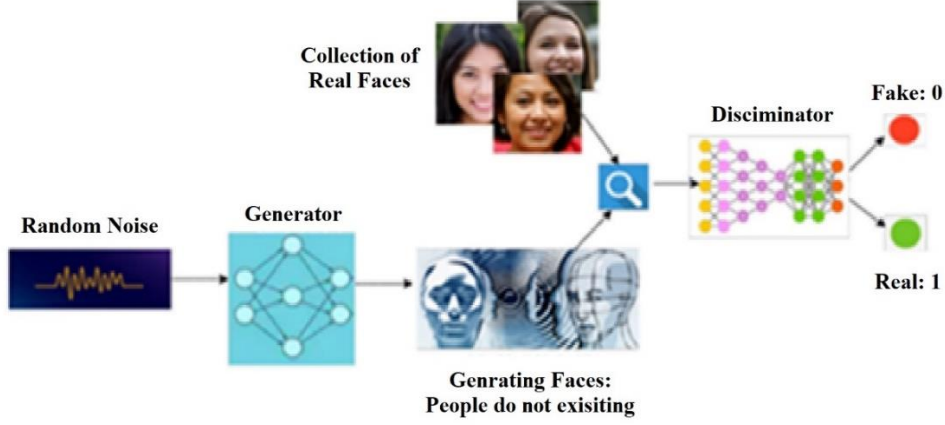
B. 2. Çekişmeli Üretici Ağ

ÇÜA (GAN) [1], Üretici (Generator) ve Ayırt edici (Discriminator) olmak üzere iki yapıdan oluşur. Üretici ağ bir gürültüden (random sayılar) sahte imajlar üreten ağıdır. Ayırt edici ağ ise gerçek ve sahte imajları alarak bunları ayırt eden ağıdır. Üretici ağ hiçbir şekilde gerçek verileri göremez. Üretici ağ sürekli yeni veriler üretmeyi öğrenirken, ayırt edici ağ ise girdi olarak kabul edilen veri seti ile üretilen verileri ayırt etmeyi öğrenir, bu süreçte her iki ağ da ne üretilip ne ayırt edeceğini kuralsız olarak kendi kendine keşfettiğinden dolayı bu bir gözetimsiz öğrenme tipidir. Her imaja 0 ile 1 arasında değer verilir. Bu değerlere göre Geri Yayılım (Backpropagation) ile her tekrar (epoch)' da ağlar, ağırlık (weight) ve kayıp (loss) değerlerini günceller ve birbiriyle yarışarak gelişir. Ağların eğitimi oto-kodlayıcılara göre daha karmaşıktır. Görüntü işlemenin birçok alanında, ses üretiminde vb. birçok konuda kullanılabilir. Yapısında Konvolüsyonel Sinir Ağları (CNN) bulunur. Hem Generator hem Discriminator birer sinir ağından oluşur. Sinir ağları aldığı veriyi öğrenebilen yapılardır.

Generator' ın gerçekçi resimler üretmesi, Discriminator' ın gerçek ve gerçek olmayan resimleri ayırt etmesi istenir. Bu ancak sinir ağları ile oluşturulabilir.

Normal yapay sinir ağları bir şeyleri tahmin etmede, bir şeyleri sınıflandırmada, problem çözmede oldukça iyidir. Ancak kendi başına yeni bir veri yaratmakta pek iyi değildir. Yapay zeka araştırmacıları uzun zamandır kendi kendine yeni ve anlamlı veriler üreten bir yapay zeka üretmeyi hayal etmiştir. Gans ise tam olarak bunu sağlıyor. Gans resimleri öğrenip neredeyse gerçekle bir farkı olmayan yeni resimler oluşturabilir. Gan' ler kullanılan nesnelere, doğadaki hayvanları, insan yüzlerini tanıyarak aslında gerçek olmayan ama ilk bakışta gerçek olmadığı fark edilemeyecek yeni resimler üretebiliyor.

Gans eğitilirken iki Network de sıfırdan eğitime başlanıyor. Yani önceden eğitilmiş gerçek ve sahte resim ayırt edebilecek bir Discriminator yoktur. Hem Generator hem Discriminator sıfırdan beraber eğitilecektir. Eğitim esnasında bu ikisi birbirleriyle bağlantılı olacaktır. Mesela gemi resimleri üretmek isteniyor olsun. Eğitim yapılırken Generator resim üretecektir. Daha sonra üretilen resim Discriminator' a verilir. Aynı zamanda Discriminator' a gerçek gemi resimleri de verilmesi gerekir. Discriminator üretilen resim gerçek resme ne kadar benziyor buna karar verecektir. Daha sonra Discriminator, Generator' a neden ürettiği resmin gemi olmadığını söyleyecektir. Bu şekilde döngü devam edecek ve zaman içerisinde Generator oldukça gerçekçi resimler üretebilecektir. Kısacası beraber öğrenip geliştireceklerdir. Gans mantıksal olarak kabaca bu şekildedir.



Şekil 2. Bir çekişmeli üretici ağ örneği [1].

Şekil 2’deki gibi üretici ağ sürekli yeni veriler üretmeyi öğrenirken, ayırt edici ağ ise girdi olarak kabul edilen veri seti ile üretilen verileri ayırt etmeyi öğrenir, bu süreçte her iki ağ da ne üretip ne ayırt edeceğini kuralsız olarak kendi kendine keşfettiğinden dolayı bu bir gözetimsiz öğrenme tipidir.

C. KULLANILAN ÇEŞİTLER

C. 1. Yüz Değiştirme

Bir bilgisayar görüşü konusu olan yüz değiştirme genel olarak imaj ya da videodaki yüzleri tanıma, hizalama, maskeleme, yüz segmentasyonu ve başka bir kişinin yüzü ile değiştirme gibi alt işlemlerin bütünüdür. Mevcut medyadaki kişinin yüzünün kaynak medyadaki kişinin yüzü üzerinde birleştirilmesi ve üst üste konması ile sahte medya üreten bir metottur. Yapay sinir ağlarını kullanır. Kullanılacak model- mimariye göre OK veya ÇÜA kullanılabilir.

C. 2. Yüz Hareketlendirme, Canlandırma

Bir bilgisayar görüşü konusu olan yüz canlandırma genel olarak kaynak yüzün şeklini hedef yüze aktarırken hedef yüzün görünümünün ve kimliğinin de korunmasını sağlar. Yani kısaca hareketin aktarılması da denebilir. Örneğin kaynak videodaki yüzde göz kırplıyorsa hedef yüzdeki imaja da göz kırpmaya uygulanır. Genellikle kaynak yüz için bir video, hedef yüz için bir veya birkaç imaj kullanılır. Genellikle ÇÜA kullanılır. Portre ve tabloların canlandırılmasında kullanılabilir. Yüz şeklini tahmin etmek için çeşitli yüz işaret algılayıcıları kullanılabilir.

C. 3. Yüz Oluşturma

Görüntü oluşturmanın bir alt alanı olan yüz oluşturma, var olan bir veri kümesinden eğitilerek yeni yüzler oluşturma ile ilgilidir. ÇÜA’ ları kullanır. Bu sayede gerçekte hiç var olmayan gerçekçi insan yüz resimleri üretilebilir. Yüz oluşturma en bilinen örneği Nvidia araştırmacılarının tanıttığı StyleGAN[4]’ dir. StyleGAN[4] Nvidia’ nın CUDA [5] yazılımını, GPU donanımlarını ve Tensorflow [6] kütüphanesini kullanır.

D. LİTERATÜR ARAŞTIRMASI

“FSGAN: Subject Agnostic Face Swapping and Reenactment” adlı çalışma Nirkin, Keller ve Hassner (2019) [7] tarafından yüz değiştirme ve yüz canlandırma için oluşturulmuş modeldir. ÇÜA yapısını ve tekrarlayan sinir ağını kullanır. Yüzler üzerinde eğitim gerektirmeden yüz çiftlerine uygulanabilir.

Model; Face reenactment and segmentation, Face inpainting, Face blending olmak üzere üç kısımdan oluşur. Face reenactment and segmentation kısmında kaynak ve hedef yüzleri yeniden hareketlendirerek yüz segmentlerini alır. Hedef görüntünün yüz ve saç segmentasyon maskesini tahmin eder. Face inpainting kısmında kaynak yüzün eksik kısımlarını boyar. Face blending kısmında segmentasyon maskesini kullanarak iki yüzü harmanlanma yapar.

“Few-Shot Adversarial Learning of Realistic Neural Talking Head Models” adlı çalışma Zakharov, Shysheya, Burkov ve Lempitsky (2019) [8] tarafından Samsung Yapay Zeka Laboratuvarlarında yapılmış bir çalışmadır. ÇÜA yapısını kullanır. Yüz üretimi konusunun bir alt dalı olan Talking Head Generation modelidir. Bir kişinin Talking Head modelinin; birkaç görüntüsünden, hatta tek bir görüntüsünden öğrenilmesi için geniş bir video veri kümesi üzerinde uzun süre meta- öğrenme gerçekleştirir. VoxCeleb (Nagrani, Chung ve Zisserman, 2017) veri setini kullanır. Few-shot learning (R. Zhang, Che, Ghahramani, Bengio ve Song, 2018) ve one-shot learning öğrenme metotlarını kullanır. Tabloların, resimlerin canlandırılmasında kullanılabilir. Modelde bir Embedder yardımıyla gerçek resimlerde renk ve yüz şekillerini, çok katmanlı algılayıcılarla Adain parametreleri olarak Generator’a verir. Generator da gürültü olarak alacağı yüz şekli input’ undan sahte resimler üretir. Sahte resim, yüz şekli ve gerçek resim Discriminator’ a verilerek loss değerleri hesaplanarak ağırlıklar güncellenir.

“One-shot Face Reenactment” adlı çalışma Y. Zhang vd. (2019) [9] tarafından yüz canlandırma modeli oluşturmak için yapılan çalışmadır. OK yapısını kullanır. One-shot learning öğrenme metodunu kullanır. “ReenactGAN: Learning to Reenact Faces via Boundary Transfer” adlı çalışma Wu, Y. Zhang, Li, Qian ve Loy (2018) [10] tarafından yüz canlandırma modeli oluşturmak için yapılan çalışmadır. OK ve ÇÜA yapılarını kullanır.

D. KULLANILAN VERİSETİ

Çalışmada veriseti olarak VoxCeleb [11] veriseti ve ayrıca çeşitli kısa video ve fotoğraflar kullanılmıştır. Donald Trump’ ın Amerikan Başkanı iken bir konuşmasından kısa bir video alınmıştır. Oyuncu Taner Ölmez’ in Covid-19 hakkındaki kısa bir kamu spotundan video alınmıştır. Ayrıca yine bu iki video gibi ünlülerin röportaj videolarından oluşan VoxCeleb veriseti kullanılmıştır. VoxCeleb [11], YouTube'a yüklenen röportaj videolarından alınan, ünlülerin kısa videolarından oluşan görsel-işitsel bir veri setidir. Ayrıca Vangogh ve Monalisa portreleri fotoğraf olarak alınmış Joe Biden vb. kişilerin konuşmalarından alınan videolar ile bu fotoğraflar canlandırılmıştır.

II. MATERYAL VE METHODLAR

Materyal olarak iki çeşit GPU kullanılmıştır. Bunlar: Tesla P40 ve Nvidia Geforce GTX 1050’ dir. Veri seti olarak çeşitli videolar kullanılmıştır. VoxCeleb[11] ve CelebA-HQ[12] veri setlerinin eğitilmiş modeller üzerindeki ağırlıkları kullanılmıştır. Yöntem olarak yüz değiştirme, yüz canlandırma ve yüz üretimi modelleri kullanılmıştır. Yüz değiştirme modeli olarak Deepfakes [13], yüz canlandırma modeli olarak First order motion model[14], yüz üretimi modeli olarak StyleGAN[4] kullanılmıştır. Deepfakes modelinde çeşitli videolar kullanılmıştır. First Order Motion Model [14]’ de Voxceleb [11] veri setinin eğitilmiş modeldeki ağırlıkları kullanılmıştır. StyleGAN [4] modelinde de CelebA-HQ [12] veri setinin eğitilmiş modeldeki ağırlıkları kullanılmıştır. Genel olarak üç uygulama gerçekleştirildi. İlk olarak Otomatik Kodlayıcı kullanılarak bir yüz değiştirme işlemi gerçekleştirildi. İkinci olarak ÇÜA kullanılarak bir yüz canlandırma işlemi gerçekleştirildi. Son olarak yine ÇÜA kullanılarak bir yüz üretimi işlemi gerçekleştirildi.

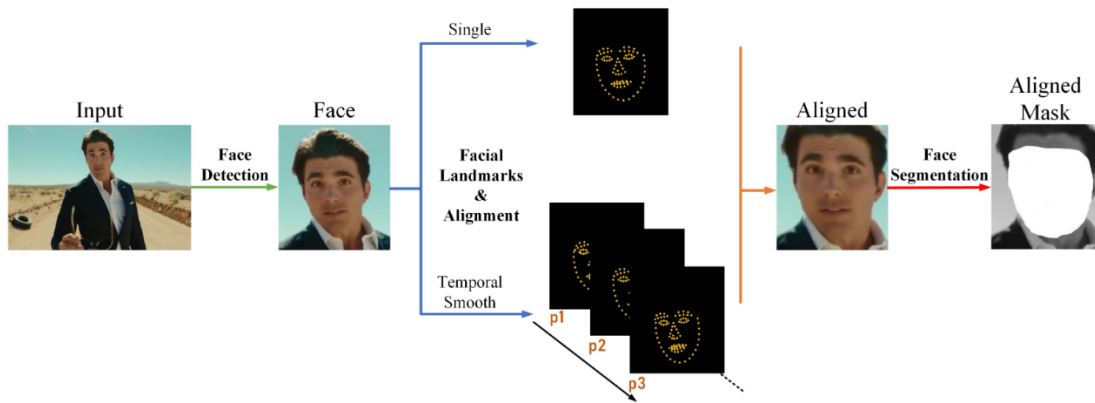
A. OTOMATİK KODLAYICI KULLANILAN YÜZ DEĞİŞTİRME

Bu metotta kullanılan DeepFaceLab [16], Ivan Perov ve arkadaşları tarafından oluşturulan açık kaynaklı bir deepfake sistemidir. Deepfakes [13] modeli genel hatlarıyla yüz hizalama modülü, yüz ayrıştırma

modülü, yüz harmanlama modülünden oluşur. Bu şekilde fotogerçekçi yüz değiştirme sonuçları elde etmeyi sağlar.

Deepfakes [13] modeli üç ana aşamadan oluşur: Yüzleri bularak çıkarma ve biriktirme aşaması, yüz çiftlerinin birbiriyle karşılıklı değiştirilecek şekilde eğitilmesi aşaması ve yüzleri eğitim sonuçlarını kullanarak dönüştürme aşaması. Yüzleri çıkarma aşaması da yüz algılama, yüz hizalama ve yüz bölümlenmeyi içerir. Bu modelde ilk olarak yüz çıkarma işlemi gerçekleştirilir. Bilindiği gibi video, görüntülerden (frame) oluşur. Bilgisayar oyunlarında da bilindiği gibi daha gerçekçi görüntü için saniyedeki görüntü sayısı (fps: frame per second) fazla olmalıdır. Modelin yüz çıkarma işleminde saniyede çıkarılması istenen sayıda görüntü çıkarılır ve görüntüdeki yüzler algılanır.

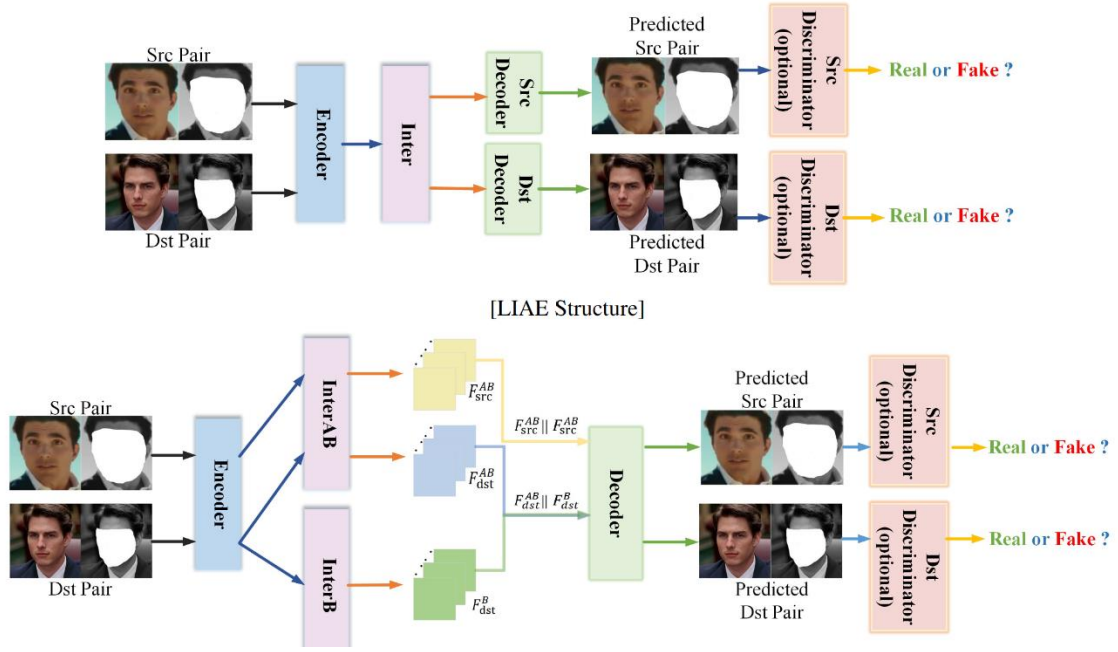
Yüz çıkarma işleminin ilk adımı yüz algılamanın amacı, verilen klasörlerde hedef yüzü bulmaktır. Belirtilen hedef için S3FD [17], RetinaFace [18], MTCNN [19] yüz algılama algoritmalarından herhangi birisi kullanılabilir. Yüz çıkarma işlemi Şekil 3 ile gösterilmektedir.



Şekil 3. Deepfakes modelinde yüz çıkarmaya genel bakış [16].

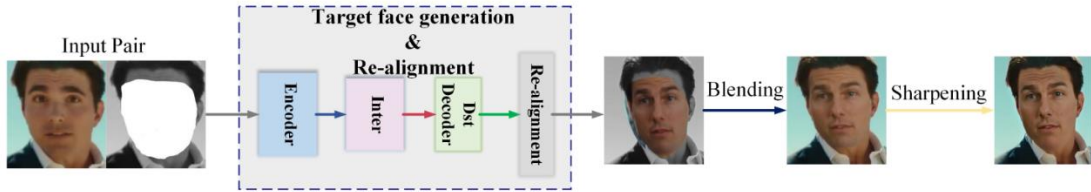
İkinci adım olarak yüz hizalama, yüz yer işaretleri algoritması ile yüzlerin hizalamasıdır. Bunu çözmek için çıkarma algoritması: (a) ısı haritası tabanlı yüz çevreleme algoritması 2DFAN [20] (normal duruşa sahip yüzler için) ve (b) PRNet [21] 3D yüz öncül bilgileri (geniş Euler açılı yüz için) kullanılabilir.

Yüz segmentasyonunda yüz hizalamadan sonra elde edilen, standart ön / yandan görünüm hizalı kesite sahip veri klasörü kullanılır. Hizalanmış klasörün üstünde bir Yüz Segmentasyon ağı (TernausNet [22]) kullanılır; bu ağ üzerinden saç, parmak veya gözlük içeren bir yüz tam olarak bölümlere ayrılabilir. Eğitim sürecinde ağı ellere, gözlüklere ve yüzleri bir şekilde kaplayabilecek diğer nesnelere karşı sağlam tutmak için düzensiz tıkanıklıkları gidermek için tasarlanmış, isteğe bağlı olarak ancak yararlıdır.



Şekil 4. Deepfakes modelinde eğitim aşamasına genel bakış [16].

Şekil 4’ te gösterildiği gibi eğitim aşamasında otomatik kodlayıcıdaki gizli uzay vektörleri birleştirilmektedir. İstenirse otomatik kodlayıcı ağ mimarisine ayırt edici ağ eklenerek Deepfakes [13] modeli için ÇÜA kullanılabilir.



Şekil 5. Deepfakes modelinde yüz dönüştürme aşamasına genel bakış [16]

Son olarak yüz dönüştürme aşamasında Şekil 5’ te görüldüğü gibi eğitim sonuçları kullanılarak çeşitli işlemlerle (harmanlama, bileme vb.) yüz diğer bir yüze dönüştürülmektedir. Burada eğitim sonuçları karşılıklı kullanılarak Umeyama’ nın [23] tersine çevrilebilirliği kullanılarak ters dönüştürme de gerçekleştirilebilir. Örneğin; A yüzü B yüzüne dönüştürülebileceği gibi, B yüzü de A yüzüne dönüştürülebilir.

Çeşitli kısa videolar alınarak Deepfakes modelinde GTX 1050 GPU ile Windows ortamında model eğitilerek yüzler birbiriyle değiştirilmiştir. Örneğin, Şekil 6’ daki çıktılar için 79 ve 195 saniyelik iki video kullanılmıştır.

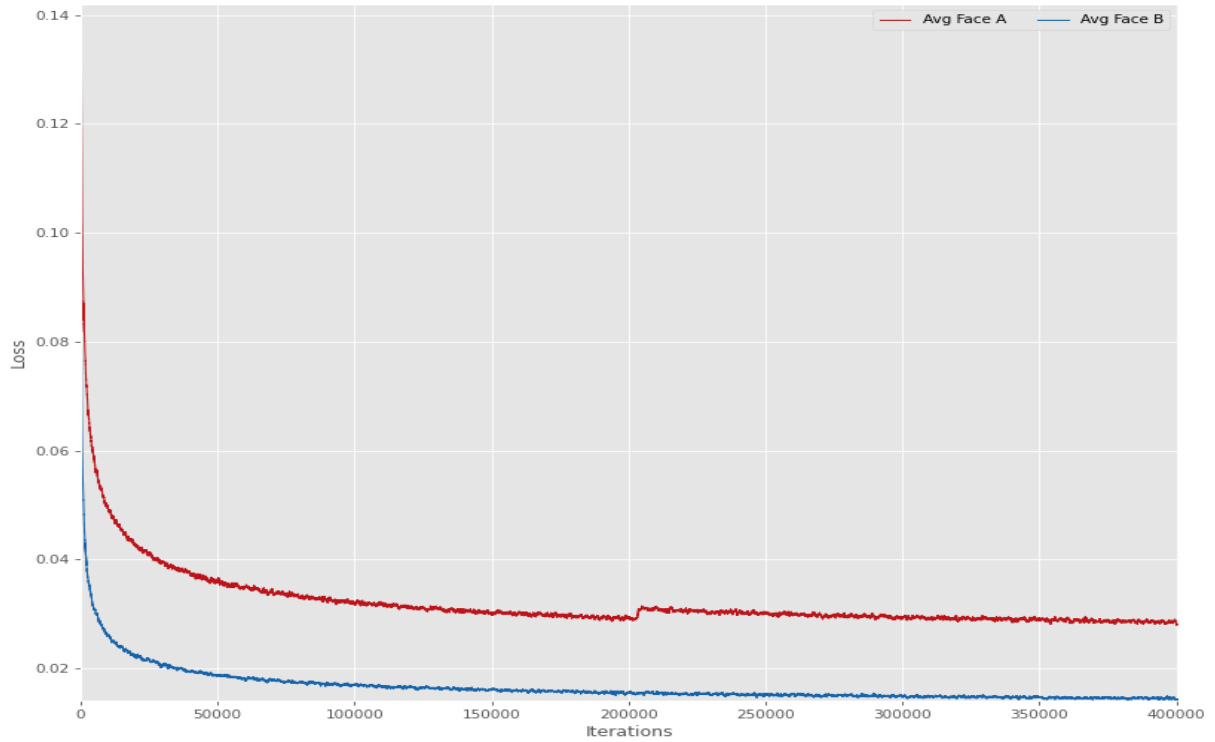
Videolardan frame’ ler çıkarılarak Yüz tanıma kısmında bu frame’ lerde yüzler S3FD [17] yüz tanıma kütüphanesi kullanılarak bulunmuştur. Bulunan yüzler 2DFAN [20] yüz çıkarma kütüphanesi kullanılarak hizalanmış ve yüz segmentasyonu çıkarılmıştır. Deepfakes modeli kullanılarak elde edilen verilerle iki yüz birbiriyle değişecek şekilde 12 batch (yığın) olarak yaklaşık 54 saat 400.000 iterasyon eğitilmiştir. Eğitilen modellerle videolar yüzler değiştirilerek yeniden oluşturulmuştur. Oluşturulan videodaki örnek frame’ ler Şekil 6’ da sağda gösterilmiştir.

Bu görüntülerde yüzün hangi kısımlardan itibaren değiştirildiğini daha net görebilmek için yüzü ve kafası birbirine benzemeyen iki kişi tercih edilmiştir. Örneğin birinin saçları siyah iken diğerinin sarı, ten renkleri farklı ve yüz özellikleri farklıdır. Normalde birini kandırmak amaçla veya kötü bir niyetle yapılan yüz değiştirmede yüz ve kafa özellikleri benzeyen kişiler tercih edilebilir.

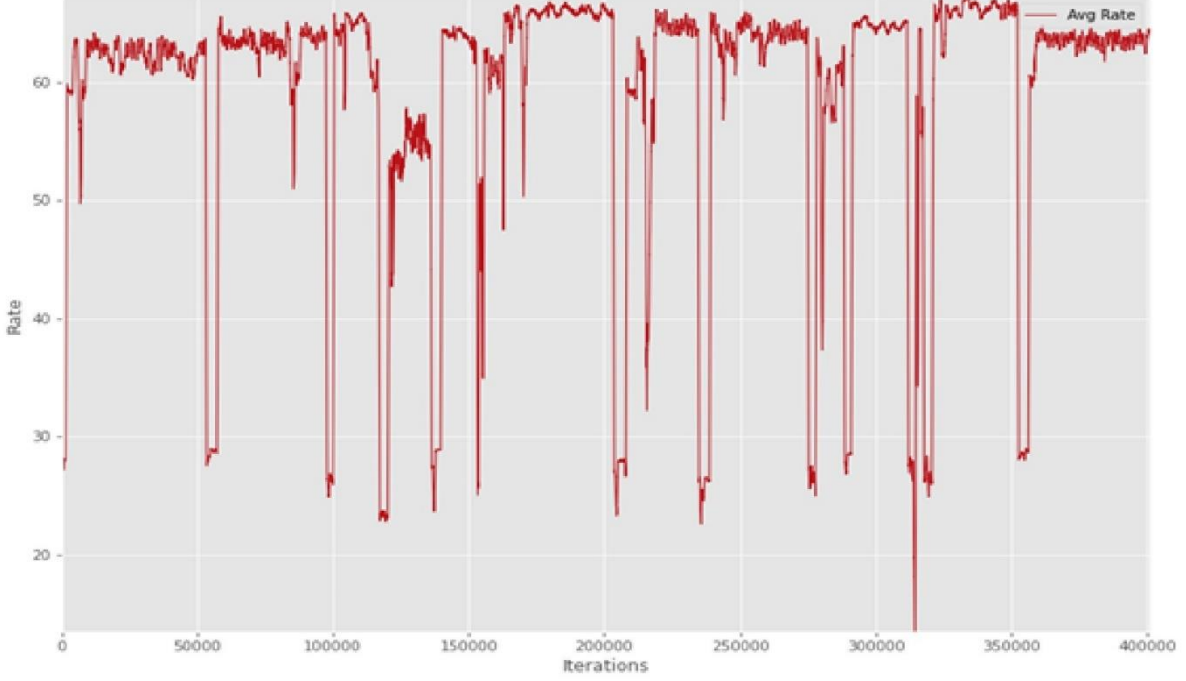


Şekil 6. Deepfakes [13] modeli kullanarak yüzlerin eğitilerek birbirine değiştirilmesi.

Learning rate ve loss değerleri grafikleri oluşturulmuştur. Burada 200.000 iterasyondan sonra bazı görüntülerin yapay zekayı iyi eğitmediği gözlemlendiğinden bu veriler çıkarılmıştır ve Şekil 7’deki gibi loss değerinde anlık bir artış olmuştur ve sonra loss değeri azalmaya devam etmiştir. Learning rate grafiği Şekil 8’de gösterilmektedir.



Şekil 7. Loss grafiği.



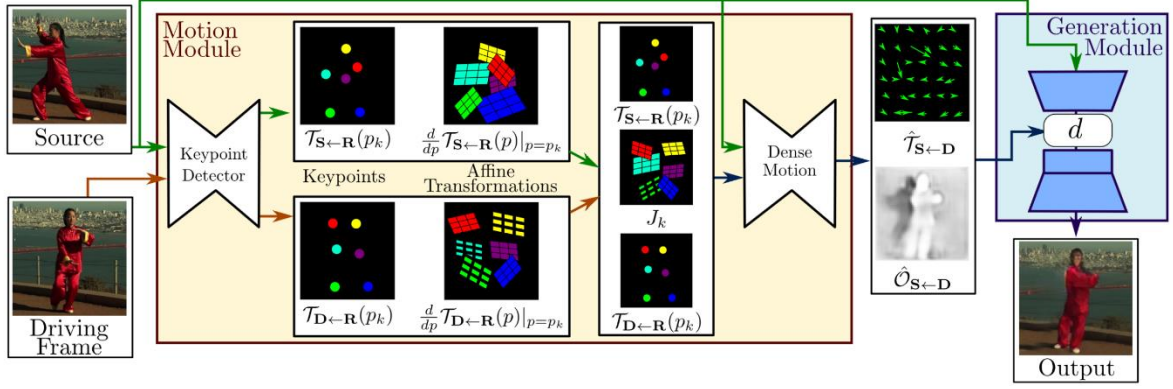
Şekil 8. Learning rate grafiği.

B. ÇEKİŞMELİ ÜRETİCİ AĞ KULLANARAK YÜZ CANLANDIRMA

Bu metotta kullanılan First Order Motion Model [14] ile görüntü canlandırma, bir kaynak görüntüdeki bir nesnenin sürüş videosunun hareketine göre canlandırılması için bir video sekansı oluşturmayı içerir. Bu modelde, canlandırılacak belirli nesne ön bilgi kullanmadan bu sorunu ele alır. Aynı kategorideki nesnelere (ör. Yüzler, insan vücutları) gösteren bir dizi video üzerinde eğitildikten sonra, yöntem bu sınıftaki herhangi bir nesneye uygulanabilir. Bunu başarmak için, kendi kendini denetleyen bir formülasyon kullanarak görünüm ve hareket bilgileri ayrılmaktadır. Karmaşık hareketleri desteklemek için, yerel afin dönüşümleriyle birlikte bir dizi öğrenilmiş anahtar noktadan oluşan bir temsil kullanılmaktadır. Bir üretici ağı, hedef hareketleri sırasında ortaya çıkan tıkanıklıkları modeller ve kaynak görüntüden çıkarılan görünümü ve sürüş videosundan türetilen hareketi birleştirir.

Şekil 9' da görüldüğü gibi çerçeve iki ana modülden oluşmaktadır: hareket tahmin modülü ve görüntü oluşturma modülü. Hareket tahmin modülünün amacı, yoğun bir hareket alanını tahmin etmektir. Soyut bir referans çerçevesi olduğunu varsayılmaktadır ve bağımsız olarak iki dönüşüm tahmin edilmektedir: referanstan kaynağa ve referanstan sürüşe. Bu seçim, kaynağı ve sürüş çerçevelerini bağımsız olarak işlemeye olanak tanır. Bu, test zamanında model, görsel olarak çok farklı olabilen farklı bir videodan örneklenmiş kaynak görüntü ve sürüş çerçevelerinin çiftlerini aldığından istenir.

İlk adımda, kendi kendini denetleyen bir şekilde öğrenilen anahtar noktaları kullanarak elde edilen seyrek yörünge setlerinden her iki dönüşümü de tahmin edilmektedir. Yerel afin dönüşümleri kullanarak her bir kilit noktanın komşuluğundaki hareketi modellenir. Yalnızca ana nokta yer değiştirmelerini kullanmakla karşılaştırıldığında, yerel afin dönüşümler, daha büyük bir dönüşüm ailesini modellemeye izin verir. İkinci adım sırasında, yoğun hareket ağı, ortaya çıkan yoğun hareket alanını elde etmek için yerel yaklaşımları birleştirir. Ayrıca, yoğun hareket alanına ek olarak, bu ağ, kaynak görüntünün eğilmesi ile hangi görüntü parçalarının yeniden yapılandırılabilirliğini ve hangi parçaların boyanması gerektiğini (bağlamdan çıkarılır) gösteren bir kapatma maskesi çıkarır. Son olarak, oluşturma modülü, sürücü videoda sağlandığı gibi hareket eden kaynak nesnenin bir görüntüsünü işler. Burada, kaynak görüntüyü yoğun harekete göre bükme ve kaynak görüntüde tıkanan görüntü parçalarını boyayan bir generator ağı kullanılır.



Şekil 9. First order motion model ' in çekişmeli üretici ağ yapısı [14].

Geometride Afin dönüşüm [24], afin uzayların aralarındaki noktalar, düz çizgiler ve düzlemler için oranların korumasını sağlayan eşlemedir. Ayrıca, paralel olan çizgi kümeleri afin dönüşümün sonrasında paralel olarak kalır. Afin dönüşümde aynı doğru üzerindeki noktaların aralarındaki mesafelerin oranları sabit kalmasına rağmen, çizgilerin aralarındaki açı ile noktaların arasında bulunan mesafeler sabit kalmayabilir.

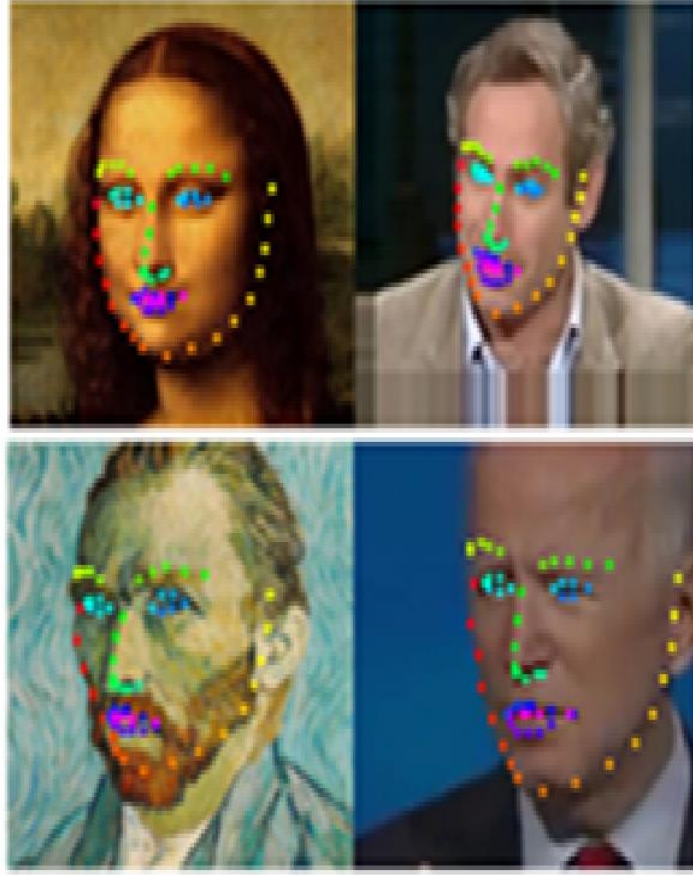
Bu modelde kilit noktalara ve yerel afin dönüşümlere dayalı yeni bir görüntü animasyonu yaklaşımı kullanılır. Yeni matematiksel formülasyon, iki çerçeve arasındaki hareket alanını tanımlar ve birinci dereceden bir Taylor genişleme yaklaşımı türetilerek verimli bir şekilde hesaplanır. Bu şekilde hareket, bir dizi anahtar nokta yer değiştirmeleri ve yerel afin dönüşümler olarak tanımlanır.

VoxCeleb [11] veri setinin First Order Motion Model [14] üzerinde eğitilmiş ağırlıkları kullanılarak çeşitli tablolar, portreler, resimler (elimizde bir pozunu bulunan resimler); alınan başka videolardaki yüz hareketleri kullanılarak Google Colab ortamında Tesla P4 GPU ile ve Windows ortamında GTX 1050 GPU ile canlandırılmıştır. Bu işlemler alınan görüntüler ve webcam üzerinde denenmiştir. Bu modelde kullanılan kaynak görüntüler yüz eşleşmelerinde solda ve driving video (hareketleri kaynak görüntüye aktarılacak videolar) sağda gösterilmiştir. Şekil 10' da gösterildiği gibi sol taraflarda bulunan Mona Lisa ve Van Gogh tabloları, hemen sağlarında bulunan kişilerin yüz hareketleri transfer edilerek canlandırılmıştır. Burada yüz çiftlerinin en uygun yüz hizası eşleşmeleri gösterilmektedir.

Buradaki eşleşen noktalarının hareketine göre tablolar üretken modellerle yeniden oluşturularak hareketlendirilmektedir. Aynı zamanda burada wav2lip modeli [15] ile arka plandaki konuşmaya göre (söylenen sözcüklere göre) dudak hareketleri de verilmiştir. Şekil 10' daki görseldeki çıktılar Tesla P4 GPU kullanılarak elde edilmiştir.

**Source
image**

**Driving
video**



Şekil 10. Tabloların canlandırılması.

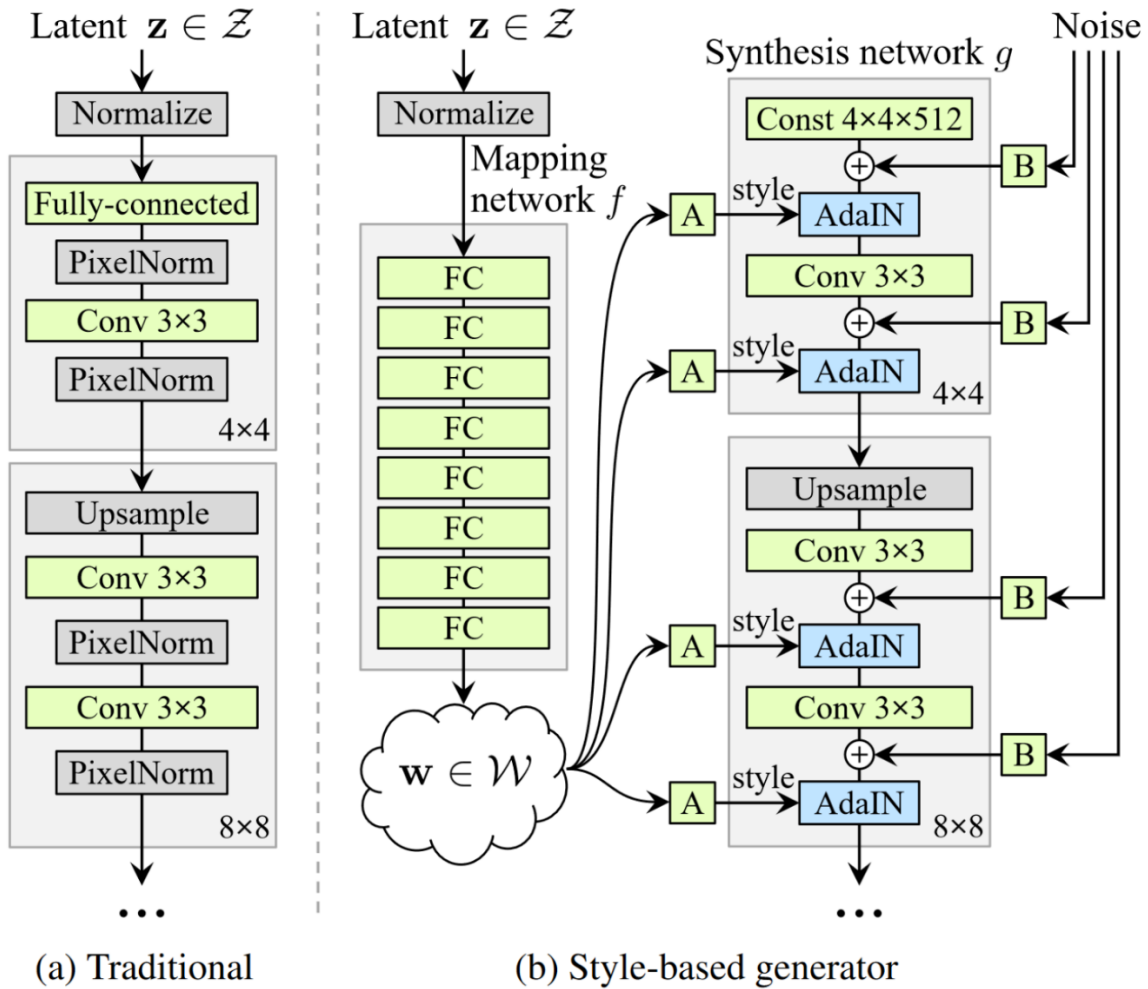
C. ÇEKİŞMELİ ÜRETİCİ AĞ KULLANARAK YÜZ OLUŞTURMA

Bu metotta kullanılan StyleGAN [4], bir tür çekişmeli üretici ağıdır. Stil transferi [25] literatüründen alıntı yaparak, özellikle, uyarlanabilir örnek normalizasyonunun kullanımı ile ÇÜA için Şekil 11' deki gibi alternatif bir üreteç mimarisi kullanır.

Stil tabanlı generator, geleneksel generator' den farklı olarak öğrenilmiş bir sabitten işleme başlar. Her bir evrişim katmanında uyarlanabilir örnek normalizasyonu (AdaIN) aracılığıyla generator kontrol edilir. Doğrusal olmama durumu değerlendirilmeden önce, her evrişimden sonra Gauss gürültüsü eklenir. Burada "A", öğrenilmiş afin dönüşümü ifade eder ve "B", gürültü girişine kanal başına öğrenilen ölçeklendirme faktörlerini uygular. Afin dönüşümleri öğrenilen w , sentez ağının her bir evrişim katmanından sonra uyarlanabilir örnek normalizasyonu (Adaptive Instance Normalization) (AdaIN) [26,27,28,29] işlemlerini kontrol eden $y = (y_s, y_b)$ stillerine özelleşir AdaIN işlemi Denklem 2' deki gibi tanımlanır.

$$\text{AdaIN}(x_i, y) = y_{s,i} \frac{x_i - \mu(x_i)}{\sigma(x_i)} + y_{b,i} \quad (2)$$

Burada her özellik haritası x_i ayrı normalleştirilir ve ardından y stilindeki karşılık gelen skaler bileşenler kullanılarak ölçeklenir.



Şekil 11. Geleneksel üretici ağ (a) ve stil tabanlı üretici ağ (b)' in karşılaştırılması [4].

StyleGan [4] üzerinde CelebA-HQ [12] veri seti ağırlıkları kullanılarak önceden eğitilmiş modelinde Google Colab ortamında Tesla P4 GPU ile birkaç yüz görüntüsü oluşturma işlemi denenmiştir. Bu model ile üretilen birkaç görüntü aşağıda Şekil 12' de gösterilmiştir.



Şekil 12. Gerçekte var olmayan insan yüz resimleri üretimi.

III. BULGULAR, TARTIŞMA, SONUÇ VE ÖNERİLER

Yapılan çalışma; kullanılan video ve fotoğraflarda yüzün karşıya (öne) dönük veya hafif sağa ya da hafif sola dönük iken, yüz hareketinin belirli bir alanda sınırlı olduğunda ve yüzün yavaş hareket ettiğinde yapay zekayı daha iyi eğittiği ve bu eğitim verileri kullanılarak oluşturulan sahte videoların daha başarılı olduğunu göstermiştir. Aynı zamanda çalışma, amaca uygun olarak çeşitli yöntem ve modellerle bir resim, birkaç resim ya da kısa bir videodan o kişiye ait sahte videolar oluşturulabileceğini göstermektedir. Deepfakes[13] modeli için videolardan çıkarılan bazı görüntülerin yapay zekayı iyi eğitemediği, bulanık bir yüzün çıkmasına neden olduğu görülmüştür. Bu görüntüler çıkarılarak aynı süre ile eğitim devam ettirildiğinde daha başarılı çıktılar elde edilmiştir. Kalan görüntülerin yapay zekayı daha iyi eğittiği görülmüştür.

Tablo 1. Deepfakes modeli için hesaplanan loss değerleri [13].

Deepfakes modeli	100.000 iterasyon loss değeri	200.000 iterasyon loss değeri	300.000 iterasyon loss değeri	400.000 iterasyon loss değeri
Avg Face A	0.0312953	0.0301418	0.0298924	0.0288525
Avg Face B	0.0164015	0.0160740	0.0153192	0.0146704

Tablo 1’ de Deepfakes [13] modelinin A kişisi ve B kişisi için hesaplanan ortalama loss değerleri gösterilmiştir. Bu tabloda görülüyor ki model zamanla daha iyi öğrenmeye başlamıştır.

Tablo 2. First Order Motion Modeli için hesaplanan loss değeri [14].

First Order Motion Modeli	VoxCeleb veri seti için
Hesaplanan loss değeri	0,0431123

Tablo 2’ de First Order Motion Modeli [14] için hesaplanan loss değeri gösterilmiştir.

Diğer çalışmalarda çok büyük veriseti ve daha yüksek hesaplama kapasiteli GPU’ lar kullanılmasına karşın biz burada mümkün olduğu kadar az resim, video ile ve daha düşük hesaplama kapasiteli GPU’ lar ile benzer çalışmalara yakın sonuçlar aldık. Bu çalışmalar laboratuvarlarda yapılmasına karşın bu çalışmada kişisel ve kısıtlı imkanlar ile yapılabilecek iyi çalışmalardan birini yapmaya çalıştık. Sonuçlar gösteriyor ki eğitim aşaması arasında kötü eğiten verilere müdahalemiz bizi daha başarılı sonuçlara götürmüştür.

Sahte videoların kötü niyetin haricinde çok çeşitli amaçlarla kullanabileceği görülmüştür. Bu durumda sahte video oluşturma teknolojisinin daha başlangıç aşamasında bu kadar geliştiği bir dönemde güvenilmeyen insanların görebileceği, bulabileceği bir şekilde video ve resim paylaşmaktan kaçınılmalıdır.

Deepfake videoları henüz ipuçlarıyla kendi başına keşfedilebilecek bir aşamadadır. Deepfake videolarda aşağıdaki özellikler ile videonun sahteliği anlaşılabilir: Titreşim hareketleri, kareler arası akışta ışıklal farklılıklar, cilt rengi değişimleri, göz hareketlerinde gariplik veya gözün kırılmaması, dudakların konuşma ile uyumsuz olması, görüntüdeki yapay dijital unsurlar.

Fakat Deepfake teknolojisi geliştikçe videolardaki sahteliğin insan gözü ile ayırt edilmesinde zorluklar olacağından gelişmiş siber güvenlik programlarından yardım alınmaya ihtiyaç duyulacaktır. Ayrıca Deepfake videolar Deepfake tespit algoritmaları kullanılarak da belirlenebilir. Videolarının gerçekliğini doğrulayabilmek için şifreleme algoritmaları, videoya belirli aralıklarla hash değeri eklemekte

faaydalanılabılır. Video deęiřtirilince hash deęeri de farklılařacaktır. Yapay zeka ile blokzincir, videoya deęiřtirilmesi imkansız sayısal bir iz ekleyebilir.

Deepfake giriřimlerini engellemek iin özüm olarak, yüz tespit yazılımlarının yararlandıęı piksel düzenlerinin gizlenmesi amacıyla videoya özel dijital "yapay unsurlar" ekleyebilecek bir program kullanılabilir. Bu unsurlar, algoritmaların hızını azaltabilir ve kalitesi düşük sonuçlar ortaya ıkabilir. Bu sayede Deepfake giriřimlerinin başarılı olma olasılıęı düşürölür. Bařka bir özüm olarak da Deepfake tespit algoritmaları kullanılabilir. Örneęin yüzün deęiřtirildięi bazı modellerde sobel filtresi yardımıyla maske bölgesel konvolüsyonel sinir aęı (Mask R-CNN) ile fotomontaj tespit algoritması kullanılabilir [30]. Ayrıca kaliteli haber kaynaklarını kullanmalı ve emin olmadan hiçbir ierięe güvenmemeliyiz.

Daha sonraki alıřmalarda yüz tanımanın alt alanları olan bu teknolojiler (yüz deęiřtirme, yüz hareketlendirme) kullanılarak iki özellięi de uygulayabilen, otomatik kodlayıcı veya ekiiřmeli üretici aę tabanlı bir üretici aęın geliřtirilmesi tasarlanmaktadır.

IV. KAYNAKLAR

- [1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville ve Y. Bengio, "Generative Adversarial Nets," *NIPS*, vol. 27, pp. 2672-2680, 2014.
- [2] Y. LeCun, Y. Bengio ve G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436-444, 2015.
- [3] D. Bank, N. Koenigstein ve R. Giryes, "Autoencoders," *CoRR*, vol. abs/2003.05991, 2020.
- [4] T. Karras, S. Laine ve T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," *CVPR*, pp. 4401-4410, 2019.
- [5] NVIDIA, Vingelmann, P., & Fitzek, F. H. P.. CUDA, release: 10.2.89. Retrieved from <https://developer.nvidia.com/cuda-toolkit>, 2020.
- [6] M. Abadi, P. Barham, J. Chen., Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard ve dięerleri, "TensorFlow: A System for Large-Scale Machine Learning," *OSDI*, pp. 265-283, 2016.
- [7] Y. Nirkin, Y. Keller ve T. Hassner, "FSGAN: Subject Agnostic Face Swapping and Reenactment", *ICCV*, pp. 7183-7192, 2019.
- [8] E. Zakharov, A. Shysheya, E. Burkov ve V. S. Lempitsky, "Few-Shot Adversarial Learning of Realistic Neural Talking Head Models," *ICCV*, 2019, pp. 9458-9467.
- [9] Y. Zhang, S. Zhang, Y. He, C. Li, C. C. Loy ve Z. Liu, "One-shot Face Reenactment", *CoRR*, vol. abs/1908.03251, 2019.
- [10] W. Wu, Y. Zhang, C. Li, C. Qian ve C. C. Loy, "ReenactGAN: Learning to Reenact Faces via Boundary Transfer," *ECCV*, pp. 622-638, 2018.
- [11] A. Nagrani, J. S. Chung ve A. Zisserman, "VoxCeleb: A Large-Scale Speaker Identification Dataset", *INTERSPEECH*, pp. 2616-2620, 2017.
- [12] T. Karras, T. Aila, S. Laine ve J. Lehtinen, "Progressive Growing of GANs for Improved Quality, Stability, and Variation," 2017.

- [13] Deepfakes (n.d.). Deepfakes/faceswap. Retrieved March 10, 2021, from <https://github.com/deepfakes/faceswap>, 2021.
- [14] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci ve N. Sebe, “First Order Motion Model for Image Animation”. CoRR, vol. abs/2003.00196, 2020.
- [15] K. R. Prajwal, R. Mukhopadhyay, V. P. Namboodiri ve C. V. Jawahar, “A Lip Sync Expert Is All You Need for Speech to Lip Generation In the Wild”, *ACM Multimedia*, pp. 484-492, 2020.
- [16] I. Perov, D. Gao, N. Chervoniy, K. Liu, S. Marangonda, C. Umé, M. Dpfks., C. S. Facenheim, L. RP, J. Jiang, S. Zhang, P. Wu, B. Zhou ve W. Zhang, “DeepFaceLab: A simple, flexible and extensible face swapping framework,” CoRR, vol. abs/2005.05535, 2020.
- [17] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang ve S. Z. Li, “S3FD: Single Shot Scale-invariant Face Detector,” CoRR, vol. abs/1708.05237, 2017.
- [18] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia ve S. Zafeiriou, “Retinaface:Single-stage dense face localisation in the wild”, arXiv preprint, arXiv:1905.00641, 2019.
- [19] K. Zhang, Z. Zhang, Z. Li ve Y. Qiao, “Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks,” *IEEE Signal Process*, vol. 23, pp. 1499-1503, 2016.
- [20] A. Bulat ve G. Tzimiropoulos, “How Far are We from Solving the 2D & 3D Face Alignment Problem? (and a Dataset of 230, 000 3D Facial Landmarks),” *ICCV* , pp. 1021-1030, 2017.
- [21] Y. Feng, F. Wu, X. Shao, Y. Wang ve X. Zhou, “Joint 3D Face Reconstruction and Dense Alignment with Position Map Regression Network,” arXiv preprint, arXiv:1803.07835, 2018.
- [22] V. Iglovikov ve A. Shvets, “TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation,” CoRR, vol. abs/1801.05746, 2018.
- [23] S. Umeyama, “Least-Squares Estimation of Transformation Parameters Between Two Point Patterns”, *IEEE Trans. Pattern Anal. Mach*, vol. 13, pp. 376-380, 1991.
- [24] M. Berger, *Geometry I*, Berlin, Springer, 1987, ISBN 3-540-11658-3.
- [25] L. A. Gatys, A. S. Ecker ve M. Bethge, “A Neural Algorithm of Artistic Style”, arXiv preprint, arXiv: 1508.06576, 2015.
- [26] X. Huang ve S. J. Belongie, “Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization”, CoRR, vol. abs/1703.06868, 2017.
- [27] V. Dumoulin, J. Shlens ve M. Kudlur, “A Learned Representation For Artistic Style”, CoRR, vol. abs/1610.07629, 2016.
- [28] G. Ghiasi, H. Lee, M. Kudlur, V. Dumoulin ve J. Shlens, “Exploring the structure of a real-time, arbitrary neural artistic stylization network,” CoRR, vol. abs/1705.06830, 2017.
- [29] V. Dumoulin, E. Perez, N. Schucher, F. Strub, H. Vries, A. Courville ve Y. Bengio, “Feature-wise transformations. Distill”, Retrieved March 10, 2021, from <https://distill.pub/2018/feature-wise-transformations>. 2, 2018.
- [30] N. E. Özmen ve E. Buluş, “Derin Sinir Ağları Yardımıyla Fotomontaj Tespiti,” *Mühendislik Bilimleri ve Tasarım Dergisi*, c. 8 s. 5, ss. 236-240, 2020.