

MÜHENDISLIK FAKÜLTESI  
YAPAY ZEKA MÜHENDİSLİĞİ   
Üretken Yapay Zeka Dersi Final Projesi

**Sesten Görüntü Üreten Üretken Yapay Zeka**

Ad Soyad: Yaren YALINBAŞ  
Öğrenci Numarası: 220212007  
20.05.2025

1. **Proje Tanımı ve Amaç**

Bu proje, ses verilerinden yapay yüz görselleri üretmeyi amaçlayan, üretken yapay zeka temelli bir çalışmadır. Özellikle bir kişinin sesinden, cinsiyetine ve ses tonundaki duygusal ifadeye uygun yüz görüntüsü oluşturulması hedeflenmiştir.

Proje iki ana aşamadan oluşur:

1. **Pretraining (ön eğitim):** FFHQ (Flickr-Faces-HQ) veri seti kullanılarak GAN mimarisiyle yüksek kaliteli yüz üretimi gerçekleştirilmiştir.
2. **Fine-tuning (ince ayar):** Kullanıcı sesiyle eşleştirilmiş görseller üzerinden, ses verisine uygun koşullandırılmış yüz üretimi sağlanmıştır.

Bu sayede, yalnızca ses verisinden özgün ve anlamlı yüzler üretilebilen bir sistem oluşturulmuştur.

1. **Kullanılan Yöntemler ve Teknolojiler**

### Kütüphaneler:

* **PyTorch:** Derin öğrenme mimarisinin kurulumu ve eğitimi için
* **Torchvision:** Görüntü işleme dönüşümleri
* **NumPy / Pandas:** Vektörel verilerin yönetimi
* **PIL (Pillow):** Görsel verilerin okunması ve dönüştürülmesi
* **Matplotlib:** Üretilen yüzlerin görselleştirilmesi

### Ortam:

* **Kaggle (GPU destekli çalışma ortamı) – x2 T4**

### Yöntem:

* **Wasserstein GAN with Gradient Penalty (WGAN-GP):** Stabil ve kaliteli üretim için tercih edildi.
* **DataParallel:** Çoklu GPU ortamında eğitim süreci hızlandırıldı.
* **DataLoader & Dataset:** Görseller ve ses vektörleri özel veri sınıfıyla eşleştirilerek işlendi.

1. **Model Mimarisi (Generator, Discriminator, Dataset)**

### Generator (Üreteç):

128 boyutlu latent vektörü alarak adım adım büyüyen bir yapıda 128x128 çözünürlüğe sahip yüz görüntüsü üretmektedir.

Yapısı şu katmanları içerir:

* Fully connected + LeakyReLU aktivasyon
* Ardışık ConvTranspose2D blokları (up-sampling)
* BatchNorm ve LeakyReLU ile desteklenmiş 5 aşamalı çözünürlük artırımı

### Discriminator (Ayrıştırıcı):

Giriş olarak aldığı yüz görselinin gerçek mi yoksa modelden mi üretildiğini öğrenmeye çalışır. Yapısı aşağıdaki gibidir:

* Çok katmanlı Conv2D blokları
* LeakyReLU aktivasyon
* En sonda sigmoid yerine WGAN-GP uyumlu düz bir skor çıktısı

### Dataset:

FFHQDataset sınıfı, thumbnails128x128 klasöründeki görselleri okuyarak normalize edip modele uygun hale getirir.

Veri ön işleme adımları:

* Görsel yeniden boyutlandırma (128x128)
* Tensor dönüşümü
* Normalize etme ([0.5, 0.5, 0.5] ortalama, [0.5, 0.5, 0.5] standart sapma)

1. **Eğitim Süreci (Veri seti, parametreler, loss fonksiyonları)**

### Pretraining:

* **Veri Seti:** FFHQ thumbnails (128x128 çözünürlükte)
* **Latent Boyut:** 128
* **Batch Size:** 16
* **Epoch:** 100
* **Optimizer:** Adam (lr = 1e-5, betas = (0.0, 0.9))
* **Cihaz:** CUDA (GPU) desteğiyle çalıştırıldı

Bu aşamada, sıfırdan yüz görselleri üretmeyi öğrenen bir Generator eğitildi.

### Fine-Tuning:

* **Veri:** Her bir ses vektörüyle eşleştirilmiş yüz görselleri
* **Amaç:** Ses vektöründeki cinsiyet bilgisini ve duygu tonunu yüz hatlarına yansıtmak
* **Yöntem:** Eğitilmiş generator modeli yüklendi, ses verilerinden gelen latent kodlar ile yeniden eğitildi
* **Early Stopping:** Generator kaybı (G loss) sabitlenince eğitim erken sonlandırıldı

1. **Sonuçlar ve Değerlendirme**

 Generator, FFHQ ile eğitildiğinde yüzler üretti.

 Fine-tune sonrası, sesin cinsiyetine uygun yüzler üretildiği gözlemlendi.

 Özellikle “neşeli” ses vektörlerinin daha yuvarlak hatlı, “ciddi” seslerin ise daha keskin hatlı yüzlere dönüştüğü görüldü.

 Model test çıktıları matplotlib ile görsel olarak değerlendirildi.

**6. Yenilikçi Yönler ve Katma Değer**

 **Ses-Görüntü Dönüşümü:** Tek başına ses bilgisinden görsel üretim yapılabilmiş olması, projeyi farklılaştıran bir özellik.

 **Koşullandırılmış Üretim:** Model yalnızca rastgele yüzler üretmiyor, sesin içerdiği bilgiler doğrultusunda yüzün yapısını şekillendiriyor.

 **Yaratıcı Uygulama Alanları:** Bu çalışma, animasyon, dijital insan tasarımı ve AI destekli avatar üretimi gibi pek çok sektöre katkı sağlayabilecek nitelikte. Ayrıca proje çıktıları korku filmi karakter üretiminde de kullanılabilir.

 **Gerçek Eğitim Süreci:** Model yalnızca teorik değil, gerçek ses-görsel eşleştirmeleri ile eğitilmiş, sonuçlar gerçek veriyle test edilmiştir.

**7.Karşılaşılan Zorluklar ve Çözümler**

 **Boyut Uyumsuzlukları:** Pretraining sırasında 128x128 görsellerle çalışılmışken, fine-tune dataseti 64x64 olduğu için model dönüştürüldü.

 **GPU Süresi Limiti:** Kaggle'daki GPU süresi sınırları nedeniyle zaman zaman eğitime ara verildi, bu durum model kaydetme–yükleme adımlarıyla dengelendi.

 **Veri Eksiklikleri:** Bazı ses–görsel dosyaları eksikti, bu nedenle try/except ve dosya kontrol mekanizmalarıyla hata yönetimi sağlandı.