



# Fake/Real Face Classification

Nickolas Bleykhman, Kensho Pilkey,  
Raheq Hassan, Kamal Deep  
Vasireddy, Kenneth Roberts

Participants in new research conducted by Royal Holloway and the Warburg Institute, both University of London, were likely to trust artificially generated faces, known as Generative Adversarial Networks faces (GAN), than real ones, making them more susceptible to misinformation and creating a cultural visual landscape towards a more deceptive society.

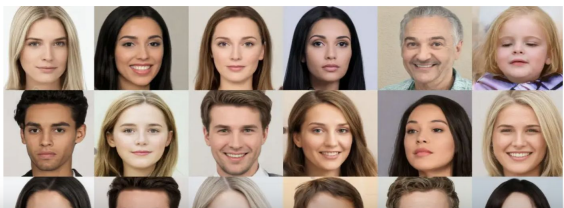


### Fake faces created by AI look more trustworthy than real people

Synthetic human faces are so convincing they can fool even trained observers, and they may be highly effective for use in scams

By [Christa Lesté-Lasserre](#)

14 February 2022



# Introduction/Motivation

## Problem:

- High-fidelity generative models (**StyleGAN, diffusion**) makes distinguishing real faces from AI-generated headshots critical
- Implications span **misinformation, identity theft, and media authenticity**

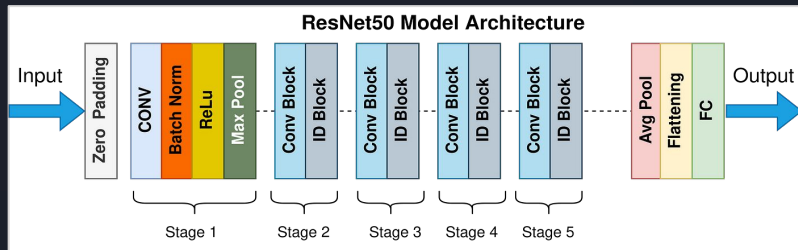
## Objective:

- To develop a **fast, reliable, and user-friendly web application** that can help individuals and organizations detect deep fake content **before it spreads.**

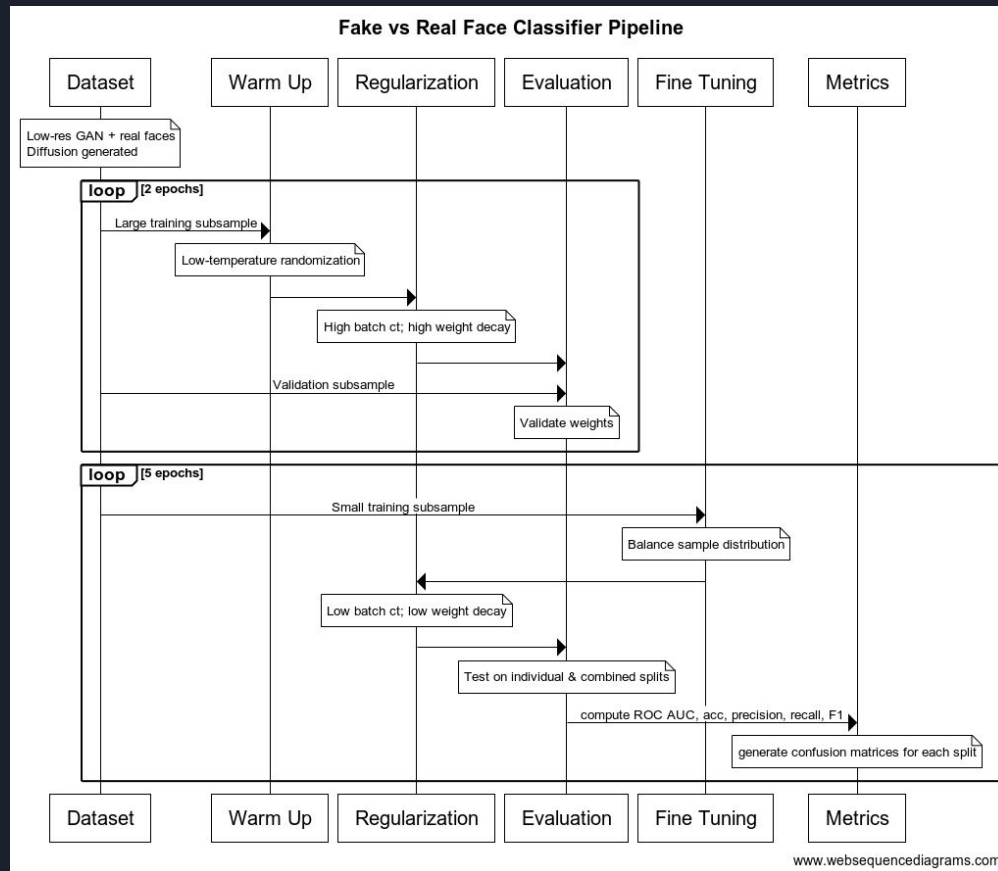
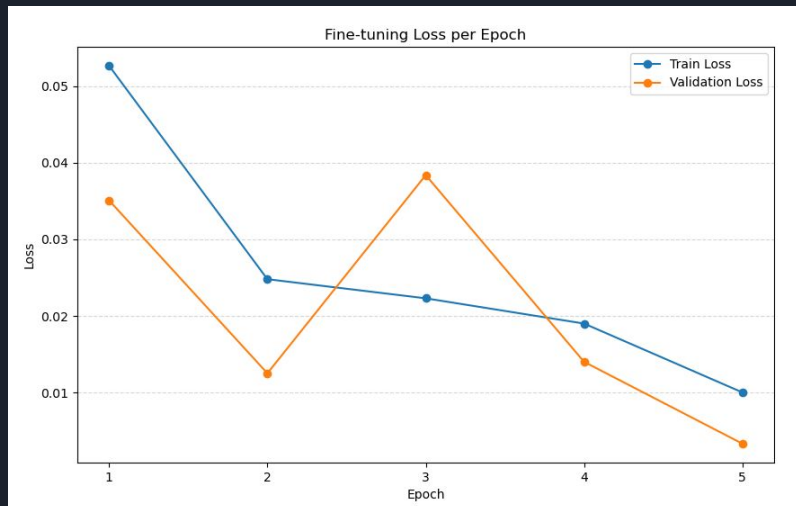
## Lots of Related Work on the Topic, Namely:

- **GAN Artifact Detection:** Marra *et al.* propose frequency analysis to spot GAN upsampling traces (Marra et al., 2019) [DOI:10.3390/app9071464]
- **Diffusion Model Forensics:** Yu *et al.* use texture and noise statistics to uncover diffusion synthesis anomalies (Yu et al., 2022) [arXiv:2203.14103].

# Methods & Experimental Design

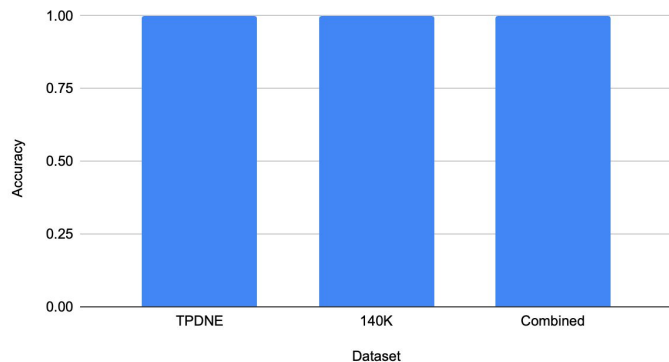


ResNet50 Architecture



# Results and Conclusions

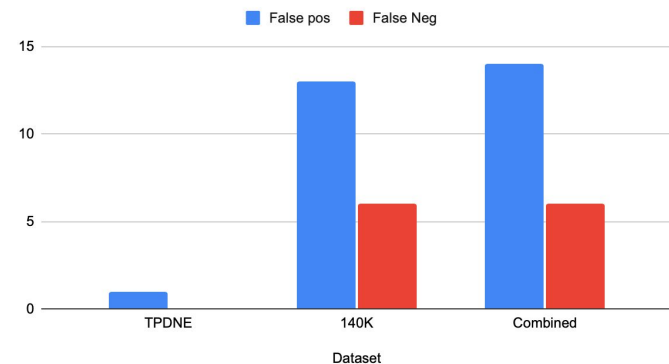
Test Accuracy per Dataset



- **TPDNE Test:** 1 misclass of 3,000 images
- **140K Test:** 19 misclass of 20,000 images
- **Combined Test:** 20 misclass over 23,000 images
- Precision & recall > 0.98 splits

Combined:  
14 false positives  
6 false negatives

Confusion per Dataset



## Future Work

- Fine-tune with samples on ResNet-101
- Fine-tune and experiment with Dall-E sample set



Live Demo

