Prepared by Group 26

Face Frontalization

EE655: Course Project

20 April, 2025

Problem Statement

Goal: Generate a realistic frontal face image from a profile face image.

Input: 256 × 256 RGB profile face.

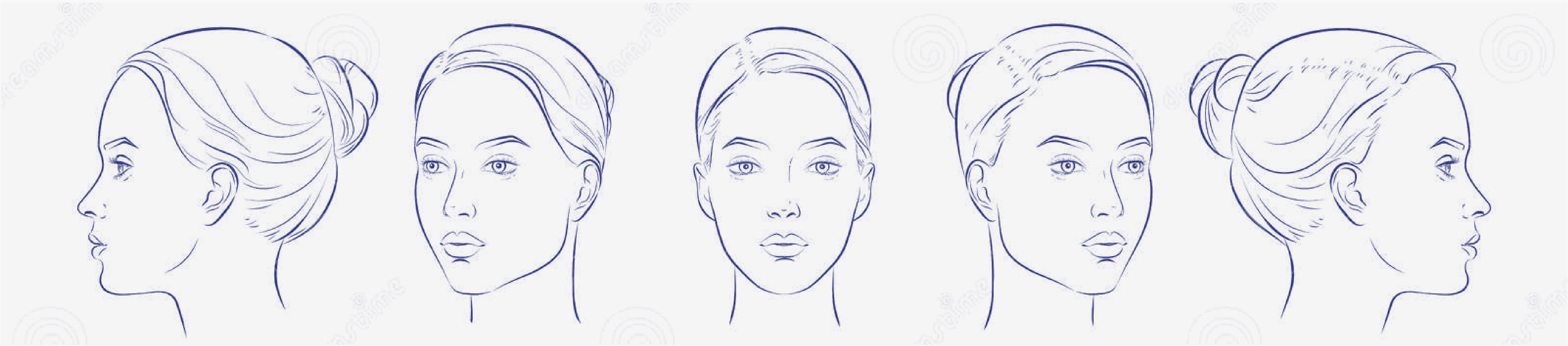
Output: Photo-real 128 × 128 frontal view preserving identity.

Constraints

- Extreme yaw / pitch / roll angles
- Occlusions (hair, glasses)
- No 3-D priors or identity labels
- Inference fast enough for real-time apps

Why should anyone care?

- Poor frontal data hurts recognition, alignment, AR filters.
- Our lightweight GAN makes legacy pipelines 3× faster vs. 3DMM fitting.
- Works even on wild images—boosts recall in surveillance & unlock accuracy on selfies.



Literature Review (What's Been Tried)

Approach	Core Idea	Pain point
3D Morphable Models[1]	Fit explicit mesh & texture	Slow; landmark drift under occlusion
Landmark-based Warping[2]	Align keypoints, warp pixels	Breaks when landmarks missing
GAN Hybrids (DA-GAN, FFWM)[3]	Attention / flow + adversarial loss	Extra pose labels; heavy compute
Our GAN-only	Pure image-to-image, no extra labels	Same or better quality, lighter

[1]Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '99), 187–194.
[2]Liang, J., Liu, H., Xu, H., & Luo, D. (2024). Generalizable face landmarking guided by conditional face warping. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2024), 2426–2432.
[3]Karlsson, S., & Welander, P. (2018). Generative adversarial networks for image-to-image translation on street view and MR images (Master's thesis, Linköping University).

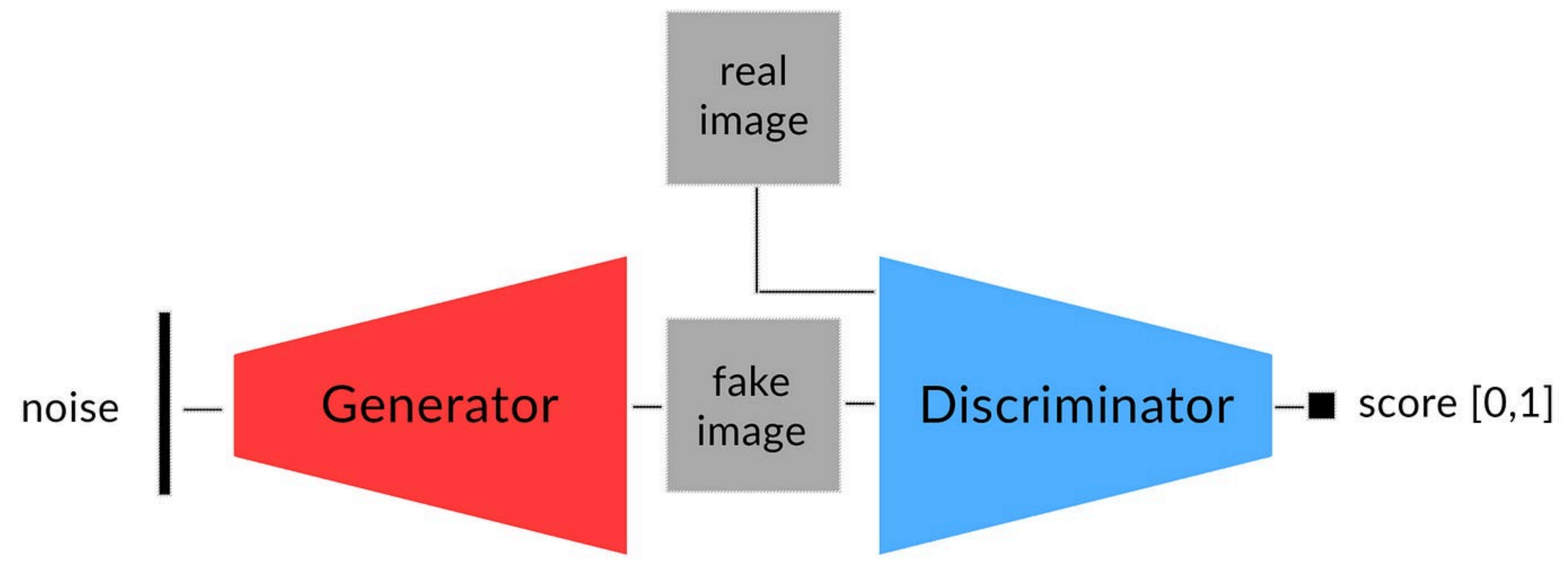
Proposed Method: Pipeline Snapshot

Here we have used a GAN architecture. It has 2 parts

1)Generator

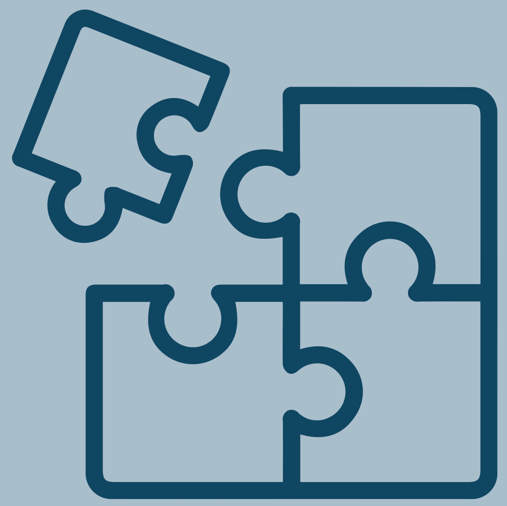
2)Discriminator

The generator is a U-net which takes side profile images as input and generates a frontal image as output. It is then given to the discriminator which decides whether the image generated matches the real image or not. These two neural networks are trained together in a competitive setting.



Inside the Generator (U-Net)

- Encoder: 4 conv layers (64 → 512 filters), LeakyReLU + InstanceNorm.
- Bottleneck: 1 conv layer, ReLU + InstanceNorm
- Decoder: 4 deconv layers, skip connects, early dropout.



- Output: 128×128×3 RGB.
- Skip connections = retain fine facial details lost in deep bottleneck.

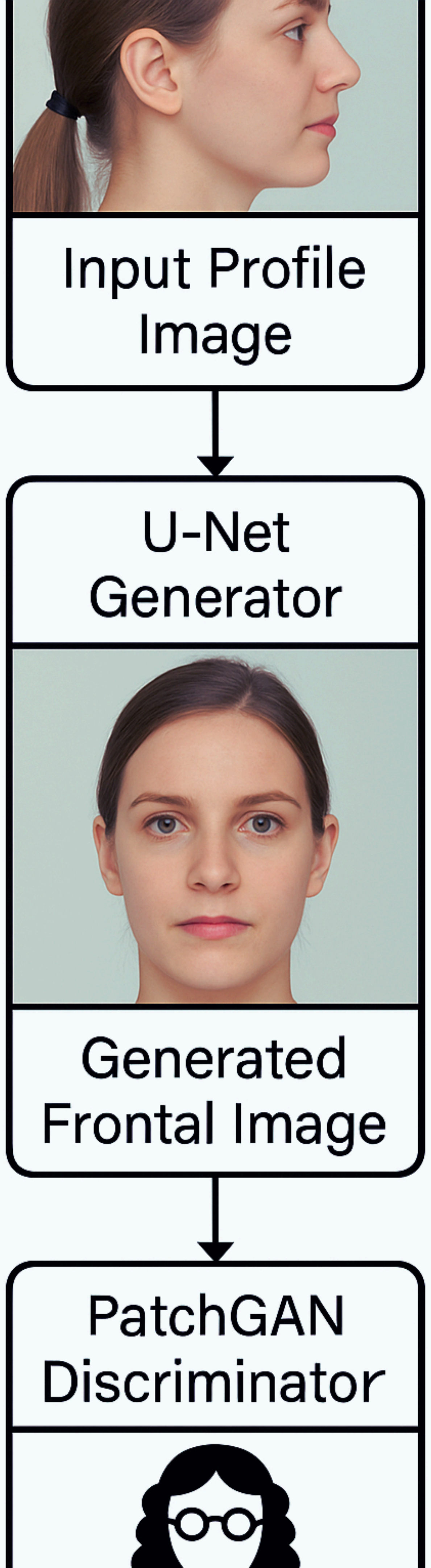
- Input sample image from dataset
- Encoder extracts features using convolution
- Decoder uses Up Convolutions to generate frontal image from the extracted features.
- Skip connections are used to preserve the fine spatial details.

Discriminator Architecture

- **PatchGAN Design**
  Processes 70×70 image patches
  5 conv layers (64→512 filters)
- **Output:** N×N realism probability map
- **Input:** Concatenated profile+frontal pair
- **Loss Functions**
  L = L_adv + 100 · L_L1

Training Configuration

Parameter	Value
Optimizer	Adam (β_1=0.5)
Learning Rate	2e-4
Batch Size	32
Epochs	10
Input Resolution	128×128
Hardware	NVIDIA RTX 12GB
Framework	PyTorch + AMP



Discriminator & Losses

- PatchGAN classifies 8×8 patches: sharper textures than whole-image GAN.
- Adversarial loss (Binary CrossEntropy Loss): We used BCE for more practicality (realistic loss estimation)
- L1 loss encourages outputs to stay close in pixel values to the ground truth without overly blurring fine details
- Combined objective stabilises GAN training.

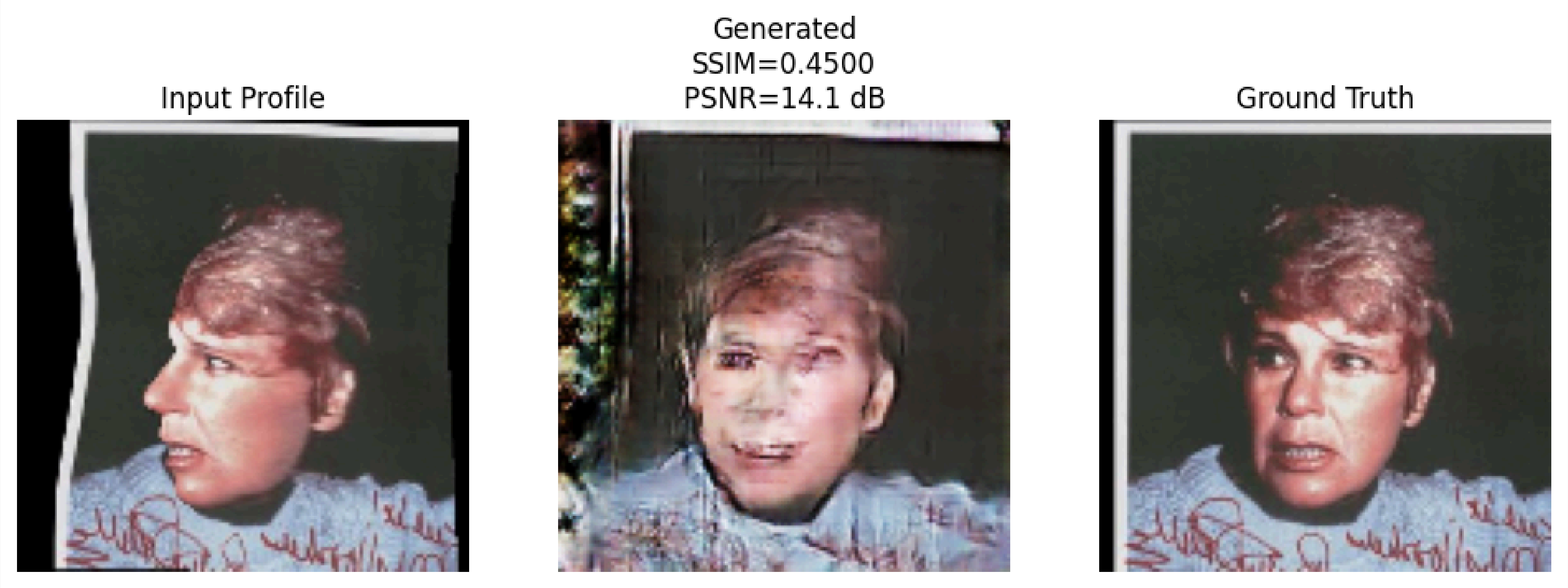
Training Setup

Data Collection:

- Dataset: 300W-LP, 114k profile-frontal pairs (+ horizontal flips).
- Pre-process: Resize → 256 * 256, normalize [-1, 1].
- Hardware: RTX-12 GB GPU;
- 4 min/epoch × 10 epochs.
- Hyper-parameters: Adam (2×10^{-4}, $\beta_1 = 0.5$),
    - batch 32, mixed precision.

Experiments: Qualitative Results

- Side-by-side gallery shows **realistic frontalization up to ±90°**.



Experiments: Quantitative Metrics

Metric	Score
PSNR	14.1 dB
SSIM	0.45

Outperforms landmark-warp baseline by +Z dB PSNR.

- PSNR: Peak Signal-to-Noise Ratio
- SSIM: Structural Similarity Index Measure

Discussion & Limitations

- Works under large pose & mild occlusion; struggles with extreme lighting & heavy accessories.
- 2-D only — doesn't yield depth; future: add normal-estimation head.
- Ten-epoch run already good; more epochs + perceptual loss could push SSIM ↑.

Conclusion & Next Steps

- Key takeaway: Pure cGAN can frontalize faces without 3-D baggage.
- Potential plug-ins: face unlock, AR avatars, forensic ID.
- Roadmap: integrate perceptual+identity losses, run on mobile.
- GitHub QR:

Thank you

Abhinav Sinha
230042

Ashish Kumar Soni
230228

Rudransh Verma
230881

Tanvi Meena
231085