Motivating Use Case

Background

Professional headshots are widely used on LinkedIn and CVs. Many people don't know how to take professional headshots leading people uploading low quality images or even postpone setting up the CV/Linkedin.

Why It Matters

Headshots influence first impressions in hiring and networking. Poor photos can negatively affect opportunities, especially for users without access to professional guidance.

Why It Is Challenging

Professionalism is subjective and depends on multiple visual factors like lighting, background, framing, and sharpness. Existing tools do not quantify or explain quality.

Current Solutions

Users rely on personal judgment, friends, or expensive photographers. Editing apps improve appearance but do not assess professionalism objectively.

Project Task Description

Formal Problem Statement

Given an image x, predict $y \in R^5$, a professionalism attribute vector.

Input

One user provided portrait image

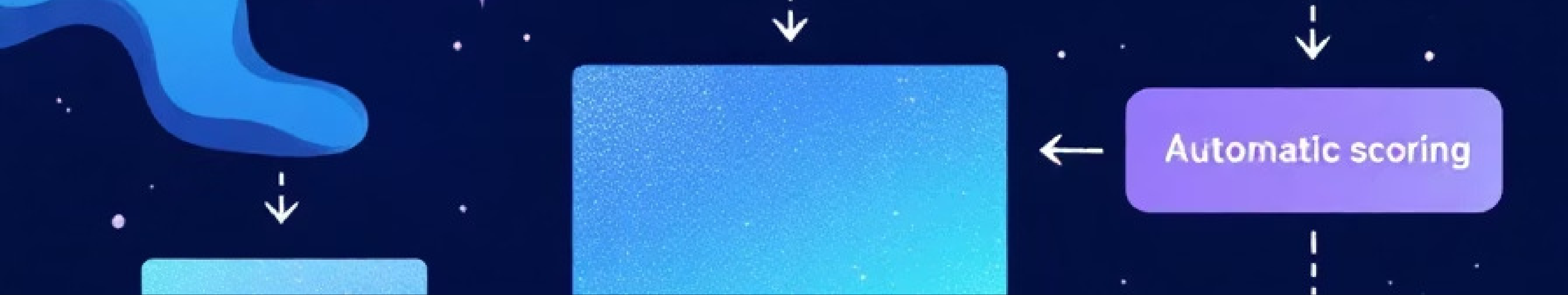
Output

A professionalism 6 attribute vector consisting of

- Lighting Even and Frontal
- Background Clean and Non-Distracting
- Business or Professional Attire Visible
- Neutral Professional Facial Expression
- Face Properly Framed and Centered
- Image Sharpness High

Novelty

We introduce a dedicated model for predicting headshot professionalism attributes directly from images, a task not explicitly addressed by existing vision models.



Models and Methods

1 Processing Pipeline

- Define a function g that maps a randomly sampled set of professionalism attributes to prompt modifications
- Apply g to base portrait images to generate synthetic headshots with known attribute degradations
- Use the applied attributes as labels

2 Models and Techniques

- Image generation: Diffusion models (WAN 2.2, FLUX, Krea 1) with prompt variations encoding professionalism attributes
- Image refinement: Super-resolution upscaling to enhance facial detail and texture
- Attribute prediction: Multi-label image classifier using a CNN or Vision Transformer (ResNet50 or ViT-B/16)

3 Adjustments and Fine Tuning

- Use a pre-trained image model to headshot images instead of training from scratch
- Fine tune the model using lightweight adaptation (e.g., LoRA).

Training and Evaluation Data

Training and Evaluation Data Requirements

- Unlabeled real portrait images to anchor the face distribution
- Multiple synthetic variations per portrait with known injected professionalism attributes
- Prompt templates that explicitly encode each professionalism attribute used for labeling

Dataset

- Base dataset: FFHQ-like real face images Used only as unlabeled portrait inputs

Labeling

- No manual labeling is performed
- Labels are assigned automatically based on the attributes injected during synthetic generation
- Each image receives binary labels for professionalism related attributes

Synthetic Data Generation

- For each base portrait, multiple image-to-image variants are generated using a diffusion model
- Prompt modifications inject specific professionalism attributes (e.g., lighting, background, framing, blur)
- Generated images inherit labels corresponding to the injected attributes

Evaluation Metrics and Protocol

Ground Truth

- Binary professionalism attribute labels defined directly by injected synthetic attributes

Model Performance Metrics

• F1-score computed separately for each professionalism attribute to see how well the model detects each issue

• An overall F1-score obtained by averaging the per-attribute F1 scores equally

Evaluation Protocol

- Split the dataset into training, validation, and test sets by base portrait to avoid leakage
- Report performance on the held-out test set

Synthetic Data Validation

• Verify that generated images match the intended prompt attributes

• Perform limited manual spot checks to confirm obvious label correctness