Professional Portrait Scoring

```
IMAGE 2: □ PERFECT MATCH

□ Lighting Even and Frontal           P:1 | T:1
□ Background Clean and Non-Distractin P:1 | T:1
□ Business or Professional Attire Vis P:0 | T:0
□ Neutral Professional Facial Express P:0 | T:0
□ Face Properly Framed and Centered   P:0 | T:0
□ Image Sharpness High                P:0 | T:0
```

Project Specifications

Input: **A single headshot image.**	• **Output: Vector with 6 binary attributes:** ▪ **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**

What Changed from the Proposal

- During analysis, we identified clearer key attributes that describe professionalism.
- Switched from the proposed ResNet50/Standard ViT to a Swin Transformer.

Bad Framing

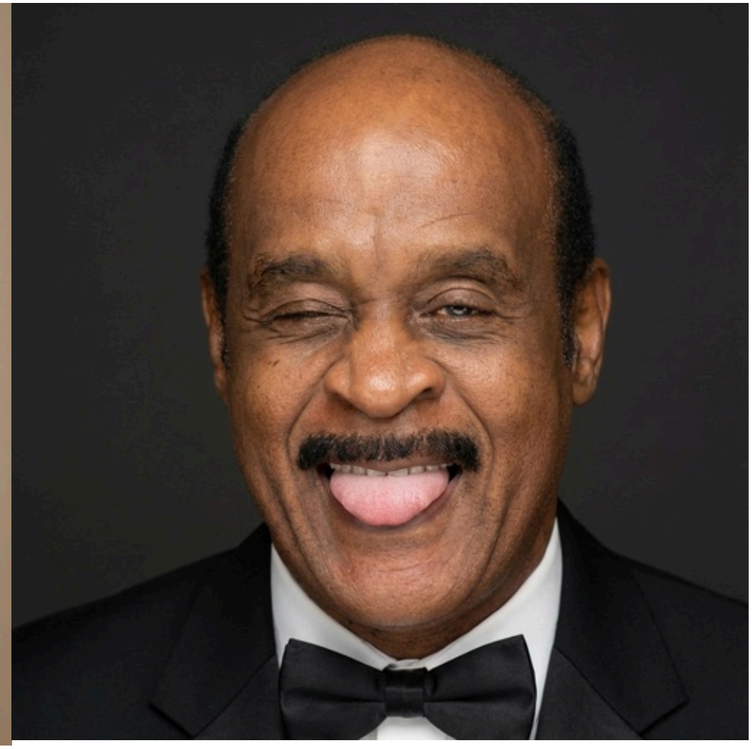
Blurred

Bad Lightning

Bad Background

Bad Attire

Bad Expression



Related Literature

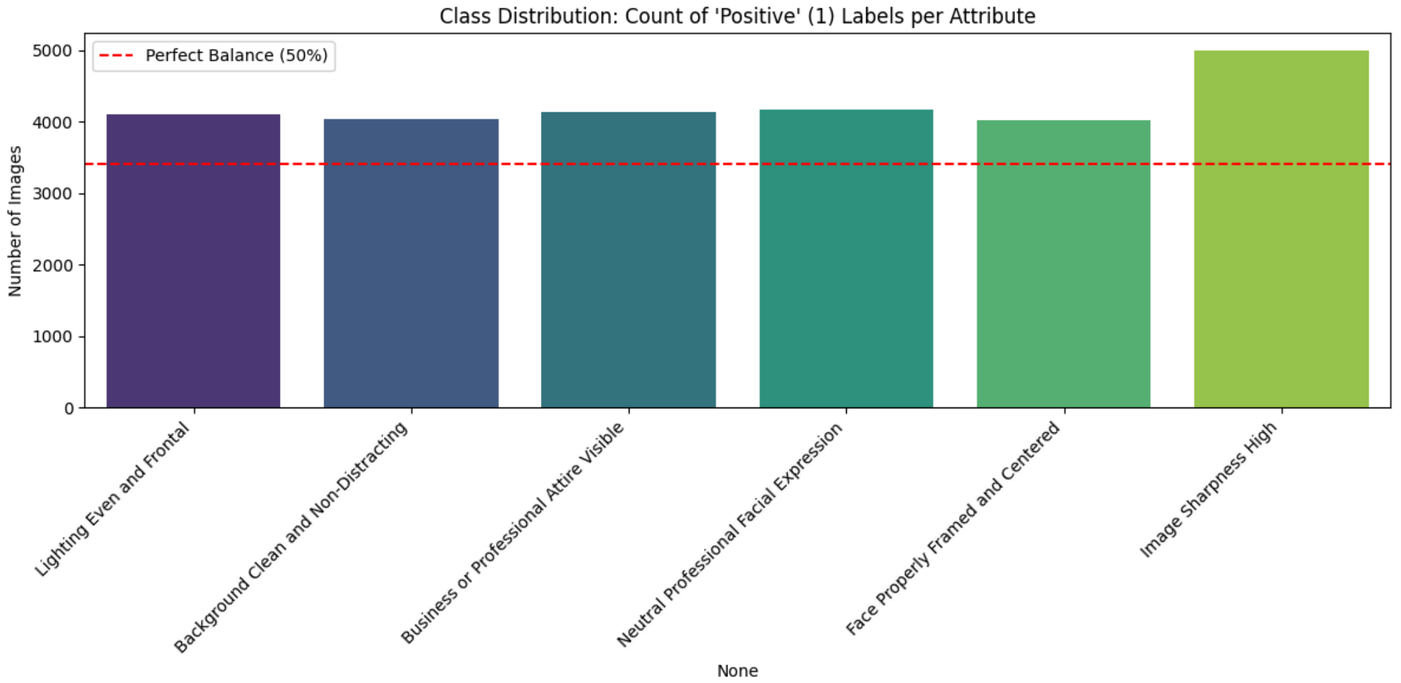
Paper	Focus	Key Idea	Relation to Our Project
Deep Learning-Based Image Aesthetic Quality Assessment – Review (2025)	General image aesthetics	Image quality depends on interpretable factors such as lighting and composition	Supports breaking professionalism into visual attributes
Face Image Quality Assessment: A Literature Survey (2022)	Face / portrait quality	Face Image Quality Assessment depends on pose, illumination, and sharpness	Grounds our focus on face-centric quality attributes such as pose, facial illumination, and sharpness, which are central to professional headshots.
Transformer for Image Quality Assessment (2021)	Image quality prediction	Transformers capture global image structure effectively	Motivates using Transformer-based architectures (like Swin) over CNNs

Dataset (FFHQ origin)

- **Dataset:** Created 6,829 synthetic professional portraits using Gemini 2.5 Flash (Image-to-Image).
- **Technique:** Generative Labeling (Labels Prompt). We first sampled random label vectors, then mapped them to dynamic prompts.
- **Sample Definition:** Each sample consists of a 512×512 RGB portrait image paired with a 6-dimensional binary attribute vector and metadata (JSON).

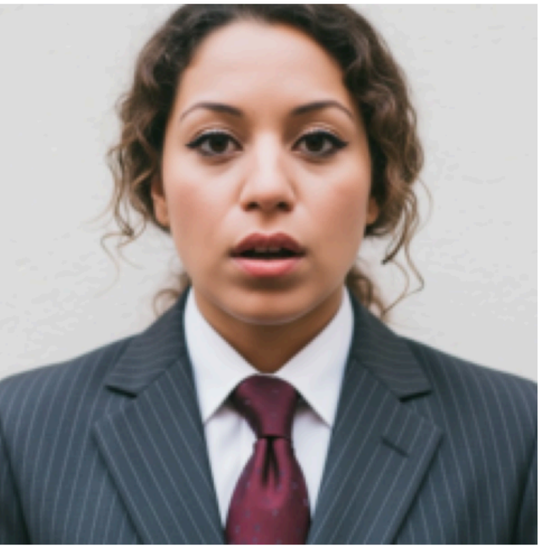
EDA

- **Class Distribution: Positive-skewed (~60%) Most labels are 1, not 0.**
- **Prompt Length: Average 40-60 words per input.**



Model:Microsoft Swin Transformer (swin_tiny_patch4_window7_224), pretrained on ImageNet.

Training: Fine-tuned for 10 epochs.
Exact Match Ratio (all 6 attributes correct): 15.67%

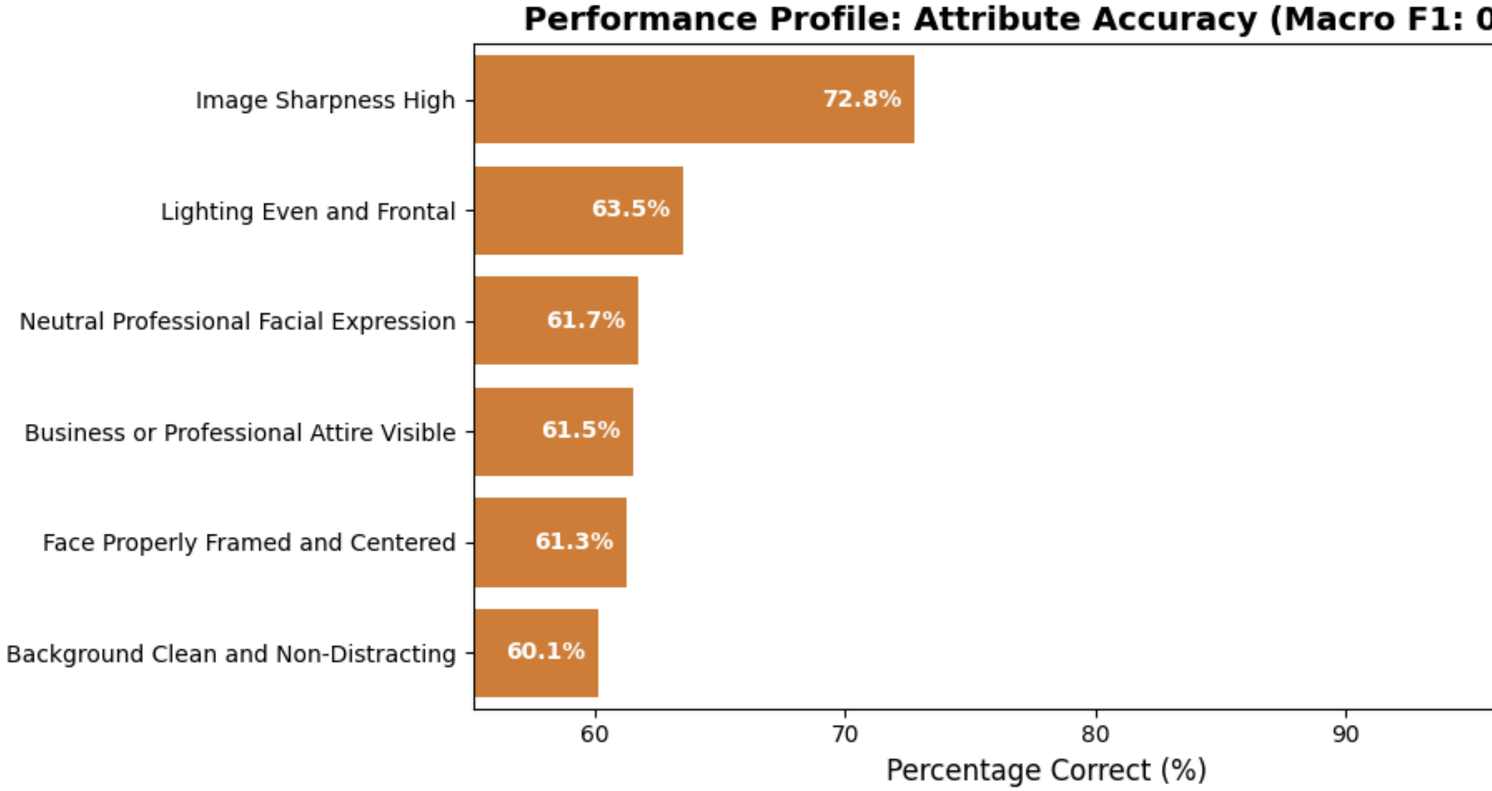


```
IMAGE 3: ⚠ HAS ERRORS

 Lighting...      | Pred: 1 | True: 1
 Background...    | Pred: 1 | True: 1
 Business...      | Pred: 1 | True: 1
 Neutral...       | Pred: 1 | True: 0
 Face...          | Pred: 1 | True: 0
 Image...         | Pred: 1 | True: 1
```

```
IMAGE 4: ⚠ HAS ERRORS

 Lighting...      | Pred: 1 | True: 1
 Background...    | Pred: 1 | True: 0
 Business...      | Pred: 1 | True: 1
 Neutral...       | Pred: 1 | True: 1
 Face...          | Pred: 1 | True: 1
 Image...         | Pred: 1 | True: 1
```



Did you achieve the desired results?

- Partially.
- The model predicts individual attributes reasonably well, but often fails to get all attributes correct together (Exact Match ≈ 15%)

What did we learn?

- The model performs better on single attributes than on joint predictions.
- Some attributes (e.g., framing, expression) are visually ambiguous.

What should be different?

- Some labels are unclear: Things like face framing are not really Yes or No. Measuring position would work better.
- Less sharpness examples.
- Better promoting, lightning prompts are not good enough

Future ideas

- Change the training settings for optimization
- See what the model looks at: Use attention maps to understand mistakes.
- Compare models: Test other models to see if this one is really better.
- Make another batch of syntatic data, with optimized prompts.
- Focus on the face, based on the research presented