Supplementary: P-Age: Pexels Dataset for Robust Spatio-Temporal Apparent Age Classification

1. Qualitative Results

In addition to the tabular results in Table 4 of the main artcile, we perform a qualitative analysis as shown in Figure 1 between the best SOTA Levi [1] age classifier and our AgeFormer on the P-Age-Face dataset. The results indicate that our AgeFormer can accurately predict age in situations such as low-resolution, obscured faces, etc.

2. Qualitative Analysis

In this section, we perform two experiments to test the model on i) video quality and ii) variation in sampling-rate. First, we perform experiments to resize the input frames to a low-resolution and bring it back to the input size of the inference 224×224 to simulate the loss of spatial quality and mimic the low-resolution input. We compare our AgeFormer with other SOTA methods, as shown in Figure 2a. The results in this figure show the robustness of our method against a wide range of low-resolution input that beat existing methods.

Second, we also experiment our method across various sampling strides to ensure that our model does not rely too much on superficial features, like a sudden change between consecutive frames as an indicator for specific classes, such as kids. Ageformer is mostly stable against these changes, as visualized in Figure 2b.

References

[1] Gil Levi and Tal Hassner. Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 34–42, 2015. 1