



Figure 13: UMAP Projection of smile and blond hair direction computed from representation gathered at different depths of the generator. A cluster of codes that correlates similarly to given attributes is more pronounced in the mid-layer for both smile direction and blond hair direction. Such cluster cannot be observed in the later layer.

9. Supplementary Material

In this section, we compare UMAP projections that are constructed from different depth of the generator in an attempt to explain the rationale behind our choice of activation layer in the generator where we extract internal representation from. We see evidence from studies like [SZ20] where linear maps from different depth of the generator give rise to different factors of variation. Also, it was studied in [KLA19] that deeper layer tends to capture localized details whereas earlier layers tend to capture coarse details. Since we are studying facial attributes that are not too coarse (such as brightness or a direction of a face) nor too localized (such as direction of a wavy hair or position of freckles), we hypothesize we would see most variability in the internal representation of latent codes with respect to the facial attributes that we consider.

In Figure 13, we compare UMAP projections of two different directions constructed using representations obtained from different depths of a generator. First looking at the blond hair direction, a cluster of smiling people can easily be identified in the early to mid layer representations. However, such cluster cannot be identified in the deeper layer. We can also see such a trend in the smile direction, where codes that smiles less tend to be grouped in the middle layer, but less so in the deeper layer. Likewise, for most of the directions that we study in this paper, mid-layer tend to show well-clustered latent codes with the attributes. We can see that the mid-layer strikes a good compromise in capturing varying semantics, and thus we choose this layer as part of our experiments.