6.1. Analysis

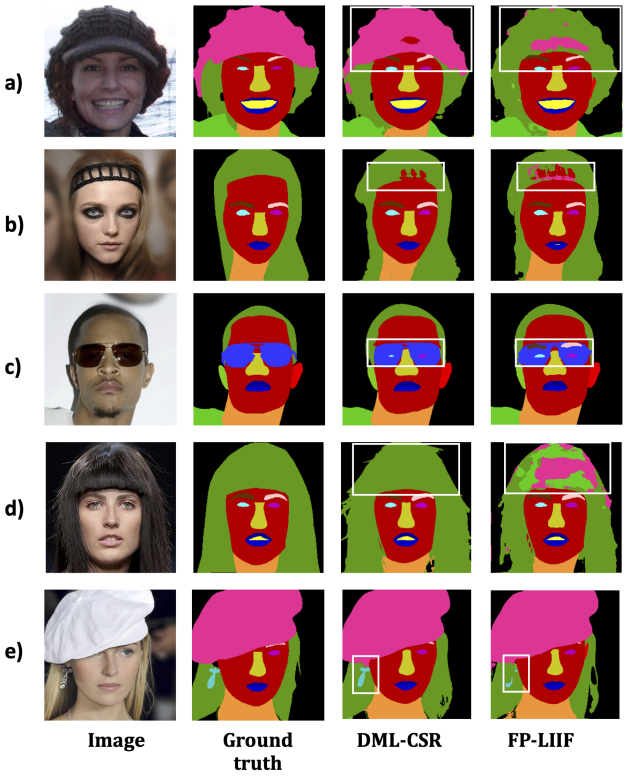

Figure 10. Few results where DML_CSR performed better than FP-LIIF on CelebAMask-HQ dataset.

The quantitative results shown in Table 1, 2 points that even though FP-LIIF fares better in mean F1, the best classwise performance is scattered across multiple models. But at the same time, the gap between the best classwise scores and FP-LIIF's classwise scores is marginal. Therefore, we try to further identify the problematic areas and include visualizations of FP-LIIF's worst-performing results compared to DML_CSR in F1 in Figure 11, 10. It can be seen from Figure 11 that rows a) and c) have negligible differences, and in the remaining rows, both are performing poorly in the problematic regions of hair and face. In Figure 10's rows b) and d), the F1 scores for these are debatable because of incorrect labeling in the ground truth. In the remaining rows, the underrepresented class of hat and earrings are bringing down our performance. Therefore the current setup of FP-LIIF is affected by a lack of data as compared to DML_CSR. This can also be corroborated by Table 2. It is also necessary to point out that the ground truth data of CelebAMask-HQ is noisy (Figure 8) and can cause problems in training and testing.

From the point of view of inference time, FP-LIIF could be used to generate segmentation at a lower resolution, and

Figure 11. Few results where DML_CSR performed better than FP-LIIF on LaPa dataset

the generated output scaled at the required higher resolution to improve inference time and hence increase fps. The generation of lower-resolution segmentation does not require any additional training and is an outcome of being an implicit neural representation network. The 128-resolution version of FP-LIIF clocked an fps of 294 compared to the regular version of resolution 256, which runs at 120 fps. This makes our model more conducive for low compute devices.

6.1.1 Variance in performance over mulplte runs

We also calculate the mean and variance of our model's F1 score for Lapa, CelebAMask-HQ and Helen in Table 7 It

	Mean	SD
F1 Lapa	92.35	0.06
F1 Celeb	85.90	0.20
F1 Helen	91.12	0.10

Table 7. Mean and Variance of FP-LIIF

should be noted that other state-of-the-art works do not report these mean and variance over multiple runs and therefore direct comparison of these numbers is not possible.