

# Supplementary Materials of *Discrete Point-wise Attack Is Not Enough: Generalized Manifold Adversarial Attack for Face Recognition*

## A. Ablation studies

We verify the validity of the three local editors and  $D_{AU}$ . Fig. 1 demonstrates the poor visual quality of the adversarial examples without three local editors. Tab. 1 illustrates that the  $D_{AU}$  and local editors help to improve the expression accuracy of the generated adversarial examples.

Table 1. MSE between the AU vector of the generated adversarial examples (detected by OpenFace) and the given AU vector.

	Without $D_{AU}$	Without local editors	Original
MSE	0.5549	0.6283	<b>0.3582</b>

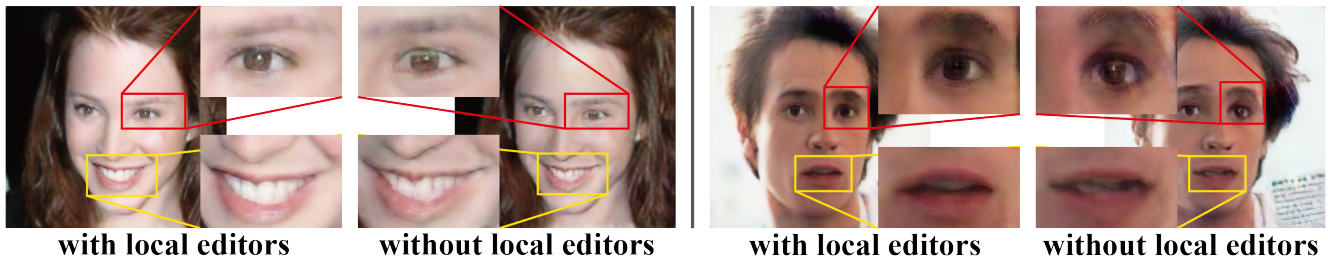
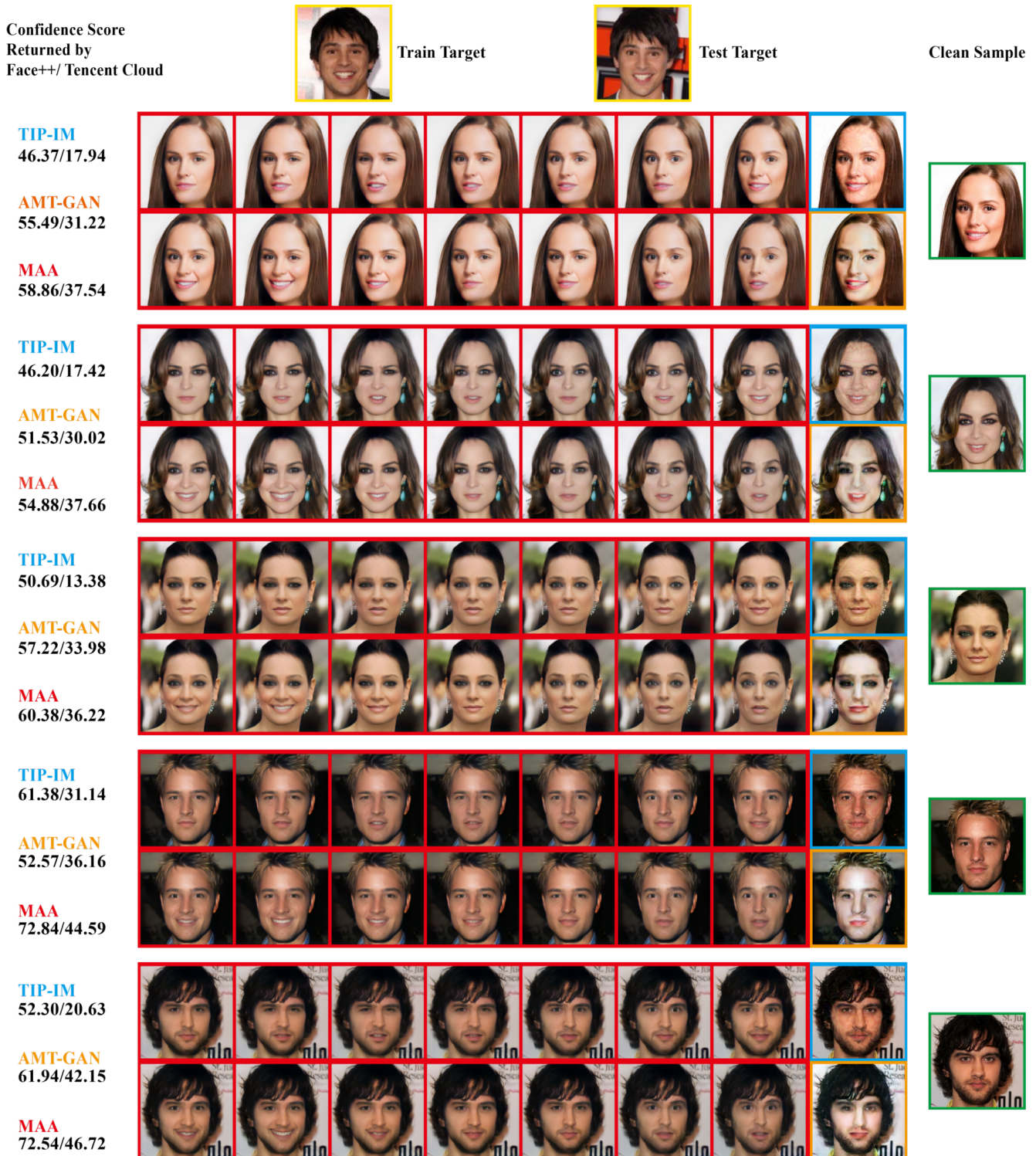


Figure 1. Ablation studies of three local editors.

## B. More visualization results

Due to the limited length of the main text, we have attached more visualization results here, which show that our model MAA outperforms the comparative models AMT-GAN and TIP-IM.

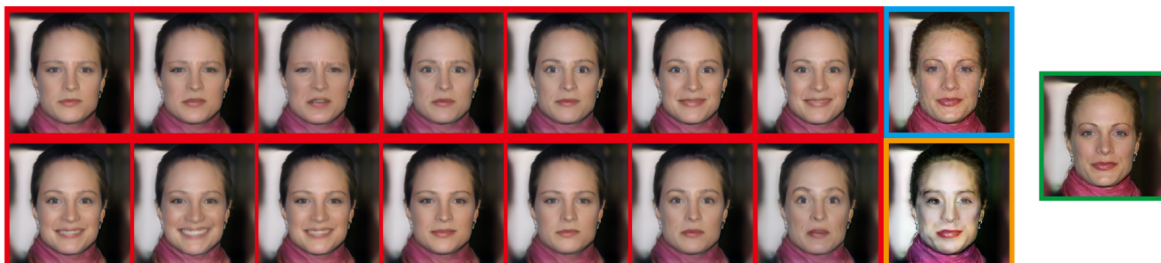




**TIP-IM**  
46.67/13.90

**AMT-GAN**  
63.89/39.19

**MAA**  
70.04/39.69



**TIP-IM**  
45.87/24.48

**AMT-GAN**  
53.36/40.08

**MAA**  
55.45/42.28



**TIP-IM**  
58.35/29.26

**AMT-GAN**  
62.65/45.07

**MAA**  
69.45/48.73



**TIP-IM**  
62.72/25.36

**AMT-GAN**  
61.82/38.71

**MAA**  
79.23/52.40



**TIP-IM**  
57.58/41.55

**AMT-GAN**  
67.53/63.02

**MAA**  
78.89/76.70



Figure 2. Visualization results of TIP-IM (highlighted by blue boxes), AMT-GAN (highlighted by orange boxes) and MAA (highlighted by red boxes). All the methods are trained to attack train target and the results returned by Face++/Tencent are the confidence score between the test target and adversarial examples. The results of MAA are the average scores of adversarial examples based on the same clean sample.