Supplementary Material for Multiple Exemplars-based Hallucination for Face Super-resolution and Editing

Kaili Wang[1,2], Jose Oramas[2], and Tinne Tuytelaars[1]

[1] KU Leuven, ESAT-PSI
{first name.last name}@esat.kuleuven.be
[2] University of Antwerp, imec-IDLab
{first name.last name}@uantwerpen.be

This documment constitutes supplementary material for the paper "Multiple Exemplars-based Hallucination for Face Super-resolution and Editing". It extends the original submssion in four fronts. First, it includes details of the network architectures used in each experiment (Sec. 1) and the hyperparameter setting. Second, we provide additional qualitative results for the face super-resolution experiment (Sec. 2). Third, we provide additional results of facial feature editing experiment. Including the exemplars used in Fig. 1 of the main manuscript. (Sec. 3).

1 Architecture and hyperparameters

1.1 Architecture

In this supplementary material, we define the convolution layer with kernel size k, padding p, stride s, the input channel c_{in} and the output channel c_{out} as $Conv(c_{in}, c_{out}, kps)$. Similarly, we define the transposed convolution layer as $Conv_{trans}(c_{in}, c_{out}, kps)$.

ISR block In the ISR block, each convolution block contains
$[Conv_{trans}(c_{in}, c_{out}, k4p1s2) \rightarrow Conv(c_{out}, c_{out}, k3p1s1)]$.

Each residual block contains
$[Conv(c_{in}, c_{out}, k3p1s1)) \rightarrow Conv(c_{out}, c_{out}, k3p1s1)]$.

The three-channel RGB images are first sent to the input block to increase the channel size but keep the spacial dimension.
$[Conv(3, 64, k3p1s1) \rightarrow Conv(64, 64, k3p1s1)]$.

Before the RGB images are generated, we apply one RGB converter block which has one convolutional layer $[Conv(c_{in}, 3, k3p1s1)]$.

The activation function is $ReLU$ except for the RGB converter block, where we use $Tanh$ to scale the pixel range to $[-1, 1]$.

For experiment with the scaling factor of 8, we use 3 convolution blocks and 1 residual block. For the convolution blocks, $c_{in} = [64+64, 128+128, 256]$, $c_{out} = [128, 256, 128]$. For the residual block, $c_{in} = 128$, $c_{out} = 64$. For experiment with the scaling factor of 16, we use 4 convolution blocks and 1 residual block. For the convolution blocks, $c_{in} = [64+64, 128+128, 256, 128]$, $c_{out} = [128, 256, 128, 64]$. For the residual block, $c_{in} = 64$, $c_{out} = 64$.

Encoder The encoder is used in the PWAve module to extract the feature maps of $I_E x$. For CelebA/Webface dataset with scaling factor of 8, the architecture is:

$[ReLU(Conv(3, 64, k3p1s1)) \rightarrow ReLU(Conv(64, 128, k3p1s1)) \rightarrow$
$ReLU(Conv(128, 256, k4p1s2)) \rightarrow ReLU(Conv(256, 128, k4p1s2)) \rightarrow$
$ReLU(Conv(128, 64, k4p1s2))]$.

For the CelebA/Webface dataset with scaling factor of 16, the architecture is:

$[ReLU(Conv(3, 64, k3p1s1)) \rightarrow ReLU(Conv(64, 128, k4p1s2)) \rightarrow$
$ReLU(Conv(128, 256, k4p1s2)) \rightarrow ReLU(Conv(256, 128, k4p1s2)) \rightarrow$
$ReLU(Conv(128, 64, k4p1s2))]$.

$\mathbf{G_S}$ For CelebA/Webface dataset with scaling factor of 8 and 16, the architecture is:

$[ReLU(Conv(3, 32, k3p1s1)) \rightarrow ReLU(Conv(32, 32, k4p1s2)) \rightarrow$
$Tanh(Conv(32, 3, k3p1s1))]$.

WNN For $WNN-1$, the architecture is

$[Conv(64+3, 128, k1p1s1) \rightarrow Conv(128, 256, k1p1s1) \rightarrow$
$Sigmoid(Conv(256, 1, k1p1s1))]$.

For $WNN-2$, the architecture is:

$[Conv(128+3, 256, k1p1s1) \rightarrow Sigmoid(Conv(256, 1, k1p1s1))]$

1.2 Hyperparameters

The hyperparameters in the total loss functions are: $\lambda_1 = 0.1$, $\lambda_2 = 0.5$, $\lambda_3 = 0.1$ for all CelebA dataset experiments. $\lambda_1 = 0.1$, $\lambda_2 = 0.1$, $\lambda_3 = 0.1$ for all Webface dataset experiments

2 Additional Qualitative results on Face Super-resolution Experiment

In this section, Fig. 1 and Fig. 2 show additional examples on the CelebA and WebFace datasets, respectively.

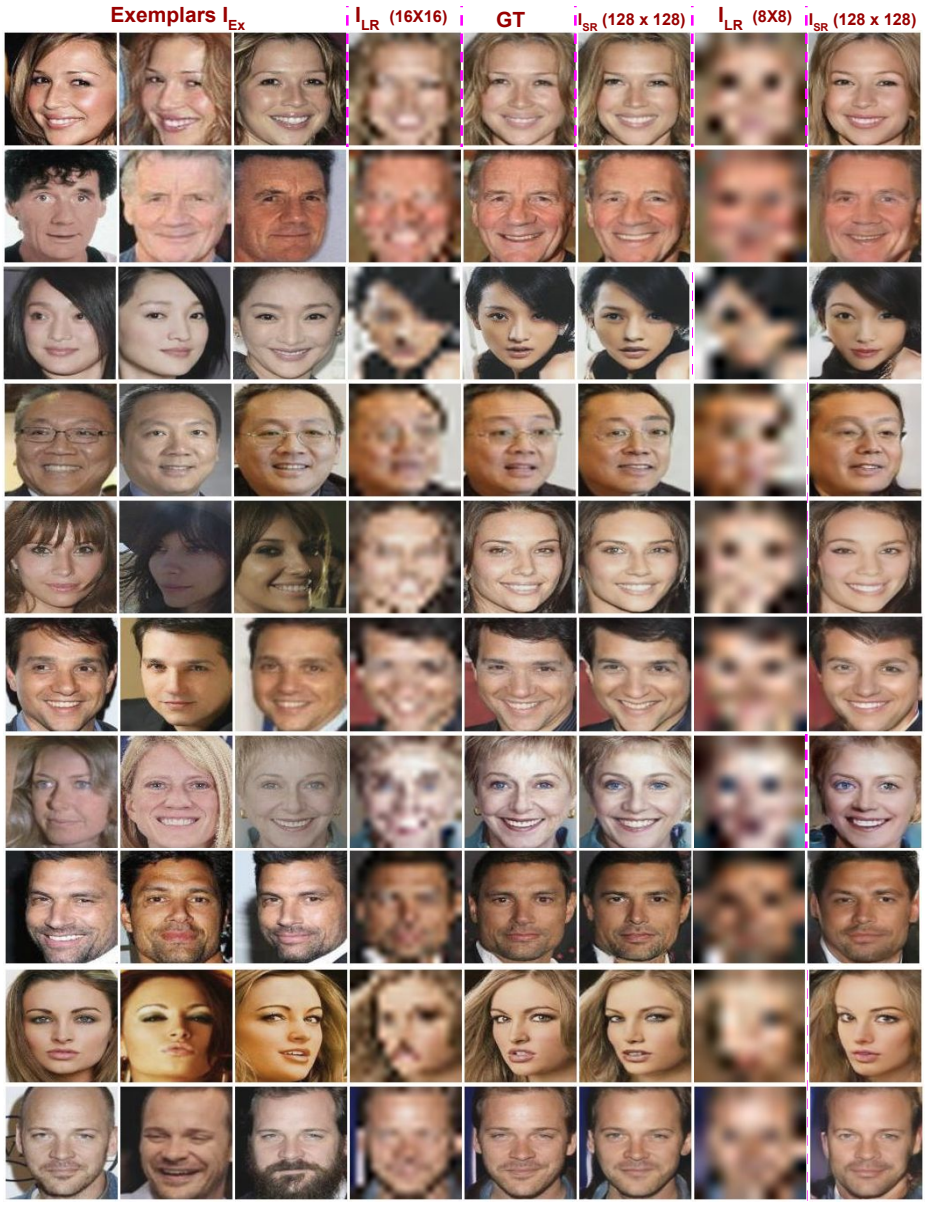

Fig. 1: Additional examples for qualitative results on the CelebA dataset. We show two scaling factors: $\times 8$ and $\times 16$. The resolution of HR images is 128×128. All the images are from testing set.

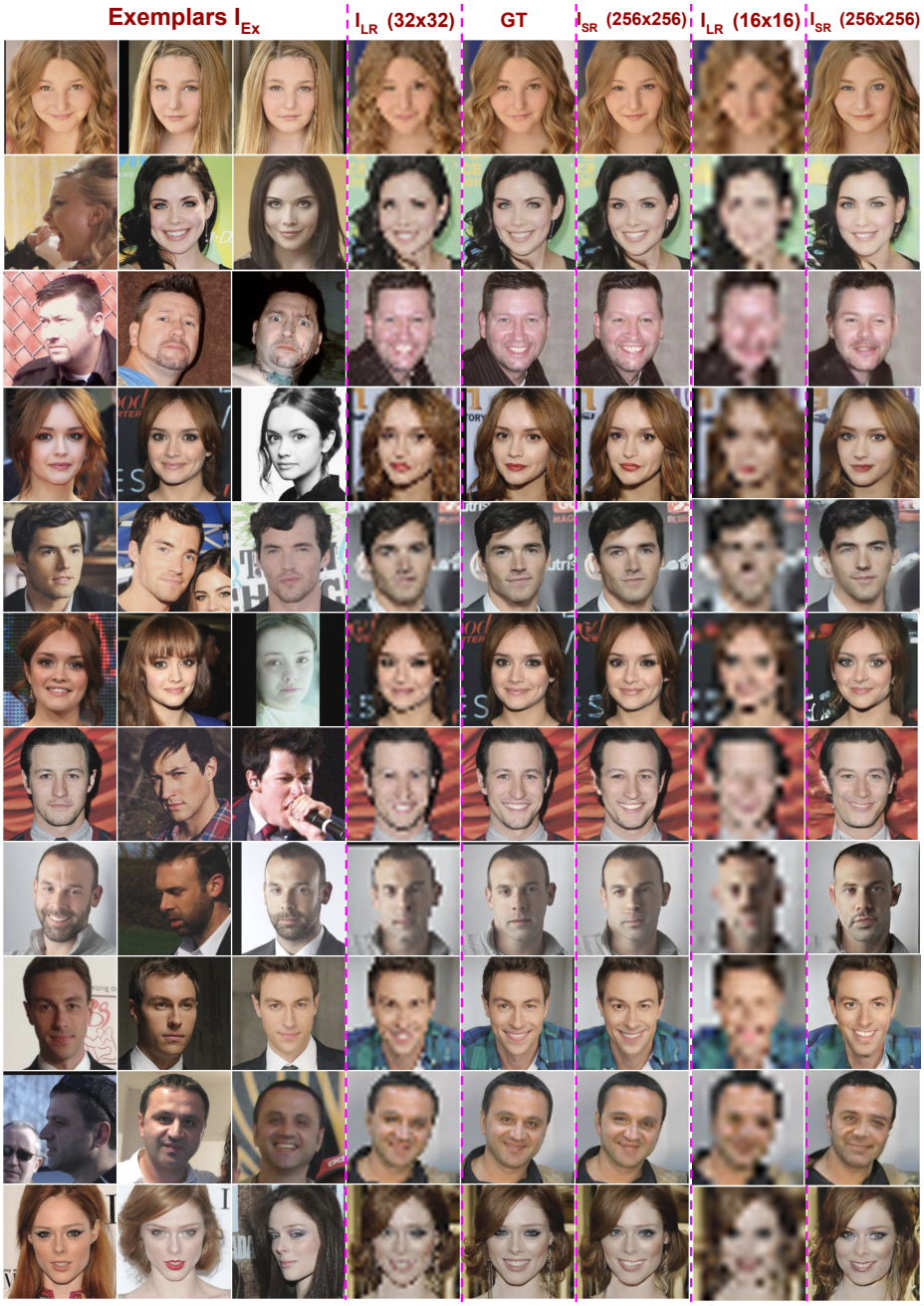

Fig. 2: Additional examples for qualitative results on the WebFace dataset. We show two scaling factors: $\times 8$ and $\times 16$. The resolution of HR images is 256×256. All the images are from testing set.

3 Additional results on Facial Feature Editing Experiment

In this section, Fig. 3 shows the full version of our teaser image in the main manuscript (i.e. including the corresponding exemplars for each super-resolved image.) Fig .4 shows more face feature editing examples.

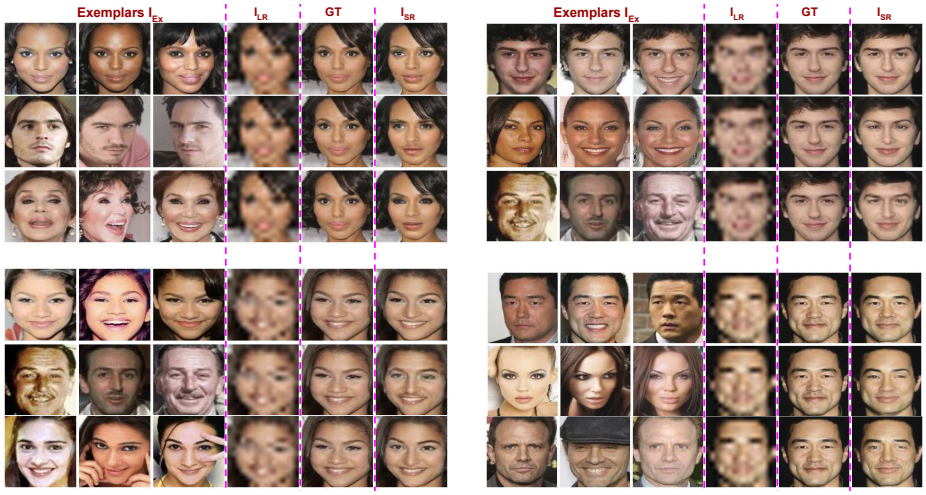

Fig. 3: Full version of the teaser image in our manuscript. In each row, the first three images are the exemplars.

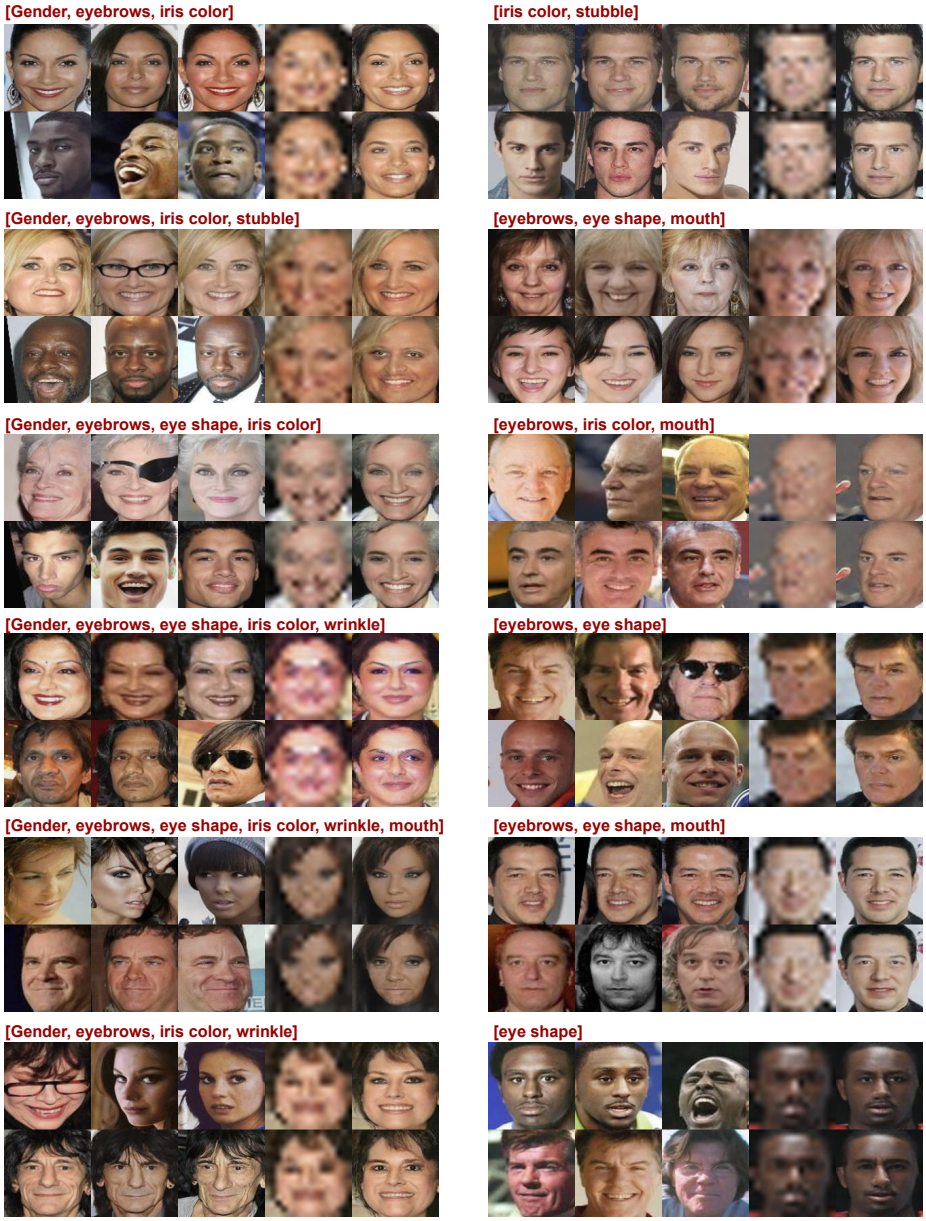

Fig. 4: Additional examples of editing/modifying facial features via exemplars. For each set, the first three images are the I_{Ex}, followed by the I_{LR} and I_{SR}. The edited facial features are displayed on the top of each set.