Supplementary Material of BBDM

Bo Li, Kaitao Xue, Bin Liu
School of Mathematics and Information Science, Nanchang Hangkong University, Nanchang, China

Yu-Kun Lai
School of Computer Sciences and Informatics, Cardiff University, Cardiff, UK

This supplementary material provides details that are not included in the main paper due to space limitations. We first fill in the derivation details of Section 3.1.3 in the paper. Then the implementation details of BBDM will be provided. After that we will provide the user study results. Finally, we will present more qualitative experiment results. The code is publicly available at https://github.com/xuekt98/BBDM.

1. Deduction Details of Training Objective

As shown in Eq.(11) in the paper:

$$q_{BB}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0,\boldsymbol{y}) = \mathcal{N}(\boldsymbol{x}_{t-1};\tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t,\boldsymbol{x}_0,\boldsymbol{y}),\tilde{\delta}_t\boldsymbol{I})$$

which can be written in the following Probability Density Function (PDF) form:

$$f(\boldsymbol{x}_{t-1}) = \frac{1}{\sqrt{2\pi\tilde{\delta}_t}} e^{-\frac{\left(\boldsymbol{x}_{t-1}-\tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t,\boldsymbol{x}_0,\boldsymbol{y})\right)^2}{2\tilde{\delta}_t}}$$

In the same way, the right part of Eq.(11) can also be represented as PDF which contains three sub-parts.
From Eq.(8), we can have $q_{BB}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\boldsymbol{y})$:

$$f(\boldsymbol{x}_t) = \frac{1}{\sqrt{2\pi\delta_{t|t-1}}} e^{-\frac{\left(\boldsymbol{x}_t-\left(\frac{1-m_t}{1-m_{t-1}}\boldsymbol{x}_{t-1}+(m_t-\frac{1-m_t}{1-m_{t-1}}m_{t-1})\boldsymbol{y}\right)\right)^2}{2\delta_{t|t-1}}}$$

The PDF of $q_{BB}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_0,\boldsymbol{y})$ and $q_{BB}(\boldsymbol{x}_t|\boldsymbol{x}_0,\boldsymbol{y})$ can also be derived based on Eq.(4):

$$f(\boldsymbol{x}_{t-1}) = \frac{1}{\sqrt{2\pi\delta_{t-1}}} e^{-\frac{\left(\boldsymbol{x}_{t-1}-\left((1-m_{t-1})\boldsymbol{x}_0+m_{t-1}\boldsymbol{y}\right)\right)^2}{2\delta_{t-1}}}$$

$$f(\boldsymbol{x}_t) = \frac{1}{\sqrt{2\pi\delta_t}} e^{-\frac{\left(\boldsymbol{x}_t-\left((1-m_t)\boldsymbol{x}_0+m_t\boldsymbol{y}\right)\right)^2}{2\delta_t}}$$

Considering that the PDF of the left part and right part of Eq.(11) should be equal, the following equation can be derived:

$$\frac{1}{\sqrt{2\pi\tilde{\delta}_t}}e^{-\frac{\left(\boldsymbol{x}_{t-1}-\tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t,\boldsymbol{x}_0,\boldsymbol{y})\right)^2}{2\tilde{\delta}_t}} = \frac{\frac{1}{\sqrt{2\pi\delta_{t|t-1}}}e^{-\frac{\left(\boldsymbol{x}_t-\left(\frac{1-m_t}{1-m_{t-1}}\boldsymbol{x}_{t-1}+(m_t-\frac{1-m_t}{1-m_{t-1}}m_{t-1})\boldsymbol{y}\right)\right)^2}{2\delta_{t|t-1}}}\frac{1}{\sqrt{2\pi\delta_{t-1}}}e^{-\frac{\left(\boldsymbol{x}_{t-1}-\left((1-m_{t-1})\boldsymbol{x}_0+m_{t-1}\boldsymbol{y}\right)\right)^2}{2\delta_{t-1}}}}{\frac{1}{\sqrt{2\pi\delta_t}}e^{-\frac{\left(\boldsymbol{x}_t-\left((1-m_t)\boldsymbol{x}_0+m_t\boldsymbol{y}\right)\right)^2}{2\delta_t}}}$$

$$= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\delta_t}{\delta_{t|t-1}\delta_{t-1}}}e^{-\frac{1}{2\frac{\delta_{t|t-1}\delta_{t-1}}{\delta_t}}\left(\boldsymbol{x}_{t-1}-\left(\frac{\delta_{t-1}}{\delta_t}\frac{1-m_t}{1-m_{t-1}}\boldsymbol{x}_t+(1-m_{t-1})\frac{\delta_{t|t-1}}{\delta_t}\boldsymbol{x}_0+(m_{t-1}-m_t\frac{1-m_t}{1-m_{t-1}}\frac{\delta_{t-1}}{\delta_t})\boldsymbol{y}\right)\right)^2}$$

Then we can have the following equations:

$$\tilde{\boldsymbol{\mu}}_t(\boldsymbol{x}_t,\boldsymbol{x}_0,\boldsymbol{y}) = \frac{\delta_{t-1}}{\delta_t}\frac{1-m_t}{1-m_{t-1}}\boldsymbol{x}_t + (1-m_{t-1})\frac{\delta_{t|t-1}}{\delta_t}\boldsymbol{x}_0 + (m_{t-1}-m_t\frac{1-m_t}{1-m_{t-1}}\frac{\delta_{t-1}}{\delta_t})\boldsymbol{y}$$

$$\tilde{\delta}_t = \frac{\delta_{t|t-1}\cdot\delta_{t-1}}{\delta_t}$$

which is equivalent to Eq.(12) and Eq.(13) in the paper.

2. Implementation Details

In this section, we provide more implementation details of BBDM, including network hyperparameters and optimization, details of training and sampling procedures.

Network hyperparameters. As mentioned in Section 4.1, we adopt the same VQGAN model and network architecture as LDM for fair comparison. In order to enable the model to be trained on a single GeForce GTX 3090 GPU, we reduced model size by modifying the total number of middle layers and channels of middle features. The network details are shown in Table 1.

model	z-shape	channel multiplier	attention resolutions	channels	total parameters	trainable parameters
BBDM-f4	$64 \times 64 \times 3$	1,4,8	32,16,8	128	292.42M	237.09M
BBDM-f8	$32 \times 32 \times 4$	1,4,8	32,16,8	128	304.81M	237.10M
BBDM-f16	$16 \times 16 \times 8$	1,4,8	16,8,4	128	327.71M	258.11M

Table 1. Network hyperparameters for both BBDM and LDM used in this paper.

Training and sampling details. In order to improve the performance of BBDM, Exponential Moving Average (EMA) was adopted in the training procedure together with ReduceLROnPlateau learning rate scheduler. The EMA and learning rate scheduler hyperparameters are reported in Tables 2 and 3.

model	EMA start step	EMA decay	EMA update interval	batch size
BBDM-f4	30000	0.995	16	8
BBDM-f8	30000	0.995	16	8
BBDM-f16	15000	0.995	8	16

Table 2. EMA hyperparameters of BBDM.

3. User Study

An additional subjective user study is designed to evaluate the performance of the proposed method against three methods with comparable FID measurement, including CDE, LDM and OASIS. 12 groups of samples are randomly selected from

model	max learning rate	min learning rate	factor	patience	cool down	threshold
BBDM-f4	1.0e-4	5.0e-7	0.5	3000	2000	1.0e-4
BBDM-f8	1.0e-4	5.0e-7	0.5	3000	2000	1.0e-4
BBDM-f16	1.0e-4	1.0e-6	0.5	3000	2000	1.0e-4

Table 3. Learning rate scheduler hyperparameters of BBDM.

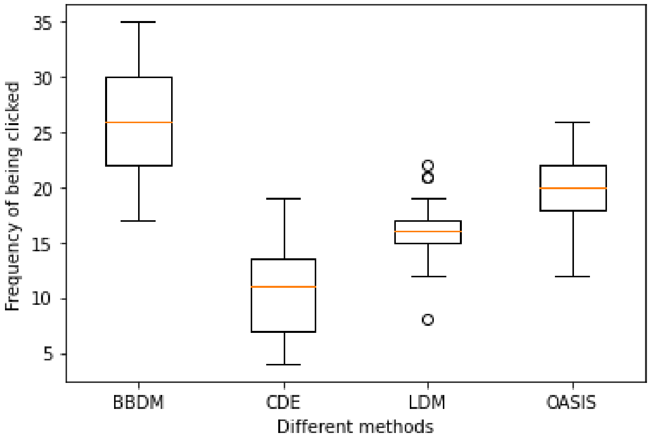

Figure 1. User study results of BBDM, CDE, LDM, OASIS on CelebAMask-HQ dataset.

CelebAMask-HQ experiments. For each sample, a pair of editing results are randomly shown to the participants. As there are 4 different editing results for each image, 72 clicks are required for each participant. 112 users with age between 20 and 50 were invited to participate in the user study. The distribution of user preference is shown in Figure 1. We can see that more users prefer the results of the proposed method.

4. Additional Qualitative Results

Finally, we provide additional qualitative results compared with other competitive methods (Figures 2, 4). More diverse samples are shown in Figures 3 and 5. Other experiment results on inpainting, colorization and face-to-label tasks can be found in Figure 6.

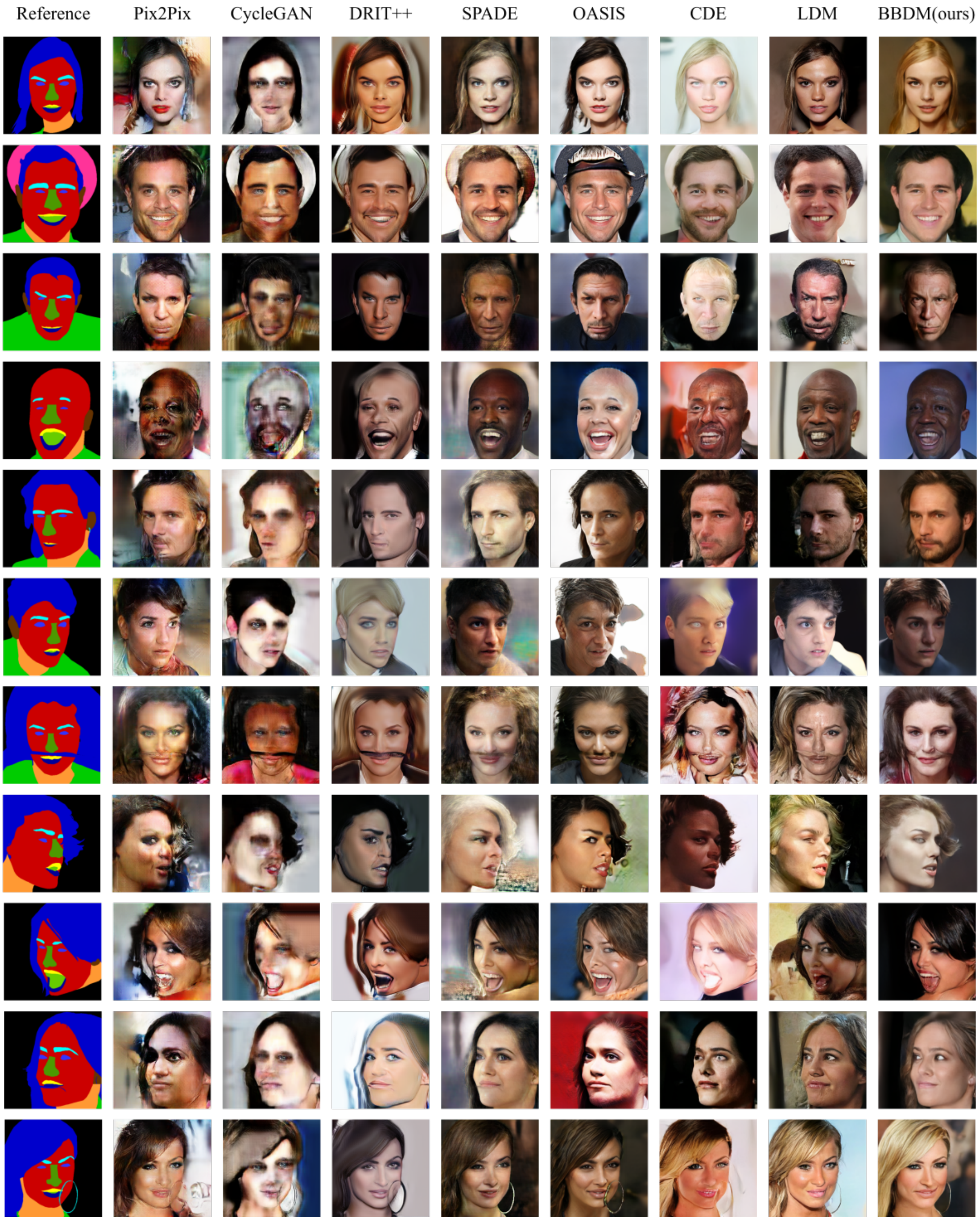

Figure 2. More qualitative results on the CelebAMask-HQ dataset.

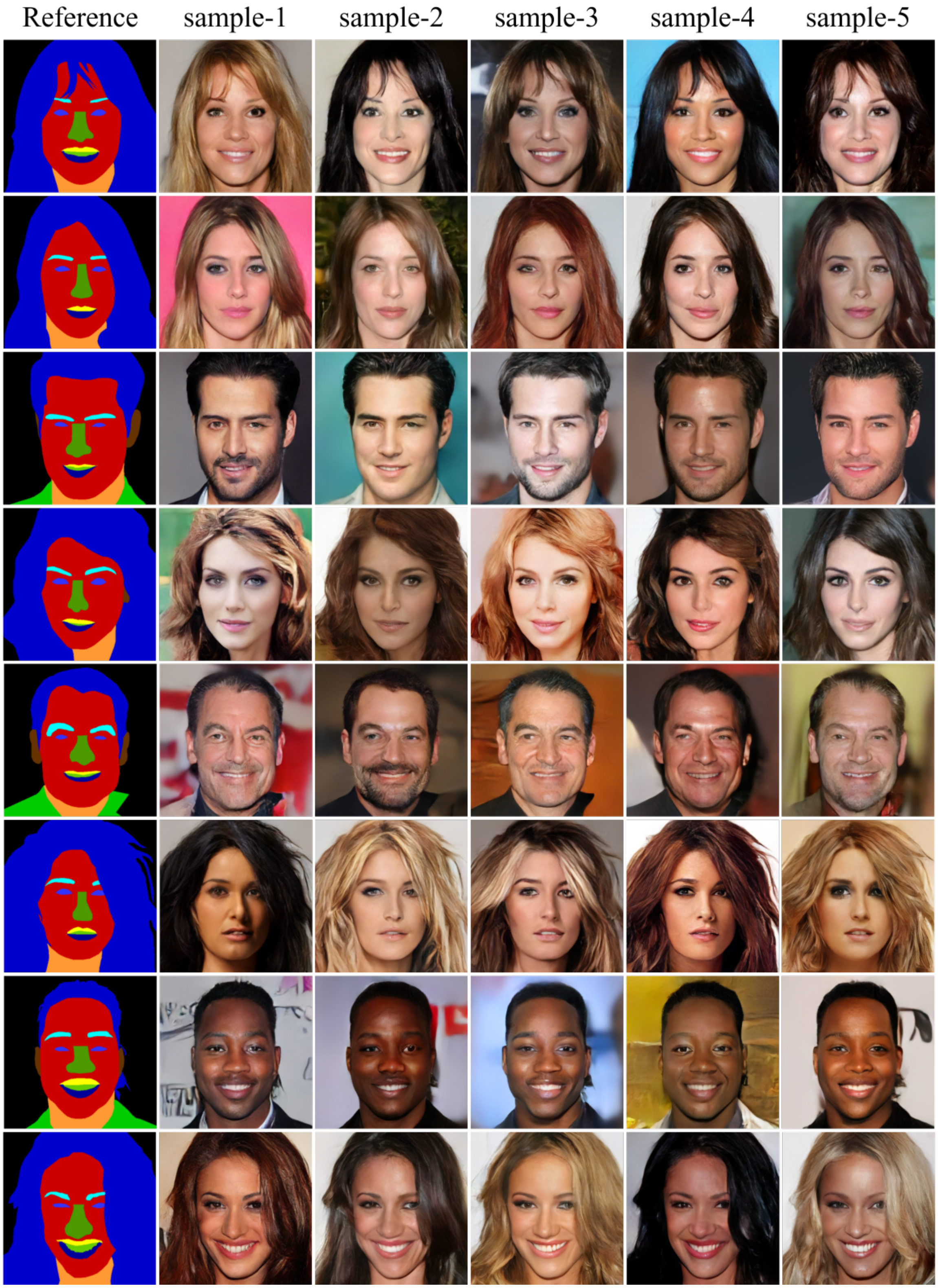

Figure 3. More diverse samples on the CelebAMask-HQ dataset.

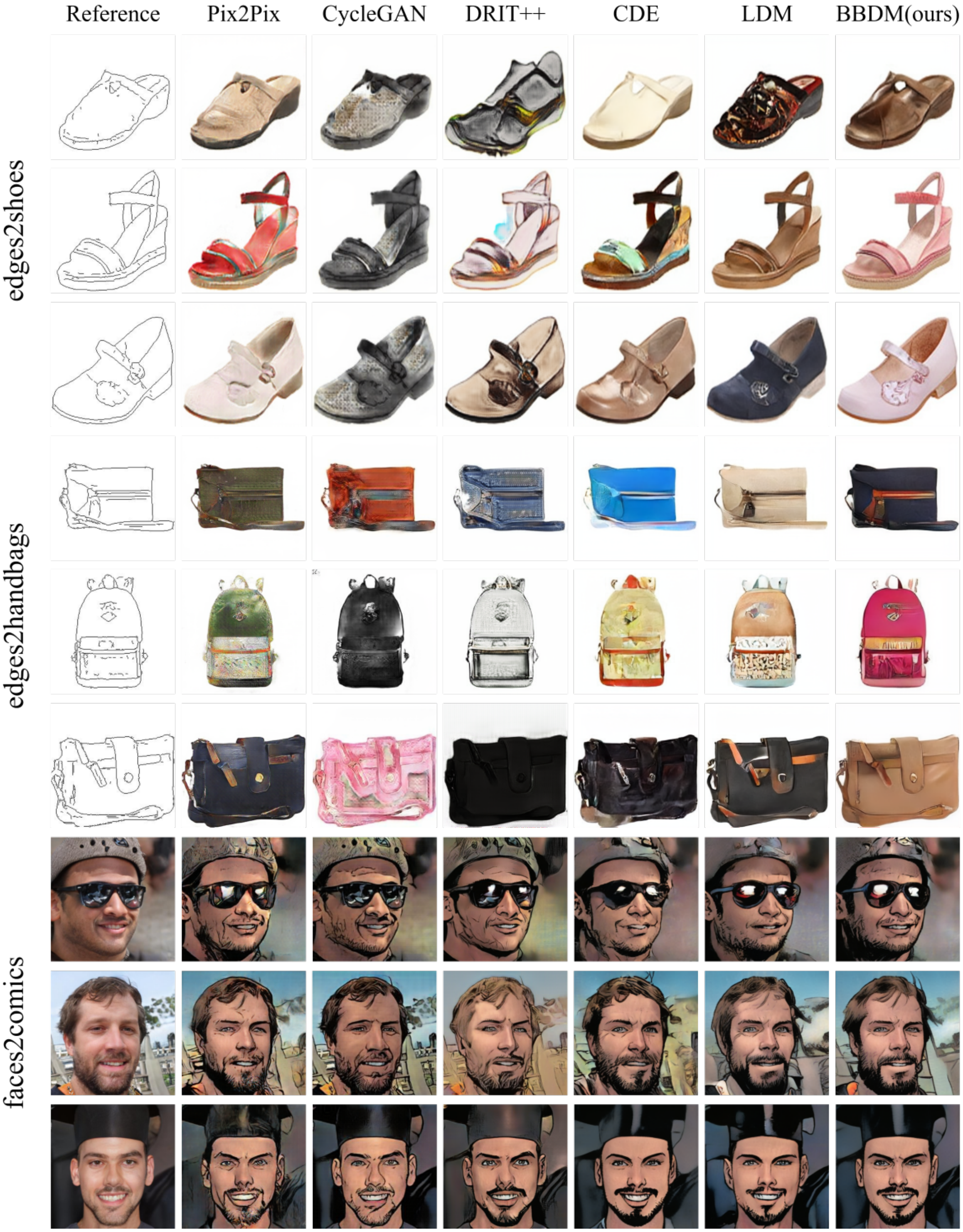

Figure 4. More qualitative results on the edges2shoes, edges2handbags and faces2comics datasets.

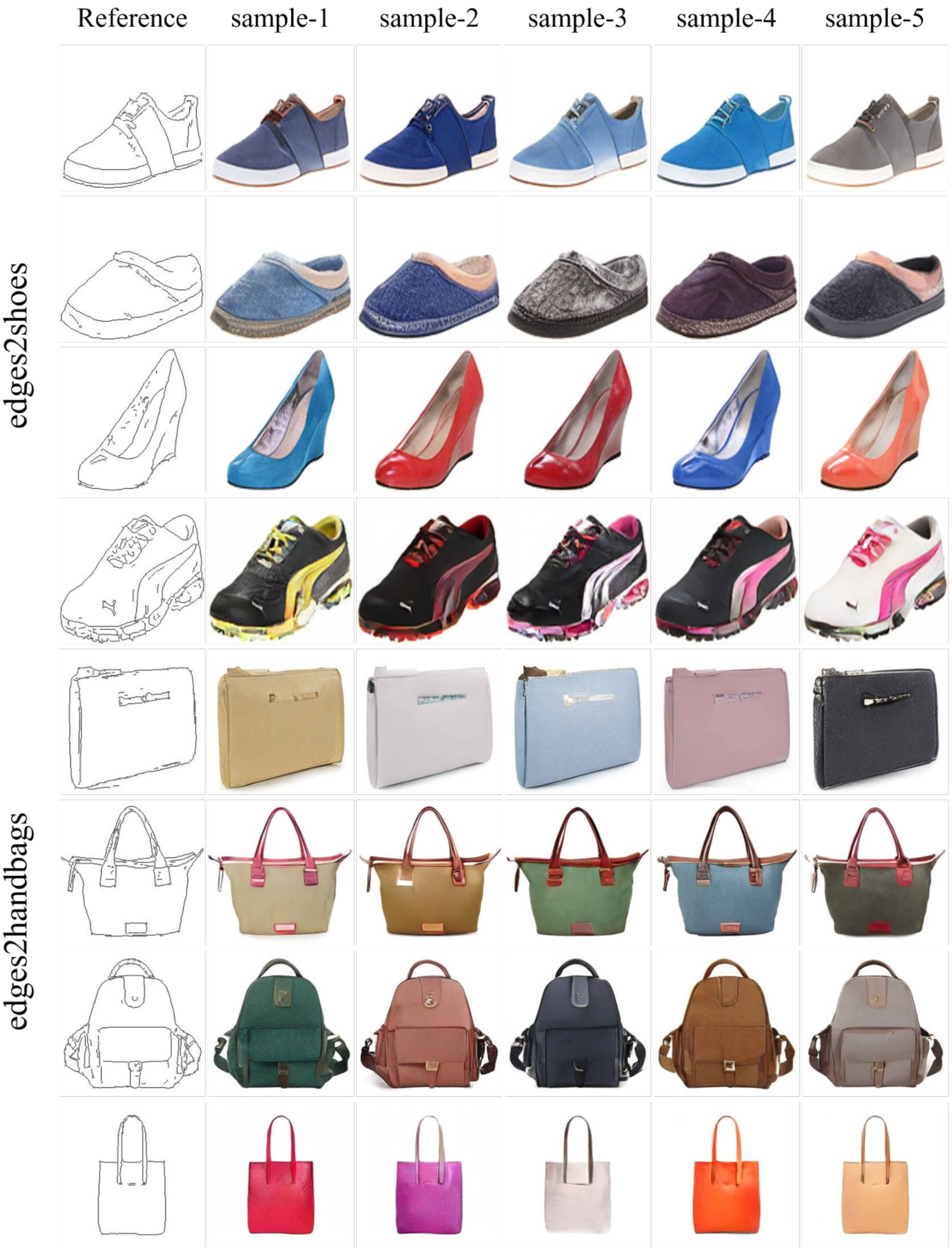

Figure 5. More diverse samples on the edges2shoes, edges2handbags datasets.

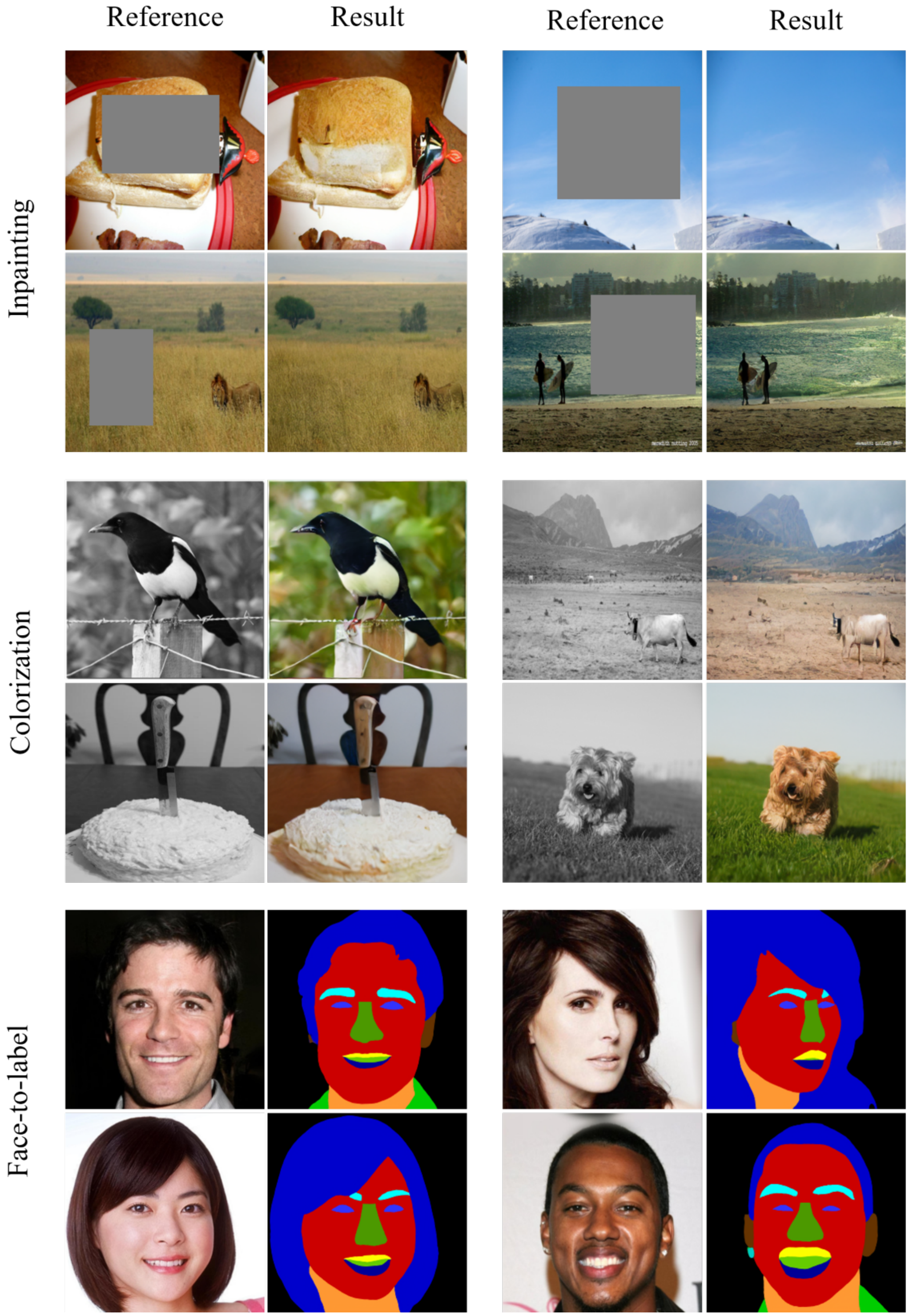

Figure 6. More inpainting, colorization and face-to-label examples generated by our method.