Recursive Cross-Domain Facial Composite and Generation from Limited Facial Parts

Yang Song [1] **Zhifei Zhang** [1] **Hairong Qi** [1]

Abstract

We start by asking an interesting yet challenging question, "If a large proportion (e.g., more than 90%) of the face/sketch is missing, can a realistic whole face sketch/image still be estimated?" Existing face completion and generation methods either do not conduct multi-domain transfer or cannot handle large missing area. For example, the inpainting approach tends to blur the generated region when the missing area is large (i.e., more than 50%). In this paper, we exploit the potential of deep learning networks in filling large missing region (e.g., as high as 95% missing) and generating realistic faces in cross domains by given patches with very limited information. We propose the recursive generation by bidirectional transformation networks (r-BTN) that recursively generates a whole face/sketch from small sketch/face patches. The large missing area and the cross domain challenge make it difficult to generate satisfactory results using a unidirectional cross-domain learning structure. On the other hand, a forward and backward bidirectional learning between the face and sketch domains would enable recursive estimation of the missing region in an incremental manner (Fig. 1) and yield appealing results. Extensive experiments have been conducted to demonstrate the superior performance from r-BTN as compared to existing potential solutions.

1. Introduction

A common scenario in law enforcement is that sketches of suspects are drawn by forensic artists or created by computer software (i.e., facial composite) based on descriptions provided by eyewitnesses or victims (Klum et al., 2013). However, the forgetting process (Ouyang et al., 2016) may cause the eyewitness to have strong impression only on some key part of the face. In these cases, the whole face sketch will have to be created (i.e., imagined) by the forensic artist. In addition, some facial composite methods (e.g., Photofit (Photofit)) synthesize faces by stitching patches from multiple domains which deteriorates the consistency and photo-reality. How to generate realistic faces/sketches that are consistent to the given sketch/face patches is still a challenging task because large missing area could lead to blurry generated images, and cross-domain filling could further deteriorate the quality of generated faces, especially when transferring from information-scarce to information-rich domains (e.g., from sketch to face). To the best of our knowledge, this work represents the first attempt to cross-filling large missing area in both face and sketch domains, and compositing realistic faces conditioned on small facial parts from multiple domains and/or subjects (See Fig. 5). Existing works that may potentially address this problem are mainly in the perspectives of face/sketch synthesis/transformation and image inpainting. The face/sketch synthesis works (Wang & Tang, 2009; Tang & Wang, 2003; Zhou et al., 2012; Song et al., 2014) synthesize target faces from the source domain through patch-wise searching of similar patches in the training set. Without the generative capability, these methods fail to render reasonable pixels for large missing areas. The generative adversarial networks (GANs) (Goodfellow et al., 2014) has shown impressive performance in face generation (Radford et al.,

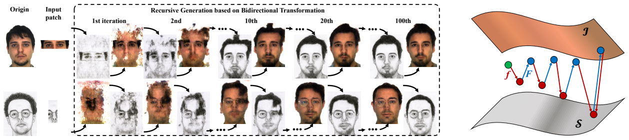


Figure 1. Examples of recursive generation from small face or sketch patches by the bidirectional transformation network. Left: Original face/sketch and the corresponding input patches extracted from them. Inside of the dashed box demonstrates the generated face/sketch at different iteration steps. Right: Illustration of transformation between the face and sketch manifolds $\mathcal{I}$ and $\mathcal{S}$, respectively. The green dot denotes a given face patch. The red and blue arrows are the learned mapping f and F, respectively. The red and blue dots are generated sketches and faces through corresponding mapping.

*Equal contribution [1]University of Tennessee, Knoxville, TN USA. Correspondence to: Yang Song <ysong18@vols.utk.edu>.

ICML 2017 Workshop on Implicit Models. Copyright 2017 by the author(s).

2015; Zhang et al., 2017), domain transformation (Zhu et al., 2016; Isola et al., 2016), and inpainting (Yeh et al., 2016; Pathak et al., 2016). However, generating faces from small patches in either single or cross domains has not been explored.

In this paper, we investigate the problem of cross-domain face/sketch composition/generation conditioned on small patches of sketch/face. We assume that faces and sketches lie on high-dimensional manifolds $\mathcal{I}$ and $\mathcal{S}$, respectively, as shown in Fig. 1 (right). The given small sketch/face patch will initially deviate from the corresponding manifold due to large amount of missing data. With the learned bidirectional transformation network (BTN), i.e., f and F, the given patch will be recursively mapped forward and backward between $\mathcal{I}$ and $\mathcal{S}$. Each mapping will yield a result progressively closing in onto either the face or sketch manifold, and eventually approaching the real whole face/sketch images as shown in Fig. 1 (left). An adversarial network is imposed on both f and F, forcing more photo-realistic faces/sketches. The rationale and benefit of the proposed r-BTN will be further discussed in section 2.

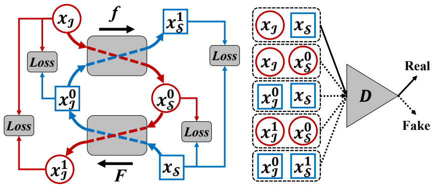
This paper makes the following contributions: 1) We tackle the challenging problem of face/sketch generation from small patches, estimating large missing area based on limited information while alleviating the blur effect suffered by existing works. 2) We propose the recursive generation by bidirectional transformation networks (r-BTN), which learns both a forward and backward mapping function between cross domains to enable a recursive update of the generated faces/sketches for more consistent and high-fidelity results even with large portions of missing data. 3) We further exploit the capacity of r-BTN in fusing multiple patches from multiple domains and multiple people to output a realistic and consistent face in a generative manner. 4) In the area of generative imaging, there is a lack of quantitative evaluation of image reality. We design relative discrimination score (RDS) for effective evaluation of image reality.

2. The Recursive Bidirectional Transformation Network

Training Stage: Fig. 2 illustrates the BTN structure where the mapping functions, f and F, are learned in a bidirectional fashion.

Given the original face/sketch pair $x_{\mathcal{I}}$ and $x_{\mathcal{S}}$, the following transformations are performed,

$$x_{\mathcal{S}}^0 = f(x_{\mathcal{I}}),\ x_{\mathcal{I}}^1 = F(x_{\mathcal{S}}^0) = F(f(x_{\mathcal{I}})),$$
$$x_{\mathcal{I}}^0 = F(x_{\mathcal{S}}),\ x_{\mathcal{S}}^1 = f(x_{\mathcal{I}}^0) = f(F(x_{\mathcal{S}})).$$



$x_{\mathcal{I}}$
f
$x_{\mathcal{S}}^1$
Loss
Loss
$x_{\mathcal{I}}^0$
$x_{\mathcal{S}}^0$
Loss
Loss
$x_{\mathcal{I}}^1$
F
$x_{\mathcal{S}}$
$x_{\mathcal{I}}$ $x_{\mathcal{S}}$
$x_{\mathcal{I}}$ $x_{\mathcal{S}}^0$
$x_{\mathcal{I}}^0$ $x_{\mathcal{S}}$
$x_{\mathcal{I}}^1$ $x_{\mathcal{S}}^0$
$x_{\mathcal{I}}^0$ $x_{\mathcal{S}}^1$
D
Real
Fake

Figure 2. Training flow of the bidirectional transformation network. $x_{\mathcal{I}}$ and $x_{\mathcal{S}}$ are the real face/sketch pair. Red and blue arrows denote the transformation paths of $x_{\mathcal{I}}$ and $x_{\mathcal{S}}$, respectively. The transformation functions f and F could be encoder-decoder networks. $Loss$ denotes the ℓ_1-norm. The discriminator D is trained on real and generated (fake) face/sketch pairs.

The objective is to learn the bidirectional transformations between $\mathcal{I}$ and $\mathcal{S}$, so that any face/sketch pair could be uniquely mapped forward and backward into another domain. Note that in training stage, all pairs are whole face instead of patches. To preserve the identity of face and sketch during transformations, we minimize the reconstruction error $\mathcal{L}_{rec}$ between real and generated faces or sketches as Eq. 1.

$$\mathcal{L}_{rec} = \sum_{i=0}^{1} \left(\|x_{\mathcal{I}} - x_{\mathcal{I}}^i\|_1 + \|x_{\mathcal{S}} - x_{\mathcal{S}}^i\|_1 \right), \qquad (1)$$

where the ℓ_1-norm instead of the ℓ_2-norm is used to avoid blurry results. Besides $\mathcal{L}_{rec}$, an adversarial constraint is employed to encourage photo-realistic face/sketch pairs. The discrimination loss can be written as

$$\mathcal{L}_{adv} = \mathbb{E}_{\omega\in\Omega}\left[\log D(\omega)\right] - \mathbb{E}_{\substack{x_{\mathcal{I}}\in\mathcal{I}\\ x_{\mathcal{S}}\in\mathcal{S}}}[\log D(x_{\mathcal{I}}, x_{\mathcal{S}})], \qquad (2)$$

where

$$\begin{aligned}\Omega &= \left\{(x_{\mathcal{I}}, x_{\mathcal{S}}^0)_j, (x_{\mathcal{I}}^1, x_{\mathcal{S}}^0)_j, (x_{\mathcal{I}}^0, x_{\mathcal{S}})_j, (x_{\mathcal{I}}^0, x_{\mathcal{S}}^1)_j\right\}\\ &= \left\{(x_{\mathcal{I}}, f(x_{\mathcal{I}}))_j, (F(f(x_{\mathcal{I}}), f(x_{\mathcal{I}}))_j,\right.\\ &\qquad \left.(F(x_{\mathcal{S}}), x_{\mathcal{S}})_j, (F(x_{\mathcal{S}}), f(F(x_{\mathcal{S}}))_j\right\}\end{aligned}$$

indicates the fake face/sketch pairs, and j indexes the fake pairs generated from the jth real pair in a mini-batch. Note that only $(x_{\mathcal{I}}, x_{\mathcal{S}})$ is the real pair. Combining Eqs. 1 and 2, the objective function is

$$\min_{f,F,D}\ \mathcal{L}_{adv} + \lambda\mathcal{L}_{rec}, \qquad (3)$$

where λ balances the adversarial loss and reconstruction loss. In optimization, f, F, and D are updated alternatively. The discriminator D is updated by minimizing $\mathcal{L}_{adv}$. The update of f and F is performed by

$$\min_{f}\ \mathbb{E}_{\omega\in\Omega_f}\left[\log D(\omega)\right] + \lambda\sum_{i=0}^{1}\|x_{\mathcal{S}} - x_{\mathcal{S}}^i\|_1, \qquad (4)$$

$$\min_{F}\ \mathbb{E}_{\omega\in\Omega_F}\left[\log D(\omega)\right] + \lambda\sum_{i=0}^{1}\|x_{\mathcal{I}} - x_{\mathcal{I}}^i\|_1, \qquad (5)$$

where

$$\begin{aligned}\Omega_f &= \{(x_\mathcal{I}, x_\mathcal{S}^0)_j, (x_\mathcal{I}^0, x_\mathcal{S}^1)_j\} \\ &= \{(x_\mathcal{I}, f(x_\mathcal{I}))_j, (x_\mathcal{I}^0, f(x_\mathcal{I}^0))_j\}, \\ \Omega_F &= \{(x_\mathcal{I}^0, x_\mathcal{S})_j, (x_\mathcal{I}^1, x_\mathcal{S}^0)_j\} \\ &= \{(F(x_\mathcal{S}), x_\mathcal{S})_j, (F(x_\mathcal{S}^0), x_\mathcal{S}^0)_j\},\end{aligned}$$

and $\Omega = \Omega_f \cup \Omega_F$. Here, j is again the index of training samples in a mini-batch.

Testing Stage: During testing, given an arbitrary patch from either domain, a whole face from the other domain could be generated in a "recursive" manner through the bidirectional transformation. The testing flow is shown in Fig. 3, which demonstrates the case of given a face patch $p_\mathcal{I}$. Similarly, if a sketch patch $p_\mathcal{S}$ is given, it will be fed to $x_\mathcal{S}$ and similar testing flow can be carried out to generate a whole face image. Given a small patch, the testing stage needs multiple iterations to gradually generate a whole face/sketch with the fixed given patch, as illustrated previously in Fig. 1. In each iteration, backpropagating the loss of D will enforce the photo-reality during the recursive generation. Repeating this procedure, the large missing area can be filled up gradually which is consistent with the given patch.

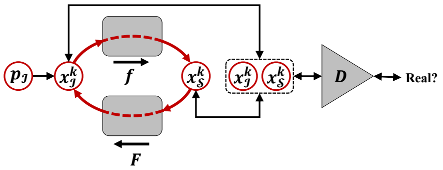


Figure 3. Testing flow of r-BTN, assuming a face patch $p_\mathcal{I}$ as the input. At step k, the generated face is $x_\mathcal{I}^k$. Replacing the corresponding area of $x_\mathcal{I}^k$ by the patch $p_\mathcal{I}$ and transforming $x_\mathcal{I}^k$ to $x_\mathcal{S}^k$, we get a face/sketch pair $\left(x_\mathcal{I}^k, x_\mathcal{S}^k\right)$. Then, this pair is adjusted by the error back propagated from D as comparing to the output of real pairs. Finally, $x_\mathcal{S}^k$ is transformed back to the face domain, generating $x_\mathcal{I}^{k+1}$ for the next iteration.

3. Experiment and Results

3.1. Implementation Details

We collect face/sketch pairs from the datasets CUHK (Wang & Tang, 2009), CUFSF (Zhang et al., 2011), AR (Martinez & Benavente, 2007), FERET (Phillips et al., 2000), and IIIT-D (Bhatt et al., 2012). All the face/sketch images are cropped and well-aligned based on the eye locations, and preprocessed to be uniform white background. The transformations f and F are implemented by the Conv-Deconv network. Details are shown in supplementary materials. After 100 epochs, we could achieve the results as shown in this paper.

3.2. Qualitative Evaluation

Comparison with Other Methods We compare the proposed r-BTN with Pix2Pix (Isola et al., 2016) and image inpainting (Pathak et al., 2016). The inpainting method compared in this paper is modified from (Pathak et al., 2016) to achieve cross-domain inpainting. Specifically, the inputs are faces/sketches with random mask (20%∼80% masked), and the outputs are the whole sketch/face. Pix2Pix and r-BTN are trained with the whole face/sketch pairs. All methods are trained on the same training dataset with the same parameter setting. The comparison results are shown in Fig. 4. The Pix2Pix and inpainting methods train face-sketch and sketch-face transformation networks independently, which is why the identity between generated sketches and faces cannot be preserved. For example, comparing the two rows labeled with "inpainting", especially the 4th-6th columns, the sketches seem female while the faces appear like male. In addition, the inpainting results present apparent discontinuity between the given patch and the estimated area. On the other hand, the results from r-BTN demonstrate higher fidelity, better consistency to given patches, and better identity preservation. Note that the main objective of this work is to generate more photo-realistic and consistent face/sketch images. The identity of the person can be preserved when the missing percentage is less than 70%. When the missing percentage is higher, the algorithm still generates photo-realistic results but the identity is not preserved. More results can be found in the supplementary materials.

Generation from Multiple Patches We explore the r-BTN to generate consistent and realistic faces from multiple patches that may be from two domains and multiple people. Examples generated from multiple patches are shown in Fig. 5, demonstrating the great versatility of r-BTN. We again observe the strong consistency and fidelity between the generated face/sketch pairs.

3.3. Quantitative Evaluation

To numerically evaluate the quality of generated faces, we design the relative discrimination score (RDS), which aims to estimate the photo-reality of generated faces. We train a discriminator to distinguish between real and generated faces. With more epochs, the discriminator output from real faces would be close to one, and that from generated faces should approach zero. If the generated faces are realistic, their discriminator output would be relatively higher and decrease slower with epochs as compared to that of unrealistic faces. Fig. 6 shows the discriminator output of each method during training the discriminator.

RDS computes the ratio of area under the curves of generated faces from certain method and real faces. Higher RDS indicates more photo-realistic faces. During train-

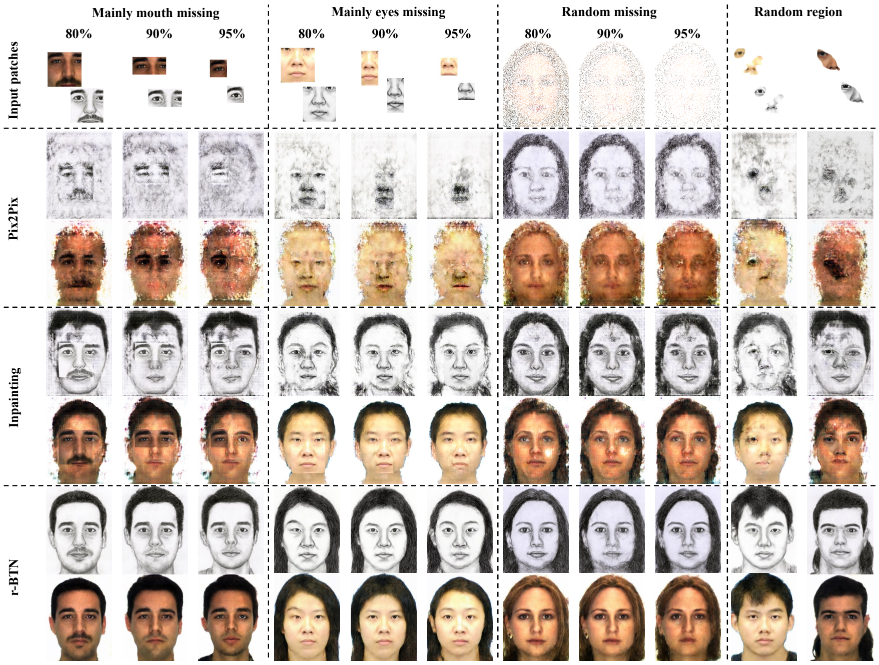


Figure 4. Comparison with other potential methods for filling large missing areas. The first row shows the input patches, and the rest rows display the results from different methods. The percentage indicates missing proportion (missing area over image area). Because Pix2Pix is for domain transfer rather than missing area filling, its results cannot compete with inpainting or r-BTN. We show them here to provide the baseline of domain transfer methods in filling large missing areas.

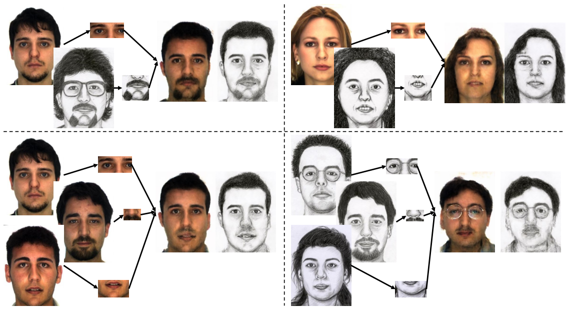


Figure 5. Examples of generated faces/sketches from multiple patches, which are from different people and/or different domains. Four examples are displayed in a 2-by-2 matrix. In each cell, the original faces and sketches are given on the left. The patches are extracted from where indicated by the arrows. The right are generated face/sketch pairs.

ing the discriminator, 900 real faces are randomly selected from the training set, and 300 faces are generated from each method. Thus, we have balanced real and generated samples. The area under a curve is computed by trapezoidal numerical integration. We observe that the RDS level of r-BTN is much higher than that of inpainting, showing more photo-realistic outputs generated from r-BTN.

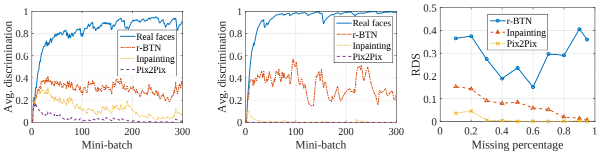


Figure 6. Averaged discriminator output at each mini-batch (30 samples) during training the discriminator that aims to distinguish real and generated faces from Pix2Pix, inpainting, and r-BTN, respectively. Left and middle are the averaged discriminator outputs by given random patches with 10% (left) and 95% (middle) missing. Right is the RDS with different missing percentage.

4. Discussion and Future works

This paper proposed and solved the challenging task of cross-domain face generation with large missing area. A novel recursive generation method by bidirectional transformation networks (r-BTN) was proposed to generate high-fidelity and consistent face/sketch even with missing area as large as 95%. The consistency is achieved by the bidirectional network while the idea of “recursive” generation demonstrated the potential capability on high-quality image generation. We also demonstrated the effectiveness of r-BTN by comparing with some potential solutions like pix2pix and inpainting. In the future, we plan to improve r-BTN to be more robust to faces/sketch misalignment.

References

Bhatt, Himanshu S, Bharadwaj, Samarth, Singh, Richa, and Vatsa, Mayank. Memetically optimized MCWLD for matching sketches with digital face images. *IEEE Transactions on Information Forensics and Security*, 7 (5):1522–1535, 2012.

Goodfellow, Ian, Pouget-Abadie, Jean, Mirza, Mehdi, Xu, Bing, Warde-Farley, David, Ozair, Sherjil, Courville, Aaron, and Bengio, Yoshua. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pp. 2672–2680, 2014.

Isola, Phillip, Zhu, Jun-Yan, Zhou, Tinghui, and Efros, Alexei A. Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:1611.07004*, 2016.

Klum, Scott, Han, Hu, Jain, Anil K, and Klare, Brendan. Sketch based face recognition: Forensic vs. composite sketches. In *International Conference on Biometrics*, pp. 1–8. IEEE, 2013.

Martinez, AR and Benavente, R. The AR face database, 1998. *Computer Vision Center, Technical Report*, 3, 2007.

Ouyang, Shuxin, Hospedales, Timothy M, Song, Yi-Zhe, and Li, Xueming. ForgetMeNot: memory-aware forensic facial sketch matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5571–5579, 2016.

Pathak, Deepak, Krahenbuhl, Philipp, Donahue, Jeff, Darrell, Trevor, and Efros, Alexei A. Context encoders: Feature learning by inpainting. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2536–2544, 2016.

Phillips, P Jonathon, Moon, Hyeonjoon, Rizvi, Syed A, and Rauss, Patrick J. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10): 1090–1104, 2000.

Photofit. Photofit. http://www.open.edu/openlearn/body-mind/photofit-me. [Online].

Radford, Alec, Metz, Luke, and Chintala, Soumith. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

Song, Yibing, Bao, Linchao, Yang, Qingxiong, and Yang, Ming-Hsuan. Real-time exemplar-based face sketch synthesis. In *European Conference on Computer Vision*, pp. 800–813. Springer, 2014.

Tang, Xiaoou and Wang, Xiaogang. Face sketch synthesis and recognition. In *IEEE International Conference on Computer Vision*, pp. 687–694. IEEE, 2003.

Wang, Xiaogang and Tang, Xiaoou. Face photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):1955–1967, 2009.

Yeh, Raymond, Chen, Chen, Lim, Teck Yian, Hasegawa-Johnson, Mark, and Do, Minh N. Semantic image inpainting with perceptual and contextual losses. *arXiv preprint arXiv:1607.07539*, 2016.

Zhang, Wei, Wang, Xiaogang, and Tang, Xiaoou. Coupled information-theoretic encoding for face photo-sketch recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 513–520. IEEE, 2011.

Zhang, Zhifei, Song, Yang, and Qi, Hairong. Age progression/regression by conditional adversarial autoencoder. *arXiv preprint arXiv:1702.08423*, 2017.

Zhou, Hao, Kuang, Zhanghui, and Wong, Kwan-Yee K. Markov weight fields for face sketch synthesis. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1091–1097. IEEE, 2012.

Zhu, Jun-Yan, Krähenbühl, Philipp, Shechtman, Eli, and Efros, Alexei A. Generative visual manipulation on the natural image manifold. In *European Conference on Computer Vision*, pp. 597–613. Springer, 2016.

Appendix

Network Structure

All the face/sketch images are cropped and well-aligned based on the eye locations, and preprocessed to be uniform white background. The transformations f and F are implemented by the Conv-Deconv network as shown in Table 1.

Table 1. Network structure used for transformation

Conv. (LeakyReLU)	Deconv. (ReLU)
$256^2 \times 3,\ 128^2 \times 64,$	$2^2 \times 1024,\ 4^2 \times 1024,$
$64^2 \times 128,\ 32^2 \times 256,$	$8^2 \times 1024,\ 16^2 \times 512,$
$16^2 \times 512,\ 8^2 \times 1024,$	$32^2 \times 256,\ 64^2 \times 128,$
$4^2 \times 1024,\ 2^2 \times 1024$	$128^2 \times 64,\ 256^2 \times 3$

Similarity/Diversity Analysis

Intuitively speaking, the generated faces from the patches of the same person should be similar. By contrast, patches from different persons are supposed to yield diverse faces. To verify this property, we collect 50 faces and pick patches of different size around the eyes, the nose, and the mouth. The proposed r-BTN is then applied to generate full faces from those patches. To measure the similarity/diversity between generated faces, we utilize the pre-trained VGG-Face model to extract high-level features and compute their Euclidean distance. We perform two comparisons: 1) self comparison (similarity) and 2) mutual comparison (diversity), conducting on faces generated from patches of the same and different persons, respectively.

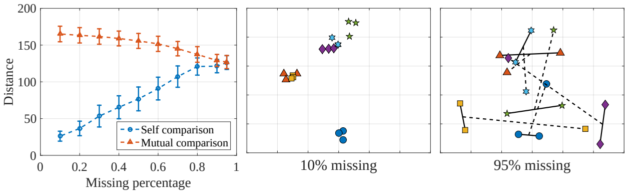


Figure 7. Left: Evaluation of similarity/diversity with increasing missing percentage. Circles/triangles are averaged distances of self comparison and mutual comparison, respectively. The bars indicate corresponding standard deviation. Middle and right: High-level feature (reduced to 2-D by PCA) of generated faces at missing percentage of 10% and 95%, respectively. Different marker types indicated different persons. There are three same markers for type (person), denoting the generated faces from patches around left eye, right eye, and mouth. In the right figure, solid lines connect the faces generated from eyes, and the dashed lines connect to the faces generated from mouth.

Fig. 7 (left) shows the averaged distance and standard deviation with respect to missing percentage. The blue circles shows the results of self comparison, and the red triangles denote mutual comparison.

With lower missing percentage, e.g., 0.1 to 0.6, the generated faces preserve relatively high intra-class (same person) similarity and inter-class (different persons) diversity. As the missing percentage increases, the two curves eventually intersect, indicating the generated faces from very small patches (e.g., 95% missing) have lost the identity of the original face. Interestingly, we discover that the generated faces from either the left or right eye of the same person still tend to be more similar as compared to those generated from nose/mouth as illustrated in Fig. 7 (right). This discovery is well in line with the quality of different biometrics where studies have shown eyes to carry more valuable cues than nose or mouth in face recognition tasks. This finding, from another perspective, demonstrates the high effectiveness of r-BTN in generating high-fidelity and realistic faces/sketches. Fig. 8 visualizes such similarity.

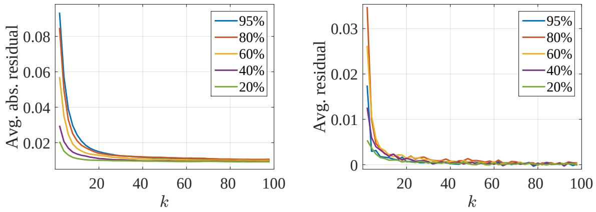


Figure 9. Convergence evaluation of the proposed r-BTN. Averaged absolute (left) and average (right) of residual with respect to iteration k are shown at missing percentage of 95%, 80%, 60%, 40%, and 20%, respectively.

Convergence Analysis

Will the generated faces/sketches converge to a certain point? How many iterations are sufficient to achieve a photo-realistic result? This section mainly answers these two questions.

We first define the residual in the face domain between subsequent iterations as $r^{k+1} = \left(x_{\mathcal{I}}^{k+1} - x_{\mathcal{I}}^{k}\right)$, where $x_{\mathcal{I}}^{k}$ and $x_{\mathcal{I}}^{k+1}$ denote the kth and k+1th generated results. The convergence is mainly evaluated by calculating the averaged residual on testing samples (i.e., 300 samples generated with different missing percentage) with respect to k as shown in Fig. 9 (right). However, the average residual is not sufficient to demonstrate the convergence because some pixels may significantly increase while the other decrease with the same level. In this case, we calculate the averaged absolute residual which illustrate the changing amplitude as shown in Fig. 9 (left).

With more iterations, the averaged residual approaches zero while the averaged absolute residual stabilizes at a small value. This well demonstrates that the generated

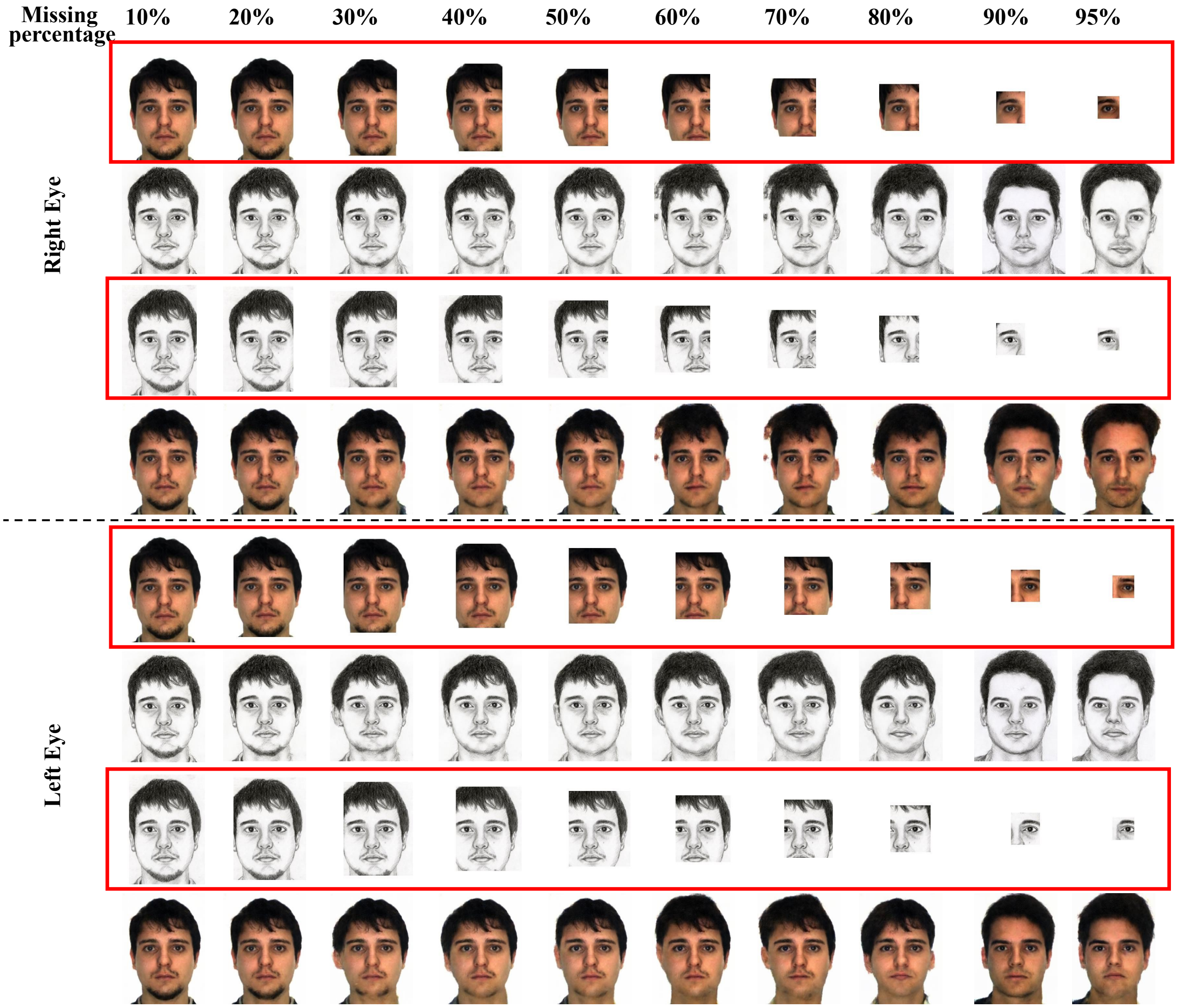


Figure 8. Given face/sketch patches (red boxes) from the same person but with different missing percentage, the proposed r-BTN generates similar sketches/faces. Please note that the given patches are mainly from one of the eyes.

faces are stable. In addition, from the experiments, the generated faces/sketches will not significantly change after 20 iterations. Therefore, we could empirically conclude that the recursive generation will converge to certain face/sketch for a given patch.

More Qualitative Results from r-BTN

Fig. 10 displays more results generated from eyes, nose, mouth, and random regions using the proposed r-BTN. In addition, we provide more quantitative results of generated faces from three methods — Pix2Pix, inpainting, and r-BTN. Fig. 11 and 12 visualize the comparison through two examples. The proposed r-BTN generates higher fidelity and more smooth results. However, the proposed method cannot preserve the identity when the missing percentage is more than 70%.

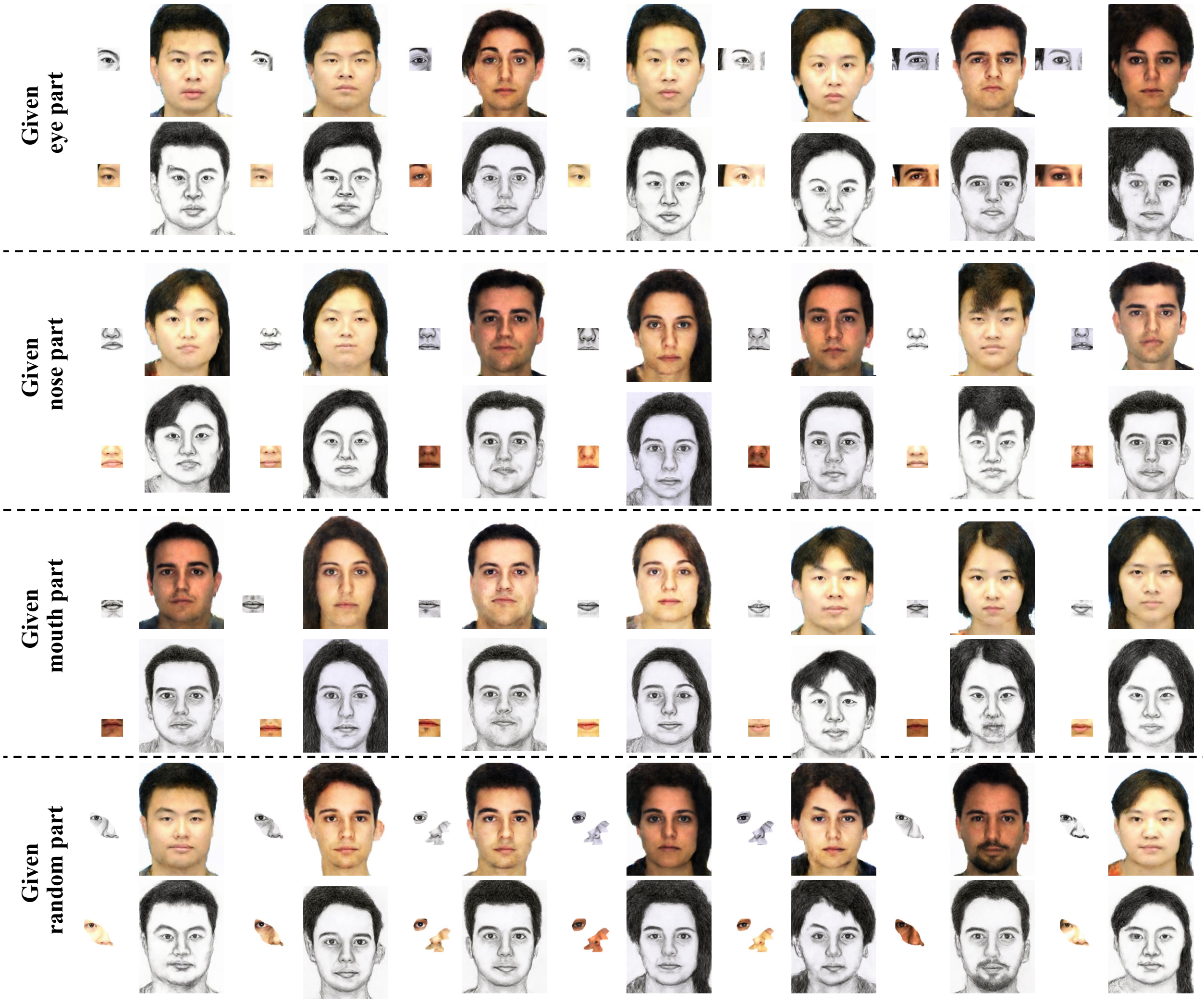


Figure 10. Generated faces/sketches from small patches of eyes, nose, mouth, and random regions by r-BTN.

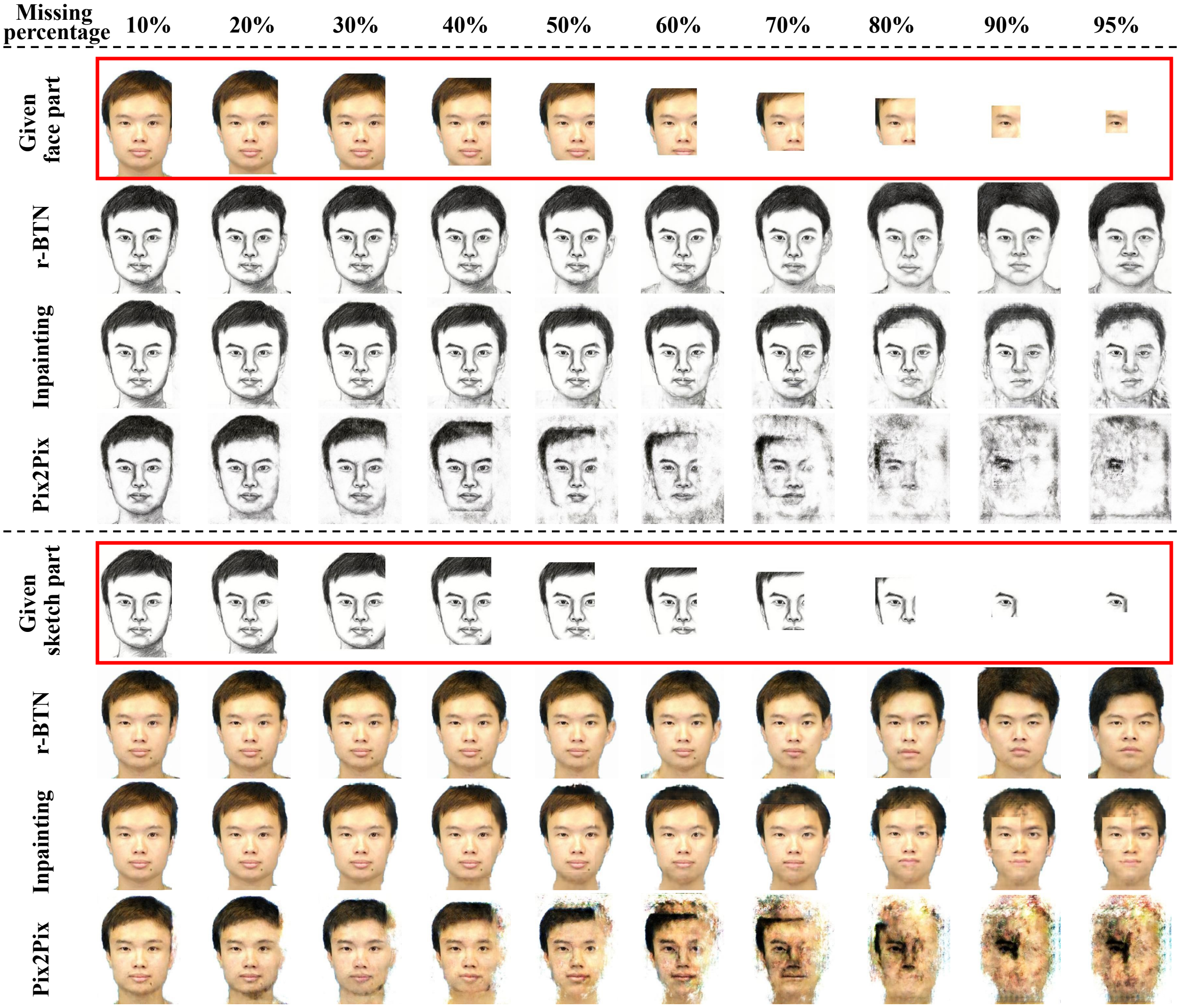


Figure 11. Example 1: Comparison of different methods in generating faces/sketches from patches with different missing percentage. The red boxes indicate the given face/sketch patches. The rest rows are correspondingly generated sketches/faces by the denoted methods. Please zoom in to see the details for small missing percentages.

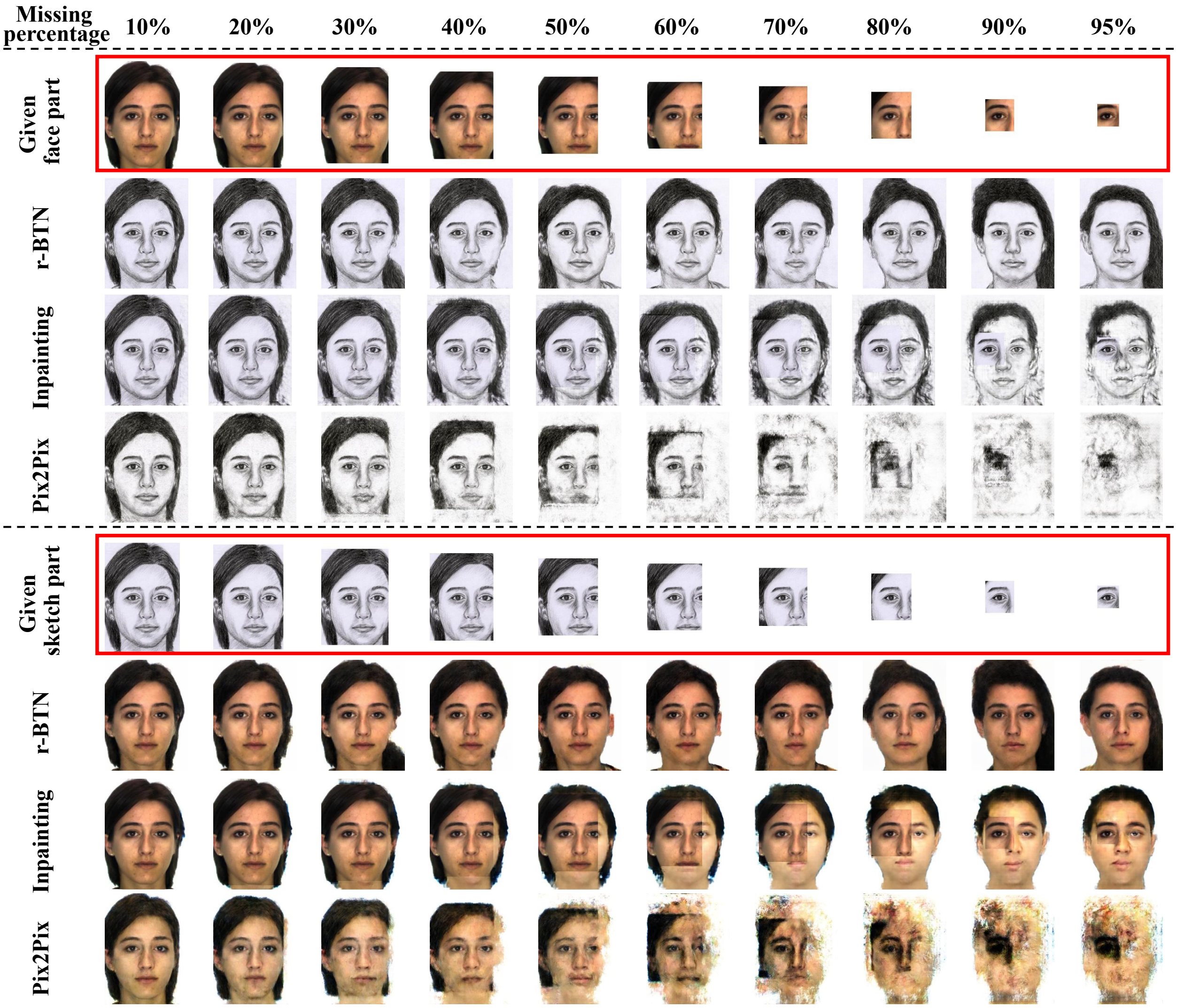


Figure 12. Example 2: Comparison of different methods in generating faces/sketches from patches with different missing percentage. The red boxes indicate the given face/sketch patches. The rest rows are correspondingly generated sketches/faces by the denoted methods. Please zoom in to see the details for small missing percentages.