Chinese Handwriting Imitation with Hierarchical Generative Adversarial Network

Jie Chang
j_chang@sjtu.edu.cn
Yujun Gu
yjgu@sjtu.edu.cn
Ya Zhang ✉
ya_zhang@sjtu.edu.cn
Yanfeng Wang
wangyanfeng@sjtu.edu.cn

Cooperative Medianet Innovation Center(CMIC).
Shanghai Jiao Tong University,
Shanghai, CHINA

Abstract

Automatic character generation, expected to save much time and labor, is an appealing solution for new typeface design. Inspired by the recent advancement in Generative Adversarial Networks (GANs), this paper proposes a Hierarchical Generative Adversarial Network (HGAN) for typeface transformation. The proposed HGAN consists of two sub-networks: (1) a *transfer network* mapping characters from one typeface to another preserving the corresponding structural information, which includes a *content encoder* and a *hierarchical generator*. (2) a *hierarchical adversarial discriminator* which distinguishes samples generated by the *transfer network* from real samples. Considering the unique properties of characters, different from original GANs, a hierarchical structure is proposed, which output the transferred characters in different phase of generator and at the same time, making the True/False judgment not only based on the final extracting features but also intermediate features in discriminator. Experimenting with Chinese typeface transformation, we show that HGAN is an effective framework for *font* style transfer, from standard printed typeface to personal handwriting styles.

1 Introduction

Designing a new Chinese Typeface is a very time-consuming task, requiring considerable efforts on manual design of benchmark characters. Automated typeface synthesis, i.e. synthesizing characters of a certain typeface given few manually designed samples, has been explored, however, usually based on manually extracted features [17, 18, 19, 22, 24]. These manual features heavily relies on preceding structural segmentation of characters, which itself is a non-trivial task and heavily affected by prior knowledge.

In this paper, we model typeface-transfer as an image-to-image transformation problem and attempt to directly learn the transformation end-to-end. Typically, image-to-image transformation involves a transfer network to map the source images to target images. A set of losses are proposed in learning the transfer network. Pixel loss is defined as pixel-wise difference between the output and the corresponding ground-truth [6, 21]. The perceptual loss [7],

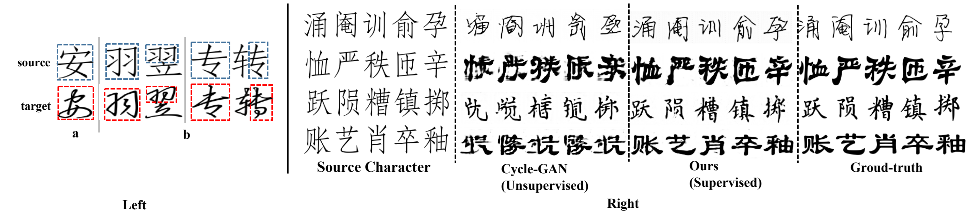


Figure 1: Left: (a) target style twists strokes in source character, making they do not share the invariant high-frequency features though they are the same character semantically. (b) The components in blue dotted box share the same radicals but their corresponding ones (in red dotted box) with target style are quite different. Right: CycleGAN [25] captures the correct style but completely fail in reconstructing the topological structure between strokes, demonstrating the supervised learning manner is necessary in this specific task.

perceptual similarity [3], style&content loss [1] and VGG loss [8] are proposed to evaluate the differences between hidden-level features and all are based on the ideology of feature matching [15]. More recently, with the proposal of GAN [4] and DCGAN [13], several variant of generative adversarial networks (e.g CycleGAN [25]), which introduce a discriminant network in addition to the transfer network for adversarial learning, have been successfully applied to image-to-image transformation including in-painting [12], de-noising [20], super-resolution [8] and others [3, 16]. While the above methods have shown great promise for various applications, they are not directly applicable to typeface transformation due to the following domain specific characteristics.

- Different from style-transfer between natural images where the source image shares high-frequency features with the target image, the transformation between two different typefaces usually leads to distortion of strokes or radicals (e.g Fig 1 Left), meaning change between different styles leads to change of high-level representations.
- For typeface transformation task, different characters may share the same radicals. This is a nice peculiarity that typeface transformation methods can leverage. However, sometimes in one certain typeface, the same radicals may appear quite differently in different characters. Fig 1 also presents two examples where certain radicals have different appearance in different styles. It will leads to severe over-fitting if we just considering the global property while ignore detailed local information.

Above characteristics in this specific task leads to the complete failure of existing state-of-the-art unsupervised image transformation method: CycleGAN [25](see Fig 1 Right). While the existing supervised image transformation methods can not directly apply to typeface transformation since their poor performance on recovering the more complicated and subtle detail in Chinese characters.

To overcome the above problems, we design a Hierarchical Generative Adversarial Network(HGAN) for Chinese typeface transformation, consisting of a *transfer network* and a *hierarchical discriminator* (Fig. 2), both of which are fully convolutional neural networks. First, in *transfer network*, a hierarchical generator which generates artificial images in multiple decoding layers, is proposed to help the decoder learn better representations in its hidden layers. Specially, the hierarchical generator attempts to maximally preserve the global topological structure in different decoding layers simultaneously considers the local features decoded in hidden layers, thus enabling the transfer network to generate close to authentic characters instead of disordered strokes. Second, we propose a *hierarchical discriminator* for adversarial learning. Specifically, the discriminator introduce additional adversarial losses, each of which employs feature representations from different hidden layers. The multi-adversarial losses constitute a hierarchical form, enabling the discriminator to dynamically measure the discrepancy in distribution between the generated domain and target domain, so that the *hierarchical generator* is trained to generate outputs with more similar statistical characteristics to the targets on different level of feature representation. The main contribution of our work is summarized as follows.

- We introduce a hierarchical generator in *transfer network* which generates multiple sets of characters based on different layers of decoded information, capturing both the global and local information for transfer. Correspondingly, a *hierarchical discriminator* which involves a cascade of adversarial losses at different layers of the network, is proposed to provide complementary adversarial capability. We have experimentally shown that the hierarchical discriminator leads to faster model convergence and generates more realistic samples.

- The proposed hierarchical generative adversarial network (HGAN) is shown to be successful for both Chinese handwriting imitation and character restoration through extensive experimental studies. The impact of proposed hierarchical adversarial loss is further investigated from different perspective including gradient propagation and the ideology of adversarial training.

- Last, experiments shows HGAN can be generalize to style-transfer in natural images.

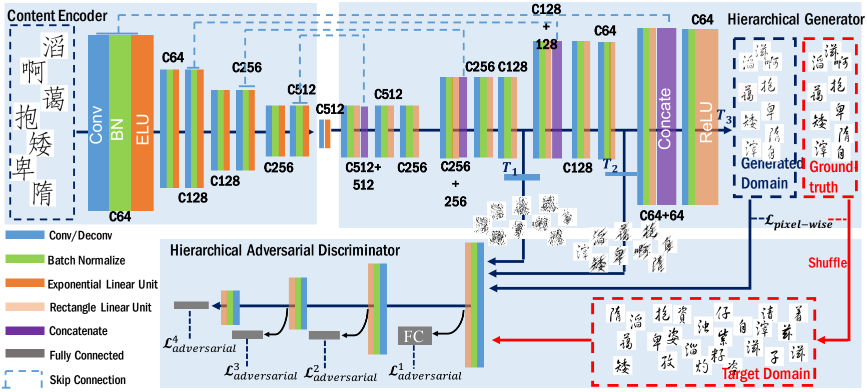


Figure 2: The proposed Hierarchical Generatvie Adversarial Network (HGAN). HGAN consists of an content encoder, a hierarchical generator and a Hierarchical Adversarial Discriminator. The content encoder follows the Conv-BatchNorm [5]-ELU [2] architecture. The hierarchical generator follows the Conv-BatchNorm-ReLU while two extra transformed characters are decoded from two intermediate features. The hierarchical adversarial discriminator is used to distinguish the transformed characters and the ground-truth from multi-level features.

2 Methods

In this section, we present the proposed Hierarchical Generative Adversarial Network (HGAN) for typeface transformation task. HGAN consists of a *transfer network* and a *hierarchical discriminator*. The former is further consists of an Content Encoder and a Hierarchical Generator.

2.1 FCN-Based Transfer Network

Content Encoder The Content Encoder maps a specified source characters to a content embedding. It shares a similar architecture to that of [14] with some modification. Because any information of relative-location is critical for Chinese character synthesis, we replace pooling operation with strided convolution in down-sampling since pooling helps reduce dimension and retains only robust activations in a receptive fields, however leading to the loss of spatial information in some degree (see Fig 2).

Hierarchical Generator Considering the domain insight of our task in Section 1. A Hierarchial Generator helps us model hierarchical representations of characters including the global topological structure and local topological of complicated Chinese characters. Specifically, different intermediate features of decoder are utilized to generate characters (T_1, T_2) too. Together with the last generated characters (T_3), all of them will be sent to the discriminator(see Fig 2). We only measure the pixel-wise difference between the last generated characters (T_3) and corresponding ground-truth. The adversarial loss produced by T_1 and T_2 helps to refine the *transfer network*. Meanwhile, the loss produced by the intermediate layers of decoder may provide regularization for the parameters in transfer network, which will relieves the over-fitting problem in some degree. In addition, for typeface transformation, the input character and the desired output are expected to share underlying topological structure, but differ in appearance or style. Skip connection [14] is utilized to supplement partial invariant skeleton information of characters with encoded features concatenated on decoded features. Both content encoder and hierarchical generator are fully convolutional networks [10].

2.2 Hierarchical Adversarial Discriminator

As mentioned in Section 1, adversarial loss introduced by discriminator is widely used in existing GAN-based image transformation task while all of them estimate the distribution consistency of two domain merely based on the final extracted features of discriminator. It is actually uncertain whether the learned features in last layers will provide rich and robust representations for discriminator. Additionally, We know the perceptual loss which penalizes the discrepancy between representations in different hidden space of images, is recently used in existing image-relative works. We combine the thought of perceptual loss and GANs, proposing a hierarchical adversarial discriminator which leverage the perceptual representations extracted from different intermediate layers of discriminator D and then distinguishes real/fake distribution between generated domain G_{domain} and target domain T_{domain}(See Fig 2). Each adversarial loss is defined as:

$$L_{d_i} = -\mathbb{E}_{f_t^i \sim p_{target}(f_t)}[\log D_i(f_t^i)] + \mathbb{E}_{s \sim p_{source}(s)}[\log D_i(f_s^i(T(s))] \tag{1}$$

$$L_{g_i} = -\mathbb{E}_{s \sim p_{source}(s)}[\log D_i(f_s^i(T(s))] \tag{2}$$

where f_t^i and $f_s^i(T(s))$ are i^{th} perceptual representations learned in *Discriminator* from target domain and generated domain respectively. D_i is branch discriminator cascaded after every intermediate layer and $i = 1, 2, ..4$ which depends on the number of convolutional layers in our discriminator D. This variation brings a complementary adversarial training for our model, which urges discriminator to find more detailed local discrepancy beyond the global distribution. Assuming L_{d_4} and its corresponding L_{g_4} reach nash equilibrium, which means the the perceptual representations f_t^4 and $f_s^4(T(s))$ are considered sharing the similar distribution, however other adversarial losses (L_{d_i}, L_{g_i}), $i \neq 4$ may have not reach nash equilibrium since these losses produced by shallow losses pay more attention on regional information during training. The still high loss promotes the model to be continuously optimized until all perceptual representations pairs $(f_t^4, f_s^4(T(s)))$, $i = 1, 2, ..4$ are indistinguishable by discriminator. Experiments shows this strategy makes the discriminator to dynamically and automatically discover the un-optimized space from various perspectives.

2.3 Losses

Pixel-level Loss L1- or L2-norm are often used to measure the pixel distance between paired images. For our typeface transformation task, each pixel in character is normalized near 0 or 1 value. So cross entropy function is selected as per-pixel loss since this character generation problem can be viewed as a logistic regression problem:

$$L_{pix-wise}(T) = \mathbb{E}_{(s,t)}[-t\lambda_w \cdot (\log \sigma(T(s))) - (1-t) \cdot \log(1 - \sigma(T(s)))], \tag{3}$$

where T denotes the transformation of *transfer network*, (s,t) is pair-wise samples where $s \sim p_{source\_domain}(s)$ and $t \sim p_{target\_domain}(t)$. σ is *sigmoid* activation.

Particularly, a weighted parameter λ_w is introduced into pixel-wise loss for balancing the ratio of positive(value 0) to negative(value 1) pixels in every typeface style. We add this trade-off parameter based on the observation that some typefaces are thin (i.e. more negative pixels) while some may be relatively thick (i.e. more positive pixels). λ_w is not a parameter determined by cross validation, it is explicitly defined by:

$$\lambda_w = 1 - \frac{\sum_{k=1}^{K} \sum_{n=1}^{N} \mathbb{1}\left\{t_k^n \geq 0.5\right\}}{\sum_{k=1}^{K} \sum_{n=1}^{N} \mathbb{1}\left\{t_k^n < 0.5\right\}}, \tag{4}$$

where N the is the resolution of one character image(here $N = 64$), K denotes the number of target characters in training set and t_k^n denotes the n^{th} pixel value of k^{th} target character.

Hierarchical Adversarial Loss For our proposed HGAN, each adversarial loss is defined by Eq 1 and Eq 2:

$$L^i_{adversarial}(D_i, T) = L_{d_i} + L_{g_i}. \tag{5}$$

Noted that here we integrate original $t \sim p_{target}(t)$ and $s \sim p_{source}(s)$ into Eq. 5 for a unified formulation, then the total adversarial losses is

$$L_{total\_adversarial}(D, T) = \sum_{i=1}^{k} \lambda_i \cdot L^i_{adversarial}(D_i, T), \tag{6}$$

where λ_i are weighted parameters to control the effect of every branch discriminator. The total loss function is formulated as follows:

$$L_{total} = \lambda_p L_{pix-wise}(T) + \lambda_a L_{total\_adversarial}(D, T), \tag{7}$$

where λ_p and λ_a are the trade-off parameters.

We optimize *transfer network* and *hierarchical adversarial discriminator* by turns.

3 Experiments

3.1 Data Set

We build a Chinese character data set by downloading large amount of .ttf scripts denoting different typefaces from the the website http://www.foundertype.com/. After preprocessing, each typeface ends up with 6000 grey-scale images in 64×64.png format.

3.2 Network Setup

The hyper-parameters relevant to our proposed network are annotated in Fig 2. The *content encoder* includes 8 conv-layers while the *hierarchical generator* is more deeper including 4 transform-conv layers and 8 con-layers. Every *conv* and *deconv* are followed by Conv-BatchNorm(BN) [5]-ELU [2]/ReLU structure. 4 skip connections are used on mirror layers both in *encoder* and *staged-decoder*. For the trade-off parameters in Section 2.3, λ_w is determined by Eq 4. The number of adversarial loss of HAN l is 4 and weighted parameter $\{\lambda_i\}_1^3$ is decay from 1 to 0.5 with rate 0.9, $\lambda_4 = 1.0$. λ_p and λ_a are both set to 1.0 to weight the pixel loss and adversarial loss.

3.3 Performance Comparison

To validate the proposed HGAN model, we experimentally compare the transfer performance of HGAN with two specialized Chinese calligraphy synthesis method (AEGG [11] and EMD [23]), a state-of-the-art supervised image-to-image transformation method (Pix2Pix [6]) and an unsupervised image-to-image transformation method (CycleGAN [25])

Qualitatively Performance All baselines except CycleGAN need pair the generated images with corresponding ground-truths for training. The *transfer network* of Pix2Pix shares the identical framework with that in our HGAN(see Fig 2) and the network architecture used in AEGG follows the instructions of their paper with some tiny adjustment for dimension adaptation. 50%($\tilde{3}$000) characters randomly selected from *FS* typeface as well as 50% corresponding target style characters selected from other handwriting-style typefaces are used as training set. The remaining 50% of *FS* typefaces is used for testing. We illustrate experimental results by transferring *FS* typeface to other 5 personal Chinese handwriting-style(see Fig 3). Both AEGG and Pix2Pix can capture coarse style of handwriting, however in test set, they failed to synthesize recognizable characters because most strokes in generated character are disordered even chaotic. While we observed that both of them perform well on training set but far worse on test set, which suggests the proposed hierarchical adversarial loss makes our model less prone to over-fitting in some degree. Experimental results demonstrate HGAN is superior in generating detailed component of characters so that it significantly outperforms previous work, especially on transferring cursive handwriting style.

We also compare the our HGAN with the most state-of-the-art Chinese typeface transformation model: EMD [23], a generalized style transfer framework which can even be generalized to new unseen style by separating style and content of the specified typeface. Actually, HGAN and EMD are two different learning pattern: HGAN encodes the fixed style-information in the network, while EMD dynamically encodes the style-information in the network according to the inputted target typeface denoting the specified style. For a fair comparison, we evaluate the performance of both EMD and HGAN on the training typefaces. Fig 4 shows us they perform equally well on legible typeface, while HGAN captures more accurate style and more subtle stroke detail than EMD in cursive handwriting style.

Quantitative Evaluation. Beyond directly illustrating qualitative results of comparison experiments, two quantitative measurements: Root Mean Square Error(RMSE) and Average Pixel Disagreement Ration [9](APDR) are utilized as evaluation criterion. As shown in Table 1, our HGAN leads to the lowest RMSE and APDR value compared with existing methods.

Model	$FS\rightarrow hw1$		$FS\rightarrow hw2$		$FS\rightarrow hw3$		$FS\rightarrow hw4$	
	RMSE	APDR	RMSE	APDR	RMSE	APDR	RMSE	APDR
AEGG [11]	22.671	0.143	28.010	0.211	24.083	0.171	22.110	0.131
Pix2Pix [6]	29.731	0.231	27.117	0.225	26.580	0.187	24.135	0.180
CycleG [25]	29.602	0.253	29.145	0.234	28.845	0.241	25.632	0.191
EMD [23]	21.435	0.163	25.230	0.207	25.190	0.180	22.005	0.130
HAN(strong)	**19.498**	**0.118**	**23.303**	**0.181**	**22.266**	**0.162**	**19.528**	**0.110**

Table 1: Quantitative Measurements

3.4 Analysis of Hierarchical Generative Adversarial Loss

We analyze each adversarial loss, $\{L_{d_i}\}_{i=1}^{4}$ and $\{L_{g_i}\}_{i=1}^{4}$, defined in Section 2.2. As shown in Fig 5, the generator loss gen_4 produced by the last conv-layer in hierarchical discriminator fluctuates greatly and then gen_3 produced by the penultimate layer, $\{gen\_2, gen\_1\}$ produced by shallower conv-layers are relatively gentle because λ_4 is set larger than $\{\lambda_i\}_{i=1}^{3}$ so that network mainly optimizes gen_4. However for discriminator loss, $\{dis\_4, dis\_3, dis\_1\}$ derived from D_4, D_3,D_1 are mostly numerical approach. We further observed that the trend of increase or reduction among various discriminator losses are not always consistent. We experimentally conclude that adversarial losses produced by intermediate layers can assist training: when D_4 is severely cheated by real/fake characters, D_3 or D_2 or D_1 can still give a high confidence of differentiating, which means True/False discrimination based on different representations can be compensated each other(see Fig 5 for more details) during training. Our hierarchical adversarial discriminator actually plays an implicitly role of fitting distribution from two domains instead of fitting hidden features from paired images to be identical compared with existing methods, which can relieve over-fitting during training.

We further explore the influence brought by our hierarchical adversarial loss. By removing the effect of hierarchical architecture from our HGAN model, we run another contrast experiment, Single Generative Adversarial Network (SGAN). The detail of network follows Fig 2 and we set trade-off parameters $\lambda_1 = \lambda_2 = \lambda_3 = 0.5$ and $\lambda_4 = 1$ in loss function of HGAN, while we set $\lambda_1 = \lambda_2 = \lambda_3 = 0$, $\lambda_4 = 1$ as well as cutting off the data flow of T_1, T_2 for SGAN in order to remove the influence of extra 3 adversarial losses and hierarchical generator. Characters generated during different training period are illustrated in Fig 7 from which we can see qualitative effect of proposed hierarchical GAN. Our proposed HGAN generates more clear characters compared with SGAN at the same phase of training period, which suggests HGAN converge greatly faster than SGAN. We also run 3 parallel typeface-transfer experiments then calculate RMSE along with the iterations of training on train set. Left loss-curves in Fig 7 demonstrates that hierarchical adversarial architecture assists to accelerate convergence and leads to lower RMSE value.

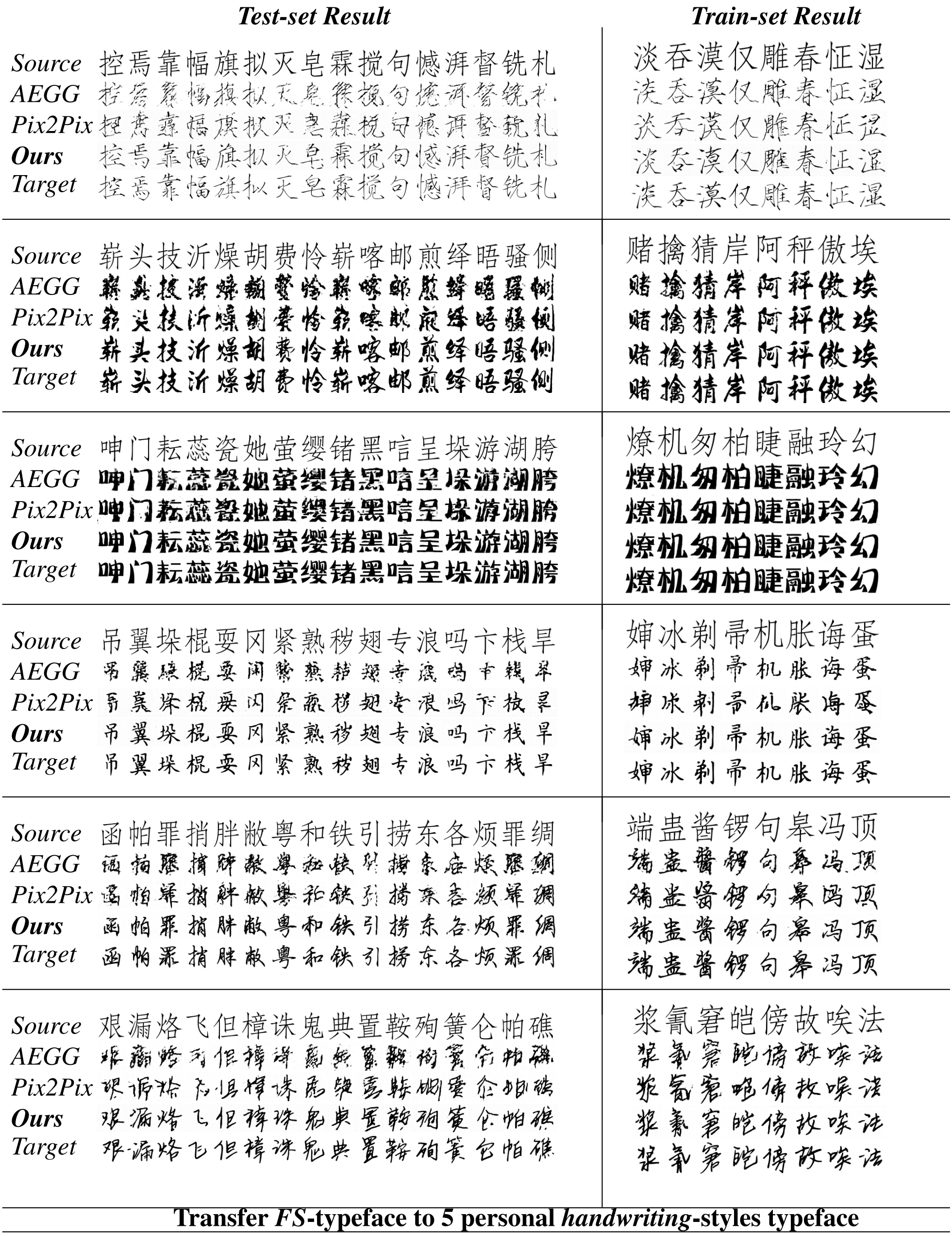

Test-set Result
Train-set Result
Source
AEGG
Pix2Pix
Ours
Target
Transfer FS-typeface to 5 personal handwriting-styles typeface

Figure 3: Performance of transferring *FS* typeface (Source) to other 5 personal *handwriting*-style typefaces. We compare our *HGAN* with a specialized model proposed for Chinese typeface transformation: AEGG [18] and a state-of-the-art image-to-image transformation model: Pix2Pix [6].

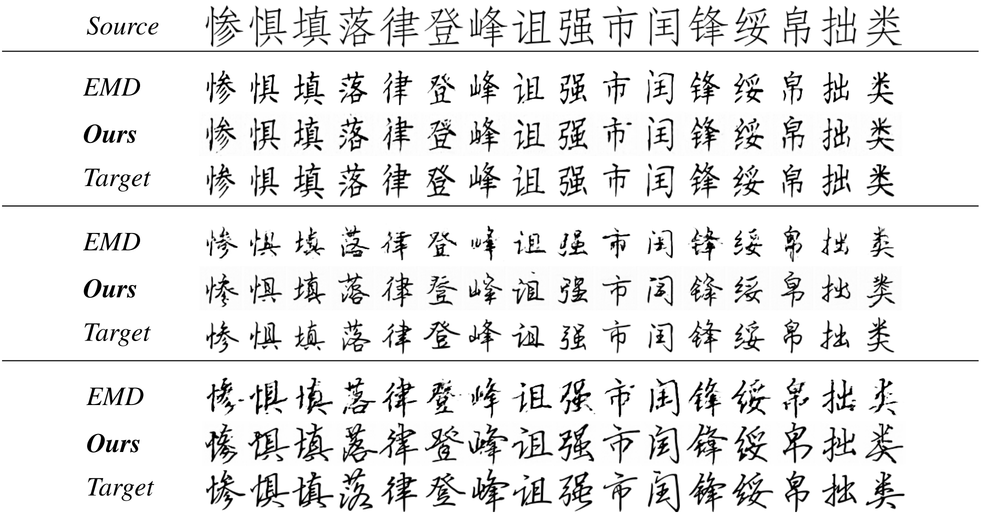
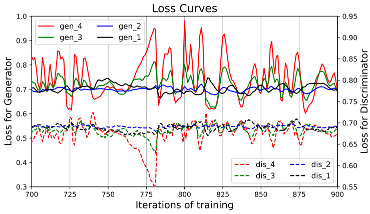


Figure 4: Performance of transferring *FS* typeface (Source) to other 3 personal *handwriting*-style typefaces. Our HGAN performs as well as EMD [23] on legible typeface (top column), while HGAN greatly performs better than EMD on cursive typefaces (middle and bottom columns).



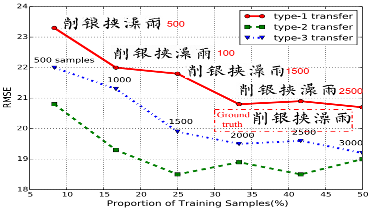


Figure 5: Each generator loss and discriminator loss during ste 700 to 900.

Figure 6: The RMSE evaluation under different size of trainging samples.

3.5 Impact of Training Set Size

Last, we experiment at least how many handwriting characters should be given in training to ensure a satisfied transfer performance. So we experiment three typeface-transfer tasks(type-1, type-2 and type-3) with different proportion of training samples and then evaluate on each same test set. As the synthesized characters shown in Fig **??**, the performance improves along with increase size of training samples. RMSE evaluation curves suggest when the training size is not less than 35%(2000 samples) of whole dataset, the performance will not be greatly improved.

3.6 Character Restoration and Image Style Transfer

As illustrated in Fig 6, we also applied our HGAN model to character restoration and art-style transfer for natural images. We randomly mask 30% region on every handwriting characters in one typeface's training set and our HGAN is able to correctly reconstruct the missing part

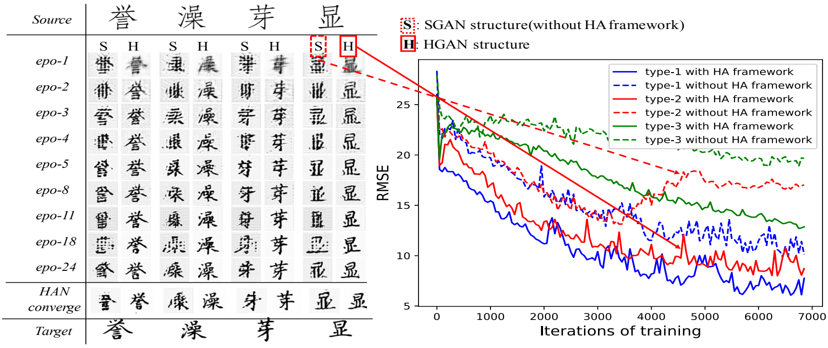


Figure 7: Contrast experiments for HGAN and SGAN. Left: Characters generated by HGAN are far more better than that by SGAN in same training epoch. *HGAN converge* row shows characters generated when our HGAN model converges. Right: The RMSE evaluation loss along with the training iterations under HGAN and SGAN which shows HGAN leads to more lower value than SGAN.

of one character on test set. HGAN also obtains good performance on art-style transfer.

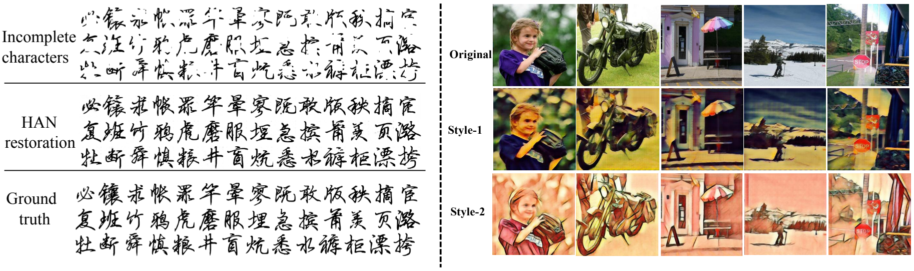


Figure 8: Performance of character restoration (Left) and art-style transfer (Right).

4 Conclusion and Future Work

In this paper, we propose a hierarchical generative adversarial network for typeface transformation. The proposed *hierarchical generator* and *hierarchical adversarial discriminator* can dynamically estimate the consistency of two domains from different-level perceptual representations, which helps our HGAN converge faster and better. Experimental results show our HGAN can synthesize most handwriting-style typeface compared with existing natural image-to-image transformation methods. Additionally, our HGAN can be applied to handwriting character restoration.

Acknowledgements

This work is supported by The High Technology Research and Development Program of China (2015AA015801), NSFC (61521062), and STCSM (18DZ2270700).

References

[1] Dongdong Chen, Lu Yuan, Jing Liao, Nenghai Yu, and Gang Hua. Stylebank: An explicit representation for neural image style transfer. *arXiv preprint arXiv:1703.09210*, 2017.

[2] Djork-ArnÃľ Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *Computer Science*, 2015.

[3] Alexey Dosovitskiy and Thomas Brox. Generating images with perceptual similarity metrics based on deep networks. In *Advances in Neural Information Processing Systems*, 2016.

[4] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *International Conference on Neural Information Processing Systems*, pages 2672–2680, 2014.

[5] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Computer Science*, 2015.

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

[7] Justin Johnson, Alexandre Alahi, and Fei Fei Li. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711, 2016.

[8] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, and Zehan Wang. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802*, 2016.

[9] Ming-Yu Liu and Oncel Tuzel. Coupled generative adversarial networks. In *Advances in neural information processing systems*, pages 469–477, 2016.

[10] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[11] Pengyuan Lyu, Xiang Bai, Cong Yao, Zhen Zhu, Tengteng Huang, and Wenyu Liu. Auto-encoder guided gan for chinese calligraphy synthesis. *arXiv preprint arXiv:1706.08789*, 2017.

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

[13] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *Computer Science*, 2015.

[14] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 9351, pages 234–241, 2015.

[15] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. In *Advances in Neural Information Processing Systems*, 2016.

[16] Chaoyue Wang, Chang Xu, Chaohui Wang, and Dacheng Tao. Perceptual adversarial networks for image-to-image transformation. *arXiv preprint arXiv:1706.09138*, 2017.

[17] Jianguo Xiao, Jianguo Xiao, and Jianguo Xiao. Automatic generation of large-scale handwriting fonts via style learning. In *SIGGRAPH ASIA 2016 Technical Briefs*, page 12, 2016.

[18] Songhua Xu, Hao Jiang, Tao Jin, Francis C. M. Lau, and Yunhe Pan. Automatic generation of chinese calligraphic writings with style imitation. *IEEE Intelligent Systems*, 24(2):44–53, 2009.

[19] Songhua Xu, Tao Jin, Hao Jiang, and Francis C. M. Lau. Automatic generation of personal chinese handwriting by capturing the characteristics of personal handwriting. In *Conference on Innovative Applications of Artificial Intelligence, July 14-16, 2009, Pasadena, California, Usa*, 2010.

[20] He Zhang, Vishwanath Sindagi, and Vishal M. Patel. Image de-raining using a conditional generative adversarial network. *arXiv preprint arXiv:1701.05957*, 2017.

[21] Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorization. In *European Conference on Computer Vision*, pages 649–666, 2016.

[22] Xu-Yao Zhang, Fei Yin, Yan-Ming Zhang, Cheng-Lin Liu, and Yoshua Bengio. Drawing and recognizing chinese characters with recurrent neural network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[23] Yexun Zhang, Ya Zhang, and Wenbin Cai. Separating style and content for generalized style transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, 2018.

[24] Baoyao Zhou, Weihong Wang, and Zhanghui Chen. Easy generation of personal chinese handwritten fonts. In *IEEE International Conference on Multimedia and Expo*, pages 1–6, 2011.

[25] Jun Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*, 2017.