Xinhao Wang, Shuai Yang, Wenjing Wang, and Jiaying Liu

Artistic Text Style Transfer

An overview of state-of-the-art methods and datasets

Word art, which is text rendered with properly designed appealing artistic effects, has been a popular form of art throughout human history. Artistic text effects are of great aesthetic value and symbolic significance. Decorating with appropriate effects not only makes text more attractive but also significantly enhances the atmosphere of a scene. Thus, artistic text effects are widely used in publicity and advertising. Some text effects are simple, such as colors and shadows, while some can be complex, such as the burning flames in Figure 1 and exquisite decorations in Figure 2. Manually creating vivid text effects requires lots of time and a series of complicated operations: observing the target glyphs, designing appropriate artistic effects, warping the texture to match the character shapes, and so on. It consumes hours of time even for well-trained designers. To produce word art more conveniently and efficiently, artistic text style transfer has been proposed recently to automatically render text with given artistic effects.

In this article, we provide a comprehensive overview of current advances in artistic text style transfer. First, we formulate the task. Second, we investigate and classify state-of-the-art methods into nondeep- and deep-based methods, introduce their core ideas, and discuss their strengths and weaknesses. Third, we present several benchmark datasets and evaluation methods. Finally, we summarize the current challenges in this field and propose possible directions for future research.

Task formulation

Artistic text style transfer aims at automatically turning plain text into fantastic artworks with given artistic

Digital Object Identifier 10.1109/MSP.2022.3196763
Date of current version: 27 October 2022

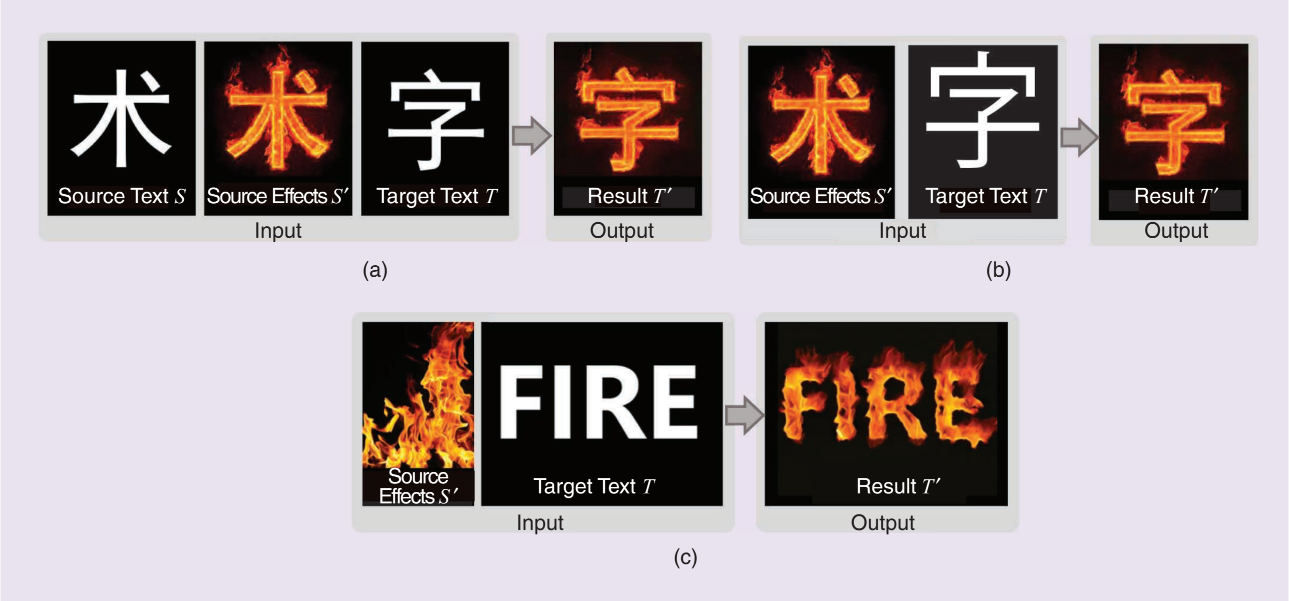

FIGURE 1. An overview of application scenarios of text style transfer methods. (a) Transfer of text effects with supervision. (b) Joint transfer of text effects and fonts without supervision. (c) Transfer of effects from arbitrary style images without supervision.

1053-5888/22©2022IEEE

effects. According to the input, we divide the artistic text style transfer problem into three categories from easy to difficult: supervised effect transfer, unsupervised effect transfer, and joint font and effect transfer.

For a supervised effect transfer, as shown in Figure 1(a), the source effects S' in addition to the corresponding nonstylized image S are required. The algorithms learn the transformation between them, then apply it to the target text T to synthesize the result T'. An unsupervised effect transfer, on the other hand, gets rid of the dependency on S and directly builds the transformation by extracting the proper features of source effects S' and target text T, as shown in Figure 1(b). Since S is not required, the constraints on S' are further relaxed, where S' can be arbitrary style images besides text effects [Figure 1(c)]. As for joint font and effect transfer, it considers fonts as a part of the style, thus aiming at transferring artistic effects and text fonts jointly.

Nondeep text effect transfer methods

Early research mainly regards text effects as textures. These methods adopt image patches to model artistic styles and use a technology called *patch matching* [1] to synthesize textures according to the text shapes. Intuitively, patch matching [1] divides the source effect image S' and target text image T into overlapped patches. For each patch of T, a set of best-matched patches from S' is collected, among which one is chosen according to a certain criterion and used to generate the texture of the target patch. Since the texture and style information come directly from the source effect image, these patch-based methods are able to generate rich texture details.

In this section, we review some nondeep text effect transfer methods according to the four typical application scenarios in Figure 3: effect transfer with/without supervision, effect transfer with human interaction, and dynamic effect transfer. We first introduce each method and then discuss their strengths and weaknesses.

Supervised text effect transfer

Yang et al. [2] first raised the brand-new topic of text effect transfer and designed a nondeep algorithm T-Effect specialized for rendering awesome word art. Compared to a general image style transfer, the authors pointed out three new challenges that text effect transfer faces. 1) Text effects and text shapes are extremely diverse. 2) It is challenging to compose the glyph and style elements properly. 3) The input text images are too simple to instruct the placement of different subeffects. As an attempt to solve these difficulties, Yang et al. initially investigated and analyzed well-designed text effects and summarized their key characteristics: *high correlation between patch patterns (e.g., color and scale) and their distances to text skeletons*. Based on this clue, the patch-matching algorithm [1]

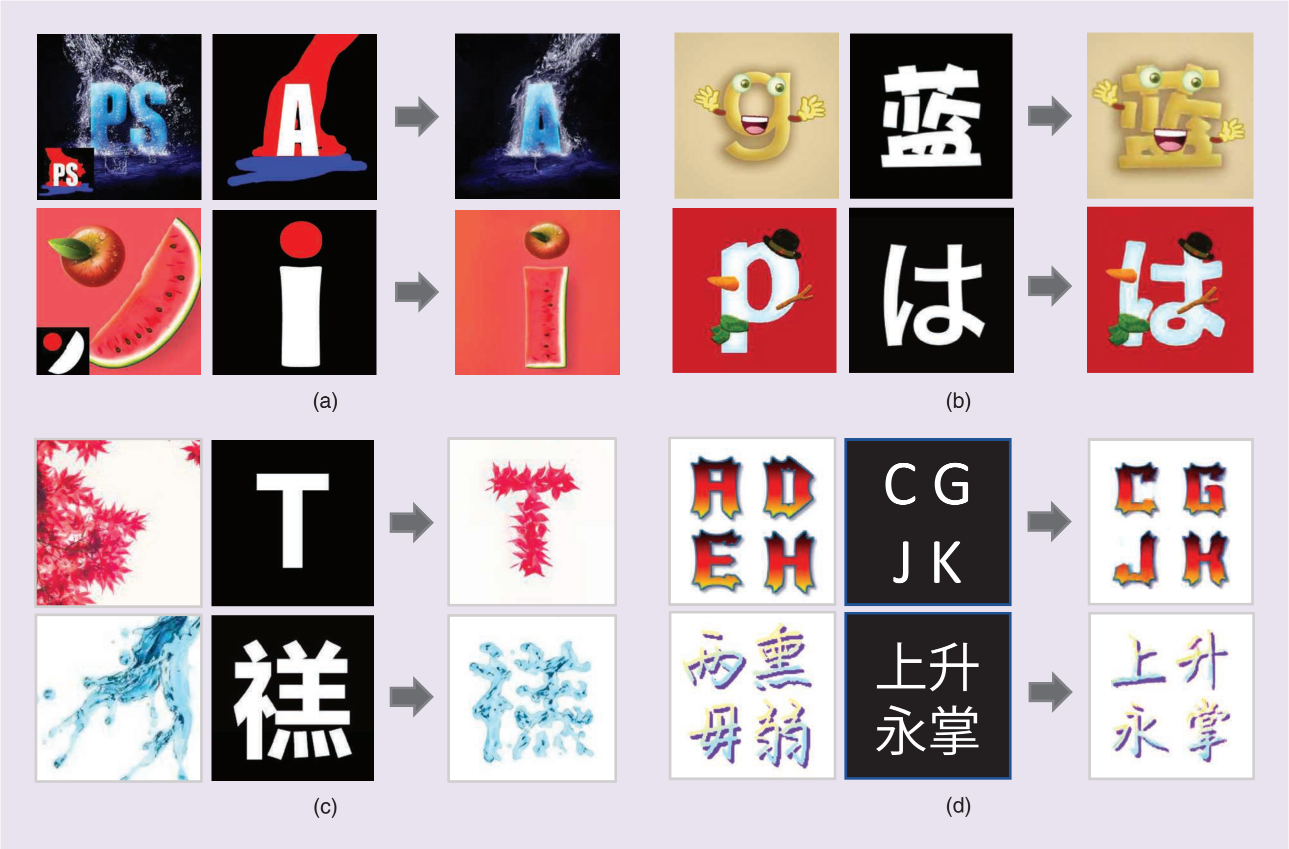

FIGURE 2. Some results of different text style transfer methods. (a) User-interactive effects transfer. (b) Decorative elements effects transfer. (c) Shape-matching effects transfer. (d) Joint effects and fonts transfer.

is improved from two perspectives: patch preparation and a matching strategy. For patch preparation, T-Effect automatically detects the optimal patch scale to depict texture patterns around each pixel. In the process of matching, T-Effect processes effects and glyphs at the same time, takes the distance from texture patches to text skeletons into consideration, and adds a psychovisual term to avoid texture overrepetitiveness. However, this method requires a corresponding nonstylized image S, as shown in Figure 3(a), to learn the transformation between text and artistic text. Unfortunately, such a pair of inputs is often unavailable in practice, which limits its application scenarios.

Unsupervised text effect transfer

To tackle the aforementioned issue, Yang et al. [3] further presented a novel unsupervised algorithm UT-Effect to stylize the text only with a target text image T and an arbitrary source effects image S', as in Figure 3(b). To build up the mapping relation between different modalities, UT-Effect [3] first extracts the main structural imagery of the style image with the help of a texture-removal algorithm [4] and superpixel extraction [5]. The extracted structure naturally serves as a guidance map and builds a preliminary mapping to the text. The mapping is then refined by a bidirectional legibility-preserving structure transfer algorithm, which adds shape characteristics of the source style to the text while maintaining the main structure of the text stroke. Based on the mapping, a patch-matching [1] texture transfer algorithm guided by saliency constraints is proposed, yielding satisfying results without any supervision. This method can be further extended to graphic design, where text is inserted into an image, matching the foreground and harmonizing with the background. The authors designed a new context-aware synthesis framework to determine where to place and how to embed the text, making the algorithm a powerful tool for computer-aided poster design.

Interactive effect transfer

An artistic text effect transfer can also introduce user interaction, in which the texture is transferred under user guidance. Men et al. [6] established a general framework of user-guided texture transfer for multiple tasks. To control the spatial distribution of stylized texture, this method requires users to annotate semantic maps for both source and target texts, as shown in Figure 3(c). Although introducing semantic maps provides a more accurate semantic mapping relationship, the texture arrangement inside each semantic area is not taken into consideration, making some existing structural textures easily lost in the transfer process. To successfully transfer these unmarked structural textures, the authors used content-aware saliency detection to extract them and transfer them through matching key contour points. These pretransferred structural textures act as a prior in the following step. Combining semantic and structure information enables an improved patch-matching algorithm [1] to provide high-quality texture with content awareness and low-level details. As shown in Figure 2(a), this method is able to precisely control the spatial distributions of text effects with user guidance. The combined semantic and structure guidance also invests this method with great flexibility; therefore, it can further generalize to other shapes besides text.

Dynamic text effect transfer

With the development of mobile Internet and social media, videos and graphic interchange formats, which are more vivid informative carriers of visual information, have become widespread. Extending the static effect transfer on a single image to a dynamic effect transfer on videos has great potential applications. Compared with the supervised static effect transfer in Figure 3(a), the dynamic effect transfer replaces the single image S' with a set of video frames, as shown in Figure 3(d). The result T' correspondingly becomes a set of video frames. The first method for dynamic effects was proposed by Men et al. [7]. To achieve temporal consistency and preserve the spatial texture continuity, this algorithm simultaneously carries

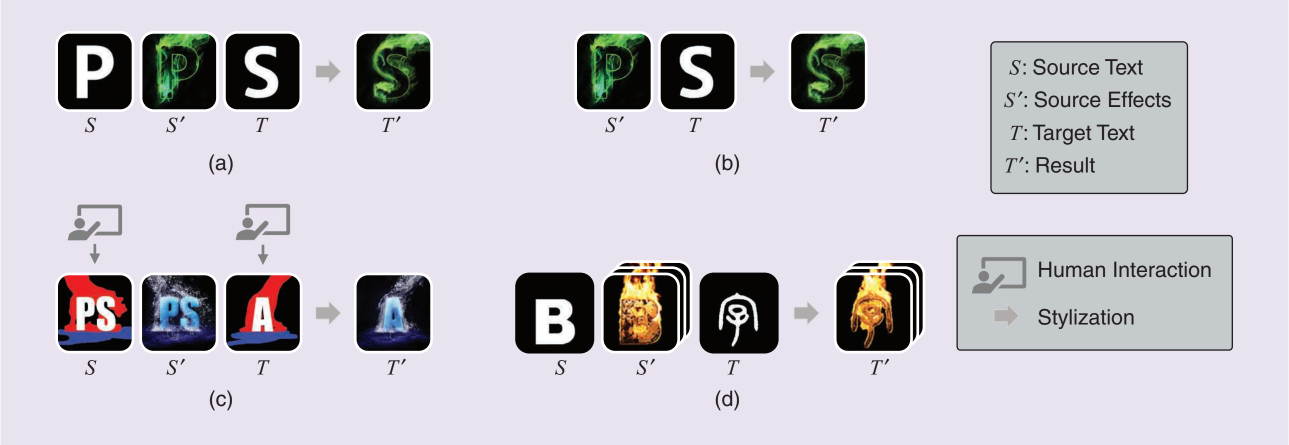

FIGURE 3. A brief introduction to nondeep text style transfer methods. (a) Supervised effects transfer. (b) Unsupervised effects transfer. (c) Supervised effects transfer with human interaction. (d) Supervised dynamic effects transfer.

out patch matching across all keyframes, during which common nearest-neighbor patches are found for all frames. In addition, the authors introduced simulated annealing for deep direction-guided propagation to ensure that complicated effects can be completely synthesized. However, this method also requires the corresponding nonstylized image of the dynamic effects video, which is usually not available in practice.

Deep-based text effect transfer methods

The aforementioned nondeep algorithms are capable of faithfully transferring prescribed style effects and synthesizing appealing stylized text. They are all based on patch matching [1], whose essence is rearranging image patches. Consequently, nondeep algorithms typically have limited flexibility and a high time complexity because of the iterative patch-matching procedure. The need to address these limitations gives birth to deep-based algorithms. Instead of operating on the image patches, deep-based methods automatically extract features of artistic text effects and text shapes. These high-dimensional features learned from data disentangle the characteristics of word art and are easier to adjust. Thus, by operating in the feature space, deep-based methods serve as a more flexible and powerful editing tool.

A generative adversarial network (GAN) is the basis of most deep-based text effect transfer methods. In a GAN, a subnetwork called a *discriminator* distinguishes real text effect images from fake ones generated by the text effect transfer network. The two networks are trained together in an adversarial manner until they achieve equilibrium. Based on the applicability, deep-based methods can be divided into two categories: multistyle-per-model and per-style-per-model. Multistyle-per-model methods are able to process multiple effects with a single trained network but usually rely on a large amount of training data. A basic framework is shown in Figure 4(a); the network takes a pair of source effects S' and target text T as input and produces the stylized text T'. During training, a discriminator is utilized to improve the visual quality of synthesized results. Per-style-per-model methods, on the other hand, are mainly designed for a more challenging one-shot learning task, where the network is only supervised by a single training image pair. One representative framework is demonstrated in Figure 4(b). As shown in the top part, the network uses the given sample to learn the transfer of structure and texture during training. Then, as shown in the bottom part, the learned model can be applied to the target text during testing. The limitation is that per-style-per-model methods only can learn one style in the training process. In this section, we introduce several representative deep-based text effect transfer methods.

Disentangle the text and effect

One intuitive path toward text effect transfer is to characterize the glyphs and styles separately and realize flexible effect transfer by disentangling and recombining them. This idea was first utilized by Yang et al. [8], who developed an encoder–decoder-based texture effects transfer GAN (TET-GAN) to support stylization and destylization at the same time. The authors designed three tasks to simultaneously train TET-GAN: text autoencoding, stylization, and destylization. In this way, the encoders are capable of disentangling content and style features, while decoders are able to recombine content and style features to synthesize stylized text. In addition, a self-stylization training scheme can be used for one-shot finetuning when facing an unseen effect. This method has a good performance when the spatial distribution of the texture in text effects is highly related to its distance from the text skeleton. But it still produces poor results when the network fails to recognize the glyph in style images.

Transfer the effects with decorative elements

All of the aforementioned mentioned approaches assume that the styles are uniform within or outside the text, thus

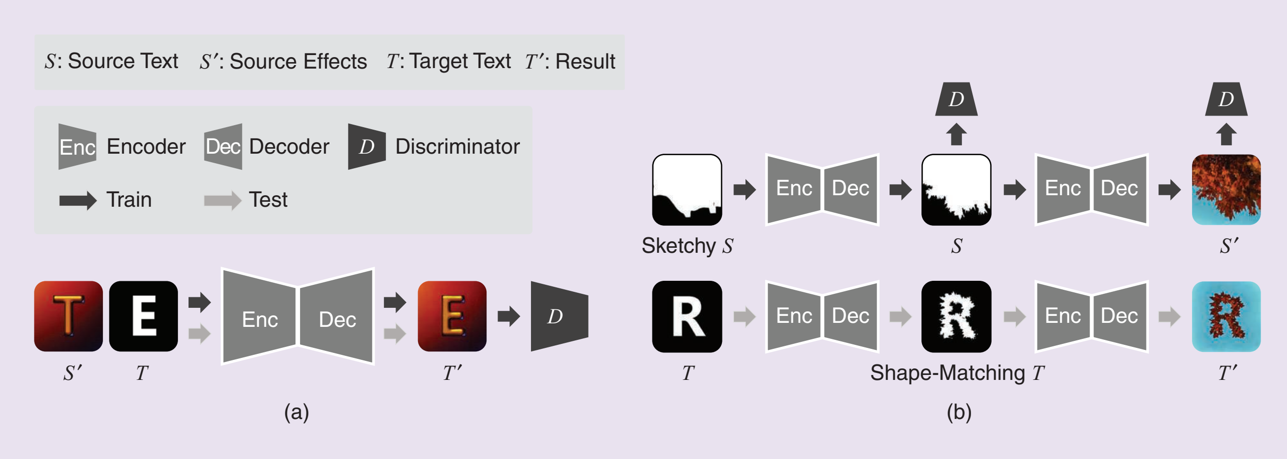

FIGURE 4. The two categories of deep-based text style transfer methods. (a) Multistyle-per-model methods. (b) Per-style-per-model methods.

failing to render the exquisite decorative elements that are commonly used in artistic text design and resulting in visual quality degradation. To address this problem, Wang et al. [9] proposed to detect, separate, and recombine these important embellishments. First, a segmentation network is trained to detect the decorative elements. Then, based on the segmentation results, the decorative elements are separated from basal text effects, and a text style transfer network infers the basal text effects for the target text. Finally, cues for spatial distributions and element diversities are characterized to jointly integrate the decorative elements onto the target text. Similar to [8], a one-shot fine-tuning scheme is proposed to empower the network to extend to a new style with only one example. This algorithm performs especially well when the reference style image contains elaborate decorative items. Some representative results are shown in Figure 2(b).

Controllable text effect transfer

Text is significantly different from and more structured than nontext images. When using arbitrary artistic images as style references, the glyph should deform to better resemble the style subject, but overdeformation degrades the legibility. In other words, there is a tradeoff between glyph legibility and stylistic degree. To find a delicate balance, Yang et al. [10] proposed a controllable artistic text style transfer algorithm, Shape-Matching GAN, which supports real-time control of glyph deformations. There are two main challenges for glyph deformations: how to change the text shapes to match the reference style and how to control the deformation degree. For the former, the authors proposed a novel bidirectional shape-matching strategy, which establishes a shape mapping between the source style and target text through both backward and forward transfers. The authors first simplified the style input into a sketchy structure S, named *backward structure transfer*. Then, models are trained to map the sketchy S to the original S and further to S' to learn the forward structure and texture transfers. Discriminators are applied to enhance the generation effects.

For the latter, the authors built a scale-controllable module to empower the network to learn the style features on a continuous scale. In this way, the deformation degree can be simply controlled with a weight parameter. As a representative per-style-per-model method, the framework of Shape-Matching GAN is demonstrated in Figure 4(b). After being trained with the bidirectional shape-matching strategy, the networks can render target text with the learned effects to generate appealing results. This method requires only one style example for training and outperforms previous algorithms in deforming the shape of text to match the style image, as shown in Figure 2(c), thanks to their bidirectional shape-matching strategy. However, one limitation is that Shape-Matching GAN is a per-style-per-model method, which means it needs retraining when facing a new style.

Controllable dynamic text effect transfer

Shortly after the work in [10], Yang et al. extended the previous method to the dynamic text effects transfer algorithm Shape-Matching GAN++ [11]. Shape-Matching GAN++ characterizes the short-term consistency of motion patterns via shape matching within consecutive frames. By repeatedly predicting the next frame according to a few previously generated frames, it achieves effective long-term consistency. Shape-Matching GAN++ can generate appealing artistic text animation that characterizes large-scale motion patterns while preserving temporal consistency at the same time. It shares the same limitations as Shape-Matching GAN [10].

Joint font and text effect transfer methods

The above-mentioned algorithms, whether nondeep or deep based, focus mainly on transferring artistic effects (such as color distributions and texture) while keeping the font unchanged. However, the font shape of a reference word art usually contains rich aesthetic implications that should not be neglected. To capture the full style information, some subsequent works are proposed to jointly transfer both font and text effects. Since it is rather difficult to extract the characteristics of a font with only one sample, these joint font and text effect transfer methods usually need several reference source effect images as input. Some of the methods are introduced and discussed in this section.

Few-shot transfer for English alphabet

Multicontent GAN (MC-GAN) [12] is the first end-to-end solution for synthesizing ornamented glyphs. The authors divided the problem into two parts, modeling the overall glyph shape and then synthesizing the final appearance with color and texture. To enable this, they developed a stacked conditional GAN to predict the coarse glyph shapes and an ornamentation network to predict the color and texture of the final glyphs. The two networks are trained jointly and specialized for each typeface using a very small number of observations. One limitation is that MC-GAN is specially designed for processing English letters, and the input is limited to concatenated images of English letters. Besides, the ornamentation network performs in a per-style-per-model fashion, which requires retraining for every new style.

Few-shot transfer for arbitrary glyph

Some language systems (e.g., Chinese, whose official character set GB18030 contains 27,533 characters) contain massive characters that disable character-set-specified methods like MC-GAN [12]. To overcome this limitation, Gao et al. developed the first few-shot learning algorithm, AGIS-Net [13], to transfer both shape and texture styles to arbitrarily large numbers of characters and generate high-quality synthesis results. The proposed AGIS-Net is a simple yet effective model that exploits two parallel encoder–decoder branches. Apart from this, Gao et al. additionally designed a novel and computationally efficient local texture refinement loss, which is helpful in improving the quality of synthesis results. Example results are demonstrated in Figure 2(d). With only four reference images available, AGIS-Net successfully transfers both fonts and effects and generates appealing word art.

Few-shot transfer for glyphs and nontext objects

MC-GAN [12], AGIS-Net [13], and other related researches mostly treat typeface and visual effects as two isolated attributes, thus requiring multiple subnetworks to separately transfer shape styles and texture styles. From another perspective, Li et al. [14] first proposed that the typeface should be considered with the visual effects as a whole. On this basis, Li et al. developed the simple unified framework FET-GAN [14], which contains only one encoder–decoder branch. Given a few samples in the same style for reference, FET-GAN can effectively translate the original effects of an image to the referenced effects while maintaining the global structure unchanged. A few-shot fine-tuning strategy is additionally designed for generalizing to unseen effects. Except for the superior synthesis performance, another exclusive advantage is that FET-GAN can be generalized to nontext objects of any shape since it considers typeface and other effects as a whole.

Dataset and evaluation

Benchmark datasets

There are multiple available benchmark datasets for text style transfer, as summarized in Table 1. These datasets differ mainly in three major aspects: 1) whether font migration is taken into consideration, 2) whether the reference text style contains special elements, and 3) the kinds of character types that are involved. For the first aspect, the first four datasets, MC-GAN-GrayScale [12], MC-GAN-Color [12], AGIS-Net-C [13], and AGIS-Net-P [13], consider fonts along with text effects. For the second aspect, only TextEffects-Decor [9] collects effects with decorative elements; examples are shown in Figure 2(b). For the third aspect, TET-GAN [8], TE141K-C [15], and TE141K-S [15] contain characters of multiple languages, while others only collect either English letters or Chinese characters. Among all of the datasets, AGIS-Net-P has the widest character types, and MC-GAN-Color has the largest number of text styles.

Performance evaluation

Assessing the quality of stylized text is of great value in allowing users to efficiently search for high-quality results as well as guiding the designing of text style transfer algorithms. However, not enough attention has been paid to quality assessment. The evaluation of

Table 1. A summary of the benchmark datasets for the text effect transfer task.

	Dataset Type	Style	Glyph	Style Type	Glyph Type	Images
MC-GAN-GrayScale [12]	Joint font and effects	10,000	26	Gray-scale fonts for English letters	English letters	260,000
MC-GAN-Color [12]	Joint font and effects	20,000	26	Colorful fonts for English letters	English letters	520,000
AGIS-Net-C [13]	Joint font and effects	2,460	639	Synthetic artistic fonts	Chinese characters	1,571,940
AGIS-Net-P [13]	Joint font and effects	35	7,326	Professional-designed fonts	Chinese characters	256,410
TET-GAN [8]	Text effects	64	837	Text effects collected from websites	775 Chinese characters, 52 English alphabets, 10 Arabic numerals	53,568
TextEffects-Decor [9]	Text effects	60	988	Text effects with decorative elements	52 English letters of 19 different fonts	59,280
TE141K-E [15]	Text effects	67	988	Professional-designed text effects	52 English letters of 19 different fonts	66,196
TE141K-C [15]	Text effects	65	837	Professional-designed text effects	775 Chinese characters, 52 English alphabets, 10 Arabic numerals	54,405
TE141K-S [15]	Text effects	20	1,024	Professional-designed text effects	56 special symbols, 968 letters in Japanese, Russian, and so on	20,480

text style transfer algorithms still remains an open and important problem in this field.

In general, there are two major types of evaluation methodologies: qualitative evaluation and quantitative evaluation. The most widely used qualitative evaluation is the aesthetic judgments of observers. Most research studies evaluate different algorithms by comparing their corresponding user preference ratio. The evaluation results are related to many factors (e.g., the age, gender, and occupation of participants). Therefore, this metric is subjective, and it is difficult to reproduce a user study result. For quantitative evaluation, there is currently no appropriate metric. TE141K [15] and other studies mainly choose widely used image quality assessment metrics as an alternative, such as the structural similarity index, peak signal-to-noise ratio, and perceptual loss [16]. These metrics are designed for natural images and therefore cannot effectively evaluate the quality of text effect transfer results.

To make up for the lack of quantitative evaluation metrics in this field, Yan et al. [17] provided a solution that can automatically assess the quality of stylized glyphs, where a multitask attentive network is proposed to imitate the human visual evaluation process. The network is trained to accomplish multiple tasks of style autoencoding, content autoencoding, stylization, and destylization, through which it learns to characterize robust style features and content features. Furthermore, visual attention modules are incorporated to simulate the process of human high-level visual judgment, paying more attention to areas of interest. This model can serve as an effective tool to assist quality assessment. The limitation is that the text effects are too diverse to cover with one dataset. As a result, this method can only handle a limited number of styles and therefore lacks applicability.

Applications

Because of the visually plausible stylized results and wide application scenarios, the text style transfer research and algorithms have already led to many successful industrial applications and begun to deliver commercial benefits. In this section, we summarize some applications and propose some potential usages.

Choose from style libraries

To ensure the visual quality of generated text with various artistic effects, most existing word-art creation tools do not support user-specified styles. Instead, they construct a large style library and allow users to choose styles according to their preferences. Microsoft Office provides users with convenient operation and a large number of fonts with special effects. Adobe Spark, a more powerful tool, contains dynamic effects and more professional-designed styles. It also supports creating customized posters with stylized text.

Create with user-specified style

For a broader application scope, another kind of tool focuses on synthesizing artistic text with arbitrary user-specified styles. For fonts, FlexiFont supports constructing a user-specified Chinese typeface with several samples provided by users. For text effects, because of the difficulty in collecting datasets and the limited number of reference works, there are no popular applications that apply the technique of transferring arbitrary reference text effects. However, with the development of text effect transfer algorithms and increasing user demand, we believe that it will be a promising potential creation tool in the future. Some other fully functional tools can also serve as an alternative; e.g., Photoshop and Adobe Spark are capable of creating customized stylized text under simple user operations. Another potential usage is to assist painters and designers in creating word art, especially when working on computer-made artistic text.

Graphic design

Artistic text effect transfer can also be potentially used in graphic design. UT-Effect [3] provides an effective tool for combining stylized text and background images. Combining this method with visual–textual layout generation [18] may enable automatic poster design.

Future challenges

Although the advances in the field of text style transfer are inspiring, and the current algorithms are capable of generating satisfactory results, there are still several challenges and open issues. In this section, we summarize some key challenges and discuss possible strategies to solve them.

Evaluation methodology

Performance evaluation for stylized text is an important issue and is becoming increasingly critical as text effect transfer algorithms constantly develop. On the one hand, there is no standard benchmark dataset. Existing datasets suffer from insufficient style or character types. Moreover, different researchers typically use their own collected datasets, making it difficult to compare the performance of different text effect transfer methods. More efforts in collecting a large-scale benchmark dataset are necessary. On the other hand, as stated in the section "Performance Evaluation," existing assessment metrics are not appropriate for text effect transfer. User studies are frequently adopted, but the result varies greatly among different observers and is hard to reproduce. More reliable evaluation criteria are needed to assess the performance of text effect transfer approaches.

Time complexity and generalization ability

The existing methods cannot achieve a high processing speed and strong generalization ability at the same time. The nondeep algorithms introduced in the section "Nondeep Text Effect Transfer Methods" have slow processing speeds caused by the iterative optimization process. The deep-based feedforward methods in the sections "Deep-Based Text Effect Transfer Methods" and "Joint Font and Text Effect Transfer Methods" all learn text effects from training data and thus are incapable of transferring unseen styles without specialized fine-tuning schemes. To achieve both a low time complexity and strong generalization ability, perhaps a feedforward algorithm that can generalize to arbitrary styles could be designed with more flexible style transfer modules and larger datasets.

Conclusions

Over the past several years, the topic of text style transfer has gained wide attention and become an inspiring research area. Numerous works, whether optimization based or deep based, have been proposed to achieve surprising results in each subtopic. Despite the great progress and achievements in recent years, the area of text style transfer is far from maturity. A lack of reliable evaluation metrics, inconsistent benchmark datasets, and other issues still remain to be solved in future research. We believe that subsequent studies will solve these existing problems and continuing to contribute to this developing research area.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Contract 62172020 and was a research achievement of Key Laboratory of Science, Technology, and Standard in Press Industry (Key Laboratory of Intelligent Press Media Technology). The corresponding author is Jiaying Liu.

Authors

Xinhao Wang (wxh0510@pku.edu.cn) received his B.S. degree in intelligence science at the Wangxuan Institute of Computer Technology, Peking University, Beijing, 100080 China, in 2022. His current research interests include style transfer and deep learning.

Shuai Yang (shuai.yang@ntu.edu.sg) received his B.S. and Ph.D. degrees (Hons.) in computer science from Peking University, Beijing, China, in 2015 and 2020, respectively. He is currently a postdoctoral research fellow with the Artificial Intelligence Corporate Laboratory, Nanyang Technological University, Singapore 639798 Singapore. He was a visiting scholar with Texas A&M University from September 2018 to September 2019. He received the IEEE International Conference on Multimedia and Expo 2020 Best Paper Award and the IEEE International Workshop on Multimedia Signal Processing 2015 Top10% Paper Award. His current research interests include image stylization and image generation. He is a Member of IEEE.

Wenjing Wang (daooshee@pku.edu.cn) received her B.S. degree in data science from Peking University, Beijing, 100080 China, in 2019, where she is currently pursuing her Ph.D. degree at the Wangxuan Institute of Computer Technology. Her current research interests include image enhancement, image synthesis, and deep learning. She is a Student Member of IEEE.

Jiaying Liu (liujiaying@pku.edu.cn) received her Ph.D. degree in computer science from Peking University. She is currently an associate professor (Boya Young Fellow) with the Wangxuan Institute of Computer Technology, Peking University, Beijing, 100080 China. She has authored numerous articles and holds 60 patents. She has served on the Multimedia Systems and Applications Technical Committee and Visual Signal Processing and Communications Technical Committee in the IEEE Circuits and Systems Society. She has been associate editor of *IEEE Transactions on Image Processing, IEEE Transactions on Circuits and Systems for Video Technology,* and *Journal of Visual Communication and Image Representation.* Her current research interests include multimedia signal processing, compression, and computer vision. She is a Senior Member of IEEE.

References

[1] Y. Wexler, E. Shechtman, and M. Irani, "Space-time completion of video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 463–476, 2007, doi: 10.1109/TPAMI.2007.60.

[2] S. Yang, J. Liu, Z. Lian, and Z. Guo, "Awesome typography: Statistics-based text effects transfer," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 7464–7473.

[3] S. Yang, J. Liu, W. Yang, and Z. Guo, "Context-aware unsupervised text stylization," in *Proc. ACM Int. Conf. Multimedia*, 2018, pp. 1688–1696, doi: 10.1145/3240508.3240580.

[4] L. Xu, Q. Yan, Y. Xia, and J. Jia, "Structure extraction from texture via relative total variation," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 1–10, 2012, doi: 10.1145/2366145.2366158.

[5] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, 2012, doi: 10.1109/TPAMI.2012.120.

[6] Y. Men, Z. Lian, Y. Tang, and J. Xiao, "A common framework for interactive texture transfer," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6353–6362.

[7] Y. Men, Z. Lian, Y. Tang, and J. Xiao, "DynTypo: Example-based dynamic text effects transfer," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5870–5879, doi: 10.1109/CVPR.2019.00602.

[8] S. Yang, J. Liu, W. Wang, and Z. Guo, "TET-GAN: Text effects transfer via stylization and destylization," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 1238–1245, doi: 10.1609/aaai.v33i01.33011238.

[9] W. Wang, J. Liu, S. Yang, and Z. Guo, "Typography with decor: Intelligent text style transfer," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5889–5897.

[10] S. Yang, Z. Wang, Z. Wang, N. Xu, J. Liu, and Z. Guo, "Controllable artistic text style transfer via shape-matching GAN," in *Proc. Int. Conf. Comput. Vis.*, 2019, pp. 4442–4451.

[11] S. Yang, Z. Wang, and J. Liu, "Shape-matching GAN++: Scale controllable dynamic artistic text style transfer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3807–3820, 2021, doi: 10.1109/TPAMI.2021.3055211.

[12] S. Azadi, M. Fisher, V. G. Kim, Z. Wang, E. Shechtman, and T. Darrell, "Multi-content GAN for few-shot font style transfer," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7564–7573.

[13] Y. Gao, Y. Guo, Z. Lian, Y. Tang, and J. Xiao, "Artistic glyph image synthesis via one-stage few-shot learning," *ACM Trans. Graph.*, vol. 38, no. 6, pp. 1–12, 2019, doi: 10.1145/3355089.3356574.

[14] W. Li, Y. He, Y. Qi, Z. Li, and Y. Tang, "FET-GAN: Font and effect transfer via k-shot adaptive instance normalization," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 1717–1724, doi: 10.1609/aaai.v34i02.5535.

[15] S. Yang, W. Wang, and J. Liu, "TE141K: Artistic text benchmark for text effect transfer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3709–3723, 2020, doi: 10.1109/TPAMI.2020.2983697.

[16] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 694–711, doi: 10.1007/978-3-319-46475-6_43.

[17] K. Yan, S. Yang, W. Wang, and J. Liu, "Multitask attentive network for text effects quality assessment," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2020, pp. 1–6, doi: 10.1109/ICME46284.2020.9102871.

[18] X. Yang, T. Mei, Y.-Q. Xu, Y. Rui, and S. Li, "Automatic generation of visual-textual presentation layout," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 12, no. 2, pp. 1–22, 2016, doi: 10.1145/2818709.

SP