

RESEARCH

Open Access

Medical image processing with contextual style transfer



Yin Xu, Yan Li and Byeong-Seok Shin*

*Correspondence:
bsshin@inha.ac.kr
Department of Electrical
and Computer Engineering,
Inha University, 100 Inha-ro,
Michuhol-gu, Incheon 22212,
Korea

Abstract

With recent advances in deep learning research, generative models have achieved great achievements and play an increasingly important role in current industrial applications. At the same time, technologies derived from generative methods are also under a wide discussion with researches, such as style transfer, image synthesis and so on. In this work, we treat generative methods as a possible solution to medical image augmentation. We proposed a context-aware generative framework, which can successfully change the gray scale of CT scans but almost without any semantic loss. By producing target images that with specific style / distribution, we greatly increased the robustness of segmentation model after adding generations into training set. Besides, we improved 2–4% pixel segmentation accuracy over original U-NET in terms of spine segmentation. Lastly, we compared generations produced by networks when using different feature extractors (Vgg, ResNet and DenseNet) and made a detailed analysis on their performances over style transfer.

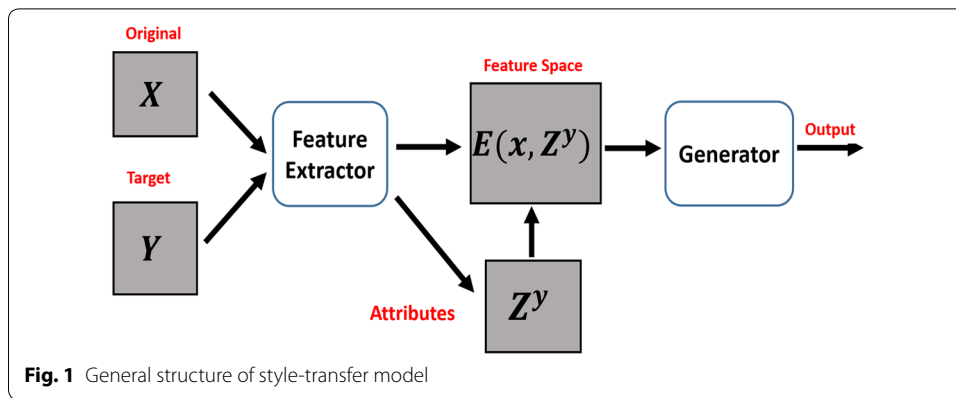
Keywords: Medical Image, Contextual transfer, Deep learning, Segmentation

Introduction

Style transfer has received more and more attention in the area of image processing. Literally, it means using certain methods to change the style of original images. Then with stylized generations, we can make many interesting applications, such as colorization [1], young to old [2]. When looking backing on its development history, we can conclude that most of them following such pipeline:

As demonstrated in Fig. 1, assuming we are expected to transfer the style of original input X into that of Y :

- Firstly, X is feed into a feature extractor, which help us access to its feature space (or latent space)
- Then we also input the target Y into the same extractor. But differ from that from X , we see Y 's feature space as a set of its style attributes, controlling the “destination” of style transformation.
- Having absorbing style information from Y (with mathematical operation), the mixed feature set would be input into a generative model to produce the generations.



Despite so many research and studies have been put in this area, there are three points should be noted: (1) The choice of feature extractor, whether the network you use can totally preserve contents of X , if not, it can be predicted that the final output would be a little different from X in terms of semantics. (2) The way when extracting style attributes from Y . Mathematically speaking, all information extracted from images with network is a set of vectors, the problem is how to quantitative the “style” attributes from complicated signal set. (3) The generative method. A good generative can not only improve the quality of generations but also bring better industrial prospective. In this section, we would detailly discussion possible solutions to challenges of research of style transfer facing.

Motivated by developments of deep learning advances, great progresses have been achieved by applying AI-based techniques. With studies of network architectures, there are more and more encoder-like models used to hele people access, understand, even control the feature space of input signal. Recent years, models like VGG, ResNet [3], DenseNet [4] are preferred by many studies [5–7]. On the face of it, encoder-like architectures may perform better in term of feature extraction and all research mentioned above originate from studies of object detection. We do not talk about their performances in practical style transfer tasks here but note this trend deserving focus. Differ from detection tasks, transformation effects (quality of generations) and semantic preservation should be given top priority in current relevant research.

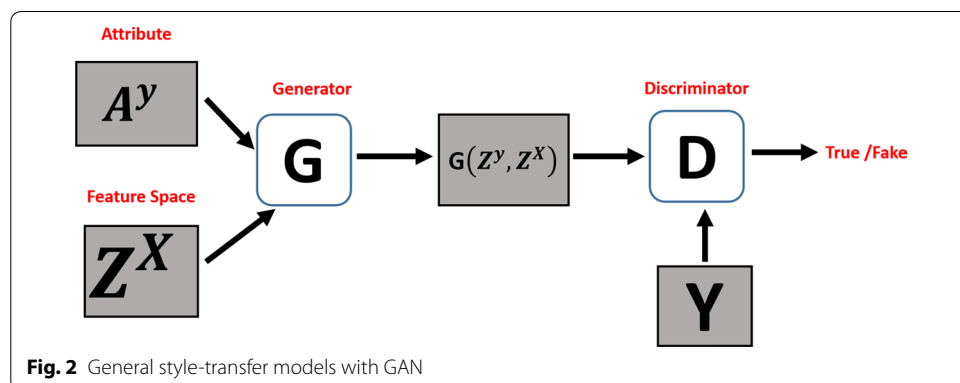
Apart from the choice of feature extractor, the way of learning style information during transformation is also received great attention. Learning from adapted loss function seems a potential solution, Gatys [8]. etc. introduced a layer-wise style loss, they provided a layer list where saved general style information, trying to learning “style” attribute from loss function. On the hand other hand, Huang [9] etc. attempted to acquire “style” by operating batch normalization. Both types of research achieved state-of-art transformation effects, considering almost all studies learning “style” from network, the only point is that they have different understanding of feature spaces. But it is worth noting that we have consider about the flexibility of solution. It means the speed of processing should be attached with great importance and the case where multiple style learning also deserves attention.

We mentioned that deep learning greatly pushed the development of generative models, especially the appearance of generative adversarial network (GAN).

Researches originated from this technique have been preferred recent years, like CycleGAN [10], UNIT [11]. As shown in Fig. 2, we provide a general structure of GAN-based model, they try to produce final outputs by the adversarial process of the generator and discriminator network. It is true that the quality of generations can be improved by adversarial GAN. But problems of such frameworks are also obvious: long training process and extra computational resource brought by discriminator, also, the difficulty of multi-style transformation should be taken into consideration.

All introductions on current solutions to style transfer are targeting on general images. At present time, applications of this kind of research are also popular with medical image analysis, such as visualization (CT to MRI), diagnosis classification and so on. With better-quality of generations, doctors would make a clearer understanding of patients' condition. But unlike those general ones, preventing context loss should be put at the first place, even a little semantic loss of medical images may result incorrect diagnosis. What's more, there are much more low-resolution but important signals existing in medical images, the preservation of which greatly increase the difficulty of transformation.

In this paper, we would focus on applications of style transfer for medical images. Relevant works about style transfer and medical image processing would be introduced in section II. Then in section III, we propose a context-aware framework for medical image processing, which based on the advances of style transfer. Unlike GAN-based research [12–16], we followed a traditional generative idea which can help avoid unnecessary training costs. To maintain semantics of input as much as possible, we introduced a context-aware loss when training. Besides, to accelerate the speed of processing, the model is designed to learn style by batch normalization operation instead of loss learning, enabling the entire model applicable for massive production. In section IV, we designed a two-part experiment, one for testing the quality of generations while another for organ segmentation after adding outputs into training set. Targeting evaluating the performance of feature preservation, we experimented VGG19, DenseNet and ResNet respectively. Last but not least, we concluded that the proposed framework can produce high-quality of medical images. Apart from it, compared with existing style transfer methods, the proposed framework can improve 2~4% segmentation accuracy of U-NET [17], the highest among current research.



Contributions we made in this paper are as follows:

- Following traditional generative idea, we proposed a style-transfer model for medical images. To prevent possible context loss, we design a context-aware loss [18] to enforce the semantic preservation in transformation process.
- Our model learns target style attributes with introduced Adain [9] (Adaptive Instance Normalization), which enables the model can absorb “style” from single target input and learn multiple styles at the same time.
- Organ (spine) segmentation results [17] showed that our framework largely improved the pixel accuracy after adding outputs into training.
- We made a detailed analysis on current feature extractors in terms of their performance over context maintenance, then we concluded that VGG19 [18] would be the best choice for medical image style transfer.

Related work

In this section, we focus on advanced works about style transfer and introduce its application in medical image analysis:

Style transfer

Style transfer means change the style of given images into another with certain methods. Since Gatys [8] etc. introduced their transformation framework based on convolutional network, using deep learning techniques seem a trend in this area. Similar to image synthesis, current research tends to treat style information as a factor which is independent from contexts. Under such circumstances, the entire transformation job would become an attribute-learning task, aiming to learn target “style” attributes.

Assuming images as a combination of context and style codes, Gatys learned style information by training. With pre-visualization intermediate results, they selected possible layers which totally preserve context and style information. In this way, the whole training loss would be divided into two parts: context loss and style loss. With pre-recorded layer names, Gatys [8]. adapted original one-time loss into a layer-wise one and researchers can control the transformation effects by adjusting the weight of style loss. Such advanced research greatly inspired other studies [9, 19]. Many similar works have been introduced, at the same time, the rapid development of network architecture and training techniques, people began to apply the idea of “transformation” to other tasks, such as colorization [1], data augmentation [15], resolution improvement [20].

Facing the challenge from industrial application, original transformation strategy is complained for long training time and limits. Many unique solutions have been introduced [9, 19]. Actually, it is almost impossible to enable “style” information a single unit, being independent from contexts. Researchers can only simulate “style” vectors and then quantify them with certain mathematical operations.

Inspired by the idea of batch normalization, Huang introduced an adaptive instance normalization layer, which enable feature codes to absorb extra style information by normalization. They performed style transfer inside original feature space, shifting and

matching target style codes in a channel wise way [9, 21]. Huang's work has completed state-of-art transformation effects and we would follow their work and compared with the proposed model in terms of performances over medical image generation.

On the other hand, with the scaled-up size of computation and increasingly complicated image input, traditional pixel-wise context loss has to deal with the threat of industrial demand. Especially when processing medical images, countless issues and anomaly structure are extremely to quantify. To present context loss in a clear way, Mechrez et al. [19] designed a context-aware loss that measured the similarity between images by feature-wise comparisons. Experiments shown that their proposed context loss is more applicable than mean square error (MSE) when dealing with complicated signals. We introduced it as the contextual loss and applied it to balance training.

Medical image analysis

Actually, using generative models to tackle with challenges facing in medical analysis has a long history. It is worth to point out that comparing with common images, medical ones are normally low-resolution and have more noise [22]. Furtherly, when applying generative solutions to medical areas, we should put semantic preservation at the first place, otherwise possible information loss may result misdiagnosis.

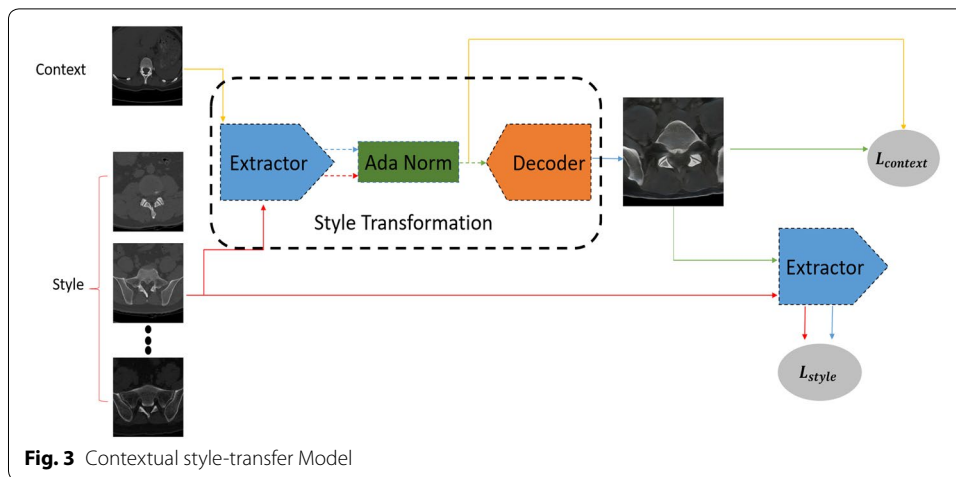
At present time, there are many related studies targeting on medical image style transformation, particularly on specific imaging technologies (CT, MRI etc.). Yang [22] built a GAN-based model and successfully transfer low-dose CT scans to high-dose ones, which greatly reduced the noise from original images. Besides, style transfer techniques also play a considerable role in data augmentation. To increase the diversity of size of training set, Frid [15] introduced a multi-resolution transformation model and added generations into training, greatly improved the accuracy of liver lesion classification.

All medical research mentioned above are based on the idea of "style transfer + generative model". We concluded that main usage of style-transfer technology focus on diagnosis assistance and data augmentation. In this work, we aim to focus on models' performances when processing CT scans, all generations would be used to scale up the training set.

Context-ware style transfer model

Since this research focus on massive medical image processing, particularly cases where multi-style transformation. We followed the traditional generative architecture, using decoding network to avoid extra computational cost, rather than adversarial process.

As illustrated in Fig. 3, the proposed style-transfer model followed similar pipeline that shown in Fig. 1. We let X be context input while $Y : \{Y_1, Y_2 \dots Y_n\}$ stands for images which has target styles we aim to learn. Firstly, both X and Y are feed into the same feature extractor E . Both feature codes (outputs from the extractor) can be represented as $E(X)$ and $E(Y)$. Then with the help of adaptive normalization layer, we align the mean and variance of $E(X)$ to those of $E(Y)$ channel by channel (details would be discussed in next section). In this way, all channels of feature maps in $E(X)$ can learn the style information from $E(Y)$, what help enforce the transformation effects. Then the output $Adain(E(X), E(X))$ would be input into decoder network to produce final generations.



But during training process, having got the output from decoder network, the generations would then be input into extractor again, intermediate results of which could be used to measure the gap between original context and target style. Throughout the entire experiment, we still use two-part loss function (context and style loss, represented as $L_{context}$ and L_{style}).

Considering the specific of medical images, we declare the context preservation of them should be attached with greater importance than that of general ones. Inspired by Mechrez [19], we design a context-aware loss $L_{context}$. In this way, we reduce semantic loss by matching feature vectors in generations with those from $E(x)$. Next, details of context-aware loss and Adain would be introduced.

Adaptive instance normalization

Having got both feature spaces $E(X)$ and $E(Y)$, the rest job is to learning style information from target input as much as possible. According to the original stylization architecture, batch normalization layer is used after each convolution layer, late research began to build a specific normalization operation for style transformation. In this part, we introduce the adaptive instance normalization layer that accelerates the speed of stylization with only single style input.

With given $E(x)$ and $E(Y)$, we try to extract style vectors by contrast normalization. Firstly, we computed the mean and variance of $E(X)$ and $E(Y)$ as follows (denoted with τ and σ respectively):

$$\tau_{NC}(E(Y)) = \frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W E(Y)_{nchw} \tag{1}$$

$$AdaIN(S, T) = \sigma(T) \left(\frac{S - \tau(S)}{\sigma(S)} \right) + \tau(T)$$

$$\sigma_{NC}(E(Y)) = \sqrt{\frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W (E(Y)_{nchw} - \tau_{NC}(E(Y)))^2 + \varepsilon} \tag{2}$$

where we let $E(X)$, $E(Y)$ be 4-D input tensor ($E(X), E(Y) \in R^{N \times C \times H \times W}$), and Let $E(Y)_{nchw}$ denote its inside elements. But differ from traditional BN [], here $\tau(Y)$ and $\sigma(Y)$ are computed cross dimensions independently for each channel.

Then we formulated the layer as:

$$\text{AdaIN}(E(X), E(Y)) = \sigma(E(Y)) \left(\frac{E(X) - \tau(E(X))}{\sigma(E(X))} \right) + \tau(E(Y)) \tag{3}$$

In which we scale up the normalized context feature map set $E(X)$ with $\sigma(E(Y))$, and shift the result with $\tau(E(Y))$. Intuitively, from Eq. (3), we can see the entire normalization do not need any extra learnable weights, indicating faster stylization speed.

As for the benefits of channel-wise computation, we insist that the style of an image results from the intersection of all channels. When detecting a certain style information, $E(Y)$ would produce a high activation when processed with normalization action. The output of AdaIN have the same average activation value for each channel but preserves the context of $E(X)$ at the same time.

Contextual-aware loss

In this section, we would concentrate on models' ability of context preservation. The proposed model follows the setting of two-part loss: context loss and style loss [8]. Considering countless small but complicated signals existing in medical images, even a little mismatch between generations and original inputs may cause possible misunderstanding when diagnosing. Despite many relevant works using pixel-wise MSE [8, 21], we furtherly highlight the importance of preventing semantic loss [19] during transformation process. Besides, we maintain contexts by matching feature vectors instead of pixel values.

Assuming generation G and original input X having the same number of features. Then both them can be defined as:

$$G = \{g_i\}, X = \{x_j\} \tag{4}$$

In which g_i and x_j stand for the feature vectors in $E(G)$ and $E(X)$ and $|G| = |X| = N$. Next, we represent the image similarity between G and T as:

$$CA(G, T) = \frac{1}{N} \sum_j \max_i CA(g_i, x_j) \tag{5}$$

The $CA(g_i, x_j)$ denotes the vector similarity between g_i and x_j . For each x_j , we search all g_i in G to find which is most close to. Then we get average feature similarity value, what can be used to stand for the image similarity between G and X .

As for the details of vector similarity $CA(g_i, x_j)$, we introduce the Cosine distance [22] $CosD_{i,j}$, the distance between g_i and x_j is formulated as:

$$CosD_{i,j} = \left(1 - \frac{(g_i - \mu_x)(x_j - \mu_x)}{\|g_i - \mu_x\|_2 \|x_j - \mu_x\|_2} \right) \tag{6}$$

In which $\mu_x = \frac{1}{N} \sum_j x_j$. When $D_{i,j} \ll D_{i,k}, \forall k \neq j$, we see vector g_i and x_j as similar. Besides, in practical experiment, to quickly find the minimum $CosD_{i,j}$ for each x_j , we start with distance normalization:

$$\widetilde{CosD}_{i,j} = \frac{CosD_{i,j}}{\min_k CosD_{i,j} + \sigma} \tag{7}$$

where the σ ($\sigma = 1e - 5$) denotes a smooth parameter that helps normalization. Next, we turn the distance into the similarity metric by exponentiation:

$$E_{i,j} = \exp \left(\frac{1 - \widetilde{CosD}_{i,j}}{w} \right) \tag{8}$$

The w is a band-width parameter ($w > 0$). Lastly, we adapt the vector similarity into a scale version (for ease of large-scale calculation):

$$CA(g_i, x_j) = \frac{E_{i,j}}{\sum_k E_{i,k}} \tag{9}$$

In this way, the whole image contextual loss [19] between G and T can be formulated as:

$$L_{context}(G, X, L) = -\log (CA(\varphi^L(G), \varphi^L(X))) \tag{10}$$

The parameter φ stands for the feature extractor (would be talked in next section). While L denotes the layer list which is pre-set by feature map visualization.

Context-aware model

As illustrated in Fig. 3, we follow the two-part loss function setting.

$$L = L_{context}(G, X, L) + \vartheta L_{style} \tag{11}$$

where the ϑ is the weight that used for balancing training. As for style loss, with commonly used Gram matrix loss [8] and pre-recorded layer list [19, 21], we define it as:

$$L_{style} = \sum_{i=1}^{\hat{L}} \tau \left(\varphi^{\hat{L}}(g(Y)) - \varphi^{\hat{L}}(G) \right)_2 + \sum_{i=1}^{\hat{L}} \sigma \left(\varphi^{\hat{L}}(g(t)) - \varphi^{\hat{L}}(s) \right)_2 \tag{12}$$

The \hat{L} here is also a list which records possible layers' name that preserve style information of Y.

Experiment and discussion

This work focus style transfer applications on medical images. Supported by Soul National University Hospital, we were given over 50 thousand CT scans of spine, expected to complete organ segmentation with deep learning technology. Soon we found

that there are several gray scales existing among those CT images, even all of them are produced by the same machines and processed by the same staff. Limited by the size of training set, we decide apply the proposed model to increase the diversity of given, furtherly improving the generalizability of original segmentation model [17].

On the other hand, although image semantics can be maintained by setting specified loss function, we observe that either style or context loss are computed over feature maps, which are accessed by extractor networks. It means a reliable feature extractor plays a significant role. Looking back on previous relevant research, we see most of them [3, 4, 18] make object-detection networks as their choice, like VGG19 [18], DenseNet [4] and ResNet [3]. How about their practical performances over medical images? Secondly, the selection context and style layer candidate are determined by the pre-visualization of intermediate results, does it really work?

Facing above challenges, our experiments here are divided into three parts: (1) style transfer (2) semantic segmentation (3) extractor analysis.

Three-part experiment

Style transfer

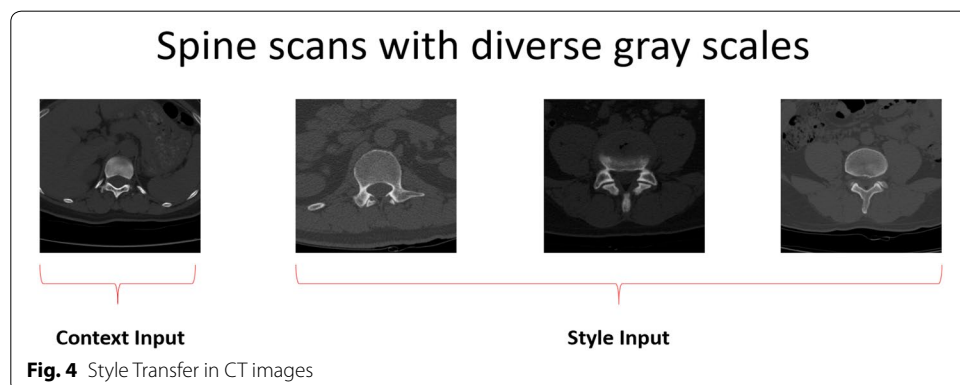
Resulting from the difference of imaging conditions or staff error, CT scans produced by the same machine have different gray scales, which poses a great threat to late processing. If train with such imbalance dataset, it is certain that models have poor performances no matter segmentation or classification.

As shown in Fig. 4, great distribution difference can be observed even with eyes. In this case, the proposed contextual model is expected to learn the style of last three image but produces images with the first one's context.

Semantic segmentation

Our goal is to increase the diversity of dataset and then furtherly improve the generalizability of models that trained with augmented training set.

We pick U-Net [17] as the baseline in this part. By comparing segmentation performances before and after adding generations, we make a clear understanding of usages of style transfer techniques.



Extractor analysis

In previous sections, we mentioned that the selection of context / style layer based on visualization of feature maps. It means the either context or style loss totally relies on the architecture of extractor.

We conclude five extractor architectures mainly used in current works: VGG19, ResNet50, ResNet101, DenseNet121, DenseNet169, DenseNet201. All models have demonstrated good ability of classification and detection. But how about their performance over medical images?

Experiments in this section would focus on encoding. By comparing their transformation performance when used as feature extractor, we try to choose the best architecture for style transfer research.

Baselines and metrics

We have built a three-part experiment, baselines and quantitative metrics are introduced as follows:

Metrics

Style transfer LPIPS distance We aim to produce diverse gray scale CT generations. To better evaluate models' performance over context preservation and style transformation, we introduced the Learned Perceptual Image Patch Similarity distance (LPIPS) distance [13] as a numerical metric. A lower distance indicates greater similarity between paired input (context + style).

Conditional inception score (CIS) This metric can provide a numerical value over images' performances over a classifier [13]. With fine-tuned Inception-V3 [23], a lower CIS value means a poorer ability of style transformation.

Semantic segmentation In the segmentation part, we make pixel accuracy (PA) and the mean intersection over union (MIoU) as our metrics to evaluate models' segmentation ability.

Baselines

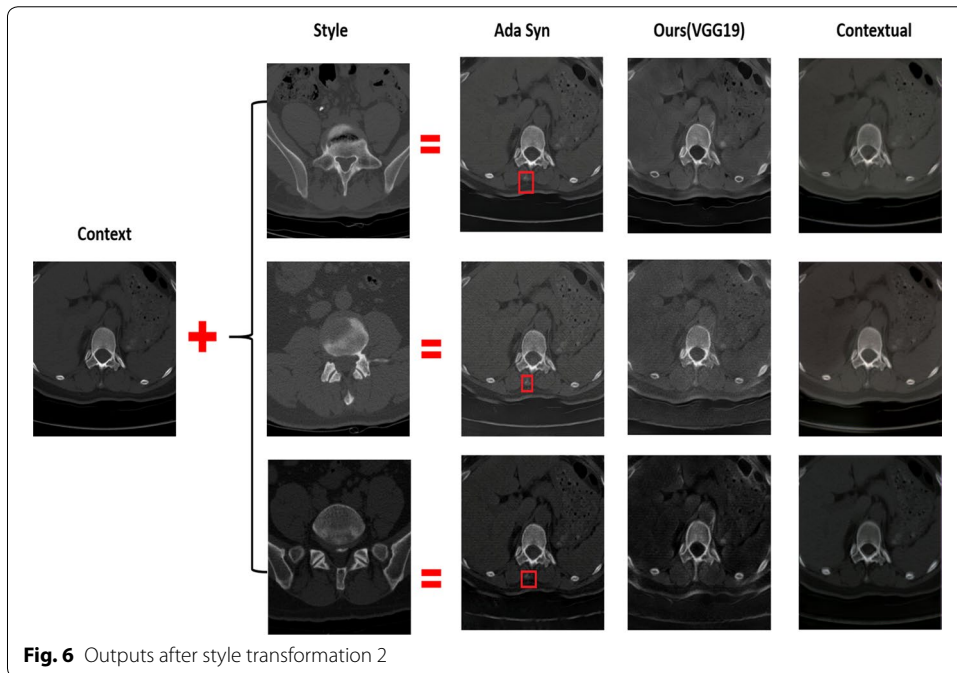
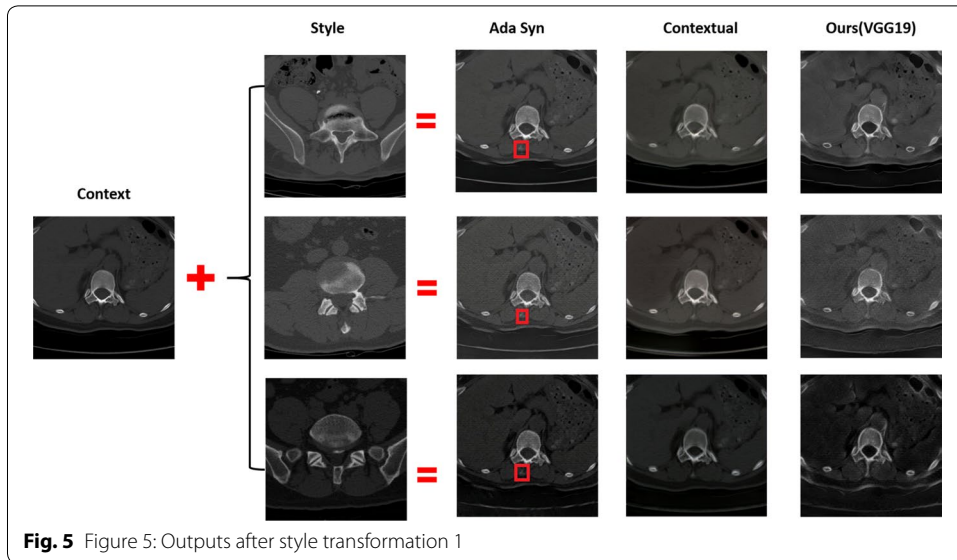
Adain synthesis [9] The AdaIN synthesis model realizes style transfer by using Adaptive method but uses a MSE as context loss.

Contextual transformation [19] This generative model is trained with contextual loss and enables style transfer with unpaired input.

Results and discussions

Style transfer

According to generations produced above (Figs. 5, 6), we found all three generative models above have good performances over medical image style transfer. Comparing with those from Adain method, no great style difference found over generations. But as labeled with red circles, we observe clear semantic loss in generations of Adain [9] that using MSE to compute context loss. While for methods which use context-aware loss (contextual transformation and ours), all semantics from context input are



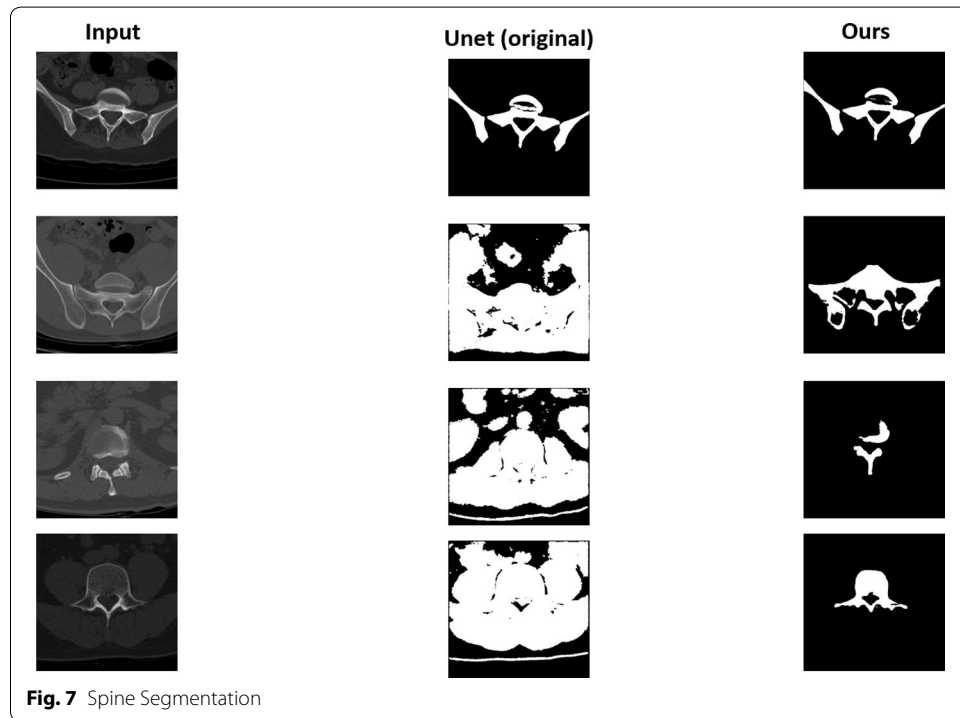
preserved well. But it turns to the style learning, it is clear that outputs from contextual transformation doesn't learn well, with unclear structure and great style difference from those from others.

Table 1 provides the diversity and similarity comparison among three methods. Considering we aim to test the performances of style learning and context preservation, the lower CIS and higher LPIPS mean better style transfer performance over CT image processing.

At the same time, both methods (Contextual transformation [19] and ours) have lower LPIPS value than that using MSE, indicating the contextual loss is better at semantic

Table 1 Numerical comparisons for generations

Methods	LPIPS	CIS
Ours	0.427	0.124
AdalN Synthesis	0.516	0.137
Contextual transformation	0.457	0.214



protection. As for style learning, the Adain method performs better not only at its speed of style normalization but numerical evaluation over CIS. Numerical evaluations in Table 1 confirm with the visual comparison in Figs. 5, 6 and it can be concluded our context-aware style transfer model outperforms existing works over medical image processing.

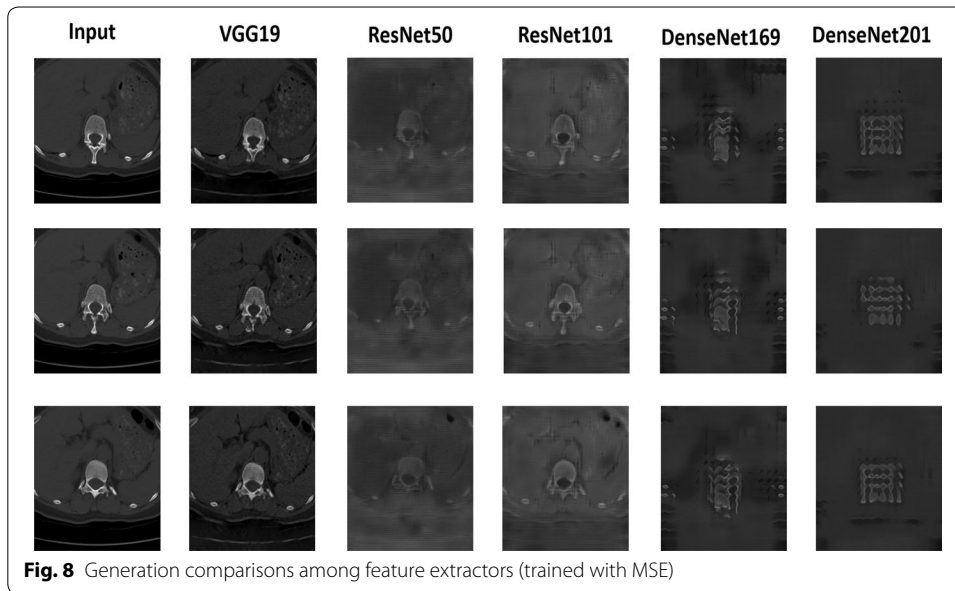
Semantic segmentation

From Fig. 7, a great improvement on segmentation result can be seen after adding generations of style transfer into original training set which has fixed size have single grayscale images, it can be viewed as a kind of data augmentation. It can not only solve class imbalance but improve the generalizability of model. Although U-NET [17] achieved good segmentation results on images following a certain style, it is not able to segment scans that have different grayscales.

As shown in Table 2, PA and MIoU greatly improved after augmentation, indicating a better segmentation quality. It means style transfer techniques can be a potential choice for data augmentation.

Table 2 Segmentation evaluation

Models	PA	MIoU
U-NET	77.71	31.66
U-NET (augmented)	95.80	69.81

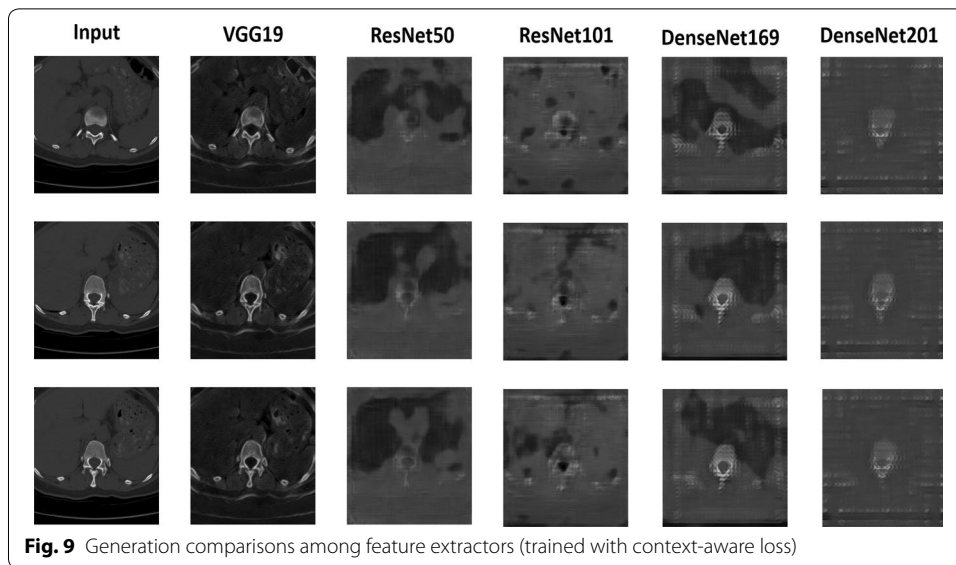


Extractor analysis

We mentioned the choice of feature extractor plays a determining role in research of style transfer, either on the quality of final generations or practical training. But not all encoder-like architectures are applicable for medical images processing, even through some studies have applied them on general images.

In this work, we firstly experimented with VGG19 that has made good achievements on relevant research. Next, we made ResNet/DenseNet-based networks as possible candidates, exploring their performances over medical images.

Figures 8 and 9 demonstrate extractor candidates’ performance over medical image style transfer when trained with MSE and context-aware loss respectively. Considering context/style layers are selected with pre-visualization on feature maps, we assume the way of layer-wise calculation has nothing with final generations. When observing performances above, it is clear that VGG19 performs much better than other two types of candidates (ResNet-based and DenseNet-based), no matter in which loss function they are trained. Despite of style learning, for ResNet50 and ResNet101, both them can barely maintain the structure of context input (spine) during transformation process, at the cost of small issues. While for DenseNet-based networks, semantics of context inputs are totally destroyed, resulting the failure of generations. From experiments in this work, we think VGG19 is the best in term of semantic preservation among all encoder-like architectures.



Conclusion

To summarize, following the traditional style transfer pipeline, we proposed a context-aware generative model. In this model, we design a new loss function that help prevent semantic loss. Also, with introduced adaptive normalization method, we greatly accelerate the speed of stylization and enable the entire model can learn style information from single style input.

Experiments show that our work can produce better quality medical images than existing research. We also treat this work as a new way of data augmentation. With increased data set, we greatly improve the segmentation ability of U-Net.

On the other hand, by experimenting on feature extractors (ResNet50, ResNet101, DenseNet169 and DenseNet201), we find that although ResNet and DenseNet improved that both them have better ability over feature extraction [3, 4] than VGG, VGG19 is still the best feature extractor for medical images.

We conclude that with development of deep learning, encoder-like networks are becoming better and better at extracting high-level signals and using high frequencies to hide low-level ones which they think not important and make all signals imperceptible to humans [24]. It means the encoding ability of neural networks is increasingly improved, that is why people can continually make advances in many advance visual tasks. But it is not a good news for generative research, especially for medical images that have numberless low-signal signals and noises. Up to now, we conclude that VGG19 still be best choice for medical image processing.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (No.NRF-2019R1A2C1090713).

Authors' contributions

Conceptualization: XY, YL, and BSS; Methodology: XY; Conceive and design of the experiment: XY and YL; Writing—original draft: XY; Writing—review and editing: YL and BSS. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Agree.

Consent for publication

Agree.

Competing interests

The authors declare that they have no competing interests.

Received: 4 August 2020 Accepted: 18 October 2020

Published online: 10 November 2020

References

- Zhang R, Isola P, Efros AA (2016) Colorful image colorization. In: Leibe B, Matas J, Sebe N, Welling M (eds) *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part III*. Springer, Cham, pp 649–666
- Galkin F, Aliper A, Putin E et al (2018) Human microbiome aging clocks based on deep learning and tandem of permutation feature importance and accumulated local effects. *bioRxiv*. <https://doi.org/10.1101/507780>
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 770–778
- Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 4700–4708
- Nam H, Kim H-E (2018) Batch-instance normalization for adaptively style-invariant neural networks. *Advances in Neural Information Processing Systems*
- He W, Xie Z, Li Y, Wang X, Cai W (2019) Synthesizing depth hand images with GANs and style transfer for hand pose estimation. *Sensors* 19(13):2919
- Chang B, Zhang Q, Pan S, Meng L (2018) Generating handwritten chinese characters using cyclegan. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, pp 199–207
- Gatys LA, Ecker AS, Bethge M (2016) Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 2414–2423
- Huang X, Belongie S (2017) Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pp 1501–1510
- Zhu J-Y, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pp 2223–2232
- Liu M-Y, Breuel T, Kautz J (2017) Unsupervised image-to-image translation networks. In *Advances in neural information processing systems*, pp 700–708
- Wolterink JM, Dinkla AM, Savenije MHF, Seevinck PR, van den Berg CAT, Išgum I (2017) Deep MR to CT synthesis using unpaired data. In *International workshop on simulation and synthesis in medical imaging*. Springer, Cham, pp 14–23
- Huang X, Liu M-Y, Belongie S, Kautz J (2018) Multimodal unsupervised image-to-image translation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp 172–189
- Zhu J-Y, Zhang R, Pathak D, Darrell T, Efros AA, Wang O, Shechtman E (2017) Toward multimodal image-to-image translation. In *Advances in neural information processing systems*, pp 465–476
- Frid-Adar M, Klang E, Amitai M, Goldberger J, Greenspan H (2018) Synthetic data augmentation using GAN for improved liver lesion classification. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, IEEE, pp 289–293
- Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, Kalra MK, Zhang Y, Sun L, Wang G (2018) Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Trans Med Imag* 37:1348–1357
- Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF (eds) *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015 Proceedings, Part III*. Springer, Cham, pp 234–241
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*
- Mehrez R, Talmi I, Zelnik-Manor L (2018) The contextual loss for image transformation with non-aligned data. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y (eds) *Computer Vision—ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part XIV*. Springer, Cham, pp 768–783
- Zhao C, Ng TK, Wei N, Prabaswara A, Alias MS, Janjua B, Shen C, Ooi BS (2016) Facile formation of high-quality InGaN/GaN quantum-disks-in-nanowires on bulk-metal substrates for high-power light-emitters. *Nano Lett* 16(2):1056–1063
- Li Y, Fang C, Yang J, Wang Z, Lu X, Yang M-H (2017) Universal style transfer via feature transforms. In *Advances in neural information processing systems*, pp 386–396
- Liao H, Xu Z (2015) Approaches to manage hesitant fuzzy linguistic information based on the cosine distance and similarity measures for HFLTSS and their application in qualitative decision making. *Expert Syst Appl* 42:5328–5336
- Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 2818–2826
- Chu C, Zhmoginov A, Sandler M (2017) Cyclegan, a master of steganography. *arXiv preprint arXiv:1712.02950*

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
