2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)
April 3-7, 2020, Iowa City, Iowa, USA

TRANSFER-GAN: MULTIMODAL CT IMAGE SUPER-RESOLUTION VIA TRANSFER GENERATIVE ADVERSARIAL NETWORKS

Yao Xiao[⋆], *Keith R. Peters*[†], *W. Christopher Fox*[‡], *John H. Rees*[†], *Dhanashree A. Rajderkar*[†], *Manuel M. Arreola*[†], *Izabella Barreto*[†], *Wesley E. Bolch*[⋆], *Ruogu Fang*[⋆*]

[⋆] Department of Biomedical Engineering, [†] Department of Radiology, [‡] Department of Neurosurgery
University of Florida, Gainesville, FL, USA

ABSTRACT

Multimodal CT scans, including non-contrast CT, CT perfusion, and CT angiography, are widely used in acute stroke diagnosis and therapeutic planning. While each imaging modality has its advantage in brain cross-sectional feature visualizations, the varying image resolution of different modalities hinders the ability of the radiologist to discern consistent but subtle suspicious findings. Besides, higher image quality requires a high radiation dose, leading to increases in health risks such as cataract formation and cancer induction. In this work, we propose a deep learning-based method Transfer-GAN that utilizes generative adversarial networks and transfer learning to improve multimodal CT image resolution and to lower the necessary radiation exposure. Through extensive experiments, we demonstrate that transfer learning from multimodal CT provides substantial visualization and quantity enhancement compare to the training without learning the prior knowledge.

***Index Terms*—** Image Super-Resolution, Multimodal CT, Transfer Learning, Generative Adversarial Network

1. INTRODUCTION

Multimodal computed tomography (CT) scans, including non-contrast CT (NCCT), CT Perfusion (CTP), and CT Angiography (CTA), are widely used in acute stroke protocols. NCCT scan, with its high dose scan characteristics, provides radiologists comprehensive anatomical brain structures and a general sense of the severity of stroke. The CTP scan is a functional imaging technique that provides substantial information regarding hemodynamics of the brain parenchyma, and is usually conducted by a lower dose setting but much longer acquisition time. The CTA scan provides better visualization of the vasculature system, which is of vital importance in diagnostic decision making. While each imaging modality has its own advantages in brain cross-sectional feature visualization, the varying image quality of different modalities due to different scanning settings impede the ability of the radiologist to discern subtle suspicious findings. Besides, to obtain higher image quality requires higher radiation dose as the image quality is positively correlated to radiation exposure, leading to increases in health risks such as cataract formation [1] and cancer induction [2]. Thus, it is highly crucial to develop an approach to improve multimodal image quality for better visualization in lowering the radiation exposure "as low as reasonably achievable."

Strategies for improving CT image resolution can be summarized in two aspects: hardware-oriented and software-oriented. The hardware-oriented solutions include refining focal spot size x-ray tubes, small image receptors, and better mechanical precision. These sophisticated hardware components are generally hard to adjust and require a longer setup time and upgrade cycle. Therefore, the software-oriented methods are more attractive by reconstructing high-resolution (HR) images directly from the low-resolution (LR) images. As image super-resolution (SR) is an ill-posed inverse problem, how to preserve the critical visual geometry such as edge information and shape details of the image structures are still an open question [3]. It is especially challenging in multimodal CT SR to achieve diagnostic image quality accuracy.

In recent years, deep learning methods, especially generative adversarial network (GAN), achieve realistic textures generation and better visual quality in both natural and medical single image SR [4, 5, 6, 7], providing us an opportunity to reconstruct HR CT images. In this work, we aim to address multimodal CT image SR and demonstrate the feasibility of our GAN based transfer learning in integrating the shared and complementary information from different modalities to achieve high diagnostic image quality.

Contributions: The contributions of this work are three-

[*]Corresponding author, ruogu.fang@bme.ufl.edu. This work is supported by projects funded by the National Science Foundation under Grant No. (NSF 1908299) and the National Center for Advancing Translational Sciences of the National Institutes of Health under the University of Florida Clinical and Translational Science Awards UL1TR000064 and UL1TR001427. The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We also acknowledge the University of Florida Integrated Data Repository (IDR) and the UF Health Office of the Chief Data Officer for providing the analytic data set for this project.

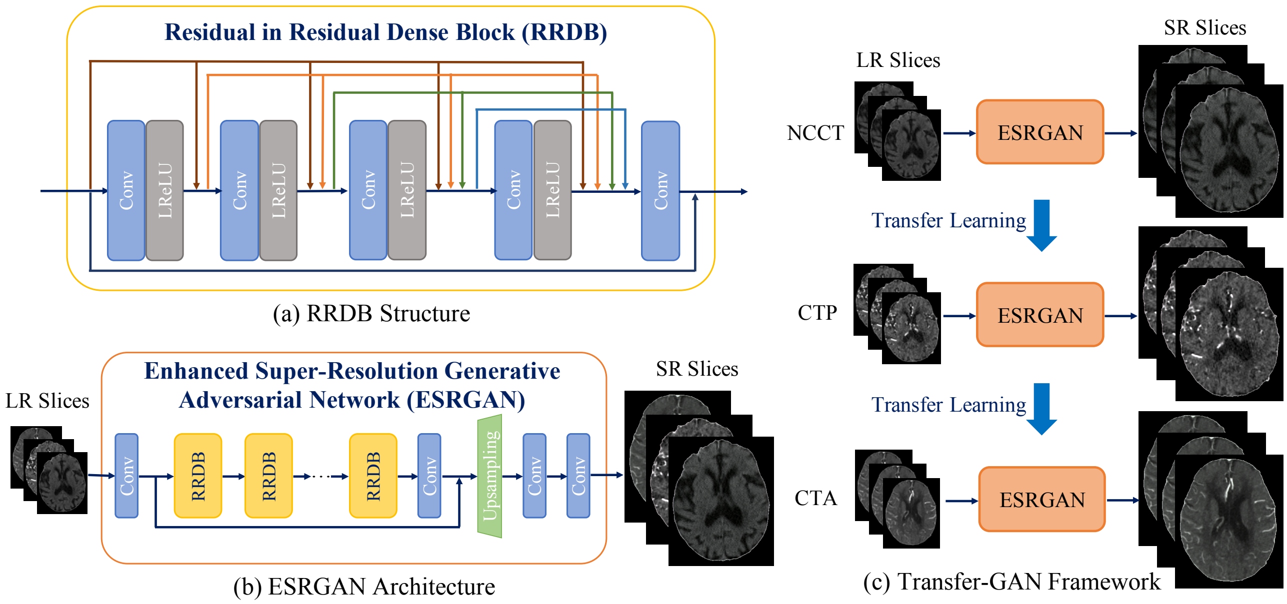

Fig. 1: (a). The structure for Residual in Residual Dense Block (RRDB). (b). The architecture for ESRGAN. (c). The proposed Transfer-GAN framework.

fold: (1) It is the first integration (to the best of our knowledge) of transfer learning with GAN to enhance stroke patient's multimodal CT image quality, where the SR results provide realistic textures and produce comparable visual quality with HR images. (2) The transfer learning strategy explores the complementary information among modalities of the same subject and boosts the performance in multimodal CT SR, and which is a novel solution for image quality enhancement in multimodal CT imaging instead of a single modality. As multimodal CT imaging of the same patient is highly correlated in structure features, the integration of prior knowledge from different modalities is beneficial for achieving high diagnostic image quality. (3) We demonstrate the effectiveness and accuracy of the proposed method by visual comparisons and quantitative evaluation on peak-signal-to-noise-ratio (PSNR) and structural similarity (SSIM) index. The results show that our proposed method can significantly improve the resolution for the images that are four times smaller than the original images.

2. METHODS

We propose Transfer-GAN, a learning-based method by using GAN with transfer learning to produce realistic multimodal CT images to achieve high diagnostic image quality. With the hypothesis that multimodal images from the same patient are highly correlated in structural features, transferring and integrating the shared and complementary information from different CT series can be beneficial for high-resolution multimodal CT image generation. For instance, NCCT, the static anatomical brain imaging modality at a high spatial resolution, can contribute towards the restoration of CTP, a spatial-temporal dynamic imaging modality to capture both the anatomical structure at lower resolutions and blood flow dynamics (a.k.a. the functionality of the brain) over time. As for CTA images that require better vasculature visualization, CTP images at peak perfusion time can provide detailed information about blood flow, which can be useful for enhancing CTA image quality.

The overall design of the proposed Transfer-GAN framework is described in Fig. 1 (c). Inspired by the Enhanced Super-Resolution GAN (ESRGAN) [5], in Fig. 1 (b), which achieves state-of-the-art performance in natural image SR, we continue to explore the application of GAN in the medical imaging domain. The ESRGAN architecture of our Transfer-GAN framework consists of three parts, a generator, a discriminator, and a loss calculator. The aim of the generator of the proposed GAN structure is to synthesize HR images that are similar to the ground truth HR images. We use Residual in Residual Dense Block (RRDB [5]) as our basic building unit in the GAN generator, as more layers and the denser of the connections will boost the performance [8]. In the experiments, we concatenate 23 RRDB blocks. The structure of RRDB can be seen in Fig. 1 (a). This dense block contains convolution (Conv) layer and leaky ReLU (LReLU) only. It consists of 4 pairs of Conv-LReLU layers and a convolution layer at the end. Each Conv-LReLU pair consists of one filter sized of 3×3 Conv layer and followed by one LReLU layer. When $x<0$, the leaky ReLU will remain a negative small slope instead of making the function into zero in ReLU. The relativistic discriminator is used to predict the probability of a real image to determine whether it is relatively more real-

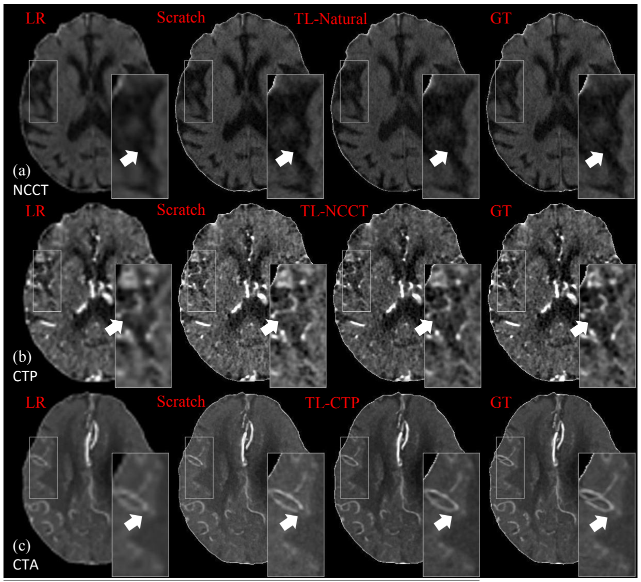

Fig. 2: Visual comparisons of NCCT (a), CTP (b), and CTA (c) images reconstructed by different methods. LR: The low-resolution image from bicubic interpolation. Scratch: No pre-training. TL-Natural: Transfer learning from natural images. TL-NCCT: Transfer learning from non-contrast CT images. TL-CTP: Transfer learning from CTP images. GT: The original high-resolution image.

istic than a fake image. We apply VGG network [9] as our relativistic discriminator, which is based on the idea that comes from [10]. We calculate the perceptual loss by constraining features before activation rather than after activation. Based on the perceptual similarity idea presented in [11, 12], the perceptual loss is defined as the minimized distance between two activated features. However, there are two drawbacks to conventional perceptual loss. The activated features become sparse when the network becomes deeper, which will provide weak activation and lead to inferior performance. Another drawback is that the features after activation may cause inconsistent reconstruction brightness compared to the ground truth image. Therefore, using the features before activation layers is more convincing.

3. EXPERIMENTS AND RESULTS

Our GAN model is evaluated on a dataset with IRB approval and HIPPA-compliant that contains 4,382 images collected from nine stroke patients' multimodal CT images, including 415 NCCT slices, 3,696 CTP slices, and 271 CTA slices. All images are with size 512×512 pixels where they are from the same protocol with the scanning sequence of NCCT, CTP, and CTA. The brain region has a 0.43 mm spatial resolution (in-plane resolution) on the XY-plane. All the CT slices are preprocessed by a brain mask to extract the brain regions out based on the brain window in Hounsfield Units (HU) for different modalities. We randomly split the nine patients into a training set (4 patients), a validation set (2 patients), and a testing set (3 patients). To create LR dataset, we down-sample the images into a quarter of the original size on both spatial dimensions by using the MATLAB bicubic kernel function.

After the preprocessing, we divide the experiments into two stages: using transfer learning and without using transfer learning to compare the performance for each modality. All the experiments are conducted by a GPU workstation that contains four NVIDIA Titan XP (Pascal) GPUs. We set the batch size to 16, and the spatial size of the HR patch to 128 in model training as a larger receptive field is helpful in capturing the semantic details. We terminate the training process when the training iterations reach 20,000 for all modalities.

We firstly train the NCCT, CTP, and CTA images, respectively, with initial weight distributions as the baseline models. As the primary goal of this work is to improve the overall image quality for multimodal CT imaging, we continue the training on the other modalities by fine-tuning on the baseline models. More specifically, following the scanning sequence in acute stroke protocol, we train the TL-CTP model with NCCT images, then fine-tune on the CTP images. For the TL-CTA model, we further fine-tune the CTP model with CTA images; thus, the knowledge learned from the previous modality can be utilized in the following modality image reconstruction. As the NCCT is the first scanning modality, we train the TL-NCCT with natural images (DIV2K dataset [13]) first, then fine-tune on the NCCT images. The experimental results show that the performance has significantly improved from the baseline models.

The model performance is evaluated by both visual and quantitative (PSNR and SSIM) comparisons. As shown in Fig. 2 (visual) and Table. 1 (quantitative), both comparisons show that our method is supremum than training the individual CT modality respectively from scratch, which demonstrates the effectiveness and accuracy of the proposed Transfer-GAN method. The quantity comparison is shown in Table. 1 (a-c), which is calculated as an average result from 184 NCCT, 882 CTP, and 107 CTA test images. The best performance is highlighted in bold font. We perform one-tailed paired t-tests with $\alpha = 0.05$ to compare the performance improvements of PSNR and SSIM for multimodal CT images. Through transfer learning of GAN, there is a significant improvement ($p<0.05$) for both PSNR and SSIM for transferred from NCCT to CTP images than directly training for CTP images, and there is a significant improvement for CTA images by transfer learning from CTP images.

For the visual comparisons in Fig. 2, we enlarge the region of interests and display the enlarged ones on the side. As pointed by the white arrows, we show that the pointed area is better in transfer learning than learning from scratch and LR with bicubic interpolation. The details can be reconstructed clearly with higher contrast, and the edges are preserved much better. Therefore, our experimental results support our hy-

Table 1: Quantity comparisons of NCCT (a), CTP (b), and CTA (c) images reconstructed by different methods. The quantity comparison is calculated as an average result from 184 NCCT, 882 CTP, 107 CTA test images. The best performance is highlighted in bold font. Scratch: Training from random initialization. TL-Natural: Transfer learning from natural images. TL-NCCT: Transfer learning from NCCT images. TL-CTP: Transfer learning from CTP images.

(a)		Scratch	Var	TL-Natural	Var
NCCT	PSNR	23.76	1.79	**27.99**	4.62
	SSIM	0.78	7.4e-3	**0.81**	8.9e-3
(b)		Scratch	Var	TL-NCCT	Var
CTP	PSNR	23.91	1.12	**25.73**	1.21
	SSIM	0.79	3.6e-3	**0.82**	2.8e-3
(c)		Scratch	Var	TL-CTP	Var
CTA	PSNR	25.80	1.90	**26.12**	2.56
	SSIM	0.88	1.4e-3	**0.90**	6.8e-4

pothesis that transferring and integrating the shared and complementary information from different modalities is practical for high-resolution multimodal CT image generation.

4. CONCLUSION

In this paper, we proposed Transfer-GAN, an end-to-end multi-modal image super-resolution network with the transfer learning strategy. The experiment result indicates our approach can improve NCCT image quality by learning from the natural images, can improve CTP image quality by learning from NCCT images, and can improve CTA image quality by learning from CTP images, thus, provides a practical solution for multimodal CT image quality enhancement. This work is the first-time for transfer learning and GAN being integrated for multimodal CT image super-resolution. With the shared and complementary information in NCCT, CTP, and CTA images, integrating the features from different scans are beneficial to achieve high diagnostic imaging quality, which provides a novel solution for image quality enhancement in multimodal CT imaging instead of single modality for the general population. This work also provides a potential solution for maintaining high image quality in support of radiation dose optimization in multimodal CT scanning, providing a safer multimodal CT scan strategy for comprehensive brain imaging.

5. REFERENCES

[1] Gabriel Chodick, Nural Bekiroglu, Michael Hauptmann, et al., “Risk of cataract after exposure to low doses of ionizing radiation: a 20-year prospective cohort study among US radiologic technologists,” *American journal of epidemiology*, vol. 168, no. 6, pp. 620–631, 2008.

[2] Amy Berrington de González, Mahadevappa Mahesh, Kwang-Pyo Kim, et al., “Projected cancer risks from computed tomographic scans performed in the United States in 2007,” *Archives of internal medicine*, vol. 169, no. 22, pp. 2071–2077, 2009.

[3] Jing Tian and Kai-Kuang Ma, “A survey on super-resolution imaging,” *Signal, Image and Video Processing*, vol. 5, no. 3, pp. 329–342, 2011.

[4] Dwarikanath Mahapatra, Behzad Bozorgtabar, and Rahil Garnavi, “Image super-resolution using progressive generative adversarial networks for medical image analysis,” *Computerized Medical Imaging and Graphics*, vol. 71, pp. 30–39, 2019.

[5] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, et al., “ESRGAN: Enhanced super-resolution generative adversarial networks,” in *European Conference on Computer Vision*. Springer, 2018, pp. 63–79.

[6] Chenyu You, Yi Zhang, Xiaoliu Zhang, Guang Li, et al., “CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE),” *arXiv preprint arXiv:1808.04256*, 2018.

[7] Jin Zhu, Guang Yang, and Pietro Lio, “How can we make gan perform better in single medical image super-resolution? A lesion focused multi-scale approach,” *arXiv preprint arXiv:1901.03419*, 2019.

[8] Yulun Zhang, Yapeng Tian, Yu Kong, et al., “Residual dense network for image super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2472–2481.

[9] Karen Simonyan and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[10] Alexia Jolicoeur-Martineau, “The relativistic discriminator: a key element missing from standard gan,” *arXiv preprint arXiv:1807.00734*, 2018.

[11] Leon Gatys, Alexander S Ecker, and Matthias Bethge, “Texture synthesis using convolutional neural networks,” in *Advances in neural information processing systems*, 2015, pp. 262–270.

[12] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *European conference on computer vision*. Springer, 2016, pp. 694–711.

[13] Eirikur Agustsson and Radu Timofte, “NTIRE 2017 challenge on single image super-resolution: Dataset and study,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.