CT2RGB: Seeing The True Colors of Computed Tomography Images

Hung-Ju Liao[*1] HUNGJULIAO@GAPP.NTHU.EDU.TW
[1] *National Tsing Hua University*
Jin-Siang Lin[*1] LINJINSIANG@GAPP.NTHU.EDU.TW
Huan-Chih Wang[2] JESSEHCWANG@NTU.EDU.TW
[2] *National Taiwan University*
Min Sun[1] MINSUN@EE.NTHU.EDU.TW

Editors: Under Review for MIDL 2022

Abstract

Human anatomy plays a pivotal role in modern medicine as many diseases or pathological conditions take the form of anatomical derangements. Nowadays, state-of-the-art medical imaging technologies such as Computed Tomography (CT) are widely used to evaluate these conditions in a noninvasive fashion. However, these medical images are typically displayed as grayscale images, which require a well-trained professional to interpret. Color cues, which are sometimes essential to facilitate diagnosis and preoperative planning, are clearly lacking in these images. We hereby propose a framework to bridge the gap between CT images and cross-sectional cryosection represented by an RGB image. We formulate the problem as an image-to-image translation task, where both the CT and structural information corresponding to critical contours (i.e., fat and bone border) are the inputs, and the cross-sectional RGB image is the output. We train our model with an adversarial training scheme to overcome the lack of paired training input and output data. Moreover, two procedures are designed to force training focusing on the interior of the body and critical contours. Specially, we compute Sobel image gradient in the ranges of CT values corresponding to fat and bone. Although our model is trained on a small dataset due to a lack of training data, our approach is designed to generate realistic cross-sectional color images given unseen CT images. Our experimental results demonstrate the effectiveness of our proposed framework in both quantitative and qualitative aspects for both within the dataset and cross-dataset settings. To investigate the proposed framework's applicability, we also acquire feedback from both doctors and general people by conducting a human perceptual study. The feedback shows a strong potential for educational and clinical communication purposes between doctors and students, and doctors and patients, respectively.

Keywords: Generative Model, Image-to-Image Translation, CT Image, Image Colorization

1. Introduction

Human anatomy is the foundation stone of modern medicine; a better understanding of the anatomy and improved reading of medical images may further translate into a superior clinical care quality, especially during surgical or interventional procedures. Medical students learn human anatomy first by numerous color atlas and correlate the memorized

[*] Contributed equally

knowledge with cadavers or live human bodies (Turney, 2007). Nevertheless, these kinds of anatomical information, especially the color and texture of tissues, could not be obtained from non-invasive imaging modalities preoperatively, most commonly computed tomography (CT) and magnetic resonance imaging (MRI). Cadavers, on the other hand, provide a valuable opportunity for hands-on learning of human anatomy. The extent of anatomical structures, the texture, and the color (although the preservation measures of cadavers might alter them) could be appreciated from cadavers. However, the cadavers could not be indefinitely accessed by medical students and trainees due to their scarcity and cost. As stated previously, learning from the non-invasive images lacks the information of tissue color and texture, which might be troublesome for less-experienced medical students and residents to correlate the image findings with live surgical scenes. It has come to our mind that would it be possible to transform CT images into realistic colorized images that would further help in medical education or even in clinical communication with patients?

In this paper, we aim to transform the CT image into the corresponding cadaveric cross-sectional cryosection represented by an RGB image, which is denoted as **CT2RGB**. This task is similar to Image-to-Image translation (Zhang et al., 2016; Isola et al., 2017; Zhu et al., 2017) with two unique challenges. There is a considerable difference in intensity range between CT and RGB images (i.e., $[-1000 \sim 3096]$ vs. $[0 \sim 255]$), and RGB images of cadaveric cross-sectional cryosection are severely lacking. Inspired by EdgeConnect (Nazeri et al., 2019), we leverage an intermediate representation corresponding to the critical contours to bridge the gap between CT and RGB domains. Specifically, we compute Sobel image gradient in the ranges of CT values[1] corresponding to fat and bone. Our framework combines the Sobel in ranges of CT and the CT images as the input. In this way, the image gradient provides strong guidance for the network to inpaint the human tissue in between fat and bone regions. Moreover, we train our model with an adversarial training scheme to alleviate the lack of paired training data. The trained model is expected to generalize across datasets.

We conduct experiments on the male dataset provided from the Visible Human Project (Ackerman, 1998) and achieve the best performance compared with the other baselines. Furthermore, we test the cross-dataset ability on the female dataset of Visible Human, the DeepLesion (Yan et al., 2018), and CT-ORG (Blaine Rister and Rubin., 2019) dataset. The qualitative results show that our proposed framework can transform CT into RGB images precisely in both within the dataset and cross-dataset settings. In addition, to investigate our approach's applicability, we conduct user studies with both doctors and general people. The feedback indicates that our framework can help facilitate the education of our internal body and communication with patients.

2. Related Work

Generative Adversarial Network. Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) consist of two main modules: the generator and the discriminator. The key to the success of GANs is due to the design of the adversarial loss term (Gui et al., 2021). It trains the discriminator to classify whether an input image is real or not, while

1. CT image is reconstructed based on the tissue density.

simultaneously encouraging the generator to generate a better quality image to fool the discriminator.

Image-to-Image Translation. Problem converting one input representation of an image to another representation can be formulated as an image-to-image translation task. To be more specific, given an input domain X, and an output domain Y, it aims to learn a mapping $G: X \rightarrow Y$. The usage of image-to-image translation is broad, such as image style transfer (Gatys et al., 2016; Isola et al., 2017; Zhu et al., 2017; Johnson et al., 2016), photo colorization (Zhang et al., 2016) and medical image synthesis (Wolterink et al., 2017; Jin et al., 2019).

Medical Image Colorization. Medical images like CT and MRI are mostly rendered in grayscale. However, since colorful images present more evident contour and describe better content information than grayscale ones, humans often recognize RGB images better. Previous works have already applied colorization on medical images (Khan et al., 2017; Zeng et al., 2020). Nevertheless, these works only segment the region of medical images with different colors rather than colorizing medical images to the native color of the human part (e.g., the color of cadavers slides). Compared with previous works, our approach directly learns the mapping between the CT image and the native color of the cadaveric cryosection, which could potentially provide important color cues for faster learning and more effective communication with patients.

3. Method

We propose **CT2RGB** that colorizes a cross-sectional CT slice to the corresponding cadaveric cross-sectional cryosection represented by an RGB image. There are two unique challenges: 1) a considerable difference in dynamic range between CT and RGB images (i.e., $[-1000 \sim 3096]$ vs. $[0 \sim 255]$), and 2) lacking of paired CT and RGB images. In Sec. 3.1, we design preprocessing steps (Fig. 1-(a)) to guidance our model focusing on critical contours which is important under considerable difference in dynamic range. To address

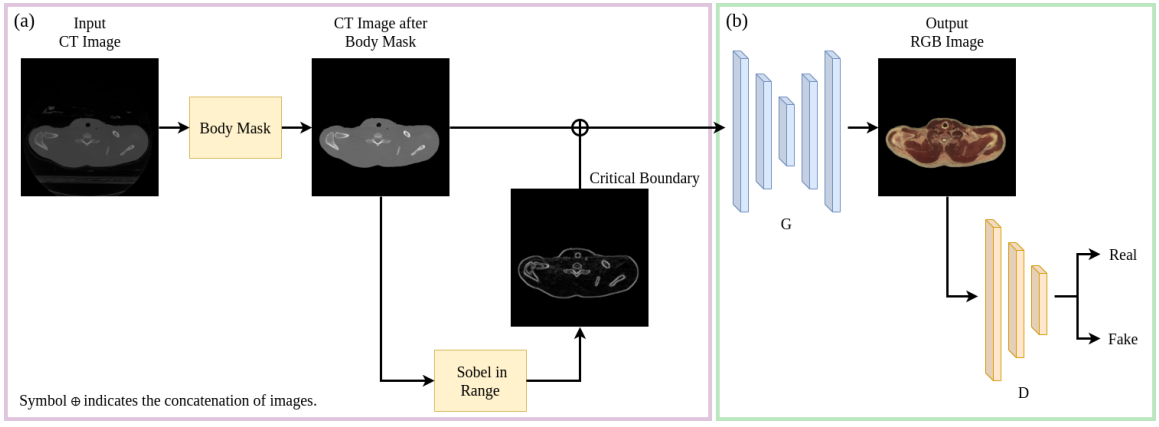

Figure 1: The system architecture of **CT2RGB**.

the lack of paired data, we train a generator G with a discriminator D under several losses including an adversarial loss (see Sec. 3.2 and Fig. 1-(b)).

3.1. Preprocessing

Sobel in Range. Inspired by EdgeConnect, G takes the image gradient of the CT image computed by the Sobel operator (Kanopoulos et al., 1988) (I) as additional input to reduce the difference in dynamic range between CT and RGB images. Instead of using the entire contour of the CT images, we only select the range of the CT values where critical tissues (i.e., fat and bone) locate to compute the image gradient.

Body Mask. Since we only want to generate the RGB value of the human body, we use Otsu's method (Otsu, 1979) to automatically select the corresponding threshold to filter out the background. In the following, the background is ignored during both training and inference.

3.2. Model Training

The generator G translates the input x, concatenation of CT image and the critical contours, into an RGB image. The mapping is denoted as $\hat{y} = G(x, I)$, where $\hat{y}$ is the predicted RGB image. With the adversarial training scheme, the objective of the generator G is to generate an indistinguishable output to confuse the discriminator D, while the discriminator keeps confronting it. The adversarial loss which G tries to minimize while D tries to maximize is formulated as

$$L_{adv}(G, D) = \mathbb{E}_{x,y}[\log(D(x, y))] + \mathbb{E}_{x,y}[\log(1 - D(x, \hat{y}))] \ , \tag{1}$$

where y is the ground truth RGB cadaveric cryosection image.

We also add the perceptual loss (Johnson et al., 2016) to make sure the output image is perceptually close to the ground truth image. We use the ResNet-50 (He et al., 2016) pre-trained on ImageNet dataset (Deng et al., 2009) to extract the perceptual features. The perceptual loss is defined as

$$L_{perc}(G) = \mathbb{E}_{x,y}\left[\sum_{i \in L_p} \|\phi_i(y) - \phi_i(\hat{y})\|_1\right] , \tag{2}$$

where L_p is the set of layers we extract and ϕ_i is the extracted feature of specific layer i in ResNet-50. In order to prevent a degenerated solution, the standard reconstruction loss L_{rec} is used as defined as

$$L_{rec}(G) = \mathbb{E}_{x,y}[\|y - \hat{y}\|_1] \ . \tag{3}$$

The final objective is formulated as

$$\min_G \max_D L_{adv}(G, D) + \lambda_{perc} L_{perc}(G) + \lambda_{rec} L_{rec}(G) \ , \tag{4}$$

where λ_{perc} and λ_{rec} are regularization parameters. In our experiments, we set $\lambda_{perc} = 10$ and $\lambda_{rec} = 10$.

4. Experiments

Dataset and Implementation Detail. We use the dataset of the male cadaver (Spitzer et al., 1996) from the Visible Human Project, which provides pairwise 2D CT images and corresponding cadaveric cross-sectional cryosections. The dataset is divided into two parts: the part of the head and the part of the body. Each part is trained on a distinct model, and both parts were split into 95% for training and 5% for testing. As for the implementation of the network, G and D are based on the pix2pix (Isola et al., 2017), but we change the number of filters of G into 128. To stabilize the training, we replace the adversarial loss in Eq. 1 with the least-squares loss proposed by (Mao et al., 2017). We use Adam optimizer on training with learning rate $= 1 \times 10^{-4}$ for both G and D. The model was trained with a batch size of 4 for 1000 epochs.

Comparison and Evaluation Metrics. We compare our method, **CT2RGB**, with two ablation variants of our method: 1) the method denoted as **CT2RGB-direct**, which directly maps a CT image into an RGB image without "Sobel in range" and 2) the method same as **CT2RGB-direct** but without the adversarial loss term, which is denoted as **CT2RGB-supervised**. We evaluate the quantitative performance using two traditional methods: the root mean square error (RMSE) and the structural similarity (SSIM) index (Zhou Wang et al., 2004). We also measure the perceptual similarity by the learned perceptual image patch similarity (LPIPS) metric (Zhang et al., 2018) (with version 0.1).

4.1. Quantitative Results

Table 1 shows the quantitative results. Although **CT2RGB-supervised** performed well on the traditional evaluation metrics (RMSE and SSIM), it failed at the perceptual similarity (LPIPS) compared to our method (see Fig. 2). Since there is no discriminator to capture the high-frequency structure, the **CT2RGB-supervised** tends to produce much more blurry images (Isola et al., 2017; Pathak et al., 2016). On the other hand, with adversarial training and critical contour information, our approach achieves the best LPIPS performance, which shows that the human perceptual similarity of our generated results is close to the ground truth images.

4.2. Qualitative Results

Fig. 3 (Top) shows the qualitative result of our method. The "Robel in Range" image gradient extracts the critical contour information from the input CT slice such that our generator leverages it to colorize the input CT slice into a corresponding realistic-like RGB

Figure 2: Comparison between our method and **CT2RGB-supervised**.

Table 1: Quantitative result of **CT2RGB-supervised**, **CT2RGB-direct**, and **CT2RGB** on male cadaver from the Visible Human Project. (The text highlighted in **bold** indicates the best results.)

Part	Method	RMSE (×100)	SSIM	LPIPS (×100)
Head	**CT2RGB-supervised**	$\mathbf{1.782 \pm 0.012}$	$\mathbf{0.9838 \pm 0.0003}$	1.192 ± 0.024
	CT2RGB-direct	1.955 ± 0.022	0.9815 ± 0.0004	0.854 ± 0.016
	CT2RGB	1.998 ± 0.010	0.9804 ± 0.0001	$\mathbf{0.837 \pm 0.033}$
Body	**CT2RGB-supervised**	$\mathbf{3.617 \pm 0.023}$	$\mathbf{0.9257 \pm 0.0007}$	5.945 ± 0.016
	CT2RGB-direct	4.239 ± 0.033	0.9033 ± 0.0015	4.267 ± 0.098
	CT2RGB	4.245 ± 0.066	0.9030 ± 0.0010	$\mathbf{3.811 \pm 0.162}$

image. We also compare the qualitative performance of our method with **CT2RGB-direct**. Fig. 3 (Bottom) shows that even though **CT2RGB-direct** achieves competitive quantitative performance, it generated inaccurate details. Since there is a significant difference in the intensity range between CT and color images, these details cannot be recovered without the guide of the contour information. On the other hand, our method colorizes the CT image based on the contour information by leveraging the image gradient, which provides contour cues for our generator. Considering medical images used for educational and clinical communication purposes, they should be precise to avoid misleading the doctors and students or patients, the qualitative results show our method can colorize the CT slices into corresponding RGB cadaveric cryosections meticulously and precisely.

4.3. Cross-dataset Setting

The model trained on the male cadaver is applied on unseen datasets to evaluate its generalization ability. We use the female cadaver from the Visible Human Project, the CT-ORG, and the DeepLesion as the unseen datasets. Fig. 4 shows the qualitative results. The results on Female and DeepLesion are reasonably good. However, some results on CT-ORG suffer more with blurry contour. In order to further improve the generalization performance, unsupervised domain adaptation can be applied and studied in the future.

Evaluation of Body Mask and Sobel in Range. Our preprocessing steps are important for achieving reasonably good generalization ability on Female and DeepLesion datasets. In Fig. 5, we show results in different settings of the body mask and the Sobel in range. "Without Both" uses the median of the CT value in the 10 pixels by 10 pixels square on the top left corner, where the background noises are usually located, as the threshold to filter out the background noise and uses the entire contour of the CT images as the additional input. "Body Mask" is "Without Both" but uses the value found by Otsu's method as the threshold. "Both" uses both Otsu's method and Sobel in range.

4.4. Human Perceptual Study

We acquired the human perceptual study feedback from 13 doctors (all of them are attending physicians) and 36 general people. Given the prediction of our framework with the corresponding ground truth RGB image, the tester had unlimited time to choose which one

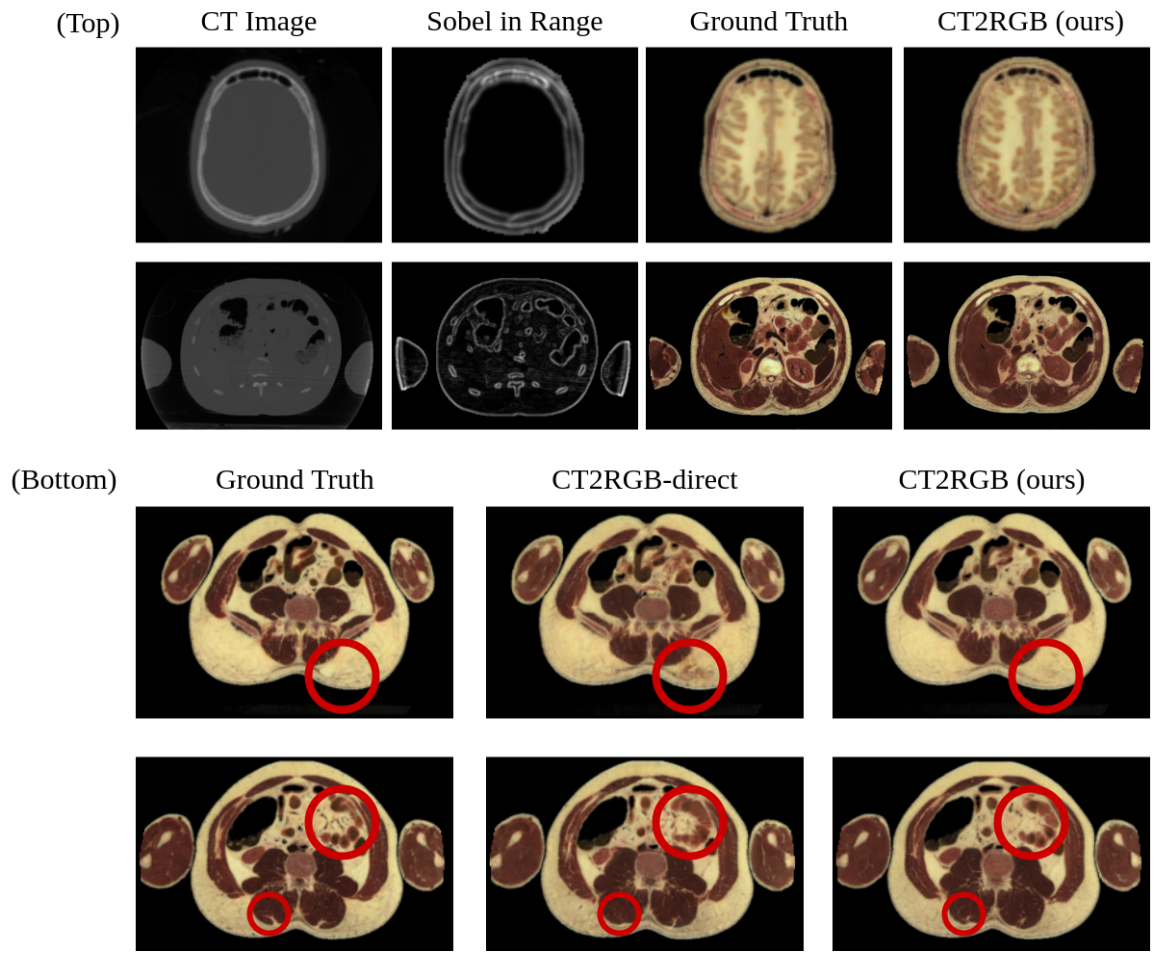

Figure 3: **(Top)** Qualitative results of our method.
(Bottom) Qualitative comparison between our method and **CT2RGB-direct**. The result shows that compared with the **CT2RGB-direct**, our method generates detail much precisely and stays in line with the ground truth RGB output.

is the "real" corresponding RGB image. The output of our framework is realistic enough that the doctors incorrectly choose our prediction image as their answers on **71.4**% of trials; on the other hand, the error rate of the general people is **56.7**%. There are 11 (**84.6**% of) doctors and 27 (**75**% of) people saying that this work helps the doctor to explain patients' health condition to them. Moreover, there are 13 (all of) doctors saying that this technology helps to learn medical images. One doctor mentioned, "Cadaveric cross-sectional cryosection images can help one not familiar with CT images to learn human Anatomy faster." One general person mentioned, "Although I don't understand exactly the RGB images, it encourages me to know more about my own body." We will release the full feedback with the publication.

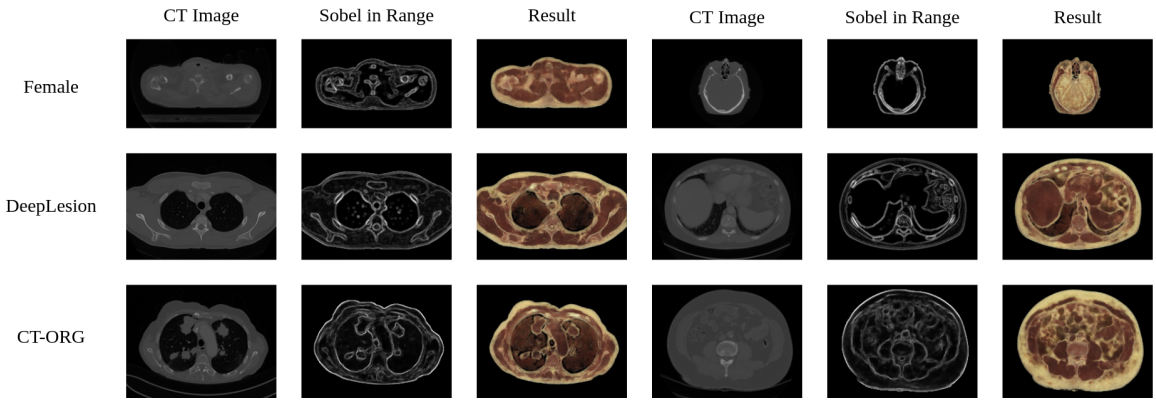

Figure 4: Qualitative results of our method evaluated in the cross-dataset setting.

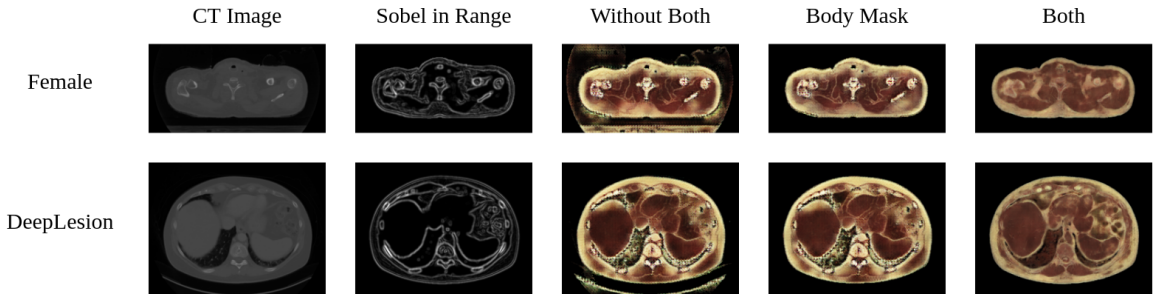

Figure 5: The importance of body mask and Sobel in range in the cross-dataset setting.

5. Conclusion

In this paper, we propose CT2RGB, a novel approach to colorize a 2D CT slice into a corresponding cross-sectional cryosection represented by an RGB image. Our framework extracts the critical contour information from the original input CT image and further leverages it to synthesize the RGB output. The experiment's results demonstrate the effectiveness of our method compared with other ablation variants of our method in both dataset and cross-dataset settings. To prove the potential of our framework, we further conduct a human perceptual study. The feedback shows a strong potential for educational and clinical communication purposes between doctors and students, and doctors and patients, respectively.

References

M. J. Ackerman. The visible human project. *Proceedings of the IEEE*, 1998.

Tomomi Nobashi Blaine Rister, Kaushik Shivakumar and Daniel L. Rubin. Ct-org: Ct volumes with multiple organ segmentations, 2019.

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

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

Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–1, 2021. doi: 10.1109/TKDE.2021.3130191.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

Cheng-Bin Jin, Hakil Kim, Mingjie Liu, Wonmo Jung, Seongsu Joo, Eunsik Park, Young Saem Ahn, In Ho Han, Jae Il Lee, and Xuenan Cui. Deep ct to mr synthesis using paired and unpaired data. *Sensors*, 2019.

Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision (ECCV)*, 2016.

N. Kanopoulos, N. Vasanthavada, and R. L. Baker. Design of an image edge detection filter using the sobel operator. *IEEE Journal of Solid-State Circuits*, 1988.

Muhammad Usman Ghani Khan, Yoshihiko Gotoh, and Nudrat Nida. Medical image colorization for better visualization and segmentation. In *Medical Image Understanding and Analysis*, 2017.

Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *IEEE International Conference on Computer Vision (ICCV)*, 2017.

Kamyar Nazeri, Eric Ng, Tony Joseph, Faisal Qureshi, and Mehran Ebrahimi. Edgeconnect: Structure guided image inpainting using edge prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, Oct 2019.

Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979. doi: 10.1109/TSMC.1979.4310076.

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

Victor Spitzer, Michael J. Ackerman, Ann L. Scherzinger, and David Whitlock. The Visible Human Male: A Technical Report. *Journal of the American Medical Informatics Association*, 1996.

Ben W Turney. Anatomy in a modern medical curriculum. *Annals of the Royal College of Surgeons of England*, 2007.

Jelmer M Wolterink, Anna M Dinkla, Mark HF Savenije, Peter R Seevinck, Cornelis AT van den Berg, and Ivana Išgum. Deep mr to ct synthesis using unpaired data. In *International workshop on simulation and synthesis in medical imaging*, 2017.

Ke Yan, Xiaosong Wang, Le Lu, and Ronald M Summers. Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *Journal of medical imaging*, 2018.

Xianhua Zeng, Shiyue Tong, Yuzhe Lu, Liming Xu, and Zhiwei Huang. Adaptive medical image deep color perception algorithm. *IEEE Access*, 2020.

Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *European Conference on Computer Vision (ECCV)*, 2016.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE International Conference on Computer Vision (ICCV)*, 2017.

Appendix A. Quantitative result on cross-dataset

We present one more quantitative result on the female part of the Visible Human Project as shown in Table 2. (Only it has the axial anatomical images among the cross-datasets we used.)

Table 2: Quantitative result of **CT2RGB-supervised**, **CT2RGB-direct**, and **CT2RGB** on female cadaver from the Visible Human Project. (The text highlighted in **bold** indicates the best results.)

Part	Method	RMSE (×100)	SSIM	LPIPS (×100)
Head	**CT2RGB-supervised**	10.773 ± 0.143	0.7489 ± 0.0034	22.948 ± 0.653
	CT2RGB-direct	11.358 ± 0.206	0.7471 ± 0.0082	26.436 ± 0.633
	CT2RGB	$\mathbf{9.751 \pm 0.033}$	$\mathbf{0.8504 \pm 0.0001}$	$\mathbf{13.092 \pm 0.016}$
Body	**CT2RGB-supervised**	20.731 ± 0.174	0.3678 ± 0.0022	47.109 ± 0.555
	CT2RGB-direct	20.838 ± 0.171	0.3350 ± 0.0009	50.146 ± 0.183
	CT2RGB	$\mathbf{17.860 \pm 0.071}$	$\mathbf{0.6804 \pm 0.0001}$	$\mathbf{29.872 \pm 0.019}$