TU Delft

Delft University of Technology

Local Implicit Neural Representations for Multi-Sequence MRI Translation

Chen, Yunjie; Staring, Marius; Wolterink, Jelmer M.; Tao, Qian

DOI
10.1109/ISBI53787.2023.10230409

Publication date
2023

Document Version
Final published version

Published in
2023 IEEE International Symposium on Biomedical Imaging, ISBI 2023

Citation (APA)
Chen, Y., Staring, M., Wolterink, J. M., & Tao, Q. (2023). Local Implicit Neural Representations for Multi-Sequence MRI Translation. In *2023 IEEE International Symposium on Biomedical Imaging, ISBI 2023* (Proceedings - International Symposium on Biomedical Imaging; Vol. 2023-April). IEEE. https://doi.org/10.1109/ISBI53787.2023.10230409

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

https://www.openaccess.nl/en/you-share-we-take-care

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI) | 978-1-6654-7358-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISBI53787.2023.10230409

LOCAL IMPLICIT NEURAL REPRESENTATIONS FOR MULTI-SEQUENCE MRI TRANSLATION

Yunjie Chen[⋆]*, Marius Staring*[⋆]*, Jelmer M. Wolterink*[†]*, Qian Tao*[⋆,‡]

[⋆]Department of Radiology, Leiden University Medical Center, the Netherlands
[†]Department of Applied Mathematics, Technical Medical Center, University of Twente, the Netherlands
[‡]Department of Imaging Physics, Delft University of Technology, the Netherlands

ABSTRACT

In radiological practice, multi-sequence MRI is routinely acquired to characterize anatomy and tissue. However, due to the heterogeneity of imaging protocols and contra-indications to contrast agents, some MRI sequences, e.g. contrast-enhanced T1-weighted image (T1ce), may not be acquired. This creates difficulties for large-scale clinical studies for which heterogeneous datasets are aggregated. Modern deep learning techniques have demonstrated the capability of synthesizing missing sequences from existing sequences, through learning from an extensive multi-sequence MRI dataset. In this paper, we propose a novel MR image translation solution based on *local implicit neural representations*. We split the available MRI sequences into local patches and assign to each patch a local multi-layer perceptron (MLP) that represents a patch in the T1ce. The parameters of these local MLPs are generated by a hypernetwork based on image features. Experimental results and ablation studies on the BraTS challenge dataset showed that the local MLPs are critical for recovering fine image and tumor details, as they allow for local specialization that is highly important for accurate image translation. Compared to a classical pix2pix model, the proposed method demonstrated visual improvement and significantly improved quantitative scores (MSE 0.86×10^{-3} vs. 1.02×10^{-3} and SSIM 94.9 vs 94.3).

***Index Terms*—** implicit neural representation, MR image translation, generative adversarial network, hypernetwork

1. INTRODUCTION

Magnetic resonance imaging (MRI) is one of the most important image modalities in clinical radiology because of its excellent soft tissue contrast and versatility. One distinguishing advantage of MRI is the availability of various sequences that provide complementary information to help clinicians characterize the anatomy in a comprehensive way. Recent research on MRI-based regression and segmentation also demonstrated the value of multi-sequence MRI in machine learning studies [1]. However, due to the heterogeneity of imaging protocols and contra-indications of contrast agents, some MRI sequences are not always acquired, especially those requiring contrast administration. This may influence the usability of machine learning models originally trained on multi-sequence MRI.

A number of methods have been developed to address this missing sequence problem [2]. The basic idea of earlier attempts is to build a common latent space shared by all modalities such as Hetero-Modal Image Segmentation (HeMIS) [3]. Such methods tend to focus on specific tasks like regression or segmentation and offer accurate results without explicitly generating the missing images, which however damages the reliability and interpretability of the improvement achieved. Nowadays, generative adversarial networks (GANs) are commonly used to directly synthesize missing modalities or MRI sequences from existing images. For instance, Sharma [4] proposed a Multi-Modal Generative Adversarial Network (MM-GAN) which accepts arbitrary MRI sequences as input to generate missing sequences. In spite of the promising results these studies have shown, the synthesis of new information is still less than satisfactory, especially when synthesizing enhanced lesion regions from non-enhanced images [2].

To address these limitations, we introduce implicit neural representations (INRs) as the generator in a GAN framework for multi-sequence MRI translation [5]. INRs is a multi-layer perceptron (MLP) that takes as input a continuous coordinate and provides an estimation of a quantity at that point as output. In addition, we propose to encode prior knowledge from the dataset using a hypernetwork which can enhance generalization and performance [6]. The main contributions of this work are as follows: 1) We propose a GAN-based model using INRs as generator which includes meta-learning and locally specific MLPs to synthesize a T1ce from a T2-weighted image (T2), T1-weighted image (T1) and Fluid attenuation inversion recovery (Flair); 2) We show that our method improves translation performance in the BraTS challenge dataset both qualitatively and quantitatively compared to a traditional GAN which uses a U-Net shaped generator; 3) We design an ablation study and further demonstrate the role of the locally specific MLPs for MRI translation.

Authorized licensed use limited to: TU Delft Library. Downloaded on October 10,2023 at 08:30:39 UTC from IEEE Xplore. Restrictions apply.

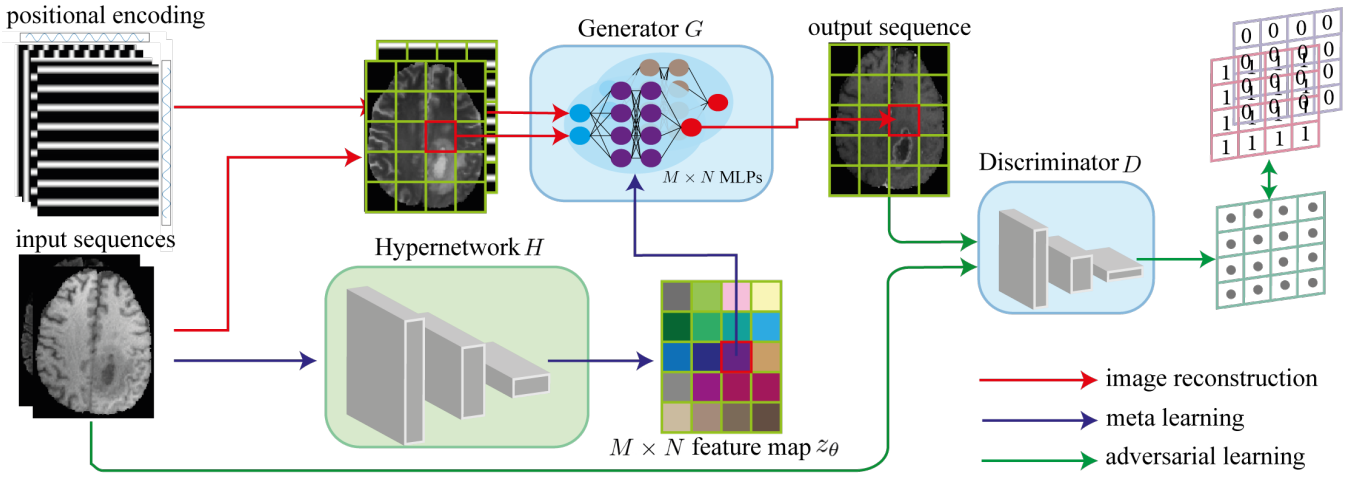

Fig. 1: Illustration of our method. The input image and encoded coordinates are split into patches.$M \times N$ MLPs, which use the output of a hypernetwork as weights, are assigned to each image patch and output a corresponding patch in the target domain. During training, we use a patch discriminator for adversarial learning to improve translation quality.

2. METHODS

2.1. Generative adversarial network

For simplicity, we formulate our method with single input and single-output, but our method can be extended to a multi-input setting by concatenating multiple sequences as a multi-channel input. The goal is to learn a non-linear mapping between two image domains. Given a set of paired images: T2 (domain $\mathcal{S}$) and T1ce (domain $\mathcal{T}$), we aim to learn the mapping $G : s \rightarrow t$, with $s \in \mathcal{S}$ and $t \in \mathcal{T}$. GAN-based models like pix2pix [7][8] are widely used for this:

$$L_{\text{GAN}}(G, D) = \mathbb{E}_{s,t}[\log D(s, t)] + \mathbb{E}_{s,t} \log(1 - D(s, G(s, \boldsymbol{z}))), \quad (1)$$

where G is the generator, D is the discriminator, and $\boldsymbol{z}$ is the latent code which is usually sampled from a prior Gaussian distribution. In a pix2pix model, a U-Net is used as a generator, often with a patch discriminator (PatchGAN) [7]. In the following sections, we introduce the main idea of implicit neural representations and how this technique can be applied to GAN-based multi-sequence MRI translation.

2.2. Implicit neural representations

Recent research on INRs has shown that complex real-world signals can be represented as a continuous function by a fully connected network [9]. In our case, we use an MLP to represent the intensity values of a T1ce. The MLP can be conditioned, such that the function that it represents depends on other information. Here, we condition the MLP on the source image intensity at its input coordinate. Then for each pixel, the image translation mapping function G can be cast into a pixel-wise mapping as follows: $G(\boldsymbol{x}, s_{\boldsymbol{x}}) = t'_{\boldsymbol{x}}$, where $\boldsymbol{x} = (x, y) \in \mathbb{R}^2$ is the 2D coordinate, $s_{\boldsymbol{x}} \in \mathbb{R}$ is the pixel value in the input image and $t'_{\boldsymbol{x}} \in \mathbb{R}$ is the pixel value in the output image. In addition, we introduce a latent code $\boldsymbol{z}$ which encodes prior knowledge which is not present in the input image, e.g. contrast-enhanced information of the tumor. To do this, we use a convolutional neural network H as a hypernetwork with trainable parameters θ which takes the source image s as input, to output the weights used by the MLP. As a result, the mapping function will be conditioned on every different input instance:

$$G(\boldsymbol{x}, s_{\boldsymbol{x}}; \boldsymbol{z}_\theta(s)) = t'_{\boldsymbol{x}}. \quad (2)$$

The overall pipeline is shown in Fig. 1.

2.3. Locally specific MLPs

To achieve a focused local representation, we adopt the idea of multiple locally specific MLPs from ASAP-Net [10], instead of using a single MLP to represent the entire image. As shown in Fig. 1, the input image and coordinates are split into $M \times N$ patches and each patch is assigned an unique MLP. The hypernetwork H downsamples the input image to the same resolution and outputs weights used by these MLPs. This enables these MLPs to be conditioned on the whole input image while still being able to encapsulate unique information from the local region.

Since an MLP with ReLU activation functions that directly operates on Cartesian coordinates has limited representation power due to a spectral bias, we employ a positional encoding $\gamma : \mathbb{R}^2 \rightarrow \mathbb{R}^M$ to map the coordinates into a higher dimensional space [9]. Unlike ASAP-net, in which coordinates have the same positional encoding in each patch, we first globally normalize the coordinates into $[0, 1]$, and

Authorized licensed use limited to: TU Delft Library. Downloaded on October 10,2023 at 08:30:39 UTC from IEEE Xplore. Restrictions apply.

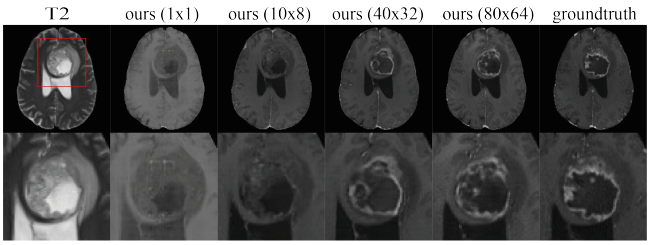

Fig. 2: The influence of choices of $M \times N$. Whole brain MRI are shown in the top row and corresponding zoomed-in results are shown in the bottom row.

then apply sinusoids to each of the two normalized coordinates separately: $\gamma(\boldsymbol{x}) = [\sin(2^0\pi\boldsymbol{x}), \cos(2^0\pi\boldsymbol{x}), \ldots, \sin(2^{i-1}\pi\boldsymbol{x}), \cos(2^{i-1}\pi\boldsymbol{x})]$, with i being the frequency parameter dependant on the task. This provides the network with the ability to fit high-frequency signals, while also giving every position in the image a unique representation.

2.4. Objective

In addition to implicit neural representation, we employ an adversarial learning strategy to further improve the quality of the synthesized image [7]. The hypernetwork and MLPs are jointly optimized using the GAN loss (1). For an image translation task, the output of the generator should be as close as possible to the target image. We use the ℓ_1 loss as the reconstruction loss, which can produce sharper results compared to mean squared error (MSE): $L_{\text{rec}}(G) = \|t^{'} - t\|_1 = \|G(s; \boldsymbol{z}_\theta(s)) - t\|_1$. The final objective function is then:

$$\theta = \arg\min_{\theta}\max_{D} L_{\text{GAN}}(G, D) + \lambda_{\text{rec}} L_{\text{rec}}(G). \quad (3)$$

For more stable training and faster convergence, we add Gaussian noise to the real and fake images before we pass them to the discriminator [11].

3. EXPERIMENTS AND RESULTS

3.1. Dataset

We evaluated our methods on the BraTS2018 challenge dataset [12], which provides co-registered MRI sequences at 1 mm^3 isotropic resolution. Each patient has four sequences: T1, T2, T1ce, and Flair. We randomly selected 296 patients for training and 76 patients for testing. Before training, all image intensities are normalized to [-1,1] and non-zero cropping is performed to reduce the size of images.

3.2. Experiments

In our experiment, we synthesize T1ce using two input settings: single-input (T2) and multi-input (T1, T2, Flair). We compare the performance of our method with pix2pix [7], with the only difference being the generator. The discriminator consists of three 4×4 convolutional blocks with stride 2 and two 4×4 convolutional blocks with stride 1. The hypernetwork contains a downsampling module, seven 3×3 convolutional blocks, and a final convolutional layer to output the desired number of parameters of the MLPs. For a fair comparison, we used an 8-layer MLP (100K parameters) for models with only one global MLP, and a 5-layer MLP (18K parameters) for models with locally specific MLPs. So in total, our models contain 100 million parameters (hypernetwork and MLPs). The models were optimized using Adam with an initial learning rate of 10^{-4} with a step decay schedule for 60 epochs. We set the batch size to 8, λ_{rec} to 100 and i to 6 for postional encoding. During training, random flipping is performed and every image is randomly cropped to 160×128. We quantitatively evaluate our methods using mean squared error (MSE), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR). The results of the two translation models are compared statistically using the Wilcoxon signed-rank test.

Table 1: Quantitative T1ce translation results comparing pix2pix with our methods. $*$ indicates significant difference ($p < .001$) compared to pix2pix using the same inputs.

methods	$M \times N$	MSE ($\times 10^{-3}$) $\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$
pix2pix (m=1)	\	1.02 ± 1.25	94.3 ± 1.3	30.98 ± 2.58
ours (m=1)	1×1	$1.63 \pm 1.42^*$	$93.2 \pm 1.5^*$	$28.76 \pm 2.50^*$
ours (m=1)	10×8	1.05 ± 1.18	94.4 ± 1.3	30.80 ± 2.51
ours (m=1)	20×16	$0.92 \pm 1.20^*$	$94.9 \pm 1.2^*$	$31.55 \pm 2.71^*$
ours (m=1)	40×32	$0.90 \pm 1.11^*$	$94.9 \pm 1.2^*$	$31.61 \pm 2.79^*$
ours (m=1)	$\mathbf{80 \times 64}$	$\mathbf{0.86 \pm 0.91}^*$	$\mathbf{94.9 \pm 1.2}^*$	$\mathbf{31.82 \pm 2.82}^*$
pix2pix (m=3)	\	0.77 ± 0.88	95.1 ± 1.1	32.32 ± 2.85
ours (m=3)	$\mathbf{40 \times 32}$	$\mathbf{0.74 \pm 1.01}^*$	$\mathbf{95.6 \pm 1.2}^*$	$\mathbf{32.76 \pm 3.15}^*$

3.3. MRI translation

We first compared models with different numbers of MLPs. We set $M \times N$ to 1×1 (one global MLP), 10×8, 20×16, 40×32, 80×64 and the results are reported in Table 1. The results show that by using multiple locally specific MLPs, we significantly improve the performance. Compared to the pix2pix model, our model gets comparable results when we set $M \times N$ to 10×8. It can be noted that the models uses more MLPs have higher translation quality and we can get best results when $M \times N$ is set to 80×64. As shown in Fig. 2, the model with 40×32 MLPs and the model with 80×64 MLPs are almost visually equivalent. Given the week-long training time of the 80×64 model using Nvidia Tesla V100, we set $M \times N$ to 40×32 for the subsequent experiments.

The overall quantitative results of models using T2 only ($m = 1$) and using T1, T2, Flair jointly ($m = 3$) as input are reported in Table 1. The results show that our method outperforms pix2pix in all metrics with both input settings. All results between the two models show statistically significant differences $p < .001$ using the Wilcoxon signed-rank test. As

Authorized licensed use limited to: TU Delft Library. Downloaded on October 10,2023 at 08:30:39 UTC from IEEE Xplore. Restrictions apply.

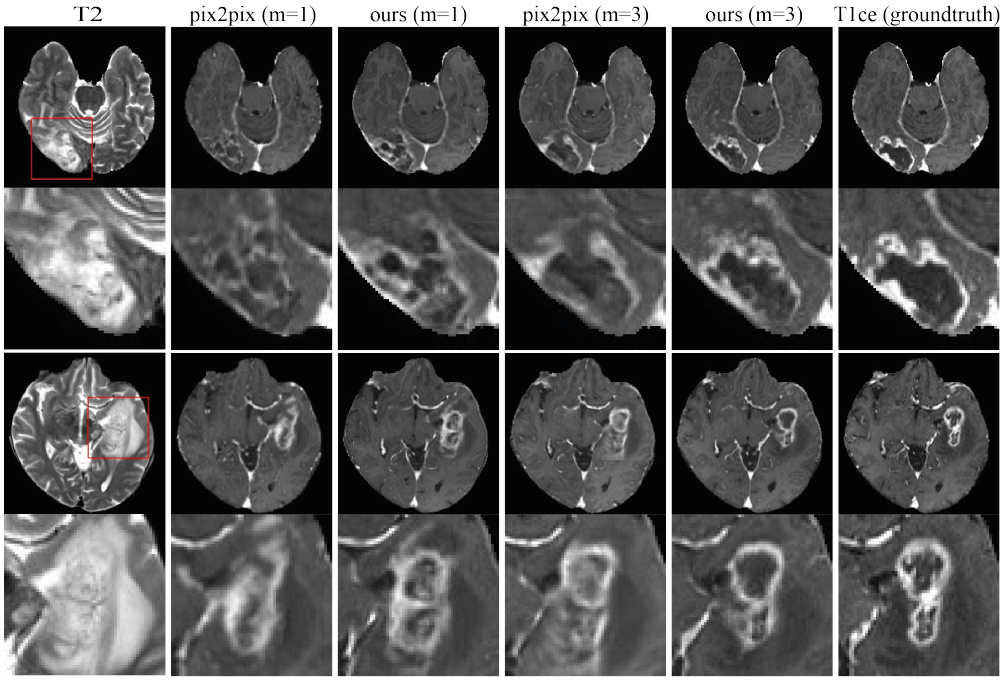

Fig. 3: Visual results of pix2pix and our method. Zoomed-in areas are indicated by red rectangles.

shown in Fig. 3, we can observe that by using INRs, the results are less blurred and exhibit finer details, showing a more obviously enhanced tumor.

3.4. Locality of MLPs

To verify the role of the locally specific MLPs in our model, we performed experiments to forward all pixels through one of the MLPs. As shown in Fig. 4, when selecting an MLP inside the white matter (top row), most of the local details of the image are lost. When selecting an MLP located in the cerebrospinal fluid (CSF) of the ventricles (middle row), the locally learned pattern is repeated periodically. When we select a background MLP (bottom row), the prediction plainly failed. This shows that we get specialized MLPs depending on the local image context in our method. On the contrary, the MLPs using local positional encoding tend to output periodic results which indicates the lack of locality.

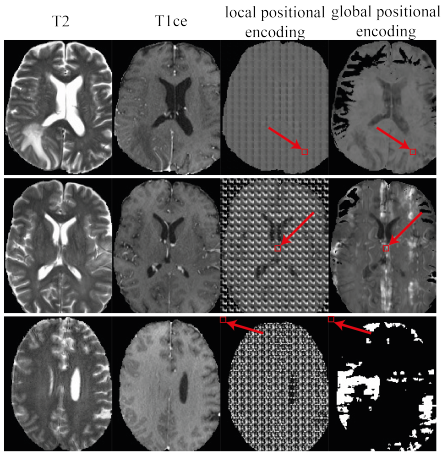

Fig. 4: The role of locally specific MLPs: using a single selected MLP, depicted with red rectangle, for generating the entire image.

4. DISCUSSION AND CONCLUSION

In this paper, we have proposed a novel GAN-based model using implicit neural representations as the generator for multi-sequence MRI translation. Our results on the BraTS dataset show that our methods improved the quality of synthesized T1ce both visually and quantitatively. Compared to the pix2pix method which uses a U-Net shaped generator, our results are overall sharper and recover more details in the enhanced tumor region. We designed an ablation study that demonstrated the specific function of locally specific MLPs in the model, which appeared critical for restoring image detail. Future extensions of this work could be leveraging the synthesized image in a downstream task like segmentation. Furthermore, by using INRs we get a unique representation for each input data, which makes it promising to introduce prior knowledge from follow-up scans in a longitudinal study. In summary, our preliminary experiments have shown promising results in using INRs in MRI translation.

Authorized licensed use limited to: TU Delft Library. Downloaded on October 10,2023 at 08:30:39 UTC from IEEE Xplore. Restrictions apply.

5. COMPLIANCE WITH ETHICAL STANDARDS

The research was conducted using human subject data made available in open access. Ethical approval was not required as confirmed by the license attached with the open access data.

6. ACKNOWLEDGMENTS

This work is funded by China Scholarship Council (grant 202008130140).

7. REFERENCES

[1] A. Cherubini, M. E. Caligiuri, P. Péran, U. Sabatini, C. Cosentino, and F. Amato, "Importance of multimodal MRI in characterizing brain tissue and its potential application for individual age prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 5, pp. 1232–1239, 2016.

[2] R. Azad, N. Khosravi, M. Dehghanmanshadi, J. Cohen-Adad, and D. Merhof, "Medical image segmentation on MRI images with missing modalities: A review," *arXiv preprint arxiv:2203.06217*, 2022.

[3] M. Havaei, N. Guizard, N. Chapados, and Y. Bengio, "HeMIS: Hetero-modal image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 469–477.

[4] A. Sharma and G. Hamarneh, "Missing MRI pulse sequence synthesis using multi-modal generative adversarial network," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 1170–1183, 2020.

[5] Y. Xie, T. Takikawa, S. Saito, O. Litany, S. Yan, N. Khan, F. Tombari, J. Tompkin, V. sitzmann, and S. Sridhar, "Neural fields in visual computing and beyond," *Computer Graphics Forum*, vol. 41, no. 2, pp. 641–676, 2022.

[6] S. M. Jayakumar, W. M. Czarnecki, J. Menick, J.n Schwarz, J. Rae, S. Osindero, Y. W. Teh, T. Harley, and R. Pascanu, "Multiplicative interactions and where to find them," in *International Conference on Learning Representations*, 2020.

[7] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, July 2017.

[8] J. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A Efros, O. Wang, and E. Shechtman, "Toward multimodal image-to-image translation," in *Advances in Neural Information Processing Systems*, 2017, vol. 30.

[9] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," *Commun. ACM*, vol. 65, no. 1, pp. 99–106, dec 2021.

[10] T. R. Shaham, M. Gharbi, R. Zhang, E. Shechtman, and T. Michaeli, "Spatially-adaptive pixelwise networks for fast image translation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2021, pp. 14882–14891.

[11] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," in *International Conference on Learning Representations*, 2017.

[12] B. H. Menze, A. Jakab, S. Bauer, et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

Authorized licensed use limited to: TU Delft Library. Downloaded on October 10,2023 at 08:30:39 UTC from IEEE Xplore. Restrictions apply.