One-sided unsupervised medical image synthesis with normalized edge consistency

Vincent Jaouen, Pierre-Henri Conze, Dimitris Visvikis

ONE-SIDED UNSUPERVISED MEDICAL IMAGE SYNTHESIS WITH NORMALIZED EDGE CONSISTENCY

V. Jaouen [⋆,•] *P.-H. Conze* [⋆,•] *D. Visvikis* [•]

[⋆] IMT Atlantique, Brest, France
[•] LaTIM, UMR Inserm 1101, Brest, France

ABSTRACT

We introduce a novel one-sided approach for unsupervised image-to-image (I2I) translation, referred to as Normalized Edge Consistency (NEC), to address some limitations of methods like CycleGAN and CUT, which can produce realistic images at the cost of limited faithfulness (or structural consistency) with respect to the source. Inspired by the Normalized Gradient Field similarity used in image registration, NEC incorporates a normalized gradient vector field loss combined to an adversarial objective to maintain attachment to local orientation of structures while allowing for realistic contrast changes. NEC is easy to implement and is governed by a unique parameter that controls its sensitivity to noise. We demonstrate NEC's (suprising) efficacy in T1-to-T2 brain MRI translation using unpaired data from the BraTS 2023 dataset, rivaling with supervised methods like Pix2Pix.

***Index Terms*—** Unsupervised learning, image synthesis, MRI, image-to-image translation, NGF.

1. INTRODUCTION

In the last years, with the growing successes of generative models in computer vision, deep image-to-image translation (I2I) has become a vivid area of research in medical imaging [1, 2, 3]. I2I methods can be characterized as monomodal or multimodal. Multimodal methods involve cross-modality image synthesis such as generation of pseudo X-ray Computed Tomography (CT) images from Magnetic Resonance Imaging (MRI) for CT-free radiation therapy treatment planning [1], attenuation estimation in Positron Emission Tomography (PET) [4], or MRI sequence synthesis for unseen domain segmentation [5, 6]. Single modality applications include image denoising and enhancement [7] and domain adaptation to reduce the domain shift at the image level [8, 9].

Typically, when paired data (i.e. co-registered, same patient) is available between the source and target image domains, a pixel-to-pixel *one-sided mapping* from source to domain can be directly learned in a supervised fashion using e.g. convolutional architectures like U-Net [10] under a reconstruction training objective such as the ℓ_1 loss. Following the success of the Pix2Pix Generative Adversarial Network (GAN) model [11], reconstruction losses are often combined to GAN losses for improved realism and blur mitigation [1]. When data from both modalities are unpaired (i.e. unsupervised case), direct reconstruction losses become unuseable. As a workaround, GANs can be combined to cycle-consistent reconstruction [12], where the generated image is backtranslated into its original domain, resulting in less stable, more data demanding *two-sided mapping* due to the simultaneous training of two generators and two discriminators. Cycle-consistency is still very popular and has been included in more recent denoising diffusion-based synthesis approaches [13]. However, in addition to increased complexity, cycle-consistent models perform generally less well than direct reconstruction [14]. A canonical example is signal inversion of the Cerebro-Spinal Fluid (CSF) in T1-to-T2 brain MRI translation (Fig.2a-b), which may be difficult to achieve with CycleGANs, especially in low data regimes. Alternatives have thus been proposed to allow for unsupervised I2I with one-sided mappings [15, 16, 17, 18]. For instance, the Contrastive Unpaired Translation (CUT) model [16] employs a spatially localized contrastive *InfoNCE* loss [19] at the feature level named *PatchNCE*, to maintain Mutual Information (MI) between source and synthetic image patches, at the cost of relatively higher implementation complexity, especially in 3D.

Models like CycleGAN and CUT were originally proposed for natural images. However successful, an often overlooked aspect in medical imaging is that they allow for *realistic* translations (that fool the discriminator), without necessarily maintaining local *faithfulness* to the source image [18], i.e. without preserving local structural information. This enables flexible I2I in the sense that local structure is not exceedingly constrained, with applications such as dog-to-cat, or map-to-satellite translation [16], which involve some degree of feature hallucination. However, such a loose judgment of image faithfulness may not be suited to every application in medical imaging, as maintaining structural integrity remains often essential. This is particularly true for applications such as domain shift mitigation for predictive modelling or segmentation of anatomical or pathological structures.

In response to these challenges, this paper introduces a new one-sided approach to unsupervised I2I named Normalized Edge Consistency (NEC). NEC enforces structural

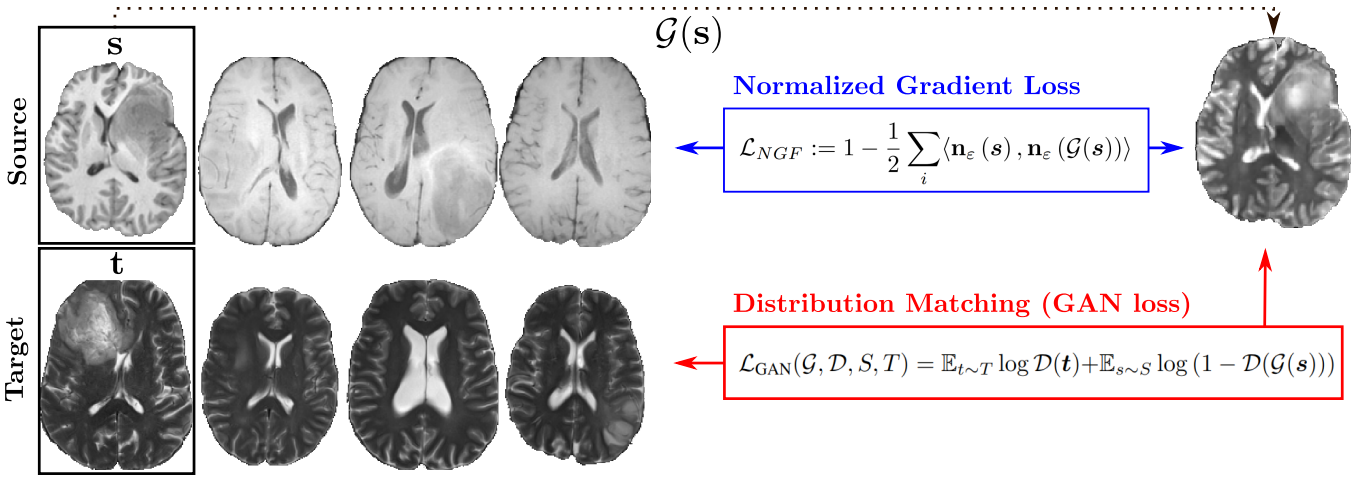


Fig. 1. NEC is a simple unsupervised one-sided model that learns a single generator using a two-term compound loss. The GAN loss promotes *realistic* images while the NGF loss promotes *faithfulness* by preserving the alignment of normalized gradients.

preservation of source image features in a stricter sense than existing approaches for better local faithfulness. Inspired by the Normalized Gradient Field (NGF) for image registration [20], the core idea of NEC is to leverage a normalized gradient vector loss combined to an adversarial objective for enforcing local structures preservation while still allowing for flexible and realistic contrast changes, including global contrast inversion such as seen in e.g. T1-to-T2 MRI translation tasks. In particular, we show that our unsupervised model using unpaired data can surpass CycleGAN and compete with supervised models with paired data on the BraTS 2023 dataset. The method can be straightforwardly extended to 3D with limited complexity.

2. METHODS

Similarly to the CycleGAN or CUT setting, our objective is to translate images from a source domain $\mathcal{S} \subset \mathbb{R}^d$, where $d = (2, 3)$ is the image dimensionality, to a target domain $\mathcal{T} \subset \mathbb{R}^d$ using a dataset of unpaired instances $\mathbf{S} = \{\boldsymbol{s} \in \mathcal{S}\}$, $\mathbf{T} = \{\boldsymbol{t} \in \mathcal{T}\}$. Similarly to CUT, we train a single, one-sided generator function $\mathcal{G}(\boldsymbol{s})$ from $\mathcal{S}$ to $\mathcal{T}$.

Our global training objective is a two-term compound loss (Fig.1), expressed as:

$$\mathcal{L} = \mathcal{L}_{\text{GAN}}(\mathcal{G},\mathcal{D},S,T) + \mathcal{L}_{\text{NGF}}(\boldsymbol{s},\mathcal{G}(\boldsymbol{s})), \qquad (1)$$

where $L_{\text{GAN}}(\mathcal{G},\mathcal{D},S,T)$ is an adversarial loss that promotes realistic pseudo target outputs from the generator $\mathcal{G}$ through its competition with a PatchGAN discriminator $\mathcal{D}$ [11, 12]:

$$\mathcal{L}_{\text{GAN}} = \mathbb{E}_{t\sim T}\log\mathcal{D}(\boldsymbol{t}) + \mathbb{E}_{s\sim S}\log[1-\mathcal{D}(\mathcal{G}(\boldsymbol{s}))]. \qquad (2)$$

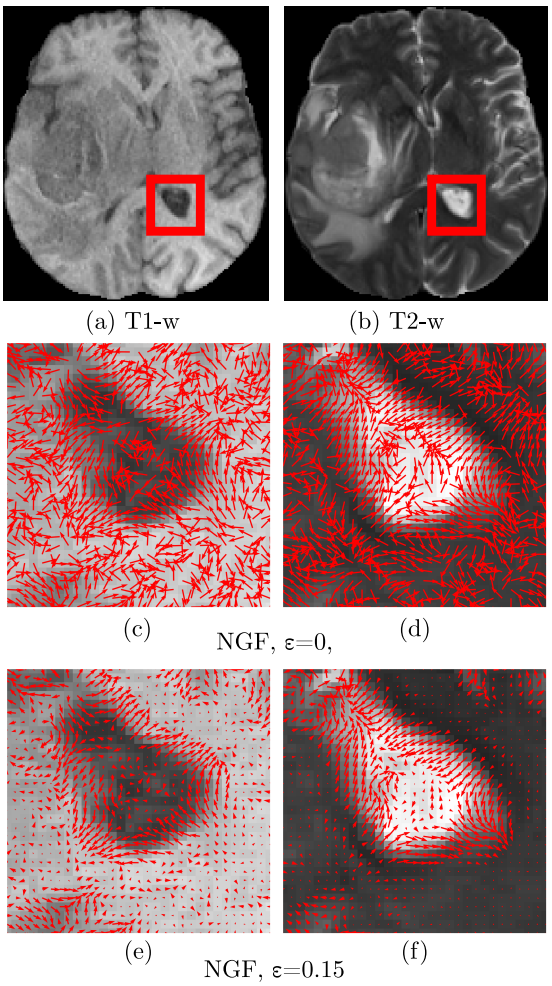
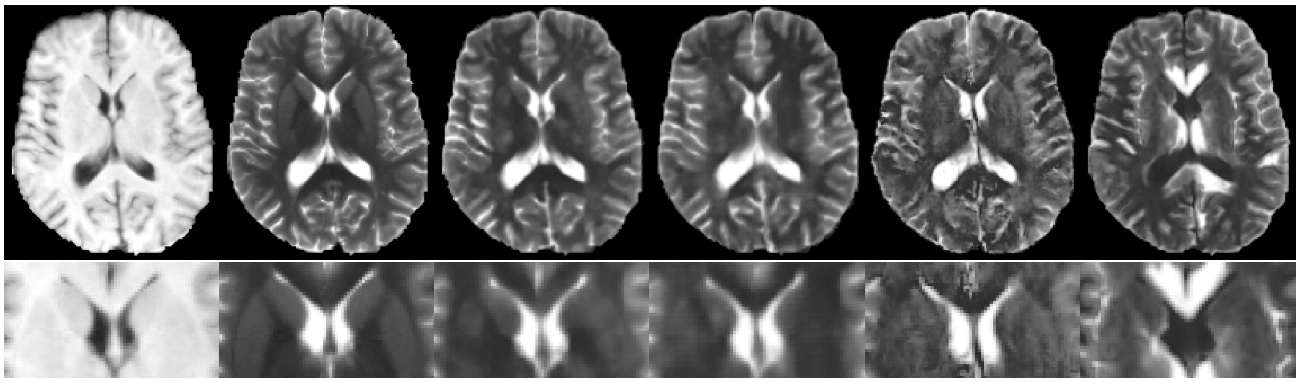


Fig. 2. Normalized gradient fields in one subject from the BraTS 2023 dataset. The unique noise threshold parameter ε controls the degree of acceptance of weak gradients.



(a) Real T1 (b) Real T2 (c) NEC 25.01 dB/SSIM 0.91 (d) Pix2Pix 23.59 dB/SSIM 0.90 (e) CycleGAN 21.13 dB/SSIM 0.80 (f) CUT 16.00 dB/SSIM 0.65

Fig. 3. T1-to-T2 synthesis results on patient 01152-000 of the BraTS 2023 dataset with a focus on the striatal region. The CUT model failed at inverting the CSF signal. Note Pix2Pix is trained on paired data, contrary to the other models.

While GANs can achieve *realistic* outputs, i.e. generate samples similar to the distribution of $\mathcal{T}$, they have no explicit mechanism to maintain *faithfulness* with respect to $\mathcal{S}$ [18]. This can be made for instance through the addition of cycle-consistent [12] or PatchNCE [16] loss terms. However, contrary to natural image processing, it is paramount for many medical imaging applications to be consistent with the source in a stricter, more local fashion than e.g. CUT that maximizes MI between features at the patch level only.

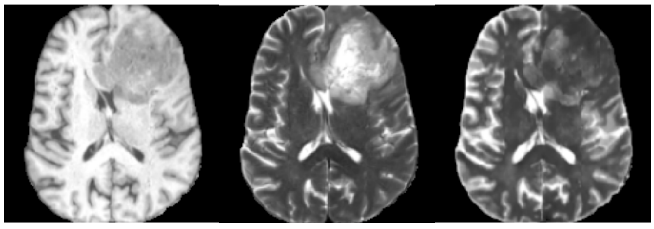
In this work, we hypothesize that the directions of the image gradient vectors must be preserved during translation. At the same time, local contrast variations (i.e. change of gradient magnitude) or even inversion of contrast (i.e. change of gradient sense) should not be penalized. To this end, we constrain the geometric alignment of image gradient vectors during training through a Normalized Gradient Field (NGF) loss. NGF similarity was originally proposed for medical image registration, where it showed advantages over maximization of MI [20]. The NGF loss if defined as :

$$\mathcal{L}_{NGF} := 1 - \frac{1}{2}\langle \boldsymbol{n}_\varepsilon(\boldsymbol{s}), \boldsymbol{n}_\varepsilon(\mathcal{G}(\boldsymbol{s}))\rangle^2, \tag{3}$$

where $\langle \boldsymbol{n}_\varepsilon(\boldsymbol{s}), \boldsymbol{n}_\varepsilon(\mathcal{G}(\boldsymbol{s}))\rangle$ is the dot product between NGFs of the source and generated images:

$$\boldsymbol{n}_\varepsilon(I) = \frac{\nabla I(\boldsymbol{x})}{\|\nabla I(\boldsymbol{x})\| + \varepsilon}, \quad \boldsymbol{x} \in \mathbb{R}^d, \tag{4}$$

with ϵ being a noise threshold scalar parameter that controls the influence of weak edges, with higher values corresponding to higher suppression (Fig.2c-f). The rationale behind normalization is to encourage gradient vector alignment while allowing changes in gradient magnitude during translation, which would be prevented through simple gradient similarity. The squaring makes the method insensitive to tissue contrast inversion between modalities, as in e.g. T1-to-T2 MRI synthesis. NEC is easy to implement as it only needs the computation and normalization of image gradients, compared to sophistications like double-sided mapping in CycleGAN or multi-layer feature-level PatchNCE loss in CUT.



(a) Real T1 (b) Real T2 (c) NEC - 18.53 dB

Fig. 4. Example failure case for NEC around the tumor region while maintaining high overall faithfulness to the T1 source.

3. EXPERIMENTS

We evaluated the performance of NEC on T1-to-T2 MRI synthesis on the RSNA-ASNR-MICCAI BraTS 2023 adult glioma challenge training dataset composed of 1251 MRI images [21]. We compared our approach to unsupervised methods CycleGAN and CUT, but also to a supervised Pix2Pix model trained on paired data. To facilitate comparative experiments at a relatively low data regime, we used a single axial slice per patient (slice #80/155). The first 400 images, the next 400, the next 10, and the next 441 in lexicographic order were used as T1 training set, T2 training set, validation set and testing set, respectively. For Pix2Pix, the first 400 images were also used for T2 training due to the need for paired data.

All methods used a residual 8-layer U-Net generator architecture with 9 residual blocks using 256×256 trilinearly resampled image inputs. Our discriminator was a 3-layer,

3-scale patch discriminator. We implemented our method in PyTorch using MONAI v1.3 loaders and classes including `nets.Unet`, `losses.PatchAdversarialLoss` and `generative.MultiScalePatchDiscriminator`. We used the same network backbone for Pix2Pix. Our re-implementation of CycleGAN failed at achieving CSF contrast inversion in this data regime. We thus used the implementation[1,2] of the authors for CycleGAN and CUT with an adapted data loader to handle NIFTI files. Images were clipped to the [1%-99%] percentile range and linearly rescaled to the interval [-1,1], following common practice. The NGF noise threshold parameter was set to $\varepsilon = 0.15$.

Image translation quality was evaluated using PSNR and SSIM [22]. Prior to evaluation, images were rescaled to their original range using histogram matching to the real image to avoid any bias in the metrics owing to intensity shifts and to facilitate visual comparison. A small posterior Gaussian filtering of scale 0.5mm was applied for all methods that slightly improved figures of merit. Weights maximizing SSIM on average on the validation set were selected for testing.

4. RESULTS AND DISCUSSION

Table 1. T1-to-T2 quantitative synthesis results.

	PSNR (dB)	SSIM
NEC (proposed)	24.47 ± 2.25	0.893 ± 0.032
CycleGAN	21.91 ± 1.58	0.825 ± 0.028
CUT	15.98 ± 1.53	0.656 ± 0.034
Pix2Pix (supervised)	24.62 ± 1.85	0.894 ± 0.031

Fig.3 compares visual synthesis results in one patient of the test set. In this example, NEC was better able to preserve cortical structures from the source T1 image than the other methods. This is particularly visible in the striatal region, with caudate and putamen showing sharper boundaries (Fig.3c). Despite using the authors implementation, CUT failed at CSF signal inversion (Fig.3f). Note that our re-implementation of CycleGAN also failed at contrast inversion (not shown). This can be explained by the relatively low data regime (400 2D images) and high domain shift due to the multi-center setting of BraTS. CUT nevertheless led to visually crisper results compared to noisier CycleGAN results (Fig.3e). Pix2Pix trained on paired data produced globally accurate translations, which were on average slightly blurrier than NEC. NEC showed more variability in the results, with a number of failed translations. In these cases, the tumoral region was in general not accurately translated to hypersignal in T2, despite showing high faithfulness with respect to T1 and realism for the rest of the image (Fig.4c).

[1] `github.com/junyanz/pytorch-CycleGAN-and-Pix2Pix`
[2] `github.com/taesungp/contrastive-unpaired-translation`

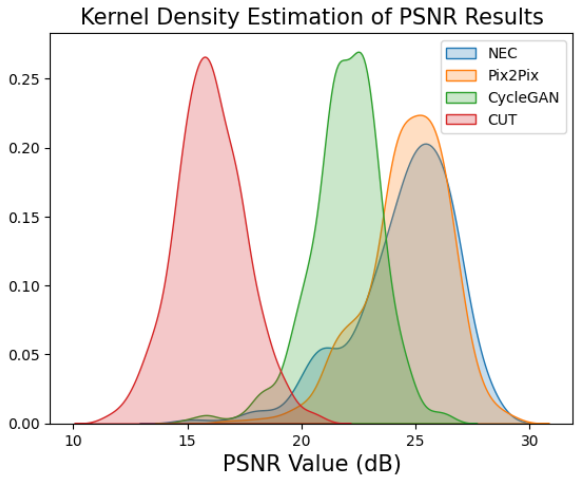


Fig. 5. Distribution of PSNR on the test set.

Regarding quantitative metrics, Tab.1 summarizes the quantitative results obtained on the 441 images of the test set. NEC (average PSNR 24.5/SSIM 0.89) clearly outperformed its unsupervised competitors CUT (average PSNR 15.98/SSIM 0.66) and CycleGAN (average PSNR 21.9/SSIM 0.83) using the same training data. Moreover, NEC achieved results competitive results with respect to Pix2Pix (PSNR 24.6/SSIM 0.89), despite being trained on unpaired data. Fig.5 shows the distributions of PSNR scores using kernel density estimation. The variability of the results was slightly higher for NEC (i.e. slightly heavier left-sided tail in Fig.5).

5. CONCLUSION

In this work, we have proposed a new one-sided unsupervised image-to-image translation approach based on the maximization of the alignment of normalized gradient vectors between the source and the generated image, named Normalized Edge Consistency (NEC). This surprisingly simple and computationally efficient idea demonstrated a clear advantage over CUT and CycleGAN in T1-to-T2 MRI translation, where structural consistency between source and target domains is prevailing. NEC reached performances comparable to the supervised method Pix2Pix trained on paired data. Future investigations will target other cross-modality applications and anatomies, such as MR to CT translation for radiation therapy treatment planning in prostate cancer.

6. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by the organizers of the BraTS challenge [21]. Ethical approval was not required, as confirmed by the license attached with the open access data.

7. REFERENCES

[1] J. M. Wolterink et al., "Deep MR to CT synthesis using unpaired data," in *SASHIMI 2017, MICCAI 2017*. Springer, 2017, pp. 14–23.

[2] S. A. Tsaftaris et al., *Simulation and Synthesis in Medical Imaging: SASHIMI 2017, MICCAI 2017*, vol. 10557, Springer, 2017.

[3] C. Zhao et al., *Simulation and Synthesis in Medical Imaging: SASHIMI 2022, MICCAI 2022*, vol. 13570, Springer Nature, 2022.

[4] K. Gong et al., "MR-based attenuation correction for brain PET using 3-D cycle-consistent adversarial network," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 5, no. 2, pp. 185–192, 2020.

[5] R. Dorent et al., "CrossMoDA 2021 challenge: Benchmark of cross-modality domain adaptation techniques for vestibular schwannoma and cochlea segmentation," *Med. Image Anal.*, vol. 83, pp. 102628, 2023.

[6] G. Sallé et al., "Cross-modal tumor segmentation using generative blending augmentation and self training," *arXiv preprint arXiv:2304.01705*, 2023.

[7] J. M. Wolterink et al., "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans. Med. Imaging*, vol. 36, no. 12, pp. 2536–2545, 2017.

[8] C. Hognon et al., "Standardization of multicentric image datasets with generative adversarial networks," in *IEEE NSS MIC 2019*, 2019.

[9] F Tixier et al., "Evaluation of conventional and deep learning based image harmonization methods in radiomics studies," *Physics in Medicine & Biology*, vol. 66, no. 24, pp. 245009, 2021.

[10] O. Ronneberger et al., "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI 2015, Part III*. Springer, 2015, pp. 234–241.

[11] P. Isola et al., "Image-to-image translation with conditional adversarial networks," in *IEEE CVPR*, 2017, pp. 1125–1134.

[12] J.-Y. Zhu et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE ICCV*, 2017, pp. 2223–2232.

[13] M. Özbey et al., "Unsupervised medical image translation with adversarial diffusion models," *IEEE Trans. Med. Imaging*, 2023.

[14] P. Klages et al., "Patch-based generative adversarial neural network models for head and neck MR-only planning," *Med. Phys.*, vol. 47, no. 2, pp. 626–642, 2020.

[15] H. Fu et al., "Geometry-consistent generative adversarial networks for one-sided unsupervised domain mapping," in *IEEE/CVF CVPR*, 2019, pp. 2427–2436.

[16] T. Park et al., "Contrastive Learning for Unpaired Image-to-Image Translation," *arXiv preprint arXiv:2007.15651*, 2020.

[17] C. Zheng et al., "The spatially-correlative loss for various image translation tasks," in *IEEE/CVF CVPR*, 2021, pp. 16407–16417.

[18] M. Zhao et al., "Egsde: Unpaired image-to-image translation via energy-guided stochastic differential equations," *Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 3609–3623, 2022.

[19] A. van den Oord et al., "Representation learning with contrastive predictive coding," in *Int. Conf. Mach. Learn.*, 2018, pp. 4854–4863.

[20] E. Haber and J. Modersitzki, "Intensity gradient based registration and fusion of multi-modal images," in *MICCAI 2006, Part II*. Springer, 2006, pp. 726–733.

[21] U. Baid et al., "The RSNA-ASNR-MICCAI BRATS 2021 benchmark on brain tumor segmentation and radiogenomic classification," *arXiv preprint arXiv:2107.02314*, 2021.

[22] Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.