



Unsupervised Joint Image Transfer and Uncertainty Quantification Using Patch Invariant Networks

Christoph Angermann^(✉), Markus Haltmeier, and Ahsan Raza Siyal

Department of Mathematics, University of Innsbruck,
Technikerstraße 13, 6020 Innsbruck, Austria
christoph.angermann@uibk.ac.at
<http://applied-math.uibk.ac.at/>

Abstract. Unsupervised image transfer enables intra- and inter-modality image translation in applications where a large amount of paired training data is not abundant. To ensure a structure-preserving mapping from the input to the target domain, existing methods for unpaired image transfer are commonly based on cycle-consistency, causing additional computational resources and instability due to the learning of an inverse mapping. This paper presents a novel method for unidirectional domain mapping that does not rely on any paired training data. A proper transfer is achieved by using a GAN architecture and a novel generator loss based on patch invariance. To be more specific, the generator outputs are evaluated and compared at different scales, also leading to an increased focus on high-frequency details as well as an implicit data augmentation. This novel patch loss also offers the possibility to accurately predict aleatoric uncertainty by modeling an input-dependent scale map for the patch residuals. The proposed method is comprehensively evaluated on three well-established medical databases. As compared to four state-of-the-art methods, we observe significantly higher accuracy on these datasets, indicating great potential of the proposed method for unpaired image transfer with uncertainty taken into account. Implementation of the proposed framework is released here: <https://github.com/anger-man/unsupervised-image-transfer-and-uq>.

Keywords: Unsupervised image transfer · Uncertainty quantification · Generative adversarial network · Patch invariance · Modality propagation · Accelerated MRI · Radiotherapy

1 Introduction

Image transfer, e.g., within and across medical imaging modalities, has gained a lot of popularity in the last decade of research [1]. The application range of inter- and intra-modality image translation is multifaceted and can help to overcome key weaknesses of an acquisition method. For example, modality propagation in

magnetic resonance imaging (MRI) is of high interest since acquisition of multiple contrasts is crucial for better diagnosis for many clinical protocols [2]. Especially acquiring T1-weighted (T1w) and T2-weighted (T2w) contrasts increases scanning time significantly and thus is often formulated as a translation task from T1w to T2w. Another example is accelerated MRI, where medical costs and patient stress are minimized by decreasing the amount of k-space measurements [3, 4]. Such methods for MRI reconstruction of undersampled measurements allow the use of MRI in applications where it is currently too time and resource intensive. A third application to mention here is automated computed tomography (CT) synthesis based on MRI images, which allows for MRI-only treatment planning in radiation therapy. A MRI-to-CT synthesizing method can eliminate the need for CT simulation and therefore improves the treatment workflow and reduces radiation exposure for the patient during radiotherapy [1, 5–8].

Just discussed applications correspond to (un)supervised image transfer, which targets at translating an image from one domain to another. Supervised approaches exploit the inter-domain correspondence between input and output data [9]. These methods rely on large amounts of paired data and perfectly registered scans of the same patient, which are not always abundant in medical applications [5]. Unsupervised approaches are commonly built on a generative adversarial network (GAN) [10] that assimilates the distribution of the generated samples to the real distribution of the target domain by employing an adversarial discriminator network. To ensure that the synthesized output does not become irrelevant to the input, additional constraints may be added to the generator loss [11–14]. Especially cycle-consistency is a well-received method for structure preservation in fully unsupervised medical image transfer [6–9]. However, a cycle-consistent GAN (cycleGAN) requires the parallel learning of an inverse mapping. As a result, the training time is significantly increased and the final performance depends on the inverse transfer function. Furthermore, cycle-consistent GANs compare the reconstruction on a whole-image base and therefore pay less attention to fine structures and high-frequency details.

Although current methods provide powerful tools for high-dimensional image transfer, the generator loss calculation usually assumes that the learned mappings are correct. This represents a significant source for instability and erroneous predictions when considering out-of-distribution (OOD) data during training [9, 15]. Tackling data-dependent uncertainty in deep computer vision has raised a lot of interest in the last years of research and provided effective tools to check the reliability of a model’s predictions in supervised applications. However, research on modeling uncertainty inherent from data in a completely unsupervised setting is still limited [9, 16] and needs a deeper investigation.

This paper presents a novel GAN approach to fully unpaired medical image transfer, including prediction of data-dependent uncertainty and invariance over patches. To be more exact, a Wasserstein generative adversarial network (WGAN) [17] is leveraged to an uni-directional image transfer model. Structural correspondence between input and target modality is guaranteed using a novel generator loss that enforces invariance over image patches. Further-

more, the patch-based residuals are assumed to follow a zero-mean Laplace distribution with the scale parameter being a function of the input. As a consequence, the generator is allowed to predict uncertainty that operates as a learned loss attenuation and can be used to indicate the quality of a transferred image in the absence of ground truth data. The proposed model and training strategy is evaluated in three different unsupervised scenarios: modality propagation using T1w and T2w brain MRI from the IXI [18] database; accelerated MRI enhancement using emulated single-coil knee MRI from the FastMRI [3] database; MRI-to-CT synthesis using head CT scans from the CQ500 [19] database. The proposed framework is benchmarked against state-of-the-art works for unidirectional and bi-directional image translation [9, 11, 12, 14]. We not only evaluate accuracy on unseen test data but further investigate robustness to perturbed inputs.

Contributions:

- We present an unidirectional framework that enables fully unsupervised image transmission of medical data while preserving fine structures.
- Structural correspondence between different characteristics and modalities is ensured by an improved generator loss based on patch invariance. This also yields implicit data augmentation for the critic and generator networks.
- In addition to the image transmission, the model provides an uncertainty map that correlates with the prediction error, indicating the quality of a mapped instance.

2 Related Work

2.1 Generative Adversarial Networks

The GAN architecture [10] is composed of a generator network $G : \mathcal{Z} \rightarrow \mathcal{X}$ and an adversarial part $f : \mathcal{X} \rightarrow [0, 1]$. The generator maps from a latent space \mathcal{Z} to image space \mathcal{X} , where the parameters of G are adapted such that the distribution of the synthesized examples assimilates to the real data distribution on \mathcal{X} . Simultaneously, the adversarial f is trained to distinguish between synthesized and real instances. In a two-player min-max game, generator parameters are updated to fool a steadily improving discriminator [20]. Improving the joint loss functional of the generating and the adversarial part yielded improved modifications of the initial GAN framework, like WGAN [17], improved WGAN [21], LSGAN [22] or SNGAN [23]. While GANs reach outstanding performance in image synthesis [24, 25], they are also well accepted in improving prediction quality in supervised image applications such as super-resolution [26], paired image translation [27, 28], and medical image enhancement [5, 29].

2.2 Unpaired Image Transfer and Domain Mapping

Unpaired image translation maps an image from input to target domain where corresponding samples from both spaces are hard to obtain or applied registration methods yield too much misalignment. In these cases, cycleGAN [11]

has become the gold standard since it learns an inverse mapping from target domain back to input space. The core idea of cycleGAN is that the synthesized image must retain enough detail of the input instance in the target domain to allow for reconstruction. Especially in the medical sector, a structure-preserving transfer function is of high priority. Wolterink et al. [7] utilized cycle-consistency for MRI-only treatment planning in radiotherapy. Hiasa et al. [8] improved the cycleGAN architecture in this application area by adding a gradient consistency loss to pay more attention to the edges in the image. Yang et al. [6] added to cycleGAN a structure-consistency loss based on a modality independent neighborhood descriptor.

Learning an inverse GAN framework simultaneously (bi-directional) in order to ensure input-output consistency increases hardware requirements and introduces an additional instability if the inverse generator is not trained sufficiently or if the transfer mapping is not injective. Fu et al. [12] investigated geometry-consistent GAN (gcGAN), an uni-directional approach that enforces consistency when applying geometric transformations (rotation, flipping) before and after propagation through the generator network. Benaim et Wolf [13] considered GAN in combination with a distance constraint, where the distance between two samples from the input domain should be preserved after mapping to the target domain. A very recent and successful uni-directional approach to unpaired domain mapping was made by Park et al. [14] that uses contrastive unpaired translation (CUT), i.e., structure-consistency is preserved by matching patches of the input and the synthesized instance using an additional classification step.

2.3 Uncertainty Quantification

Uncertainty quantification methods have been applied to solve a variety of real-world problems in computer vision where in addition to the model's response also a measure on its confidence is provided [30]. In general, two broad categories of uncertainty are considered: aleatoric uncertainty captures noise inherent in the data and epistemic/model uncertainty describes uncertainty in the model parameters [15]. The latter type of uncertainty occurs in finite data settings and thus can be explained away providing a sufficient amount of data. Bayesian models provide a mathematically grounded framework that can account for model uncertainty in combination with Bayesian inference techniques. Gal et Ghahramani [31] set up a theoretical framework that casts the dropout technique as approximate Bayesian inference, enabling a rather simple calculation of epistemic uncertainty by multiple network forward passes. The works of Saatci et Wilson [16] as well as Palakkadavath et Srijith [32] leverage this framework to Bayesian GANs and show that considering Bayesian learning principles can address mode collapse in image synthesis. Kendall et Gal [15] have explored the benefits of modeling aleatoric and epistemic uncertainty simultaneously in image segmentation and regression and concluded that the two types of uncertainty are not mutually exclusive, but in fact complementary in different data scenarios. Upadhyay et al. modeled aleatoric uncertainty for MRI image enhancement [2]

and unsupervised image transfer [9] by introducing uncertainty-aware generalized adaptive cycleGAN (UGAC). Therefore, the latter work will also serve as a benchmark method for the proposed uncertainty-aware uni-directional image transfer approach.

3 Method

3.1 Preliminaries and GAN Architecture

The underlying structure of the proposed uncertainty-aware domain mapping is a GAN combined with a patch invariant generator term. Let $\mathcal{X} \subset \mathbb{R}^{d \times d \times c_{in}}$ and $\mathcal{Y} \subset \mathbb{R}^{d \times d \times c_{out}}$ denote the input and the target domain, respectively. For simplicity we consider quadratic instances with the number of image pixels equal to d^2 . Furthermore, let $X := \{x_1, \dots, x_M\}$ be the set of M given input images and $Y := \{y_1, \dots, y_N\}$ the set of N available but unaligned target images. $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ denote the distributions of the images in both domains. The proposed image transfer is built on a generator function $G_{\theta}: \mathcal{X} \rightarrow \mathcal{Y}$, which aims to map an input sample to a corresponding instance in the target domain. The generator function is approximated by a convolutional neural network (CNN), which is parameterized by a weight vector θ . By adjusting θ , the distribution P_{θ} of generator outputs may be brought closer to the real data distribution $P_{\mathcal{Y}}$ in the target domain. The distance between the generator distribution and the real distribution is estimated with the help of the critic $f_{\omega}: \mathcal{Y} \rightarrow \mathbb{R}$, which is parameterized by weight vector ω and is trained simultaneously with the generator network since P_{θ} changes after each update to the generator weights θ [20].

We choose a network critic based on the Wasserstein-1 distance [17, 20, 33]. The Wasserstein-1 distance between two distributions P_1 and P_2 is defined as $\mathcal{W}_1(P_1, P_2) := \inf_{J \in \mathcal{J}(P_1, P_2)} \mathbb{E}_{(x,y) \sim J} \|x - y\|$, where the infimum is taken over the set of all joint probability distributions that have marginal distributions P_1 and P_2 . The Kantorovich-Rubinstein duality [33] yields

$$\mathcal{W}_1(P_1, P_2) = \sup_{\|f\|_L \leq 1} \left[\mathbb{E}_{y \sim P_1} f(y) - \mathbb{E}_{y \sim P_2} f(y) \right], \quad (1)$$

where $\|\cdot\|_L \leq C$ denotes that a function is C -Lipschitz. Equation (1) indicates that a good approximation to $\mathcal{W}_1(P_{\mathcal{Y}}, P_{\theta})$ is found by maximizing $\mathbb{E}_{y \sim P_{\mathcal{Y}}} f_{\omega}(y) - \mathbb{E}_{y \sim P_{\theta}} f_{\omega}(y)$ over the set of CNN weights $\{\omega \mid f_{\omega}: \mathcal{Y} \rightarrow \mathbb{R} \text{ 1-Lipschitz}\}$, where the Lipschitz continuity of f_{ω} can be enhanced via a gradient penalty [21]. Given training batches $\mathbf{y} = \{y_n\}_{n=1}^b$, $y_n \stackrel{\text{iid}}{\sim} P_{\mathcal{Y}}$ and $\mathbf{x} = \{x_n\}_{n=1}^b$, $x_n \stackrel{\text{iid}}{\sim} P_{\mathcal{X}}$, this yields the following empirical risk for critic f_{ω} :

$$\ell_{\text{cri}}(\omega, \theta, \mathbf{y}, \mathbf{x}, p) := \frac{1}{b} \sum_{n=1}^b f_{\omega}(G_{\theta}(x_n)) - f_{\omega}(y_n) + p \cdot \left((\|\nabla_{\tilde{y}_n} f_{\omega}(\tilde{y}_n)\|_2 - 1)_+ \right)^2, \quad (2)$$

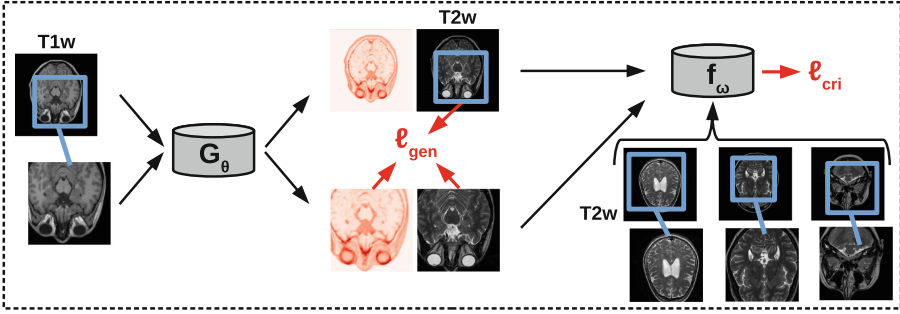


Fig. 1. Utilizing patch invariance for unsupervised MRI propagation and uncertainty quantification. The T1w input and a corresponding random patch on a finer scale are fed to the generator G_θ , which outputs the synthesized T2w counterparts and corresponding scale maps (red). The synthesized patch and the corresponding patch of the full-size output are compared. Loss attenuation is introduced by the scale map of the synthesized patch. The generator is additionally updated using the Wasserstein-1 distance, estimated with the help of f_ω . (Color figure online)

where p denotes the influence of the gradient penalty, $(\cdot)_+ := \max(\{0, \cdot\})$ and $\tilde{y}_n := \epsilon_n \cdot G_\theta(x_n) + (1 - \epsilon_n) \cdot y_n$ for $\epsilon_n \stackrel{\text{iid}}{\sim} \mathcal{U}[0, 1]$. Since only the first term of the functional in (2) depends on θ and the goal for the generator is to minimize the Wasserstein-1 distance, the adversarial empirical risk for generator G_θ simplifies as follows:

$$\ell_{\text{gen}}(\theta, \omega, \mathbf{x}) := -\frac{1}{b} \sum_{n=1}^b f_\omega(G_\theta(x_n)). \quad (3)$$

3.2 Patch Invariance

In the frame of medical image translation, it is not sufficient to ensure that the output samples lie in the target domain. Great attention should be paid that a model also preserves global structure as well as fine local details. Let $x \in \mathbb{R}^{d \times d \times c}$ and $\Phi := \{(\rho, j_1, j_2) \in [0.7, 1] \times [0, d]^2 \mid j_1 + \rho d \leq d \wedge j_2 + \rho d \leq d\}$. We define the patch operator $\mathcal{P} : \Phi \times \mathbb{R}^{d \times d \times c} \rightarrow \mathbb{R}^{d \times d \times c}$ as follows:

$$\mathcal{P}(\rho, j_1, j_2)(x) := \mathcal{R}_{d \times d \times c}(x[j_1 : j_1 + \rho d, j_2 : j_2 + \rho d, :]), \quad (4)$$

where $(\rho, j_1, j_2) \in \Phi$ and $\mathcal{R}_{d \times d \times c}$ resizes the patch to original image size $d \times d \times c$. The patch operator \mathcal{P} chooses a quadratic patch of 70% to 100% the input size and conducts resampling to original size (cf. Fig. 1). For resampling, we use bicubic interpolation¹.

The basic intuition now is: if we take a patch of the input image and propagate it through the generator, than it should be equal to the corresponding patch

¹ https://www.tensorflow.org/api_docs/python/tf/image/resize.

of the transferred full-size image. We choose the 1-norm for comparing the corresponding patches and ensure realistic synthesized patches by adding the patch operator also to the Wasserstein-1 critic. This yields the following improvements for the critic and the generator risks:

$$\ell_{\text{cri}}(\omega, \theta, \mathbf{y}, \mathbf{x}, p, \vec{\phi}) := \frac{1}{b} \sum_{n=1}^b \left[f_{\omega}(G_{\theta}(x_n)) - f_{\omega}(y_n) + f_{\omega}(\mathcal{P}_{\phi_n}(G_{\theta}(x_n))) - f_{\omega}(\mathcal{P}_{\phi_n}(y_n)) + p \cdot (\dots)^2 \right], \quad (5)$$

$$\ell_{\text{gen}}(\theta, \omega, \mathbf{x}, \vec{\phi}, \lambda) := \frac{1}{b} \sum_{n=1}^b \left[-f_{\omega}(G_{\theta}(x_n)) - f_{\omega}(\mathcal{P}_{\phi_n}(G_{\theta}(x_n))) + \lambda \cdot \underbrace{d^{-2} \|G_{\theta}(\mathcal{P}_{\phi_n}(x_n)) - \mathcal{P}_{\phi_n}(G_{\theta}(x_n))\|_1}_{\ell_{\text{patch}}(\theta, x_n, \phi_n)} \right], \quad (6)$$

where λ controls the influence of the patch loss and the patch extraction settings $\phi = \{\phi_n\}_{n=1}^b$, $\phi_n \in \Phi$ are chosen randomly at each risk calculation.

This approach yields some practical advantages: The generator is forced to be consistent over smaller patches, which prevents the network to generate modes with highest similarity to the real data (mode collapse). Furthermore, the generator is prevented from learning arbitrary mappings between input and target domain (e.g., mapping a T1w MRI of an old lady to a T2w MRI of a young boy), because this memorized mapping would then also have to be fulfilled for all smaller patches, i.e., the transfer would also have to be memorized on arbitrary scales. The patch extractor can also be viewed as a magnification and cropping operation. This yields a higher penalty when comparing fine structures that may not have much effect on the loss function when compared on full image scale. Finally, patch extraction causes implicit data augmentation and can help to avoid critic overfitting where the critic is tempted to memorize training samples.

3.3 Uncertainty by Loss Attenuation

We consider now Eq. (6) from a probabilistic point of view. For $x \in \mathcal{X}$, let $a = \mathcal{P}_{\phi}(G_{\theta}(x))$ and $b = G_{\theta}(\mathcal{P}_{\phi}(x))$ for a patch configuration $\phi \in \Phi$. If we force the patch invariance via the 1-norm, the underlying assumption is that every pixel of the residual $\epsilon := a - b$ should follow a zero-mean and fixed-scale Laplace distribution [9]. Consider residual pixel $\epsilon_j \sim \text{Laplace}(0, \sigma)(\epsilon_j) = \frac{1}{2\sigma} \exp(-|\epsilon_j|/\sigma)$ where σ represents the scale parameter of the distribution. Maximum likelihood (ML) optimization on the full image (note a and b are functions of θ) yields

$$\max_{\theta} \prod_{j=1}^d \frac{1}{2\sigma} \exp(-|a_j - b_j|/\sigma). \quad (7)$$

Applying the negative logarithm and dividing by factor d^2 results in

$$\min_{\theta} \frac{1}{d^2} \sum_{j=1}^{d^2} |a_j - b_j|/\sigma + \log(2\sigma), \quad (8)$$

which is equivalent to minimizing $\ell_{\text{patch}}(\theta, x, \phi)$ in Eq. (6) when considering a fixed scale σ . The assumption of a fixed scale for the pixel-wise residuals is quite strong and may not hold in the presence of OOD data. The idea now is to consider individual scales for every pixel. Inspired by [9, 15], we make the scale σ a function of input x , i.e., we split the generator $G_{\theta}(x) = [G_{\theta}^I(x), G_{\theta}^{\sigma}(x)]$ in the output branch and return two images, the transferred image $G_{\theta}^I(x)$ and the corresponding pixel-wise scale map $G_{\theta}^{\sigma}(x)$ for the residuals. This results in

$$\ell_{\text{patch}}(\theta, x, \phi) = \frac{1}{d^2} \sum_{j=1}^{d^2} \frac{|G_{\theta}^I(\mathcal{P}_{\phi}(x))_j - \mathcal{P}_{\phi}(G_{\theta}^I(x))_j|}{G_{\theta}^{\sigma}(\mathcal{P}_{\phi}(x))_j} + \log(2 \cdot G_{\theta}^{\sigma}(\mathcal{P}_{\phi}(x))_j). \quad (9)$$

This can be seen as a loss attenuation as we get high values in $G_{\theta}^{\sigma}(x)$ for image regions with high absolute residuals. At the same time, the logarithmic term discourages the model to predict high uncertainty for all pixels. The proposed generator loss for patch invariant and uncertainty-aware image transfer is obtained by inserting ℓ_{patch} (9) into ℓ_{gen} (6).

3.4 Implementation Details

In this work, the generator is a U-net [34] with five downsampling operations and approximately 10.7×10^6 parameters. After the last upsampling operation, the U-Net is split into two branches to generate two responses, the transferred image $G_{\theta}^I(\cdot)$ and the corresponding uncertainty map $G_{\theta}^{\sigma}(\cdot)$, c.f. (9). A non-negative scale map is enforced by applying the softplus activation function $\text{softplus}(x) := \log(\exp(x) + 1)$ to the latter output branch. A decoding network for the Wasserstein critic is built following the DCGAN critic [35] with 5 downsampling steps and approximately 4.7×10^6 parameters. Detailed information on critic and generator implementation can be found in the github repository. All models are trained using the Adam optimizer [36] with $\beta_1 = 0$, $\beta_2 = 0.9$ and minibatch size 8. The learning rate is set to 5×10^{-5} for the generator and 2×10^{-5} for the critic network. No learning rate scheduler or further data augmentation techniques are applied. The total amount of generator updates is 15k and we iterate between 15 critic updates and 1 generator update. Gradient penalty parameter p equals 10, the influence λ of the patch constraint is chosen for each data set individually by a grid search.

4 Experiments

4.1 Datasets

We consider three different tasks in medical image-to-image translation.

Modality Propagation: The IXI [18] database consists of registered T1w and T2w scans of 577 patients. We want to demonstrate the plausibility of our model for unpaired modality propagation and thus build a model for T1w to T2w transfer. To do so, we remove 10% of the patients for evaluation. The remaining patients are randomly split into input and target data, where no patient contributes to both domains at the same time. This is done to simulate a scenario where no paired slices are available throughout the entire training process. We only use the core 60% of all axial slices, which yields approximately 20k train slices for input, 20k train slices for target, and 4k pairs for evaluation. The spatial dimension d equals 256.

Accelerated MRI Enhancement: The FastMRI [3] database consists of more than 1500 multi-coil diagnostic knee MRI scans and corresponding emulated single-coil data. Our experiments are based on a subset of nearly 800 coronal proton-density weighted scans without fat-suppression from the official train and validation single-coil releases. We remove 10% of the patients for evaluation and split the remaining patients into input and target data. While the target domain consists of slices from fully-sampled MRI scans, we consider 4x acceleration (only 25% of k-space measurements) for the slices by using the subsampling scheme discussed in [3,4]. This yields 7.1k train slices for input and 6.9k for target. We are aware that enhancement of accelerated MRI can also be considered as a supervised task since generation of paired instances is feasible. However, this experiment should demonstrate possible applicability of the proposed framework to inverse problems in general.

MRI-to-CT Synthesis: The CQ500 [19] consists of CT scans of nearly 500 patients, where we use a randomly selected subset of 80 patients for the target domain. Furthermore, we make use of T1w MRI scans of 144 randomly selected patients taken from IXI [18] as input data. For each dataset we use the core 60% of all axial slices, which yields 11.4k train slices for input and 10.4k train slices for target. All slices are subsampled to spatial size $d = 256$. Note that in this experiment input and target data is coming from completely separated datasets (inconsistent head orientations, different brain areas, resolution, etc.). An additional challenge are artifacts outside the skull caused by the CT table and the measurement equipment in CQ500. We step away from any preprocessing here and investigate how the method reacts to this kind of artifacts. Since no ground truth data is available for the two separated databases only qualitative evaluation is conducted.

4.2 Compared Methods and Scenarios

We compare our approach to a variety of state-of-the-art methods for unsupervised image transfer that have already been introduced in Sect. 2.2 and Sect. 2.3: cycle-consistent GAN (cycleGAN) [11], uncertainty-aware generalized adaptive cycleGAN (UGAC) [9], geometry-consistent GAN (gcGAN) [12] with horizontal flip and contrastive unpaired translation (CUT) [14]. We test two versions of our approach: the first version utilizes only patch invariance (PI, cf. (6)) and the

second version utilizes uncertainty-aware patch invariance (UAPI, cf. (9)). For cycleGAN, UGAC and gcGAN we make use of the same generator and critic architecture and training configurations as described in Sect. 3.4 to guarantee a fair comparison. For CUT, we use the publicly available github repository². All slices are normalized to range $[0, 255]$ and handled as grayscale images. During optimization, the images are scaled to $[-1, 1]$ to speed up training while evaluation metrics, the structural similarity index (SSIM) [37] and the peak-signal-to-noise-ratio (PSNR), are calculated on original image scale.

For each of the three applications, we want to test not only performance on unseen assessment data, but also robustness to different types of perturbations. All approaches are trained on unaffected images (without additional noise) and evaluated in the following scenarios: GN0 (original test images); GN5 (adding Gaussian noise, deviation 5% of image range); GN10 (adding Gaussian noise, deviation 10%); GN20 (adding Gaussian noise, deviation 20%); IP2 (impulse perturbation, 2% of pixels replaced by random values); IP5 (5% of pixels replaced); IP10 (10% of pixels replaced).

4.3 Quantitative Evaluation

Table 1. Quantitative evaluation of our approach and four compared methods on two datasets under seven different noise scenarios. The reported metrics are the structural similarity index and the peak-signal-to-noise-ratio (SSIM/PSNR, higher is better).

Methods	GN0	GN5	GN10	GN20	IP2	IP5	IP10
IXI test data							
cycleGAN	78.99/20.90	77.04/20.71	72.53/20.14	63.87/19.19	69.81/20.39	64.51/19.99	60.08/19.50
UGAC	79.31/21.21	77.24/21.05	72.53/20.43	64.03/19.37	70.15/20.68	64.88/20.22	61.01/19.78
gcGAN	81.00/21.52	76.51/20.66	69.96/19.50	57.75/17.90	73.79/20.30	69.24/19.55	63.91/18.78
CUT	78.67/21.15	76.10/21.22	71.43/20.58	61.27/19.67	74.11/21.19	71.78/20.94	68.97/18.65
PI(ours)	82.53/22.18	69.82/21.49	51.90/20.64	36.96/19.26	54.85/20.92	44.10/20.24	38.73/19.70
UAPI(ours)	79.99/22.62	78.76/22.20	74.49/21.70	62.24/20.34	77.44/21.93	74.29/21.56	69.35/21.14
FastMRI test data							
cycleGAN	81.60/21.48	70.71/19.45	59.59/17.89	40.87/15.80	62.95/18.35	57.50/17.74	50.80/17.02
UGAC	85.29/22.62	72.17/19.91	59.64/17.99	43.21/16.24	72.01/20.03	63.97/18.73	56.36/17.71
gcGAN	80.73/20.76	68.67/18.13	59.23/16.60	48.31/15.26	71.51/18.61	65.22/17.54	58.66/16.61
CUT	89.36/23.37	85.67/22.71	83.32/22.15	76.63/20.58	86.61/23.02	84.45/22.45	82.59/21.86
PI(ours)	85.85/23.37	76.22/20.85	66.44/18.94	53.46/17.19	75.77/20.80	67.61/19.28	59.44/18.04
UAPI(ours)	90.37/24.74	81.08/21.37	69.97/18.41	55.72/16.07	79.18/19.81	71.69/17.97	65.44/16.98

The quantitative metrics in Table 1 obtained on unseen test data indicate superior performance of our approach for the modality propagation task on IXI. Considering evaluation on clean test data, usage of patch invariance gives an increase in SSIM and PSNR metrics compared to the bi-directional (cycleGAN, UGAC) and uni-directional (gcGAN, CUT) benchmarks. As compared to the benchmark methods, consideration of uncertainty-aware patch invariance even

² <https://github.com/taesungp/contrastive-unpaired-translation>.

improves results, also for the scenarios with perturbed test data (GN5 to IP10). Robustness of UAPI is also established visually in Fig. 2. Especially for scenarios with high perturbations (GN10, GN20, IP5, IP10), we observe a performance advantage when using the uncertainty-aware method UAPI. Interestingly, usage of PI without uncertainty awareness on perturbation scenarios GN5 to IP10 yields a significant decrease for the SSIM but not for the PSNR metric. This rather unexpected observation needs further investigation in future research.

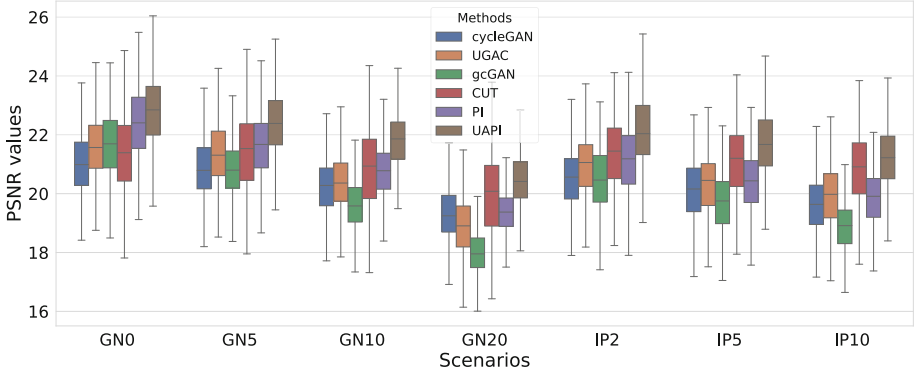


Fig. 2. Visual analysis of the PSNR values on IXI under seven different test scenarios, obtained by our two approaches (PI, UAPI) and 4 compared methods (cycleGAN, UGAC, gcGAN, CUT).

In Table 1 we observe for the accelerated MRI enhancement task on FastMRI that our method UAPI significantly outperforms other benchmark on unaffected test data but gives modest accuracy when additional noise and perturbed pixels are added. Figure 3 shows the superior performance of CUT in terms of the PSNR metric for noisy data. This is quite interesting since that has not been the case for the previous modality propagation application. The task of accelerated MRI enhancement strongly differs from the other two applications. While the goal of modality propagation and MRI-to-CT synthesis is to come up with a completely new image, the aim of accelerated MRI enhancement is to improve quality of a already existing image. In fact, the methods cycleGAN, UGAC, gcGAN, PI and UAPI depend on a rather simple U-Net [34] implementation and a standard DCGAN critic [35] with the aim to demonstrate plausibility of different transfer approaches on easy-to-implement frameworks. The CUT method is a benchmark where the publicly available source code had to be used, consisting of a ResNet-based generator [11] and built-in data augmentation techniques that may better compensate for noisy input data. Nevertheless, our methods PI and UAPI seemingly achieve better results compared to the U-Net based benchmarks. We will take up investigation of robustness of our methods in combination with different network architectures as a future goal.

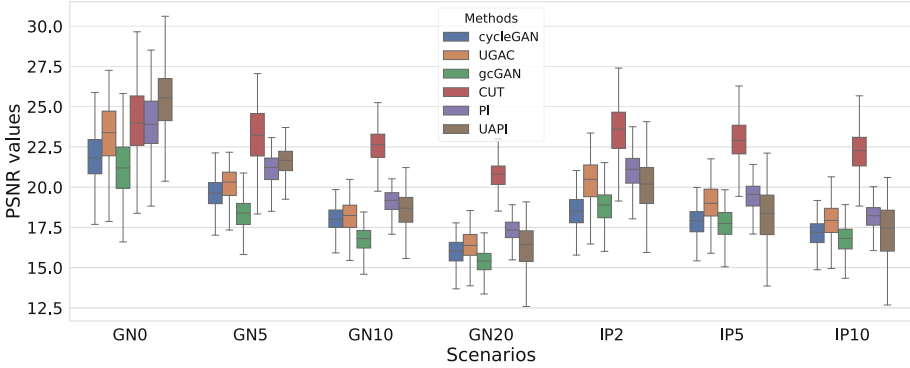


Fig. 3. Visual analysis of the PSNR values on FastMRI under seven different test scenarios, obtained by our two approaches (PI, UAPI) and 4 compared methods (cycleGAN, UGAC, gcGAN, CUT).

4.4 Qualitative Evaluation

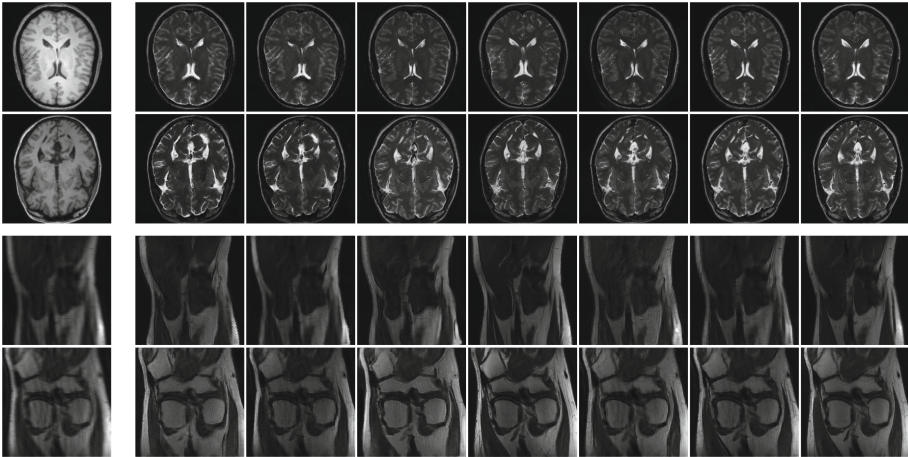


Fig. 4. Evaluation samples of our approach and four compared methods on IXI and FastMRI for scenario GN0 (test data without perturbations). From left to right: input, images transferred by cycleGAN, UGAC, gcGAN, CUT, PI, UAPI and ground truth.

In Fig. 4 we analyze the prediction quality of our and compared approaches in a qualitative way. Considering modality propagation in MRI, we see that usage of uncertainty-aware patch invariance (UAPI) gives a better detailed weighting of the cerebrospinal fluid in the middle of the brain. In general, employing patch invariance yields better preservation of fine structures. This observation also applies to accelerated MRI enhancement. In particular, CUT and UAPI provide

comparatively sharper knee images with more high-frequency details than the other methods.

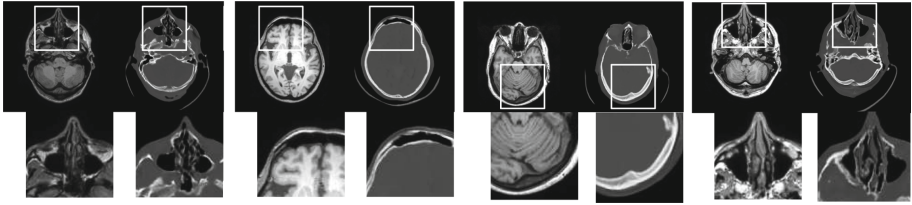


Fig. 5. Evaluation samples of the UAPI method on unseen brain MRI slices. For every data pair, the input slice and the corresponding UAPI prediction are visualized on the left and the right side, respectively. The first row contains the images on original scale, the second row selected patches to visualize the prediction quality for detailed structures.

Qualitative evaluation plays an important role for the third investigated application, namely MRI-to-CT synthesis, where quantitative comparison is not possible due to lack of ground truth data. Satisfying results were obtained with the UAPI method, which are visualized in Fig. 5. Cavities and brain shapes are well preserved by our method although we used two completely independent and unaligned head datasets for this experiment. UAPI synthesizes brain table artifacts that are also visible in CQ500. A proper evaluation on cleaned CT data is necessary and thus will be considered as a future working step.

4.5 Uncertainty Scores

Additional to improved accuracy we demonstrate the efficacy of estimating the scale maps with the proposed method. The input-dependent non-negative scale maps are derived from the second output branch G_{θ}^{σ} , see (9). Indeed, the predicted scale maps are able to model uncertainty inherent from data. This can be observed in Fig. 6, where in addition to the transferred images also the predicted scale maps and the absolute residuals between predicted and ground truth images are displayed. Obviously, uncertainty is relatively greater in regions with higher residual values. From the scale maps it can be deduced for which positions the generator is comparatively uncertain in its prediction, such as the cerebral cortex and eye sockets in head MRI or the lateral knee ligaments in knee MRI.

The correspondence between residual and scale maps suggests that the latter can be used as an approximation to a prediction’s residuals that are not available due to the lack of ground truth data in unsupervised learning. In order to quantitatively study this relationship we visualize mean absolute residual score and mean uncertainty maps for 512 randomly selected unseen test images in a scatter plot (see Fig. 7). Moreover, we compare our uni-directional method UAPI also to the relations observed by UGAC that models uncertainty with the help

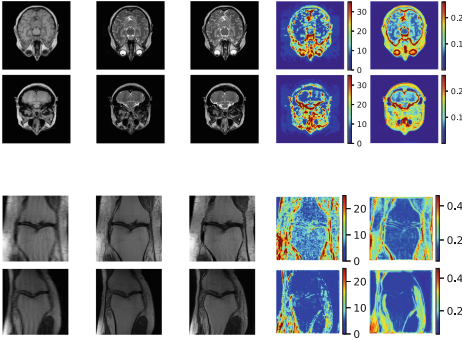


Fig. 6. Position-based relation between abs. residuals and predicted scale maps on IXI (top) and FastMRI (bottom). Left to right: input, ground truth, prediction by UAPI, abs. residuals and predicted scale map.

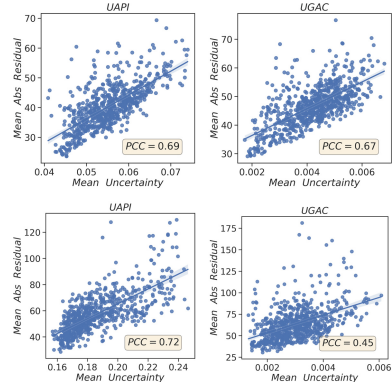


Fig. 7. Scatter plot between abs. residual and scale map values on IXI (top) and FastMRI (bottom). The predictions are generated by UAPI (left) and UGAC (right).

of a bi-directional cycleGAN [9]. For modality propagation as well as accelerated MRI enhancement we visually observe an approximate positive linear correlation between mean absolute residual scores and mean uncertainty scores. We calculate the Pearson correlation coefficient (PCC) to obtain a quality estimate for the linear correlation and compare between UAPI and UGAC. Our method returns a slightly higher PCC on IXI (UAPI: 0.69, UGAC: 0.67). The discrepancy between both methods even increases on FastMRI (UAPI: 0.72, UGAC: 0.45). This further encourages the idea that scale maps derived from our approach can be used to indicate the overall quality of a transferred image.

5 Conclusions

In this paper we proposed a WGAN-based approach using patch invariance to employ joint image transfer and uncertainty quantification in a fully unsupervised manner. We demonstrate superior performance of our uni-directional method for modality propagation and accelerated MRI enhancement compared to four state-of-the-art benchmarks in unpaired image translation. Moreover, the method reaches qualitatively satisfying results for MRI-to-CT synthesis using completely unaligned databases during training. The predicted uncertainty can be representative of the residual maps and thus indicate the quality of a transferred image in the absence of ground truth data. Further investigation of the network architecture and improvement in robustness represents an important goal for future research. Future work will also include the application of the uncertainty-aware and patch invariant network to other unpaired image-to-image applications outside the medical sector.

References

1. Wang, T., et al.: A review on medical imaging synthesis using deep learning and its clinical applications. *J. Appl. Clin. Med. Phys.* **22**(1), 11–36 (2021). <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/acm2.13121>
2. Upadhyay, U., Sudarshan, V.P., Awate, S.P.: Uncertainty-aware GAN with adaptive loss for robust MRI image enhancement. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), pp. 3248–3257 (2021)
3. Zbontar, J., Knoll, F., et al.: fastMRI: an open dataset and benchmarks for accelerated MRI (2018). <https://arxiv.org/abs/1811.08839>
4. Hyun, C.M., Kim, H.P., Lee, S.M., Lee, S., Seo, J.K.: Deep learning for undersampled MRI reconstruction. *Phys. Med. Biol.* **63**(13), 135007 (2018). <https://doi.org/10.1088/1361-6560/aac71a>
5. Lei, Y., et al.: MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. *Med. Phys.* **46**(8), 3565–3581 (2019)
6. Yang, H., et al.: Unpaired brain MR-to-CT synthesis using a structure-constrained CycleGAN. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 174–182. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_20
7. Wolterink, J.M., Dinkla, A.M., Savenije, M.H.F., Seevinck, P.R., van den Berg, C.A.T., Išgum, I.: Deep MR to CT synthesis using unpaired data. In: Tsaftaris, S.A., Gooya, A., Frangi, A.F., Prince, J.L. (eds.) SASHIMI 2017. LNCS, vol. 10557, pp. 14–23. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68127-6_2
8. Hiasa, Y., et al.: Cross-modality image synthesis from unpaired data using CycleGAN. In: Gooya, A., Goksel, O., Oguz, I., Burgos, N. (eds.) SASHIMI 2018. LNCS, vol. 11037, pp. 31–41. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00536-8_4
9. Upadhyay, U., Chen, Y., Akata, Z.: Robustness via uncertainty-aware cycle consistency. In: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (eds.) Advances in Neural Information Processing Systems, vol. 34, pp. 28261–28273. Curran Associates, Inc. (2021). <https://proceedings.neurips.cc/paper/2021/file/ede7e2b6d13a41ddf9f4bdef84fdc737-Paper.pdf>
10. Goodfellow, I., et al.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K. (eds.) Advances in Neural Information Processing Systems, vol. 27. Curran Associates, Inc. (2014). <https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf>
11. Zhu, J., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2242–2251 (2017)
12. Fu, H., Gong, M., Wang, C., Batmanghelich, K., Zhang, K., Tao, D.: Geometry-consistent generative adversarial networks for one-sided unsupervised domain mapping. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2427–2436 (2019)
13. Benaim, S., Wolf, L.: One-sided unsupervised domain mapping. *Adv. Neural Inf. Process. Syst.* **30** (2017)
14. Park, T., Efros, A.A., Zhang, R., Zhu, J.-Y.: Contrastive learning for unpaired image-to-image translation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12354, pp. 319–345. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58545-7_19

15. Kendall, A., Gal, Y.: What uncertainties do we need in Bayesian deep learning for computer vision? *Adv. Neural Inf. Process. Syst.* **30** (2017)
16. Saatchi, Y., Wilson, A.G.: Bayesian GAN. *Adv. Neural Inf. Process. Syst.* **30** (2017)
17. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, pp. 214–223. PMLR (2017)
18. Robinson, E.C., Hammers, A., Ericsson, A., Edwards, A.D., Rueckert, D.: Identifying population differences in whole-brain structural networks: a machine learning approach. *Neuroimage* **50**(3), 910–919 (2010)
19. Chilamkurthy, S., et al.: Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet* **392**(10162), 2388–2396 (2018)
20. Angermann, C., Schwab, M., Haltmeier, M., Laubichler, C., Jónsson, S.: Unsupervised single-shot depth estimation using perceptual reconstruction. *CoRR abs/2201.12170* (2022). <https://arxiv.org/abs/2201.12170>
21. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of wasserstein GANs. In: Guyon, I., et al. (eds.) *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc. (2017). <https://proceedings.neurips.cc/paper/2017/file/892c3b1c6dccc52936e27cbd0ff683d6-Paper.pdf>
22. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2794–2802 (2017)
23. Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y.: Spectral normalization for generative adversarial networks. In: *International Conference on Learning Representations* (2018)
24. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4401–4410 (2019)
25. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation. In: *International Conference on Learning Representations* (2018)
26. Ledig, C., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4681–4690 (2017)
27. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134 (2017)
28. Arslan, A.T., Seke, E.: Face depth estimation with conditional generative adversarial networks. *IEEE Access* **7**, 23222–23231 (2019)
29. Armanious, K., et al.: MedGAN: medical image translation using GANs. *Comput. Med. Imaging Graph.* **79**, 101684 (2020)
30. Abdar, M., et al.: A review of uncertainty quantification in deep learning: techniques, applications and challenges. *Inf. Fusion* **76**, 243–297 (2021)
31. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: Balcan, M.F., Weinberger, K.Q. (eds.) *Proceedings of The 33rd International Conference on Machine Learning. Proceedings of Machine Learning Research*, vol. 48, pp. 1050–1059. PMLR, New York (2016). <https://proceedings.mlr.press/v48/gal16.html>
32. Palakkadavath, R., Srijith, P.: Bayesian generative adversarial nets with dropout inference. In: *8th ACM IKDD CODS and 26th COMAD*, pp. 92–100 (2021)

33. Villani, C.: *Optimal Transport: Old and New*, vol. 338. Springer, Heidelberg (2008). <https://doi.org/10.1007/978-3-540-71050-9>
34. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
35. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks (2015)
36. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization (2014). <https://arxiv.org/abs/1412.6980>
37. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)