



# Semi-supervised Learning Approach to Generate Neuroimaging Modalities with Adversarial Training

Harrison Nguyen<sup>(✉)</sup>, Simon Luo, and Fabio Ramos

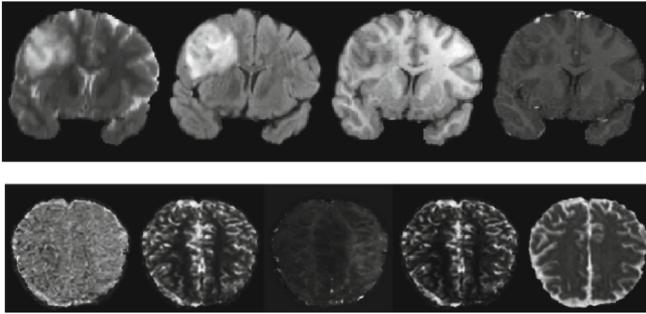
The University of Sydney, Sydney, NSW 2006, Australia  
{harrison.nguyen,fabio.ramos}@sydney.edu.au, sluo4225@uni.sydney.edu.au

**Abstract.** Magnetic Resonance Imaging (MRI) of the brain can come in the form of different modalities such as T1-weighted and Fluid Attenuated Inversion Recovery (FLAIR) which has been used to investigate a wide range of neurological disorders. Current state-of-the-art models for brain tissue segmentation and disease classification require multiple modalities for training and inference. However, the acquisition of all of these modalities are expensive, time-consuming, inconvenient and the required modalities are often not available. As a result, these datasets contain large amounts of *unpaired* data, where examples in the dataset do not contain all modalities. On the other hand, there is smaller fraction of examples that contain all modalities (*paired* data) and furthermore each modality is high dimensional when compared to number of datapoints. In this work, we develop a method to address these issues with semi-supervised learning in translating between two neuroimaging modalities. Our proposed model, Semi-Supervised Adversarial CycleGAN (SSA-CGAN), uses an adversarial loss to learn from *unpaired* data points, cycle loss to enforce consistent reconstructions of the mappings and another adversarial loss to take advantage of *paired* data points. Our experiments demonstrate that our proposed framework produces an improvement in reconstruction error and reduced variance for the pairwise translation of multiple modalities and is more robust to thermal noise when compared to existing methods.

**Keywords:** GAN · MRI · Semi-supervised learning

## 1 Introduction

Magnetic Resonance Imaging (MRI) of the brain has been used to investigate a wide range of neurological disorders and depending on the imaging sequence used, can produce different modalities such as T1-weighted images, T2-weighted images, Fluid Attenuated Inversion Recovery (FLAIR), and diffusion weighted imaging (DWI). Each of these modalities produce different contrast and brightness of brain tissue that could reveal pathological abnormalities. Many of the advances in the use of data-driven models in Alzheimer's disease classification [17], brain tumour segmentation [9] and skull stripping methods [18], rely on deep



**Fig. 1. Top:** A coronal slice of a low grade glioma (brain tumour) in the BraTs dataset in different modalities. From left to right: T2, Fluid-attenuated inversion recovery (FLAIR), T1 and T1c. **Bottom:** Axial slices of modalities of a CT perfusion scan of an ischemic stroke lesion patient in the ISLES dataset. From left to right: Mean Transit Time (MTT), cerebral blood flow (CBF), time to peak of the residue function (Tmax), cerebral blood volume (CBV), Apparent diffusion coefficient (ADC).

convolutional neural networks (DCNN). In particular, datasets such as BraTs [23] and ISLES [19] have been focusing on the evaluation of state-of-the-art methods for the segmentation of brain tumours and stroke lesions respectively. These methods do not require the use of hand designed features and instead are able to learn a hierarchy of increasingly complex features. However, they require multiple neuroimaging modalities for high performance and improved sensitivity [4] (See Fig. 1). Collecting multiple modalities for each patient can be difficult, expensive and not all of these modalities are available in clinical settings. In particular, *paired* data, where an example has all modalities present, is difficult to access, making these data dependent models more difficult to train or reduce their applicability during inference.

To ensure each modality is present, the missing modality could be imputed through a domain adaptation model where characteristics of one image set is transferred into another image set (*e.g.* T1-weighted to T2-weighted) that has been learned from existing *paired* examples. However, since this *paired* data is limited in the neuroimaging context, learning from examples that do not have all modalities (*unpaired* data) is valuable as this form of data is more readily available.

There has been significant interest in unsupervised image-to-image translation where *paired* training data is not available but two distinct image sets. Methods proposed by Zhu *et al.* [37] and Hoffman *et al.* [11] assume the two image collections are representations of some shared, underlying state. They use adversarial training which discriminates at the image level to guide the transformation between the domains. Furthermore, the translations between these two sets should have approximately invertible solutions and should be *cycle consistent*—where the mapping of a particular source domain to the target domain and back should yield the original source at the pixel level. Alternative methods extract domain invariant features with DCNNs and discriminate the feature distributions of source/target domains [32].

One work in recent literature that exploits the two distinct image sets of *unpaired* data, in order to improve the performance on tasks with a scarcity of *paired* data is the Cycle Wasserstein Regression GAN [22] (CWRG). The CWRG uses the  $l_2$ -norm as a penalty term for the reconstruction of *paired* data along with the adversarial signal and cycle-loss of the CycleGAN. However, the CWRG demonstrated its performance on ICU timeseries data and transcriptomics data and not on image data.

Our proposed method, the *Semi-Supervised Adversarial CycleGAN (SSA-CGAN)* further extends the application of leveraging *unpaired* data and *paired* data to MRI image translation, where the dimensionality of the examples is orders of magnitude larger. Our method uses multiple adversarial signals for semi-supervised bi-directional image translation. Our experimental results have demonstrated that our proposed approach has superior performance compared to the CycleGAN and CWRG in terms of average reconstruction error and variance and as well as robustness to noise when evaluated using the BraTs and ISLES dataset.

## 2 Related Work

General adversarial networks (GAN) have received significant attention since the work by [8] and various GAN-based models have achieved impressive results in image generation [5] and representation learning [28]. These models learn a generator to capture the distribution of real data by introducing a competing model, the discriminator, that evolves to distinguish between the real data and the fake data produced by the generator. This forces the generated image to be indistinguishable from real images.

Various conditional GANs (cGAN) have been adapted to condition the image generator on images instead of a noise vector to be used in applications such as style transfer from normal maps to images [33]. Isola *et al.*'s [12] work in particular, uses labeled image pairs to train a cGAN to learn a mapping between the two image domains. On the other hand, there have been significant works that have tackled image-to-image translation in the unpaired setting. The CycleGAN [37] uses a *cycle consistency loss* to ensure the forward mapping and back results in the original image. It has demonstrated success in tasks where paired training data is limited *e.g.* in painting style and season transfer. The Dual GAN, being inspired by dual learning in machine translation used a similar loss objective, where the reconstruction error is used to measure the disparity between the reconstructed object and the original [36]. Unlike the previous two frameworks, the CoGAN [16] and cross-modal scene networks [1] does not use a cycle consistency loss but instead, uses weight sharing between the two GANs, corresponding to high level semantics to learn a common representation across domains.

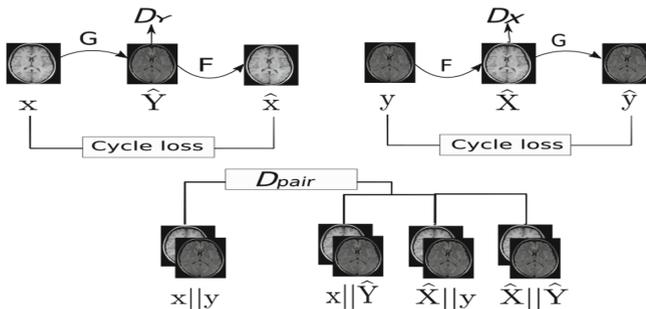
GANs have been used in the semi-supervised learning (SSL) context as the visually realistic images generated can be used as additional training data. Salimans *et al.* [29] proposed techniques to improve training GANs which included learning a discriminator on additional class labels which can be used for SSL. Mayato *et al.* [24] modified the adversarial objective to a regularization method

based on virtual adversarial loss. The method probabilistically produces labels that are unknown to the user and computes the adversarial direction based on the virtual labels. Park *et al.* [26] improves upon the performance of virtual adversarial training by using adversarial dropout which maximizes the divergence between the training supervision and the outputs from the network with the dropout.

GANs have been used in a range of applications in biomedical imaging such as the generation of multi modal MRI images and retinal fundus images [2], to detect anomalies in retinal OCT images [30] and image synthesis of MR and CT images [35]. Adversarial methods have also been extended to domain adaptation for medical imaging. Chen *et al.* [3] recently developed the Synergistic Image and Feature Adaptation framework that enhances domain-invariance through feature encoder layers that are shared by the target and source domain and uses additional discriminator to differentiate the feature distributions. Perone *et al.* forgoes the use of adversarial training and instead demonstrates application of self ensembling and mean teacher framework [27].

The CycleGAN has been recently applied to the biomedical field for translating between sets of data. Welander *et al.* [34] investigated the difference between the CycleGAN and UNIT [15] for the translation between T1 and T2 MRI modalities and found the CycleGAN was the better alternative if the aim was to generate visually realistic images as possible. McDermott *et al.* [22] on the other hand, tackled domain adaptation in the semi-supervised setting by proposing Wasserstein CycleGANs coupled with a  $l_2$  regression loss function on paired data. The semi-supervised setting for this paper is similar to McDermott *et al.*, however we propose an adversarial training signal for paired data instead of the  $l_2$  loss. We demonstrate our method produces better reconstructions with lower variance and is more robust to noise in the context of translating between neuroimaging modalities compared to existing methods.

### 3 Methods



**Fig. 2.** Our model is composed of the CycleGAN architecture and an auxiliary discriminator which takes as input concatenated paired examples and the concatenation of generators' various transformations.

### 3.1 CycleGAN

The CycleGAN [37] learns to translate points between two domains  $X$  and  $Y$ . Given two sets of unlabeled and *unpaired* images,  $\{x_i\}_{i=1}^N$  where  $x_i \in X$  and  $\{y_j\}_{j=1}^M$ ,  $y_j \in Y$ , two generators,  $F$  and  $G$ , are trained to learn mapping functions  $G : X \rightarrow Y$  and  $F : Y \rightarrow X$ , where  $F$  and  $G$  are usually represented by DCNNs. Furthermore, two discriminators  $D_X$  and  $D_Y$  are trained where  $D_X$  learns to distinguish between images  $\{x\}$  and  $\{F(y)\}$  and  $D_Y$  discriminates between  $\{y\}$  and  $\{G(x)\}$ . Instead of the original GAN loss, the CycleGAN trains discriminators using the least squares loss function proposed by Mao *et al.* [20]. For example,  $D_X$  minimises the following objective function:

$$\mathcal{L}_{D_X} = \mathbb{E}_{\mathbf{x} \sim P(\mathbf{x})} [(D_X(\mathbf{x}) - 1)^2] + \mathbb{E}_{\mathbf{y} \sim P(\mathbf{y})} [(D_X(F(\mathbf{y})))^2]. \quad (1)$$

Conversely the generator,  $F$ , for example is trained according to the following *adversarial loss*,

$$\mathcal{L}_{F_{adv}} = \mathbb{E}_{\mathbf{y} \sim P(\mathbf{y})} [(D_X(F(\mathbf{y})) - 1)^2], \quad (2)$$

as well as a *cycle-consistency loss* where reconstruction error between the inverse mapping and the original point is minimised [37],

$$\mathcal{L}_{cyc} = \mathbb{E}_{\mathbf{x} \sim P(\mathbf{x})} [\|F(G(\mathbf{x})) - \mathbf{x}\|_1] + \mathbb{E}_{\mathbf{y} \sim P(\mathbf{y})} [\|G(F(\mathbf{y})) - \mathbf{y}\|_1]. \quad (3)$$

The overall loss function for the generator is therefore given as

$$\mathcal{L}_F = \mathcal{L}_{F_{adv}} + \lambda \mathcal{L}_{cyc}, \quad (4)$$

where  $\lambda$  controls the relative strength between the adversarial signal and the cycle-consistency loss.

### 3.2 Semi-Supervised Adversarial CycleGAN

We extend the CycleGAN through the Semi-Supervised Adversarial CycleGAN (SSA-CGAN) to take advantage of *paired* training data. In our scenario we have additional information in the form of  $T$  paired examples  $\{\mathbf{x}_p, \mathbf{y}_p\}_{p=1}^T$ , a subset  $P \subseteq X \times Y$ . We seek to take advantage of this paired information through an auxiliary adversarial network,  $D_{pair}$  (See Fig. 2).  $D_{pair}$  takes as input, only the paired examples from  $P$  and the concatenations of the following transformations: a)  $\mathbf{x}_p$  and  $\mathbf{y}_p$ , b)  $\mathbf{x}_p$  and  $G(\mathbf{x}_p)$ , c)  $F(\mathbf{y}_p)$  and  $\mathbf{y}_p$ , d)  $F(\mathbf{y}_p)$  and  $G(\mathbf{x}_p)$ .  $D_{pair}$  attempts to discriminate between the ground-truth pairs,  $\{\mathbf{x}_p, \mathbf{y}_p\} \in P$ , as real and the transformation of the image and its respective real image as fake. Therefore, the paired discriminator minimises

$$\begin{aligned} \mathcal{L}_{D_{pair}} = & \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim P_{pair}(\mathbf{x}, \mathbf{y})} [(D_{pair}(\mathbf{x}, \mathbf{y}) - 1)^2] + \frac{1}{3} \left[ \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim P_{pair}} [D_P(\mathbf{x}, G(\mathbf{x}))^2] \right. \\ & \left. + \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim P_{pair}} [D_{pair}(F(\mathbf{y}), \mathbf{y})^2] + \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim P_{pair}} [D_{pair}(F(\mathbf{y}), G(\mathbf{x}))^2] \right] \end{aligned} \quad (5)$$

and  $F$ 's loss is

$$\mathcal{L}_{F_{semi}} = \mathcal{L}_{F_{adv}} + \lambda \mathcal{L}_{cyc} + \alpha \mathcal{L}_{pair}, \quad (6)$$

where  $\mathcal{L}_{pair}$  is given as

$$\begin{aligned} \mathcal{L}_{pair} = & \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim P_{pair}} [(D_{pair}(\mathbf{x}, G(\mathbf{x})) - 1)^2] + \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim P_{pair}} [(D_{pair}(F(\mathbf{y}), \mathbf{y}) - 1)^2] \\ & + \mathbb{E}_{\mathbf{x}, \mathbf{y} \sim P_{pair}} [(D_{pair}(F(\mathbf{y}), G(\mathbf{x})) - 1)^2]. \end{aligned} \quad (7)$$

and  $\alpha$  and  $\lambda$  control the relative weight of the losses. The third loss term can be seen as further regularisation of the generators where its forward and backward transformations are pushed towards the joint distribution of  $X$  and  $Y$ .

## 4 Experiments

### 4.1 Dataset

We evaluate our method using BraTS and ISLES datasets which have been used to evaluate state-of-the-art methods for the segmentation of brain tumours and lesions respectively. BraTS utilizes multi-institutional pre-operative MRI scans and focuses on the segmentation of intrinsically heterogeneous (in appearance, shape, and histology) brain tumors, namely gliomas. This proposed method is trained and tested on the BraTs 2018 dataset. The training dataset contains 285 examples including 210 High GradeGlioma (HGG) cases and 75 cases with Low Grade Glioma (LGG). For each case, there are four MRI sequences, including the T1-weighted (T1), T1 with gadolinium enhancing contrast (T1c), T2-weighted (T2) and FLAIR. The dataset includes pre-processing methods such as skullstrip, co-register to a common space and resample to isotropic  $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$  resolution. Bias field correction is done on the MR data to correct the intensity in-homogeneity in each channel using N4ITK tool [31].

The dataset was divided as the following: 30% of examples was designated as *unpaired* examples of domain  $X$  (e.g. T2-weighted volumes) and 30% as *unpaired* examples of domain  $Y$  (e.g. T1-weighted), 10% was designated as *paired* training examples where each example, for example, had both T2-weighted and T1-weighted modalities. 10% was reserved as a held-out validation set for hyperparameter tuning and 20% was reserved to be a test set used for evaluation.

ISLES contains patients who have received the diagnosis of ischemic stroke by MRI. Ischemic stroke is the most common cerebrovascular disease and one of the most common causes of death and disability worldwide [25]. The stroke MRI was performed on either a 1.5T (Siemens Magnetom Avanto) or 3T MRI system (Siemens Magnetom Trio). Sequences and derived maps were cerebral blood flow (CBF), cerebral blood volume (CBV), time-to-peak (TTP), and time-to-max (Tmax) and mean transit time (MTT). The dataset included images that were rigidly registered to the T1c with constant resolution of  $2 \text{ mm} \times 2 \text{ mm} \times 2 \text{ mm}$  and automatically skull-stripped [19]. The dataset includes 38 patients in total and was divided in similar proportions as the BraTS experiment regime.

Further pre-processing for each dataset included each image modality was normalized by subtracting the mean and dividing by the standard deviation of the intensities within the volume and rescaled to values between 1 and  $-1$ . The volumes were reshaped to  $240 \times 240$  coronal and  $128 \times 128$  axial slices for the BraTS and ISLES dataset respectively. This resulted in an average of 170 slices per patient for the BraTS dataset and 18 slices per patient in ISLES.

## 4.2 Implementation

**Network Architecture:** The generator network was adapted from Johnson *et al.* [13] and Zhu *et al.* [37]. The network contains two stride-2 convolutions, 6 residual blocks [10] and two fractionally strided convolutions with stride  $\frac{1}{2}$ . The single input discriminator networks is a PatchGAN. The paired input discriminator was a two stride-2 convolution layers. It used the concatenation of feature maps from the second last layer of  $D_X$  and  $D_Y$  as inputs as a form of weight sharing with the single image discriminators.

**Training Details:** For all the experiments, we set  $\lambda = 10$  and  $\alpha = 2$  in Eq. 6 chosen by the performance on the held out validation set averaged across the pairs of MR modalities mentioned in Sect. 4.3. All networks were trained from scratch using NVIDIA V100 GPU with an initial learning rate of  $2 \times 10^{-4}$ , weights were initialised using Glorot initialization [6] and optimised using Adam [14] with a batch size of 1. The learning rate was kept constant for the first 100 epochs and was linearly decreased thereafter to a learning rate of  $2 \times 10^{-7}$ . Training was finished after 200 epochs. While standard data augmentation procedures randomly shift, rotate and scale images, the images were only augmented by random shifting during training as the volumes were normalised to the same orientation and shape due to co-registration.

## 4.3 Evaluation Metrics

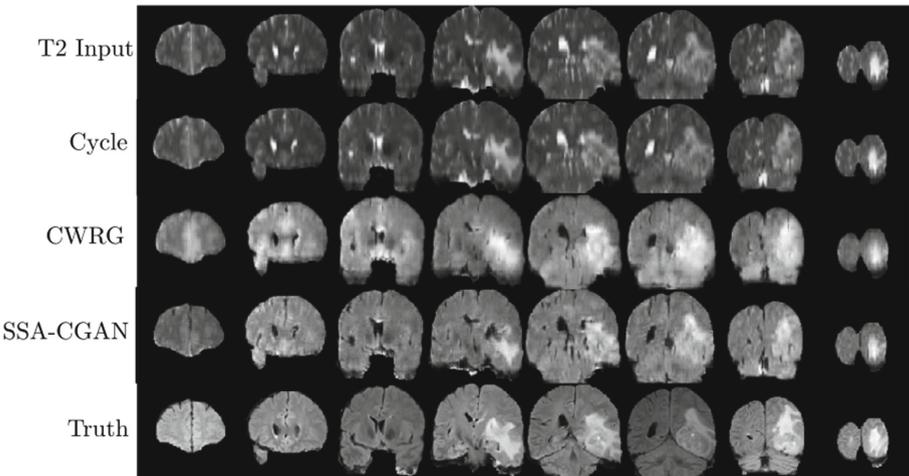
We evaluated the *SSA-CGAN* by learning a separate model for the following pairs of MR modalities: T2→T1, T2→T1c, T2→FLAIR, CBF→MTT, CBF→CBV, CBF→TTP, CBF→Tmax. For example, T2→T1 indicates the models were evaluated on the reconstruction of a T1 volume when transformed from a T2 volume. This was evaluated against the CycleGAN and the Cycle Wasserstein Regression GAN [22] (CWRG) which is currently the only other method in recent literature that combines *unpaired* and *paired* training data for translation between different modalities. We also included in our experiments using the *SSA-CGAN* framework using only *paired* data, labelled *SSA-CGAN-p*. On the other hand, our proposed method, *SSA-CGAN* uses *paired* data and leverages *unpaired* data to improve learning. The hyperparameter settings for each method is similar to the training details mentioned in Sect. 4.2. For each transformation (*e.g.* T2→T1c) and for each method, five networks were learned, each with different initialization of weights. These models were compared based on two quantitative metrics, the mean squared error (MSE) and mean absolute error (MAE) averaged across the five runs and its standard deviation.

#### 4.4 Results

Results for the performance of *SSA-CGAN* are shown in Table 1. We observe that the *SSA-CGAN* yields from a 8.32% reduction from the CycleGAN (T2 to T1) up to a 89.6% decrease in MSE in the case of CBF to CBV with an average reduction of 33.8% and 46.0% in MAE and MSE respectively across all transformations. The consistent out-performance of our method over the CycleGAN demonstrate there is potential gains when the information from paired data points can be leveraged. This is further emphasised by the improvement over *SSA-CGAN-p* which has been trained using only *paired* data. By leveraging *unpaired* data during training, the *SSA-CGAN* produces a reduction of 18.02% and 28.16% in MAE and MSE on average when compared to *SSA-CGAN-p*. *SSA-CGAN* produces a lower MSE in most cases despite CWRG includes a loss component that minimises the  $l_2$  norm. Furthermore, *SSA-CGAN* produces lower variance compared to other methods demonstrating that our method is less sensitive to different weight intializations and improves the stability of training and convergence.

**Table 1.** MSE and MAE for various paired transformations averaged across five runs with one standard deviation.

	Method	T1	T1c	FLAIR	MTT	rCBV	TTP	Tmax
MSE	Cycle	0.0314 ± 0.0006	0.5301 ± 0.4880	0.7072 ± 0.3956	0.1280 ± 0.1603	0.2437 ± 0.3111	0.0616 ± 0.0017	0.1887 ± 0.1565
	CWRG	0.7503 ± 0.1687	0.4607 ± 0.3602	0.6145 ± 0.4279	0.5803 ± 0.2688	0.6826 ± 0.2604	0.5785 ± 0.2945	0.4825 ± 0.1722
	SSA-CGAN-p	0.0234 ± 0.0032	0.0160 ± 0.0100	<b>0.0147 ± 0.0018</b>	0.0503 ± 0.0051	0.0262 ± 0.0017	0.0443 ± 0.0085	0.0348 ± 0.0021
	SSA-CGAN	<b>0.0169 ± 0.0011</b>	<b>0.0102 ± 0.0024</b>	0.0177 ± 0.0071	<b>0.0271 ± 0.0007</b>	<b>0.0202 ± 0.0014</b>	<b>0.0210 ± 0.0011</b>	<b>0.0235 ± 0.0041</b>
MAE	Cycle	0.0608 ± 0.0041	0.4924 ± 0.4146	0.6231 ± 0.3264	0.2162 ± 0.1610	0.4236 ± 0.2957	0.1409 ± 0.0022	0.3048 ± 0.1939
	CWRG	0.6963 ± 0.3738	0.4564 ± 0.3868	0.5603 ± 0.5564	0.6819 ± 0.1240	0.7008 ± 0.1478	0.5258 ± 0.2860	0.5189 ± 0.2800
	SSA-CGAN-p	0.0508 ± 0.0037	0.0411 ± 0.0118	<b>0.0390 ± 0.0028</b>	0.1322 ± 0.0059	0.0834 ± 0.0029	0.1155 ± 0.0118	0.0837 ± 0.0048
	SSA-CGAN	<b>0.0436 ± 0.0011</b>	<b>0.0338 ± 0.0046</b>	0.0426 ± 0.0089	<b>0.0947 ± 0.0018</b>	<b>0.0720 ± 0.0043</b>	<b>0.0754 ± 0.0026</b>	<b>0.0613 ± 0.0069</b>



**Fig. 3.** A comparison of the transformation from T2 to FLAIR.

Figure 3 and 4 shows a comparison of the transformation from T2 to FLAIR and MTT to CBF respectively, of a particular chosen MR scan produced by the various models. The CycleGAN produces no noticeable change from the input image and the CWRG creates a smoothed version of the ground truth. This can be attributed to the MSE component of the objective function where the MSE pushes the generator to produce blurry images [21]. The additional adversarial component of our method forces the generator to synthesise a more visually realistic image. However, in Fig. 3 the image produced does not match the pixel intensity of the ground truth and in Fig. 4, fails to capture the high detail and edges of the CBF modality and fails to distinguish between background and low intensity areas.

### 4.5 Robustness to Noise

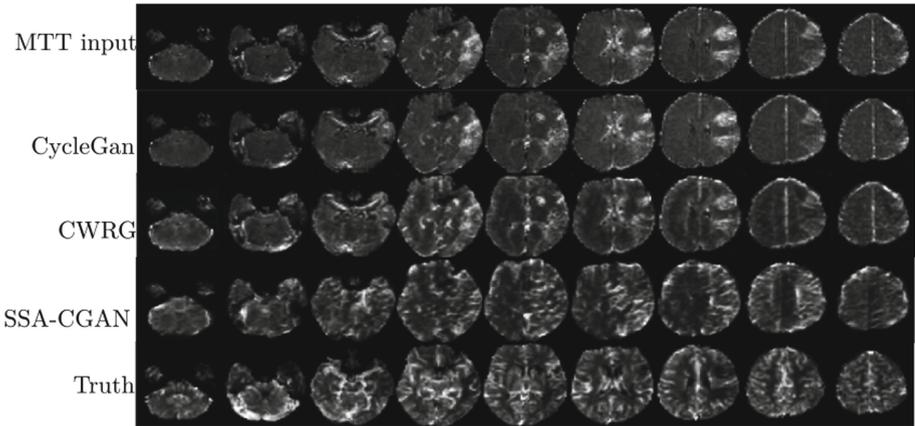


Fig. 4. A comparison of the transformation from MTT to CBF.

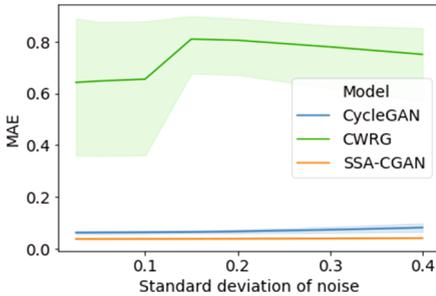
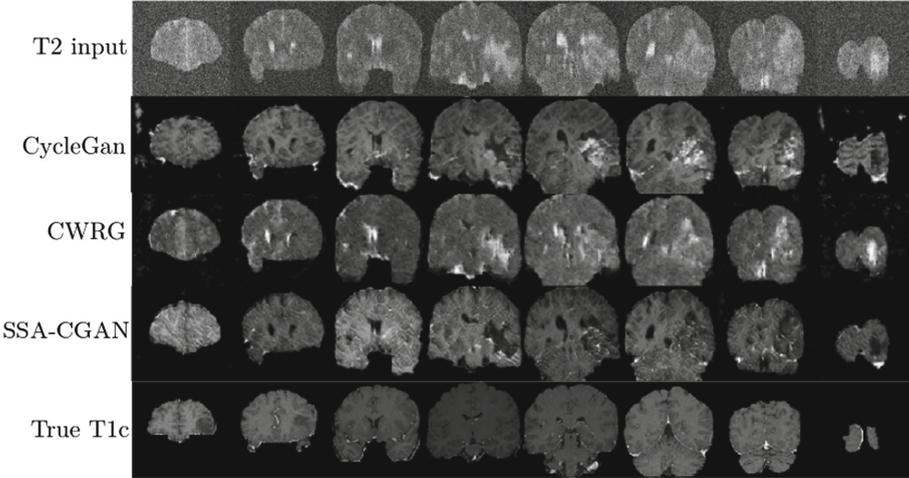


Fig. 5. Quantitative comparison of the reconstruction error by varying the amount of random noise injected to test data.

The methods were assessed by injecting random Gaussian noise into the test data to simulate thermal noise conditions to evaluate the robustness of the models, despite not being trained on noisy examples. Various levels of noise was injected to the data, ranging from a standard deviation of 0.025 to 0.4. The predictions of the models was evaluated against the ground truth. Figure 5 shows the comparison between the models, with

the MAE as the evaluation metric. At all noise levels, the *SSA-CGAN* outperforms other methods with lower variance further demonstrating the robustness of our method.



**Fig. 6.** A T2 image was corrupted with Gaussian noise and was transformed to a T1c image by the various models.

The methods were also visually evaluated under extreme simulated thermal noise conditions by adding Gaussian noise with mean 0 standard deviation of 0.2 to the input. Figure 6 shows the transformation produced by a noisy input volume to the networks. The CWRG produces noise filtered version of the T2 scan and fails to perform the transformation to T1c. Our method and the CycleGAN shows robustness under the extreme scenario and fabricates successful slices. However, it fails to hide the tumour in the T2 scan (the bright spot in bottom right) in the T1c reconstruction and instead substitutes background for that tumour.

#### 4.6 Limitations

This approach has several limitations. Due to the additional discriminator that distinguishes paired examples, additional computational time is required for training. Second, adversarial networks remain a very active area of research, and are known to be difficult to train and suffer issues such as mode collapse [7]. Further work would be to investigate the effect on performance when the fraction of paired examples changes and the point where the paired-input discriminator fails to be effective.

## 5 Conclusion

Many state-of-the-art models in brain tissue segmentation and disease classification require multiple modalities during training and inference. However, examples where all modalities are available is limited and therefore the ability to incorporate unpaired data could be important for the adoption of these methods in clinical settings or improve existing models. Furthermore, the overall data available in limited and MRI volumes are high dimensional. The *Semi-Supervised Adversarial CycleGAN* (SSA-CGAN) learns translations between neuroimaging modalities using *unpaired* data and *paired* examples through a *cycle-consistency* loss, an adversarial signal for the discrimination between generated and real images of each domain and an additional adversarial signal that discriminates between the pairs of real data and pairs of generated images. Our experimental results have demonstrated that *SSA-CGAN* has superior results in achieving lower reconstruction error and is more robust compared to all of current state-of-the-art approaches across a wide range of modality translations.

## References

1. Aytar, Y., Castrejon, L., Vondrick, C., Pirsiavash, H., Torralba, A.: Cross-modal scene networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(10), 2303–2314 (2018)
2. Beers, A., et al.: High-resolution medical image synthesis using progressively grown generative adversarial networks. arXiv preprint [arXiv:1805.03144](https://arxiv.org/abs/1805.03144) (2018)
3. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Synergistic image and feature adaptation: towards cross-modality domain adaptation for medical image segmentation. arXiv preprint [arXiv:1901.08211](https://arxiv.org/abs/1901.08211) (2019)
4. Dai, D., Wang, J., Hua, J., He, H.: Classification of ADHD children through multimodal magnetic resonance imaging. *Front. Syst. Neurosci.* **6**, 63 (2012)
5. Denton, E.L., Chintala, S., Fergus, R., et al.: Deep generative image models using a Laplacian pyramid of adversarial networks. In: *Advances in Neural Information Processing Systems*, pp. 1486–1494 (2015)
6. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–256 (2010)
7. Goodfellow, I.: NIPS 2016 tutorial: generative adversarial networks. arXiv preprint [arXiv:1701.00160](https://arxiv.org/abs/1701.00160) (2016)
8. Goodfellow, I., et al.: Generative adversarial nets. In: *Advances in Neural Information Processing Systems*, pp. 2672–2680 (2014)
9. Havaei, M., et al.: Brain tumor segmentation with deep neural networks. *Med. Image Anal.* **35**, 18–31 (2017)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
11. Hoffman, J., et al.: CyCADA: cycle-consistent adversarial domain adaptation. arXiv preprint [arXiv:1711.03213](https://arxiv.org/abs/1711.03213) (2017)
12. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. arXiv preprint (2017)

13. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 694–711. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46475-6\\_43](https://doi.org/10.1007/978-3-319-46475-6_43)
14. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
15. Liu, M.Y., Breuel, T., Kautz, J.: Unsupervised image-to-image translation networks. In: Advances in Neural Information Processing Systems, pp. 700–708 (2017)
16. Liu, M.Y., Tuzel, O.: Coupled generative adversarial networks. In: Advances in Neural Information Processing Systems, pp. 469–477 (2016)
17. Lu, D., Popuri, K., Ding, G.W., Balachandar, R., Beg, M.F.: Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer’s disease using structural MR and FDG-PET images. *Sci. Rep.* **8**(1), 5697 (2018)
18. Mahbod, A., Chowdhury, M., Smedby, Ö., Wang, C.: Automatic brain segmentation using artificial neural networks with shape context. *Pattern Recogn. Lett.* **101**, 74–79 (2018)
19. Maier, O., et al.: ISLES 2015-A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Med. Image Anal.* **35**, 250–269 (2017)
20. 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)
21. Mathieu, M., Couprie, C., LeCun, Y.: Deep multi-scale video prediction beyond mean square error. arXiv preprint [arXiv:1511.05440](https://arxiv.org/abs/1511.05440) (2015)
22. McDermott, M.B., et al.: Semi-supervised biomedical translation with cycle Wasserstein regression GANs. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
23. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**(10), 1993–2024 (2015)
24. Miyato, T., Maeda, S.I., Ishii, S., Koyama, M.: Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 1979–1993 (2018)
25. World Health Organization: Cause-specific mortality, estimates for 2000–2012 (2012). [http://www.who.int/healthinfo/global\\_burden\\_disease/estimates/en/index1.html](http://www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html)
26. Park, S., Park, J., Shin, S.J., Moon, I.C.: Adversarial dropout for supervised and semi-supervised learning. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
27. Perone, C.S., Ballester, P., Barros, R.C., Cohen-Adad, J.: Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. arXiv preprint [arXiv:1811.06042](https://arxiv.org/abs/1811.06042) (2018)
28. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint [arXiv:1511.06434](https://arxiv.org/abs/1511.06434) (2015)
29. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training GANs. In: Advances in Neural Information Processing Systems, pp. 2234–2242 (2016)
30. Seeböck, P., et al.: Unsupervised identification of disease marker candidates in retinal OCT imaging data. *IEEE Trans. Med. Imaging* **38**, 1037–1047 (2018)
31. Tustison, N.J., et al.: N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**(6), 1310 (2010)

32. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7167–7176 (2017)
33. Wang, X., Gupta, A.: Generative image modeling using style and structure adversarial networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9908, pp. 318–335. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46493-0\\_20](https://doi.org/10.1007/978-3-319-46493-0_20)
34. Welander, P., Karlsson, S., Eklund, A.: Generative adversarial networks for image-to-image translation on multi-contrast MR images—a comparison of CycleGAN and UNIT. arXiv preprint [arXiv:1806.07777](https://arxiv.org/abs/1806.07777) (2018)
35. 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](https://doi.org/10.1007/978-3-030-00889-5_20)
36. Yi, Z., Zhang, H.R., Tan, P., Gong, M.: DualGAN: unsupervised dual learning for image-to-image translation. In: ICCV, pp. 2868–2876 (2017)
37. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv preprint (2017)