CT-To-MR Conditional Generative Adversarial Networks for Improved Stroke Lesion Segmentation

Jonathan Rubin[1] JONATHAN.RUBIN@PHILIPS.COM
S. Mazdak Abulnaga[2] ABULNAGA@MIT.EDU
[1] *Philips Research North America, Cambridge, MA, USA*
[2] *Computer Science and Artificial Intelligence Lab, MIT, Cambridge, MA, USA*

Editors: Under Review for MIDL 2019

Abstract

Infarcted brain tissue resulting from acute stroke readily shows up as hyperintense regions within diffusion-weighted magnetic resonance imaging (DWI). It has also been proposed that computed tomography perfusion (CTP) could alternatively be used to triage stroke patients, given improvements in speed and availability, as well as reduced cost. However, CTP has a lower signal to noise ratio compared to MR. In this work, we investigate whether a conditional mapping can be learned by a generative adversarial network to map CTP inputs to *generated* MR DWI that more clearly delineates hyperintense regions due to ischemic stroke. We detail the architectures of the generator and discriminator and describe the training process used to perform image-to-image translation from multi-modal CT perfusion maps to diffusion weighted MR outputs. We evaluate the results both qualitatively by visual comparison of generated MR to ground truth, as well as quantitatively by training fully convolutional neural networks that make use of generated MR data inputs to perform ischemic stroke lesion segmentation. We show that segmentation networks trained with generated CT-to-MR inputs are able to outperform networks that make use of only CT perfusion input.

Keywords: Conditional adversarial networks, Image-to-Image translation, Ischemic stroke lesion segmentation, CT perfusion

1. Introduction

Ischemic stroke is caused by partial or total restriction of blood supply to part of the brain. During an acute stroke, prolonged ischemia results in irreversible tissue death. Decisions about ischemic stroke therapy are highly time-sensitive and rely on distinguishing between the infarcted core tissue and hypoperfused lesions, i.e. the penumbra. As such, automated methods that can locate and segment ischemic stroke lesions can aid clinician decisions about acute stroke treatment. Computed tomography perfusion (CTP) has been used to triage stroke patients and has advantages in cost, speed and availability over diffusion-weighted magnetic resonance imaging (DWI). CTP provides detailed information about blood flow within the brain and can determine areas that are (in)adequately perfused with blood. However, CTP has a lower signal to noise ratio compared to DWI where infarcted core brain tissue readily shows up as hyperintense regions. In this work we train generative adversarial networks to learn a conditional mapping that maps CTP infarcted core regions to more clearly delineated hyperintense areas in *generated* MR scans. We utilize a dataset

of 94 paired CT and MR scans (Cereda et al., 2016) made available as part of the ISLES 2018 Ischemic Stroke Lesion Segmentation Challenge (Maier et al., 2017; Winzeck et al., 2018). Data was collected from 63 subjects from 4 hospital sites worldwide, whereby each acute stroke patient underwent back-to-back CTP and MRI DWI imaging within 3h of each other. Each CT scan was co-registered with its corresponding DWI. Perfusion maps were derived from each CT scan including cerebral blood flow (CBF), cerebral blood volume (CBV), mean transit time (MTT) and time to peak of the residue function (Tmax).

We employ the image-to-image translation framework introduced in (Isola et al., 2016) and modify it to accept multi-modal CT perfusion maps as input. After training conditional generative adversarial networks (CGANs) to reconstruct MR from CT perfusion maps, we train fully convolutional neural networks (FCN) to performs semantic segmentation of infarcted core tissue and compare whether performance can be improved by including the *generated* MR as an extra channel of information to the network. We show that FCNs trained with combined CTP and generated MR inputs are able to outperform networks trained without extra derived MR information on a range of metrics.

2. Related Work

Generative adversarial networks (Goodfellow et al., 2014) are increasingly being utilized within medical image synthesis and analysis (Wolterink et al., 2017; Nie et al., 2017; Lau et al., 2018; Wolterink et al., 2018). Typically, the high cost of acquiring ground truth labels means that many medical imaging datasets are either not large enough or exhibit large class imbalances between healthy and pathological cases. Generative models and GANs attempt to circumvent this problem by generating synthesized data that can be used to augment datasets during model training. One such approach is described in (Lau et al., 2018) for simulating myocardial scar tissue in late-gadolinium enhancement cardiovascular magnetic resonance (CMR) scans. The goal of ScarGAN is to perform dataset augmentation by beginning with a healthy CMR scan and adding realistic scar tissue to it. The training task is broken down into sub-tasks and avoids some of the difficulties associated with traditional GAN training – such as mode collapse. In particular, their approach consists of training two conditional adversarial networks. The first generates masks augmented with scar tissue and the second refines intensity values that have been set using a domain-specific heuristic. The authors of (Lau et al., 2018) make use of the pix2pix framework, as we do in this work, and they produce simulated scar tissue that experienced physicians mistake as real.

(Wolterink et al., 2017) investigates the problem of CT generation from MR for radiation therapy planning – a task that requires both MR and CT volumes. They synthesize CT from MR scans using CycleGAN (Zhu et al., 2017), which employs a forward and backward cycle-consistency loss. The CycleGAN framework allows processing of unpaired MR and CT slices and does not rely on having co-registered images. In the forward cycle a first network generates CT from MR inputs and a second network translates synthesized CT images back to MR. A discriminator is trained to differentiate between real and synthesized CT images. The loss is given by the difference between the original and reconstructed image. A backward cycle also contributes to the overall cycle loss whereby the synthesis is reversed from CT to MR. Interestingly, (Wolterink et al., 2017) shows that a model trained using unpaired MR and CT data was able to outperform models that used paired data.

3D Conditional GANs were also used in (Jin et al., 2018) to learn shape and appearance information about pulmonary nodules in CT volumes. The authors generated synthetic nodules by training a generator on pairs of nodule-masked 3D input patches together with their corresponding ground truth. A multi-mask L1 loss function was used that considered L1 loss only at the region of the nodule mask and surrounding border through the use of a dilation operation. Heavier weighting of the L1 loss around the border avoided discontinuity artifacts that can arise. As in our work, and others (Isola et al., 2016; Lau et al., 2018), the overall loss function consisted of a combination of the L1 loss together with the conditional GAN loss (see Equation 3). The generator network in (Jin et al., 2018) was employed to augment a CT training set with nodules close to lung borders for the purpose of improving CT lung segmentation.

In addition to the works mentioned above, adversarial networks have also been used for tasks such as improving lung segmentation in chest radiographs (Dai et al., 2018), correcting motion-related artifacts in cardiac magnetic resonance (Oksuz et al., 2018), as well as in registering medical images (Mahapatra et al., 2018). The motivation for our work is to explore whether hyperintensitities in MR scans can be emulated from CT perfusion inputs with conditional adversarial networks to improve the performance of ischemic stroke lesion segmentation.

3. Contributions

The contributions of this work are as follows:

1. Given a dataset of paired CTP and MR scans, we train a conditional generative adversarial network to reconstruct MR from CT perfusion maps.

2. We combine the generated MR scans with CTP data and train fully convolutional neural networks to perform semantic segmentation of ischemic core stroke lesions.

3. We show that the learned conditional mapping from CTP input results in more clearly delineated hyperintense regions in generated MR and this leads to improved segmentation of infarcted core areas.

4. CT-To-MR Conditional GAN Architecture

Generative adversarial networks (Goodfellow et al., 2014) work by training both a generator and a discriminator network. The generator network, $G(z)$, attempts to generate outputs that resemble images, y, from a distribution of training data, where z is a random noise vector, $G : z \rightarrow y$. The discriminator network, $D(\cdot)$, is given either a real input, $D(y)$, or a generated one, $D(G(z))$, and attempts to distinguish whether the input is real (images resulting from the true underlying data distribution) or fake (images created by the generator).

For the task of generating MR, conditioned on CT perfusion inputs, we adopt the Conditional GAN formulation, introduced in (Mirza and Osindero, 2014) and further described in the pix2pix framework (Isola et al., 2016). Conditional adversarial networks alter the generator such that it is conditioned on an original input image, $G : x, z \rightarrow y$, where x is

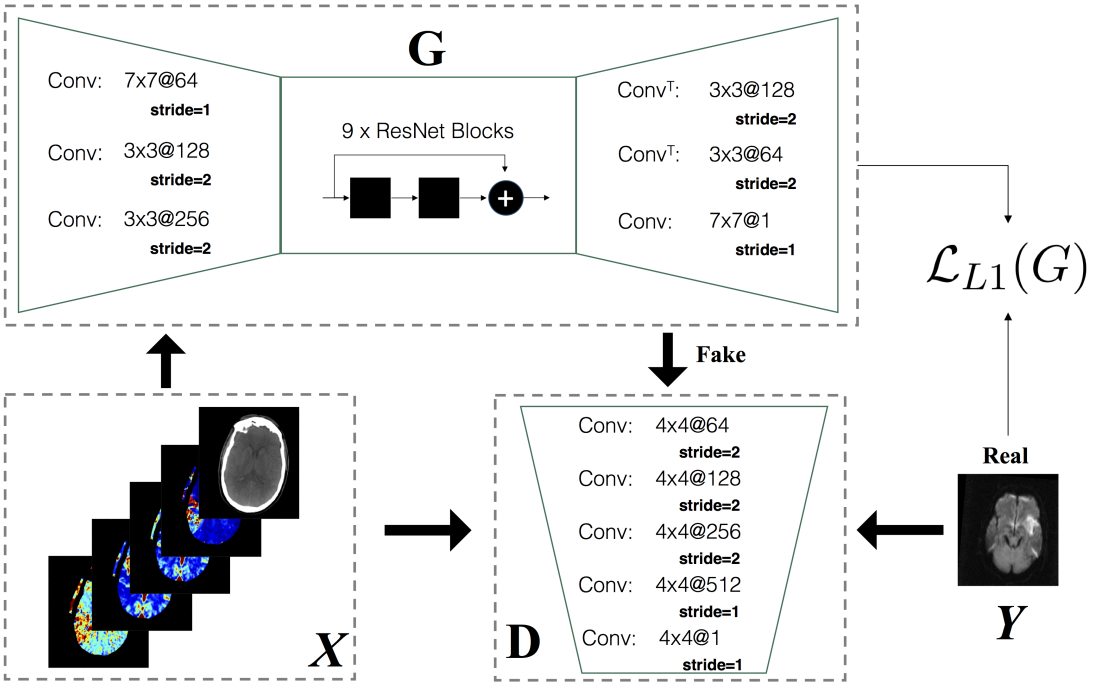

Figure 1: Overview of CT-To-MR Conditional GAN architecture

the input image and z is once again a random noise vector. The discriminator function is also updated to accept the conditional image as input, as well as the real, y, or fake input, $G(x, z)$, created by the generator. The full objective function for the conditional generator is given in Eq (1).

$$\mathcal{L}_{CGAN}(G, D) = \mathbb{E}_{x,y\sim p_{data(x,y)}}[\log(D(x, y)] + \mathbb{E}_{x\sim p_{data(x)},z\sim p_z(z)}[\log(1 - D(x, G(x, z)))] \quad (1)$$

As in (Isola et al., 2016), z is introduced into the generator network in the form of dropout at both train and test time. The final objective function for training the CT-To-MR translation model combines both the global $\mathcal{L}_{CGAN}(G, D)$ loss together with an additional L1 loss term, Eq (2), that captures the local per-pixel reconstruction error. The combined objective function is given in Eq (3), where λ is selected as a hyperparameter .

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y\sim p_{data(x,y)},z\sim p_z(z)}[\|y - G(x, z)\|_1] \quad (2)$$

$$G^* = \arg\min_G \max_D = \mathcal{L}_{CGAN}(G, D) + \lambda\mathcal{L}_{L1}(G) \quad (3)$$

4.1. Generator Architecture

A high level overview of the generator architecture, G, is shown in Fig 1. Generator inputs, x, are 5-channel 256×256 CT perfusion slices that contain the CT scan, stacked together

with the CBF, CBV, MTT, and Tmax perfusion maps. First, three initial convolution operations are applied. The size and number of convolutional kernels are shown in the figure as: $nxn@f$, where n is the kernel size and f the number of kernels. Downsampling is achieved via strided convolution. This is followed by 9 ResNet blocks, where a ResNet block is a residual block that consists of the following operations: Conv-InstanceNorm-ReLU-Dropout-Conv-InstanceNorm. Before each convolution operation (Conv) in the block, reflection padding with size 1 is added to each border of the input. The number of feature maps stays constant at 256, throughout the 9 ResNet blocks, as does their spatial resolution. Upsampling is achieved in the generator via fractionally strided convolutions (Conv^T), as shown in Fig 1. The generator output is a $1 \times 256 \times 256$ single channel derived MR slice.

4.2. Discriminator Architecture

As in (Isola et al., 2016), we utilize a convolutional PatchGAN discriminator that models high frequency image structure in local patches and penalizes incorrectness at the $N \times N$ patch-level. This is combined with the $L1$ loss term, $\mathcal{L}_{L1}$, that enforces low frequency correctness. A high level overview of the discriminator, D, is depicted in Fig 1.

The conditional discriminator accepts either real, $D(x, y)$, or generated, $D(x, G(x, z))$, MR slices, together with the original CT data and perfusion maps, $x \in \mathbb{R}^{5\times256\times256}$. CTP data and 'real' or 'fake' MR slices are stacked together in the channel dimension resulting in $6 \times 256 \times 256$ inputs being processed by the PatchGAN discriminator. All convolutions use a kernel size of 4×4, with downsampling once again being handled via strided convolution. Excluding the first and last convolution shown in Fig 1, each convolution is followed by an instance normalization operation (Ulyanov et al., 2016) and LeakyReLU activation with a negative slope coefficient of 0.2. The output of the network is a 30×30 map of discriminator activations, where each activation captures a 70×70 receptive field of overlapping patches from input channels. The final discriminator output is given by an average of this activation map.

4.3. CT-To-MR Conditional GAN Training

4.4. Data split

From the 94 available scans, a 5-fold split of the dataset was performed to create $5 \times$ 80%/20% splits. Splits were performed *by subject* to ensure that each fold of the data consisted of scans from unique subjects. To ensure the CGAN model generated "MR slices" only for scans it was not trained on, 5 CT-To-MR CGANs were created, where each model was trained on 80% of the data and produced derived MR slices for the remaining 20% of the data not seen during model training.

4.5. CGAN Implementation details

Training of the CT-To-MR CGAN took place by alternating one gradient descent step of the discriminator, followed by one gradient descent step for the generator. A batch size of 1 was used for training all networks. A dropout rate of 0.5 was applied within each ResNet block in the generator (see Fig 1). Within the final loss function, G^*, a value of $\lambda = 100$ was used to weight the combination of both L1 loss and that supplied from $\mathcal{L}_{CGAN}$. Adam

optimization used for training both the generator and discriminator with learning rates set to $2e^{-4}$ and momentum parameters $\beta_1 = 0.5$, $\beta_2 = 0.999$. Affine data transformations consisting of translation, rotation and scaling were used for augmentation. Each network was trained for a total of 200 epochs using PyTorch on a single Nvidia P100 GPU.

5. Ischemic Core Segmentation FCN Model

The final ischemic core segmentation network is based on the network architecture defined in (Abulnaga and Rubin, 2018). The model employs a fully convolutional neural network and utilizes pyramid pooling (Zhao et al., 2017) for capturing global and local context. The FCN component of the architecture relies on residual connections (He et al., 2016) to aid information flow during training and dilated convolution (Yu and Koltun, 2015) to cover larger receptive field sizes from the network inputs. Focal loss (Lin et al., 2018) is used as the loss function to attempt to learn the varying shapes of the lesion masks and effectively deal with the class imbalance between ischemic core and non-infarct areas.

The network is trained using transfer learning, beginning with weights that have been trained on natural images from the Pascal Visual Object Classes Challenge (Everingham et al., 2010). During training, data augmentations are created using standard affine transformations including rotation $[-10°, 10°]$, translation $[-10\%, 10\%]$ and scaling $[0.9, 1.1]$.

Two FCNs were trained to perform the final ischemic core segmentation. The network architecture and training details remained the same and the only difference between the networks were the inputs that were fed to them. Inputs to the first network (FCN) consisted of 5-channel 2D slices containing the CT image, together with its corresponding CBF, CBV, TTP and MTT perfusion maps. Inputs to the second network (FCN-GAN) were augmented with an extra channel of information that contained the *derived* MR slice – generated by the CT-to-MR GAN, conditioned on the 5-channel CTP input.

6. Results

We present qualitative results produced by the CT-To-MR CGAN and quantitative results of training FCN models with and without derived MR inputs. Fig 2 shows a subset of results created by the CT-To-MR CGAN model. Five MR slices are shown that were generated by conditioning on CTP input from a 20% test-set. The top row shows the ground truth MR slice and the bottom row shows the corresponding slice that was created by the CT-To-MR generator. While only a qualitative assessment can be made, it can be seen that hyperintense regions within the ground truth DWI image are approximately replicated by the conditional generator.

For quantitative analysis, Table 1 compares the 5-fold cross validation results of two ischemic core segmentation models. The first baseline model, FCN, shows the results for a FCN trained only on the 5 original CTP input modalities. FCN-CGAN refers to the same network architecture, but trained by including the generated MR slice as one of the input modalities. Dice coefficient, Hausdorff distance, average distance, precision, recall and absolute volume difference metrics are provided. It can be seen that including derived MR slices from the CT-To-MR CGAN model leads to improvements in all evaluation metrics.

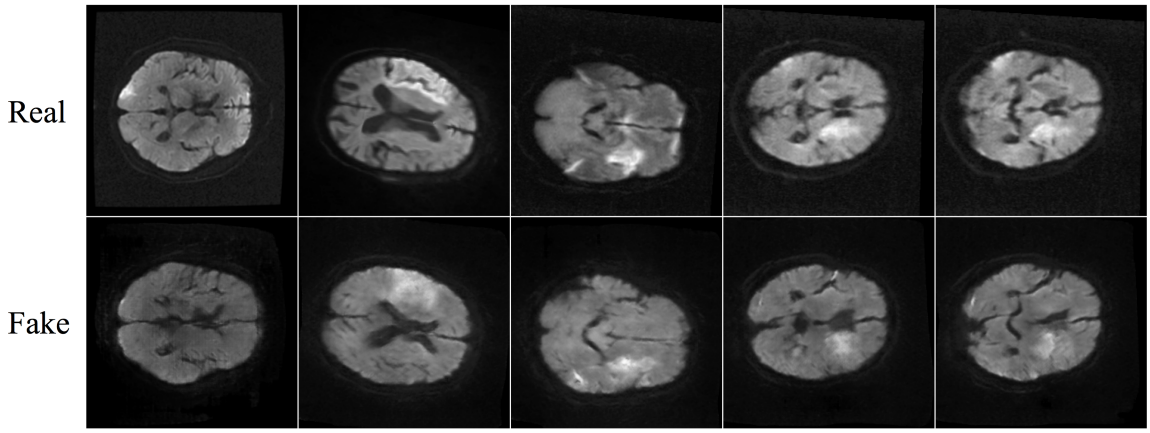

Figure 2: Examples of real and derived MR slices. The top row shows real images and the bottom row shows slices generated from the CT-To-MR Conditional GAN model

Table 1: Ischemic stroke lesion segmentation results using a FCN baseline model compared to FCN-CGAN that incorporates CT-To-MR input information learned via a conditional adversarial network. Arrows in the columns indicate whether lower or higher values are better. For all evaluation metrics, the FCN-CGAN improves upon the FCN baseline.

Metric	**FCN**	**FCN-CGAN**
↑ Dice	0.53 ± 0.25	**0.54** ± 0.23
↓ Hausdorff Distance	27.94 ± 20.27	**27.88** ± 21.00
↓ Average Distance	4.73 ± 9.88	**4.37** ± 9.35
↑ Precision	0.56 ± 0.27	**0.56** ± 0.25
↑ Recall	0.62 ± 0.27	**0.63** ± 0.25
↓ Absolute Volume Difference	11.53 ± 13.11	**10.20** ± 13.10

Figure 3 shows a sample of segmentation results produced by the FCN and FCN-CGAN approaches. The bottom row shows ischemic core segmentation masks produced by the FCN baseline model in green. Ground truth segmentations are shown in red. The top row compares segmentation results from the FCN-CGAN model in blue to ground truth, once again shown in red. The results show that, in general, the FCN-CGAN model results in predictions that cover more of the ischemic core region compared to predictions from the baseline FCN. It can also be seen in some cases (e.g. the 4th case from the right in Figure 3) that the FCN-CGAN model was able to correct for islands of false positives that were predicted by the FCN baseline. In addition to the mask predictions, also displayed in Figure 3 are the overall dice coefficient values for the corresponding scan from which the slice was taken. It can be seen in some cases (e.g. the 2nd case from the right: FCN: 0.087

vs FCN-CGAN: 0.390) that substantial improvements in dice values can be observed with the FCN-CGAN model.

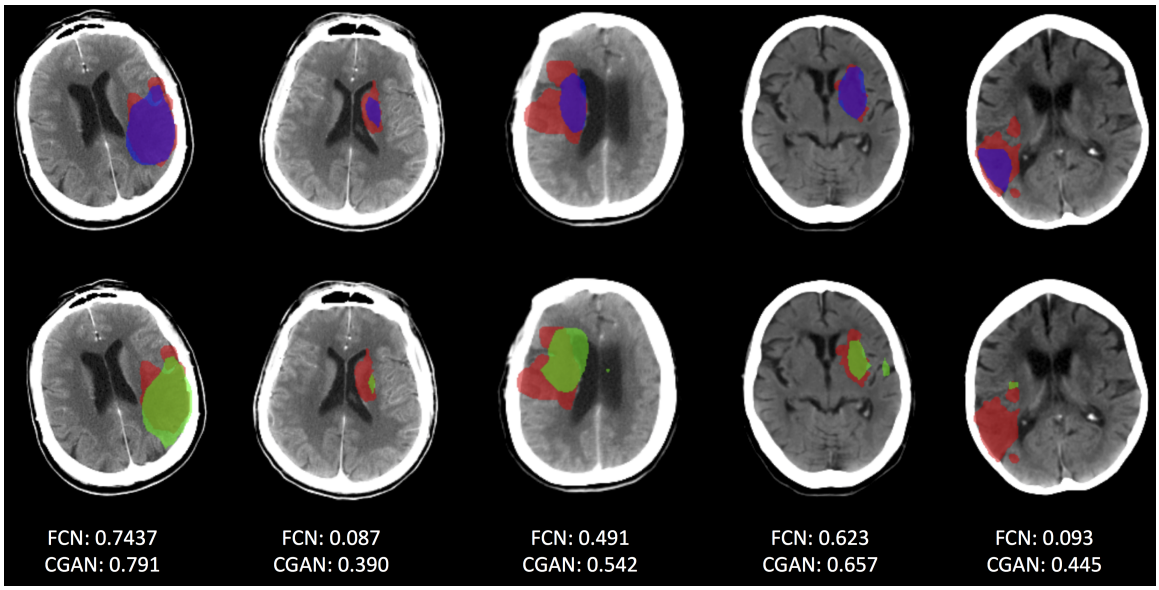

Figure 3: Comparison of segmentation results. Ground truth is shown in red. The bottom row shows results produced by the FCN baseline in green and the top row shows results produced by the FCN-CGAN in blue. Also shown are the dice coefficient values for each approach. Note: the dice score is per scan from which each slice was taken from.

7. Conclusions & Future Work

We have presented an approach that utilized conditional generative adversarial networks to improve the performance of fully convolutional stroke lesion segmentation networks. Diffusion-weighted magnetic resonance imaging is considered most accurate for early detection of acute stroke (Biesbroek et al., 2013; Gillebert et al., 2014), as infarcted brain tissue can be recognized as hyperintense regions of the DWI map compared to surrounding brain tissue. Hence, the motivation for this work was to emulate a DWI map, conditioned on a given CT perfusion input. By generating corresponding DWI maps and including them as inputs to a FCN, it was hoped to ease the learning process. Our qualitative results show that CGANs do a reasonable job of mapping infarcted core regions to hyperintensities that align with corresponding areas in ground truth MR. We also show quantitative improvements in segmentation performance, as measured by dice coefficient, Hausdorff distance, average distance, precision, recall and absolute volume difference, when including generated MR images as input to the training of segmentation networks.

In the present work, we incorporated CGAN models that act as a 1:1 mapping from CT perfusion input to corresponding generated MR. However, related works (Lau et al., 2018;

Jin et al., 2018) have utilized generative models to synthesize further training examples to better capture the wide range of shapes, locations and appearances of pathological tissue and disease that show up in medical imaging modalities. Given the small dataset used in this work, consisting of 94 scans in total, it is possible that further improvements in segmentation performance could be possible by employing CGANs rather than as 1:1 mappings from CTP to MR, but to synthesize novel examples of ischemic core lesions in generated scans to artificially bolster the training set size.

References

S Mazdak Abulnaga and Jonathan Rubin. Ischemic stroke lesion segmentation in ct perfusion scans using pyramid pooling and focal loss. *arXiv preprint arXiv:1811.01085*, 2018.

JM Biesbroek, JM Niesten, JW Dankbaar, GJ Biessels, BK Velthuis, JB Reitsma, and IC Van Der Schaaf. Diagnostic accuracy of ct perfusion imaging for detecting acute ischemic stroke: a systematic review and meta-analysis. *Cerebrovascular diseases*, 35(6): 493–501, 2013.

Carlo W Cereda, Søren Christensen, Bruce CV Campbell, Nishant K Mishra, Michael Mlynash, Christopher Levi, Matus Straka, Max Wintermark, Roland Bammer, Gregory W Albers, et al. A benchmarking tool to evaluate computer tomography perfusion infarct core predictions against a dwi standard. *Journal of Cerebral Blood Flow & Metabolism*, 36(10):1780–1789, 2016.

Wei Dai, Nanqing Dong, Zeya Wang, Xiaodan Liang, Hao Zhang, and Eric P Xing. Scan: Structure correcting adversarial network for organ segmentation in chest x-rays. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 263–273. Springer, 2018.

Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

Cline R. Gillebert, Glyn W. Humphreys, and Dante Mantini. Automated delineation of stroke lesions using brain ct images. *NeuroImage: Clinical*, 4:540 – 548, 2014.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014. URL http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:1611.07004*, 2016.

Dakai Jin, Ziyue Xu, Youbao Tang, Adam P Harrison, and Daniel J Mollura. Ct-realistic lung nodule simulation from 3d conditional generative adversarial networks for robust lung segmentation. *arXiv preprint arXiv:1806.04051*, 2018.

Felix Lau, Tom Hendriks, Jesse Lieman-Sifry, Sean Sall, and Dan Golden. Scargan: chained generative adversarial networks to simulate pathological tissue on cardiovascular mr scans. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 343–350. Springer, 2018.

Tsung-Yi Lin, Priyal Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *IEEE transactions on pattern analysis and machine intelligence*, 2018.

Dwarikanath Mahapatra, Bhavna Antony, Suman Sedai, and Rahil Garnavi. Deformable medical image registration using generative adversarial networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1449–1453. IEEE, 2018.

Oskar Maier, Bjoern H Menze, Janina von der Gablentz, Levin Häni, Mattias P Heinrich, Matthias Liebrand, Stefan Winzeck, Abdul Basit, Paul Bentley, Liang Chen, et al. Isles 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral mri. *Medical image analysis*, 35:250–269, 2017.

Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

Dong Nie, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with context-aware generative adversarial networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 417–425. Springer, 2017.

Ilkay Oksuz, James Clough, Aurelien Bustin, Gastao Cruz, Claudia Prieto, Rene Botnar, Daniel Rueckert, Julia A Schnabel, and Andrew P King. Cardiac mr motion artefact correction from k-space using deep learning-based reconstruction. In *International Workshop on Machine Learning for Medical Image Reconstruction*, pages 21–29. Springer, 2018.

D Ulyanov, A Vedaldi, and VS Lempitsky. Instance normalization: the missing ingredient for fast stylization. corr abs/1607.0 (2016), 2016.

Stefan Winzeck, Arsany Hakim, Richard McKinley, José AADSR Pinto, Victor Alves, Carlos Silva, Maxim Pisov, Egor Krivov, Mikhail Belyaev, Miguel Monteiro, et al. Isles 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral mri. *Frontiers in neurology*, 9, 2018.

Jelmer M Wolterink, Anna M Dinkla, Mark HF Savenije, Peter R Seevinck, Cornelis AT van den Berg, and Ivana Išgum. Deep mr to ct synthesis using unpaired data. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 14–23. Springer, 2017.

Jelmer M Wolterink, Tim Leiner, and Ivana Isgum. Blood vessel geometry synthesis using generative adversarial networks. *arXiv preprint arXiv:1804.04381*, 2018.

Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2881–2890, 2017.

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