CSC-GAN: Cycle and semantic consistency for dataset augmentation

Renato B. Arantes, George Vogiatzis, and Diego R. Faria

Aston University, Aston St, Birmingham B4 7ET
{180178991,g.vogiatzis,d.faria}@aston.ac.uk
https://www2.aston.ac.uk/

Abstract. Image-to-image translation is a computer vision problem where a task learns a mapping from a source domain A to a target domain B using a training set. However, this translation is not always accurate, and during the translation process, relevant semantic information can deteriorate. To handle this problem, we propose a new cycle-consistent, adversarially trained image-to-image translation with a loss function that is constrained by semantic segmentation. This formulation encourages the model to preserve semantic information during the translation process. For this purpose, our loss function evaluates the accuracy of the synthetically generated image against a semantic segmentation model, previously trained. Reported results show that our proposed method can significantly increase the level of details in the synthetic images. We further demonstrate our method's effectiveness by applying it as a dataset augmentation technique, for a minimal dataset, showing that it can improve the semantic segmentation accuracy.

Keywords: GAN · Dataset augmentation · Semantic segmentation.

1 Introduction

The sparsity of training data can hinder the performance of supervised machine learning algorithms which often require large amounts of data to train and avoid overfitting. Typical deep neural networks have hundreds of millions of parameters to learn, which requires many passes over the training data. On small datasets, running a large number of iterations can result in overfitting, which is usually remedied by one or more of the following: acquiring more data, applying regularisation and performing data augmentation. The latter approach is mostly limited to simple randomised manipulation of an existing dataset (*e.g.* affine warping, rotation or other small perturbations) [27]. In this paper, we leverage the recent success of adversarial image-to-image translation to propose a much more sophisticated approach for data augmentation that applies to semantic segmentation tasks. Under our scheme, an arbitrary number of new images is generated by 'translating' each ground truth label image in our training dataset, and the

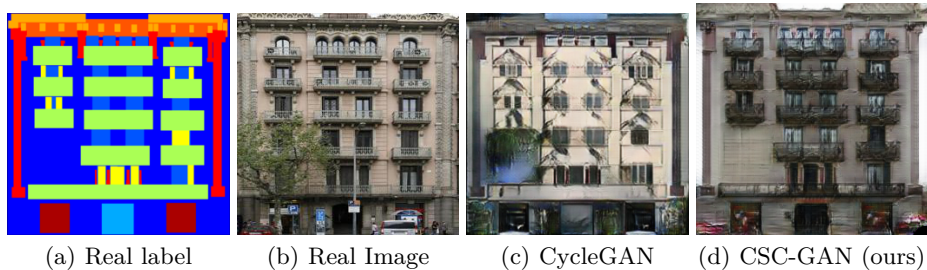(a) Real label   (b) Real Image   (c) CycleGAN   (d) CSC-GAN (ours)

Fig. 1: **Translation of facade label to a facade image.** The task is to translate from label to image. (a) and (b) are the real label and the real image, (b) is CycleGAN generated, that incorrectly translates the facade behind the tree and the balcony railings. (c) Using our CSC-GAN, that implements semantic consistency in the translation process, the result is a more realistic image.

resulting image/labelling pairs are used to augment that dataset, leading to improved accuracy. This is achieved by a new cycle-consistent, adversarially trained image-to-image translation model (CSC-GAN) that strengthens local and global structural consistency through pixel cycle-consistency and a semantic loss.

Synthesising realistic and high-resolution images is one of the most challenging tasks in computer vision and machine learning investigation areas [16]. The recent CycleGAN [30] introduced the concept of Cycle-consistency in the context of an image-to-image translation GAN model producing impressive results. In particular, it appears to be able to *translate* semantic labelling to the corresponding image, which seems perfect for data augmentation in semantic segmentation tasks. Unfortunately, CycleGAN fails to capture the relevant semantic constraints accurately (see Fig. 1(c)) and therefore, cannot be used for data augmentation.

The question we pose in this paper is: How can one maintain the semantic correspondence during the image-to-image translation process? Our answer is an additional objective that enforces semantic consistency while the cycle-consistency loss of CycleGAN helps the cross-domain transformation to retain local structural information. Our CSC-GAN also proved to be simpler than other methods [13], while generating realistic images with a significant increase on its level of details to such an extent that they can improve the performance of a state of the art semantic segmentation models like DeeplabV3 [5] and Fully-Convolutional Network (FCN) [20]. This is a significant achievement if we consider that (a) there is no new data or prior information that was fed to the model and (b) DeeplabV3 and FCN are already extremely good at extracting generalisable knowledge from any given dataset. Figure 1(d) is an example of our method's potential. We apply our CSC-GAN to an extremely sparse building facade dataset [24] that consists of only 606 images and labels. Our experiments

verify that image-to-image translation methods strengthened by semantic consistency can be used to improve pixel-level semantic segmentation models.

2   Related work

Generative adversarial networks (GANs), introduced by Goodfellow *et al.* [10], employ the concept of adversarial training, where two models are simultaneously trained: a generative model G that captures the data distribution, and a discriminator model D that estimates the probability that a sample came from the training data rather than G. The generator learns the data distribution p_g, over the data x, by creating a mapping function from a prior noise distribution $p_z(z)$ to data space as $G(z; \theta_g)$. Further, the discriminator, $D(x; \theta_d)$, outputs a single scalar representing the probability that x came from the training data instead of p_g. It has shown extraordinary results in many computer vision tasks such as style transfer [17, 28], image generation [4, 8, 21], image translation [14, 30, 7], multi-domain image translation [7] and super-resolution [19].

The GAN framework was first introduced to generate visually realistic images and, since then, many applications have been proposed, including data augmentation [27], a technique widely used to increase the volume of data available for training.

One of the areas where GANs have been employed for data augmentation is Medical Imaging. Yi *et al.* in [29], surveyed 150 published articles in the medical image synthesis area and found that GANs are employed for image reconstruction, segmentation, detection, classification and cross-modality synthesis. The main reason for this widespread use seems to be the relative sparsity of labelled datasets in the highly specialized medical image domains.

The same conclusion is reached in Bowles *et al.* [3], where the use of GANs for augmenting CT scan data is investigated. They use a Progressive Growing of GANs (PGGAN) network [16] to generate the synthetic data in the joint image-label space. Their results show that GAN augmentation works better, the more sparse the dataset. Unfortunately the PGGAN framework is more suitable for spatially registered datasets (e.g. faces, or medical imaging). Using an extensive image dataset, Sandfort *et al.* [25] trained a CycleGAN model [30] to transform contrast Computed Tomography (CT) images into non-contrast images, for data augmentation in CT segmentation tasks. According to the authors, the publicly available datasets consists universally of contrast-enhanced CT images, while real-world data contains a certain percentage of non-contrast CT images. This domain shift affects the performance of real-world applications negatively. Using CycleGAN to alleviate this issue, the authors report significant improvement in CT segmentation tasks performance.

When the available data is not uniformly distributed between the distinct classes, the accuracy of an image classification model can degenerate. In [22], Mariani *et al.* propose a balancing generative adversarial network (BAGAN), an augmentation method that can generate new minority-class images and then restore the dataset balance. Likewise, in [1], Antoniou *et al.* proposed a Data

Augmentation Generative Adversarial Network (DAGAN) architecture based on a conditional GAN, conditioned on images from a given class c, where the generator network also uses an encoder to project the condition down to a lower-dimensional manifold. Their adversarial training leads their system to generate new images from a given sample, one that appears to be within the same class but look different enough to be a diverse sample.

An impressive dataset augmentation method using GAN for semantic segmentation is introduced by Richter *et al.* in [2]. In their paper, they present an approach for creating a semantic label for images extracted from modern computer games. Using game engines to generate endless quantities of labelled data is a longstanding theme in Computer Vision research. Their experiments show that using the acquired data, to supplement real-world images, significantly increases accuracy, showing that a network trained on unrealistic data can generalise very well to existing datasets.

CyCADA, proposed by Hoffman *et al.* in [13], also explores ways of enforcing semantic consistency on image synthesis. This adversarial unsupervised adaptation algorithm aims to learn a model that correctly predicts a label for a target data. Their input is the source data X_S, the source label Y_S and the target data X_T. Their aim is to learn a model that can correctly predict the label Y_T for the target data X_T. In contrast, our work doesn't seek to predict but to increase the translated image quality with the help of a pre-defined classifier. We are provided with the source data X_S and the source label Y_S. Our purpose is to learn a model that do the translation $Y_S \rightarrow X_S$ more accurately.

3   Proposed method

In this section, we present our *Cycle and Semantic Consistent GAN* (CSC-GAN) framework. We consider the problem of image-to-image translation with cycle and semantic consistency, where we are provided with the source data X and source labels Y. The aim is to learn a stochastic model f that translates a labelling into the corresponding image (*i.e.* $Y \rightarrow X$) in such a way that the resultant images are so realistic they can improve the results of a deep semantic segmentation model [5, 20], when used as a dataset augmentation technique. To do so, we extend the CycleGAN framework by adding a new loss function $\mathcal{L}_{sem}$ that evaluates how accurate is the synthetic generated image against a previously trained semantic segmentation model g. To establish the background, we first review in the following sections, the GAN/Cycle-GAN models on which our method is based.

3.1   CycleGAN loss functions

The CycleGAN full objective is composed of an adversarial loss and a cycle consistency loss, as follows.

Adversarial loss. The adversarial loss is given by

$$\mathcal{L}_{GAN}(G, D_Y) = \mathbb{E}_{y \sim p_{data}(y)}[\log D_Y(y)] \\ + \mathbb{E}_{x \sim p_{data}(x)}[\log(1 - D_Y(G(x)))], \tag{1}$$

where G tries to generate images that look similar to images from domain Y and D_Y aims to distinguish between the translated samples $G(x)$ and real samples y. G aims to minimise this objective against an adversary D that tries to maximise it.

Cycle consistency loss. Adversarial losses individually cannot guarantee that the learned function can map a single input x to a desired output y. To additionally decrease the space of possible mapping functions, CycleGAN authors argue that the learned mapping functions should be cycle-consistent, intending to encourage the source content to be preserved during the conversion process. For each image x from domain X, a different map F should be able to bring x back to the original image, *i.e.*, $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$. This is called the *forward cycle consistency*. Similarly, for each image y from domain Y, G and F should also satisfy *backward cycle consistency*: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$. This behaviour is encouraged using a cycle consistency loss:

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)} \parallel F(G(x)) - x \parallel_1 \\ + \mathbb{E}_{y \sim p_{data}(y)} \parallel G(F(y)) - y \parallel_1 . \tag{2}$$

Full objective. The CycleGAN full objective is

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y) \\ + \mathcal{L}_{GAN}(F, D_X) \\ + \lambda \mathcal{L}_{cyc}(G, F), \tag{3}$$

where λ controls the relative importance of the two objectives.

3.2   Semantic consistency objective

As we have access to the source labelled data, we aim to encourage high semantic consistency after image translation explicitly. For this purpose, we pre-train a semantic segmentation model g, on the same *training* set used to train our CSC-GAN model, and use this model g to evaluate the synthetic images during the CSC-GAN training. By fixing the model g weights during the CSC-GAN training, we guarantee that a good segmentation result, obtained from a synthetic image, is due to an improvement in the synthetic image quality, as the model g was trained on real images. Using the segmenter model g we propose our semantic consistency objective as
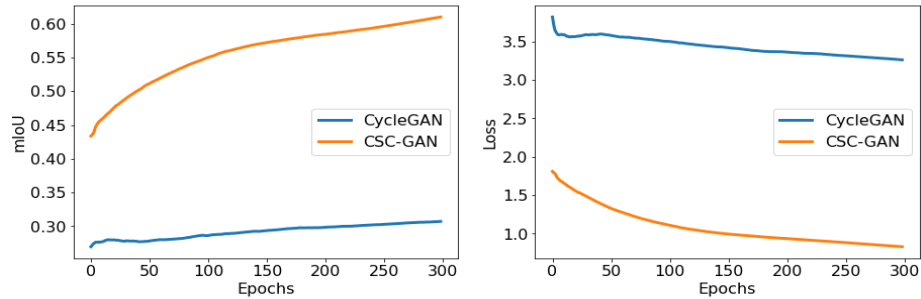
Fig. 2: Impact of the proposed semantic consistency loss over the synthetic images, during model training.

$$\mathcal{L}_{sem}(g, F) = \mathbb{E}_{y \sim p_{data}(y)} \mathcal{L}_{cs}[g(F(y)), y], \tag{4}$$

where $\mathcal{L}_{cs}[.,.]$ is the cross-entropy loss [13] comparing two segmentation masks $g(F(y))$ and y. This equation means that given a label $y \in Y$, the model g predicts the labels for $F(y)$, *i.e.*, the synthetic image generated using the real label y. The loss $\mathcal{L}_{sem}$ evaluates this prediction, and its result is added to the CycleGAN objective function, intending to help improve the overall synthetic image quality. It is interesting to note the superficial similarity between the semantic consistency loss $\mathcal{L}_{sem}$ and the third term of the cycle consistency loss (Eq. 3). Both G and g map from images to labels but (a) g is producing 1-hot encoding per label while G produces an RGB image as in [30], hence the different choice of image distance metric and (b) g is a pre-trained network (we use the DeepLabV3 architecture) while G is a network we train adversarially. We experimented with removing the trainable G network but were unable to achieve a converged solution, possibly because G introduces a convex relaxation to the optimisation problem. The full objective is then

$$\begin{aligned}
\mathcal{L}(G, F, D_X, D_Y) &= \mathcal{L}_{GAN}(G, D_Y) \\
&+ \mathcal{L}_{GAN}(F, D_X) \\
&+ \lambda \mathcal{L}_{cyc}(G, F) \\
&+ \mu \mathcal{L}_{sem}(g, F).
\end{aligned} \tag{5}$$

where λ and μ are relative important weights. To show the impact of the new semantic consistency loss, Figure 2 presents the evaluation comparison of the synthetic images, during model training, with and without our proposed loss. It shows a dramatic improvement in the mIoU score and the cross-entropy loss when the objective described in the equation 5 is used, instead of the regular CycleGAN loss defined in equation 3.
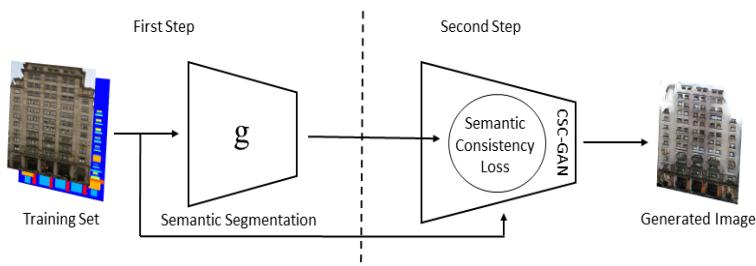
Fig. 3: **CSC-GAN two steps.** First we train a semantic segmentation model *g* on the *training* set, then we use it to evaluate the synthetic images quality during our model training.

4   Experimental results

In this section we present the attained results of our approach CSC-GAN compared against the Facade dataset [24] augmented with the regular CycleGAN model [30],SPADE model [23], which is a state-of-art label to image translation and two style transfer models [9, 15], being that [15] was designed explicitly for dataset augmentation. The datasets used on the experiments are listed in Table 1.

In order to compare these augmented datasets, for each one we trained two semantic segmentation models, DeeplabV3 [5] and Fully-Convolutional Network (FCN) [20], both with a ResNet101 [11] backbone. Each model is trained for 50 epochs with a learning rate set to 0.0002 and the Adam optimiser [18]. The performance of each model is reported using the mean intersection-over-union (mIoU) score [26] evaluated over the *test* set.

The facade dataset [24] has been split into a *training* X_S, with 80% or 484 images and a *test* X_T set, with 20% or 122 images. As shown in Table 1, there are six variations of the *training* set, but the *test* set remains the same across all experiments.

The augmentation is done as follows: For each label $y \in X_S$ a synthetic image $x = F(y)$ is generated and the pair (y, x) is added to the dataset. By the end a new *training* set $X_S^{'}$ is created with size twice as X_S. This is done because one synthetic image per each label $y \in X_S$ is added to the new *training* set $X_S^{'}$.

The CSC-GAN model is trained in two steps: First, we train the semantic segmentation model *g* on the regular facade dataset, which is used as a labeller during the CSC-GAN training, then we train the CSC-GAN model, as described above. Figure 3 illustrates these two stages.

Table 2 presents the results. Each experiment was executed 5 times and the results reported are the mean and the standard deviation for these 5 executions. The presented results show that the dataset augmented with images from our CSC-GAN model can outperform the regular facade dataset by approximately 4% when the DeeplabV3 [6] model is trained on it and by roughly 3% with the FCN [20] model, as shown in the last row on Table 2. SPADE also provided

Training set name	Size	Description
facade	484	regular training set without any augmentation. [24]
facade+CycleGAN	968	regular plus images from regular CycleGAN model [30].
facade+SPADE	968	regular plus images from SPADE model [23].
facade+Styleaug	968	regular plus images from style augmentation model [15].
facade+Arbitrary	968	regular plus images from arbitrary artistic stylization model [9].
facade+CSC-GAN	968	regular plus images from our CSC-GAN model.

Table 1: **Datasets.** Lists of all datasets used in the semantic consistency experiment.
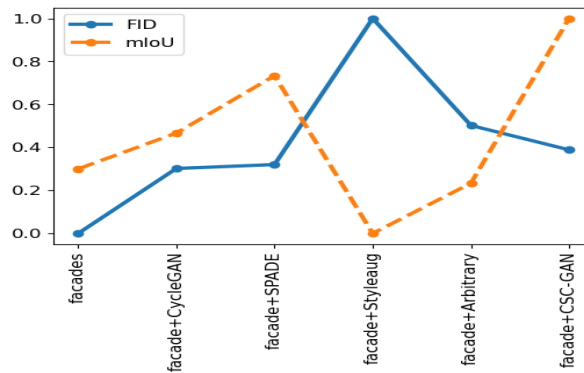

Fig. 4: **FID x MioU.** The graph suggests an inverse relationship between the two metrics. Data were normalised to fit in the same plot. The *Frechet Inception Distance* (FID) is a metric that compares statistics of real and generated images and summarises how similar the two groups of images are. A lower FID score indicates that the two groups of images are more similar.

promising results, but it was also outperformed by our method. The datasets augmented by [15] and [23] got a decrease in performance concerning the regular Facade dataset, showing that style transfer is not a good technique for dataset augmentation on a per-pixel classification scenario as semantic segmentation. Also, the FID [12] metric in the first column suggests an inverse relation with the mIoU metric, as shown in Figure 4. This inverse relationship is because the synthetic images must be within the original distribution, but it also has to look different enough to be regarded by the semantic segmentation model as a diverse sample. A low FID means that the synthetic images are too close to the original dataset distribution; therefore, they cannot be viewed as new samples. Conversely, a high FID means that the synthetic data are too diverse from the original data; hence they cannot be considered as samples from the same distribution.

To better understand the improvement brought by our CSC-GAN, in Figure 5 we compare the correctly predicted pixels of the models trained with CycleGAN

		DeeplabV3 [6]		FCN [20]	
Dataset name	FID	mIoU	Accu.	mIou	Accu.
facades [24]		0.552 (0.016)	0.706 (0.013)	0.558 (0.003)	0.711 (0.002)
facade+CycleGAN [30]	**22.742**	0.557 (0.008)	0.710 (0.007)	0.562 (0.002)	0.714 (0.002)
facade+SPADE [23]	24.107	0.565 (0.009)	0.716 (0.008)	0.572 (0.004)	0.723 (0.003)
facade+Styleaug [15]	75.307	0.543 (0.005)	0.698 (0.005)	0.561 (0.005)	0.714 (0.004)
facade+Arbitrary [9]	37.804	0.550 (0.005)	0.703 (0.004)	0.562 (0.003)	0.714 (0.002)
facade+CSC-GAN (ours)	29.268	**0.573 (0.005)**	**0.723 (0.004)**	**0.574 (0.003)**	**0.724 (0.002)**

Table 2: **Experimental results**. Accu. is the pixel accuracy; the FID [12] metric is calculated against the dataset without augmentation. Each experiment was executed 5 times, the mIoU and the pixel accuracy results reported are the mean and the standard deviation for these 5 executions.
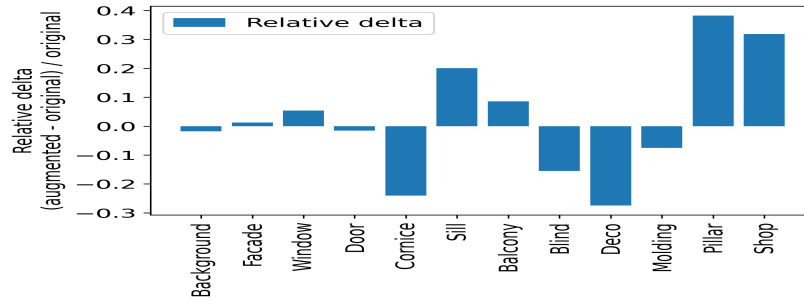
and CSC-GAN augmented datasets with the model trained with the regular facade dataset. The comparisons are made over the test set segmentation results.

Figure 5(a) presents the comparison of the regular facade with CycleGAN; the graph shows that for six out of twelve labels, there is a drop in the correctly predicted pixels, being *cornice* and *deco* labels the ones that have the most significant difference. Conversely, in Figure 5(b), except for *background*, *window* and *shop* the model trained with our CSC-GAN augmented dataset attained a substantial improvement in the correctly predicted pixels. Special attention must be taken to the *decoration* and *balcony* labels, two very challenging categories where our method presented a significant improvement over the facade dataset. The reason is the increase in level-of-detail obtained by the CSC-GAN model to the generated images, as we can see in Figure 6.

In Figure 6, column (c), note that our model is capable of learning the appearance of balcony railings and in the second row the extended model begins to show the pillar as a 3D structure, instead of just a flat shape as shown in column (b). The third row, column (c), shows that the extended model can also learn the decoration, which the regular CycleGAN model represents as a flat shape.

5 Conclusion

In this paper, we propose a cycle consistent image generation framework that combines with semantic constraints to deliver increased level-of-detail in the generated images. This improvement enables the use of the method for data augmentation in the domain of semantic segmentation. As future work we intend to apply our model to photorealistic images of general scenes. We are also planning to investigate more challenging datasets, such as geospatial images. Another promising direction to explore is the domain-transfer effect that arises from cross-training the semantic segmentation model on different datasets.

(a) CycleGAN augmented
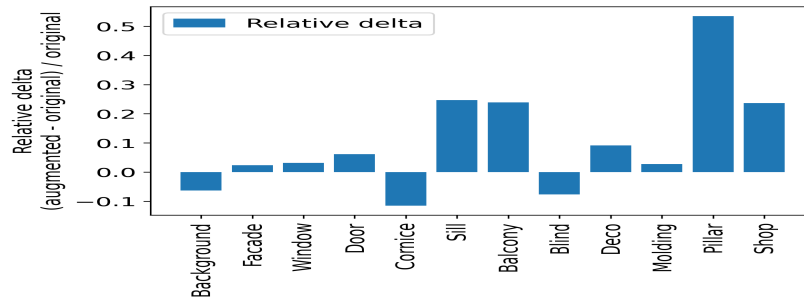
(b) CSC-GAN augmented

Fig. 5: Figures (a) and (b) compare the correctly predicted segmentation pixels between the regular facade and the augmented datasets. Figure (b) shows that for the majority classes, our CSC-GAN model can produce more accurate images for data augmentation.

References

1. Antoniou, A., Storkey, A., Edwards, H.: Data augmentation generative adversarial networks. arXiv preprint arXiv:1711.04340 (2017)
2. Antoniou, A., Storkey, A., Edwards, H.: Data augmentation generative adverdarial networks. stat **1050**, 8 (2018)
3. Bowles, C., Chen, L., Guerrero, R., Bentley, P., Gunn, R.N., Hammers, A., Dickie, D.A., del C. Valdés Hernández, M., Wardlaw, J.M., Rueckert, D.: GAN augmentation: Augmenting training data using generative adversarial networks (2018)
4. Brock, A., Donahue, J., Simonyan, K.: Large scale GAN training for high fidelity natural image synthesis (2018)
5. Chen, L., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation (2017)
6. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: ECCV (2018)
7. Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018)

(a)   (b)   (c)

Fig. 6: **Comparing images.** Real image (a), regular CycleGAN model (b) and CSC-GAN model (c). The labels were omitted.

8. Donahue, J., Simonyan, K.: Large scale adversarial representation learning (2019)
9. Ghiasi, G., Lee, H., Kudlur, M., Dumoulin, V., Shlens, J.: Exploring the structure of a real-time, arbitrary neural artistic stylization network. arXiv preprint arXiv:1705.06830 (2017)
10. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 27, pp. 2672–2680. Curran Associates, Inc. (2014)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016. pp. 770–778 (2016)

12. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: Advances in neural information processing systems. pp. 6626–6637 (2017)
13. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A., Darrell, T.: Cycada: Cycle-consistent adversarial domain adaptation. In: International conference on machine learning. pp. 1989–1998 (2018)
14. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 5967–5976 (2017)
15. Jackson, P.T., Abarghouei, A.A., Bonner, S., Breckon, T.P., Obara, B.: Style augmentation: data augmentation via style randomization. In: CVPR Workshops. pp. 83–92 (2019)
16. 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)
17. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. (2018)
18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
19. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W.: Photo-realistic single image super-resolution using a generative adversarial network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 105–114 (2017)
20. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
21. Lučić, M., Ritter, M., Tschannen, M., Zhai, X., Bachem, O.F., Gelly, S.: High-fidelity image generation with fewer labels. In: International Conference on Machine Learning (2019)
22. Mariani, G., Scheidegger, F., Istrate, R., Bekas, C., Malossi, A.C.I.: BAGAN: data augmentation with balancing GAN. CoRR **abs/1803.09655** (2018)
23. Park, T., Liu, M.Y., Wang, T.C., Zhu, J.Y.: Semantic image synthesis with spatially-adaptive normalization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2019)
24. Radim Tyleček, R.Š.: Spatial pattern templates for recognition of objects with regular structure. In: Proc. GCPR. Saarbrucken, Germany (2013)
25. Sandfort, V., Yan, K., Pickhardt, P.J., Summers, R.M.: Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks. Scientific reports **9**(1), 1–9 (2019)
26. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(4), 640–651 (Apr 2017)
27. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. Journal of Big Data **6**(1), 60 (2019)
28. Wang, X., Gupta, A.: Generative image modeling using style and structure adversarial networks. CoRR **abs/1603.05631** (2016)
29. Yi, X., Walia, E., Babyn, P.: Generative adversarial network in medical imaging: A review. Medical Image Analysis **58**, 101552 (2019)
30. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Computer Vision (ICCV), 2017 IEEE International Conference on (2017)