WeaGAN:Generative Adversarial Network for Weather Translation of Image among Multi-domain

Yating Lin, Yidong Li
School of Computer and Information Technology
Beijing Jiaotong University, Beijing 100044, China
{17120383, ydli}@bjtu.edu.cn

Haidong Cui*, Zheng Feng
Taishan Information Technology Co., Ltd.
No. 57 Mingtang Road, Tai'an City, Shandong, China
{573470610@qq.com, fengzheng@ts-it.cn}

***Abstract*—Weather translation of image refers to the task of changing the weather of an input image to desired weather while preserving the structure of the image's content, which belongs to a task of image-to-image translation. Recent works have made great process in image-to-image translation between two domains and some works have even achieved multi-domain translation within a single model. However, existing works have limited robustness in handling weather translation among multi-domain, since bad weather produces a loud noise and it is challenging to process scene images without fixed pattern in a unified model. In this paper, we propose WeaGAN based on encoder-decoder architecture and generative adversarial training process to translate the weather of image among multi-domain. In particular, We employ SE block in generator and combine adversarial loss, classification loss and content loss for visually detailed and realistic result. Experience in qualitative and quantitative aspect on synthetic dataset and real dataset show the effectiveness and competitiveness of our method compared with state-of-the-art works.**

***Index Terms*—weather translation, generative adversarial network, multi-domain translation**

I. Introduction

The task of weather translation of image is to change the weather of a given image from source domain to target domain, e.g., changing the image's weather condition from rainy to sunny(see Fig. 1). Most existing relative works solving weather translation task by handling de-fogging and de-raining respectively, because bad weather conditions such as rain and fog will have negative effect on detection and recognition algorithms in cameras. Single image de-fogging and de-raining methods are divided into prior knowledge-based approaches and learning-based approaches. Former method relies on artificial design features, for example, dark channel prior is used for de-fogging and low-rank model is applied in de-raining. However, these traditional methods with poor extendibility tend to lead to over smooth result. Recently, learning-based approaches have been designed to handling de-fogging and de-raining taking advantage of the end-to-end mapping ability of convolutional neural networks(CNNs). [1] learns to map foggy image to its correspond medium transmissions by CNN architecture to do de-fogging. [2] trains a CNN on the detail layer instead of the image domain for removing rain from individual images.

Due to the effective application of GAN and cGAN in the field of image-to-image translation, a possible method to address these issues is viewing them as image-to-image translation task which aims to translate the particular feature of an input image to desired feature. For instance, to convert a image's style from black and white to colorful without changing its content. GANs try to model image distribution of the target domain by generating fake image which are distinguishable from real image sampled from the target domain. It has wildly application such as style transfer, face attribute edit and super resolution. Based on cGAN, [3] and [4] integrate adversarial loss, pixel loss and perceptual loss into generator to achieve de-fogging and de-raining respectively.

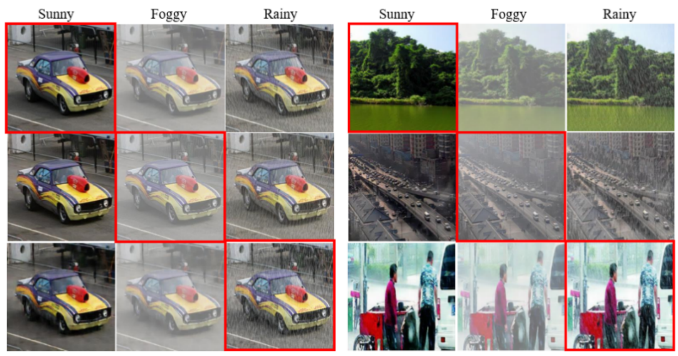

Fig. 1. Sample results of the proposed WeaGAN. Images with red bound represent input images and other images are generated results. Left: results in synthetic dataset. Right: results in real dataset.

Nevertheless, these works have limitation in practical application in which we expect to deal with these tasks simultaneously in a unified model. It is inefficient that for the purpose of learning all mapping relations among k domains, $k(k-1)$ cross-domain models need to be trained, as shown in Fig. 2. Firstly, in some device such as traffic surveillance camera, only lightweight model can be load, and for every frame of the video, multi models should be switched based on the classification of image while single model only needs original image and target label. Furthermore, weather translation tasks are not independent, they share content features which can be fully utilized in a single model while cross-domain models can only learn features from two domains. AttGAN proposed by He et al. [5] and StarGAN put forward by Choi et al. [6] have already achieved image-to-image translation among multi-domain within a single model by using reconstruction

978-1-7281-4762-8/19/$31.00 © 2019 IEEE

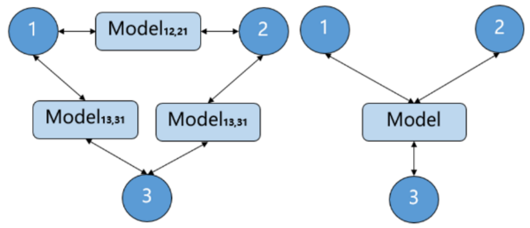

Fig. 2. To learn mapping among three domains, left: six models need to be trained in cross-domain model, right: we use a single model to handle the task.

loss and attribute classification constrain based on cGAN. They are applied in editing human face attribute, which have limitation in weather translation of complex and diversified scene images. Inspired by AttGAN, we propose WeaGAN to realize weather translation of image among multi-domain in a unified model. Our contributions are as follow:

- We preprocess images from DID-MDN [7] to synthesis images with the same content but under three different weather conditions for training.
- In WeaGAN, we add SE block [8] on generator and use content loss in the optimization to get results with better perceptual quality.
- We present a simple and efficient model for handling single image's weather translation among multi-domain, which is demonstrated to be valid in both synthesis dataset and real dataset and even outperform other models designed for mapping between two domains.

II. Related work

A. Two domains translation

GAN using mini-max game theory to fit the distribution of training data through training generator and discriminator simultaneously has shown potential ability in image generation. cGAN can synthesis image with special class via adding conditional information have been applied in image-to-image translation task recently. Pix2Pix [9] based on cGAN, which proposed by Phillip et al. is a general framework for image-to-image translation, in which training images are put into the U-Net structure generator and then generated images as well as real images are put into discriminator. These two training processes alternately iterate by optimizing the loss function combining adversarial loss and L1 loss, but the model needs paired images which sometimes may be difficult to collect. To address the difficulty of getting paired data, CycleGAN [10] couples two GANs together, utilizing symmetrical training strategy and cycle consistency loss function for optimization. [11] and [12] enhance CycleGAN by refining network structure and constraint condition for de-fogging and de-raining. However, these models can only learn mapping between source domain and target domain.

B. multi-domain translation

In order to solve image-to-image translation among multi-domain in one model, several GAN-based approaches are developed. IcGAN [13] train Encoder once the cGAN has been trained to map input image x to z representing content and y representing attribute which can be edited, then y is put into generator together with z to get x' with desired attribute. Fader Networks [14] utilize adversarial training process to force encoder to learn attribute-invariant represent which then be decoded conditioned on arbitrary attribute vector. These two models separating content represent from attribute strictly can result in information loss. StarGAN put image together with attribute vector into generator, add classifier in discriminator to classify generated image's attribute and use reconstruction loss to ensure the consistency of content. Different from StarGAN, AttGAN first encoder input image to latent representation which is concatenated with attribute represent and then be input into decoder, which is proved to preserve content information better. StarGAN and AttGAN are proposed for facial attribute editing whose input face images usually have a fixed pattern and need smaller change than scene images which need more variations on texture and color when translating. StarGAN and AttGAN use reconstruction loss to preserve attribute-excluding detail of their input images, it is time consuming when training and is not suitable for weather translation for its weak constraint.

C. Perception loss

Perceptual loss proposed by L et al. [15] measuring the discrepancy between the high-level features extracted from VGG network is proved to be able to promote perceptual quality of synthesis image and has been used for style transfer and super-resolution task. PAN [16] exploit hidden layer of discriminator instead of VGG network for image-to-image translation task.

Considering that we can synthesis training images containing the same contents while owning different weather conditions, we build our model based on AttGAN and design loss function which is composed of adversarial loss, classification loss and content loss for perceptual convincing result.

III. Proposed method

Our network architecture, shown in Fig. 3, is comprised of two basic sub-networks: a generator consisting of an encoder and a decoder, together with a discriminator consisting of an attribute classifier and a real/fake discriminator. We add Squeeze-and-Excitation block after convolution layers in generator aims to assign different weights to different feature maps to help model learn more useful features. U-Net [17] architecture which has been shown to produce high quality result in image-to-image translation is used to symmetrically connect the SE layer in encoder with that in decoder. SE layer assign different weights to different feature maps, which is used for image classification and then is proven to be effective in improving quality of generated images.

A. Adversarial loss

To make fake image indistinguishable from real image, we constrain generator and discriminator by adversarial loss as

978-1-7281-4762-8/19/$31.00

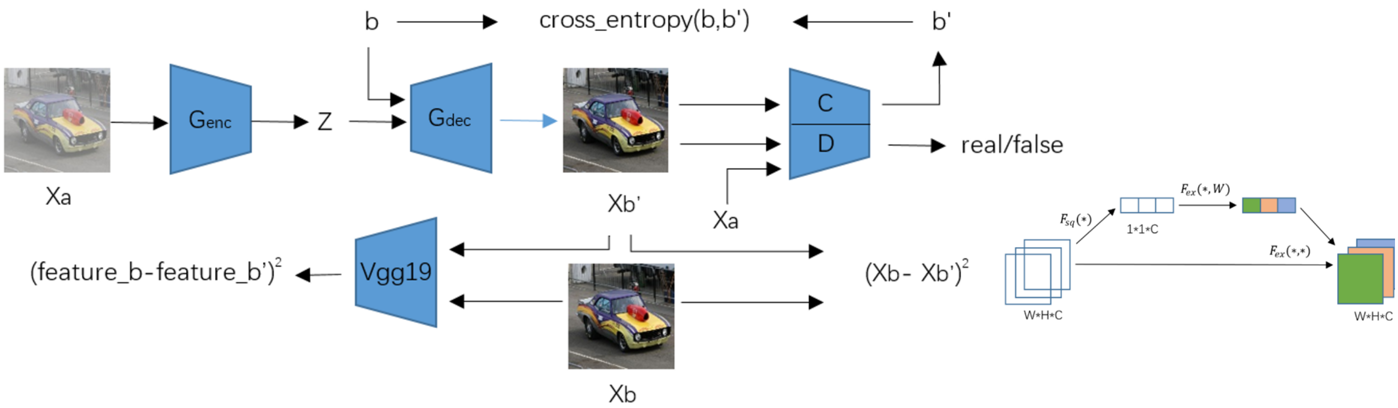

Fig. 3. Left: over view of the network of proposed WeaGAN which consist of two modules: a generator and a discriminator. Right: SE block

most GAN adopt. G and D represent generator and discriminator respectively. It is known that the training process of GAN is often unstable, so we use WGAN to replace traditional adversarial loss and optimize it by WGAN-GP [18]. It is defined as follow:

$$L_{adv} = \mathbb{E}_x(D(x)) - \mathbb{E}_{x,c}[D(G(x,c))] \\ -\lambda_{gp}\mathbb{E}_{x'}[(\|\nabla_{x'}D(x')\|_2) - 1)^2], \quad (1)$$

x and c represent input images and target domain code. x' is randomly interpolated sampled on the line of true and false samples, we use $\lambda_{gp} = 10$.

B. Domain classification loss

One-hot vector is used to represent image attribute. Given image X_a whose attribute is a and target attribute b, it is required that the generated image $X_{b'}$ is classified to the target domain b. When optimizing D and G, we impose different domain classification loss. The former one is defined as:

$$L_c^r = \mathbb{E}_a[-\log C(X_a)], \quad (2)$$

The term $C(X_a)$ represents the distribution of domain computed by classifier given real data X_a and try to fit the real domain distribution. The later one is defined as:

$$L_c^f = \mathbb{E}_b[-\log C(G(X_a, b))], \quad (3)$$

Generator try to generate image $X_{b'}$ which is classified to target domain b.

C. Content loss

Since the generated image $X_{b'}$ have reference image X_b, we add perceptual loss which is a criticism for generator. It compares the Euclidean distance between the feature maps extracted from pre-trained VGG19 network of X_b and $X_{b'}$.

$$L_{VGG}(X_b, X_{b'}) = \frac{1}{H_i W_j}\|\phi_j(X_b) - \phi_j(X_{b'})\|_2^2, \quad (4)$$

ϕ_j represents the $j-th$ layer of VGG19, we use $j = 11$. H_i, W_j are the height and width of feature map. We also calculate the Manhattan distance of generated image and ground-truth image for shaper result, so the content loss is defined as :

$$L_{con} = \lambda_{VGG}L_{VGG} + \|X_b - X_{b'}\|_1, \quad (5)$$

λ_{VGG} represents hyper-parameter of perceptual loss, we set 0.006 in our experience.

D. Full object

The object functions to optimize G and D are defined respectively as:

$$L_D = -L_{adv} + \lambda_{cls\_D}L_c^r, \quad (6)$$

$$L_G = L_{adv} + \lambda_{cls\_G}L_c^f + \lambda_{con}L_{con}, \quad (7)$$

In our experience, $\lambda_{cls\_D} = 1.0$, $\lambda_{cls\_G} = 10$, $\lambda_{con} = 1000$.

IV. Experience

A. Dataset

In order to get images of the same content under different weather conditions, we use paired rainy-sunny images(add rain streak to sunny images by Photoshop) from DID-MDN, then we adjust the light condition $A = 0.89$, and the scattering coefficient $\beta = 0.63$ of sunny images to synthesis their corresponding foggy images. Finally, we get 11400 training images and 600 test images which are scaled to 256×256 pixel and composed of three categories[sunny/foggy/rainy] of average.

B. Qualitative evaluation

To validate the performance of the proposed algorithm, we evaluate the quality of generated images using two classical criterions to compare them with corresponding ground-truth: Point Signal-to-Noise(PSNR) and Structural Similarity Index(SSIM). Meanwhile, we compare our method (WeaGAN and WeaGAN-SE) with PAN, CycleGAN, StarGAN and AttGAN and all of them are trained on our dataset. Table I presents the comparison results for different methods mentioned above. It can be observed that our method can generate competitive results whose PSNR and SSIM are higher than StarGAN and AttGAN, although it is lower in some terms

978-1-7281-4762-8/19/$31.00

TABLE I
QUANTITATIVE RESULTS COMPARISON. F-S: FOGGY IMAGE TO SUNNY IMAGE; R-S: RAINY IMAGE TO SUNNY IMAGE, THE REST CAN BE DEDUCED BY ANALOGY. RED: HIGHEST AMONG ALL METHODS. BLUE: HIGHEST AMONG METHODS BELONGING TO THE SAME MODEL TYPE.

		F-S		R-S		S-F		S-R		F-R		R-F	
		ssim	*psnr*	*ssim*	*psnr*	*ssim*	*psnr*	*ssim*	*psnr*	*ssim*	*psnr*	*ssim*	*psnr*
6 Models	PAN	0.74	23.36	0.62	21.64	0.84	28.84	0.42	19.60	0.40	18.91	0.78	28.84
	CycleGAN	0.89	26.97	0.76	24.21	0.95	32.91	0.58	21.64	0.53	21.32	0.89	32.91
Single Model	StarGAN	0.77	23.53	0.65	22.47	0.79	26.48	0.37	19.05	0.38	19.28	0.72	26.63
	AttGAN	0.85	22.93	0.79	25.42	0.89	28.85	0.48	21.02	0.46	20.84	0.81	27.59
	Ours	0.95	33.57	0.83	27.83	0.89	33.00	0.50	21.86	0.50	21.91	0.83	30.63
	Ours-SE	0.95	34.08	0.84	27.84	0.90	33.37	0.50	21.96	0.50	21.97	0.84	31.18

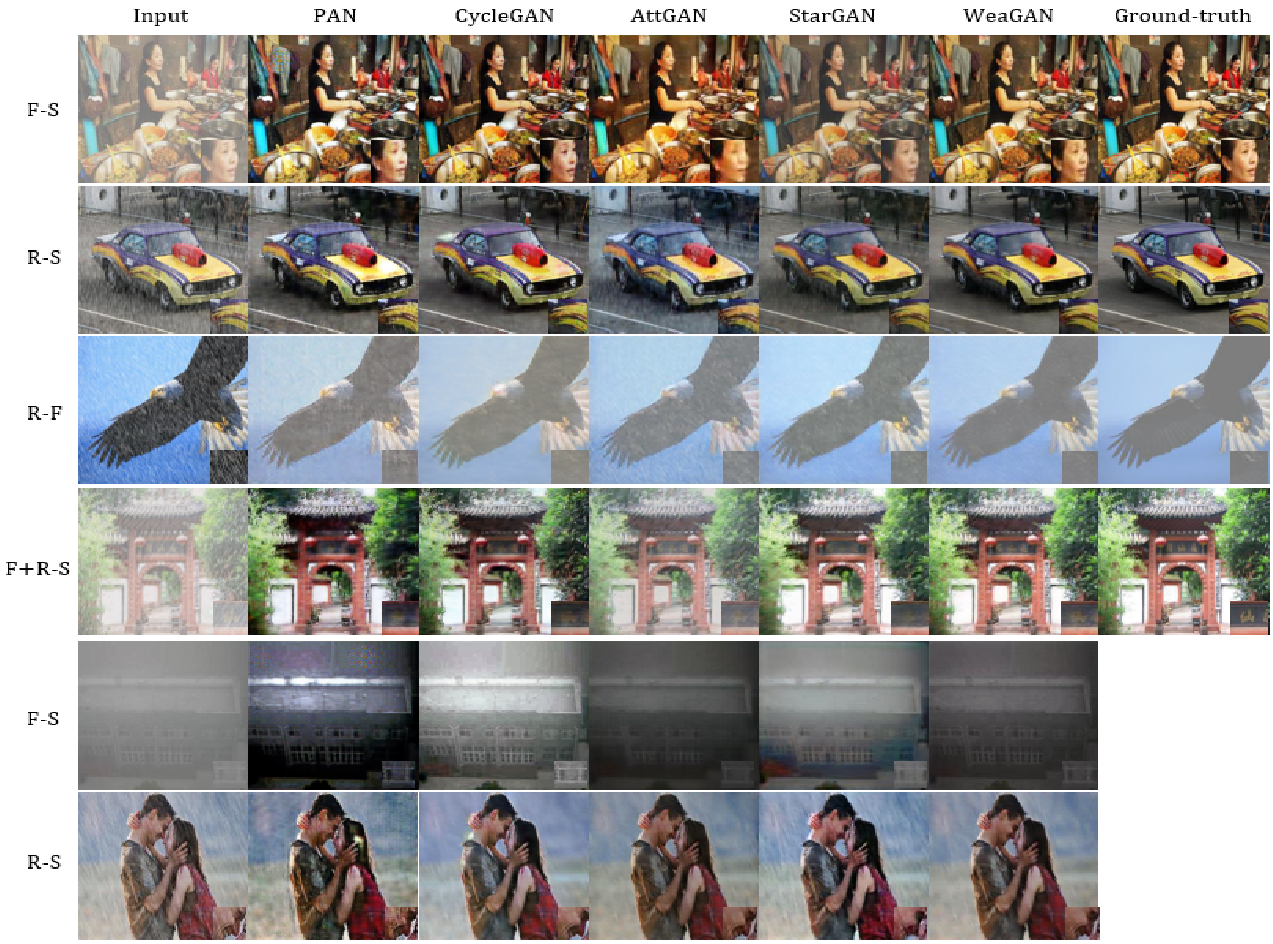

Fig. 4. Comparison samples of weather translation. 1, 2, 3 and 4 lines show synthesis images, 5, 6 lines show real world images.

such as S-F compared with CycleGAN but very close and WeaGAN just need train a model while CycleGAN need train 6 models. Fig. 4 shows generated images comparison results. It can be seen that our method can generate images with sharper, detailed and realistic structure.

V. EXTENSION

Considering the weather condition raining and fogging at the same time, we add an additional category: foggy-rainy day. Similarly, we adjust the light condition and the scattering coefficient of rainy images with the same parameter as that used in sunny images, then we trained these four classes[sunny/foggy/rainy/foggy-rainy] in several models for comparison. Table II shows the quality comparison results of the translation from foggy to sunny, rainy to sunny and foggy-rainy to sunny because they are more often used in practical application. It can be seen that our method has excellent results in image attribute translations from bad weather(foggy/rainy/foggy-rainy) to good weather(sunny).

VI. CONCLUSION

In this paper we propose an encoder-decoder based network using adversarial training process to realize weather translation of image among multi-domain in a single model. We add SE

TABLE II
QUANTITATIVE RESULTS COMPARISON OF IMAGE TRANSLATION FROM BAD WEATHER TO GOOD WEATHER. F+R-S: FOGGY-RAINY IMAGE TO SUNNY IMAGE; RED: HIGHEST AMONG ALL METHODS.

		F-S		**R-S**		**F+R-S**	
		ssim	*psnr*	*ssim*	*psnr*	*ssim*	*psnr*
6 Models	PAN	0.74	23.36	0.62	21.64	0.62	21.53
	CycleGAN	0.89	26.97	0.76	24.21	0.75	23.15
Single Model	StarGAN	0.74	24.69	0.62	22.96	0.65	22.04
	AttGAN	0.85	22.54	0.80	25.54	0.73	18.89
	Ours-SE	0.96	35.45	0.84	28.80	0.84	28.75

block in generator to enhance generalization ability and adding perceptual loss calculating distance of features extracted from pre-trained VGG19 network to get detailed result. Compared with several state-of-the-art methods, our network is proved to have superior performance. However, it still suffer from some defects such as paired training data are demanded. In the future, we aim to design more robust network which can handle unpaired images.

ACKNOWLEDGMENT

This work is supported in part by the Natural Science Foundation of China (NSFC) under Grants 61672088, 61790575 and 61790573. The corresponding author is Haidong Cui.

REFERENCES

[1] J. K. Cai B, Xu X, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[2] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley, "Clearing the skies: A deep network architecture for single-image rain removal," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2944–2956, 2016.

[3] R. Li, J. Pan, Z. Li, and J. Tang, "Single image dehazing via conditional generative adversarial network," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[4] Z. He, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," 2017.

[5] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, "Arbitrary facial attribute editing: Only change what you want," 2017.

[6] J. W. Ha, Y. Cho, M. Choi, M. Kim, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," 2017.

[7] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network," 2018.

[8] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," 2017.

[9] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2016.

[10] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE International Conference on Computer Vision*, 2017.

[11] D. Engin, A. Genc, and H. K. Ekenel, "Cycle-dehaze: Enhanced cyclegan for single image dehazing," 2018.

[12] J. Pu, X. Chen, L. Zhang, Q. Zhou, and Y. Zhao, "Removing rain based on a cycle generative adversarial network," in *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2018, pp. 621–626.

[13] G. Perarnau, J. V. D. Weijer, B. Raducanu, and J. M. lvarez, "Invertible conditional gans for image editing," 2016.

[14] G. Lample, N. Zeghidour, N. Usunier, A. Bordes, L. Denoyer, and M. Ranzato, "Fader networks: Manipulating images by sliding attributes," 2017.

[15] J. Johnson, A. Alahi, and F. F. Li, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*, 2016.

[16] C. Wang, C. Xu, C. Wang, and D. Tao, "Perceptual adversarial networks for image-to-image transformation," *IEEE Transactions on Image Processing*, vol. PP, no. 99, pp. 1–1, 2018.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.

[18] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein gans," 2017.