GAN based data augmentation for histopathological image segmentation

Florian Allender*, Rémi Allègre*, Cédric Wemmert*, Jean-Michel Dischler*

* ICube, Université de Strasbourg
CNRS, France

Abstract In the context of kidney transplant, histopathological images and deep learning are powerful tools to help keep track of diseases related with transplant rejection. However, the training of neural networks require huge amounts of data that are not always available due to the lack of annotation. The use of synthetic data to train algorithms has been proven effective even in medical imaging. We aim at providing a pipeline composed of Generative Adversarial Networks to produce high resolution glomeruli patches that can be used to train a segmentation network. As it is still a work in progress, we focus in this article on the second part of the pipeline : generating images from segmentation masks. We use image translation techniques and in particular the Pix2Pix network. We show that adding structure maps to the input and a regularizing loss helps mitigate the issue of mode collapse and produce good looking results.

1 Introduction

Every year, thousands of patients around the world undergo kidney transplant surgery, while other thousands die while on the waiting list. In this context, the study of kidney transplant rejection is crucial in order to save more lives. Interstitial Fibrosis and Tubular Atrophy (IFTA) and glomerulosclerosis are two pathologies associated with chronic kidney transplant rejection. The study of these pathologies could lead to a better understanding of the mechanisms behind transplant rejection, and thus help reduce the loss of transplanted organs. To do so, researchers and practitioners can rely on Whole Slide Imaging (WSI) and the development of digital histopathology. The objects of interest in these extremely high resolution images are glomeruli, clusters of blood vessels allowing blood filtration. To this end, we need to detect and segment glomeruli on patches extracted from the complete images.

Deep Learning has revolutionized the area of image processing, providing powerful tools to automatically accomplish various tasks such as image recognition (Krizhevsky et al., 2012), voice generation (van den Oord et al., 2016) or self-driving cars (Santana and Hotz, 2016), with great success, sometimes outperforming humans (He et al., 2015; Mnih et al., 2015). In the field of medical images, Deep Learning is closing the gap between clinicians and AI performances (Liu et al., 2019) allowing them to process those images faster, with more accuracy. However the application of Deep Learning on medical images comes with its own set of limitations, one of the main being the lack of annotated data. This is in particular true for

histopathological images and the challenge we are tackling. The segmentation of many WSI requires expert knowledge and is an extremely time consuming task, and as a result an expensive one. Moreover, privacy issues make it difficult for researchers to share their data, making it difficult to collect large databases on which we could test the algorithms of the community, providing a common reference to evaluate their performances.

To cope with the lake of data, new architectures and training procedures have been proposed. A standard procedure in Deep Learning is to artificially augment the size of databases is the use of random affine transformations (rotations, translations, scaling) on the training set (Krizhevsky et al., 2012). In (Ronneberger et al., 2015), the authors propose a deep neural network, U-Net, to segment medical images, as well as elastic transformations to augment their database. U-Net can be train with less data while outperforming previous architectures on the ISBI challenge. We will detail the architecture in section 3. In (Lampert et al., 2019), the authors propose a procedure to learn the segmentation of glomeruli on images with different staining. To go further, a new approach has aroused in recent years : training networks with synthetic data (Nikolenko, 2019).

This approach has been made far more efficient with the introduction of Generative Adversarial Networks (GANs) in 2014 (Goodfellow et al., 2014). Their goal is to learn an implicit representation of a dataset distribution that can be sampled to produce new data. GANs are composed of a pair of Neural Networks : a generator and a discriminator. The generator takes random noise as input and output a data (an image in our case). The discriminator takes a data (real or fake) as input, and outputs a digit indicating if the data is real or fake. Both networks are trained together in a competitive manner, until G learns the data distribution and D is not able to differentiate between fake and real data anymore. GANs have made quick progress and took many forms, from Convolutional GANs (Radford et al., 2015) to Conditional GANs (Odena et al., 2016). They are now able to synthesize high resolution images with astonishing realism (Karas et al., 2018; Karras et al., 2018; Park et al., 2019). GANs still suffer from a few drawbacks : they are notoriously difficult to train, require careful hyper-parameter tuning and have difficulties to produce diverse results (Goodfellow, 2017; Lucic et al., 2018; Salimans et al., 2017; Arjovsky and Bottou, 2017). Indeed, they have a tendency to focus on only a few mode of the data distribution. This issue is known as mode collapse. Many tricks have been proposed in the literature to solve it (Metz et al., 2016; Mao et al., 2019; Che et al., 2016; Arjovsky et al., 2017; Srivastava et al., 2017), but they seem to work only with specific architectures or datasets.

GANs have been used in medical images to help enhance the results of Deep Learning methods on various tasks (reconstruction, registration, segmentation, detection...), as highlighted by the recent review (Yi et al., 2018). Conditional GANs and in particular the ones performing image domain translation such as Pix2Pix (Isola et al., 2016) and CycleGAN (Zhu et al., 2017) are popular in the field. The use of synthetic data to train classification or segmentation algorithms in the medical field has been proven effective (Mahmood et al., 2018; Xiao et al., 2019; Frid-Adar et al., 2018; Hou et al., 2017; Senaras et al., 2018). Our aim in this paper is to propose a new pipeline able to generate high resolution synthetic glomeruli patches. We validate our results by two different means : a perception study with experts, and the training of a U-Net with our synthetic data. [1]

1. The first part of the pipeline and the complete analysis of the results are still a work in progress and will not be shown in this version of the paper. The conclusion will be modified accordingly.

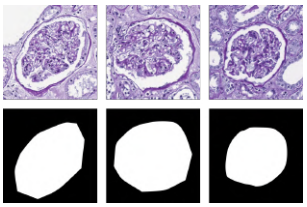

FIG. 1 – *Example of glomeruli patches with associated segmentation masks.*

2 Materials

We have at our disposal 10 annotated WSI from different patients and with different staining, distributed between training, testing and validating set. We extracted patches centered around glomeruli from those images with the corresponding annotation, which is a segmentation mask. See figure 1 for examples. These patches are usually used to train a U-Net for segmentation purposes (Lampert et al., 2019). In this article, we will use the training set to train GANs and then use the produced synthetic data to train the network from (Lampert et al., 2019). To simplify the problem and ease the conduct of our experimentations, we focused on only one staining. In the end, we have 660 pairs of glomeruli and mask with a 256*256 size.

Glomeruli patches are difficult to synthesize because the size of the object of interest is large with respect to the size of the image. we have to reproduce the global structure and local patterns. Thus we can not process it as a stochastic texture only. To do so, GANs are promising tools that may help us captures semantic information at different scale levels.

In order to use our synthetic data to train a U-Net, we need to generate a glomeruli patch and the corresponding mask at the same time. To do so, we first imagined to seperate the problem into two parts : generating a new mask, then generating a new glomerulus with respect to this mask. In this article, we focus on the second step. To generate a glomerulus that respects the constraints imposed by a mask, we use Pix2Pix (Isola et al., 2016), as it takes a semantic label map as input and outputs the corresponding image. However Pix2Pix severely suffers from mode collapse, has already mentioned in the original paper. As a result, the generator always outputs the same image, even when feed with different inputs, as highlighted by figure 2.

In order to tackle the issue of mode collapse, we propose to add information to the input of our generator, in the form of structure maps. Those structure maps are binary images obtained by thresholding the glomeruli images, as show in figure 3. Our input is now the concatenation of a mask and a structure map. By this mean we give more constraints to the generator and completely avoid mode collapse during the image translation phase, as will be shown in the Results section. The issue is not completely solved as we merely moved the burden of generating diversity to a structure and mask generation phase.

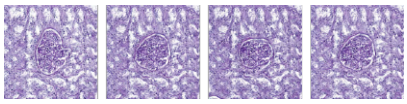

FIG. 2 – *Illustration of mode collapse. The generator always outputs the same texture content, no matter the input.*

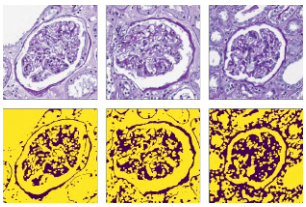

FIG. 3 – *Example of glomeruli patches with associated structure map.*

3 Methods

3.1 U-Net description

U-Net (Ronneberger et al., 2015) is a network with two branches : one that performs data compression with downsampling layers and the other that upscale the data back to its original resolution, just like auto-encoders (Rumelhart et al., 1986). The difference between auto-encoders and U-Net is the presence of skip connections that connect the two branches, allowing information to flow from one side to the other. See figure 4 for details on the original architecture.

3.2 GAN description

A GAN is composed of two networks : a generator (G) and a discriminator (D). We denote p_{data} the distribution of the data and p_g the distribution learned by G. We note p_z the distribution from which we sample a random vector z, used as input for G. G and D play a zero sum game, described by the following equation (Goodfellow et al., 2014) :

$$\min_G \max_D L(D,G) = \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[logD(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_z(\mathbf{z})}[log(1-D(G(\mathbf{z})))] \quad (1)$$

Theoretically there exists a optimum for this game when $p_g = p_{data}$, but it is notoriously difficult to find, especially with the stochastic gradient descent techniques used to train neural networks.

3.3 Generating images using a U-Net : mode seeking Pix2pix

Pix2pix is an image translation model, a type of Conditional GAN that takes a label map as input and produce an image as output. It uses a U-Net as generator and can be used to generate images from sketches. The GAN equation is then modified. G now learns a mapping from an image x to an image y. In our case, x is the concatenation of a mask and a structure map.

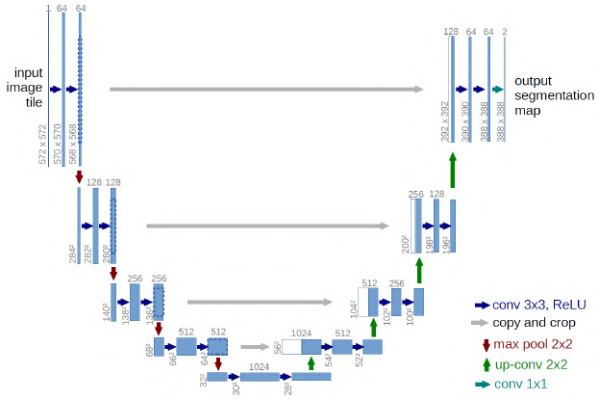

FIG. 4 – *Original U-Net architecture. (Ronneberger et al., 2015)*

In the Conditional GAN framework, D usually receives as input the concatenation of x and y, in order to give poor rating to an image that does not correspond with the condition, not matter how realistic it may be. However not feeding x to D helps reduce mode collapse, so we only feed y as for classic GANs, at the price of a slight loss in visual quality. It gives us the following modified equation :

$$\min_G \max_D L_{cGAN}(D,G) = \mathbb{E}_{\mathbf{x},\mathbf{y}\sim p_{data}(\mathbf{x},\mathbf{y})}[logD(\mathbf{y})] + \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[log(1 - D(G(\mathbf{x})))] \quad (2)$$

The authors also found that adding a L1 distance between the output of G and the real image from the dataset helps to reduce blurring :

$$L_{L1}(G) = \mathbb{E}_{\mathbf{x},\mathbf{y}\sim p_{data}(\mathbf{x},\mathbf{y})}[||\mathbf{y} - G(\mathbf{x})||_1] \quad (3)$$

To that we add a regularizing loss inspired by Mode Seeking GANs (Mao et al., 2019). It aims at reducing mode collapse in the case of conditional GANs by penalizing the generator if it produces similar images for two different condition masks x_1 and x_2 :

$$L_{ms}(G) = \mathbb{E}_{\mathbf{x}_1\sim p_{data}(\mathbf{x}_1),\mathbf{x}_2\sim p_{data}(\mathbf{x}_2)} \frac{||\mathbf{x}_1 - \mathbf{x}_2||_1}{||G(\mathbf{x}_1) - G(\mathbf{x}_2)||_1} \quad (4)$$

Which gives us the expression of our optimal generator :

$$G* = \underset{G}{argmin}\, \underset{D}{max} L_{cGAN}(D,G) + \lambda L_{L1}(G) + \mu L_{ms}(G) \quad (5)$$

We set $\lambda = 10$ and $\mu = 10$. We train both our networks with the Adam optimizer, with the following parameters : learning rate $= 0.0002$, $\beta_1 = 0.5$ and $\beta_2 = 0.999$, batch size $= 1$. The complete model of G and D is given in the appendix.

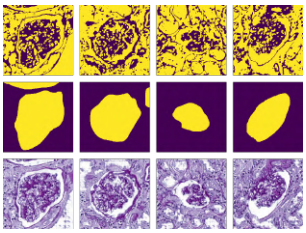

FIG. 5 – *Results from our modified Pix2Pix tested on real structure maps and masks. First row : input structure maps, second row : input masks, third row : results.*

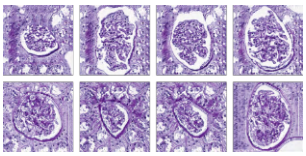

FIG. 6 – *First row : results when training with no regularizing loss. Second row : results when training with no structure maps. In both cases, mode collapse still occurs and the visual quality of the generated samples is not convincing enough.*

4 Results

4.1 Images generated by our mode seeking Pix2Pix

We visually selected twelve epochs with appealing results during training and tested the generator on unseen data. Those data are real masks and structure maps coming from the test set. We show in figure 5 four randomly selected samples from the results. As can be seen here and in the appendix, the images are of very high visual quality and mode collapse is non-existent. The texture created by the generator perfectly matches the structure given as input.

4.2 Ablation study

The use of structure maps and the L_{ms} term are both mandatory to obtain good results. We trained two different models, one using only masks and L_{ms} and the other using masks and structure maps but not L_{ms}. As we show in figure 6, both models collapse and output poor looking results.

We add results obtained when feeding the discriminator with the condition as mention in section 3. The mode collapse effect is not as prominent as in 6, but still visible, especially near the center of the image.

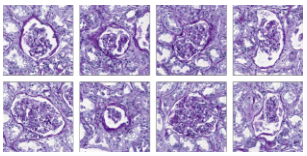

FIG. 7 – *Results when feeding the condition to the discriminator. Mode collapse is still an issue in this case, even if not highly visible.*

5 Conclusion

In this article we exposed the first part of a work in progress. We showed how we can modify an existing image translation model (Pix2Pix) to produce high quality images and alleviate the mode collapse issue. This improvement is made possible by two key ingredients : the use of structure maps to give more constraints to the input of our generator and a regularizing loss term to stabilize the model convergence. This constitutes the second part of our planned pipeline.

The first part of the pipeline is still worked on and will be crucial to our application. We have now to synthesize masks and structure maps of enough quality to feed our modified Pix2Pix, but also with enough diversity so that we can use the images produced by the complete pipeline as a training set.

When the pipeline is complete, we will be able to validate our method by two different means. First a perception study with histopatholgy experts to see if our images can fool the human eye, then a quantitative study to check if a segmentation network trained by our synthetic images can generalize well on real images.

References

Arjovsky, M. and L. Bottou (2017). Towards principled methods for training generative adversarial networks. *ArXiv abs/1701.04862*.

Arjovsky, M., S. Chintala, and L. Bottou (2017). Wasserstein gan. cite arxiv:1701.07875.

Che, T., Y. Li, A. P. Jacob, Y. Bengio, and W. Li (2016). Mode regularized generative adversarial networks. *CoRR abs/1612.02136*.

Frid-Adar, M., I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan (2018). Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *CoRR abs/1803.01229*.

Goodfellow, I. J. (2017). NIPS 2016 tutorial: Generative adversarial networks. *CoRR abs/1701.00160*.

Goodfellow, I. J., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio (2014). Generative Adversarial Networks. *ArXiv e-prints*.

He, K., X. Zhang, S. Ren, and J. Sun (2015). Deep residual learning for image recognition. *CoRR abs/1512.03385*.

Hou, L., A. Agarwal, D. Samaras, T. M. Kurç, R. R. Gupta, and J. H. Saltz (2017). Unsupervised histopathology image synthesis. *ArXiv abs/1712.05021*.

Isola, P., J.-Y. Zhu, T. Zhou, and A. A. Efros (2016). Image-to-image translation with conditional adversarial networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5967–5976.

Karas, T., T. Aila, S. Laine, and J. Lehtinen (2018). Progressive growing of gans for improved quality, stability, and variation. *International Conference on Learning Representations*.

Karras, T., S. Laine, and T. Aila (2018). A style-based generator architecture for generative adversarial networks. *CoRR abs/1812.04948*.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25*, pp. 1097–1105. Curran Associates, Inc.

Lampert, T., O. Merveille, J. Schmitz, G. Forestier, F. Feuerhake, and C. Wemmert (2019). Strategies for training stain invariant cnns. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 905–909.

Liu, X., L. Faes, A. U. Kale, S. K. Wagner, D. J. Fu, A. Bruynseels, T. Mahendiran, G. Moraes, M. Shamdas, C. Kern, J. R. Ledsam, M. K. Schmid, K. Balaskas, E. J. Topol, L. M. Bachmann, P. A. Keane, and A. K. Denniston (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health 1*(6), e271 – e297.

Lucic, M., K. Kurach, M. Michalski, O. Bousquet, and S. Gelly (2018). Are gans created equal? a large-scale study. In *Proceedings of the 32Nd International Conference on Neural Information Processing Systems*, NIPS'18, USA, pp. 698–707. Curran Associates Inc.

Mahmood, F., R. Chen, and N. J. Durr (2018). Unsupervised reverse domain adaptation for synthetic medical images via adversarial training. *IEEE Transactions on Medical Imaging 37*(12).

Mao, Q., H. Lee, H. Tseng, S. Ma, and M. Yang (2019). Mode seeking generative adversarial networks for diverse image synthesis. *CoRR abs/1903.05628*.

Metz, L., B. Poole, D. Pfau, and J. Sohl-Dickstein (2016). Unrolled generative adversarial networks. *ArXiv abs/1611.02163*.

Mnih, V., K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis (2015). Human-level control through deep reinforcement learning. *Nature 518*(7540), 529–533.

Nikolenko, S. I. (2019). Synthetic data for deep learning. *ArXiv abs/1909.11512*.

Odena, A., C. Olah, and J. Shlens (2016). Conditional image synthesis with auxiliary classifier gans. In *ICML*.

Park, T., M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu (2019). Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Radford, A., L. Metz, and S. Chintala (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *CoRR abs/1511.06434*.

Ronneberger, O., P. Fischer, and T. Brox (2015). U-net: Convolutional networks for biomedical image segmentation. *CoRR abs/1505.04597*.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, and C. PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1*, pp. 318–362. Cambridge, MA, USA: MIT Press.

Salimans, T., I. J. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen (2017). Improved techniques for training gans. *CoRR abs/1606.03498*.

Santana, E. and G. Hotz (2016). Learning a driving simulator. *CoRR abs/1608.01230*.

Senaras, C., M. K. K. Niazi, B. Sahiner, M. P. Pennell, G. Tozbikian, G. Lozanski, and M. N. Gurcan (2018). Optimized generation of high-resolution phantom images using cgan: Application to quantification of ki67 breast cancer images. *PLOS ONE 13*(5), 1–12.

Srivastava, A., L. Valkov, C. Russell, M. U. Gutmann, and C. Sutton (2017). Veegan: Reducing mode collapse in gans using implicit variational learning. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30*, pp. 3308–3318. Curran Associates, Inc.

van den Oord, A., S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. W. Senior, and K. Kavukcuoglu (2016). Wavenet: A generative model for raw audio. *CoRR abs/1609.03499*.

Xiao, Y., E. Decencière, S. Velasco-Forero, H. Burdin, T. Bornschlögl, F. Bernerd, E. Warrick, and T. Baldeweck (2019). A NEW COLOR AUGMENTATION METHOD FOR DEEP LEARNING SEGMENTATION OF HISTOLOGICAL IMAGES. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI)*, Venise, France.

Yi, X., E. Walia, and P. Babyn (2018). Generative adversarial network in medical imaging: A review. *ArXiv abs/1809.07294*.

Zhu, J., T. Park, P. Isola, and A. A. Efros (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251.

Appendix 1 : Model description

Layer	Output shape	Number of parameters
input	(1, 256, 256, 2)	0
conv2d	(1, 128, 128, 64)	1664
leaky ReLU	(1, 128, 128, 64)	0
conv2d	(1, 64, 64, 128)	205184
leaky ReLU	(1, 64, 64, 128)	0
conv2d	(1, 32, 32, 256)	819968
leaky ReLU	(1, 32, 32, 256)	0
conv2d	(1, 16, 16, 512)	3278336
leaky ReLU	(1, 16, 16, 512)	0
conv2d	(1, 8, 8, 512)	6555136
leaky ReLU	(1, 8, 8, 512)	0
conv2d	(1, 4, 4, 512)	6555136
leaky ReLU	(1, 4, 4, 512)	0
conv2d	(1, 2, 2, 512)	6555136
leaky ReLU	(1, 2, 2, 512)	0
conv2d	(1, 1, 1, 512)	655513
ReLU	(1, 1, 1, 512)	0
deconv2d	(1, 2, 2, 512)	6555136
dropout	(1, 2, 2, 512)	0
concatenation	(1, 2, 2, 1024)	0
ReLU	(1, 2, 2, 1024)	0
deconv2d	(1, 4, 4, 512)	13108736
dropout	(1, 4, 4, 512)	0
concatenation	(1, 4, 4, 1024)	0
ReLU	(1, 4, 4, 1024)	0
deconv2d	(1, 8, 8, 512)	13108736
dropout	(1, 8, 8, 512)	0
concatenation	(1, 8, 8, 1024)	0
ReLU	(1, 8, 8, 1024)	0
deconv2d	(1, 16, 16, 512)	13108736
concatenation	(1, 2, 2, 1024)	0
ReLU	(1, 16, 16, 1024)	0
deconv2d	(1, 32, 32, 256)	6554368
concatenation	(1, 2, 2, 512)	0
ReLU	(1, 32, 32, 512)	0
deconv2d	(1, 64, 64, 128)	1638784
concatenation	(1, 2, 2, 256)	0
ReLU	(1, 64, 64, 256)	0
deconv2d	(1, 128, 128, 64)	409792
concatenation	(1, 2, 2, 128)	0
ReLU	(1, 128, 128, 128)	0
deconv2d	(1, 256, 256, 3)	9603

TAB. 1 – *Pix2Pix generator model*

Layer	Output shape	Number of parameters
input	(1, 256, 256, 3)	0
conv2d	(1, 128, 128, 64)	4864
leaky ReLU	(1, 128, 128, 64)	0
conv2d	(1, 64, 64, 128)	205184
leaky ReLU	(1, 64, 64, 128)	0
conv2d	(1, 32, 32, 256)	819968
leaky ReLU	(1, 32, 32, 256)	0
conv2d	(1, 32, 32, 512)	3278336
leaky ReLU	(1, 32, 32, 512)	0
reshape	(1, 524288)	0
linear	(1, 1)	0

TAB. 2 – *Pix2Pix discriminator model*

Appendix 2 : Additional results

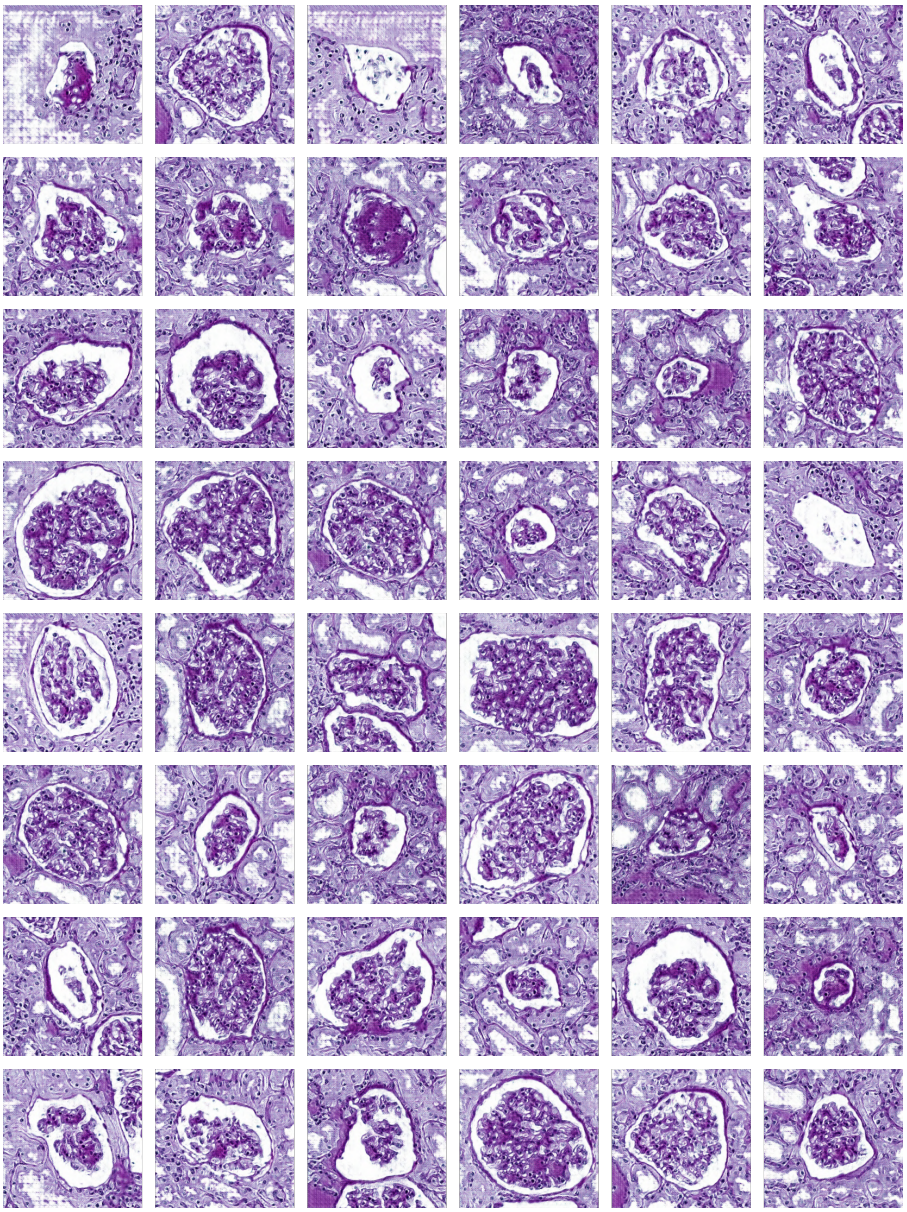

FIG. 8 – *Batch of 48 randomly selected results.*

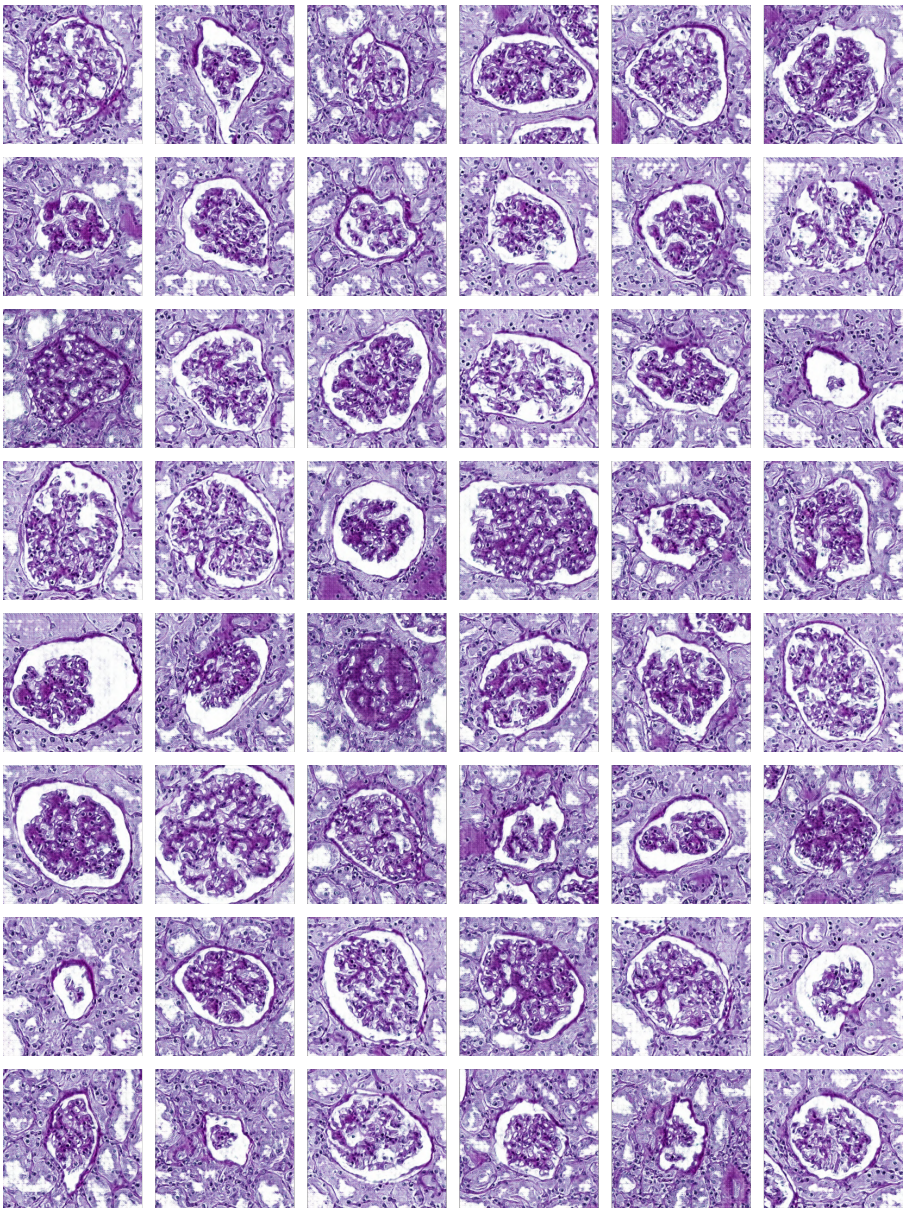

FIG. 9 – *Second batch of 48 randomly selected results.*