University of Nevada, Reno

# Generative Adversarial Networks for Synthesizing Medical Images of Multiple Modalities

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in
Computer Science and Engineering

by

Sharif Amit Kamran

Dr. Alireza Tavakkoli – Thesis Advisor
Dr. Sal Baker – Thesis co-Advisor
December 2020

THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

**SHARIF AMIT KAMRAN**

entitled

**Generative Adversarial Networks for Synthesizing Medical Images of Multiple Modalities**

be accepted in partial fulfillment of the
requirements for the degree of

**MASTER OF SCIENCE**

Alireza Tavakkoli, Ph.D.
*Advisor*

Sal Baker, Ph.D.
*Co-advisor*

Mircea Nicolescu, Ph.D.
*Committee Member*

Tin Nguyen, Ph.D.
*Committee Member*

Kenton M. Sanders, Ph.D.
*Graduate School Representative*

David W. Zeh, Ph.D., Dean
*Graduate School*

December, 2020

## Abstract

Deep learning architectures have revolutionized the field of medical image analysis and consistently achieve state-of-the-art accuracy by learning from high volumes of data. Despite these advances, differential diagnoses could still potentially require a multitude of complicated procedures with potentially adverse effects. For example, fluorescein angiography (FA) for detecting retinal neurovascular abnormalities requires the injection of an exogenous dye to capture retinal vascular structures. This procedure could cause potentially life-threatening allergic reactions to the dye. Although there is only one non-invasive procedure for generating retinal vasculature is optical coherence tomography-angiography (OCTA), this technology is very expensive and only suitable in limited cases – e.g. small macula regions. In addition, limitations in medical imaging technologies result in low-quality and noisy data. As another example, Calcium imaging is used to monitor $Ca^2+$ transient activities in pacemaker population in the gut, termed: interstitial cells of Cajal (ICC). $Ca^2+$ spatio-temporal maps (STMaps) are utilized to effectively quantity $Ca^2+$ signal events in these cells. These maps contain high levels of sensory noise in these and manual generation and processing of STMaps is intractable. These challenges introduce the necessity for novel approaches to process medical image for developing non-invasive screening protocols with potential for automation. This thesis investigates novel approaches to fundamentally overcome these problems. The main challenge is extracting a representative feature manifold from which higher level of information can be obtained. Accurately establishing a shared feature manifold will potentially improve outcome in the absence of certain measurements, the availability of potentially conflicting data, and under high signal-to-noise ratios (SNR). Generative networks are investigated in this thesis to establish a theoretical platform for extracting shared features from different

sensory domains. The proposed architecture incorporates multiple attention-based skip connections in generators and comprises novel residual blocks for both generators and discriminators. Additionally, it employs reconstruction, feature-matching, and perceptual loss along with adversarial training to fundamentally learn shared features across domains. This work is the first in the literature to employ the proposed pipeline for multiple medical imaging modalities while investigating the use of decoding to further enhance semantic segmentation. To showcase the proposed architecture, anatomically accurate fluorescein angiography images are produced from fundus images. In addition, the framework shows significant promise in accurately segmenting extremely noisy $Ca^2+$ STMaps for calcium imaging.

# Dedication

To my late father, Sharif Kafi, for instilling the dream to pursue research and not giving up. To my mother, Shahana Sultana, for nurturing and paving the path for my education. And lastly but most importantly, my wife, Fria, for having my back in the most troublesome time, motivating me to finish the task at hand, and always driving me to the limits.

# Acknowledgments

Firstly, I would like to thank my advisors, Professor Alireza Tavakkoli and Professor Sal Baker, for giving me the opportunities to pursue research on pivotal and practical problems over the past one and a half years. Dr. Ali and Dr. Sal, was there to guide and help me in conducting essential experiments without thinking twice. They let me determine my own goals and advised me on the way. From both of them, I have acquired a tremendous amount of knowledge related to the vast field of Computer Vision and Medical Imaging and also the difficulties of research, academic duties, and publishing scholarly works in prestigious venues.

Next, I would also like to acknowledge my colleagues, Khondker Fariha Hossain, Wesley Leigh, and Guillermo Del Valle, for sharing cognizant ideas and timely cooperation. I went back and forth on many ideas and experimentation with my colleagues that were a fundamental part of our research and publications. Without their tireless efforts, I wouldn't have been able to improve our proposed techniques and methods. Consequently, I have gained both practical and theoretical knowledge across multiple subfields of Machine Learning and Biomedical Imaging, which would be quite impossible to acquire without their constant involvement.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

In recent times, there has been a surge in use of automated computer-aided medical diagnosis [1–4]. Computer-aided diagnosis incorporates advanced machine learning methods combining computer vision and image processing techniques [5]. The majority of medical experts utilize computer-aided medical imaging for differential diagnosis. To diagnosis more precisely, deep learning methods have shown tremendous potential to generate promising results, outperforming experts in the field of medical imaging [6]. Deep learning also improves the effectiveness of interpreting various modalities of data due to its computational and automated feature extraction abilities. Most of the medical images are of higher dimensions (Medical Resonance Imaging, Optical Coherence Tomography and Computed tomography), and it is challenging as well as time-consuming to annotate such a huge array of images manually. Another such problems with medical images are, these are not widely available, and difficult to annotate. The devices through which the images are obtained may differ in quality. So, its less beneficial for deploying neural networks for this task, which requires a huge amount of data for extracting features and generalization. As a result, it directly restricts the development of systems for automated medical diagnosis. Conse-

quently, we need an end-to-end architecture that can synthesize medical images from one modality to another while extracting and learning representations of the manifold features. To alleviate this, we propose three such Generative architectures for two modalities of medical images. For the first one, we synthesize Fluroscein Angiography images from Fundus images. On the second task, we segment Ca2+ transient events from Ca2+ imaging to Binary Segmentation.

Fluorescein Angiography (FA) coupled with Retinal Funduscopy is the go to technique for diagnosing retinal vascular and pigment epithelial-choroidal diseases [7]. The technique utilizes the insertion of fluorescent dye which appears in the optic vein within 8-12 seconds and stays up to 10 minutes [8]. Although deemed safe, there have been numerous cases of mild to severe complications ocurring due to allergic responses to the dye [9]. Common side effects can range from nausea, vomiting, heart attack, to anaphylactic shock and even death [10]. Additionally, leakage of fluorescein in the intravenous space can also happen. Nevertheless, the concentration of fluorescein dye doesn't have any direct influence on the unfavorable effects mentioned above. [11].

Various automated methods have been introduced over the years, for diagnosing underlying conditions and diseases from fundus photos. Most of them consists of different image processing methods and machine learning models [10, 12–14]. Currently, there is no computationally inexpensive option for generating safe and reproducible fluorescein angiography images. Retinal fundoscopy is the only alternative for differential diagnosis that is readily accessible and financially viable. A combination of Optical coherence Tomography combined and image processing [15] can also be employed for the diagnosis of retinal disease but is too costly and not broadly available in poorer countries. As it stands, a cost-effective and quicker method is imperative for circumventing any inherent dangers associated with invasive fluorescein angiography.

Calcium imaging is used for understanding $Ca^2+$ signals in a variety of cellular

populations and their activities. The method captures Ca2+ signals which take the form of $Ca^2+$ waves that diffuse over cells or cell-to-cell and spread for distances of 100 µm or more and last for seconds or even minutes [16–21]. To gather in-depth information and better understand neural activities we need to semantically segment $Ca^2+$ activities, which helps to identify the source of calcium transients from raw data. Moreover, intracellular $Ca^2+$ behavior is an essential study to aid in understanding the multitude of complex and physiologically essential cellular activities. $Ca^2+$ ST-Maps is a popular technique through which $Ca^2+$ dynamics are analyzed and quantified in a range of both single-cell and intact tissues by plotting intracellular $Ca^2+$ transients as a function of space occupied over time [20, 22–25]. As a result, it can provide a platform to effectively quantify cellular $Ca^2+$ behaviors and its correlated parameters such as duration of activity, spatial spread of $Ca^2+$, frequency of event spikes, event angle, and event intensity [26–30].

Extracting quantifiable determinators of $Ca^2+$ transients from STMaps is challenging as this information is most often manually defined in traditional forms of analysis. Possibilities for user error are prevalent, which transmutes to inaccuracy when quantifying STMap $Ca^2+$ events. These fundamental issues arise from the requirement of user designation of single-pixel measurement lines that are assumed to be representative of the entire $Ca^2+$ event. This user-dependent manual process is highly variable, time-consuming, and labor-intensive.

To overcome this inconsistency, we need an efficient end-to-end architecture to achieve automated, fast, and accurately characterized Spatio-temporal $Ca^2+$ activity patterns as segmented images. Only a handful of works has been done using machine learning to segment these $Ca^2+$ events. An approach with the fast random-forest was proposed very recently by our group, for clustering and segmenting the $Ca^2+$ Spatio-temporal data [31]. However, the technique requires hand-picked features for

learning and semantically segment such activities, and also the precision is low for extracting quantifiable information. Therefore, for extracting feature-rich information from $Ca^2+$ imaging we need an end-to-end architecture that can generate accurate STMaps and execute quantification of activities with high precision. By automating this procedure can move towards a robust standardization and high-throughput analysis of cellular $Ca^2+$ dynamics.

In this work, we propose three novel conditional Generative Adversarial Network (GAN) for different medical imaging uses cases. For the task of synthesizing fluorescein angiograms from retinal fundus images, we propose two architectures, namely Fundus2Angio [32] and Attention2Angio [33]. And for segmenting Spatio-temporal events from Ca2+ transient events, which susceptible to noise, we propose CalciumGAN. The synthesis procedure is automated and does not require any human intervention. For quantitative evaluation of Retinal Image synthesis, we compare our proposed architecture with some recent state-of-the-art GAN architectures [34–37]. Moreover, we evaluate using Frechet inception Distance (FID) [38] and Kernel Inception Distance (KID) [39] to quantify image quality and calculate similarity with real angiograms. For qualitative comparison, expert ophthalmologists were directed to differentiate between fake angiograms from a random set of balanced real and fake angiograms over two trials. Results confirm that the angiograms produced by our proposed architecture are quite indistinguishable from genuine FA images. As for $Ca^2+$ events segmentation, we compare CalciumGAN with a recent method involving machine learning [31]. We also provide quantitative comparison using Mean-IOU and SSIM metrics for STMap segmentation. Our model outperforms these architectures for both use-cases in terms of qualitative and quantitative measurement. We believe, all of our architecture will be an important contribution for automating disease diagnosis by learning manifold features from multiple modalities of medical imaging

data.

# Chapter 2

# Background

Recently, there has been a surge of Generative adversarial networks(GAN) based applications ranging from image translation [40, 41], editing [42, 43] and image style transfer [35, 44]. GANs are categorized into two architectures i.e. Generator and Discriminator. Both of them are pitted against each other like using a min-max game to improve their learning. The purpose of the discriminator is to classify the input image as a real or fake image, while the generator tries to synthesize images that are as real as possible to fool the discriminator. GANs suffer from training instability, as is to tries to find Nash Equilibrium by reaching the global minima and thus halts the training process as a result. GANs can potentially extract and learn fine and coarse information from images by combining multiple architectures having multi-scale resolution [45, 46]. Such examples are wide-spread in both Conditional [47, 48] and Unconditional GAN settings [49, 50]. By incorporating multiple high-resolution architectures, they can learn distinct domain-specific features with high precision and robustness.

Emphasizing image to image translation, numerous prior work has been proposed, where they focused on architectural changes to acquire a higher quality re-

sult. To illustrate, pix2pixHD [35] utilized PatchGAN as a multi-scale discriminator to achieve better visual representation containing local and global information. On the other hand, U-GAT-IT [36], an unsupervised learning architecture, extracts local features and texture by incorporating AdaLIN (Adaptive Layer-Instance Normalization). While most architectures ensure high quality of images, Stargan-V2 [37], a style-transfer network, focuses on the domain-specific features. By doing so, they allow diverse and scalable image-to-image translation in multiple domains within a single model.

Cross-domain image translation has been employed for medical image inpainting by learning from incomplete information in standard statistical analysis [51–53], or to improve standard steps of examination such as image registration [54, 55], fusing information [56–58], image segmentation [59–62], image construction [63, 64] and disease diagnosis [65, 66]. Each of these techniques converts images of one modality to another interchangeably as medical resonance imaging (MRI), optical coherence tomography (OCT), Spectral-domain OCT, positron emission tomography (PET), and ultrasound imaging. In ophthalmology, generative networks have been employed for retinal vessel segmentation [67–69]. One such approach used sparsity regularization for detecting anomalies in optical coherence tomography [70]. Moreover, the GAN based retinal vessel segmentation architectures are derivative of DC-GAN [71], consisting of one generator and one discriminator with learning conducted by an adversarial loss. No techniques have been incorporated, to learn manifold features by incorporating coarse-to-fine generators or multi-scale discriminators for coarse and fine feature learning. On the other hand, an autoencoder based framework has been proposed for OCT angiography to OCT flow image generation [72]. For de-noising, a U-net based Auto-encoder has also shown good results for de-noising optical nerve head from OCT images [73]. However, no end-to-end system has been proposed, for

synthesizing fluorescein angiograms from retinal fundus images.

Calcium imaging for STMap analysis mainly builds upon existing image processing and mathematical modelling techniques [74–76]. Very recently, machine learning and deep learning approaches has been adopted for calcium imaging for STMap analysis. One such technique called Weka Segmentation incorporates Fast-random forest and selective feaure learning for $Ca^2+$ event segmentation [31]. However, it suffers from noise artifacts and false positive prediction due to periodic noise.

Most of the image-to-image translation models are either focused on domain level transformation or combining style and textures of two images. For instance, high-resolution images generated by U-GAT-IT [36] and Stargan-V2 [37] use attention modules to extract local features information and don't utilize perceptual loss. Whereas StyleGAN [77], pix2pixHD [35] emphasizes more on incorporate perceptual loss with different styles of target images. To adapt to these changes and for our particular medical imaging uses-case we propose an architecture, where we combine perceptual loss and multi-scale discriminator to retain global information like the shape of optic-disc, color, contrast, etc for retinal image synthesis and spread and duration of the calcium event. On the other hand, we utilize feature matching loss and introduce new multi-attention modules to retain local features like retinal venules, arteries, protein buildup, and microaneurysm for angiograms and also co-localization and boundary of calcium transient events. The visual representation and quantitative result prove that our proposed technique surpasses state-of-the-architectures and tricks, experts, to think they are authentic.

# Chapter 3

# Methodology

## 3.1 The Proposed Methodology

This thesis proposes three generative adversarial network (GAN) comprising of novel residual blocks for generators and discriminators. Moreover, the training includes perceptual, feature matching, and reconstruction loss to synthesize more vivid looking angiograms from retinal fundus images and ST-map segmentation from noisy images. First, we discuss about coarse and fine generators in section 3.1.1. Next, we elaborate our building blocks in Section 3.1.2, 3.1.3, 3.1.4. We then delve into the multi-scale discriminators and their interconnection with the generators to define the whole end-to-end pipeline for the generative network in sections 3.1.5. Ultimately, in section 3.1.6, we discuss the loss minimizing and maximizing function and loss weight distributions for different losses interrelated to each of the separate architecture that forms the proposed models.

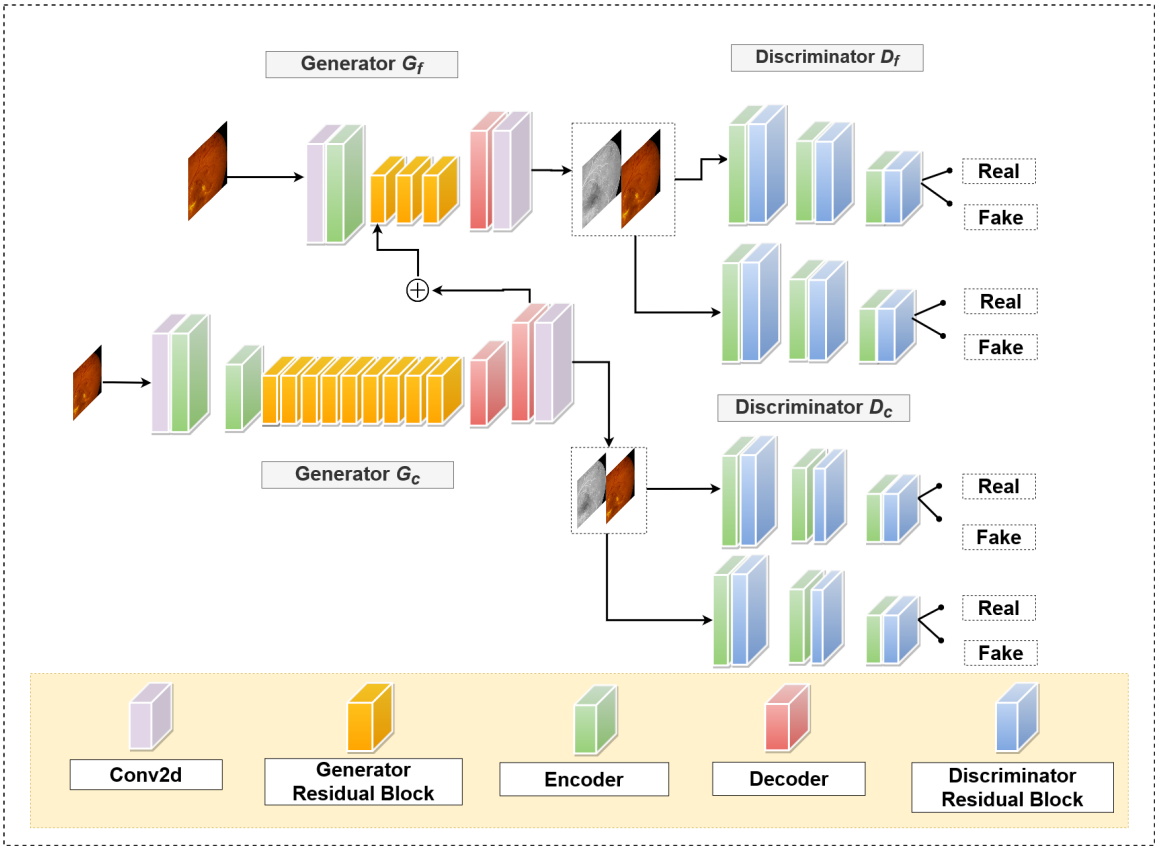


Figure 3.1: Proposed Generative Adversarial Network, Fundus2Angio, consisting of two Generators $G_{coarse}$, $G_{fine}$, and four discriminators $D1_{coarse}$, $D1_{fine}$, $D2_{fine}$, $D2_{coarse}$. The generators take Fundus image as input and outputs FA image. Whereas, the discriminators take both Fundus and FA images as input and outputs if the pairs are real or fake.

### 3.1.1 Coarse and Fine Generators

Coupling coarse-to-fine generator for image translation tasks results in very pristine and high quality images , as witnessed in recent architectures, such as pix2pixHD [35], SPADE [77], and Starganv2 [37]. We incorporate this technique in our architectures by using two generators ($G_{fine}$ and $G_{coarse}$), as illustrated in Fig. 3.1, Fig. 3.2 and Fig. 3.3. In Fig. 3.1, we visualized our first architecture Fundus2Angio, and in the consecutive ones in Fig. 3.2, we illustrated a newer version with some architectural changes called Attention2Angio. For retinal image synthesis, the generator $G_{fine}$ synthesizes smooth FA from fundus images by learning local information such as retinal venules, blood vessels, hemorrhages, exudates, and protein buildup. On the contrary, the generator $G_{coarse}$ tries to extract and preserve global information, such as the structures of the macula, optic disc, color intensity, contrast, and illumination, while producing less detailed angiograms. The same is true for our CalciumGAN, where $G_{fine}$ learns local information of the event's spread and duration, whereas $G_{coarse}$ learns the boundary between multiple events in the generated Spatio-temporal maps as illustrated in Fig. 3.3. In Fig. 3.1. the generators consist of multiple upsampling and downsampling blocks, residual blocks, and a feature fusion block between the fine and coarse generator. But there are inherent manifold features that can be extracted from the image. To address that we also incorporate the proposed attention block as illustrated in Fig. 3.2 and Fig. 3.3. $G_{fine}$ has an input dimension of $512 \times 512$ and produces outputs with the same resolution. Likewise, $G_{coarse}$ takes an image with half the resolution ($256 \times 256$) and synthesizes an image with the same size. Additionally, the $G_{coarse}$ outputs a feature vector of the size $256 \times 256 \times 64$ that is combined with one of the intermediate layers of $G_{fine}$ using the fusion operation. In the following sections, we elaborate on each of these blocks in detail.

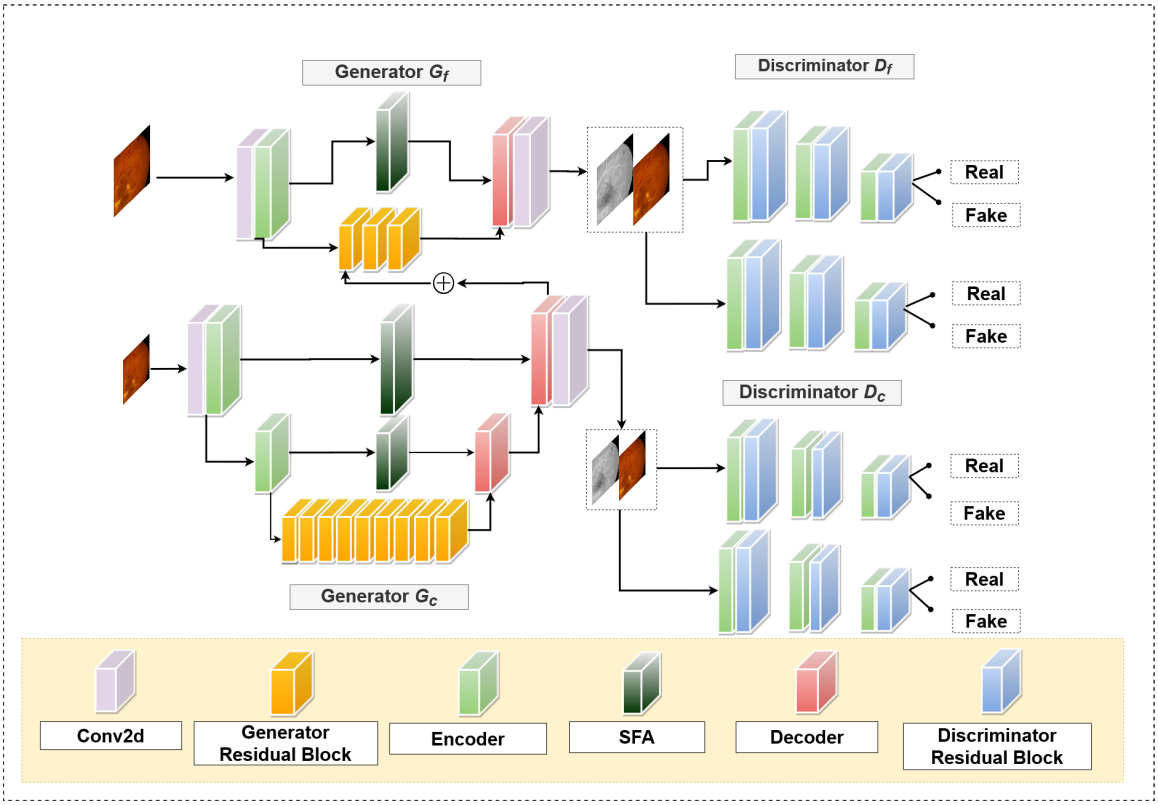


Figure 3.2: Proposed Generative Adversarial Network, Attention2Angio consisting of two Generators $G_{coarse}$, $G_{fine}$, and four discriminators $D1_{coarse}$, $D1_{fine}$, $D2_{fine}$, $D2_{coarse}$. The generators take Fundus image as input and outputs FA image. Whereas, the discriminators take both Fundus and FA images as input and outputs if the pairs are real or fake. The architecture also contains attention blocks in both the generators for extracting manifold features and synthesizing realistic images.

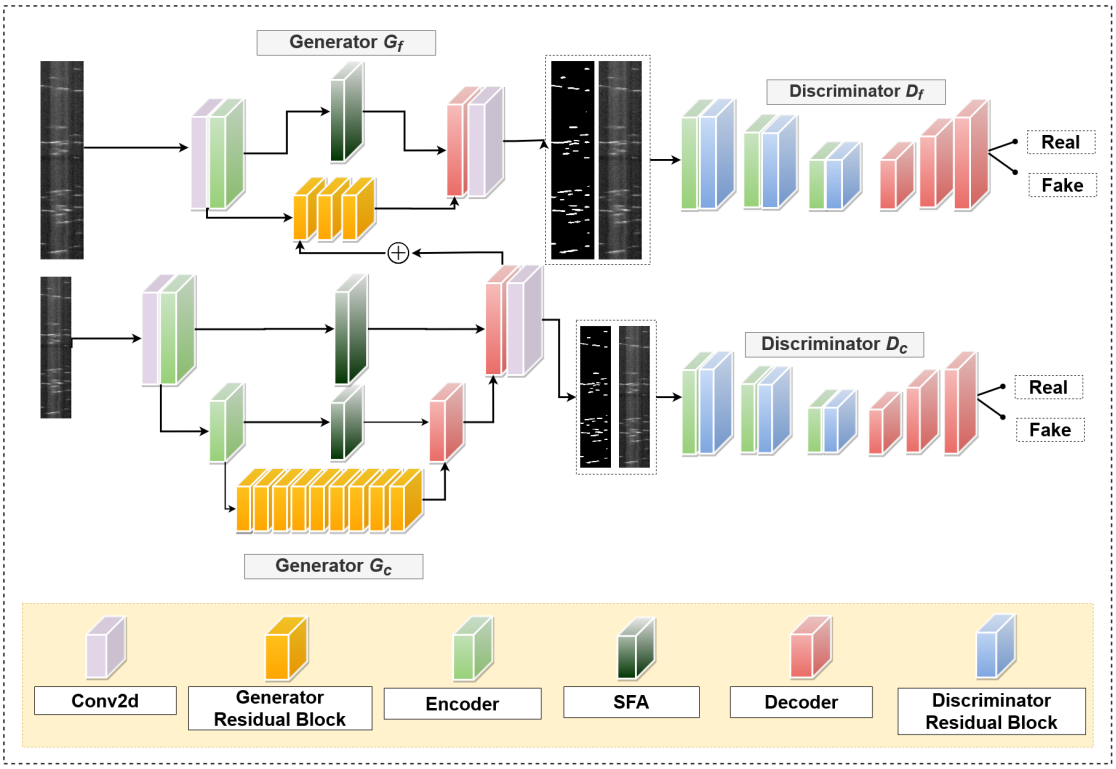


Figure 3.3: Proposed Generative Adversarial Network, Attention2Angio consisting of two Generators $G_{coarse}$, $G_{fine}$, and two discriminators $D_{coarse}$, $D_{fine}$. The generators take Calcium ST-maps as input and outputs segmented ST-maps. Whereas, the discriminators take both Calcium ST-maps and segmented ST-maps as input and outputs if the pairs are real or fake. The architecture also contains attention blocks in both the generators for extracting manifold features and synthesizing realistic images. The discriminators use auto-encoders for

### 3.1.2 Upsampling and Downsampling Blocks

For our first architecture, Fundus2Angio, we used Convolution layer for downsampling and Transposed Convolution layers for upsampling the features as illustrated in Fig. 3.1. Further investigation and experimentation confirmed that certain enhancements can help us to retain intrinsic features and tackle the problem of information loss due to spatial feature compression. Based on this we introduce two new blocks for upsampling and downsampling as can be seen in Fig. 3.4(b)(i) and Fig. 3.4(b)(ii). Both generators and discriminators of our Attention2Angio and CalciumGAN, incorporate the downsampling blocks and only the generators use decoder blocks for upsampling to get the desired feature maps and output. The downsampling block consists of a convolution layer followed by a batch-norm layer [78] and Leaky-ReLU activation function, as illustrated in Fig. 3.4(b)(i). In contrast, the upsampling block comprises of transposed convolution layer and successive batch-norm [78] and Leaky-ReLU activation 3.4(b)(ii). Interestingly, $G_{coarse}$ is downsampled twice ($\times 2$) using the downsampling. After successive residual blocks, the upsampling blocks are used twice to get the original spatial dimension. For $G_{fine}$, the downsampling is utilized once, and after the repetition of residual blocks, a single upsampling block is used to get the same spatial dimension of the output. We use a kernel size, $k = 3$ and stride, $s = 2$ for both of our convolution, and transposed convolution layers.

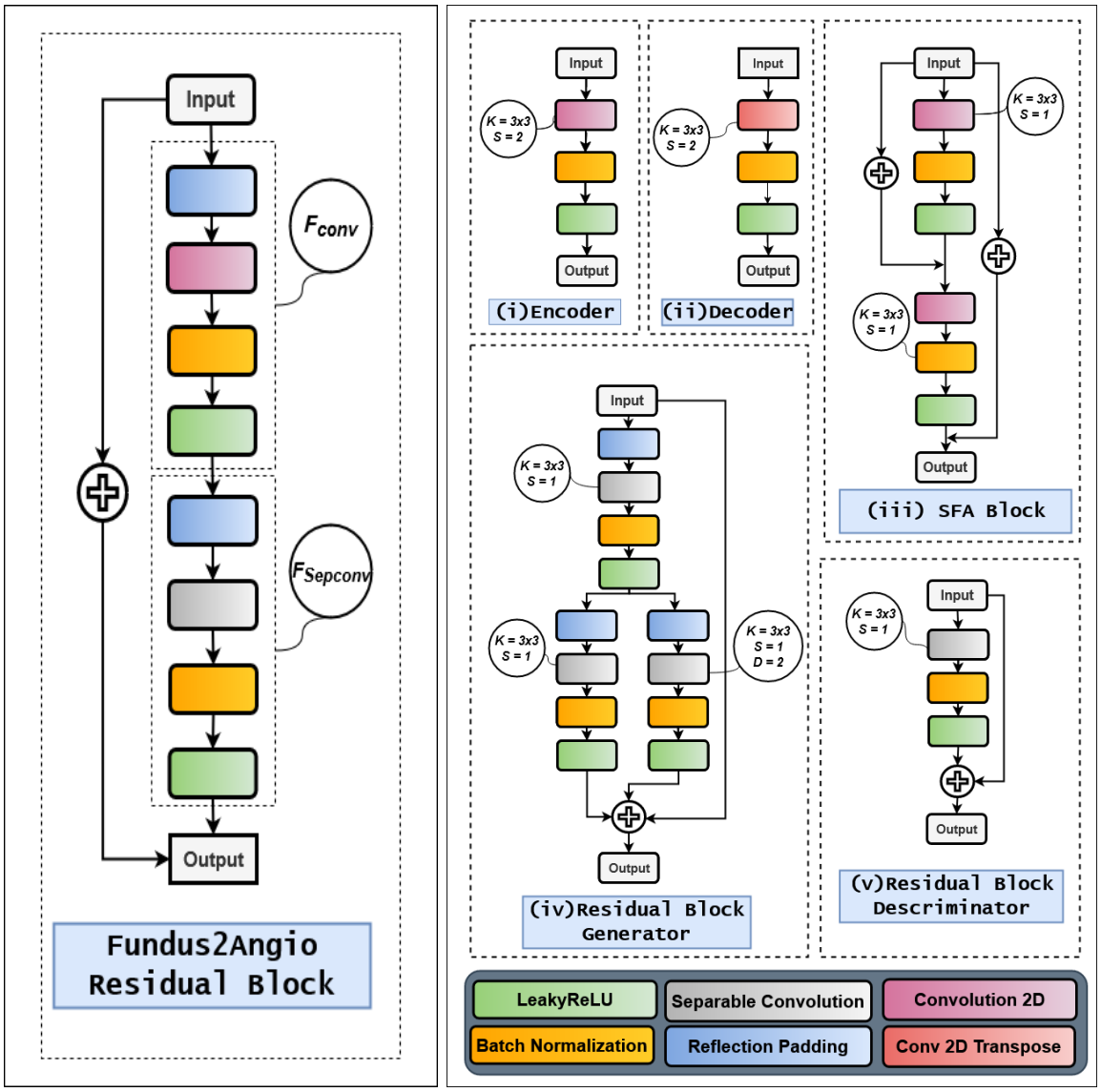


Figure 3.4: (a) Residual block for Fundus2Angio consisting of two residual units $F_{conv}$ and $F_{SepConv}$. First one cosists of Reflection padding, Convolution, Batch-Normalization and Leaky-ReLU layers. The second one has some layers except has Separable Convolution instead of Vanilla Convolution. (b) Individual blocks of our proposed Attention2Angio and CalciumGAN architecture consisting of (i) Encoder, (ii) Decoder, (iii) Attention block, (iv) Residual Block for Generator and (v) Residual Block for Discriminator where K stands for kernel size, S is for stride and D is for Dilation rate.

### 3.1.3 Novel Residual Blocks for Generator and Discriminator

Lately, residual blocks have become the standard for generative models accomplishing image-to-image translation, image inpainting, and semantic segmentation tasks [35, 77]. The fundamental design consists of a residual unit with two consecutive convolution layers and a skip connection that adds feature tensor of the input with the output. Regular convolution layers are computationally inefficient as opposed to separable convolution [79]. Separable convolution consists of a depth-wise convolution followed by a point-wise convolution. By doing so, it extracts and retains the depth and spatial information through the network. Recent studies show that combining separable convolutional layers with dilation allows for more robust feature extraction [15]. We incorporate this idea to design a novel residual block to retain both depth and spatial information, decrease computational complexity, and ensure efficient memory usage as illustrated in Fig. 3.4(a). As illustrated in Fig. 3.4(a), we replace the last convolution operation with a separable convolution. We also use Batch-normalization and Leaky-ReLU as post-activation mechanism after both convolution and separable Convolution layers. For better results, we incorporate reflection padding as opposed to zero-padding before each convolution operation. After carrying out more experimentation, we found out that certain features require a wider receptive field to understand neighboring feature information. To address this, we augment our residual block by having two branches of separable convolution layers with different dilation rates as illustrated in Fig. 3.4(b)(iv). The difference is that one branch consists of a separable convolution with a dilation rate of, $d = 1$ and the other with a dilation rate, $d = 2$. We use a kernel size, $k = 3$ and stride, $s = 1$ for all of our separable convolution layers. Each separable convolution is preceded by a Reflection padding layer and succeded by a Batch-Normalization and Leaky-ReLU activation layer. The skip connection and output of the two branches are all added together

to produce the final output. Moreover, we incorporate this technique to design a distinct novel residual block for our discriminators, as shown in Fig. 3.4(b)(v). In contrast, the residual block for the discriminator consists of a Separable convolution layer, followed by Batch-Normalization and Leaky-ReLU activation function. The separable convolution has a kernel size of $k = 3$ and stride, $s = 1$.

### 3.1.4 Attention block

Next, we elaborate on our proposed attention block, as illustrated in Fig. 3.4(b)(iii). The reason for using this block is, due to consecutive downsampling and upsampling layers there might be a loss of spatial information from the manifold feature space [80,81]. To retain those intrinsic features, we use the skip connections in combination with attention block for fusing feature information from the bottom layers of the network with the top layers of the network as illustrated in Fig. 3.2 and Fig. 3.3. The block consists of two successive residual units, Convolution, BatchNorm, and Leaky-ReLU layers. Both convolution layer has kernel size of $k = 3$ and stride, $s = 1$. Other than that, there are two skip connections one coming from the input and being added to the output of the first residual unit. The other one is coming from the input and summed with the output of the last residual unit. $G_{coarse}$ comprises of two attention block, coming out of the with two encoders and being added with the two decoders successively. In contrast, the $G_{fine}$ has only one attention block between the encoder and decoder.

### 3.1.5 Multi-scale Markovian Discriminators

GAN discriminators need to adjust to coarse and fine generated outputs for distinguishing between real and synthesized images. To solve this underlying issue, we

need a dense network with a huge amount of computable parameters. Alternatively, convolution with a wider receptive field can be utilized for extracting spatial information. This can easily lead to overfitting while training the model. To address this issue, we exploit the idea of using two Markovian discriminators, first introduced in a technique called PatchGAN [82]. The method consists of discriminators with variable sized input resolution and can help with the overall adversarial training of the architecture as observed in [35]

For Fundus2Angio and Attention2Angio, we use four discriminators that incorporates almost the same network structure but operate at two different resolutions as illustrated in Fig. 3.1 and Fig. 3.2. We organize the four discriminators into two sets, $D_{fine} = [D1_f, D2_f]$ and $D_{coarse} = [D1_c, D2_c]$ as illustrated in Fig. 3.1. We resize each of the coarse and fine angiograms and fundus with size $512 \times 512$ and $256 \times 256$ by a factor of 2 using the Lanczos filter [83]. $D2_f$ and $D2_c$ have a unique average pooling layer right after the input which resizes the resolution to $256 \times 256$ and $128 \times 128$. Other than that, all four discriminators have identical layers consisting of three repetitive encoder and residual block pairs (in Fig. 3.2(a) and Fig. 3.2(e). Lastly, convolution layer is used for getting spatial dimension of $64 \times 64$ and $32 \times 32$ for $D1_f, D2_f$ and $32 \times 32$ and $16 \times 16$ for $D1_c, D2_c$ as outputs.

The coarse discriminators one that learns feature at a lower resolution tries to convince the coarse generator to retain more global features such as the macula, spherical optic disc, appearance, and illumination. On the other hand, the fine discriminators dictate the fine generator to produce more detailed local features such as retinal vessels, arteries, exudates, etc. By doing this we fuse features from both generators while training them autonomously with their joined multi-scale discriminators.

For CalciumGAN, we use two discriminators, one for $G_c$ and one for $G_f$ as illustrated in Fig. 3.3. We also incorporate decoders to convert the discriminators into an

auto-encoder. Unlike the patch-gan discriminator in the previous two architectures (in Fig. 3.1 & Fig. 3.2), the auto-encoder dictates if each pixel rather than patches of the generated image is fake or real. This in turn helps with pixel-level segmentation task, as more discriminative features are retained throughout the architecture and results highly accurate spatio-temporal segmentation map.

### 3.1.6 Weighted Object Function and Adversarial Loss

With the given discriminators and generators, the objective function for our whole network can be formulated as Eq. 3.1. It's a multi-objective problem of maximizing the loss of the discriminators while diminishing the loss of the generators.

$$\min_{G_f, G_c} \max_{D_f, D_c} \mathcal{L}_{adv}(G_f, G_c, D_f, D_c) \tag{3.1}$$

For Fundus2Angio, we use LSGAN [84] for calculating the loss and training our conditional GAN. The objective function for our conditional GAN is given in Eq. 3.2.

$$\mathcal{L}_{adv}(G, D) = \mathbb{E}_{x,y}\big[\ (D(x, y) - 1)^2\big]\ + \mathbb{E}_{x}\big[\ (D(x, G(x)) + 1))^2\big] \tag{3.2}$$

where the discriminators are first trained on the real fundus, $x$ and real angiography image, $y$ and then trained on the the real fundus, $x$ and fake angiography image, $G(x)$. We start with training the discriminators $D_{fine}$ and $D_{coarse}$ for couple of iterations on random batches of images. Next, we train the $G_{coarse}$ while keeping the weights of the discriminators frozen. Following that, we train the the $G_{fine}$ on a batch of random samples in a similar fashion. We use Mean-Squared-Error (MSE) for calculating the individual loss of the generators as shown in Eq. 3.3.

$$\mathcal{L}_{rec}(G) = \mathbb{E}_{x,y}\|G(x) - y\|^2 \tag{3.3}$$

where, $\mathcal{L}_{rec}$ is the reconstruction loss for a real angiogram, $y$, given a generated angiogram, $G(x)$. We use this loss for both $G_{fine}$ and $G_{coarse}$ so that the model can generate high quality angiograms of different scales. Previous techniques have also exploited this idea of combining basic GAN objective with a MSE loss [85]. From Eq. 3.2 and 3.3 we can formulate our final objective function for Fundus2Angio as given in Eq. 3.4.

$$\begin{aligned}\min_{G_{fine},G_{coarse}} \max_{D_{fine},D_{coarse}} \mathcal{L}_{cGAN}(G_{fine}, G_{coarse}, D_{fine}, D_{coarse}) \\ +\lambda\big[\ \mathcal{L}_{L2}(G_{fine}) + \mathcal{L}_{L2}(G_{coarse})\big]\end{aligned} \tag{3.4}$$

Here, $\lambda$ dictates either to prioritize the discriminators or the generators. For our architecture, more weight is given to the reconstruction loss of the generators and thus we pick a large $\lambda$ value.

For Attention2Angio and CalciumGAN, we use Hinge-Loss [80, 86] for adversarial training as illustrated in Eq. 3.5 and Eq. 3.6. Effectively, all the fundus images and their corresponding angiogram pairs, and ST maps and its corresponding masks, are normalized to $[-1, 1]$. This in turn helps with widening the gap between the pixel intensities of the real and synthesized angio images. In Eq. 3.7 we add them and use $\lambda_{adv}$ as weight multiplier with the $\mathcal{L}_{adv}(G)$.

$$\mathcal{L}_{adv}(D) = -\mathbb{E}_{x,y}\big[\ \min(0, -1 + D(x, y))\big]\ - \mathbb{E}_{x}\big[\ \min(0, -1 - D(x, G(x)))\big] \tag{3.5}$$

$$\mathcal{L}_{adv}(G) = -\mathbb{E}_{x,y}\big[(D(G(x), y))\big] \tag{3.6}$$

$$\mathcal{L}_{adv}(G, D) = \mathcal{L}_{adv}(D) + \lambda_{adv}(\mathcal{L}_{adv}(G)) \tag{3.7}$$

Here, In Eq. 3.5 and Eq. 3.6 the discriminators are first trained on the real fundus, $x$ and real angiogram, $y$, and then trained on the real fundus, $x$ and synthesized angiogram, $G(x)$. We begin by batch-wise training the discriminators $D1_f, D2_f$, and $D1_c, D2_c$ for a couple of iterations on randomly sampled data. After that, we train the $G_c$ while keeping the weights of the discriminators frozen. In the same manner, we train the $G_f$ on a batch of random images while keeping weights of all the discriminators frozen.

The generators incorporate the reconstruction loss same as before as shown in Eq. 3.3. However, to extract perceptual information we use perceptual loss [87] as given in Eq. 3.8. By utilizing these losses we ensure the synthesized images contain more realistic color, contrast, and vascular structure. We also employ feature matching loss [35] with all our discriminators and as given in Eq. 3.9.

$$\mathcal{L}_{perc}(G) = \mathbb{E}_{x,y} \sum_{i=1}^{k} \frac{1}{M} \| F^i_{vgg}(y) - F^i_{vgg}(G(x)) \| \tag{3.8}$$

$$\mathcal{L}_{fm}(G, D_n) = \mathbb{E}_{x,y} \sum_{i=1}^{k} \frac{1}{N} \| D^i_n(x, y) - D^i_n(x, G(x)) \| \tag{3.9}$$

In Eq. 3.8, $\mathcal{L}_{perc}$ calculates the difference between real and fake angio features extracted by pushing both of successively in VGG19 architecture [88]. Lastly, Eq. 3.9 is calculated by taking the features from intermediate layers of the discriminator by first inserting the real and fake angiograms consecutively. Here, $M$ and $N$ stands for the number of feature layers extracted from VGG19 and the discriminators consecutively.

By incorporating Eq. 3.3, 3.7, 3.8 and 3.9 we can formulate our final objective function for Attention2Angio and CalciumGAN as given in Eq. 3.10 and Eq. 3.11.

$$\min_{G_f,G_c} \Big( \max_{D_f,D_c} (\mathcal{L}_{adv}(G_f, G_c, D_f, D_c)) + \lambda_{rec}\big[\, \mathcal{L}_{rec}(G_f, G_c)\big] + \lambda_{fm}\big[\, \mathcal{L}_{fm}(G_f, G_c, D_f, D_c)\big] + \lambda_{perc}\big[\, \mathcal{L}_{perc}(G_f, G_c)\big] \Big) \quad (3.10)$$

$$\min_{G_f,G_c} \Big( \max_{D_f,D_c} (\mathcal{L}_{adv}(G_f, G_c, D_f, D_c)) + \lambda_{rec}\big[\, \mathcal{L}_{rec}(G_f, G_c)\big] + \lambda_{fm}\big[\, \mathcal{L}_{fm}(G_f, G_c, D_f, D_c)\big] \Big) \quad (3.11)$$

Here, $\lambda_{adv}$, $\lambda_{rec}$, $\lambda_{perc}$ and $\lambda_{fm}$ signifies loss weighting that are multiplied with their respective losses. The loss weighting dictates which networks to prioritize while training. For our architecture, more weight is given to the $\mathcal{L}_{adv}(G)$, $\mathcal{L}_{rec}$, $\mathcal{L}_{perc}$, and thus we select bigger $\lambda$ values for those.

# Chapter 4

# Experiments

In the next section, we detail our model experiments and evaluate our architecture based on qualitative and quantitative metrics. First, we elaborate on the structuring and pre-processing of our dataset in Sec. 4.0.1. Then detail our hyper-parameter selection and tuning in Sec. 4.0.2. Next, we describe our adversarial training scheme in Sec. 4.0.3. Also, we compare our architecture with existing state-of-the-art generative models based on some qualitative evaluation metrics in Sec. 4.0.5. Lastly, in Sec. 4.0.4, we analyze the quantification done by experts, by distinguishing between real and synthesized angiograms.

### 4.0.1 Dataset

We use the fundus and angiography data-set provided in [89]. It consists of thirty image and twenty-nine pairs of the healthy and unhealthy fundus and angiogram images, collected from fifty-nine individual patients. Next, we clean the dataset by taking only seventeen pairs of images based on one-to-one alignment between the fundus and angiogram pairs. These image pairs are either accurately aligned or

almost aligned. The original image size is $576 \times 720$, but we take 50 overlapping crops of $512 \times 512$ sized samples from each. By doing so, we end up having 850 images in total for training. The fundus images are in RGB format, and angiograms are in a Gray-scale format. For testing, we take fourteen image pairs and crop four overlapping quadrants of the image to generate a test set of fifty-six test images.

For ST-map segmentation we train on 6 images with varied resolution taking $64 \times 64$ overlapping crops with stride 8, ending with 7316 images in total. Furthermore, we follow the same procedure to test on 6 images and end up having 3599 crops from them. The images are in Gray-scale format, and the segmentation are in binary format.

### 4.0.2 Hyper-parameter tuning

For adversarial training, we used lsgan [84] for Fundus2Angio and hinge loss for Attention2Angio and CalciumGAN [80, 86]. We picked $\lambda_{adv} = 10$ (Eq. 3.4) and $\lambda_{rec} = 10$, $\lambda_{perc} = 10$, $\lambda_{fm} = 1$ (Eq. 3.8). For optimizer, we used Adam [90], with learning rate $\alpha = 0.0002$, $\beta_1 = 0.5$ and $\beta_2 = 0.999$. We train with mini-batches with batch size, $b = 2$ for 200 epochs while training Fundus2Angio and Attention2Angio. Whereas for CalciumGAN we take a batch of $b = 64$ and train for 100 epochs. It took approximately 48 hours to train Fundus2Angio and Attention2Angio on NVIDIA P100 GPU. On the other hand, it took 12 hours to train on CalciumGAN on the same GPU.

### 4.0.3 Training procedure

In this section, we elaborate one of our model's (Attention2Angio) detailed algorithm provided in Algorithm 1. To train our model, we start by initializing all the hyper-

**Algorithm 1** Attention2AngioGAN training

**Input:** $x_i \epsilon X$, $y_i \epsilon Y$

**Output:** $G_f$, $G_C$

**Initialize hyper-parameters**:
$max\_epoch$, $b$, $max\ d\_iter$, $\omega_D^f$, $\omega_D^c$, $\omega_G^f$, $\omega_G^c$ $\alpha_D^f$, $\alpha_D^c$, $\alpha_G^f$, $\alpha_G^c$, $\beta_D^f$, $\beta_D^c$, $\beta_G^f$, $\beta_G^c$, $\lambda_{rec}$, $\lambda_{per}$, $\lambda_{fm}$, $\lambda_{adv}$

**for** $e = 0$ *to* $max\_epoch$ **do**

  Sample $x_f, x_c, y_f, y_c$, using batch-size $b$

  **for** $d\_iter = 0$ *to* $max\ d\_iter$ **do**

    $\mathcal{L}_{adv}(D) \leftarrow D_c(x_c, y_c),\ D_f(x_f, y_f)$

    $\mathcal{L}_{adv}(D) \leftarrow D_c(x_c, G_c(x_c)),\ D_f(x_f, G_F(x_f))$

    $\mathcal{L}_{adv}(G) \leftarrow G_c(x_c),\ G_F(x_f)$

    $\mathcal{L}_{adv}(G, D) \leftarrow \mathcal{L}_{adv}(D) + \lambda_{adv}(\mathcal{L}_{adv}(G))$

    $\omega_D^c \leftarrow \omega_D^c + Adam(D_c, G_c, \omega_D^c, \alpha_D^c, \beta_D^c)$

    $\omega_D^f \leftarrow \omega_D^f + Adam(D_f, G_f, \omega_D^f, \alpha_D^f, \beta_D^f)$

  **end for**

  **Freeze** $\omega_D^c, \omega_D^f$

  Sample $x_f, x_c, y_f, y_c$, using batch-size $b$

  $\mathcal{L}_{rec}(G_c) \leftarrow G(x_c), y_c,$

  $\mathcal{L}_{rec}(G_f) \leftarrow G(x_f), y_f$

  $\mathcal{L}_{perc}(G_c) \leftarrow F_{vgg}^c(y), F_{vgg}^c(G(x))$

  $\mathcal{L}_{perc}(G_f) \leftarrow F_{vgg}^f(y), F_{vgg}^f(G(x))$

  $\omega_G^c \leftarrow \omega_G^c + Adam(G_c, \omega_G^c, \alpha_G^c, \beta_G^c)$

  $\omega_G^f \leftarrow \omega_G^f + Adam(G_f, \omega_G^f, \alpha_G^f, \beta_G^f)$

  **Unfreeze** $\omega_D^c, \omega_D^f$

  $\mathcal{L}_{fm}(D_c) \leftarrow D_n^c(x_c, y_c), D_n^f(x_c, G(x_c))$

  $\mathcal{L}_{fm}(D_f) \leftarrow D_n^f(x_f, y_f), D_n^f(x_f, G(x_f))$

  $\omega_D^c \leftarrow \omega_D^c + Adam(D_c, G_c, \omega_D^c, \alpha_D^c, \beta_D^c)$

  $\omega_D^f \leftarrow \omega_D^f + Adam(D_f, G_f, \omega_D^f, \alpha_D^f, \beta_D^f)$

  **Freeze** $\omega_D^c, \omega_D^f$

  $\mathcal{L}_{adv}(D) \leftarrow D_c(x_c, y_c),\ D_f(x_f, y_f)$

  $\mathcal{L}_{adv}(D) \leftarrow D_c(x_c, G_c(x_c)),\ D_f(x_f, G_F(x_f))$

  $\mathcal{L}_{adv}(G) \leftarrow G_c(x_c),\ G_F(x_f)$

  $\mathcal{L}_{adv}(G, D) \leftarrow \mathcal{L}_{adv}(D) + \lambda_{adv}(\mathcal{L}_{adv}(G))$

  $\omega_D^c \leftarrow \omega_D^c + Adam(D_c, G_c, \omega_D^c, \alpha_D^c, \beta_D^c)$

  $\omega_D^f \leftarrow \omega_D^f + Adam(D_f, G_f, \omega_D^f, \alpha_D^f, \beta_D^f)$

  Save weights and snapshot of $G_f, G_c$

  $\mathcal{L} \leftarrow \mathcal{L}_{adv} + \lambda_{rec}(\mathcal{L}_{rec}) + \lambda_{fm}(\mathcal{L}_{fm}) + \lambda_{perc}(\mathcal{L}_{perc})$

**end for**

parameters. Next, we sample a batch of the real fundus and angiogram $x, y$ images. We train the real fundus and fake angiogram pairs with $D_f, D_c$. After that, we use $G_f, G_c$ to synthesize fake angiograms and use the real fundus and fake angiograms, $x, G(x)$ to train discriminators $D_f, D_c$. Following that, we calculate the adversarial loss, $\mathcal{L}_{adv}(D, G)$, and update the weights. We freeze the weights of the discriminators. Next, we train the generators. We calculate the $\mathcal{L}_{rec}(G)$, $\mathcal{L}_{perc}(G)$ losses, if reconstruction and perceptual loss are used in the training and update both generator's weights. Subsequently, we unfreeze both discriminators weights, calculate the feature matching loss $\mathcal{L}_{fm}(D)$ and update the discriminator weights. In the final stage, we freeze both discriminator's weights and jointly fine-tune all the discriminator and generators. We calculate the total loss by adding and multiplying with their relative weights. For testing, we save the snapshot of the model and its weights.

### 4.0.4 Qualitative Evaluation

For assessing the performance of our network, we used 14 test samples and cropped four quadrants of the image with a size of $512 \times 512$. We conducted two sets of tests to evaluate our networks. First, for estimating the accurate visual representation without transforming the image. Next, for global and local changes due to transformation and distortion of the image. By doing so, we measured the network's ability to adjust to structural changes to the vascular patterns and formation of the retinal subspace. We used the GNU Image Manipulation Program (GIMP) [91] for carrying out transformation and distortion on the images.

For the first experiment, we train Fundus2Angio, three variants of Attention2Angio, and three other state-of-the-art image-to-image translation architecture using the same number of epochs and batches of images. Next, we evaluate them using the same test sample. A side by side comparison of the results is illustrated in Fig. 4.1.

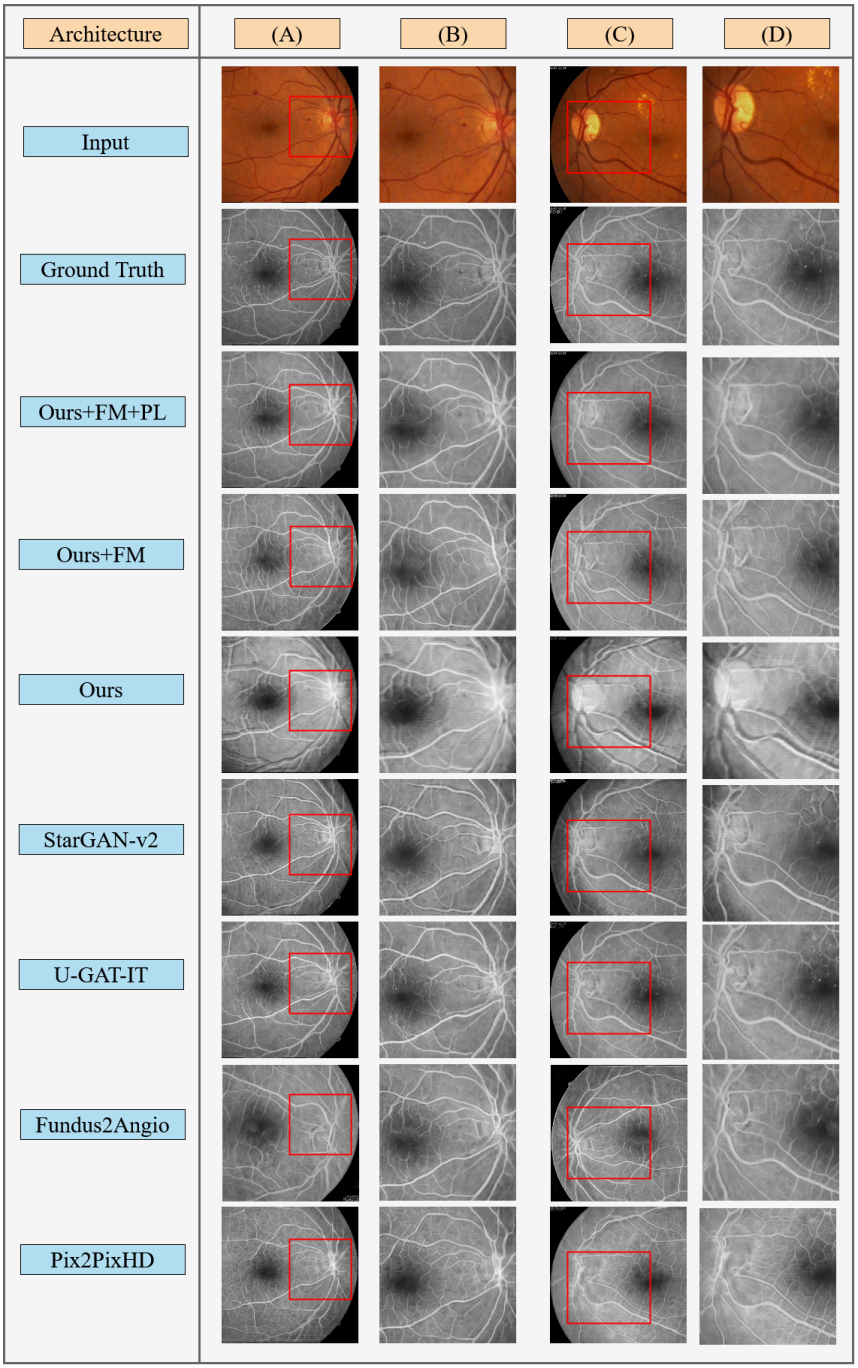


Figure 4.1: Comparativie results of different Angiograms generated using different state-of-the-art architectures. Column (A) and (C) represents two samples of real fundus, real angio and predicted angio images. Whereas column (B) and (D) represents the red rectangle box to show zoomed in local venular structures corresponding.

Column A & C in Fig. 4.1 shows the global changes while column B & D are zoomed-in to display local vascular structure and other local information. We use FM and PL to denote feature-matching and perceptual loss. By the looks of it, both of our Attention2Angio models using the FM loss (with or without PL) produces vivid and convincing results. On the other hand, the visual result of our network without FM and PL produces distorted local structure due to not learning contrast and color of the optic disc using perceptual loss. Fundus2Angio and StarGANv2 also produce impressive results. However, if seen in the closed-up versions in columns B & D, we can witness the right upper portion of the optic discs contains fewer blood vessels compared to our ones. U-GAT-IT and Pix2PixHD also fail to generate rich venules, exudates, and protein buildup.

In the second set of experiments, three transformations and two distortions were applied on the fundus images: 1) blurring to represent severe cataracts, 2) sharpening to represent dilated pupils, 3) signal noise to represent machine impedance during fundoscopy, 4) pinch, to visualize the pulled/pushed vascular formation, 5) whirl, for representing distortions caused by increased intraocular pressure (IOP). Improved robustness and adaption are represented by the generated angiograms similarity to the real FA image since these transformations and distortions may or may not affect the vascular structure of the retina. A side by side comparison of different architecture's predictions on these transformed images is illustrated in Fig. 4.2. As it can be observed from the image, our proposed Attention2ANgio architecture produces images very similar to the ground-truth (GT). On the other hand, Fundus2Angio prroduces better images than Pix2piXHD and U-GAT-IT.

In the case of **blurred** fundus images, Attention2ANgio with and without PL, is less affected compared to other state-of-the-art models, as seen in (row 6 to 9 of column 1) of Fig. 3.4. The venular and cellular structures are better conserved as

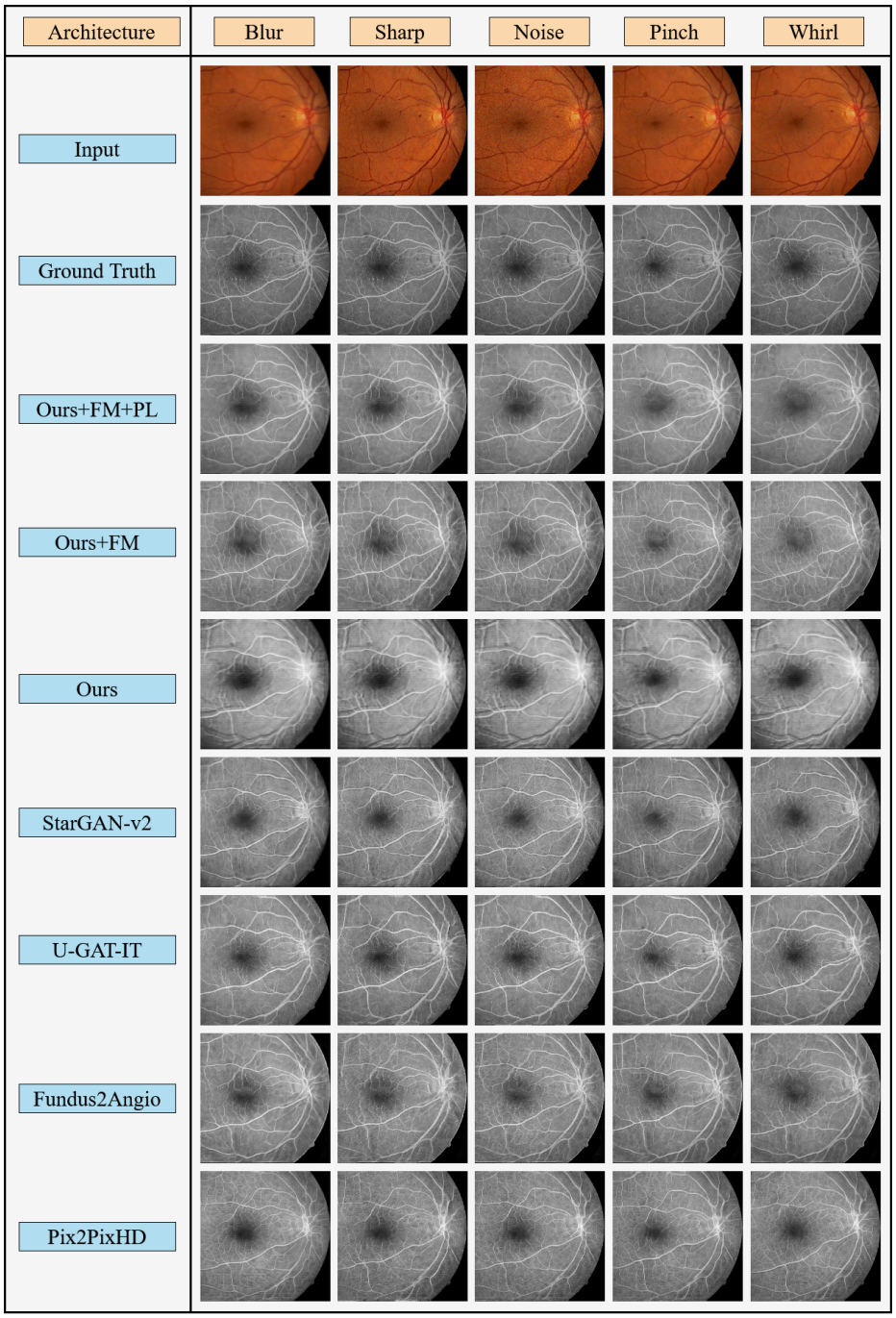


Figure 4.2: Angiogram generated from transformed and distorted Fundus images with natural changes, imaging errors and biological markers.

opposed to StarGANv2 and Pix2PixHD. For **sharpened** fundus, the angiogram produced by UGATIT and Fundus2Angio (row 7 and 8 of column 2) exhibits grainy artifacts around the blood vessels, which are not present in our model with and without PL. For **noisy** images, Attention2Angio's result with and without PL, is still unaffected with this pixel-level modification. However, all other state-of-the-art models (row 6 to 9 of column 3) fail to generate thin and small venular formations by failing to extract local features from the retinal subspace. On the contrary, Attention2Angio model without PL and FM produces jittery motion artifacts and high contrast around the border of the optic disc for all these transformations.

For distortions like **Pinch** and **Whirl**, our experimental result with and without perceptual loss shows the versatility and reproducibility of the proposed network to uncover the changes in vascular structure as seen in Fig. 3.4 (row 3 and 4 of column 4 and 5). Compared to our Attention2Angio, only StarGANv2 and U-GAT-IT maintains the flattening condition and manifestation of vascular changes but loses the overall smoothness in the process (row 6 to 7 of column 4 and 5). As seen in Fig. 3.4 Attention2Angio with and without PL network encodes the feature information of vessel structures and is much less affected by both kinds of contortion. The other architectures failed to generate microvessel structure due to IOP or vitreous changes as can be seen in Fig. 3.4.Contrarily, Attention2Angio model without any perceptual and feature-matching loss fails to encode this information and vascular changes. Consequently, For all kinds of transformation and distortion our Attention2Angio model with and without perceptual triumphs over existing state-of-the-art image-to-image translation models.

In the next experiment, we evaluate our CalciumGAN with the current state-of-the-art segmentation model for Spatio-temporal map segmentation based on WEKA segmentation model [31]. The model utilizes a fast-random-forest under the hood with

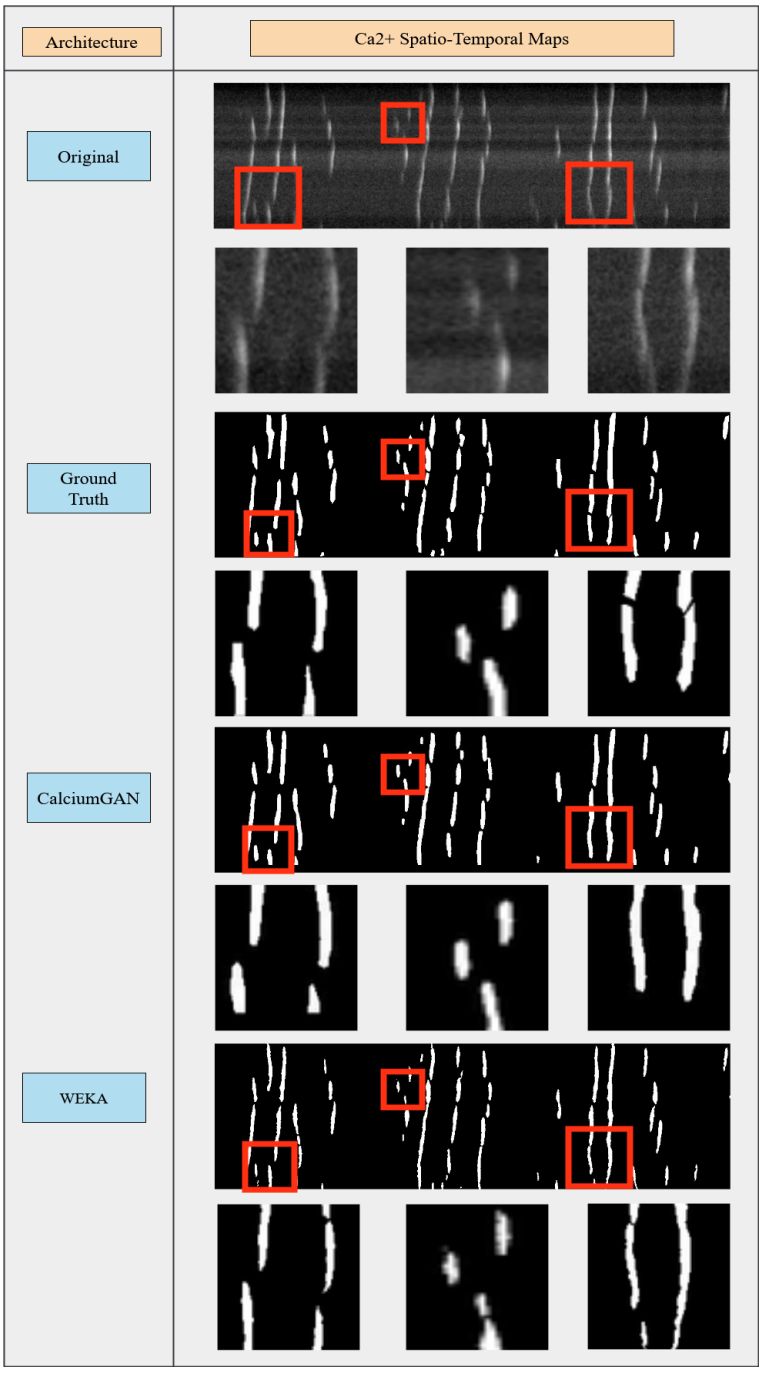


Figure 4.3: Spatio-temporal maps segmented using CalciumGan for quantification of Ca2+ transient events

hand-selected features to generate binary segmentation of the Ca2+ transient events. A side by side comparison of WEKA and our model's predictions is given in Fig. 4.3. It can be observed from the image, that the output of our proposed architecture produces accurate segmentation maps quite similar to the ground-truth (GT). We can easily infer the spread and duration of the events from our predicted output. Contrarily, WEKA produces broken patches of Ca2+ events and, the boundaries of the events are not uniformly spread. This, in turn, makes CalciumGAN, the better model out of the two.

Table 4.1: Test results for different architectures for Retinal Angiogram generation

| **Fréchet Inception Distance (FID)** | | | | | | |
|---|---|---|---|---|---|---|
| Architecture | Orig. | Noise | Blur | Sharp | Whirl | Pinch |
| **A2AGAN + PL[1]+ FM[2]** | 24.6 | 21.6 (3.0↓) | 30.0 (5.4↑) | 25.6 (1.0↑) | 40.0 (15.4↑) | 24.9 (0.3↑) |
| **A2AGAN + FM[2]** | **20.7** | **20.8** (0.1↑) | **23.5** (2.8↑) | **24.9** (4.2↑) | **27.8** (7.1↑) | **19.5** (1.2↓) |
| **A2AGAN** | 47.5 | 43.1 (4.4↓) | 49.8 (2.3↑) | 50.5 (3.5↑) | 61.9 (14.5↑) | 46.7 (0.8↓) |
| StarGAN-v2 [37] | 27.7 | 35.1 (7.4↑) | 32.6 (4.9↑) | 27.4 (0.3↓) | 32.7 (5.0↑) | 26.7 (1.0↓) |
| U-GAT-IT [36] | 24.5 | 26.0 (1.5↑) | 30.4 (5.9↑) | 26.8 (2.3↑) | 33.0 (9.5↑) | 29.1 (4.6↑) |
| **Fundus2Angio** | 30.3 | 41.5 (11.2↑) | 32.3 (2.0↑) | 34.3 (4.0↑) | 38.2 (7.9↑) | 33.1 (2.8↑) |
| Pix2PixHD [35] | 42.8 | 53.0 (10.2↑) | 43.7 (1.1↑) | 47.5 (4.7↑) | 45.9 (3.1↑) | 39.2 (3.6↓) |
| **Kernel Inception Distance (KID)** | | | | | | |
| Architecture | Orig. | Noise | Blur | Sharp | Whirl | Pinch |
| **A2AGAN + PL[1]+ FM[2]** | **0.00087** | **0.05045** | **0.00235** | **0.05162** | 0.05390 | **0.04575** |
| **A2AGAN + FM[2]** | 0.00392 | 0.05390 | 0.00505 | 0.05301 | 0.05657 | 0.05341 |
| **A2AGAN** | 0.00595 | 0.05237 | 0.00617 | 0.05298 | 0.05613 | 0.05419 |
| StarGAN-v2 [37] | 0.00118 | 0.05274 | **0.00235** | 0.05331 | 0.05539 | 0.05271 |
| U-GAT-IT [36] | 0.00131 | 0.05610 | 0.00278 | 0.05533 | 0.05815 | 0.05719 |
| **Fundus2Angio** | 0.00184 | 0.05328 | 0.00272 | 0.05267 | **0.05278** | 0.04985 |
| Pix2PixHD [35] | 0.00258 | 0.05613 | 0.00254 | 0.05788 | 0.06029 | 0.05838 |

[1] PL = Perceptual Loss; FM = Feature-Matching Loss
[2] FID: Lower is better; KID: Lower is better

### 4.0.5 Quantitative Evaluations

For quantitative evaluation, we performed two experiments. In the first experiment we use the Fréchet inception distance (FID) [38] and Kernel Inception distance (KID) [39] which has been used to evaluate similar style-transfer GANs [36, 37]. We computed the FID and KID scores for different architectures on the generated FA image and original angiogram, including the five transformations and distortions. The results

Table 4.2: Results of Qualitative with Undisclosed Portion of Fake/Real Experiment on Retinal Angiograms

| | | Results | | Average | | |
|---|---|---|---|---|---|---|
| Architectures | | Correct | Incorrect | Missed[1] | Found[1] | Precision[2] |
| A2AGAN + FM + PL | Fake | 10% | 90% | 55% | 45% | **47.1%** |
| | Real | 80% | 20% | | | |
| A2AGAN + FM | Fake | 12% | 88% | 53% | 47% | **48.2%** |
| | Real | 82% | 18% | | | |
| Fundus2Angio | Fake | 15% | 85% | 53% | 48% | **48.5%** |
| | Real | 80% | 20% | | | |

[1] Missed higher is better; Found lower is better
[2] Precision Lower is better

are reported in Table. 4.1. It should be noted that lower FID and KID score means better results.

From Table. 4.1, out of our four networks, the best FID is achieved for Attention2AngioGAN without PL. And it achieves the lowest scores among out of all other architecture, for both with and without distortions. For KID, Attention2Angio model with PL achieves the lowest score for four out of five types of distortions. Fundus2Angio scores lower KID for distorted images using whirl. Other than that, StarGAN-v2 achieves the same score as our network having a KID of 0.00235.

In the next experiment, we assess the quality of the synthesized angios by asking three expert ophthalmologists to identify fake angios among a collection of 50 balanced (50%, 50%) and randomly set of mixed angiograms. For this experiment, the exact number of fake and real images was not known by the experts. By not disclosing this information we tried to evaluate following criterion: 1) Correct fake and real angios found by the experts, where lower is better, 2) Incorrect fake and real angios missed by the experts, where higher is better and 3) The average precision representing the effective the identification is by the experts, where lower is better. The detailed results are shown in Table 4.2.

Table 4.3: Test results for Spatio-temporal map segmentation

| Method | Mean-IOU | SSIM |
|---|---|---|
| Weka | 78.25% | 88.17% |
| **CalciumGAN** | 86.63% | 95.66% |

As it can be seen from Table 4.2, experts assigned 90% and of 88% of the fake angiograms as real, for images generated by two of our Attention2Angio models. Whereas for Fundus2Angio they assinged 85% of the fake angiogram as real. The result also shows that experts had difficulty in identifying fake images, while they easily identified real angiograms with 80% and 82% certainty for Attention2Angio and 80% for Fundus2Angio. On average, the experts misclassified 55% and 53% of all images for Attention2Angio models consecutively and 53% for Fundus2Angio model. The average precision diagnosis of the experts are 47.1% and 48.2% for Attention2Angio and 48.5% for Fundus2Angio. Consequently, our model Attention2Angio with feature-matching and perceptual loss achieves the best result by fooling the experts to identify fake angios as real.

Additionally, for Spatio-temporal generation we use two other metrics for segmenting calcium transient event duration and spread, namely Mean-IOU (Jaccard Similarity Coefficient) and Structural Similarity Index [92]. We chose Mean-IOU because its the gold standard for measuring segmentation results for many Semantic Segmentation Challenges such as Pascal-VOC2012 [93], MS-COCO [94], Cityscapes [95]. Contrarily, SSIM is a standard metric for evaluating GANs for image-to-image translation tasks. As illustrated in the Table. 4.3, CalciumGAN outperforms Weka [31] which is the state-of-the-art atchitecture, in terms of Mean-IOU, and SSIM, the two main metrics for this task.

# Chapter 5

# Conclusion and Future Work

To conclude, we proposed three new image-to-image translation architecture called Fundus2Angio, Attention2AngioGAN and CalciumGAN. The first two architecture synthesizes high quality and vivid looking angiograms from fundus images without any expert intervention. Additionally, we illustrated its robustness, and flexibility by producing high-quality angiograms from transformed and distorted images, which imitates biological markers seen in real fundus images. As a result, the two proposed networks can be efficiently employed to generate precise FA images of patients developing disease overtime. This is best suited for disease progression monitoring to predict the development of diseases. The third architecture synthesizes calcium STMap segmentation without manual intensive annotations to overcome limitations of previous techniques. Moreover, we showed both qualitative side-by-side comparison and quantitative evaluation with two standard metrics for semantic segmentation task. Hence, this architecture can be used for automatic segmentation of Ca2+ STMaps and quantification of single Ca2+ event parameters. In future, we hope to extend this work to other retinal and calcium imaging modalities.

# Bibliography

[1] H. R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. M. Summers, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1170–1181, 2015.

[2] A. Teramoto, H. Fujita, O. Yamamuro, and T. Tamaki, "Automated detection of pulmonary nodules in pet/ct images: Ensemble false-positive reduction using a convolutional neural network technique," *Medical physics*, vol. 43, no. 6Part1, pp. 2821–2827, 2016.

[3] Y. Cai, M. Landis, D. T. Laidley, A. Kornecki, A. Lum, and S. Li, "Multi-modal vertebrae recognition using transformed deep convolution network," *Computerized medical imaging and graphics*, vol. 51, pp. 11–19, 2016.

[4] B. D. de Vos, J. M. Wolterink, P. A. de Jong, M. A. Viergever, and I. Išgum, "2d image classification for 3d anatomy localization: employing deep convolutional neural networks," in *Medical imaging 2016: Image processing*, vol. 9784. International Society for Optics and Photonics, 2016, p. 97841Y.

[5] C. Becker, R. Rigamonti, V. Lepetit, and P. Fua, "Supervised feature learning for curvilinear structure segmentation," in *International conference on medical*

*image computing and computer-assisted intervention*. Springer, 2013, pp. 526–533.

[6] Y. Ganin and V. Lempitsky, "$n^4$-fields: Neural network nearest neighbor fields for image transforms," in *Asian Conference on Computer Vision*. Springer, 2014, pp. 536–551.

[7] V. S. Mary, E. B. Rajsingh, and G. R. Naik, "Retinal fundus image analysis for diagnosis of glaucoma: a comprehensive survey," *IEEE Access*, 2016.

[8] N. Mandava, E. Reichel, D. Guyer *et al.*, "Fluorescein and icg angiography," *St Louis: Mosby*, vol. 106, pp. 800–808, 2004.

[9] K. Brockow and M. Sánchez-Borges, "Hypersensitivity to contrast media and dyes," *Immunology and Allergy Clinics*, vol. 34, no. 3, pp. 547–564, 2014.

[10] R. P. C. Lira, C. L. d. A. Oliveira, M. V. R. B. Marques, A. R. Silva, and C. d. C. Pessoa, "Adverse reactions of fluorescein angiography: a prospective study," *Arquivos brasileiros de oftalmologia*, vol. 70, no. 4, pp. 615–618, 2007.

[11] L. A. Yannuzzi, K. T. Rohrer, L. J. Tindel, R. S. Sobel, M. A. Costanza, W. Shields, and E. Zang, "Fluorescein angiography complication survey," *Ophthalmology*, vol. 93, no. 5, pp. 611–617, 1986.

[12] N. Gurudath, M. Celenk, and H. B. Riley, "Machine learning identification of diabetic retinopathy from fundus images," in *2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. IEEE, 2014, pp. 1–7.

[13] H. Fu, J. Cheng, Y. Xu, C. Zhang, D. W. K. Wong, J. Liu, and X. Cao, "Disc-aware ensemble network for glaucoma screening from fundus image," *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2493–2501, 2018.

[14] R. Poplin, A. V. Varadarajan, K. Blumer, Y. Liu, M. V. McConnell, G. S. Corrado, L. Peng, and D. R. Webster, "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning," *Nature Biomedical Engineering*, vol. 2, no. 3, p. 158, 2018.

[15] S. A. Kamran, S. Saha, A. S. Sabbir, and A. Tavakkoli, "Optic-net: A novel convolutional neural network for diagnosis of retinal diseases from optical tomography images," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 2019, pp. 964–971.

[16] F.-X. Boittin, N. Macrez, G. Halet, and J. Mironneau, "Norepinephrine-induced ca2+ waves depend on insp3 and ryanodine receptor activation in vascular myocytes," *American Journal of Physiology-Cell Physiology*, vol. 277, no. 1, pp. C139–C151, 1999.

[17] S. A. Baker, G. W. Hennig, A. K. Salter, M. Kurahashi, S. M. Ward, and K. M. Sanders, "Distribution and ca2+ signalling of fibroblast-like (pdgfr$\alpha$+) cells in the murine gastric fundus," *The Journal of physiology*, vol. 591, no. 24, pp. 6193–6208, 2013.

[18] G. W. Hennig, N. J. Spencer, S. Jokela-willis, P. O. Bayguinov, H.-t. Lee, L. A. Ritchie, S. M. Ward, T. K. Smith, and K. M. Sanders, "Icc-my coordinate smooth muscle electrical and mechanical activity in the murine small intestine," *Neurogastroenterology & Motility*, vol. 22, no. 5, pp. e138–e151, 2010.

[19] S. V. Straub, D. R. Giovannucci, and D. I. Yule, "Calcium wave propagation in pancreatic acinar cells: functional interaction of inositol 1, 4, 5-trisphosphate receptors, ryanodine receptors, and mitochondria," *The Journal of general physiology*, vol. 116, no. 4, pp. 547–560, 2000.

[20] H. Cheng, M. R. Lederer, W. Lederer, and M. Cannell, "Calcium sparks and [ca2+] i waves in cardiac myocytes," *American Journal of Physiology-Cell Physiology*, vol. 270, no. 1, pp. C148–C159, 1996.

[21] B. T. Drumm, R. J. Large, M. A. Hollywood, K. D. Thornbury, S. A. Baker, B. J. Harvey, N. G. McHale, and G. P. Sergeant, "The role of ca2+ influx in spontaneous ca2+ wave propagation in interstitial cells of cajal from the rabbit urethra," *The Journal of physiology*, vol. 593, no. 15, pp. 3333–3350, 2015.

[22] H.-T. Lee, G. W. Hennig, K. J. Park, P. O. Bayguinov, S. M. Ward, K. M. Sanders, and T. K. Smith, "Heterogeneities in icc ca2+ activity within canine large intestine," *Gastroenterology*, vol. 136, no. 7, pp. 2226–2236, 2009.

[23] M. A. Colman, C. Pinali, A. W. Trafford, H. Zhang, and A. Kitmitto, "A computational model of spatio-temporal cardiac intracellular calcium handling with realistic structure and spatial flux distribution from sarcoplasmic reticulum and t-tubule reconstructions," *PLoS computational biology*, vol. 13, no. 8, p. e1005714, 2017.

[24] C. J. Roome and B. Kuhn, "Simultaneous dendritic voltage and calcium imaging and somatic recording from purkinje neurons in awake mice," *Nature communications*, vol. 9, no. 1, pp. 1–14, 2018.

[25] R. Waadt, M. Krebs, J. Kudla, and K. Schumacher, "Multiparameter imaging of calcium and abscisic acid and high-resolution quantitative calcium measurements using r-geco1-mturquoise in arabidopsis," *New Phytologist*, vol. 216, no. 1, pp. 303–320, 2017.

[26] S. A. Baker, B. T. Drumm, C. A. Cobine, K. D. Keef, and K. M. Sanders, "Inhibitory neural regulation of the ca2+ transients in intramuscular interstitial cells of cajal in the small intestine," *Frontiers in physiology*, vol. 9, p. 328, 2018.

[27] B. T. Drumm, G. P. Sergeant, M. A. Hollywood, K. D. Thornbury, N. G. McHale, and B. J. Harvey, "The role of camp dependent protein kinase in modulating spontaneous intracellular ca2+ waves in interstitial cells of cajal from the rabbit urethra," *Cell Calcium*, vol. 56, no. 3, pp. 181–187, 2014.

[28] S. Fedigan, E. Bradley, T. Webb, R. J. Large, M. A. Hollywood, K. D. Thornbury, N. G. McHale, and G. P. Sergeant, "Effects of new-generation tmem16a inhibitors on calcium-activated chloride currents in rabbit urethral interstitial cells of cajal," *Pflügers Archiv-European Journal of Physiology*, vol. 469, no. 11, pp. 1443–1455, 2017.

[29] M. Sancho, E. Bradley, A. Garcia-Pascual, D. Triguero, K. D. Thornbury, M. A. Hollywood, and G. P. Sergeant, "Involvement of cyclic nucleotide-gated channels in spontaneous activity generated in isolated interstitial cells of cajal from the rabbit urethra," *European Journal of Pharmacology*, vol. 814, pp. 216–225, 2017.

[30] G. P. Sergeant, L. Johnston, N. G. McHale, K. Thornbury, and M. Hollywood, "Activation of the cgmp/pkg pathway inhibits electrical activity in rabbit urethral interstitial cells of cajal by reducing the spatial spread of ca2+ waves," *The Journal of physiology*, vol. 574, no. 1, pp. 167–181, 2006.

[31] W. A. Leigh, G. Del Valle, S. A. Kamran, B. T. Drumm, A. Tavakkoli, K. M. Sanders, and S. A. Baker, "A high throughput machine-learning driven analysis of ca2+ spatio-temporal maps," *Cell Calcium*, vol. 91, p. 102260, 2020.

[32] S. A. Kamran, K. F. Hossain, A. Tavakkoli, and S. L. Zuckerbrod, "Fundus2angio: A novel conditional gan architecture for generating fluorescein angiography images from retinal fundus photography," in *International Symposium on Visual Computing*. Springer, 2020.

[33] ——, "Attention2angiogan: Synthesizing fluorescein angiography from retinal fundus images using generative adversarial networks," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2020.

[34] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[35] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8798–8807.

[36] J. Kim, M. Kim, H. Kang, and K. Lee, "U-gat-it: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation," *arXiv preprint arXiv:1907.10830*, 2019.

[37] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "Stargan v2: Diverse image synthesis for multiple domains," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8188–8197.

[38] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in neural information processing systems*, 2017, pp. 6626–6637.

[39] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton, "Demystifying mmd gans," *arXiv preprint arXiv:1801.01401*, 2018.

[40] W. Chen and J. Hays, "Sketchygan: Towards diverse and realistic sketch to image synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9416–9425.

[41] P. Sangkloy, J. Lu, C. Fang, F. Yu, and J. Hays, "Scribbler: Controlling deep image synthesis with sketch and color," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5400–5409.

[42] J.-Y. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros, "Generative visual manipulation on the natural image manifold," in *European Conference on Computer Vision*. Springer, 2016, pp. 597–613.

[43] T. Dekel, C. Gan, D. Krishnan, C. Liu, and W. T. Freeman, "Sparse, smart contours to represent and edit images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3511–3520.

[44] W. Xian, P. Sangkloy, V. Agrawal, A. Raj, J. Lu, C. Fang, F. Yu, and J. Hays, "Texturegan: Controlling deep image synthesis with texture patches," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8456–8465.

[45] P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on communications*, vol. 31, no. 4, pp. 532–540, 1983.

[46] M. Brown, D. G. Lowe *et al.*, "Recognising panoramas." in *ICCV*, vol. 3, 2003, p. 1218.

[47] X. Huang, Y. Li, O. Poursaeed, J. Hopcroft, and S. Belongie, "Stacked generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 5077–5086.

[48] E. L. Denton, S. Chintala, R. Fergus *et al.*, "Deep generative image models using a laplacian pyramid of adversarial networks," in *Advances in neural information processing systems*, 2015, pp. 1486–1494.

[49] Q. Chen and V. Koltun, "Photographic image synthesis with cascaded refinement networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1511–1520.

[50] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 5907–5915.

[51] G. Van Tulder and M. de Bruijne, "Why does synthesized data improve multi-sequence classification?" in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 531–538.

[52] A. V. Dalca, K. L. Bouman, W. T. Freeman, N. S. Rost, M. R. Sabuncu, and P. Golland, "Medical image imputation from image collections," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 504–514, 2018.

[53] K. Eilertsen, L. Nilsen Tor Arne Vestad, O. Geier, and A. Skretting, "A simulation of mri based dose calculations on the basis of radiotherapy planning ct images," *Acta Oncologica*, vol. 47, no. 7, pp. 1294–1302, 2008.

[54] J. E. Iglesias, E. Konukoglu, D. Zikic, B. Glocker, K. Van Leemput, and B. Fischl, "Is synthesizing mri contrast useful for inter-modality analysis?" in *International*

*Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 631–638.

[55] C. Wang, G. Yang, G. Papanastasiou, S. A. Tsaftaris, D. E. Newby, C. Gray, G. Macnaught, and T. J. MacGillivray, "Dicyc: Gan-based deformation invariant cross-domain information fusion for medical image synthesis," *Information Fusion*, 2020.

[56] J. Du, W. Li, K. Lu, and B. Xiao, "An overview of multi-modal medical image fusion," *Neurocomputing*, vol. 215, pp. 3–20, 2016.

[57] Q. He, X. Li, D. N. Kim, X. Jia, X. Gu, X. Zhen, and L. Zhou, "Feasibility study of a multi-criteria decision-making based hierarchical model for multi-modality feature and multi-classifier fusion: Applications in medical prognosis prediction," *Information Fusion*, vol. 55, pp. 207–219, 2020.

[58] K. Wang, M. Zheng, H. Wei, G. Qi, and Y. Li, "Multi-modality medical image fusion using convolutional neural network and contrast pyramid," *Sensors*, vol. 20, no. 8, p. 2169, 2020.

[59] S. Roy, A. Carass, and J. Prince, "A compressed sensing approach for mr tissue contrast synthesis," in *Biennial International Conference on Information Processing in Medical Imaging*. Springer, 2011, pp. 371–383.

[60] A. Chartsias, T. Joyce, G. Papanastasiou, S. Semple, M. Williams, D. Newby, R. Dharmakumar, and S. A. Tsaftaris, "Factorised spatial representation learning: Application in semi-supervised myocardial segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 490–498.

[61] L. Li, X. Zhao, W. Lu, and S. Tan, "Deep learning for variational multimodality tumor segmentation in pet/ct," *Neurocomputing*, vol. 392, pp. 277–295, 2020.

[62] J. Liu, C. Shen, T. Liu, N. Aguilera, and J. Tam, "Active appearance model induced generative adversarial network for controlled data augmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 201–208.

[63] N. Cordier, H. Delingette, M. Lê, and N. Ayache, "Extended modality propagation: image synthesis of pathological cases," *IEEE transactions on medical imaging*, vol. 35, no. 12, pp. 2598–2608, 2016.

[64] O. Commowick, S. K. Warfield, and G. Malandain, "Using frankenstein's creature paradigm to build a patient specific atlas," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2009, pp. 993–1000.

[65] R. Li, W. Zhang, H.-I. Suk, L. Wang, J. Li, D. Shen, and S. Ji, "Deep learning based imaging data completion for improved brain disease diagnosis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2014, pp. 305–312.

[66] T. Zhou, K.-H. Thung, M. Liu, F. Shi, C. Zhang, and D. Shen, "Multi-modal latent space inducing ensemble svm classifier for early dementia diagnosis with neuroimaging data," *Medical Image Analysis*, vol. 60, p. 101630, 2020.

[67] K.-B. Park, S. H. Choi, and J. Y. Lee, "M-gan: Retinal blood vessel segmentation by balancing losses through stacked deep fully convolutional networks," *IEEE Access*, vol. 8, pp. 146 308–146 322, 2020.

[68] T. Yang, T. Wu, L. Li, and C. Zhu, "Sud-gan: Deep convolution generative adversarial network combined with short connection and dense block for retinal vessel segmentation," *Journal of Digital Imaging*, pp. 1–12, 2020.

[69] C. Wu, Y. Zou, and Z. Yang, "U-gan: Generative adversarial networks with u-net for retinal vessel segmentation," in *2019 14th International Conference on Computer Science & Education (ICCSE)*. IEEE, 2019, pp. 642–646.

[70] K. Zhou, S. Gao, J. Cheng, Z. Gu, H. Fu, Z. Tu, J. Yang, Y. Zhao, and J. Liu, "Sparse-gan: Sparsity-constrained generative adversarial network for anomaly detection in retinal oct image," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1227–1231.

[71] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[72] C. S. Lee, A. J. Tyring, Y. Wu, S. Xiao, A. S. Rokem, N. P. Deruyter, Q. Zhang, A. Tufail, R. K. Wang, and A. Y. Lee, "Generating retinal flow maps from structural optical coherence tomography with artificial intelligence," *Scientific reports*, vol. 9, no. 1, pp. 1–11, 2019.

[73] S. K. Devalla, G. Subramanian, T. H. Pham, X. Wang, S. Perera, T. A. Tun, T. Aung, L. Schmetterer, A. H. Thiéry, and M. J. Girard, "A deep learning approach to denoise optical coherence tomography images of the optic nerve head," *Scientific reports*, vol. 9, no. 1, pp. 1–13, 2019.

[74] A. Giovannucci, J. Friedrich, P. Gunn, J. Kalfon, B. L. Brown, S. A. Koay, J. Taxidis, F. Najafi, J. L. Gauthier, P. Zhou *et al.*, "Caiman an open source tool for scalable calcium imaging data analysis," *Elife*, vol. 8, p. e38173, 2019.

[75] P. Zhou, S. L. Resendez, J. Rodriguez-Romaguera, J. C. Jimenez, S. Q. Neufeld, A. Giovannucci, J. Friedrich, E. A. Pnevmatikakis, G. D. Stuber, R. Hen *et al.*, "Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data," *Elife*, vol. 7, p. e28728, 2018.

[76] A. Giovannucci, J. Friedrich, M. Kaufman, A. Churchland, D. Chklovskii, L. Paninski, and E. A. Pnevmatikakis, "Onacid: Online analysis of calcium imaging data in real time," in *Advances in neural information processing systems*, 2017, pp. 2381–2391.

[77] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2337–2346.

[78] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[79] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[80] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *International Conference on Machine Learning*, 2019, pp. 7354–7363.

[81] X. Chen, C. Xu, X. Yang, and D. Tao, "Attention-gan for object transfiguration in wild images," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 164–180.

[82] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *European conference on computer vision*. Springer, 2016, pp. 702–716.

[83] C. E. Duchon, "Lanczos filtering in one and two dimensions," *Journal of applied meteorology*, vol. 18, no. 8, pp. 1016–1022, 1979.

[84] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2794–2802.

[85] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2536–2544.

[86] J. H. Lim and J. C. Ye, "Geometric gan," *arXiv preprint arXiv:1705.02894*, 2017.

[87] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.

[88] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[89] S. Hajeb Mohammad Alipour, H. Rabbani, and M. R. Akhlaghi, "Diabetic retinopathy grading by digital curvelet transform," *Computational and mathematical methods in medicine*, vol. 2012, 2012.

[90] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[91] G. Team *et al.*, *GIMP: GNU Image Manipulation Program*. GIMP Team., 2019.

[92] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[93] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International journal of computer vision*, vol. 111, no. 1, pp. 98–136, 2015.

[94] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[95] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.