Synthetic Fundus Fluorescein Angiography using Deep Neural Networks

Florian Schiffers[1,2], Zekuan Yu[1], Steve Arguin[1], Andreas Maier[2], Qiushi Ren[1]

[1]Department of Biomedical Engineering, Peking University, China
[2]Pattern Recognition Lab, University of Erlangen-Nuremberg, Germany.
florian.schiffers@fau.de

Abstract. Fundus fluorescein angiography yields complementary image information when compared to conventional fundus imaging. Angiographic imaging, however, may pose risks of harm to the patient. The output from both types of imaging have different characteristics, but the most prominent features of the fundus are shared in both images. Thus, the question arises if conventional fundus images alone provide enough information to synthesize an angiographic image. Our research analyzes the capacity of deep neural networks to synthesize virtual angiographic images from their conventional fundus counterparts.

1 Introduction

The human retina converts incoming light into a neural signal for further processing in the brain. Because the tissue is metabolically active, early symptoms of diseases such as diabetes are detectable from retinal analysis. Though fundus cameras are widely used for retinal imaging, conventional color fundus cameras are not able to image the functional state of retinal circulation [1].

Fluorescent angiographic methodology, in contrast, augments the capability of conventional fundus imaging. With angiographic imaging, an intravenous, fluorescent dye bounds to leukocytes, which excites the molecules when exposed to blue light. This, in turn, produces a narrow yellow-green light. The enhanced image highlights different features of the fundus. Thus, it is a routine diagnostic tool for diseases such as pseudophakic cystoid macular edema and diabetic macular edema [2]. Despite the diagnostic benefits, physicians are increasingly reluctant to use angiographic imaging technology because of its severe potential side effects [3].

Conventional color fundus imaging and fluorescence angiography are significantly different in appearance. However, many features, such as vessels or granular structures are shared between both methods. The question then emerges if an angiographic image can be efficiently estimated purely from the conventional color image. If possible, this would have the potential to further enhance diagnostic capabilities without an increase in patient risk. A successful synthetization of an angiographic image could reduce or potentially eliminate the need for actual angiographic imaging. In addition, a successful outcome in this area also serves as a potential solution to the shortage of publicly available angiographic images. Modern algorithms for image enhancement or segmentation currently cannot be efficiently trained without access to a large database of angiographic images. If these images were to be synthesized, an indefinite amount of them could be produced and therefore allow for greater research in this area.

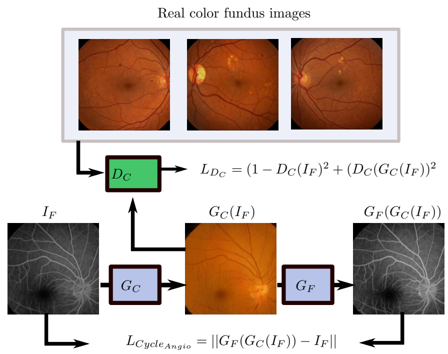
Image synthesis and translation between different modalities has long been an area of research in the medical sciences. For example, tomographic images are a tool not only in diagnosis, but also in dose planning for cancer treatment. This method, however, often has dangerous potential side effects due to dose deposition. To circumvent this risk of harm, researchers have explored methods to generate synthetic CT images from MRI images, as MRI does not pose a risk to patients. In several studies, deep neural networks have proven to work well for typical image-translation tasks in medical imaging [4,5].

Recently, similar image translation methods have also been applied to fundus imaging. Because access to fundus images is often restricted, Costa *et al.* [6] propose to synthesize fundus images from binary vessel trees in order to create large databases for other machine-learning tasks. However, to the best of our knowledge, no current algorithm exists which can estimate fundus fluorescence angiographic images from conventional color fundus images. In this work, we address this question by applying the image translation method of Zhu *et al.* [7] to this problem. They demonstrate image-to-image translation without the necessity of paired images from both modalities. While other methods using large databases of paired images yield superior results [8], similar-sized datasets of paired conventional and angiographic fundus images are not available. For this reason, a generative model using paired images cannot be employed.

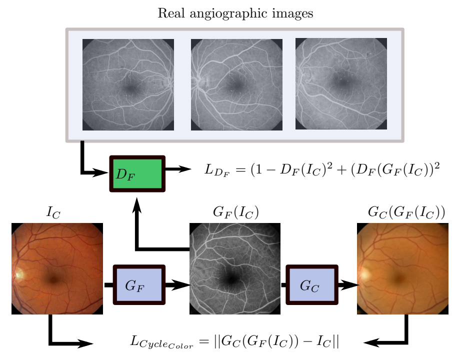
2 Material and Methods

In computer vision, generative models were long investigated to perform image synthesis. These networks were outperformed by generative adversarial networks (GANs) proposed by Goodfellow *et al.* [9]. Here, the generator is augmented by a discriminator, which discerns real and synthesized images. During training, the generator and the discriminator compete in a min-max game, similar to game theory. The generator network gradually refines its ability to fool the discriminator while the discriminator network gradually fine-tunes its filter to detect synthesized images. Thus, the GAN can eventually synthesize images which are indistinguishable from real images.

Recently, Zhu *et al.* [7] proposed a novel architecture, namely CycleGAN, translating images between two image domains A and B, without the need for tightly-coupled pairs. Unlike previous work, this setup is trained solely on the generated image quality specified by the discriminator. However, this problem is highly underdetermined and is hardly optimized. This problem is overcome by enforcing "cycle-consistency" in the sense that an image from the output domain should also translate correctly to the input domain. This backwards translation is ensured by training both a second generator and a discriminator network. Just



(a) Cycle consistency for angiographic images.



(b) Cycle consistency for color images.

Fig. 1. The two figures visualize the composition of the loss term used for the training process of the cycleGAN architecture. I_F and I_C are the input images for the color fundus image generator G_C and the angiographic image generator G_F, respectively. Similarly, D_C and D_F denote the respective discriminator networks. Cycle consistency is enforced so that the backwards translation resembles the input image for both ways, see $L_{Cycle_{Angio}}$ and $L_{Cycle_{Color}}$. The adversarial loss, i.e. the capacity of the network to distinguish between real and fake images, is modeled by L_{D_C} and L_{D_F}.

as in a standard GAN, the second generator synthesizes from domain B to A, while the discriminator then discerns a real image B from a synthesized image B. For our problem of conventional color and angiographic fundus imaging, this architecture is visualized in Fig. 1.

2.1 Database

This study includes conventional color and angiographic fundus images from two datasets. The first provided by Hajeb *et al.* [10] is publicly available and contains in total 60 image pairs of 30 normal and 30 abnormal cases with a resolution of 720×576 each. The second dataset is comprised of an unpaired dataset of conventional and fluorescent images provided by the people's hospital of Jiangmen City, China. It contains 319 color and 219 fluorescent images where the resolution varies between 1380×1150 and 2800×2300.

In total, 379 color and 279 angiographic images were available in this study. From this, 365 color and 265 angiographic images are used for training, and 14 images from each group are used for testing. The test images were manually chosen to be the image pairs with the best visual alignment. Thus, the most salient image features are visible in both images.

2.2 Preprocessing

For further processing, all images were cropped to be a square with the new image center being the center of the fundus image. Subsequently, all images were downsampled to a resolution of 256×256 to keep the network from overfitting due to the lack of available training data. Color images and angiographic images are saved in the .JPG file format. Since deep learning networks require large datasets, data augmentation is a standard tool to synthetically increase the size of the dataset. In this work, each image was rotated by 90, 180, and 270 degrees, leading to a fourfold increase of training data. For further data augmentation, training samples are upscaled to 286×286 and subsequently randomly cropped to 256×256 during training.

2.3 Network Architecture

This study uses the publicly available PyTorch implementation of CycleGANs provided by Zhu *et al.* [7]. Both generator networks are fully convolutional networks and employ the same architecture. Similarly, both discriminator networks use the same architecture. Generator networks G_C and G_F first process the image through two consecutive convolutional layers with a stride of 2. Thus, these convolutions down-sample the photo twice. The image is then subjected to six residual blocks. Subsequently, the image is brought back to its original size via two consecutive fractionally-strided convolutions with a stride of 1/2. The discriminator networks D_C and D_F apply 70×70 PatchGANs randomly selected on the full resolution images [7]. By learning smaller patches, less parameters have to be determined, thus making the training process of the discriminator more robust. G_C takes a color image with three channels as input, then produces a grayscale image with only one channel. Vice versa, G_F inputs a grayscale image with a three channel color image as output. Respective input configurations are valid for the two discriminator networks.

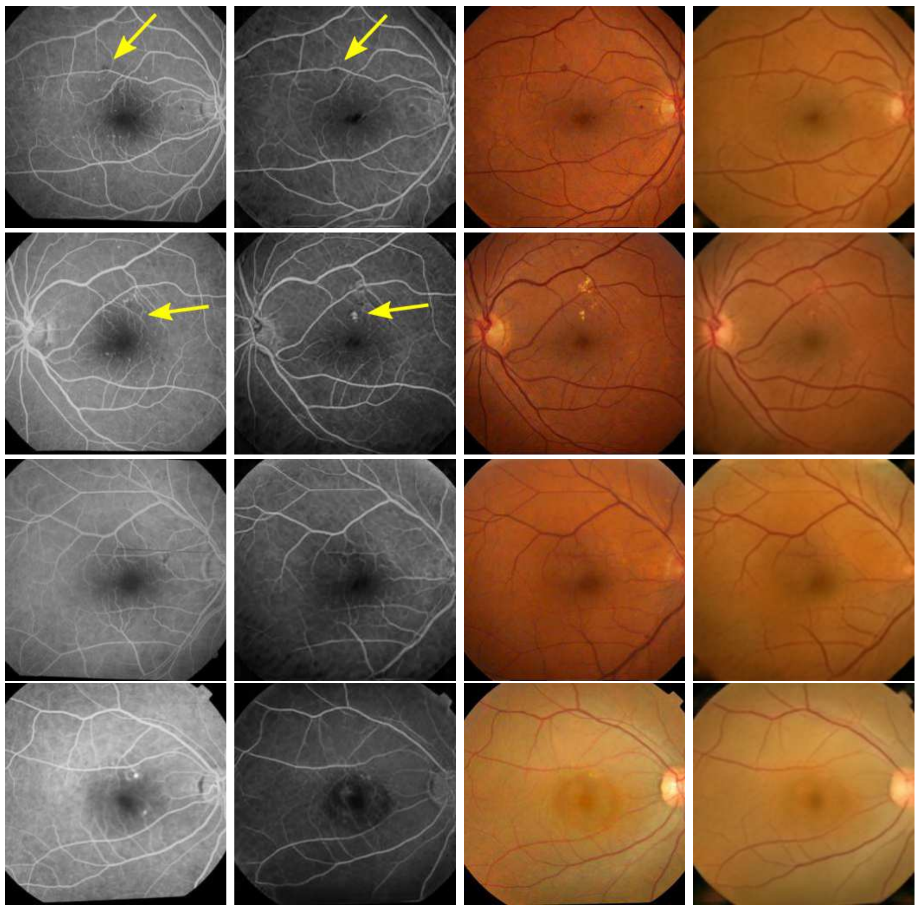
The network was trained using ADAM with a batch size of 1 [7]. The learning rate was 0.0002 for the first 100 epochs, and then linearly decreased to reach 0 with epoch 200. Training with the dataset and configuration described above took about 27 hours using a single GTX 1080.

3 Results

We base our evaluation on how well synthesized images resemble their ground truth counterparts. Figure 2 shows four instances taken from the test set, where an accurate registration between the image pairs were available. Each row of Fig. 2 shows from left to right the real and generated angiographic image, the input color fundus and last the backwards translated color fundus image to demonstrate the cycle consistency.

The generated images are hardly to be identified as synthetic images for a non-medically trained human. Some structures such as vessels are clearly enhanced compared to the input color image. However, there are several instances where patterns or structures in the authentic angiographic images are not synthesized correctly. For example, fine vessel structures, that are clearly visible in the real angiographic images, are unclear or not present within the synthesized

(a) Real FFA (b) Fake FFA (c) Real RGB (d) Fake RGB



(a) Real FFA (b) Fake FFA (c) Real RGB (d) Fake RGB

Fig. 2. Each row shows from left to right the real and generated angiographic image, the authentic color image and the reconstructed color image to show cycle consistency. The first three rows are from dataset [10], the remaining is taken from our own data.

images. Some local structures are located at different positions in the image, as indicated by the yellow arrow. Furthermore, the overall image brightness and contrast between ground truth and synthesized images differ.

4 Discussion

We have demonstrated that image translation between color fundus images and angiographic images is principally possible. For this, the CycleGAN architecture proposed by [7] was applied on two unpaired datasets containing conventional color and angiographic fundus images. The network was trained with downsampled and unpaired fundus images with a resolution of 256×256. A qualitative evaluation with registered image pairs independent from the training set reveals high overall resemblance with the ground truth, while some small details cannot be synthesized correctly. We are optimistic that this technique suffices to engineer robust algorithms for angiographic images by creating large synthetic databases. However, it remains unclear whether this has the potential to provide the same level of utility to a medical practitioner. This matter will be subject to a future clinical study.

Additionally, our research will focus on increasing the generated image resolution from 256×256 to state-of-the-art resolution used in medical imaging. A naive increase of the generator network's capacity to directly synthesize high-resolution data will not lead to a success, since only an insufficient amount of training samples are available. We will investigate patch-based approaches employed by similar work [6,4] as well as more sophisticated data-augmentation methods.

References

1. Abramoff MD, Garvin MK, Sonka M. Retinal imaging and image analysis. IEEE Rev Biomed Eng. 2010; p. 169–208.
2. Shoughy SS, Kozak I. Selective and complementary use of Optical Coherence Tomography and Fluorescein Angiography in retinal practice. Eye and Vision. 2016;.
3. Musa F, Muen W, Hancock R, Clark D, et al. Adverse effects of fluorescein angiography in hypertensive and elderly patients. Acta Ophthalmol. 2006;.
4. Nie D, Cao X, Gao Y, Wang L, Shen D; Springer. Estimating CT image from MRI data using 3D fully convolutional networks. 2016; p. 170–178.
5. Wolterink JM, Dinkla AM, Savenije MH, Seevinck PR, van den Berg CA, Isgum I. Deep MR to CT Synthesis using Unpaired Data. 2017;.
6. Costa P, Galdran A, Meyer MI, Abràmoff MD, Niemeijer M, Mendonça AM, et al. Towards Adversarial Retinal Image Synthesis. arXiv:170108974. 2017;.
7. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. 2017;.
8. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. arXiv preprint arXiv:161107004. 2016;.
9. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. Adv Neural Inf Process Syst. 2014; p. 2672–2680.
10. Hajeb Mohammad Alipour S, Rabbani H, Akhlaghi MR. Diabetic retinopathy grading by digital curvelet transform. Comput Math Methods Med. 2012;2012.