

Fundus2Angio: A Conditional GAN Architecture for Generating Fluorescein Angiography Images from Retinal Fundus Photography

吴卓儒

摘要

利用荧光素血管造影技术 (Fluorescein Angiography) 进行眼底视网膜血管病变临床诊断是一个耗时的过程。荧光素血管造影技术需要给病人注射荧光素钠 (Sodium Fluorescein) 或让病人口服荧光染料, 这可能会导致患者严重的不良反应, 甚至可能致命。目前, 在临床治疗中是没有能够为病人生成眼底荧光素血管造影图像的非侵入性系统的。然而, 视网膜眼底彩照成像是一种非侵入性的成像技术, 可以在几秒钟内完成。为了尽可能消除对 FA 的需求, 本文提出了一个有条件的生成对抗网络 (GAN) 来将眼底彩色图像转换为 FA 图像。所提出的 GAN 由一个能够生成高质量 FA 图像的新型残差块组成。这些图像是鉴别诊断视网膜疾病的重要工具, 凭借本文所提出架构, 病人无需进行可能有副作用的侵入性检查方法。本文的实验表明, 所提出的架构实现了 30.1 的低 FID 分数, 并且优于其他最先进的生成网络, 生成的图像与真实的眼底荧光素血管造影图像几乎无法被区分。

关键词: 生成对抗网络; 图像翻译; 荧光素血管造影技术; 视网膜眼底镜像检查

1 引言

长期以来, 荧光素血管造影技术 (Fluorescein Angiography) 和视网膜眼底镜像检查 (Retinal Funduscopy) 一直被组合运用于视网膜血管和色素上皮脉络膜疾病的诊断中^[1]。这个过程需要给患者注射一种荧光染料。根据年龄和眼睛的心血管结构, 这种染料将会在 8 至 12 秒内出现在视静脉中, 并停留长达 10 分钟^[2]。虽然一般认为这样的检查手段是安全的, 但也有报道称, 由于患者对染料的过敏反应, 出现了轻度到严重的并发症^[3]。常见的副作用包括恶心、呕吐、过敏反应、心脏病发作、过敏性休克和死亡等^[4]。此外, 荧光素在体内区域的渗漏也很常见, 但荧光素溶液的浓度对上述不良反应没有直接影响^[5]。

考虑到这种检查手段的并发症和风险, 一种新的、无创的、可负担得起的以及计算有效的检查方法是非常必要的。目前可用光学相干断层扫描技术 (Optical Coherence Tomography) 和其他基础图像处理技术来达到代替荧光素血管造影技术的目的, 但成本较高。没有一个计算有效和经济可行的机制来生成可靠和可复制的荧光素血管造影图像, 唯一的替代方案是利用视网膜眼底镜像检查来鉴别诊断。尽管已经提出了由数字图像处理和机器学习算法组成的自动化系统, 用于从眼底图像诊断潜在的条件和疾病^[6], 但没有从视网膜照片中有效生成荧光素血管造影图像。本文提出了一种新的条件生成对抗网络 (GAN), 称为 Fundus2Angio, 能够从视网膜眼底图像中合成荧光素血管造影图像。该程序完全自动化, 不需要任何人为干预。我们使用定量评价指标来测试所提出的体系结构。我们将所提出的体系

结构与其他最先进的条件 GANs 进行了比较^[7-9]，我们的模型优于这些网络。结果表明，该网络生成的荧光素血管造影图像与真实的荧光素血管造影图像几乎很难被区分。

2 相关工作

生成对抗网络已经彻底影响和改变了许多图像处理任务的解决方法，例如图像编辑^[10]、图像风格生成^[11]和图像风格转移^[8-9]。多分辨率结构是计算机视觉中常见的实践架构，而耦合结构具有从图像中结合精细和粗略图像信息的能力^[12]。最近，关于条件 GANS^[13]和无条件 GANS^[14]的技术探索实践了针对特定领域任务体系结构中组合分辨率的思想。受此启发，我们提出了一个在不同尺度上提取特征的架构。

有些方法还使用多尺度鉴别器进行风格转换^[8]。然而，它们只在处理精细特征的生成器中附加了鉴别器，而完全忽略了处理粗略特征生成器中的鉴别器。为了在最粗糙的尺度上学习有用的特征，分离的多尺度鉴别器是必要的。我们提出的体系结构将其用于粗略生成器和精细生成器。

为了获得高质量的图像合成，前人还提出了一种具有多对鉴别器和生成器的金字塔网络，称为 Singan^[15]。虽然它能产生高质量的合成图像，但该模型只适用于未配对的图像。为了解决这个问题，每个生成器的输入都是前一个生成器产生的合成输出。因此，它不能用于满足条件的成对图像训练。为了缓解这个问题，需要建立一个可以将特征从粗略生成器传播到精细生成器的连接。本文提出了一种在粗略生成器和精细生成器之间具有特征附加机制的结构，使其成为具有多尺度鉴别器的两级金字塔网络，如图 1所示。Fundus2Angio 由两个生成器 G_{coarse} 、 G_{fine} 和四个判别器 $D1_{coarse}$ 、 $D1_{fine}$ 、 $D2_{fine}$ 和 $D2_{coarse}$ 组成。两个生成器以眼底彩照图像为输入进而输出荧光素血管造影图像，四个判别器以眼底彩照图像和对应的荧光素血管造影图像为输入进而输出生成器输出图像是真或假的判断结果。Fundus2Angio 所设计单独的多尺度鉴别器可以在最粗糙的尺度上学习到有用特征。

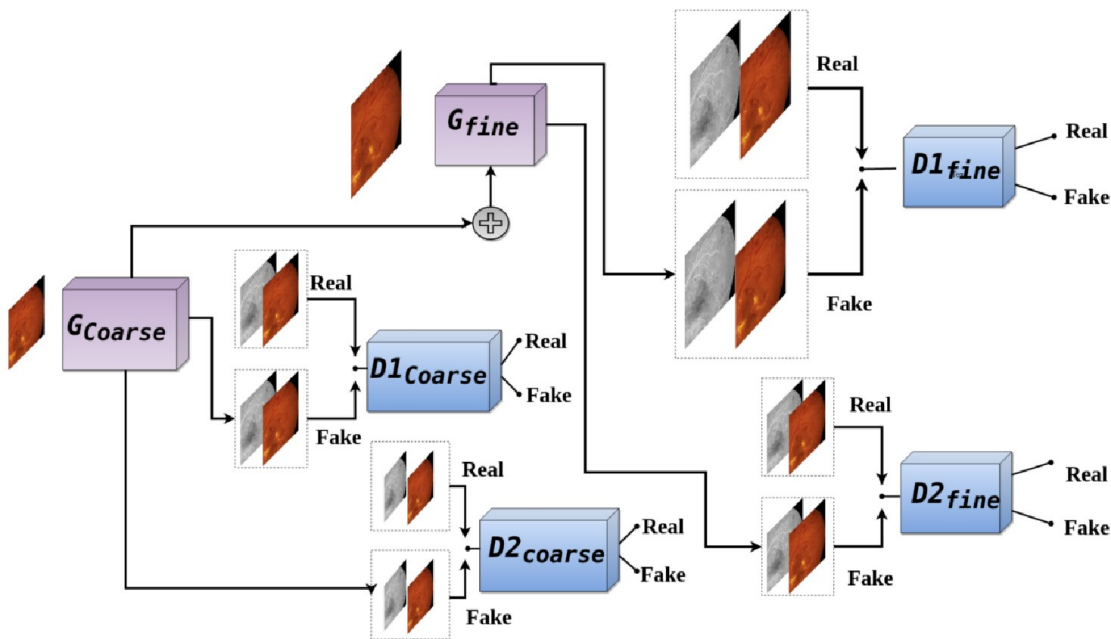


图 1: Fundus2Angio 整体框架

3 本文提出方法

本文介绍了一种新的条件生成对抗网络 (GAN)，它由新型残差模块组成，用于从视网膜眼底图像中生成尽可能真实的荧光素血管造影图像。首先，文章在 3.1 中介绍了新型残差模块。然后，在 3.2 和 3.3 中深入研究了所提出的包含精细生成器和粗糙生成器以及四个多尺度鉴别器的条件 GAN。最后，在 3.4 中讨论了所提出架构中所用到的目标函数和损失权重分布。

3.1 新型残差模块

最近，残差模块被频繁运用于许多图像分类、检测和分割任务^[16]。图像到图像转换任务、图像超分辨率任务等有趣应用均使用了残差模块^[8,17]。一个残差单位包括两个连续的卷积层，第二个卷积层的输出被附加在输入中，进而允许实现更深的网络。在计算成本上来看，普通卷积 (regular convolution layers) 的成本高于可分离卷积 (separable convolution)^[18]。这里的可分离卷积就是执行深度可分离卷积，进行深度 Depthwise 卷积再进行点 Pointwise 卷积，这种设计有利于通过网络来提取和保留深度和空间的信息。此外，interspersing 散布卷积层可以实现更高效更准确的网络^[19]。我们结合了这个想法来设计了一种新的残差块，以求保留深度和空间信息，降低计算复杂度，确保有效的内存使用，如图 2 所示。

Residual Block	Equation	Activation	No. of Parameters [†]
Original	$[R_i \otimes F_{Conv} \otimes F_{Conv}] + R_i$	ReLU (Pre) [7]	18,688
Proposed	$[R_i \otimes F_{Conv} \otimes F_{SepConv}] + R_i$	Leaky-ReLU (Post)	10,784

[†] F_{Conv} and $F_{SepConv}$ has kernel size $K = 3$, stride $S = 1$, padding $P = 0$ and No. of channel $C = 32$.

图 2: 新型残差模块与普通残差模块参数对比

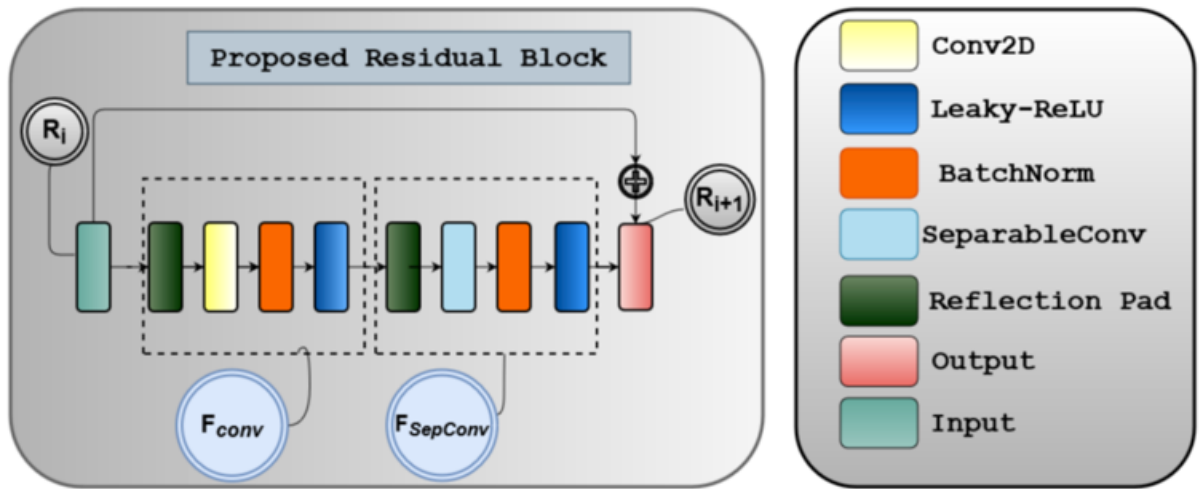


图 3: 所提出的新型残差模块操作细节

如图 3 所示，利用一个可分离卷积代替最后一个普通卷积运算操作。在普通卷积层和可分离卷积层之后，还使用批次归一化和 Leaky-Relu 作为后激活机制。为了获得更好的效果，在每次卷积运算操作之前都采用了反射填充操作，而不采用零填充操作。整个操作流程的计算公式如下所示：

$$\begin{aligned}
 R_{i+1} &= [R_i \otimes F_{Conv} \otimes F_{SepConv}] + R_i \\
 &= F(R_i) + R_i
 \end{aligned} \tag{1}$$

在计算公式中， \otimes 指卷积运算操作， F_{Conv} 和 $F_{SepConv}$ 分别代表端到端的普通卷积和可分离卷积运算操作。Leaky-ReLU 与普通卷积以及可分离卷积的组合使用可以使得两个不同特征图（空间和深

度信息) 被用于生成精细的荧光素血管造影图像。

3.2 粗糙和精细生成器

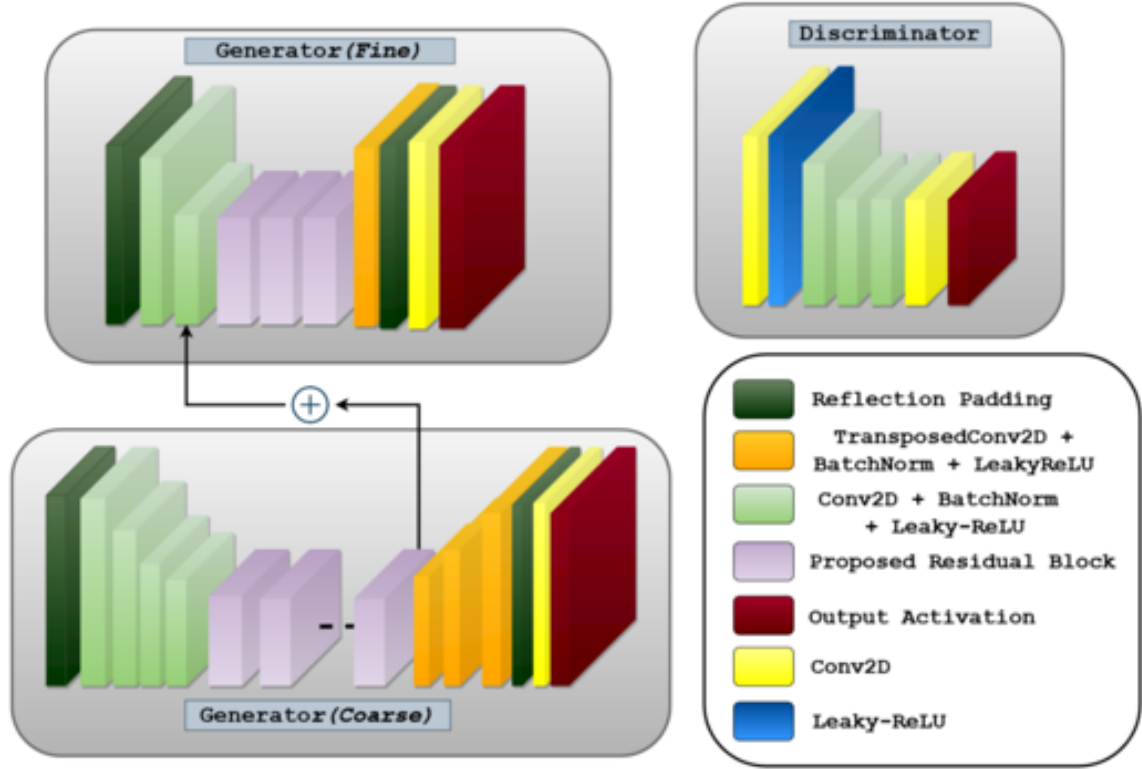


图 4: Fundus2Angio 生成器和鉴别器结构细节

在架构 PIX2PIXHD^[8]和 SINGAN^[15]中观察到, 对于有条件和无条件的 GANs, 使用由粗糙到精细的生成器可以产生非常高质量的图像。受此启发, 我们在所提出的 Fundus2Angio 网络中采用 G_{coarse} 和 G_{fine} 两个生成器。如图 4所示, G_{fine} 生成器由两个编码块(浅绿色)、三个残差块(紫色)和一个解码块(橙色)组成, G_{coarse} 生成器由四个编码块(浅绿色)、九个残差块(紫色)和三个解码块(橙色)组成。两个生成器都有中间特征普通卷积层、反射填充层和输出激活函数 Tanh。鉴别器由三个编码块(浅绿色)组成, 鉴别器都有中间特征普通卷积层、Leaky-ReLU 层和输出激活函数 Sigmoid。 G_{fine} 生成器通过学习局部信息, 例如视网膜小静脉、小动脉、出血、渗出物和微血管瘤等, 来从眼底图像生成精细的荧光素血管造影图像。 G_{coarse} 生成器在生成粗糙的荧光素血管造影图像过程中尝试去提取和保存全局信息, 例如黄斑、视盘、颜色、对比度和亮度的整体结构。

G_{fine} 生成器的输入输出大小均为 512×512 , G_{coarse} 生成器的输入输出大小均为 256×256 。 G_{coarse} 生成器输出的大小为 $256 \times 256 \times 64$ 的特征向量被加进 G_{fine} 生成器其中一个中间层里, 混合的生成器可以共享局部和全局信息^[8,15,17]。两个生成器都利用卷积层进行下采样, 再利用转置卷积层进行上采样。需要注意的是, G_{coarse} 生成器在使用转置卷积层进行两次上采样之前先进行下采样两次。两个生成器在进行最后一次下采样之后以及进行第一次上采样之前都使用了提出的新型残差模块。此外, 在 G_{fine} 生成器中进行了一次下采样之后, G_{coarse} 生成器输出的特征向量被加进来, 再执行残差块以及进行上采样, 最后会得到精细的荧光素血管造影图像。所有卷积和转置卷积操作之后都是由批次归一化和 Leaky-ReLU 来激活。为了训练生成器, 首先对随机样本进行一次批量训练 G_{coarse} 生成器, 然后在一个新的随机样本集合上训练 G_{fine} 生成器。在这期间, 鉴别器的权重被冻结, 它们不是可被训练的。最后我们将所有鉴别器和生成器联合微调以此来训练 GAN。

3.3 多尺度 PatchGAN 鉴别器

为了合成荧光素血管造影图像，GAN 判别器需要适应于所生成的粗糙和精细图像，判断是真还是假。解决这个问题需要更深的结构或者感受野更大的卷积核。但这两种方法都可能会导致模型过拟合和增加模型参数量。此外计算所有参数需要大量处理能力。为了解决这个问题，本文运用了两个 Markovian 鉴别器的思想。该鉴别器在 PatchGAN 中被首次引入^[20]，这种技术旨在利用不同尺度的图像输入，SinGan^[15]和 pix2pixHD^[8]中都有运用。

本文使用了四个具有相似网络结构但操作在不同图像尺度的鉴别器。利用 Lanczos 采样方法对真实荧光素血管造影图像和生成的伪荧光素血管造影图像进行下采样操作，下采样 2 倍来创建具备三个尺度（原图、两次下采样和四次下采样）的图像金字塔。将四个鉴别器分成两组， $D_{fine} = [D1_{fine}, D2_{fine}]$ 和 $D_{coarse} = [D1_{coarse}, D2_{coarse}]$ 。这些鉴别器将在三个不同分辨率尺度上被训练，来区分真实和生成的荧光素血管造影图像。

$D1_{fine}$ 和 $D2_{fine}$ 的输出分别为 64x64 和 32x32， $D1_{coarse}$ 和 $D2_{coarse}$ 的输出分别为 32x32 和 16x16，利用给定的鉴别器，损失函数可以表示为公式 (2)。这是一个多任务的问题，即求最大的鉴别器损失，同时求最小的生成器损失。

$$\min_{G_{fine}, G_{coarse}} \max_{D_{fine}, D_{coarse}} \mathcal{L}_{cGAN}(G_{fine}, G_{coarse}, D_{fine}, D_{coarse}) \quad (2)$$

尽管鉴别器具有相似的网络结构，但在较低分辨率下学习特征的鉴别器具有更宽的感受野，它试图提取和保留更多的全局特征，如黄斑、视盘、颜色和亮度等，以生成更好的粗糙图像。与此相反，以原始分辨率学习特征的鉴别器要求生成器产生精细的特征，如视网膜静脉和动脉、渗出液等。这样，我们将全局和局部尺度的特征信息结合起来，同时用它们的配对多尺度鉴别器独立地训练生成器。

3.4 加权目标函数和对抗损失

本文使用 LSGAN^[21]来计算损耗和训练条件 GAN。条件 GAN 的目标函数表示为公式 (3)。

$$\mathcal{L}_{cGAN}(G, D) = E_{x,y}[(D(x, y) - 1)^2] + E_x[(D(x, G(x)) + 1)^2] \quad (3)$$

鉴别器首先在真实眼底图像 x 和真实造影图像 y 上训练，然后再在真实眼底图像 x 和生成的血管造影图像 $G(x)$ 上训练。从训练鉴别器 D_{fine} 和 D_{coarse} 开始，对随机批次的图像进行两次迭代，然后冻结鉴别器的权重来训练 G_{coarse} ，接下来我们以同样的方式在随机样本的一个批次上训练 G_{fine} ，使用 Mean-Squared-Error(MSE) 来计算单个生成器的损失，计算公式如下所示：

$$\mathcal{L}_{L2}(G) = E_{x,y}|G(x) - y|^2 \quad (4)$$

其中， \mathcal{L}_{L2} 是关于真实荧光素血管造影图像 (y) 和生成荧光素血管造影图像 ($G(x)$) 的重建损失， G_{fine} 和 G_{coarse} 都运用了这种损失设计，这样可以让模型生成不同尺度的高质量荧光素血管造影图像。先前的技术也采用了将 MSE 损失运用于基础 GAN 相结合的想法^[22]。从公式 (3) 和公式 (4) 可以计算出最终的目标函数，计算公式如下所示：

$$\min_{G_{fine}, G_{coarse}} \max_{D_{fine}, D_{coarse}} \mathcal{L}_{cGAN}(G_{fine}, G_{coarse}, D_{fine}, D_{coarse}) + \lambda[\mathcal{L}_{L2}(G_{fine}) + \mathcal{L}_{L2}(G_{coarse})] \quad (5)$$

λ 指定了鉴别器或者生成器的优先级。因为就本文目标而言，更多的权重被赋予了生成器的重构损失，所以 λ 值在本文是比较大的。

4 复现细节

4.1 与已有开源代码对比

本文在复现过程中参考了原文作者所开源的 Fundus2Angio 的 Tensorflow 版本代码。此后，再结合原文进行思路理解后，自己动手实现了 Fundus2Angio 的 Pytorch 版本全部代码。对于复现原文中作为 baseline 的 PIX2PIX^[7]和 PIX2PIXHD^[8]，参考了英伟达 Nvidia 在 GitHub 上的开源代码。

4.2 实验设置

本文选用开源的配对的眼底彩照 (Color Fundus) 和荧光素血管造影 (Fundus Fluorescein Angiogram) 图像数据集，来自伊斯法罕医科大学的医学图像和信号处理研究中心 MISP，由 Hajeb 等人提供^[23]。该数据集采集自 60 位病人，包含 30 对具有糖尿病视网膜病变的配对图像和 30 对正常的配对图像。因为经过观察数据集，并不是所有的眼底彩照图像和荧光素血管造影图像均完全对齐。本文基于配准对齐原则，选择了 17 对眼底彩照-荧光素血管造影图像进行实验，所选的图像对要么完全对齐，要么几乎对齐。眼底彩照-荧光素血管造影图像分辨率均为 576x720，眼底彩照图像格式为 RGB，荧光素血管造影图像格式为灰度图。因为数据不足的情况，本文从每幅图像中随机裁剪出分辨率为 512x512 的区域，重复操作 50 次，最终得到 850(17*50) 个可用于模型训练的样本。

LSGAN^[21]被发现能有效地为我们的任务生成所需的合成图像。在实验超参数设置上，我们设置公式 (5) 中 λ 为 10。对于优化器，我们使用学习速率 $\alpha=0.0002$ 、 $\beta_1=0.5$ 和 $\beta_2=0.999$ 的 Adam 优化器。我们设定 batchsize 为 4，并训练 100 个 epoches，在 NVIDIA GeForce GTX TITAN X GPU 上训练模型花了大约 13 个小时。

5 实验结果分析

为了评估模型的性能，本文采集了 14 幅图像，从图像的每个象限裁剪 4 个截面，尺寸为 512×512，最终得到 56(14*4) 个可用于模型测试的样本。为了定量评价，本文使用 Fréchet 起始距离 (FID) 作为评价指标^[24]，它已经被用来评估相同风格转移的 GANs^[25]。计算真实荧光素血管造影图像与生成荧光素血管造影图像之间的 FID 分数，FID 越低代表生成效果越好，实验结果如表 1 所示。

Architecture(Version)	FID
PIX2PIX(Tensorflow)	48.6
PIX2PIXHD(Tensorflow)	42.8
Fundus2Angio(Tensorflow)	30.3
PIX2PIX(Pytorch)	49.9
PIX2PIXHD(Pytorch)	43.2
Fundus2Angio(Pytorch)	29.7

表 1: 模型实验结果

Fundus2Angio 的 Pytorch 版本模型生成荧光素血管造影图像示例如图 5 所示，该模型在测试集上计算出 FID 为 29.7，与 FID 为 30.3 的 Fundus2Angio 的 Tensorflow 版本模型性能相近并且更优，亦均优于 PIX2PIX 和 PIX2PIXHD。此外，复现的 Pytorch 版本的 PIX2PIX 和 PIX2PIXHD 模型基本能分别获得与 Tensorflow 版本 PIX2PIX 和 PIX2PIXHD 模型一样的性能，尽管 FID 分数稍低。经过比较与分

析猜测，batchsize 的设定大小和 GPU 的型号均会影响模型的训练效果。

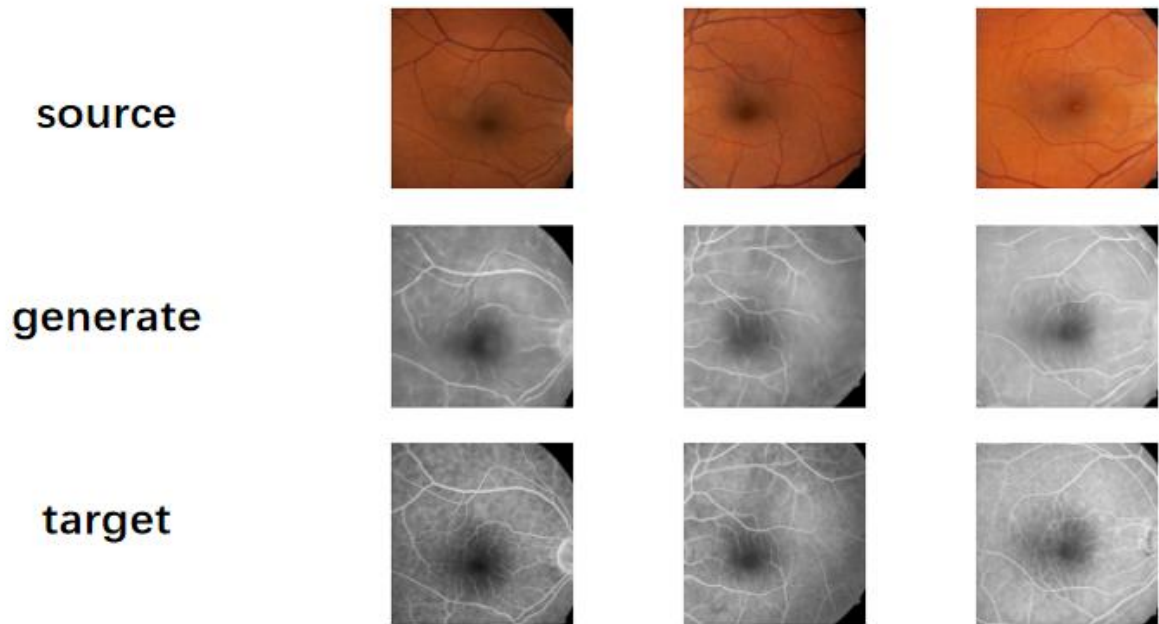


图 5: Fundus2Angio 生成图像示例

6 总结与展望

本文提出了一种全新的有条件的生成对抗网络 Fundus2Angio，在不同尺度上提取特征，同时关注粗糙和细节特征，并建立起将特征从粗略生成器传播到精细生成器的连接，使得粗略生成器和精细生成器之间具有特征附加机制，成为具有多尺度鉴别器的两级金字塔网络，它能够有效地根据同一患者的视网膜眼底彩照图像来生成精确的荧光素血管造影图像。文章方法可以更好地协助临床医生掌握病人的疾病进展，监测或帮助发现新发展的疾病状况。文章的一个未来改进方向是与视网膜血管分割任务以及渗出物定位任务进行有效结合。

参考文献

- [1] MARY V S, RAJSINGH E B, NAIK G R. Retinal fundus image analysis for diagnosis of glaucoma: a comprehensive survey[J]. IEEE Access, 2016.
- [2] MANDAVA N, REICHEL E, GUYER D, et al. Fluorescein and icg angiography[J]. St Louis: Mosby, 2004, 106: 800-808.
- [3] BROCKOW K, SÁNCHEZ-BORGES M. Hypersensitivity to contrast media and dyes[J]. Immunology and Allergy Clinics, 2014, 34(3): 547-564.
- [4] LIRA R P C, OLIVEIRA C L D A, MARQUES M V R B, et al. Adverse reactions of fluorescein angiography: a prospective study[J]. Arquivos brasileiros de oftalmologia, 2007, 70: 615-618.
- [5] YANNUZZI L A, ROHRER K T, TINDEL L J, et al. Fluorescein angiography complication survey[J]. Ophthalmology, 1986, 93(5): 611-617.
- [6] POPLIN R, VARADARAJAN A V, BLUMER K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning[J]. Nature Biomedical Engineering, 2018, 2(3): 158-164.

- [7] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1125-1134.
- [8] WANG T C, LIU M Y, ZHU J Y, et al. High-resolution image synthesis and semantic manipulation with conditional gans[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8798-8807.
- [9] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.
- [10] ZHU J Y, KRÄHENBÜHL P, SHECHTMAN E, et al. Generative visual manipulation on the natural image manifold[C]//European conference on computer vision. 2016: 597-613.
- [11] CHEN W, HAYS J. Sketchygan: Towards diverse and realistic sketch to image synthesis[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 9416-9425.
- [12] BROWN M, LOWE D G, et al. Recognising panoramas.[C]//ICCV: vol. 3. 2003: 1218.
- [13] HUANG X, LI Y, POURSAEED O, et al. Stacked generative adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 5077-5086.
- [14] CHEN Q, KOLTUN V. Photographic image synthesis with cascaded refinement networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 1511-1520.
- [15] SHAHAM T R, DEKEL T, MICHAELI T. Singan: Learning a generative model from a single natural image[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 4570-4580.
- [16] HE K, ZHANG X, REN S, et al. Identity mappings in deep residual networks[C]//European conference on computer vision. 2016: 630-645.
- [17] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual losses for real-time style transfer and super-resolution [C]//European conference on computer vision. 2016: 694-711.
- [18] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1251-1258.
- [19] KAMRAN S A, SAHA S, SABBIR A S, et al. Optic-net: A novel convolutional neural network for diagnosis of retinal diseases from optical tomography images[C]//2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA). 2019: 964-971.
- [20] LI C, WAND M. Precomputed real-time texture synthesis with markovian generative adversarial networks[C]//European conference on computer vision. 2016: 702-716.
- [21] MAO X, LI Q, XIE H, et al. Least squares generative adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2794-2802.

- [22] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context encoders: Feature learning by inpainting [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2536-2544.
- [23] HAJEB MOHAMMAD ALIPOUR S, RABBANI H, AKHLAGHI M R. Diabetic retinopathy grading by digital curvelet transform[J]. Computational and mathematical methods in medicine, 2012, 2012.
- [24] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. Gans trained by a two time-scale update rule converge to a local nash equilibrium[J]. Advances in neural information processing systems, 2017, 30.
- [25] KARRAS T, LAINE S, AILA T. A style-based generator architecture for generative adversarial networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4401-4410.