

# 郑州大学毕业设计（论文）

题 目： 基于 U-GAT-IT 的细胞虚拟染色方法研究

指导教师： 朱睿杰 职称： 副教授

学生姓名： 吕怡霖 学号： 20178461417

专 业： 计算机科学与技术

院（系）： 国际学院

完成时间： 2021 年 5 月 19 日

2021 年 5 月 19 日

## 摘要

急性淋巴细胞性白血病等血液疾病已成为威胁人类健康的重要疾病，目前，对于外周血细胞研究主要借助于病理切片的细胞学分析方法，也是血液病诊断的金标准。但是病理切片染色是一个十分费时费力的流程，使用的化学试剂在染色之后也会对细胞造成不可逆的破坏，并且人工化学染色会出现误染现象从而导致误诊。本文在上述背景下，研究人体外周血细胞的虚拟染色方法，本课题主要工作分为以下几步：

(1) 数据集的采集。通过光学显微镜中 $\times 40$  的物镜观测经过瑞氏吉姆萨染色前后的细胞切片，利用显微镜成像系统记录两种切片在显微镜中的视野形成 2560\*1920 像素的图像。

(2) 数据集的处理。数据集被分为训练集和测试集，且根据不同的需求对采集到的图像进行数据增强。在片段染色时采用裁剪的方式使数据集数量扩充为原来的 70 倍，整体染色时采用旋转的方式使数据集数量增加至原来的 4 倍。

(3) 网络模型的提出。采用一种基于无监督的图像到图像翻译的模型 U-GAT-IT，将训练集放入模型进行训练，完成后就可以实现虚拟染色，即输入一张未染色的白片，一秒内就可以生成虚拟染色图像。

(4) 结果评价。展示 U-GAT-IT 的染色效果，对比五种不同的 GAN 模型在此数据集上的表现，并通过三种不同的评估指标对这五种 GAN 模型进行分组评估，从而显示 U-GAT-IT 的优越性。

虚拟染色的提出解决了当前人工染色出现的主要问题，节约了大量的时间和人力成本，还是一种精度高、完全自动、非主观的方法。循环生成对抗网络无需配对的数据集，尽管目前准确度并不理想，但拥有巨大的发展潜力，可以作为建立全国范围内的虚拟染色系统的基本方法。

**关键词：**虚拟染色；外周血细胞；瑞氏吉姆萨染色法；循环生成对抗网络；Pytorch

## **Abstract**

Blood diseases such as acute lymphocytic leukemia have become an important threat to human health. For peripheral blood cell research, cytological analysis methods mainly rely on pathological slices are also the gold standard for the diagnosis of blood diseases. But pathological slicing is a time-consuming and laborious process, the use of chemical reagents after dyeing can also cause irreversible damage to cells, and artificial chemical staining will be misfected phenomenon leading to misdiagnosis. In this paper, we try to explore the virtual staining method of human peripheral blood cells in the above context, the main work of this topic is divided into the following steps:

- (1) Collection of data sets. Through the optical microscope  $\times 40$  objective observation through the Wright-Gemsa dyeing before and after the cell slice, using the microscope imaging system to record the two slices in the microscope field of view to form a 2560 x 1920 pixel image.
- (2) The processing of the data set. Data sets are divided into training and test sets, and the captured images are data enhanced according to different needs. The number of data sets is expanded by 70 times when the fragment is dyed by cropping, and the number of data sets is increased by 4 times when the overall staining is rotated.
- (3) The proposed network model. Using a model U-GAT-IT based on unsupervised image-to-image translation, the training set is put into the model for training, and virtual staining can be achieved after completion, i.e. an unstained white sheet can be entered, and virtual staining image can be generated in one second.
- (4) Evaluation of results. The staining effect of U-GAT-IT is demonstrated, the performance of five different GAN models is compared on this data set, and the five GAN models are evaluated in groups with three different evaluation indicators to show the superiority of U-GAT-IT.

The introduction of virtual staining solves the main problems of artificial dyeing, saves a lot of time and labor costs, and is also a high precision, complete automation and non-subjective method. Cycle generative adversarial network does not need the paired images, although the accuracy is not ideal at the moment, has great potential for development and can be used as a basic method for building virtual staining systems across the country.

**Key Words:** Virtual staining; peripheral blood cells; Wright-Giemsa staining method; Cycle generative adversarial network; Pytorch

# 目录

摘要.....	I
Abstract.....	II
目录.....	IV
第1章 前言.....	1
1.1 研究背景.....	1
1.2 研究意义.....	1
第2章 生成对抗网络介绍.....	3
2.1 生成对抗网络 (GAN) .....	3
2.1.1 概述.....	3
2.1.2 原理.....	4
2.1.3 损失函数.....	4
2.2 循环生成对抗网络 (CycleGAN) .....	5
2.3 基于 cGAN 的图像翻译 (Pix2pix) .....	6
第3章 U-GAT-IT 模型方法.....	8
3.1 概述.....	8
3.2 网络框架.....	8
3.3 模型创新.....	11
3.4 模型实现过程.....	14
3.5 双向损失函数.....	17
第4章 数据集与实现.....	19
4.1 数据集.....	19
4.2 实施细节.....	19
第5章 结果与评价.....	20
5.1 染色结果.....	20
5.1.1 U-GAT-IT 模型染色结果.....	20
5.1.2 染色结果对比.....	25
5.1.3 染色结果分析.....	27
5.2 评价指标.....	28
5.3 评价结果.....	31
第6章 总结与展望.....	32
6.1 总结.....	32
6.2 讨论.....	32
6.3 问题及今后的研究方向.....	34
参考文献.....	36
成果.....	37
致谢.....	38

# 第1章 前言

## 1.1 研究背景

细胞形态学又称细胞学检查，是传统的临床诊断方法，历史悠久。它的方法简便、实用以及可通过人体视觉系统直接所见的特性，在很长的一段时间内作为血液病研究最重要的方法。

随着光学显微镜的发明和染色技术的发现和改进，血细胞学应运而生。1898年，Ehrlich 和 Lazarus 发明了一种名为三酸混合染色液的混合物，这是一个跨时代的染料，它主要用于外周血切片的染色，并且能准确清晰地区分白细胞的种类。在之后，Wright、Giemsa、Jenner 等人对天青伊红混合物染色剂做出了改进，取名为 Wright-Giemsa 混合染色液，得到了相对良好的染色效果，这种染色剂如今也被广泛使用。

随着 GAN 于 2014 年被 Goodfellow 等人提出，这个被称为近十年来最有趣的想法将图像生成领域推向了一个新的高度。在各种 GAN 模型不断被提出后，近几年一些研究团队也将注意力放在了染色领域，认识到了化学染色领域中的一些瓶颈，他们着手使用一些常用的 GAN 模型进行虚拟染色。2019 年 4 月，来自加州大学的 Yair Rivenosn 和 Hongda Wang 等人在 *nature* 上发表了一篇有关虚拟组织染色的论文引起了巨大反响，随后有越来越多的团队尝试通过不同的 GAN 模型实现虚拟染色，虚拟染色领域也正式作为一种新的发展方向被越来越多医学专家所重视。

## 1.2 研究意义

在细胞形态学中染色细胞样品的标准过程制备是十分耗时耗力的，因为它是劳动密集型的工作，需要专业的染色室环境，化学试剂和训练有素的技术人员。在不同的染色室或技术人员的处理下，具有不确定性的细胞染色可能导致观察受阻甚至误诊，同样地，繁琐的步骤带来了较长的染色时间。因此我们提出了一种低成本、高效率的虚拟细胞染色方法，相对与人工化学染色方法，虚拟染色是完全自动化且非主观的，消除了人工染色中因操作不当导致的染色模糊、误染等问题，最显著的优势是利用深度学习中训练好的神经网络将染色时间从原来的小时级降至毫秒级，极大地提升了诊断效率。

目前世界上的大部分有关虚拟染色的研究都是基于 CycleGAN 这种无监督（染色图像与未染色图像的细胞分布位置不同且没有联系）的网络模型开展的，用作研究公开的数据集也十分有限，因此我们希望通过使用其他的 GAN 模型并结合我们自己采集的数据集实现虚拟染色，为虚拟染色领域今后的研究提供一些帮助。

在未来虚拟染色领域，科学家们更愿意用更少的训练成本去实现精准染色，所以一些基于 CycleGAN 的无监督改进模型备受关注。U-GAT-IT 是一种无监督的图像到图像翻译的新方法，于 ICLR2020 国际会议被提出，可以满足纹理和图像差别很大的两个图像域之间的转换问题。它与基本的 CycleGAN 类似，在训练模型时我们不需要像素级对应的数据集即可实现两个图像域之间的转换，实际上我们目前的条件还未达到专业实验室的水准，无法获取完全对应的两张图像，因此选择 U-GAT-IT 的一大好处就是能够省去大量制作数据集的时间精力，从而快速搭建好完整的虚拟染色训练模型。我们选择 U-GAT-IT 的另一个原因是其模型的优越性，它可以通过无监督学习实现两个图像域间纹理差别很大时的风格转换，并且在细节处理方面优于普通的 CycleGAN，因此我们认为 U-GAT-IT 能在风格转换和细节提升方面也有更好的的表现。

## 第2章 生成对抗网络介绍

### 2.1 生成对抗网络 (GAN)

#### 2.1.1 概述

GAN 的产生受博弈论中的二人零和博弈启发，规定双方一个为生成器(Generator)而另一个为判别器(Discriminator)，生成器被输入噪声向量并寻找真实数据样本的潜在分布，最终生成新的数据样本；判别器则是一个二分类器，对它输入两种样本，让其判断输入的数据是真实的还是生成的。为了达成均衡状态，双方需要不断地优化，分别增强自身的生成和判别能力，而这个学习过程就是寻找二者之间平衡的过程。

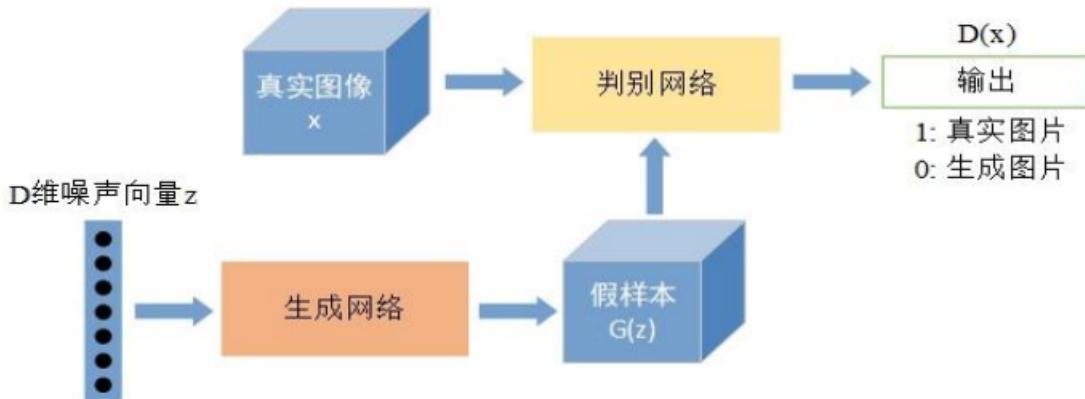


图 2.1 GAN 的结构

GAN 的计算流程如图 2.1 所示。生成网络接收一个随机的 D 维噪声向量  $z$ , 通过最大似然估计算法去尽可能的生成样本图像，记为  $G(z)$ 。判别网络接收一张输入图片  $x$ , 判断该图像是真实样本还是网络生成的假样本  $G(z)$ ，判别网络的输出  $D(x)$ , 其中  $x$  为真实图片的概率。如果  $D(x)$  的值为 1, 说明判别网络判定该输入是真实图片, 如果  $D(x)$  的值为 0, 则说明其判定该输入是经过生成网络后生成的样本  $G(z)$ 。

在训练过程中，两个网络互相对抗并形成一个动态的平衡，用公式可以被描述为：

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}(x)} [\log(D(x))] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (2-1)$$

在最理想的情况下，生成器  $G$  可以生成与真实图片极其相似的样本  $G(z)$ ，此时判别器  $D$  很难判定这张生成的图片的真实性，对图片的真假进行随机估计，得到  $D(G(z))=0.5$ 。

## 2.1.2 原理

(1) 生成器模型中的最大似然估计。最大似然估计的作用是利用已知的样本结果，倒推出现这种结果最可能的参数，上文中提到生成器需要通过输入的噪声向量去生成一个假样本，那么我们就可以利用最大似然估计的特性达成目的。

(2) 生成器通过函数  $G$  努力学习一个分布  $P_g(x)$ ，让它接近数据的真实分布  $P_d(z)$ ，鉴别器尽可能学习到  $P_g(x)$  和  $P_d(z)$  之间的差异  $V(D, G)$ ，GAN 的目标是计算  $G = \arg \min_G \max_D V(G, D)$ ，鉴别器目标是使两个分布间的差异  $V(D, G)$  最大，生成器目标是使得两个分布间的差异  $V(D, G)$  最小，则最终优化目标是： $G = \arg \min_G \max_D V(G, D)$ 。

## 2.1.3 损失函数

将概述中提到的生成模型和判别模型训练优化过程归结为如下目标函数表达式：

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}(x)} [\log(D(x))] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (2-2)$$

其中  $V(D, G)$  为分布差异（交叉熵损失函数）， $x$  为输入的真实图像， $z$  为生成器输出的图像， $D(x)$  为判别器对真实图像的判别结果值（通常以 label 表示，0 或 1）， $D(G(z))$  为判别器对生成样本的判别结果值。

GAN 模型的优化学习是一个交替迭代的过程，正常情况下我们无法首先优化生成器，所以我们先对判别器进行优化，即保持生成器  $G$  不变，训练判别器  $D$ ，表达式如下：

$$\max_D V(D, G) = E_{x \sim p_{\text{data}}(x)} [\log(D(x))] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (2-3)$$

GAN 模型中  $D$  的作用是判别输入样本的真实性，我们希望它的判别结果值越接近 1 越好，我们采用交叉熵作为其损失函数，整个判别器的损失函数是真实图像和噪声的交叉熵之和。对于第一项  $E$  因为输入的是真实图像所以我们期望  $D(x)$  趋近 1，也就是第一项更大才能让损失函数更小，同理第二项  $E$  输入的是  $G$  生成的假图像，我们期望  $D(G(z))$  趋近 0，也是说第二项更大才能使损失函数更小，所以只有令  $D(x)$  最大时才能保证输出的损失函数值最小，这时判别结果才更接近 1。

优化判别器后我们再对生成器  $G$  进行优化，此时保证判别器  $D$  不变，训练  $G$ ，表达式如下：

$$\min_G V(D, G) = E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (2-4)$$

GAN 模型中生成器  $G$  的作用是生成假样本以欺骗判别器  $D$ ，最终达到判别器无法分辨真假的均衡状态，由于我们设置当标签 `label` 等于 1 时代表为真实图像，那么我们就需要尽可能让  $D(G(z))$  趋近 1，也就是让上述表达式中的  $E$  项越小越好，所以只有令  $G(z)$  最小时才能保证其损失函数值最小，此时生成结果才更接近 1。

生成器  $G$  期望  $D(G(z))$  更大且趋近于 1，此时  $V(D, G)$  会变小；判别器  $D$  期望  $D(x)$  更大且趋近于 1，并且期望  $D(G(z))$  更小且趋近于 0，此时  $V(D, G)$  会变大，从而达到动态平衡。

## 2.2 循环生成对抗网络 (CycleGAN)

CycleGAN 可以将一类风格的图片转换成另一类风格图片，它将两个 GAN 模型结合起来，从而实现非成对数据集之间的转换，特别适用于图像风格迁移。它对数据集质量要求较低，只需要两种不同风格的图像就能实现风格转换。

为了增强 GAN 的稳定性并解决单一映射的问题，CycleGAN 引入了双向映射机制。假设目前有两个样本域  $X$  和  $Y$ ，我们要把  $X$  域中的样本转换成  $Y$  域中的样本，那么 CycleGAN 的目标就是学习两个域之间的相互映射。映射为  $F$  对应着生成对抗网络中的生成模型，且  $F$  可以将  $X$  域中的图片  $x$  转换为  $Y$  域中的图片  $F(x)$ 。同时我们还需要生成对抗网络中的判别器来判别生成的图片是否为真实的，假设这个判别器为  $D_Y$ ，那么我们可以下面这个表达式来表示 GAN 损失：

$$\min_F \max_D V(D, G) = E_{Y \sim P_{data}(y)} [\log D_Y(y)] + E_{X \sim P_{data}(x)} [\log (1 - D_Y(F(x)))] \quad (2-5)$$

事实上这个损失函数与上文提到的通用对抗网络的损失函数是相同的，但是仅仅采用这一个损失是不能直接训练的，因为映射  $F$  会将所有的  $x$  映射至  $Y$  域中的同一张图片，这时损失函数就变得无效，即使进行迭代也无法优化。因此 CycleGAN 提出了循环一致性损失函数。

再假设一个可以将  $Y$  域中的样本  $y$  转换为  $X$  域中样本  $G(y)$  的关系映射  $G$ ，CycleGAN 同时学习  $F$  和  $G$  两个映射，并通过一个 L1 损失函数来要求  $F(G(y)) \approx y$  以及  $G(F(x)) \approx x$ 。这样将  $X$  域样本转换到  $Y$  域后，还可以重新转换回  $X$  域中，从而可以杜绝 GAN 模型把所有的  $x$  都转换为  $Y$  域中的某一张图片。根据  $F(G(y)) \approx y$  以及  $G(F(x)) \approx x$ ，这个基于 L1 损失（平均绝对误差）的循环一致性损失函数就定义为：

$$L_{cycle} = E_{x \sim P_{data}(x)} [\| G(F(x)) - x \|] + E_{y \sim P_{data}(y)} [\| F(G(y)) - y \|] \quad (2-6)$$

同样，我们为映射  $G$  也引入一个判别器  $D_x$ ，由此可以定义一个 GAN 的损失  $L_{GAN}(F, D_x)$ ，最终的损失就由三部分组成：

$$\begin{aligned} L &= L_{GAN}(F, D_Y) + L_{GAN}(G, D_X) + L_{Cycle}(F, G) \\ &= E_{Y \sim P_{data}(y)} [\log D_Y(y)] + E_{X \sim P_{data}(x)} [\log(1 - D_Y(F(x)))] + E_{X \sim P_{data}(x)} [\log D_X(x)] \\ &\quad + E_{Y \sim P_{data}(y)} [\log(1 - D_X(G(y)))] + E_{X \sim P_{data}(x)} [\| G(F(x)) - x \|] + E_{Y \sim P_{data}(y)} [\| F(G(y)) - y \|] \end{aligned} \quad (2-7)$$

## 2.3 基于 cGAN 的图像翻译 (Pix2pix)

Pix2pix 方法在 CVPR2017 上被提出，将 GAN 应用于有监督的图像翻译领域。与 CycleGAN 除了网络模型不同外，它也需要严格的一一配对数据集，但它的训练效率更高，生成图像也更逼真。

原始 GAN 中生成器使用的是 encoder-decoder 结构，而 Pix2pix 则采用的是 U-Net 网络结构，能够充分融合样本特征。在之前先降采样到低纬度（降至低分辨率），再升采样到原始分辨率的编解码结构网络中，下采样时我们利用最大池化层或平均池化层保留特征，但这样就会导致一些边缘特征的损失，这些损失在后面的上采样过程中是无法弥补的，因为我们常常采用填零法进行像素扩充。就此缺陷，U-net 引入了跳跃连接层，把编码对应的特征映射和解码之后的同样大小的特征映射按通道连接一起，这样就可以保存图像在各种分辨率下的细节特征，从而进行像素匹配。此外这个跳跃连接层也有一定的减压作用，它避免了 encoder-decoder 网络结构的弊端，即解码时所有的细节特征信息都会被堆积在一个卷积层内，可能会引起网络堵塞甚至发生错误。判别器方面 Pix2pix 采用马尔可夫判别器 (PatchGAN)，将生成的图像分为一块块的感受域，这时判别器只需要判定感受域的真假即可，大大减少了输入维度和参数量。

由于 pix2pix 输入的是图片样本而非噪声，那么它需要采用 cGAN 网络模型的损失函数，即：

$$L_{cGAN}(G, D) = E_{x,y} [\log D(x, y)] + E_{x,z} [\log(1 - D(x, G(x, z)))] \quad (2-8)$$

与此同时，pix2pix 和 CycleGAN 类似，也为生成器添加了 L1 损失函数：

$$L_{L1}(G) = E_{x,y,z} [\| y - G(x, z) \|] \quad (2-9)$$

那么 pix2pix 最终的损失函数由两部分组成：

$$G = \arg \min_G \max_D L_{cGAN}(G, D) + \lambda L_{L1}(G) \quad (2-10)$$

其中  $\lambda$  代表加权因子，用以控制权重，权重越高，L1 损失函数就越有意义。

## 第3章 U-GAT-IT 模型方法

### 3.1 概述

从信息流的角度来看，GAN 的设计一般是单向的。在没有未配对图像的情况下，GAN 可能导致不同的未染色图像被映射到相同的组织染色图像。与 GAN 相比，CycleGAN 是一种先进的未配对图像合成算法，它是两个对称且反向的 GAN 网络拼接构成的一个大的环形网络，这个网络中共享了两个生成器  $G_{\text{Blank} \rightarrow \text{WG}}$  和  $G_{\text{WG} \rightarrow \text{Blank}}$  并各自有一个判别器（两个相关的对抗性判别器  $D_{\text{WG}}$  和  $D_{\text{Blank}}$ ）。

基于 Cycle-GAN 的基本概念，我们的最终目标是训练一个生成器  $G_{\text{Blank} \rightarrow \text{WG}}$ ，该生成器仅用从两个结构域中提取的不配对样本将图像从未染色结构域（Blank）转移到瑞氏吉姆萨染色结构域（Wright-Giemsa）。我们的框架由两个发生器和鉴别器对组成：第一对 ( $G_{\text{Blank} \rightarrow \text{WG}}$  和  $D_{\text{WG}}$ ) 和第二对 ( $G_{\text{WG} \rightarrow \text{Blank}}$  还有  $D_{\text{Blank}}$ )。

U-GAT-IT 是基于 CycleGan 的改进模型，除了需要使用 Adversarial loss、Cycle loss 和 Identity loss 这三种在 Cyclegan 模型中常见的损失函数之外，U-GAT-IT 还提出了在生成器和判别器中使用一个分类器，训练这个分类器也需要一个损失函数，且这个函数受类激活映射启发，因此被称为 CAM loss。除此之外，U-GAT-IT 还采用了注意力模块并引入了新的归一化方式 AdaLIN。

### 3.2 网络框架

真实的白片通过  $G_{\text{Blank} \rightarrow \text{WG}}$  得到假的染色图像之后再通过  $G_{\text{WG} \rightarrow \text{Blank}}$  得到重建的白片，即通过逆向网络还原的虚拟重组白片。同样，真正的 W-G 染片通过  $G_{\text{WG} \rightarrow \text{Blank}}$  得到假的白片然后通过  $G_{\text{Blank} \rightarrow \text{WG}}$  重建染片，得到一个虚拟重组染色图像。判别器方面，与普通的 GAN 类似， $D_{\text{Blank}}$  负责判别真实的白片和通过  $G_{\text{WG} \rightarrow \text{Blank}}$  生成的假白片之间的真实性，同理， $D_{\text{WG}}$  负责判别真实的 W-G 染色图像和通过  $G_{\text{Blank} \rightarrow \text{WG}}$  生成的假染片之间的真实性，如图 3-1 所示。

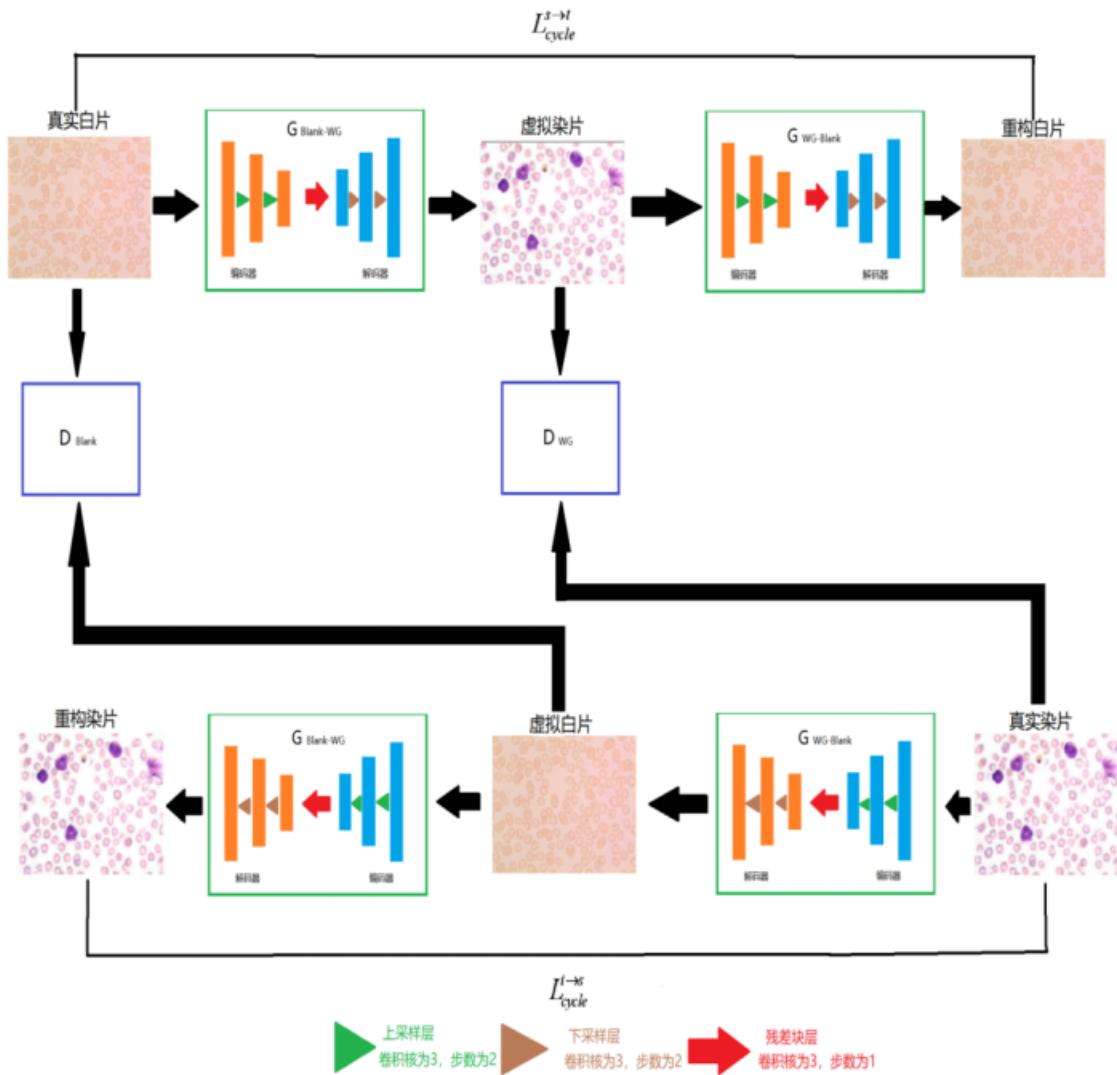


图 3-1 U-GAT-IT 网络框架

U-GAT-IT 的网络架构如下表格 3-1 和表格 3-2 所示。与原始 U-GAT-IT 模型不同的是，由于论文开放的数据集中图像分辨率为  $64 \times 64$ ，然而我们想要使用的数据集分辨率为  $256 \times 256$ ，因此为了适应新的数据集，我人为调整了模型中生成器的结构，令该生成器的编码器和解码器内原本的四个残差块都变为六个残差块，采样层数保持不变。值得一提的是，在更改网络时我将上下采样的层数由二变为三，但生成图像的效果与两层采样层一致，为了减轻网络模型对显存的压力，最终保持了两层采样层不变。对于判别器网络我们采用了两种不同比例的 PatchGAN（即 multi-scale PatchGAN），它将局部( $70 \times 70$ )和全局( $286 \times 286$ )图像进行真假分类。激活函数方面，在生成器中使用 ReLU 函数，判别器中使用的则是斜率为 0.2 的 leaky-ReLU 函数。假设训练迭代次数为原文中设定的 100 万次，那么训练前 50 万代学习率固定为  $1e-4$  ( $10^{-4}$ )，然后在后 50 万代线性衰减。

表格 3-1 U-GAT-IT 生成器架构

模块	输入→输出状态	网络层信息
编码器 下采样层	(h,w,3)→(h,w,64)	CONV-(N64,K7,S1,P3),IN,ReLU
	(h,w,64) → ( $\frac{h}{2}, \frac{w}{2}, 128$ )	CONV-(N128,K3,S2,P1),IN,ReLU
	( $\frac{h}{2}, \frac{w}{2}, 128$ ) → ( $\frac{h}{4}, \frac{w}{4}, 256$ )	CONV-(N256,K3,S2,P1),IN,ReLU
编码器 残差块层	( $\frac{h}{4}, \frac{w}{4}, 256$ ) → ( $\frac{h}{4}, \frac{w}{4}, 256$ )	ResBlock-(N256,K3,S1,P1),IN,ReLU
	( $\frac{h}{4}, \frac{w}{4}, 256$ ) → ( $\frac{h}{4}, \frac{w}{4}, 256$ )	ResBlock-(N256,K3,S1,P1),IN,ReLU
	( $\frac{h}{4}, \frac{w}{4}, 256$ ) → ( $\frac{h}{4}, \frac{w}{4}, 256$ )	ResBlock-(N256,K3,S1,P1),IN,ReLU
	( $\frac{h}{4}, \frac{w}{4}, 256$ ) → ( $\frac{h}{4}, \frac{w}{4}, 256$ )	ResBlock-(N256,K3,S1,P1),IN,ReLU
	( $\frac{h}{4}, \frac{w}{4}, 256$ ) → ( $\frac{h}{4}, \frac{w}{4}, 256$ )	ResBlock-(N256,K3,S1,P1),IN,ReLU
	( $\frac{h}{4}, \frac{w}{4}, 256$ ) → ( $\frac{h}{4}, \frac{w}{4}, 256$ )	ResBlock-(N256,K3,S1,P1),IN,ReLU
生成器的 CAM	( $\frac{h}{4}, \frac{w}{4}, 256$ ) → ( $\frac{h}{4}, \frac{w}{4}, 512$ )	Global Average & Max Pooling, MLP-(NI),Multiply the weights of MLP
	( $\frac{h}{4}, \frac{w}{4}, 512$ ) → ( $\frac{h}{4}, \frac{w}{4}, 256$ )	CONV-(N256,K1,S1),ReLU
$\gamma, \beta$	( $\frac{h}{4}, \frac{w}{4}, 256$ ) → (1,1,256)	MLP-(N256),ReLU
	(1,1,256) → (1,1,256)	MLP-(N256).ReLU
	(1,1,256) → (1,1,256)	MLP-(N256).ReLU
解码器 残差块层	( $\frac{h}{4}, \frac{w}{4}, 256$ ) → ( $\frac{h}{4}, \frac{w}{4}, 256$ )	AdaResBlock-(N256,K3,S1,P1),AdalLN,ReLU
	( $\frac{h}{4}, \frac{w}{4}, 256$ ) → ( $\frac{h}{4}, \frac{w}{4}, 256$ )	AdaResBlock-(N256,K3,S1,P1),AdalLN,ReLU
	( $\frac{h}{4}, \frac{w}{4}, 256$ ) → ( $\frac{h}{4}, \frac{w}{4}, 256$ )	AdaResBlock-(N256,K3,S1,P1),AdalLN,ReLU
	( $\frac{h}{4}, \frac{w}{4}, 256$ ) → ( $\frac{h}{4}, \frac{w}{4}, 256$ )	AdaResBlock-(N256,K3,S1,P1),AdalLN,ReLU
	( $\frac{h}{4}, \frac{w}{4}, 256$ ) → ( $\frac{h}{4}, \frac{w}{4}, 256$ )	AdaResBlock-(N256,K3,S1,P1),AdalLN,ReLU
	( $\frac{h}{4}, \frac{w}{4}, 256$ ) → ( $\frac{h}{4}, \frac{w}{4}, 256$ )	AdaResBlock-(N256,K3,S1,P1),AdalLN,ReLU
解码器 下采样层	( $\frac{h}{4}, \frac{w}{4}, 256$ ) → ( $\frac{h}{2}, \frac{w}{2}, 128$ )	Up-CONV-(N128,K3,S1,P1),LIN,ReLU
	( $\frac{h}{2}, \frac{w}{2}, 128$ ) → (h, w, 64)	Up-CONV-(N64,K3,S1,P1),LIN,ReLU
	(h, w, 64) → (h, w, 3)	CONV-(N3,K7,S1,P3),Tanh

表格 3-2 U-GAT-IT 判别器架构

模块	输入→输出状态	网络层信息
	(h,w,3) → ( $\frac{h}{2}, \frac{w}{2}, 64$ )	CONV-(N64,K4,S2,P1),SN,Leaky-ReLU
	( $\frac{h}{2}, \frac{w}{2}, 64$ ) → ( $\frac{h}{4}, \frac{w}{4}, 128$ )	CONV-(N128,K4,S2,P1),SN,Leaky-ReLU

编码器 下采样层	$(\frac{h}{4}, \frac{w}{4}, 128) \rightarrow (\frac{h}{8}, \frac{w}{8}, 256)$	CONV-(N256,K4,S2,P1),SN,Leaky-ReLU
	$(\frac{h}{8}, \frac{w}{8}, 256) \rightarrow (\frac{h}{8}, \frac{w}{8}, 512)$	CONV-(N512,K4,S1,P1),SN,Leaky-ReLU
判别器的 CAM	$(\frac{h}{8}, \frac{w}{8}, 512) \rightarrow (\frac{h}{8}, \frac{w}{8}, 1024)$	Global Average & Max Pooling, MLP-(N1), Multiply the weights of MLP
	$(\frac{h}{8}, \frac{w}{8}, 1024) \rightarrow (\frac{h}{8}, \frac{w}{8}, 512)$	CONV-(N512,K1,S1),Leaky-ReLU
分类器	$(\frac{h}{8}, \frac{w}{8}, 512) \rightarrow (\frac{h}{8}, \frac{w}{8}, 1)$	CONV-(N1,K4,S1,P1),SN

### 3.3 模型创新

(1) 新的 Attention 机制。

U-GAT-IT 中引入了新的 Attention 模块，能够引导生成器关注那些区分源域与目标域的更重要的区域，从而提高生成器的性能，并让生成器对于图像整体的改变有更好的处理能力。

图 3-2 展示了注意力机制下的类激活映射实现过程，图像经过下采样层和残差块得到的编码特征映射（Encoder Feature map）经过平均池化（Global average pooling）和最大池化（Global max pooling）后，得到两个不同的但都与通道数一致的特征描述。通过解码器隐藏层和编码器隐藏层之间的计算，我们为编码特征映射中的每个通道都分配了一个权重（weight），这个权重代表了通道的重要性，令权重和编码特征映射做对位相乘后做拼接，随后通过一个卷积层变成输入时的通道数，再继续前馈进行 AdaIN 归一化操作。

对于经过全连接层得到的两个  $B \times 1$  维特征向量（B 代表 BatchSize，一次训练选取的样本数），我们把它们做拼接后形成一个二通道的特征向量，再送入辅助分类器，做初始域和目标域的无监督分类。当分类器在注意力模块的协助下能够对源域和目标域的输入作出区分时，就可以令网络模型更好的确定在哪个区间进行集中转换。

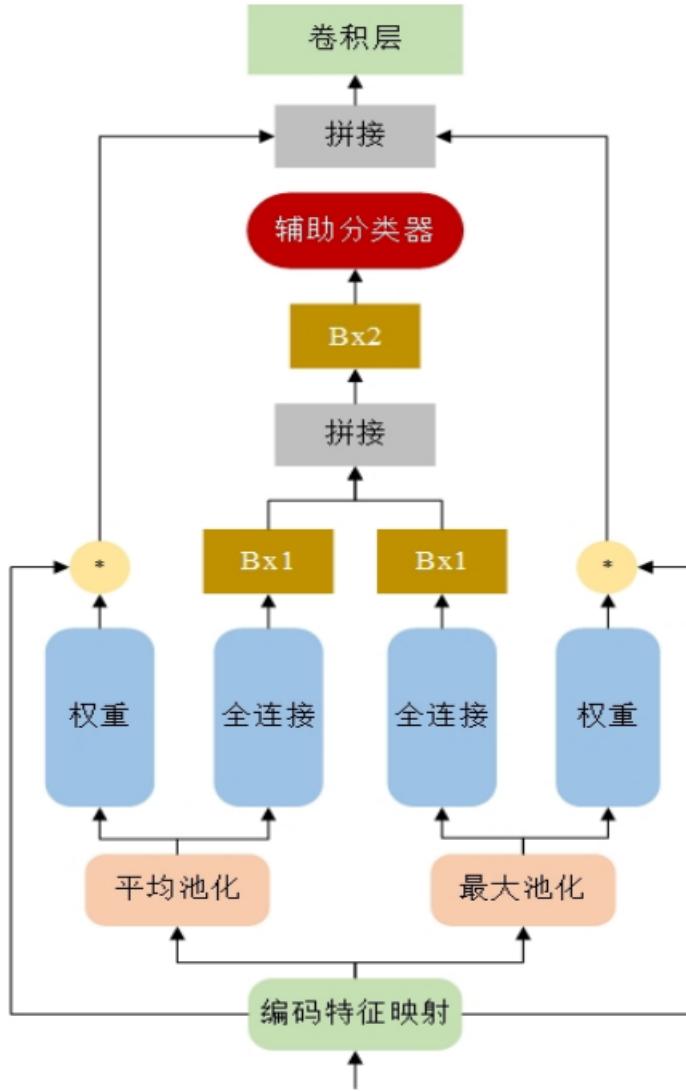


图 3-2 新的 Attention 机制

## (2) 引入新的归一化方式 AdaLIN(自适应 Layer Normalization 与 Instance Normalization)

归一化是深度学习过程中的一项重要技术，它能够加快梯度下降的速度，从而使模型更快的收敛，也就是能够以更少的迭代次数达到同样的正确率。神经网络学习过程中，如果没有对数据进行归一化操作，那么可能会导致网络难以学习到数据的分布情况，因此我们在 U-GAT-IT 模型中添加了归一化操作，与传统 CycleGAN 不同的是，我们对内部模型分别采用了不同的归一化方式，对生成器中的编码器使用 IN、对生成器中解码器的自适应残差层使用 AdaLIN、对生成器中解码器的上采样层使用 LIN、对判别器采用 SN。下面我们主要讨论两种常见的归一化方法以及 U-GAT-IT 模型的创新性归一化处理方法 AdaLIN。

IN 适用于风格化迁移。因为图片生成的结果主要依据某些图像实例（提取图像内部特征的均值和方差代表着图像风格），对整个 Batch 进行归一化操作并不适合图像风

格化迁移的任务，因此在风格迁移中采用 IN 除了能保证图像间的独立性外，还可以加速模型的收敛。实例归一化这种操作能够规范图像的内部特征，也就是说如果我们保持或改变代表图像风格的参数就能成功延续或迁移图像风格。同时我们的方法中还涉及到 AdaIN，这个方法无需学习放射参数，相比 IN 更易计算和实现。

与此同时，BN 大多适用于如图像分类这种的判别模型，因为 BN 需要对每个 Batch 都进行归一化操作，从而保证数据分布的一致性，而数据整体分布恰好决定了判别模型的结果。但是 BN 对 Batch Size（批尺寸）要求严格，因为每次计算方差和均值都是在一个 Batch 上，所以当批尺寸太小时，计算的方差和均值不足以代表整个数据的分布情况。LN 实际是 BN 的一种转化，根据深度网络中某一层的所有神经元进行归一操作，它同层的神经元具有相同的方差和均值，输入样本间存在不同的方差和均值，且 LN 不依赖 Batch 的大小（适用于批尺寸等于 1 时的归一化操作），因此我们选用 LN 作为归一化的其中一个自适应结合目标。

AdaLIN 是 AdaIN 和 LN 的结合，事实上它的设计也保留了 AdaIN 和 LN 的亮点。我们知道 AdaIN 需要保证通道之间的不相关性，并且它仅对图像自身进行归一化，因此它在残差块中会保留较多的风格特征与内容结构，但在上采样层时无法反卷积回残差块中的全局样式，所以会得到相对较差的风格迁移结果。LN 没有设计关于通道相关性的操作而是做了整体归一化，它在上采样层能够更好地实现风格转换，但缺点是在残差块中不能较好的保留特征。这时我们将两者线性结合，通过调整一个参数  $\rho$  的值就能很好的处理不同的任务，AdaLIN 表达式如下：

$$AdaLIN(\alpha, \gamma, \beta) = \gamma \cdot (\rho \cdot \alpha_I + (1 - \rho) \cdot \alpha_L) + \beta \quad (3-1)$$

其中  $\alpha_I = \frac{\alpha - \mu_I}{\sqrt{\sigma_I^2 + \epsilon}}$ ， $\alpha_L = \frac{\alpha - \mu_L}{\sqrt{\sigma_L^2 + \epsilon}}$ ，且  $\mu_L$ 、 $\sigma_L^2$ 、 $\mu_I$ 、 $\sigma_I^2$  分别对应 LN 和 IN 的均值

和方差， $\gamma$  和  $\beta$  是在全连接层计算得到的动态参数。

值得注意的是，在 U-GAT-IT 模型中只在解码器中采用了 AdaLIN，而在编码器中使用的是 IN，辅助分类器使用的则是 LN。原因在于解码器中的残差块可以融合特征，上采样块则负责图像合成，利用 AdaLIN 能够保留更多结构和特征的优势可以生成更具风格化的图像，但 AdaLIN 在编码器中实现的意义不大，反而会令网络架构更加复杂，易在训练网络时出现显存不足等问题。

### 3.4 模型实现过程

我们利用 U-GAT-IT 单方向上的生成对抗网络说明其实现过程，即从空白域到瑞氏吉姆萨染色域（实际上 U-GAT-IT 还有一个反向的网络，与正向的网络模型架构相同）

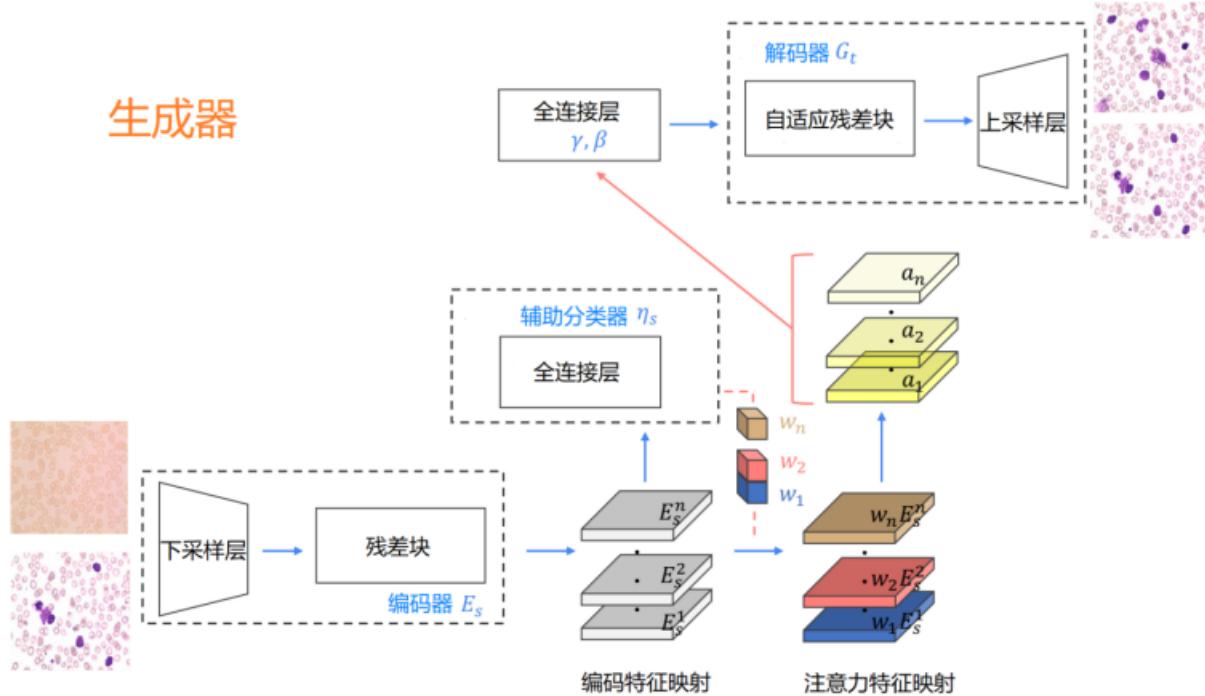


图 3-3 U-GAT-IT 生成器结构

生成器模型实现过程：

(1) 分别输入源域 source (空白域) 图像  $x \sim X_{Blank}$  与目标域 target (瑞氏吉姆萨染色域) 图像  $x \sim X_{Wright-Giemsa}$ 。

(2) 经过编码器  $E_s$  (两次下采样与六个残差块) 得到输出的特征：

$$E_s(x) = [E_s^1 E_s^2 E_s^3 \dots E_s^n]$$

(3) 经过 GAP 和 GMP，再经过辅助分类器  $\eta_s$  进行分类 (判断初始域和目标域)，利用创建的权重  $w_s = \{w_s^1, w_s^2, \dots, w_s^n\}$  作为注意力参数，可以得到引入注意力参数后的特征：

$$d_s^k = w_s^k E_s^k \quad (3-2)$$

注意力特征映射经过激活层变为激活映射  $a = [E_s^1 E_s^2 E_s^3 \dots E_s^n]$ 。

(4) 将黄色方块代表的激活映射继续前馈，通过 GAP 和 GMP 后进入全连接层，预测  $2c$  个参数， $c$  代表特征映射的通道数，且  $c$  个是作为缩放因子  $\gamma$ ， $c$  个是作为偏移量  $\beta$ ，用于对后面的特征做 AdaLIN 正则化：

① 分别通过 IN 和 LN 求出均值  $(\mu_I, \mu_L)$  和标准差  $(\sigma_I, \sigma_L)$ ，令特征标准化：

$$\hat{a}_I = \frac{a - \mu_I}{\sqrt{\sigma_I^2 + \epsilon}} \quad (3-3)$$

$$\hat{a}_L = \frac{a - \mu_L}{\sqrt{\sigma_L^2 + \epsilon}} \quad (3-4)$$

② 加权求和，

$$AdaLIN(\alpha, \gamma, \beta) = \gamma \cdot (\rho \cdot \hat{a}_I + (1-\rho) \cdot \hat{a}_L) + \beta \quad (3-5)$$

其中  $\rho$  为学习权重。

(5) 继续前馈正则化后的特征，通过解码器  $G_t$ （六个残差块和两次上采样）得到输出。

## 判别器

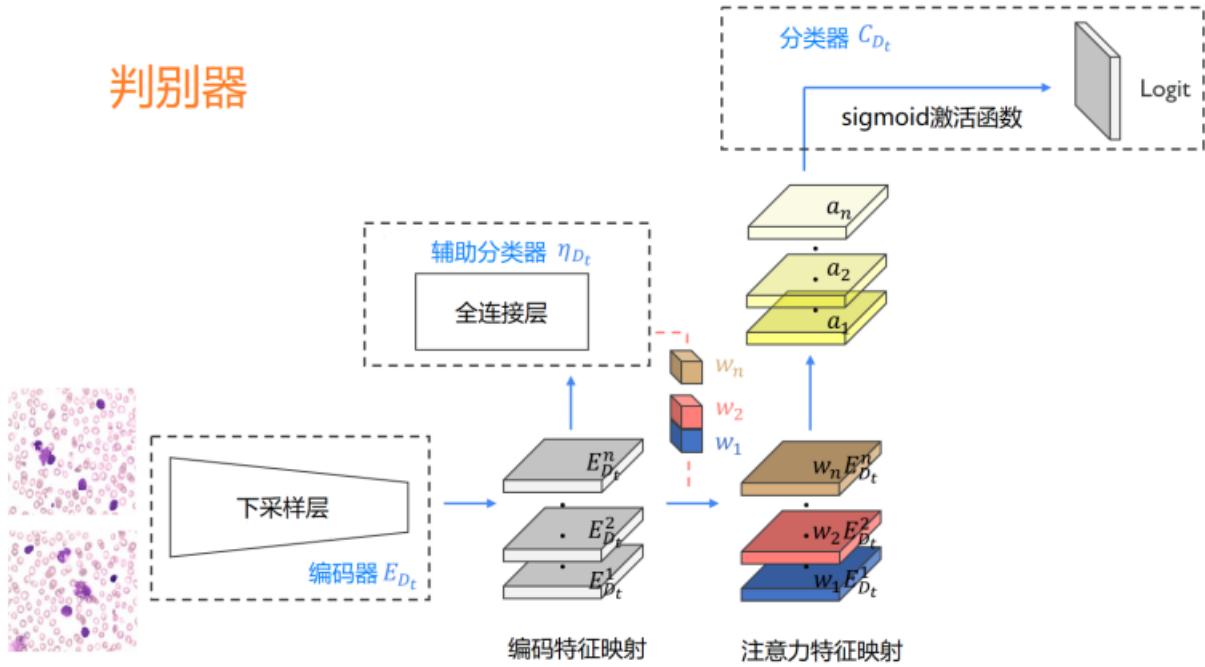


图 3-4 U-GAT-IT 判别器结构

判别器模型实现过程：

- (1) 分别输入生成器生成的图像  $x \sim G_{Blank-WG}$  与目标域（瑞氏吉姆萨染色域）图像  $x \sim X_{Wright-Giemsa}$ 。
- (2) 经过编码器  $E_{D_t}$  得到输出的特征  $E_{D_t}(x) = [E_{D_t}^1 \ E_{D_t}^2 \ E_{D_t}^3 \dots E_{D_t}^n]$ 。
- (3) 经过 GAP 和 GMP，再经过辅助分类器  $\eta_{D_t}$  进行分类（判断初始域和目标域），利用创建的权重  $w_{D_t} = \{w_{D_t}^1, w_{D_t}^2, \dots, w_{D_t}^n\}$  作为注意力参数，可以得到引入注意力参数后的特征：

$$a_{D_t}^n = w_{D_t}^n E_{D_t}^n \quad (3-6)$$

注意力特征映射经过激活层变为激活映射  $a = [E_{D_t}^1 \ E_{D_t}^2 \ E_{D_t}^3 \dots E_{D_t}^n]$ 。

- (5) 将黄色方块代表的激活映射继续前馈进入分类器  $C_{D_t}$ （经过 sigmoid 激活函数进入 Logit 模块）。

### 3.5 双向损失函数

#### (1) GAN 的对抗损失 (GAN Loss)

U-GAT-IT 的对抗损失中并没有使用传统 CycleGAN 中交叉熵 log 误差损失函数, 而是采用了 MSE 函数。

$$L_{GAN}^{s \rightarrow t} = E_{x \sim X_{Blank}} \left[ (D_t(x))^2 \right] + E_{x \sim X_{Wright-Giemsa}} \left[ (1 - D_t(G_{s \rightarrow t}(x)))^2 \right] \quad (3-7)$$

$$L_{GAN}^{t \rightarrow s} = E_{x \sim X_{Wright-Giemsa}} \left[ (D_s(x))^2 \right] + E_{x \sim X_{Blank}} \left[ (1 - D_s(G_{t \rightarrow s}(x)))^2 \right] \quad (3-8)$$

#### (2) 循环一致性损失 (Cycle Loss)

$$L_{cycle}^{s \rightarrow t} = E_{x \sim X_{Blank}} \left[ \|x - G_{t \rightarrow s}(G_{s \rightarrow t}(x))\| \right] \quad (3-9)$$

$$L_{cycle}^{t \rightarrow s} = E_{x \sim X_{Wright-Giemsa}} \left[ \|x - G_{s \rightarrow t}(G_{t \rightarrow s}(x))\| \right] \quad (3-10)$$

#### (3) 身份损失 (Identity Loss)

身份损失函数的作用在于保证生成器能够正常生成图像风格及色调。我们想要生成目标图像风格, 就要保证在生成器输入一个目标风格的图像时, 输出能保持不变。然而事实上, 如果不添加身份损失, 模型中的生成器往往会改变图像的颜色, 产生不理想的视觉效果。

$$L_{identity}^{s \rightarrow t} = E_{x \sim X_{Wright-Giemsa}} \left[ \|x - G_{s \rightarrow t}(G_{t \rightarrow s}(x))\| \right] \quad (3-11)$$

$$L_{identity}^{t \rightarrow s} = E_{x \sim X_{Blank}} \left[ \|x - G_{t \rightarrow s}(G_{s \rightarrow t}(x))\| \right] \quad (3-12)$$

#### (4) CAM 损失 (CAM Loss)

CAM 损失函数分为两种, 分别对应模型中的生成器和判别器。它们的主要功能是训练各自的辅助分类器, 进而提升注意力机制的表现。生成器中 CAM 损失的作用主要是将图像域分类, 令空白域和瑞氏吉姆萨染色域完全分开。判别器中 CAM 损失的作用是方便在注意力特征映射上进一步区分真实图像和生成图像, 从而获得最终的目标函数。

生成器部分的 CAM 损失函数:

$$L_{cam}^{G_{s \rightarrow t}} = - \left( E_{x \sim X_{Blank}} \left[ \log(\eta_s(x)) \right] + E_{x \sim X_{Wright-Giemsa}} \left[ \log(1 - \eta_{ta}(x)) \right] \right) \quad (3-13)$$

$$L_{cam}^{G_{t \rightarrow s}} = - \left( E_{x \sim X_{Wright-Giemsa}} \left[ \log(\eta_t(x)) \right] + E_{x \sim X_{Blank}} \left[ \log(1 - \eta_s(x)) \right] \right) \quad (3-14)$$

判别器部分的 CAM 损失函数：

$$L_{cam}^{D_s} = E_{x \sim X_{w r i g h t-G i e m s a}} \left[ (\eta_{D_s}(x))^2 \right] + E_{x \sim X_{B l a n k}} \left[ \log \left( 1 - \eta_{D_s}(G_{s \rightarrow t}(x)) \right)^2 \right] \quad (3-15)$$

$$L_{cam}^{D_t} = E_{x \sim X_{B l a n k}} \left[ (\eta_{D_t}(x))^2 \right] + E_{x \sim X_{w r i g h t-G i e m s a}} \left[ \log \left( 1 - \eta_{D_t}(G_{t \rightarrow s}(x)) \right)^2 \right] \quad (3-16)$$

根据上述 4 个损失函数，通过加权求和，我们可以得到总的损失函数：

$$L_{u g a t t}^{t \rightarrow s} = \min_{G_{s \rightarrow t}, \eta_s, G_{t \rightarrow s}, \eta_t} \max_{D_s, \eta_{D_s}, D_t, \eta_{D_t}} \lambda_1 L_{G A N}^{t \rightarrow s} + \lambda_2 L_{c y c l e}^{t \rightarrow s} + \lambda_3 L_{i d e n t i t y}^{t \rightarrow s} + \lambda_4 \left( L_{c o m}^{G_{t \rightarrow s}} + L_{c o m}^{D_s} \right) \quad (3-17)$$

## 第4章 数据集与实现

### 4.1 数据集

在本工作中，我们从河南省人民医院的血液科采集了 23 张白片（未染色图像）和 37 张染片（化学染色后的图像），病理科主任批准本研究使用样本。样本是匿名的，所有与病人相关的数据和唯一的标识符都被删除。

这些采集到的数据集具有相同的图像风格和相似的细胞分布结构，但没有像素级的匹配。每张图像均是将切片放置在  $\times 40$  物镜下，利用电子显微镜拍摄到的当前视野下的切片区域，图像的像素为 2560\*1920。我们留下 6 张白片作为测试集，其余图像（17 张白片，37 张染片）作为训练集放入各个模型中训练。

在片段染色过程中，我们将原始白片及染片裁剪为等分的 70 张细胞片段，每张图像的像素为 256\*256，裁剪过后，我们的数据集中有 1190 张白片和 2590 张染片；在整体染色时，由于内存限制，我们在保证图像感受野不变的情况下，人为地降低了图像的分辨率，同时在数据增强方面我们利用几何变换法，将原图像在正方向上顺时针旋转三次，并记录每次旋转后的图像，因此在整体染色时我们的数据集的是最初的四倍，即有 68 张白片和 148 张染片。

### 4.2 实施细节

此网络在一台 IntelCorei9-9900k CPU、32GB RAM 和一个 NVIDIA GTX1080-Ti GPU 的计算机上训练和测试结果。我们在 Win10  $\times 86\_64$  系统下利用 Anaconda3 创建虚拟环境，使用 Python3.7 解释器、CUDA10.2 版本并搭配 cuDNN v7.6.5，安装配置必要的深度学习包（torch1.2.1、torchvision0.7.1、cv2、numpy 等等），在 Pycharm 平台进行调试及开发。

卷积层中的参数与 J Kim 等人在 U-GAT-IT 模型论文中的设定相同，并根据这些参数初始化网络。不同的是，为了适应数据集，我将 BatchSize 大小设置为 1，最初的学习速率设置为 0.0001，并使用指数衰减，衰减率为 0.9，衰减期为 2，我们通常将迭代次数设置为 3 万次。

## 第5章 结果与评价

### 5.1 染色结果

#### 5.1.1 U-GAT-IT模型染色结果

(1) 片段染色结果

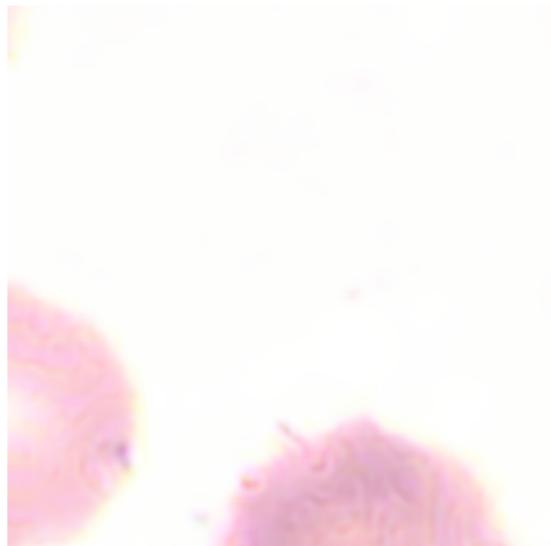


图 5-1 片段染色图 1

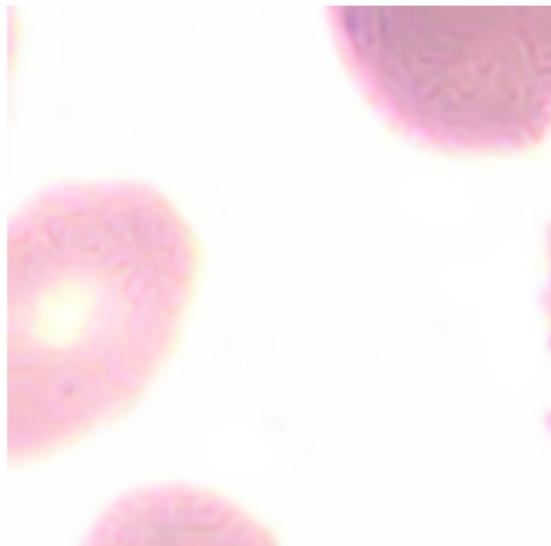


图 5-2 片段染色图 2

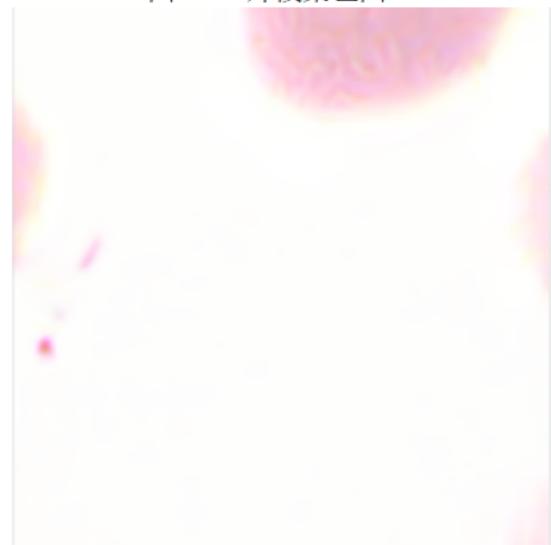


图 5-3 片段染色图 3

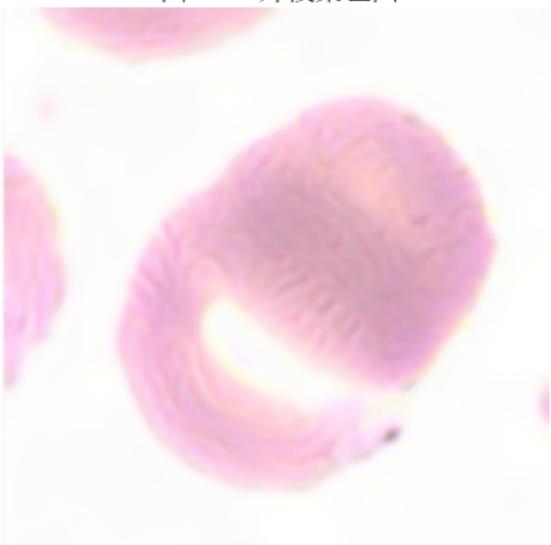


图 5-4 片段染色图 4

(2) 片段染色过程及热力图

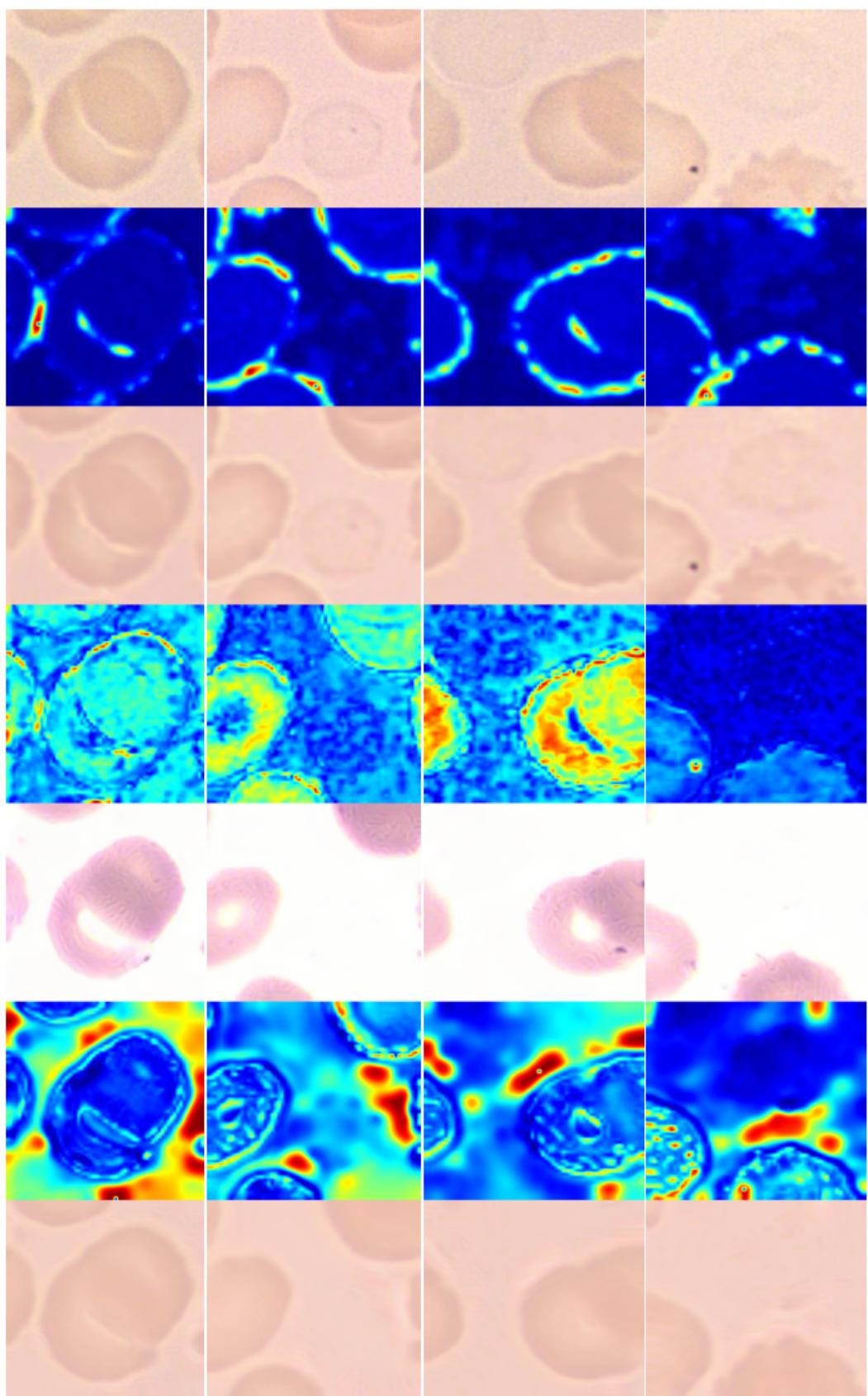


图 5-5 片段染色过程

图 5-5 中显示的染色过程分为七行四列，我们代入了四张裁剪后的白片图像，所有的图像像素均为 256\*256。第一二行展示的分别是原始白片及其对应的热力图，第三四行展示的是像素级对应的重塑白片图像及其热力图，第五六行展示的是虚拟细胞片段染色图像及其热力图，最后一行展示的则是 CycleGAN 的另一个方向的网络合成结果，即从染色图像转为未染色图像。

### (3) 整体染色结果

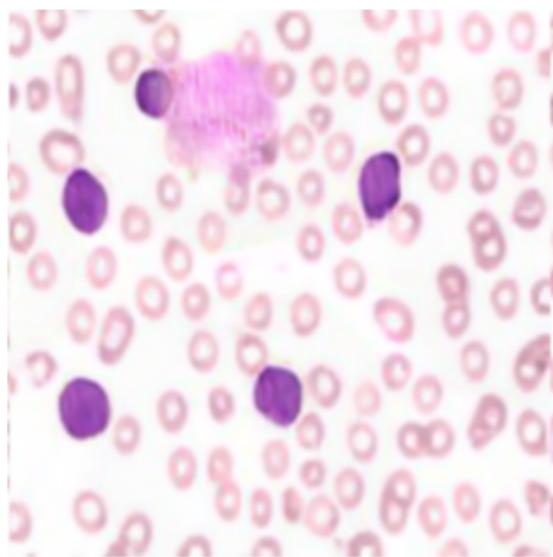


图 5-6 整体染色图 1

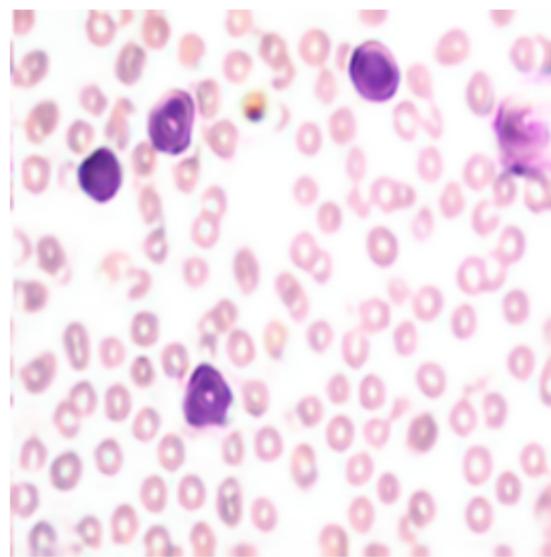


图 5-7 整体染色图 2

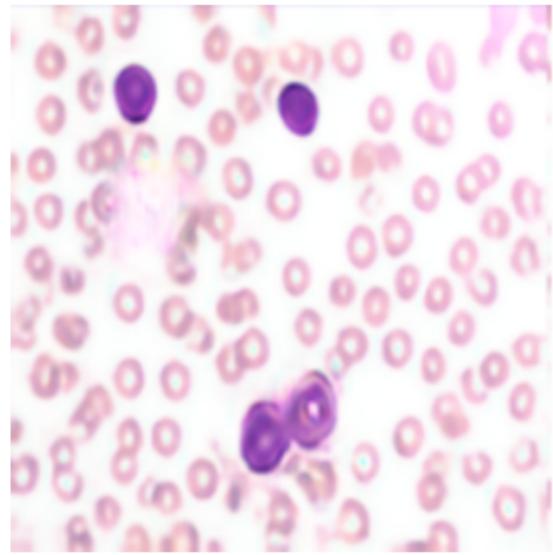


图 5-8 整体染色图 3

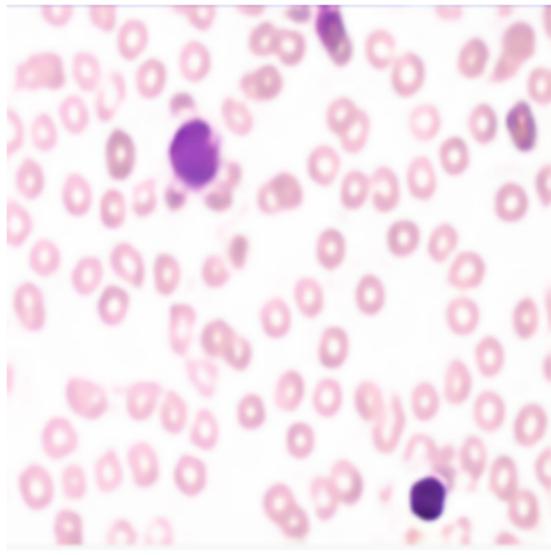


图 5-9 整体染色图 4

(4) 整体染色过程及热力图

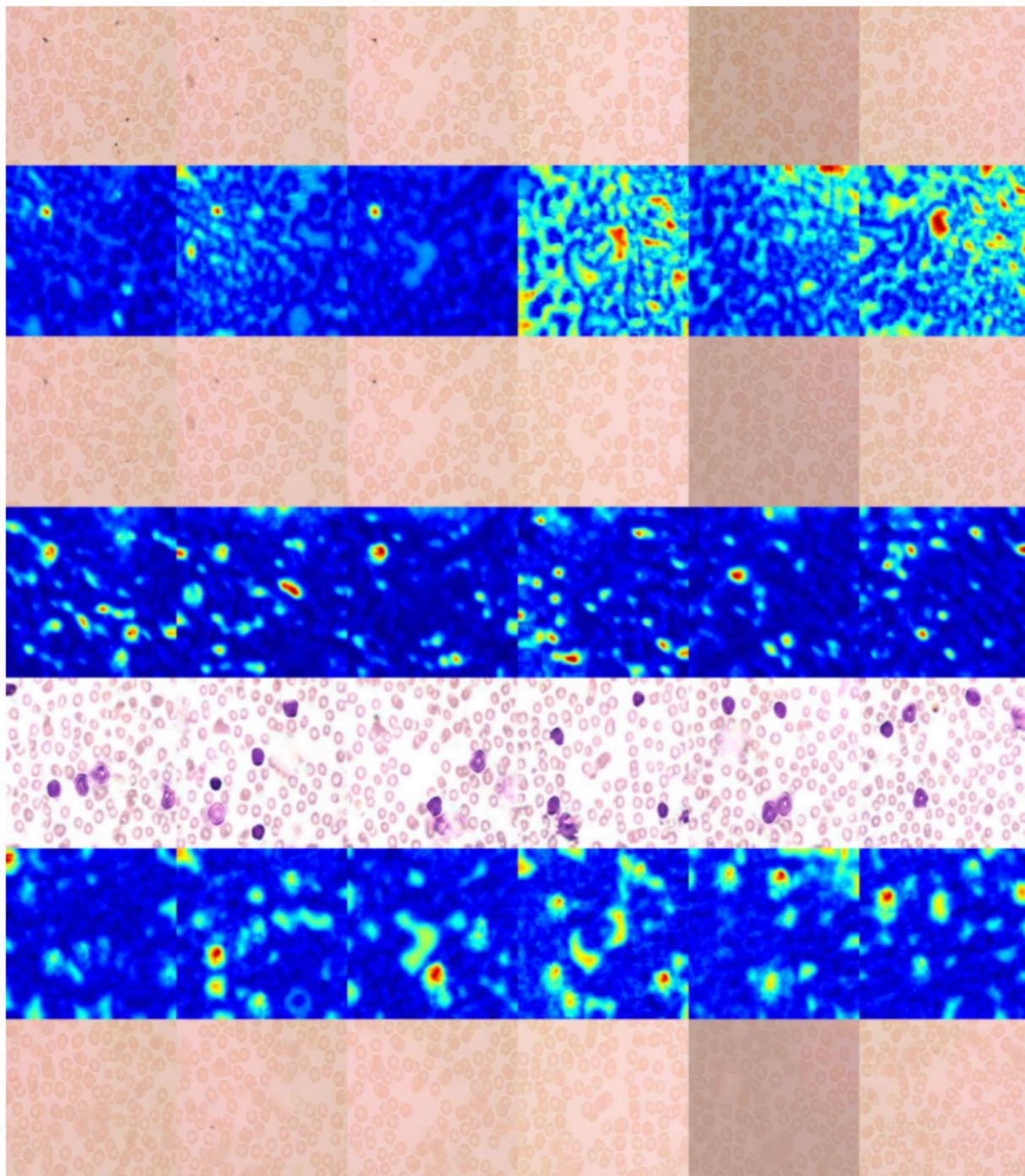


图 5-10 由白片到染色的整体染色过程

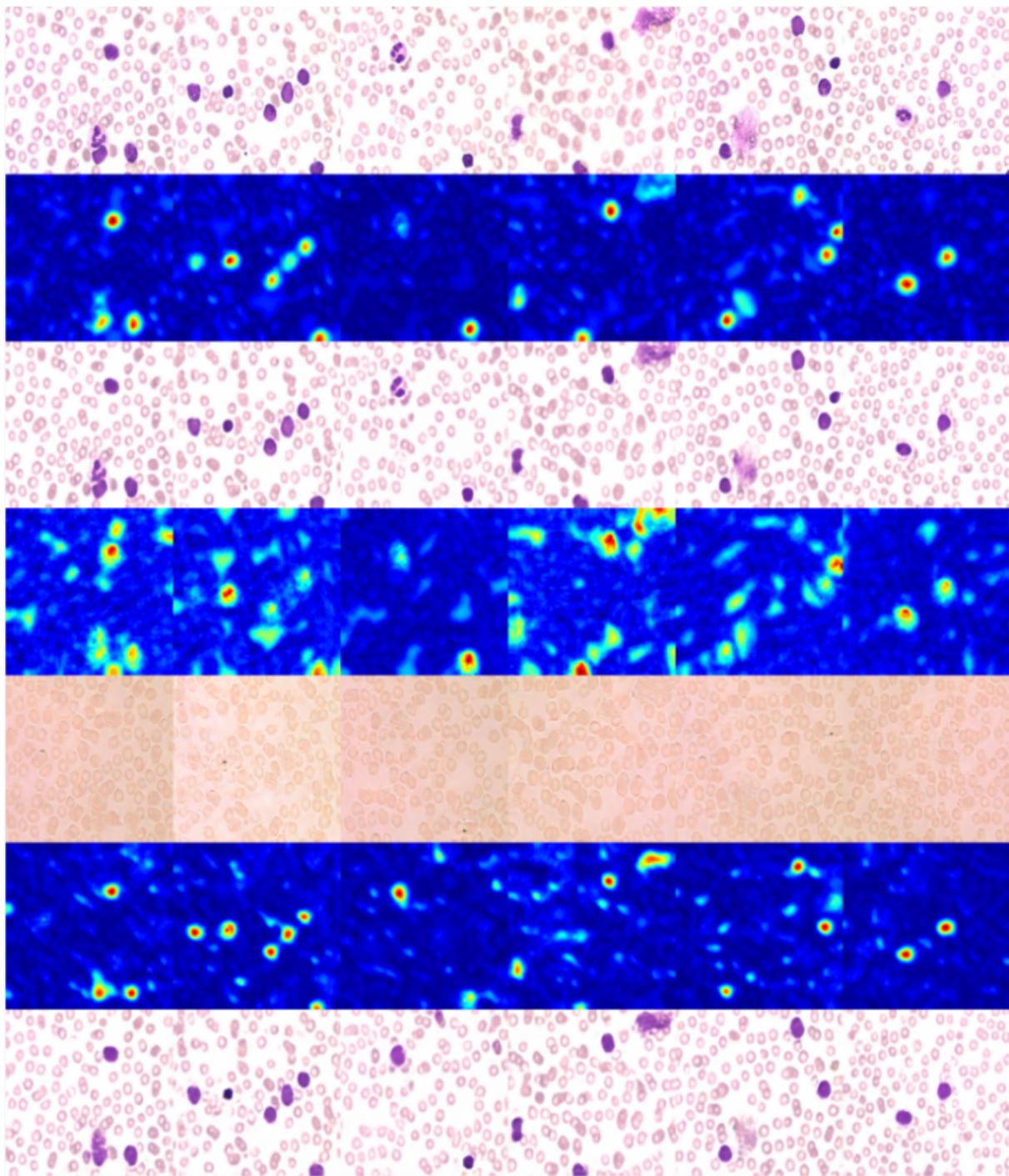


图 5-11 由染片到白片的整体染色过程

图 5-10 和图 5-11 分别展示了白片生成染片和染片重新转回白片的染色过程，类似片段染色过程，区别是我们代入训练的是六张不同的整体白片和染片图像，并且在图 5-10 中第五列我们选取了一张视野偏暗的白片作为输入，来观察染色后可能出现的变化。

## 5.1.2 染色结果对比

### (1) 片段染色的对比

我们将同一个细胞片段放入四种不同的网络模型（CycleGAN、BayesGAN、AttentionGAN、U-GAT-IT）中，观察他们之间染色的差异。

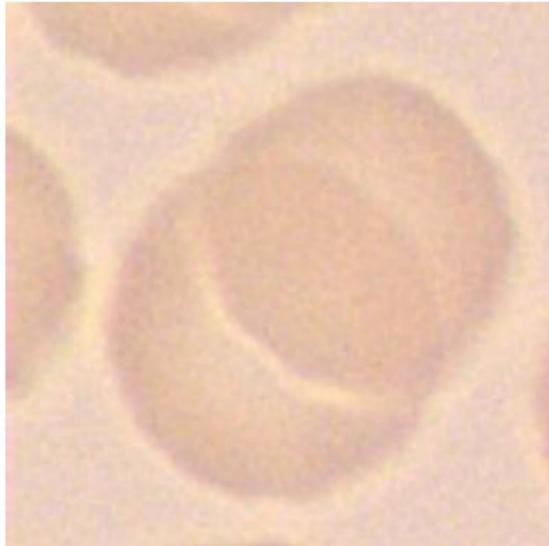


图 5-12 原始白片

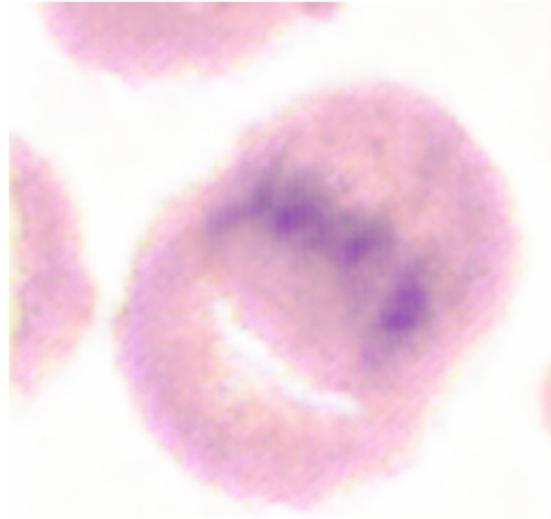


图 5-13 BayesGAN 片段染色

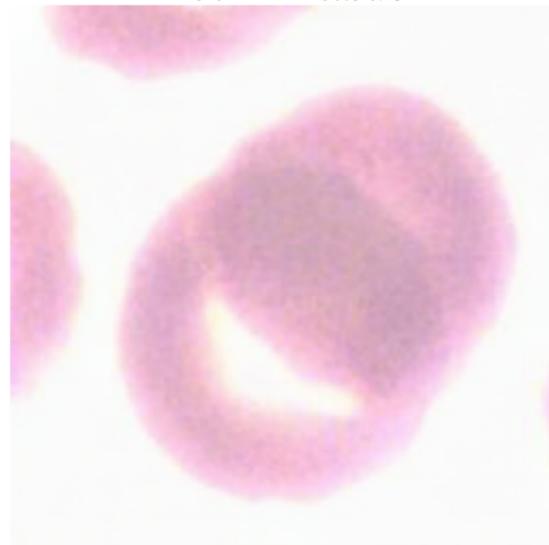


图 5-14 CycleGAN 片段染色



图 5-15 AttentionGAN 片段染色

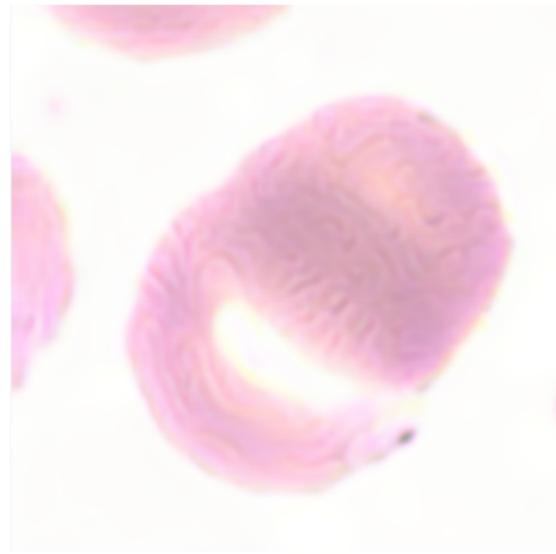


图 5-16 U-GAT-IT 片段染色

## (2) 整体染色的对比

目前由于数据集的局限性，我们将相同的细胞白片放入四种不同的网络模型(Cyclegan、AttentionGAN、Pix2pix、U-GAT-IT)中，观察他们之间染色风格的差异。

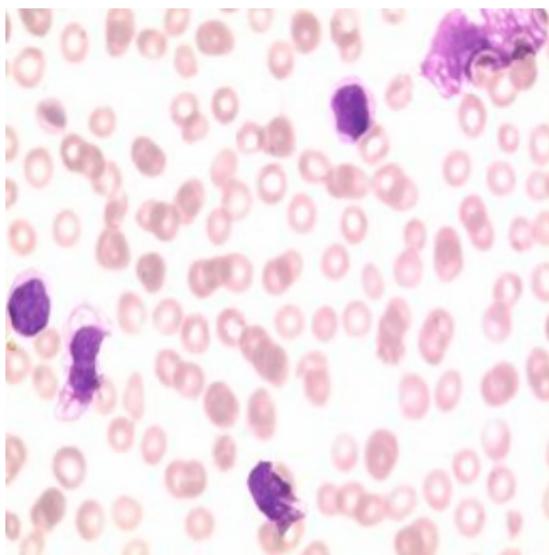


图 5-17 真实化学染色

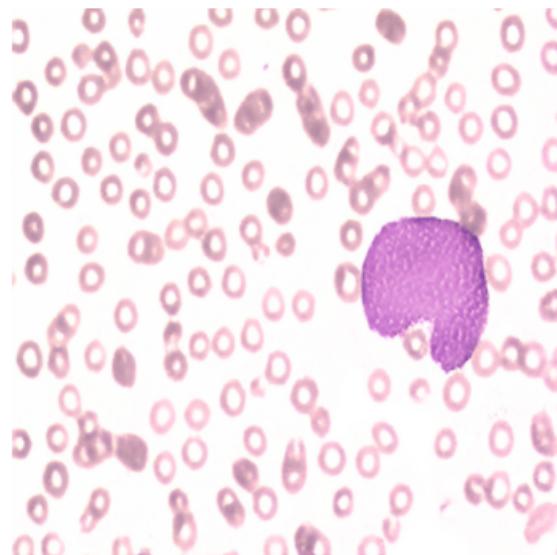


图 5-18 CycleGAN 整体染色

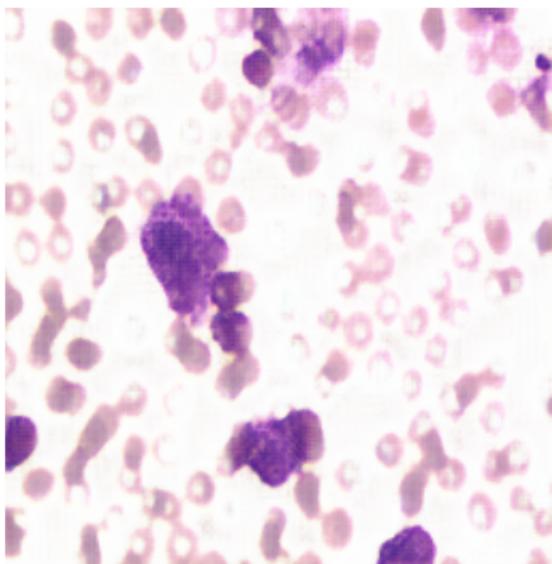


图 5-19 AttentionGAN 整体染色

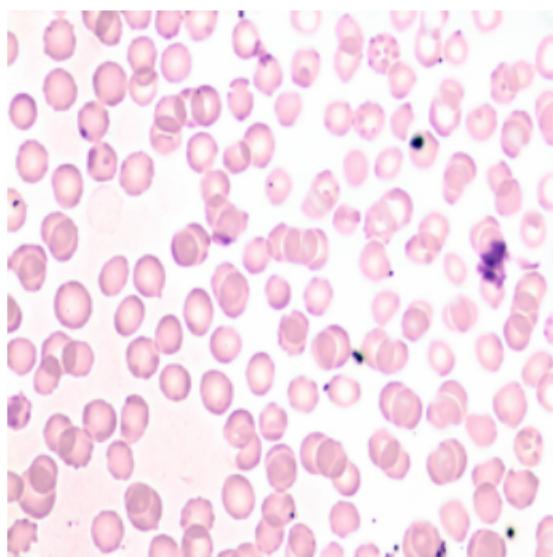


图 5-20 Pix2pix 整体染色

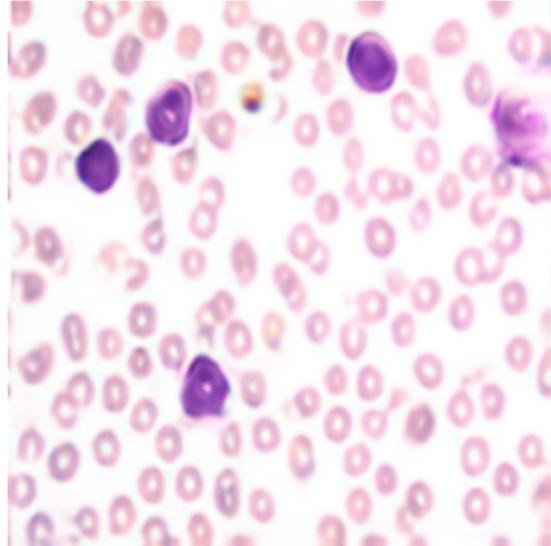


图 5-21 U-GAT-IT 整体染色

### 5.1.3 染色结果分析

我们从染色结果中的片段染色可以看出，U-GAT-IT 与 CycleGAN 的效果类似，两个细胞的重叠处会显示有更深的颜色，将真实化学染色图像放在显微镜下观察时，我们同样也会看到这样的现象。对于 BayesGAN，我们观察到它的细胞间存在一些紫色，这在实际染色中是不会存在的，因为染色中只有淋巴细胞的胞浆和细胞核蛋白与碱性染料相结合才会产生蓝紫色，所以正常的染色结果应该是对于细胞整体的染色，即整个细胞呈现蓝紫色而非只有一小部分呈现。对于 AttentionGAN，我们看到它的细胞重叠处有颜色加深，但细胞外轮廓呈现淡黄色，可能是其迭代次数太少的原因。

对整体染色效果进行分析，首先我们可以注意到 AttentionGAN 和 Pix2pix 的效果是不尽人意的。对于 Pix2pix 网络模型，他的要求是需要一一对应的数据集，即保证白片中细胞的位置与染片中细胞的位置不能发生改变，在我们现有条件下无法获取这样的细胞数据集，因此我采用的方法是将染色图像还原成灰度图，将灰度图作为生成器的输入，化学染色图作为判别器的输入，实现最终的风格转变。得到结果后，我们发现图像中染色痕迹很少，分析原因可能是还原成灰度图之后，原化学染色图像中的蓝紫色细胞会被还原成黑色，那么这时训练就会变成黑色块变为蓝紫色染色块的学习过程，然而实际上我们的白片数据集中会存在黑色杂质，因此模型会将这些杂质训练为蓝紫色的染块，如图 5-19。对于 CycleGAN 的训练结果，我们可以看到其白细胞的清晰度、胞间的细节都是最逼近于真实染色图像的，但它的缺陷就是染色结果的偏差较为严重，在真实染色中，是会存在图 5-18 中的大面积染色的，这也是化学染色中较为常见的误染现象，但正常情况下，在这样的视野区域中即使是正常人也是一定会存在淋巴细胞的，所以图中这种只有误染而其他细胞根本没有被染色的情况是在现实中几乎不存在的，但不可否认的是，CycleGAN 作为一个最基本、使用最广泛的网络模型在虚拟染色领域的表现十分良好，特别是在图像重建方面。最后是对于 U-GAT-IT 模型的结果，通过人体视觉系统可以发现，在相同的像素下，U-GAT-IT 的结果是最接近原染色图像的，这也意味着随着 U-GAT-IT 更高分辨率版本的发展，它的表现可能会更加突出。从图 5-21 中我们可以看到，U-GAT-IT 的染色结果是针对于整个细胞的，这符合化学染色的规律，我们可以注意到，在图像的右上角有一块颜色偏浅的误染现象，这是因为我们在训练整体染色的过程中的数据集较少且数据集中存在着大量误染现象（几乎是每一张图像上均有 1-2 处大面积误染），因此 U-GAT-IT 的学习过程中会学习到这些特征，从而造成了虚拟误染现象。总体来说，U-GAT-IT 和 CycleGAN 在虚拟染色领域的效果都是相对较好的，但在低分辨率下，U-GAT-IT 的视觉和染色效果更好，几乎可以达到以假乱真的程度。

## 5.2 评价指标

在以 cGAN 为代表的深度网络结果测试中，生成图像的质量评价是至关重要的。图像质量评价方式通过被分为两种：主观质量评价和客观质量评价。主观质量评价主要是由人根据个人经验和感受对看到的图像进行评价。客观质量评价则是在图像相似度等方面，利用多种评估函数所得的输出值来表示图像的质量。本研究是医学图像的生成，我

们需要让网络所生成的图像具备可用性。因此本文的评价指标主要为客观上的结构相似性和峰值信噪比，分别评价网络学习的结果。

我通过三种不同的评估指标来评估和比较我们的方法，即使用结构相似指数（SSIM）、峰值信噪比(PSNR)和均方误差(MSE) 来衡量网络生成的瑞氏吉姆萨染色图像与真实的化学染色图像之间的差异，再与 CycleGAN、Pix2pix、BayesGAN、AttentionGAN 这四种 GAN 模型的生成结果进行对比，从而突出 U-GAT-IT 的优越性。

上述三种评估函数的简单介绍：

### (1) 结构相似性指数 SSIM

SSIM (structural similarity index) 是一种判断两张图像相似度的指标，是目前国际上使用较为广泛的评估函数，本次客观评价环节也主要以 SSIM 值作为主要依据来评估 U-GAT-IT 的生成效果。

结构相似性的主要观点是图像之间存在结构性，相邻的像素间有强相关性，其相关性反映了图像的结构信息，而我们人类的视觉认知中恰恰会在观看图片时选择提取图片中这样的结构信息，因此 SSIM 也是最能契合我们视觉系统的评估函数。SSIM 值的范围是 0-1，越接近 1 说明生成图像的质量越接近于原图质量。

现在设置两个图像  $x, y$ ，那么两者的结构相似性就被定义为：

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma \quad (5-1)$$

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (5-2)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (5-3)$$

$$s(x, y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (5-4)$$

其中  $l(x, y)$  代表亮度， $c(x, y)$  代表对比度， $s(x, y)$  代表结构度， $\mu_x, \mu_y, \sigma_x, \sigma_y$  分别代表  $x$  和  $y$  的平均值和标准差， $\sigma_{xy}$  代表协方差。

### (2) 峰值信噪比 PSNR

PSNR (Peak signal-to-noise ratio) 是一个表示信号最大可能达到的功率峰值与信号内部噪声功率的比值的术语，目前作为一种测量方法广泛用于图像重建领域。同样峰值

信噪比也是世界上使用最普遍、最广泛的图像质量客观评价方法，但 PSNR 的值可能与人眼视觉系统看到的图像质量不一致，会出现 PSNR 值更高的图像看起来比值低的图像质量更差。PSNR 的值越大，代表图像失真度越低，也就意味着图像质量越好。

PSNR 常常利用两张图像的均方误差 MSE 进行定义：

$$PSNR = 10 \times \log_{10} \left( \frac{(2^n - 1)^2}{MSE} \right) \quad (5-5)$$

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \| I(i, j) - K(i, j) \|^2 \quad (5-6)$$

### (3) 均方误差 MSE

MSE(Mean Square Error)指参数估计值和参数真实值之差的平方的期望值，在图像处理领域中，它代表的就是生成前后图像像素值之差平方和的均值，反映的是图像间的差别，是一种像素级误差图像质量评价指标，用于衡量生成图像和真实图像之间的差异。MSE 的值越小，表示生成图像的质量越好。

MSE 常被定义为：

$$MSE = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (f'(i, j) - f(i, j))^2 \quad (5-7)$$

其中  $f(x, y)$  和  $f'(x, y)$  分别代表的是真实图像和生成图像，M 和 N 代表图像的长和宽。

在客观评价环节中，我会从五类 GAN 的生成图像中选取一至三张最有代表性且生成效果最好的图像，拿这些图像与真实化学染色的图像进行对比，最终得到所有的评估值。值得注意的是，BayesGAN 的生成结果是片段染色（训练过程中为了凸显细节，我们将一张原图等分裁剪成了 70 张 256\*256 的片段染色图像，这时训练出的生成图像也是同样尺寸的细胞片段），因此我会将真实染片切分为同样的尺寸，同时找到与生成图像相对应的图像，将这两张图像放入评估函数中。对于 Pix2pix、AttentionGAN 这两类 GAN，我们的生成结果是一整张图，像素为 256\*256，在评估过程中我们同样将真实染色图像尺寸调整为 256\*256，然后代入评估函数即可。对于本文中主要研究的 U-GAT-IT 方法以及最常见的 CycleGAN 方法，我会采用片段染和整体染的方法，先将片段染代入评估函数的值与 CycleGAN 和 BayesGAN 所得的值进行对比，得到表格 5.1，再把整体

染代入函数中，将得到的评估值与 Pix2pix 和 AttentionGAN 所得的值做对比，得到表格 5.2。

表格 5-1 CycleGan、BayesGAN、AttentionGAN 和 U-GAT-IT 片段染色相似性评估值

	CycleGAN	BayesGAN	AttentionGAN	U-GAT-IT
SSIM	0.8842	0.8710	0.8501	0.8908
PSNR	21.2696	21.1338	24.9985	20.4557
MSE	485.4167	585.4776	621.6693	500.8423

表格 5-2 CycleGan、Pix2pix、AttentionGAN 和 U-GAT-IT 整体染色相似性评估值

	CycleGAN	Pix2pix	AttentionGAN	U-GAT-IT
SSIM	0.3223	0.3429	0.3769	0.3844
PSNR	16.0485	17.0558	15.1259	16.5754
MSE	1615.1911	1280.8337	1997.4745	1430.6721

### 5.3 评价结果

从客观评价结果中我们可以看到，U-GAT-IT 在片段染色时的效果与 CycleGAN 几乎相同（PSNR 相差小于 1、MSE 相差小于 20）且在 SSIM 指数方面优于 CycleGAN，同样，U-GAT-IT 在两项函数（SSIM、MSE）中表现均优于 BayesGAN 以及 AttentionGAN，并且 SSIM 的值相比 BayesGAN 提高了约 0.02，相比 AttentionGAN 提高了约 0.04，证明了其在细胞染色细节处理方面具有一定的优越性。

在整体染色相似性评估中，U-GAT-IT 展示了相当良好的染色结果，其中 SSIM 指数高于其余三种 GAN 的评估值，且在 PSNR 和 MSE 方面的表现均处于第二位，仅次于像素级匹配染色法 Pix2pix（一般来说会将图片还原至像素级，但对数据集要求较高，需要一一对应的数据集，因此不易实现）。

我们需要注意的是，在 U-GAT-IT 片段染色时，我们的三种评估函数的值表现均优于整体染色，SSIM 值从片段染色时的 0.8908 降为 0.3844，PSNR 值从 20.4557 降为 16.5754，MSE 则从 500.8423 升至 1430.6721，其他 GAN 模型的评估值也存在着相同的问题。

## 第6章 总结与展望

### 6.1 总结

我们提出了一种新的基于非配对数据集的对抗性学习方法，利用一组全新的外周血细胞数据集实现了从未染色域到瑞氏吉姆萨染色域之间的转换。我们的评估实验表明，U-GAT-IT 作为一个新的 GAN 模型在风格转换和细节处理上有一定的优越性，并且更贴合人类视觉系统，能够为医师后续的病情诊断提供帮助。

从本次的课题研究中我们可以得到以下结论：

(1) 循环生成对抗网络能够很好地学习未染色病理切片和瑞氏吉姆萨染色切片之间的非线性关系，并且在实现风格转换的情况下它所需的数据集是要求最低的。基于 CycleGAN 的无监督学习在图像处理领域备受关注，类似于 U-GAT-IT 模型的各种增强版 CycleGAN 也不断在 CVPR 等国际会议上被提出，因此各种新型的无监督学习模型可能会在今后成为虚拟染色领域的主流。

(2) 在损失函数使用环节，我们除了选用 CycleGAN 中使用的常见损失函数，还添加了 CAM 损失函数，通过不断优化生成器和判别器中的辅助分类器，达到令人满意的收敛效果（相比与 CycleGAN 需要经历总迭代数的  $2/3$  才能最佳收敛效果，U-GAT-IT 仅需要总迭代数的  $1/3$  就可将损失值收敛到最优值的范围内）。选取正确的损失函数可以帮助我们有效地解决训练不稳定或长时间不收敛的问题，通过优化损失值以帮助 GAN 达到更好的效果。

(3) 在评价环节中，我们证明了 U-GAT-IT 作为一个新生的网络模型，无论是片段染色还是整体染色，最终效果都趋同甚至优于被大家广为引用的 CycleGAN 模型。其中 U-GAT-IT 片段染色的 SSIM 值达到了 0.8908，MSE 值为 500.8423，证明了在像素级对比环境下，U-GAT-IT 仍然有很好的表现。

### 6.2 讨论

在本节中我主要想讨论无监督学习网络模型的选择问题及其弊端。目前有很多研究都表明了 Pix2pix 这种像素级匹配的网络模型能够很好的实现不同组织或细胞的染色，那么根据以往的研究我们可以确定两点虚拟染色技术的特性：(1) 保存足够的细节特征和微观结构信息；(2) 不改变组织或细胞的病理特征。

由于内存限制，在 GAN 的领域内，大部分图像处理工作均是基于 256\*256 的图像进行风格转换等工作，同样我们的研究中也是采用了 256\*256 分辨率的图像作为输入输出，这样就导致了我们无法保存够足够的细节特征。例如在原始白片中会存在一些光学显微镜下只能观测到其轮廓的细胞甚至有些直接观测不到的细胞，这样的细胞在经过化学染色后可能会被染为蓝紫色，这时如果没有足够的数据集支撑其学习染色过程，那么出现这种现象的细胞在虚拟染色中是一定不会被成功染色的。因此我认为随着模型可容纳的数据集分辨率逐渐增大，一些具备细节特征的数据集被引入，用于虚拟染色的网络模型能够保存大量的细节信息。实际上 U-GAT-IT 的效果达到了上述第二个特性要求，血液病医生在观察一位病人是否具有某个病症时会利用电子显微镜观察一个染色完成的病人血细胞切片，在视野域内通过计数法来判断此病人是否有淋巴细胞百分比偏高等问题。从这个过程来看，医生依据的主要病理特征是与瑞氏吉姆萨染液中的碱性染料天青或美蓝结合的淋巴细胞个数（结合之后效果是整个细胞被染成蓝紫色，通过确定蓝紫色细胞的个数所占全部细胞个数的比例来预测可能的病情），那么根据我们目前利用 U-GAT-IT 得到的测试结果，这些虚拟染色图像中的病理特征与真实化学染色几乎相似，证明了我们的模型能够很好的维持细胞病理特征。

此外，我也想说明目前的循环生成网络模型在虚拟染色领域中尽管有很多研究证明了其准确性，但它在现实中有一定的不可靠性和不实用性。基于直观的结果，我们可以看到包括先前国外团队对 CycleGAN 研究和我对 U-GAT-IT 的研究成果，即由 CycleGAN 生成的虚拟染色图像看起来十分真实，甚至达到了以假乱真的程度。但是我在客观评价环节中注意到片段染色的准确度非常高，其 SSIM 值达到了 0.8908，但整体染色时包括 CycleGAN 在内的其余模型的 SSIM 值均大幅降低，U-GAT-IT 的值降至 0.3844，这说明我们的生成图像与真实图像之间还存在着较大差异。这是因为 CycleGAN 模型的固有弱点造成的，缺乏像素级匹配仅依靠强大的风格转移功能，并不能够生成准确的虚拟染色图像，同样我们也发现循环生成网络的结果并不稳定，其结果很容易受训练数据集影响（如我们在 5.1.3 节谈到的误染现象），然而现实中医学对数据和图像的精准度要求十分高，甚至存在一个细胞的误差都会影响诊断，因此我认为循环生成网络在虚拟染色在实际应用的发展仍然很局限，需要进一步提升模型和数据集精度。

此外，即使我们现在针对组织而非细胞进行染色，在数据集制备过程中，病理学家需要从同一组织块中切割两片相同的组织切片，一个用作白片，一个经过染色后作为标签图像，但它们之间通常有显著的结构差异，在临床实践中制作切片过程中变形是不可

避免的。因此，大量未配对的医学图像未被开发且更容易获得，而无监督学习的虚拟染色方法在许多现实应用场景中将更具竞争力。

### 6.3 问题及今后的研究方向

最后，目前类似于 U-GAT-IT、CycleGAN 这种循环生成对抗网络在虚拟染色领域只是起步阶段，我们能够做到在保存病理特征时实现染色风格的整体转换，但不能保证染色的完全正确，这一问题同样也是全球范围内虚拟染色研究团队尚未攻克的难题。

本文基于某些血液病如急性淋巴细胞性白血病的判断需染色病理切片支持的背景下，在虚拟染色方面做了调研和研究，使用了全新的网络模型并取得了一些成果。但是针对我们目前面临的问题可以发现虚拟染色还有很多待研究的方向，也是未来虚拟染色领域发展的突破点：

(1) 图像分辨率过低，不能满足医学图像的标准要求。实际上我们只要不断提高生成图像的像素值，达到医学图像标准以上，即医师可直接观察到图像中清晰的细胞结构从而区分出淋巴细胞时，我们就可以轻松地利用机器学习实现精准的虚拟染色。目前我们的图像像素最高只能达到 512\*512 (33G 显存的服务器)，平常只能维持在 256\*256 (11G 显存的服务器)，原因之一是 U-GAT-IT 的原始模型搭建时采用的网络结构是针对低分辨率图像的，我在尝试修改结构后发现目前网络模型出现梯度爆炸的问题，服务器内的显存无法支撑模型的下一次迭代。对于这个问题，我认为目前有两个创新网络的思路：一是借鉴 Pix2pixHD 中的多尺度判别器和 coarse2fine (由粗到精网络) 的生成器，可以提高生成图像的质量。多尺度判别器是指利用多个判别器分别判定不同分辨率下的真假图像。比如四尺度判别器，它会分别判定像素为 512\*512、256\*256、128\*128、64\*64 的图像（直接经过 pooling 下采样即可获得不同分辨率的图像）。由粗到精网络的生成器是要先训练一个低分辨率生成网络，训练完成后再接一个高分辨率生成网络，高分辨率网络融合低分辨率网络中的特征映射从而得到更准确的结果。二是借鉴 LAPGAN 的做法，从低分辨率开始，不断向下得到高分辨率残差，并将这些残差累加，最终得到高分辨生成结果。

(2) 循环生成对抗网络相对于像素级匹配方法会表现出较低的精度，无法适应医学的严谨性。与拥有配对图像的虚拟染色方法不同，我们想要提升循环生成对抗网络这种无监督学习的效果，就要提供大量的像素注释来设计完全监督的学习算法。通过不停地尝试不同的网络模型，改进网络结构和损失函数，我们可以获得更高精度的染色结果，

最后我认为我们也可以逐渐从无监督学习向半监督学习转变，因为这样仅需要一小部分配对的数据集即可显著提高准确度。

(3) 尽管其在节约时间人力成本上的优势显而易见，但目前的虚拟染色在现实应用受限，医生信任度低。我认为受信任度低这个问题的根本是虚拟染色的原理不被认可，通过我对虚拟染色的了解和与医师的对话，首先我想说明虚拟染色这项技术的主要依据是图像内细胞的物理特征，然而在化学染色时它依靠的是细胞的生物化学特性，这个仅从物理特征例如位置、大小就进行染色的技术虽然有着巨大的成本优势且有一定的可行性，但依旧不被医生看好，因此我认为未来虚拟染色领域除了需要不断优化神经网络模型外，还需要尽量引入细胞或组织的生物特性信息，这样才能实现高精度的染色效果。同样我们可以建立虚拟染色与临床诊断的联合机制，对于一个白片同时进行虚拟染色和化学染色后再让医生进行对比和判断，纠正虚拟染色中出现的错误并在网络中设置惩罚机制，在积累数据集的同时也能人为地改善网络性能，优化染色结果。

## 参考文献

- [1]Rivenson,Yair,Hongda Wang,et al. Virtual histological staining of unlabelled tissue autofluorescence images via deep learning[J]. Nature, 2019(3):67-95.
- [2]Rivenson,Yair,Kevin Haan,et al. Emerging Advances to Transform Histopathology Using Virtual Staining[J]. BME Frontiers, 2020(8):112-123.
- [3]Rivenson,Yair,Tairan Liu,et al. PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning[J]. CIOMP, 2019(8):2047-2059.
- [4]Thomas B,Meyke Hermse,et al. Stain-Transforming Cycle-Consistent Generative Adversarial Networks for Improved Segmentation of Renal Histopathology[J].MIDL102,2019(10):151-163.
- [5]Mercan,C.&Mooij. Virtual staining for mitosis detection in breast histopathology[J].eess.IV,2020(3).
- [6]S.Liu,X.Li,et al. The Generation of Virtual Immunohistochemical Staining Images Based on an Improved Cycle-GAN[J]. MLICOM,2020(1):137-147.
- [7]S.Liu,B.Zhang,et al. Unpaired Stain Transfer using Pathology-Consistent Constrained Generative Adversarial Networks[J]. IEEE,2021.
- [8]Kim,J&Minjae Kim. U-GAT-IT: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation[J].ICLR,2020.
- [9]陈凡.病理组织切片图像的分型识别及其并行化研究[D].贵州大学,2019.
- [10]李丹.基于条件生成对抗网络的颈动脉切片虚拟染色方法研究[D].北京交通大学,2020.
- [11]X.Zhang&C.Zhu. ECycleGAN Enhanced Cycle-Consistent Generative Adversarial Networks[R].Dalian University of Technology,2021.
- [12]Artzai Picon,Alfonso Medela,et al.Autofluorescence Image Reconstruction and Virtual Staining for In-Vivo Optical Biopsying[J].IEEE,2021(2).
- [13]David M,Michael W,et al.Stain-less staining for computed histopathology[J].Technology,2015(3).
- [14]HCho&S Lim.Neural Stain-Style Transfer Learning using GAN for Histopathological Images[J].JMLR,2017(10).
- [15]K Haan,Y.Zhang,et al.Deep learning-based transformation of the H&E stain into special stains improves kidney disease diagnosis[J].eess.IV,2020(8).
- [16]Rivenson,Yair.,Kevin Haan,et al.Emerging Advances to Transform Histopathology Using Virtual Staining[J].BME,2020(8).

## 成 果

- [1] 搭建一套染色系统“基于 U-GAT-IT 的细胞虚拟染色系统”
- [2] 撰写一篇软件著作权“基于 U-GAT-IT 的细胞虚拟染色方法研究”（学生一作）
- [3] 提交审核一篇期刊论文“一种新型细胞虚拟染色方法及未来发展趋势”（《科学与技术》，2021）

## 致 谢

在论文提交之际，我首先要感谢我的指导老师——朱睿杰老师。朱老师带领我进入虚拟染色领域并积极帮助我采集文中所需要的数据集，他严谨的工作和学习态度让我收益颇多。同时我要感谢我们课题小组的同学们，积极帮助我解决复现模型时所遇到的问题，最后感谢我所参考的国内外期刊著作的作者，没有他们的启发我无法完成本篇论文。

这篇论文的实现内容是虚拟染色领域的一项创新，但同时它还存在很多不足，希望大家批评指正。