

参赛队员姓名： 陈应涵

中学： 北京市第四中学

省份： 北京市

国家/地区： 中国

指导教师 1 姓名： 黄庆明

指导教师 1 单位： 中国科学院大学

指导教师 2 姓名： 孔丽燕

指导教师 2 单位： 北京市第四中学

论文题目： 基于生成对抗网络的山
水画着色研究

摘 要

山水画是中国传统绘画艺术形式，具有历史悠久、风格多样等特点。但是，由于颜料工艺、保存条件等限制原因，往往仅有水墨画得以完整流传，而设色山水画则可能随着时间流逝而褪色。为解决这一问题，本文构建了 LPD 数据集，其中包含了训练用的设色山水图像 8467 张、测试用的设色和水墨山水图像各 100 张。这些数据经过裁剪、标准化处理以及部分灰度化处理后形成了配对的彩色与黑白山水画样本。同时，本文运用生成对抗网络 (GAN) 进行灰度山水图像着色，使用 Resnet 和 Unet 两种网络骨干网络构建山水画着色模型，并分析对比了两类模型的效果。定量分析结果表明，本文方法可以生成真实感强、还原度高的设色山水图像。最后，本文还进行了用户调查，调查显示用户难以分辨本文方法生成结果与真实设色山水画。本文方法可以用于褪色山水画的修复、山水名画再创作等，具有广泛的实用价值。

关键词：生成对抗网络；山水画；图像上色

ABSTRACT

Landscape painting is a traditional Chinese painting art form with a long history and diverse styles. However, due to pigment technology, preservation conditions and other limitations, most of the landscape paintings that can be seen nowadays are monochrome wash drawings, while colorful landscape paintings may fade as time went by. In order to solve this problem, this research establishes a new landscape painting dataset LPD, in which 8467 colorful landscape painting images for training, 100 images of both colorful and wash landscape drawings for testing are collected. After cutting, normalization and grayscale processing for partial images, pairs of gray and colorful landscape painting samples are formed in the dataset. Meanwhile, this research uses the Generative Adversarial Networks (GAN) and pix2pix technology to colorize grayscale images. Resnet and Unet are applied to establish landscape painting colorization models, which are analyzed and compared afterwards. The results of quantitative analysis show that the method can generate aesthetically pleasing paintings with strong sense of reality and high degree of restoration. User survey show that it is difficult for human to distinguish real colorful landscape paintings from fake ones generated by the model. The model has various practical values in restoring color of faded landscape paintings and re-creation of famous landscape paintings.

Key words: GAN; landscape paintings; image colourization

目录

第 1 章	绪论	4
1.1.	研究背景	4
1.2.	问题的提出	4
第 2 章	国内外研究发展现状	5
2.1.	山水画生成和处理	5
2.2.	图像着色技术	6
2.3.	生成对抗网络 (GAN)	6
第 3 章	研究思路 and 方案	7
第 4 章	数据集	8
4.1.	数据的收集与介绍	8
4.2.	数据的处理	9
4.3.	LPD 数据集的建立	10
第 5 章	模型的建立与训练	10
5.1.	pix2pix 模型	10
5.2.	Resnet	12
5.3.	Unet	13
5.4.	模型的训练	15
第 6 章	测试与结果	15
6.1.	Resnet 模型的测试与筛选	15
6.2.	Unet 模型的测试与筛选	16
6.3.	Resnet 与 Unet 模型的对比	18
第 7 章	讨论	19
7.1.	结论	19
7.2.	创新点	21
7.3.	不足	21
7.4.	展望	21
第 8 章	参考文献	21
第 9 章	致谢	22

第1章 绪论

1.1. 研究背景

图像着色技术是当代计算机科学中非常重要的组成部分。研究表明,人认知事物的过程主要由视觉系统获得有效信息,而颜色作为图像最重要的属性之一,是构成视觉信息的基本要素。在当今信息技术发展迅速的时代背景下,人们对图像色彩的需求越来越大,图像着色技术也因此得到了快速发展。其中,生成对抗网络(Generative Adversarial Network, GAN)是近年来在图像的复杂分布上进行无监督学习最具前景的方法之一,于2014年由Ian J. Goodfellow等人提出^[1]。GAN通过生成模型G(Generative Model)和判别模型D(Discriminative Model)的相互博弈学习产生出效果良好真实的输出,可以很好地运用在图像着色领域中。而pix2pix技术由Phillip Isola等人于2017年提出^[2],是基于GAN的一种新型图像到图像转换方法,可以端到端地在两种图像分布之前学习转换,具有较高的通用性和更好的细节处理,是当今图像着色领域中较为先进和主流的技术。

山水画是最富有中国特色的、最具代表性的中国传统画法之一,它以描绘山川自然景色为主体,可以分为青绿山水、水墨山水、浅绛山水、没骨山水等。其中,青绿山水和浅绛山水是在墨线的基础上,用石青、石绿、花青、赭石等矿物质颜料上色绘得,统称为设色山水。而水墨山水则是纯用水墨、不设颜色的山水画体,用墨的浓淡干湿来表现物体的虚实远近,而这与图像着色领域中用作输入的灰度图像(Grayscale)具有类似的性质。灰度图像是指每个像素只有一个采样颜色的图像,通常显示为从最暗的黑色到最亮的白色的灰度。与黑白图像不同的是,在计算机图像领域中黑白图像只有黑白两种颜色,而灰度图像在黑色与白色之间还有许多级的颜色深度,与水墨山水中墨色的浓淡相似。因此,运用图像着色技术可以实现将水墨山水着色成为设色山水的功能。而历经千百年的岁月沧桑,部分设色山水古画也可能会出现褪色现象。因此,山水画着色技术也可以应用于褪色古画的修复工作中。

1.2. 问题的提出

山水画作为中国传统艺术形式,从魏晋南北朝时期开始发展,隋唐时期开始脱离人物画独立存在、发展,并分支出了水墨山水、青绿山水等。随着历史的发展,山水画在两宋繁荣,在元代达到了高潮,并在明清随商品经济发展走向了辉煌。山水画表现了丰富多彩的自然风光,更体现了中国人崇尚自然的审美意识。

然而,当人们想起山水画时,大部分人在第一时间想到的通常是不设色的水墨山水,即不上彩色颜料、只以墨汁作画,视觉上为黑白色调的无彩色图像,而对青绿山水等设色山水画知之甚少。本文旨在运用生成对抗网络实现图像着色,将可以理解为灰度图像的纯用水墨、不设颜色的水墨山水画转化为设色山水画。通过为水墨山水画设色,可以帮助大众更好地了解设色山水这一艺术门类,同时对比、了解水墨山水和设色山水的特点。

同时，由于古代受颜料技术、物质条件等因素的限制，古时候绘制设色山水的成本要远高于水墨山水，这可能也是传世的设色山水画远少于水墨山水画的原因之一。许多为人称道的山水名画都仅为不设色的水墨版本传世，不得不称为一种遗憾。山水画上色技术可以为没有颜色的水墨山水名画赋予新的生命，使人们从一个全新的视角欣赏水墨山水名画的魅力。

此外，历史上的很多古代山水画经过长时间光照、水分等侵损，多数会出现褪色现象。而对于设色山水画而言，除了极少数采用珍贵矿石颜料绘制的画卷，很多早已褪去原有的颜色。而运用山水画着色技术，可以通过计算机技术将这些褪色的山水画重新着上绚丽的颜色，使它们重新获得初被画就时的光彩，完成对古画的电子修复工作。

本文的研究意在运用计算机技术对山水画进行着色处理，保护、宣传山水画这门传统艺术，更好地培养大众审美水平和素质，弘扬中华民族优秀传统文化。

第2章 国内外研究发展现状

2.1. 山水画生成和处理

作为中国的传统艺术形式，山水画多年来持续受到人们的重视，其生成与处理也随着人工智能时代的到来获得了更多信息技术的支持。截至目前，计算机技术在山水画处理方面主要应用在利用黑白照片生成水墨山水画和利用彩色真实照片生成设色山水画两方面。

利用黑白照片生成水墨山水画的技术较为完善，前人从事的研究也较为丰富。李大锦于 2011 年申请的专利“一种自动生成水墨山水画图像的方法”^[3]据画家真实作画的过程，将生成水墨山水的过程划分为钩、皴、染、点四步，每个步骤分别生成一张图像，将四步分别生成的图像融合为一张最终水墨山水画的图像。这种方法生成水墨画的效率相对较高，可用于提高水墨动画制作的效率。上海大学的刘晓崇等人于 2014 年发表的论文《基于真实场景的水墨风格图像自动生成方法》^[4]，可以将适合传统水墨画的静态真实场景图像转化为水墨风格的艺术图像。该文运用了图像腐蚀、各向异性扩散和边缘信息提取叠加等方法实现图像的水墨风格转换，对水墨效果的仿真效果较好。

相比于水墨山水的生成，生成设色山水的自动技术较少，多需要运用软件手动处理。如利用 Photoshop 软件，将背景设为不同图层，运用明度、滤镜、液化等工具生成彩色山水画图像。这种方法较为不完善，需要人工处理，效率较低。且利用这种方式得出的彩色山水画有较为明显的处理痕迹，与真实的彩色山水画差别较大，仿真效果较差，改进空间较大。

普林斯顿大学的 Alice Xue 在 2020 年发表了论文《End-to-End Chinese Landscape Painting Creation Using Generative Adversarial Networks》^[5]是山水画生成领域的新突破。该文建立了素描-绘画生成对抗网络（SAPGAN），是一个可以生成可理解的、边缘清晰的中国山水画的端到端框架，并引入了新的数据集。

SAPGAN 架构包含两个步骤：第一步为素描 GAN (SketchGAN)，从潜在空间中的采样向量生成高分辨率的边缘图像作为底稿；第二步是绘画 GAN (PaintGAN)，将素描 GAN 生成的边缘图利用 pix2pix 技术生成山水画。该文运用了与本文相似的、较为先进的 GAN 和 pix2pix 技术，生成效果较为真实，但依然不能解决水墨山水画到设色山水画的转化问题。

综合来看，目前针对山水画的研究成果主要由照片生成山水画，其中基本不涉及设色山水和水墨山水之间的转化，即山水画的着色问题。且使用的技术多数较为落后，使用近几年发展出的较为先进的图像处理技术较少。因此，本文所研究的山水画着色具有较高的创新性。

2.2. 图像着色技术

图像着色的目标是给定灰度图像作为输入，生成彩色图像。这个问题长期以来具有较大的挑战性，因为单个灰度图像可能对应许多合理的彩色图像，是一种多模式的任务。因此，传统模型通常高度依赖于用户输入以及灰度图像的内容。

近年来，深度学习和神经网络推动了自动图像着色方面的显著发展，从灰度图像到彩色图像的过程中不再需要额外的人工输入。这种成功的部分原因在于深度神经网络能够捕捉和使用语义信息（即图像的实际内容），尽管目前仍无法完全弄清其内部学习原理和此类模型表现如此出色的机理，深度学习技术的确促进了图像着色领域的巨大飞跃和快速发展。

来自澳大利亚国立大学的学者 Saeed Anwar 等人于 2020 年发表了论文《Image Colorization: A Survey and Dataset》^[6]，对图像着色领域做了较为全面的总结和综述。该论文参考近百篇文献，对图像着色领域的发展进行了全面回顾，对该领域的各种方法进行了总结，并根据领域类型、网络结构、辅助输入和最终输出的差异提出了新的着色网络分类法，将网络分为 Plain Networks、User-Guided Networks、Domain-Specific Networks、Text-based Networks、Diverse Colorization Networks、Multi-Path Networks、Exemplar-based Colorization Networks 等网络类型。此外，文章还引入了名为 Natural Color Dataset (NCD) 的新基准数据集，并对该数据集上对领域内的最新模型进行了系统评估。

通过综合比较，在现有的模型中，运用了本文所使用的 GAN 和 pix2pix 技术的网络具有更好的效果，且应用更加广泛。

2.3. 生成对抗网络 (GAN)

生成对抗网络 (GAN) 是由蒙特利尔大学 Ian J. Goodfellow 在 2014 年提出的机器学习架构。GAN 包括了两套独立的网络，两者之间相互对抗、相互博弈学习产生出效果良好真实的输出。第一套网络是需要训练的判别器 D (Discriminative Model)，用来分辨样本是真实数据还是虚假数据；第二套网络是生成器 G (Generative Model)，用于生成类似于真实样本的结果。

在模型训练的过程中，生成器 G 会生成假的数据，判别器 D 会接收真实数据和 G 产生的假数据，D 的任务是判断图片的真假。针对最后输出的结果，模型可以同时调整 D 和 G 的参数进行优化。调整 G 的参数使生成器生成的假数据更加真实，调整 D 的参数可以避免下次类似的判断再度出错。这种参数的优化调整会随着模型的训练一直持续到两者进入到一个较为均衡和谐的状态，判别器和生成器的性能都能得到有效提高。

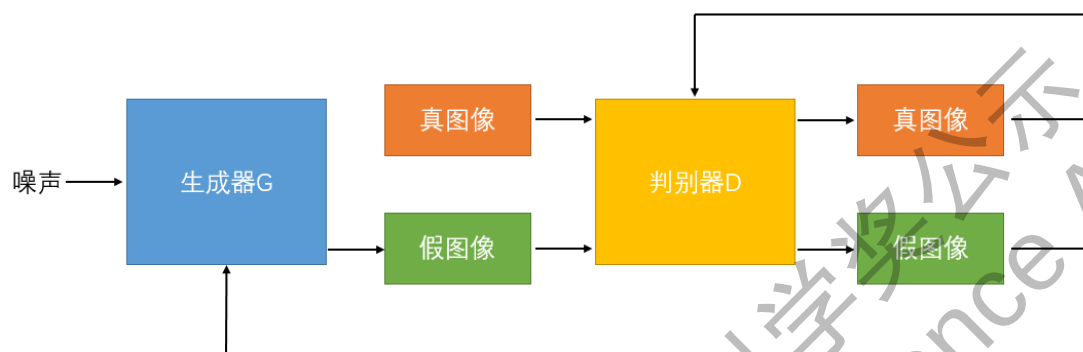


图 1：生成对抗网络（GAN）结构

GAN 类型多样，自 2014 年提出至今，在许多领域内都取得了较好的成绩，各式的变体也应运而生。例如：InfoGAN^[7]在输入噪声向量 z 的同时，联合输入了一个隐藏编码 c ，并最大化 c 与输出图像的交互信息，通过无监督的方式使 c 具有控制输出图像语义特征的能力；条件 GAN^[8]（Conditional GAN, cGAN）在生成网络和判别网络的输入中同时附加信息条件 y 来使生成网络能够基于 y 生成目标图像，使判别网络具有依据条件信息判断生成图像真假的能力；辅助分类器 GAN^[9]（Auxiliary Classifier GAN, AC-GAN）与 InfoGAN 和条件 GAN 非常相似，它的生成网络输入中包含了分类标签 c ，并在架构中包含了一个辅助分类网络，来从生成网络的输出中重建 c ，生成不同类别的图像。

第3章 研究思路 and 方案

本文通过研究、改进生成对抗网络的 pix2pix 模型，将其适用于来源于生活实际的山水画着色方面。训练所用的输入的的数据由互联网上搜索设色山水画得到，测试数据由搜索水墨山水画和设色山水画得到，经过筛选、裁剪、处理等过程投入模型进行训练。本文运用 Resnet 和 Unet 两种不同的网络框架构造的 pix2pix 模型分别进行训练，可以比较不同网络结构的特点，同时提供可能有不同效果的模型进行选择，利于选取最优模型。

本文首先针对图像着色领域的研究进行调研，充分了解了前人做过的相关工作，了解到过往的研究往往针对食物、生活用品等，而在传统绘画的着色领域鲜有人进行探索。且从文化保护的角度考虑，传统绘画自动上色技术目前确实存在着很大空缺。因此作者选定了此兼具创新性和现实意义、并且能极好地运用前沿图像着色技术的研究内容——山水画着色。

目前在图像着色领域中应用较为广泛的技术为 GAN 和 pix2pix，而生成对抗

网络的训练需要大量的数据。由于此前关于山水画着色方面的研究较少，因此很难找到现有的数据集直接使用。因此本文提出了 LPD(Landscape Paintings Dataset) 山水画数据集，该数据集由设色山水画和水墨山水画共同组成，图片经裁剪与规范化最终以 256×256 像素的 JPG 格式呈现。而训练数据集中应同时包含设色山水画和其对应的灰度图，因此训练数据中的设色山水画还需经过灰度化后才可作为 GAN 网络的输入，而原始图片则作为监督信号。测试数据则由水墨山水画和灰度化后的设色山水画共同构成。

本文在 pix2pix 工作的基础上进行了进一步研究，并分别以 Resnet 和 Unet 作为 G 网络的骨干网络对模型进行了训练。通过测试对比两种网络的最终效果，本文根据其各自的特点选择了更优的设计。

本文最终结果将以 html 网页的形式公开，其包含了真实图像、灰度图像和生成的彩色图像，希望本文的研究能为这一领域的学者提供启发。

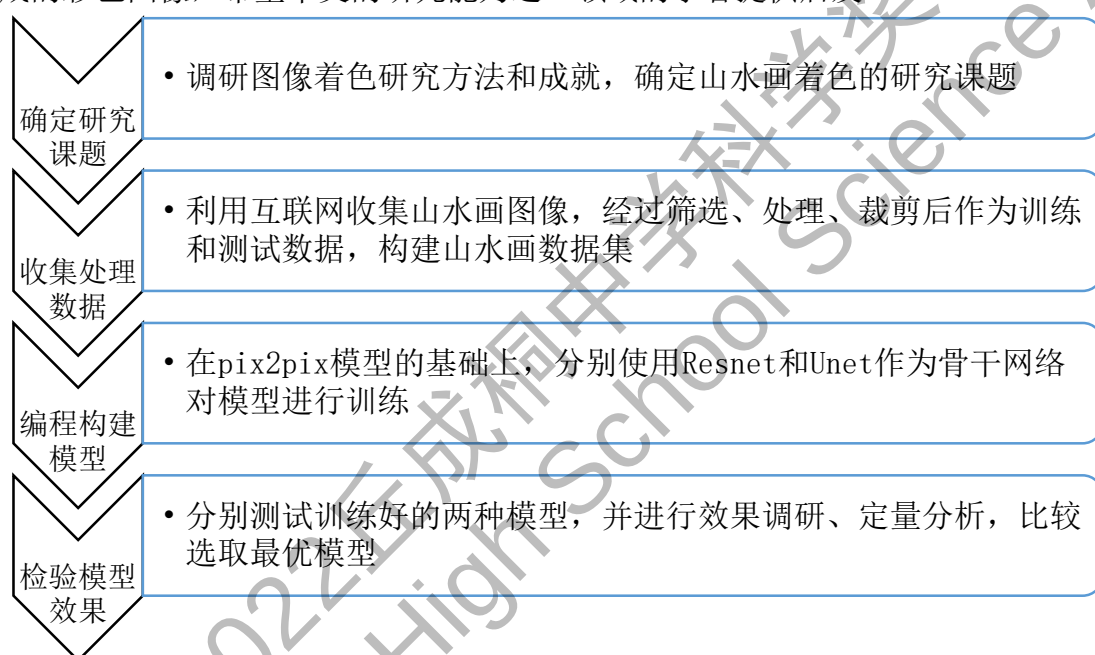


图 2：本文研究思路和路线

第4章 数据集

4.1. 数据的收集与介绍

● 数据的收集

本文所构建的生成对抗网络所需的全部数据均从互联网上搜集、下载得到，所搜集的原始图像都是清晰度高、尺寸适中、风格鲜明、有代表性的山水画电子图像，处理后即形成了一个全新的山水画数据集。

● 训练数据介绍

图像着色任务所需的训练数据为彩色图像，在本文中即设色山水画。本文从互联网上搜集了 1000 张从古至今、不同流派风格的设色山水画，包含青绿山水、

浅绛山水等不同种类，其中不乏传世名画和当代画家精品，基本可以囊括山水画的各种种类和特点。这 1000 张设色山水图片，将在处理后作为训练数据的输入。



图 3：作为训练原始数据的设色山水画

● 测试数据介绍

测试数据分为两大类：一类是 8 幅设色山水画，用于与测试中生成的彩色图像进行对比，从而检测模型的生成效果；另一类为 16 张水墨山水画，可通过模型进行着色也即风格转换，贴合本文山水画着色的研究目的，体现实用价值。测试数据所选用的山水画多为名画佳作，如设色山水中选用了王希孟《千里江山图》、赵伯驹《江山秋色图》等古代名画和诸多当代画家的佳作，水墨山水中更是选用了黄公望《富春山居图》、范宽《溪山行旅图》、李唐《万壑松风图》、王诜《渔村小雪图》、赵孟頫《江山万里图》等国宝级名画，以此更好地体现着色模型的实用价值。



图 4：测试原始数据

设色山水画——宋·赵伯驹《江山秋色图》（左）

水墨山水画——宋·李唐《万壑松风图》（右）

4.2. 数据的处理

● 数据的裁剪

从互联网上初步搜集到的数据大小不一、良莠不齐，需要通过一定的处理才能作为实验数据。同时，互联网上能搜集到的数据量较少，而本文训练所需要的

数据量大，因此需要对原始图像进行裁剪、分割。本文原始数据经由裁剪和规范化，每张图像能裁剪出 4 到 20 张不等的、大小为 256×256 的 jpg 格式图像。

● 数据的筛选

在对图像进行裁剪后人工筛除图像模糊、颜色怪异、以空白为主的图像，以及带有水印、题跋和印章的图像，保留下适于训练和测试的优质数据。完成所有裁剪、筛选等处理后，剩余设色山水图像 8467 张用作训练数据、100 张设色测试数据和 100 张水墨测试数据，投入模型的建立与检验中。

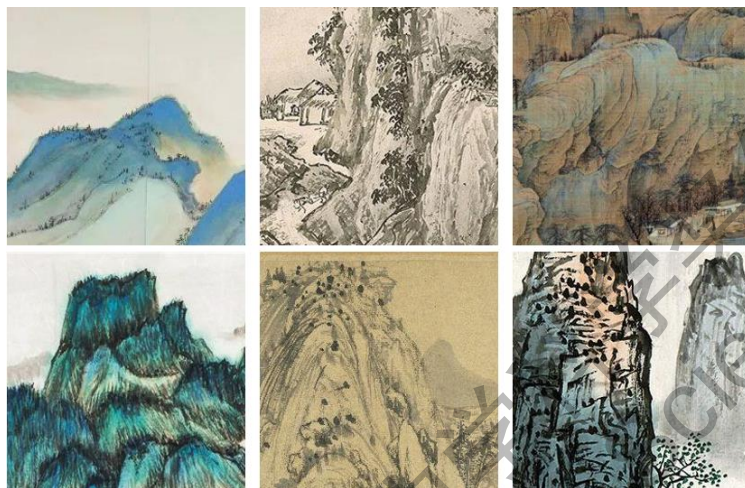


图 5：处理后的训练（左）、测试（水墨山水（中）、设色山水（右））图像

4.3. LPD 数据集的建立

经过数据的收集和处理后，本文构建了一个全新的山水画数据集，命名为 LPD (Landscape Paintings Dataset)。LPD 数据集包含了训练用的设色山水图像 8467 张、测试用的设色和水墨山水图像各 100 张，均以 256×256 像素的 JPG 格式呈现。训练数据中的设色山水画经过灰度化，设色山水原始图像和其对应的灰度图共同作为 GAN 网络的输入，而测试数据则由水墨山水画和灰度化后的设色山水画共同构成。经过处理后配对的山水画彩色和灰度图像样本，可供后人酌情选用，继续进行山水着色研究乃至其他山水画相关研究工作。

第5章 模型的建立与训练

5.1. pix2pix 模型

图像转化，即 pix2pix 技术，是基于一张输入图像得到目标输出图像的过程，也即图像间的一种映射，将高分辨率输入网格映射到高分辨率输出网格，是生成对抗网络（GAN）的重要应用方向。

Pix2pix 技术来自于加利福尼亚大学 Phillip Isola 等人在 CVPR 2017 上发表的一篇论文《Image-to-Image Translation with Conditional Adversarial Networks》，这是将生成对抗网络（GAN）应用于有监督的图像转换任务的经典论文。这篇论文

改进了原始 GAN 使用噪声作为输入的局限,将条件 GAN 引入了图像转换问题,提出了一个通用的端到端的“图像到图像转换”的框架。这篇论文提出的 pix2pix 模型相比以前的模型有诸多进步:一是相比于传统方法,对于不同类型的图像转换, pix2pix 不需要额外的算法和损失函数的设计,具有很强的通用性;二是使用的 Unet 框架可以提升图像的细节特征,利用 PatchGAN 分类器处理图像高频部则可以使生成图像的轮廓和条纹更清晰,色彩度更加鲜艳。

Pix2pix 技术主要应用了生成对抗网络中的条件 GAN 进行图像转化其可以通过添加条件信息指导图像生成。运用条件 GAN 的图像转化中以输入图像为条件,学习从输入到输出图像间的映射,从而得到目标输出图像。条件 GAN 给生成器 G 输入有 A 分布特点的数据集 A,且还需准备一个与 A 中数据样本成对的数据集 B。给 G 输入一个来自 A 的样本 X,使 G 输出一个符合数据集 B 分布的 G(X),且 G(X)要与 X 在 B 中的对应数据样本 Y 尽可能接近。这是通过条件 GAN 的损失函数(如 L1 损失函数),以及判别器 D 在判定时将<X, Y>和<X, G(X)>成对输入实现的,实现了 G 同时学习从分布 A 到分布 B 的映射关系和 X 与 Y 的配对关系,从而实现控制 G 的输入来生成特定输出的目的。

Pix2pix 技术自 2017 年诞生以来发展迅速,涌现了很多不同版本、风格各异的实现方案。由于 pytorch 框架灵活高效,本文使用了该框架搭建 pix2pix 模型,并以此为基础开展了山水画着色研究。

对于本文考虑的着色问题,输入图像和输出图像在纹理上具有强相关性,因此其间存在信息的共享。针对这样的条件,本文的网络模型中用于信息压缩的卷积部分和图像生成的反卷积部分的结构大致对齐,模型大致呈对称结构。

本文的 Pix2pix 模型中生成器 G 选用的损失函数为 L1 损失函数。G 除了要生成让 D 难以判别的图像之外,还需要输出更接近真实图片,即约束其产生的图片 G(x)与所对应的真实图片之间的接近程度。而本文的 pix2pix 模型所选用的 L1 损失函数经测试效果优于 L2 损失函数,产生的图片更加清晰。L1 损失函数的表达式如下:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x,z)\|_1]$$

而对于 cGAN 的损失函数较为通用,且同时针对 G 和 D 均有效果,即:

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{x,z}[\log(1 - D(x,G(x,z)))]$$

再与只针对生成器 G 的 L1 损失函数相结合,两部分组合后可以得到最终函数如下:

$$G^* = \operatorname{argmin}_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda \mathcal{L}_{L1}(G)$$

在这样的优化目标的约束下,模型才能实现对输出图片的具体控制,而不是仅仅满足一个真实图片集的分布。

在 pix2pix 模型的函数中,生成器 G 使目标函数最小化,判别器 D 使目标函数最大化。x 表示输入的灰度图像, y 表示真实的目标图像, z 表示生成的图像。作为生成器, G(x,z)生成的图像尽可能逼真,努力使判别器 D 认为 G(x,z)为真的图像,则 D(x,G(x,z))的值变大,尽可能接近 1,则 1-D(x,G(x,z))的值变小,目标函数变小。作为判别器,“D”会识别出 G(x,z)为假的图像,则 D(x,G(x,z))的值尽可能小,则 1-D(x,G(x,z))的值变大,目标函数变大。D(x,y)表示对真实图像的判别,其值越大代表判别器效果越好,目标函数变大。

为了能更好得对图像的局部进行判断, 本文的 Pix2pix 判别器采用 patchGAN 结构, 即使用全卷积网络作为判别器 D, 使 D 的输出为具有一定形状大小的二维张量, 而不是一个代表全图的统一数值。张量上每一处仅对应于输入上的一块图像块大小, 其数值也仅代表这一块面片的真假。训练时使这个张量与全零或全一的等形状张量作 L2loss, 来通过对每个面片真伪的学习来提升 D 对图像真伪的判断。这样做有着较多好处: 首先, patchGAN 结构使判别器 D 的结构变小, 计算量变小, 因此训练速度加快。其次, 因为生成器 G 本身是全卷积网络, 对图像尺寸没有限制。若判别器 D 用 patch 结构处理图像, 也不会对图像大小产生限制, 就可以使整个 Pix2pix 框架对图像大小没有限制, 增大了框架的扩展性。

Pix2pix 模型是 x 到 y 之间的一对一映射, 也就是对 ground truth 的重建: 先输入轮廓图, 经过编码解码 (Encode-Decode) 成对应的向量, 再解码成真实图像。这种一对一映射的应用范围较为有限, 当输入数据与训练集中的数据差距较大时, 生成的结果很可能没有意义。本文中, 希望生成图像的方向和目标较为确定, 希望生成的是传统风格的设色山水画, 指向性较强。因此, 本文所选用的训练数据均为青绿山水、浅绛山水等传统设色山水画, 可以更好地突出 pix2pix 模型的优势。

5.2. Resnet

何恺明等人于 2015 年发表的论文《Deep Residual Learning for Image Recognition》^[10]提出了深度残差网络 (Deep Residual Network, Resnet), 是卷积神经网络历史上的一件里程碑事件。

按传统经验, 网络深度与模型性能应成正相关的关系, 网络层数越多, 网络进行越复杂的特征提取, 理论上可以取得更好的效果, 但大量实验研究发现并非如此, 随着网络深度增加, 网络出现了退化问题 (Degradation Problem)。针对此问题, 何凯明等人提出了残差学习的方法, 对于一个堆积层, 将对输入 x 的学习特征记为 $H(x)$, 希望可以学习到残差 $F(x)=H(x)-x$ 。这是一种短路链接 (Shortcut Connection), 可以使堆积层在输入特征基础上学习到新的特征, 从而拥有更好的性能。

Resnet 网络参考了 VGG19 网络, 在其基础上通过短路链接加入了残差单元, 形成残差学习。Resnet 使用步幅 (Stride) 为 2 的滤波器进行下采样, 网络较浅的 18-layer 和 34-layer 的 Resnet 进行的两层间的残差学习, 当网络更深时, 将进行卷积核大小分别是 1×1 、 3×3 和 1×1 的三层间的残差学习。下图中, 左图对应的是浅层网络的残差单元, 而右图对应的则是深层网络的残差单元。

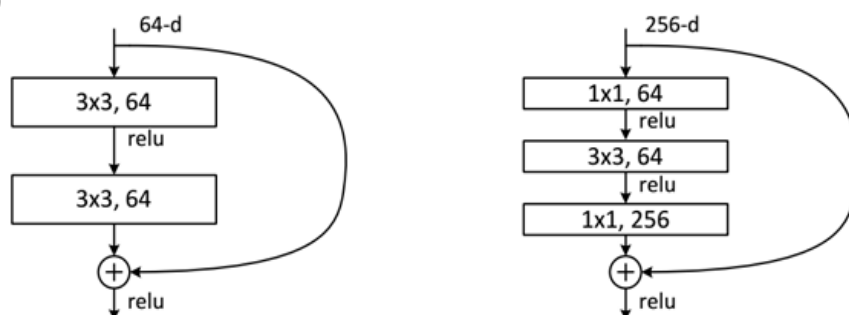


图 6: Resnet 浅层、深层网络残差单元对比

本文所选用的训练框架之一 Resnet9 是有着 9 个残差块（Residual Block）的 Resnet 网络。残差块结构具有更强的泛化能力，能更好地避免网络退化，即便网络层数超过 1000 层性能仍可以继续优化。残差块结构通过保持短路的“纯净”，使信息可以在前向传播和反向传播中平滑传递。在残差路径上，将 BN 和 ReLU 统一放在 weight 前，获得了具有易于优化且不易退化的效果的 full pre-activation 结构。本文选用 Resnet 网络作为训练框架之一，是为了进行更深层次的有效训练，以图获得更好的训练效果，以选取更适于完成山水画着色任务的模型。

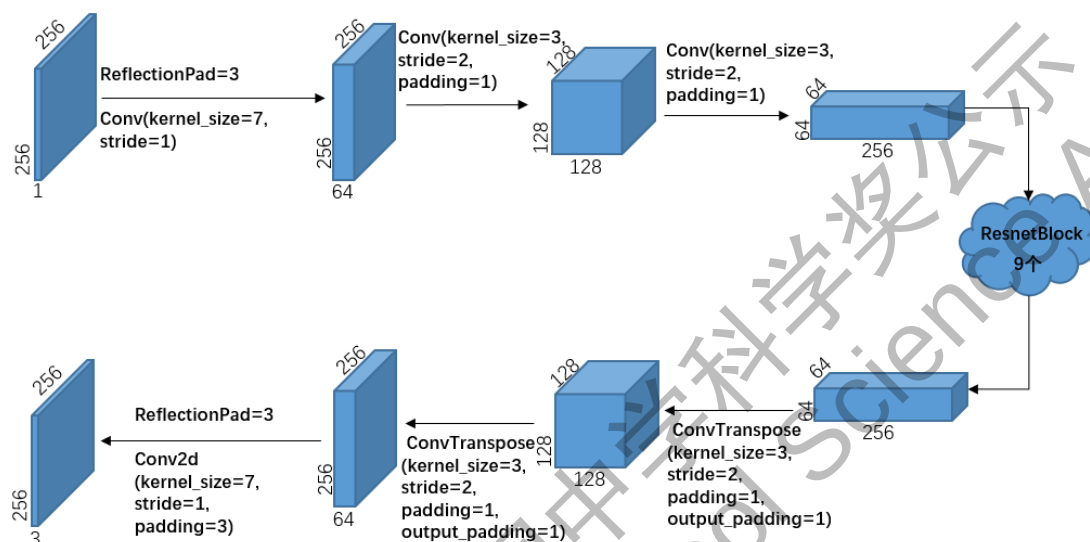


图 7: Resnet 模型网络结构

5.3. Unet

Unet 是较早使用全卷积网络进行语义分割的算法之一。Olaf Ronneberger 等人于 2015 年发表的论文《U-Net: Convolutional Networks for Biomedical Image Segmentation》^[11]中使用了当时十分先进的包含压缩路径和扩展路径的对称 U 形结构，在一定程度上影响了后续分割网络的设计思路 and 理念。

Unet 主要用来处理图像分割任务，将图像分割出一个物体的准确轮廓，并求一个映射函数。Unet 网络是典型的编码-解码结构，左侧是压缩路径（Contracting Path），是由卷积和池化构成的一系列降采样操作，由 4 个卷积块（Block）组成，每个 block 分别使用 3 个有效卷积和 1 个最大池化（Max Pooling）进行降采样，每次降采样之后特征图的个数乘 2。U 型网络的右侧部分为扩展路径（Expansive Path），同样由 4 个 block 组成，每个 block 通过反卷积将特征图的大小乘 2，同时将其个数减半（最后一层略有不同）并与裁剪到与扩展路径特征图相同尺寸的左侧相应的特征图合并归一，最终经过分类后输出两个特征图。

本文所使用的数据均为 256×256 的图像，Unet 网络通过多次步长为 2 的卷积，将输入依次变为 128×128，64×64，32×32，16×16 等不同尺寸的特征图，再对特征图进行反卷积：以 16×16 的特征图为例，对 16×16 的特征图进行反卷积，得到 32×32 的特征图，将其与之前的 32×32 的特征图进行拼接，再对拼接后的特征图进行卷积、上采样，得到 64×64 的特征图。这样经过连续的拼接、卷积、上采样，就可以得到一个与输入图像尺寸相同的生成结果。Unet 网络利用拼接层

来融合网络左右两侧对应位置上的特征图，从而更完善地恢复原始图像中的细节信息，提高分割精度。而运用跳层连接（Skip Connection）能够使网络在每一级的上采样过程中保留更多高层特征图蕴含的高分辨率细节信息，从而进一步提高图像分割精度。因此，Unet 是一种可以保留更多图像细节信息的网络结构，适于运用在本文的着色任务中。

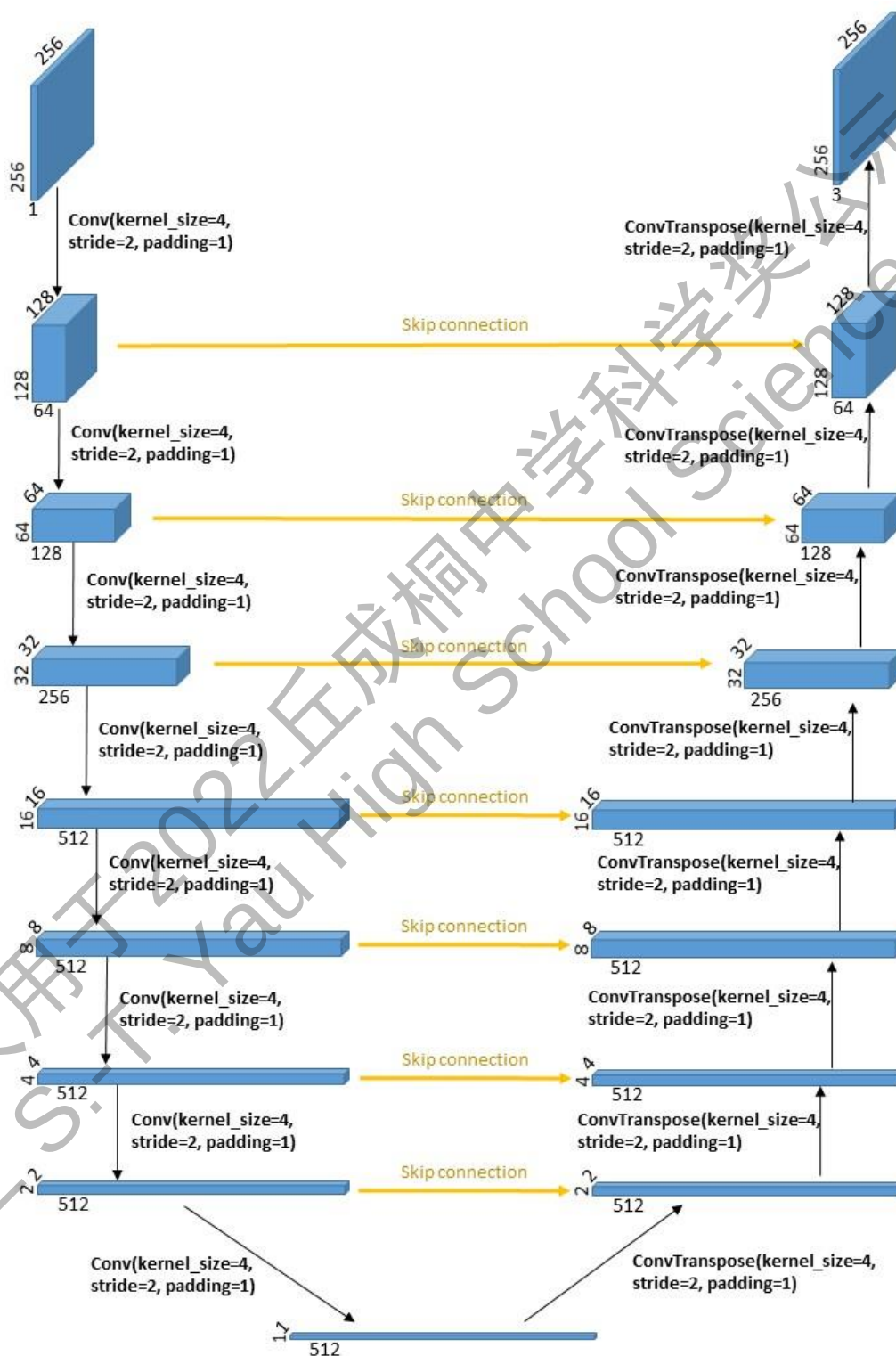


图 8: Unet 模型网络结构

5.4. 模型的训练

在模型建立完成后,开始分别使用 Resnet 和 Unet 两种框架进行模型的训练。本文所选用的激活函数是 ReLU 函数。输入图像在馈入网络之前已经调整为 256×256 的固定大小,且被转化为灰度图像馈入。训练中采用的优化器是 Adam 优化器,其中 $\beta_1 = 0.5$ 并且 $\beta_2 = 0.999$ 。本文所采用的判别器是一个感受野为 70×70 的 PatchGAN 判别器,这有利于图像细节的质量。本文分别使用 Resnet 和 Unet 两种网络框架进行训练,分别在所训练数据集上训练了 200 个 epoch,其中前 100 个 epoch 的学习速率被设定为 2×10^{-4} ,后 100 个 epoch 的学习速率线性衰减到 0。本文模型的训练在具有 Intel(R) Core(TM) i7-11800H 型号 CPU 和 NVIDIA(R) GeForce RTX(TM) 3070 型号 GPU 的平台上运行,一种网络框架训练 200 个 epoch 所用的时常为一天左右。

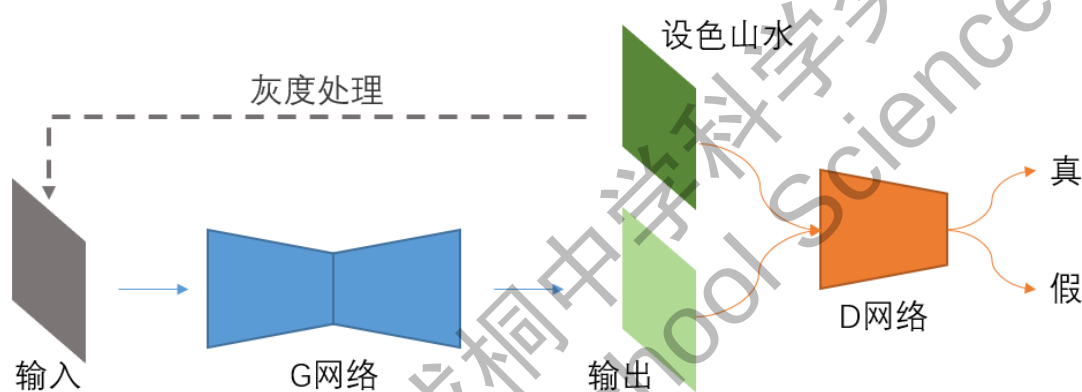


图 9: 本文模型的训练

第6章 测试与结果

在训练完成后,本文对 Resnet 模型和 Unet 模型的各个 epoch 进行了分别以设色山水图像和水墨山水图像为测试数据的测试,分别选取了 Resnet 和 Unet 两种框架各自的最佳模型,对比了两种框架的特点和优劣,并运用最佳模型实现了给水墨山水名画着色的实际应用。测试过程中,选用了 256×256 的设色山水画、水墨山水画各 100 张作为测试数据。

6.1. Resnet 模型的测试与筛选

本文对于 Resnet 模型共进行了 200 个 epoch 的训练,每 5 个 epoch 生成一个模型。将各个模型分别代入相同的数据进行测试,训练的着色水平随训练时期增大先增强再减弱。epoch5 到 epoch30 的模型测试效果较差,着色不够鲜艳浓郁,颜色较浅较暗且偏黄棕等暖色调,期待的青绿色调不够明显。这个阶段的模型还会出现明显的紫红色光斑,严重影响视觉效果,仍属于训练早期的不完善阶段。训练的 epoch35 到 epoch50 阶段,紫红色光斑基本消失,模型得到了较大的改善,但着色不充分、颜色较浅的问题依然存在,晕染略过使轮廓略微模糊,且输入图

像原本较浅的部分基本无法上色，效果依然有待改进。从 epoch55 到 epoch75，着色变得愈发鲜艳浓郁，效果愈发真实，但又出现了紫红色光斑，且图像左右两边出现了竖向的光纹。到了 epoch80 的模型，测试结果达到了最好状态，无光斑，着色鲜亮，整体为十分真实的青绿色调，轮廓清晰，晕染得当，肉眼很难与真实图片区分，训练效果好。而到了 epoch85 及以后，生成效果出现了明显的过度训练，生成图像边缘出现竖向光纹，且着色时出现大量的黄褐色，效果随进一步训练而明显下降。经过充分比较，Resnet 模型中，epoch80 的效果最为良好，适于作为使用 Resnet 结构网络中的最佳模型。

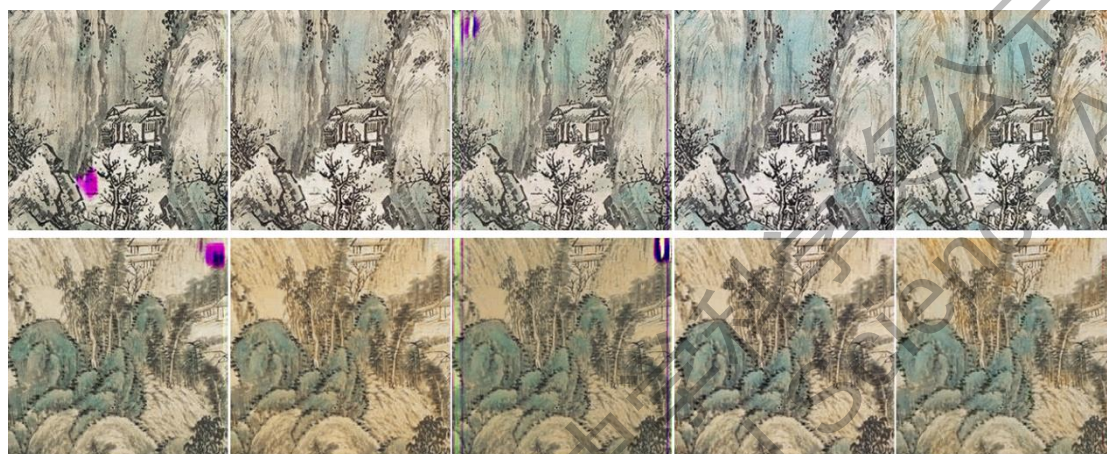


图 10: Resnet 模型在 epoch=15、35、65、80、85 时的测试效果

在选定 epoch80 的 Resnet 模型后，本文对其他数据进行了测试，发现无论使给水墨山水画着色和对设色山水画的对应的灰度图像着色与原图相对比，这个模型都取得了较为良好的效果。



图 11: Resnet 模型在水墨、设色山水画测试集上的测试效果

6.2. Unet 模型的测试与筛选

本文对 Unet 模型同样进行了 200 个 epoch 的训练，每 5 个 epoch 生成一个模型。Unet 模型生成的图像颜色丰富，青、绿、蓝、黄、红等色调兼备，自训练开

始生成的模型颜色就已经十分鲜艳。从 epoch5 到 epoch80 的模型测试结果较差，本应被着色为青绿的山水部分与树木等被着上了红色、黄色，效果不够真实，属于训练早期的不完善阶段。训练的 epoch85 到 epoch95 阶段，山体和水体开始被着色为较为真实的青绿色，而树木的枝干、木船等部分仍保留了真实的红棕色调，但生成图像的边缘出现大量黄绿色和红色的光斑，效果依然有待改进。到了 epoch100 的模型，测试结果达到了最佳状态，无光斑和光纹，着色鲜艳，山水部分为较真实的青绿色调，轮廓清晰，而木质结构、树干、土地等为真实的棕色调，树叶被着色为黄、绿等不同颜色，整体都十分真实，肉眼很难与真实图片区分，训练效果好。而到了 epoch105 至 epoch125 到 epoch200，生成效果下降，山水部分又陆续出现了大量红黄色调，并又出现了影响视觉效果的光斑等，训练到最后生成的图像颜色过于鲜亮绚烂反而影响了真实性。经过充分比较，Unet 模型中，epoch100 的效果最好，最适于作为 Unet 结构网络中的最佳模型。

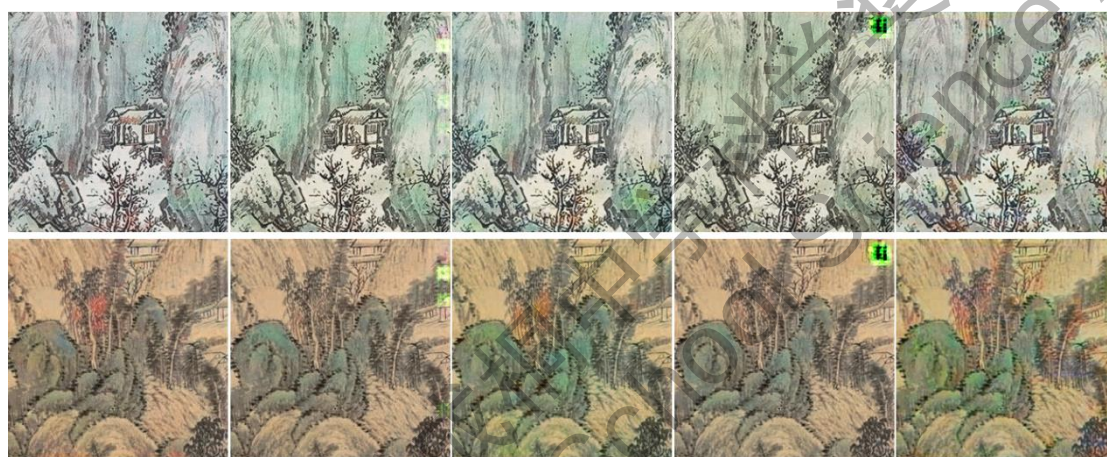


图 12: Unet 模型在 epoch=50、85、100、125、200 时的测试效果

在选定 epoch100 的 Unet 模型后，本文又对设色山水画的对应的灰度图像进行着色，与原图相对比肉眼观察无太大区别，同样取得了较为良好的效果。



图 13: Unet 模型在水墨、设色山水画测试集上的测试效果

6.3. Resnet 与 Unet 模型的对比

将 Resnet 网络生成的图像与 Unet 网络生成的图像进行直观对比，可以发现二者的风格有较大的差异。Resnet 模型生成的图像多数只带有青绿色调，且整体色彩仍较淡。经观察，该模型通常只对山水部分着青绿色，而对非山水部分（树木、船只等）不着色或只着淡棕色，效果与水墨山水画区别不大。而 Unet 生成的图像色彩丰富，细节准确，山和水被着色为青绿，树叶为黄绿，桥、船、树枝干为红棕，瀑布、云朵等不着色或者着浅白色，效果更真实，细节更逼真。因此，经过直观观察和比较，Unet 模型的效果整体优于 Resnet 模型，最佳模型为 epoch 为 100 时的 Unet 模型。

本文还对于模型测试的结果进行了广泛的调研，设计了含有 10 道题的调查问卷，每道题的四个选项中有一张图是模型生成的假图像，三张图是真实的设色山水画图像，调查要求用户选出模型生成的假图像。10 道题中，假图像的选项分别为 5 张 Resnet epoch80 和 Unet epoch100 生成的设色山水图像，用于在检测模型效果的同时进一步对比两种网络的模型质量。本调查共收集到来自 165 人的数据，这些参与调查的用户包含了学生、老师、从事其他工作的成年人，还包含了几位专业的美术教师。十道题选出模型生成的假图像的正确人数及正确率如下表所示：

表 1：模型生成设色山水画真实性的用户调研结果

Unet	正确人数	正确率	Resnet	正确人数	正确率
1	26	15.76%	6	76	46.06%
2	30	18.18%	7	15	9.09%
3	34	20.61%	8	40	24.24%
4	19	11.52%	9	25	15.15%
5	13	7.88%	10	9	5.45%

综合统计计算，165 位用户对 Resnet 模型判断的平均正确率只有 19.998%，对 Unet 模型判断的平均正确率则只有 14.79%。这样的正确率甚至低于每个选项的平均概率 25%，充分证明了模型生成的真实性之高，正常人肉眼难以分辨，甚至部分生成的比真实图像还要“真实”。而专业美术教师在辨别生成图像的过程中，正确率与其他用户相差无几，足以说明模型生成的设色山水画图像在专业角度不存在可被肉眼辨识出的缺陷，生成效果足够真实。根据用户选择，Unet 模型比 Resnet 模型调查的结果的正确率更低，则说明从正常人的视角来看 Unet 模型比 Resnet 模型生成的效果更真实美观，效果更加良好。

在直观观察和用户调查之外，本文还对两种模型进行了定量分析和比较，其中使用了 FID(Frechet Inception Distance)^[12]技术和学习感知图像块相似度(LPIPS)^[13]来定向分析模型生成的设色图像与真实的设色山水画图像的相似程度。

FID 分数由 Martin Heusel 等人于 2017 年在论文 GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium 中提出，是计算真实图像和生成图像的特征向量之间距离的一种度量。FID 从原始图像的计算机视觉特征的相似度来衡量两组图像的相似度，这种视觉特征是使用 Inception v3 图像分类模型计算的得到的。分数越低代表两组图像越相似，在最佳情况下的得分为 0，表示两组图像相同。本文将真实设色图像与 Resnet 和 Unet 模型测试生成的

图像分别计算 FID 分数,得到两个模型分别与真实图像的相似度。Resnet 模型的 FID 值为 96.77755993238543,而 Unet 模型的 FID 值为 101.58548317575247。可以看到,Resnet 模型的 FID 值略低于 Unet 模型的 FID 值,说明 Resnet 模型生成的图像较 Unet 模型生成的图像更贴近原图。

学习感知图像块相似度(LPIPS)由 Richard Zhang 等人于 2018 年在论文 The Unreasonable Effectiveness of Deep Features as a Perceptual Metric 中提出,引入了 2AFC、JND(Just Noticeable Difference)两个评估指标,设计了用于学习感知相似度(Perceptual Similarity)的 LPIPS 模型。LPIPS 的值域为 0 到 1,数值越接近 0 说明两张图像相似度越高,数值越接近 1 说明两张图像相似度越低。将 Resnet 生成的测试图像与原始设色图像计算比较 LPIPS 相似度,得到的平均结果为 0.1711;而将 Unet 生成的测试图像与原始设色图像计算比较 LPIPS 相似度,得到的平均结果为 0.1847。通过计算 LPIPS 的定量分析,可以看出两种模型生成的图像都与原图像相似度较高,且 Resnet 模型略优于 Unet 模型。

经过综合比对,可以看出:Resnet 模型经过定量计算分析,与设色山水画原图的相似度高于 Unet 模型,理论上优于 Unet 模型;Unet 模型据直观观察和用户调查生成效果更真实美观,即生成效果更能被当代人的审美接受。因此,两种模型各有优劣,应根据实际需要酌情选用。

第7章 讨论

7.1. 结论

本文运用了生成对抗网络和 pix2pix 技术,选用了 Resnet 和 Unet 两种不同的网络框架,运用大量的设色山水画图像数据训练了基于这两种网络框架的着色模型,均取得了较好的效果。Resnet 网络框架的最佳模型为训练到第 80 个 epoch 时的模型,Unet 网络框架的最佳模型则是训练到第 100 个 epoch 时的模型。

Resnet 和 Unet 模型各有各的特点:Resnet 因为网络深度更深,可能具有更强的学习能力,能提取到深层语义特征,更侧重于还原原始设色图像的忠实度,因此定量测试中的结果更加优秀;而 Unet 因为具有跳层连接,能够更加充分地利用语义信息,因此其可以保留更多的细节。山水画图像着色的方案本来就不是唯一的,方法应有多样性,原画作的着色方法只是其中的一个解,绝不是唯一解。Unet 为着色方案提供了更多的选择,广泛的用户调查足以表明,大多数用户同样赞赏 Unet 的生成结果,认为其具有真实性和艺术性。因此,Resnet epoch80 模型和 Unet epoch100 模型都是适用于山水着色的优秀模型。

本文所进行的山水着色的初衷是为了进行水墨山水的着色以及褪色设色山水画的修复,因此,本文所打造的模型需运用于水墨山水着色的实际任务中。将测试数据中的由水墨山水——南宋赵黻的《江山万里图》(局部)裁剪出的 256×256 的测试图像用 Resnet 和 Unet 着色后拼接起来,得到完整的《江山万里图》(局部)的设色版本,如下图:



图 14: Unet 模型生成的设色山水画（拼接）



图 15: Resnet 模型生成的设色山水画（拼接）

而 Resnet 模型可以用于任何尺寸的图像着色中，因此可以整体生成设色山水画，可以与由 256×256 图像拼接而成的设色结果进行对比，如下图：

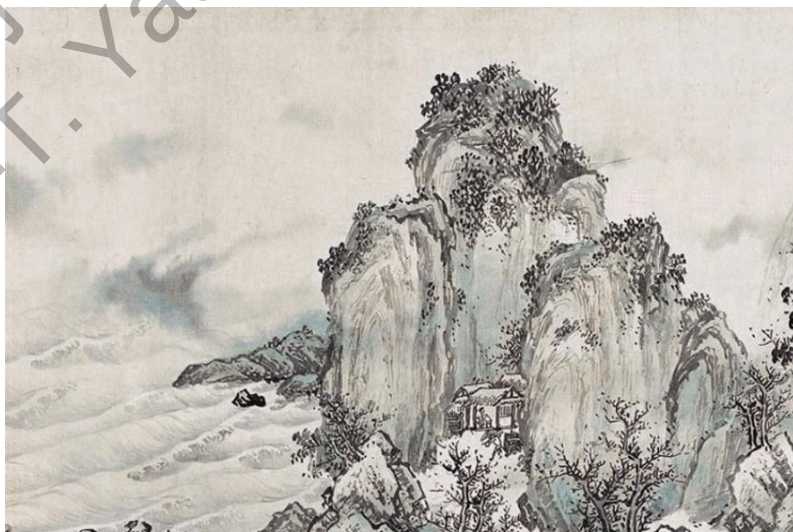


图 16: Resnet 模型生成的设色山水画（整体）

综上所述，本文运用 pix2pix 算法较为完善地成功解决了山水画着色问题，对山水画的着色真实美观，且具有现实实用意义，值得推广使用。

7.2. 创新点

本文首次创造性地提出了将生成对抗网络和 pix2pix 技术应用于山水画领域，首次进行了山水画自动着色的研究，得出了效果优秀的山水画自动着色模型。本文改良、修饰了原有的 pix2pix 模型，使 pix2pix 模型更好地用于山水画着色模型的构建。选取了适于山水画着色模型的参数进行训练，采用了不同网络框架进行训练并对结果进行对比分析，选择出了最适于水墨山水画着色的最优模型。此外，本文提出并引入了 LPD 山水画数据集，可供后续山水画着色等相关研究用于模型的训练等。

7.3. 不足

无论是 Resnet 和 Unet 网络模型，对山水画的着色依然不够鲜亮，难以达到如《千里江山图》等青绿山水画的鲜艳颜色，在对此类图像着色时生成的设色图像与原图的相似度较低。同时，尽管每一张 256×256 的图像的设色效果都较真实，但在拼接为完整山水画图像时，相邻两张图像可能会存在较大色差，导致两张图像拼接处存在较大的颜色割裂感，使整体观感下降，仍需要改善。

而 Resnet 模型生成整体图像的策略也仍存在一定问题，其整体生成的山水画效果与对 256×256 图像生成的效果差距较大，着色不够鲜亮、色调偏灰暗，仍需要继续改善其效果。

7.4. 展望

本文已经构建出了可以完成山水画着色任务的较优模型，充分论证了生成对抗网络和 pix2pix 模型用于山水画着色任务的可行性和优越性。未来，作者将继续完善山水画着色模型，使其着色更加鲜艳、生成图像更加贴近真实山水画，完善、解决同一张画中相邻切割图像的色差问题和交界问题，从而进一步提升模型的性能和效果。

在使用方面，希望本文所构建的模型能够广泛地应用于真实的山水画着色领域，实用于水墨山水画到设色山水画的风格转换和褪色古画的电子修复工作，做到科学理论研究有所实际应用，为保护宣扬传统艺术、弘扬中华优秀传统文化做出科技方面的贡献。

第8章 参考文献

- [1] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 27.
- [2] 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.

[3] 李大锦. 基于纹理合成的水墨山水画自动绘制[J]. 计算机应用, 2009, 29(09): 2406-0410.

[4] 刘晓崇, 赵凡, 张文俊. 基于真实场景的水墨风格图像自动生成方法[J]. 电视技术, 2014, 38(15): 249-252.

[5] Xue A. End-to-end chinese landscape painting creation using generative adversarial networks[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021: 3863-3871.

[6] Anwar S, Tahir M, Li C, et al. Image colorization: A survey and dataset[J]. arXiv preprint arXiv:2008.10774, 2020.

[7] Chen X, Duan Y, Houthoofd R, et al. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets[C]// Proceedings of the International Conference on Neural Information Processing Systems. 2016: 2172-2180.

[8] Mirza M, Osindero S. Conditional generative adversarial nets[J]. arXiv preprint arXiv:1411.1784, 2014.

[9] Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier gans[C]//

Proceedings of the International Conference on Machine Learning. 2017.

[10] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[11] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[12] 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.

[13] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 586-595.

第9章 致谢

本文研究的顺利完成首先要衷心感谢北京市青少年拔尖人才培养计划和北京青少年科技后备人才培养计划提供的进入大学实验室进行学习和科研工作的机会,感谢中国科学院大学计算机科学与技术学院数字媒体技术研究实验室提供的学习和科研机会。感谢实验室黄庆明老师的指导,感谢苏荔老师的细心指导,她给我耐心讲解深度学习、图像处理 and 图像着色的相关知识,指导学习相关理论,并对项目的整体方向、研究操作提出了宝贵的建议。

感谢北京四中的科技活动指导老师孔丽燕老师、高增老师，帮助我安排各项科技学习活动，并对我确定选题和项目进展提出了宝贵的建议。感谢北京四中艺术组于明老师为本文涉及美术方面的部分提供了专业的指导。

最后，还要感谢家人和朋友们在我研究遇到困难时给予的关心和支持，感谢身边的家人、老师和同学们积极帮助填写调查问卷，极好地反馈了本文的研究效果。

仅用于2022丘成桐中学科学奖公示
2022 S.-T. Yau High School Science Awards