**《Web信息处理与应用》大作业**

**——网络图片风格迁移探究**

李家郡

2019104205

**【摘要】**

由于经典画作毕竟是少数，为了得到满足用户需要的图片，有必要将原有的一些图片的风格迁移到其他画作或者是照片中。为探究现有的风格迁移发展，本次作业通过比较两个具有代表性的风格迁移实验，Neural Style Transfer、GANILLA，探究现有的风格迁移未来的发展可能性。其次，本课程也是希望将现有的机器学习的技术应用到现实当中。结合后羿采集器，本次实验探索了对网络图片的采集过程。最后，由于出于本人对画师WLOP的喜爱，也希望能通过本次实验创造一些具有WLOP风格的图片，并对风格迁移的适用对象进行探索。由于风格迁移的不易量化性，在实验结果评定时，采用用户打分制。

关键词： 风格迁移，特征提取，图像翻译

1. 引言

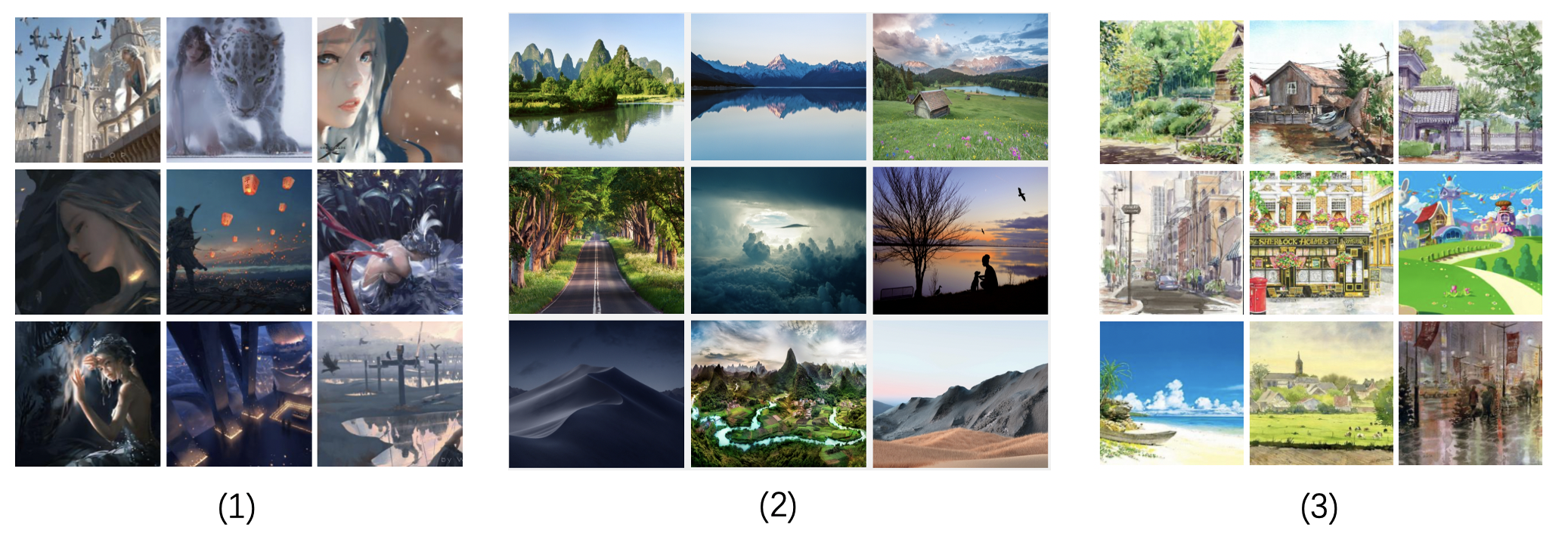
风格迁移指的是将一个图片的风格转换到另一个图片中。风格的定义则比较模糊，这也是风格迁移任务无法定义一个比较合适的量化评价指标的原因。画作风格大多数是从艺术的角度出发，由主观决定的一种属性。常见的作画风格有现实主义、印象派、现代主义等。除了公认的作画风格派别外，部分有个人特色的画家也是可以看出明显风格的，例如凡高、齐白石等。风格迁移主要由Gatys等人开始。自从对抗神经网络的提出以来，风格迁移得到很好的发展。其迁移的目标也从迁移风格转化为迁移各种性质，如动物迁移，地理照片转化为地图的简图等等。本次实验，主要关注于对比起始的Gatys的风格迁移与最新的利用GAN做插画迁移的模型GANILLA。希望从中得到一些风格迁移的发展脉络和趋势。

Web信息主要包含文字、图片、视频，其中图片因为其直观和便利性，其实使用最为广泛。对于图片信息的采集也是比较容易的。本次实验也希望能通过风格迁移的任务，探索采集和处理Web图片的过程，总结出一套利于训练的数据收集模式。

本次实验主要关注于中国画师WLOP的风格转化。WLOP的画作极具个人风格，但其作品并不是很多。希望能将插画数据集和风景数据集尝试转化一部分为WLOP风格的作品。并通过用户打分的方式，比较经典Neural Style Transfer和最新的GANILLA两种模型的效果。

1. Web数据
   1. 数据采集

后羿采集器是一款极其容易操作的爬取网络数据的软件，只需要输入对应的网址，即可根据需要爬取对应的表格、段落和图片。其比较显著的优势有两点，其一是操作简单，极易实现，操作指南比较详细。其二是容易导出爬取的数据，且无需收费。



图表 1数据集部分图片例子

如图表1，本文使用的(1)WLOP画作均从其主页爬取[[1]](#footnote-1)，(3)插画数据集则是由本人过去从网络收集所得，(2)风景照片数据集从彼岸图网[[2]](#footnote-2)爬取。

收集所得的图片数据数量如下表所示：

|  |  |  |
| --- | --- | --- |
| 数据名 | 总数量（张） | 大小(MB) |
| WLOP | 371 | 75.9 |
| Landscape | 1028 | 134.7 |
| Illustration | 1037 | 165.7 |

表格 1图片数据规模

图片数据集将放在以下百度网盘中：

链接: https://pan.baidu.com/s/19JUViUynlu3wfGesF-Fvvg 提取码: vnja

* 1. 数据预处理

网络图片数据通常有维度不统一，大小不一致的问题。同时由于现有的WLOP的图片数据比较少，为了将数据集稍微扩充，直接使用切割舍弃的方式是比较不合理的，直接将图片缩放为正方形会影响用户直观的评价。所以本次实验采用先切割后缩放的方式，既能将数据集扩充，也不影响图片的形状。

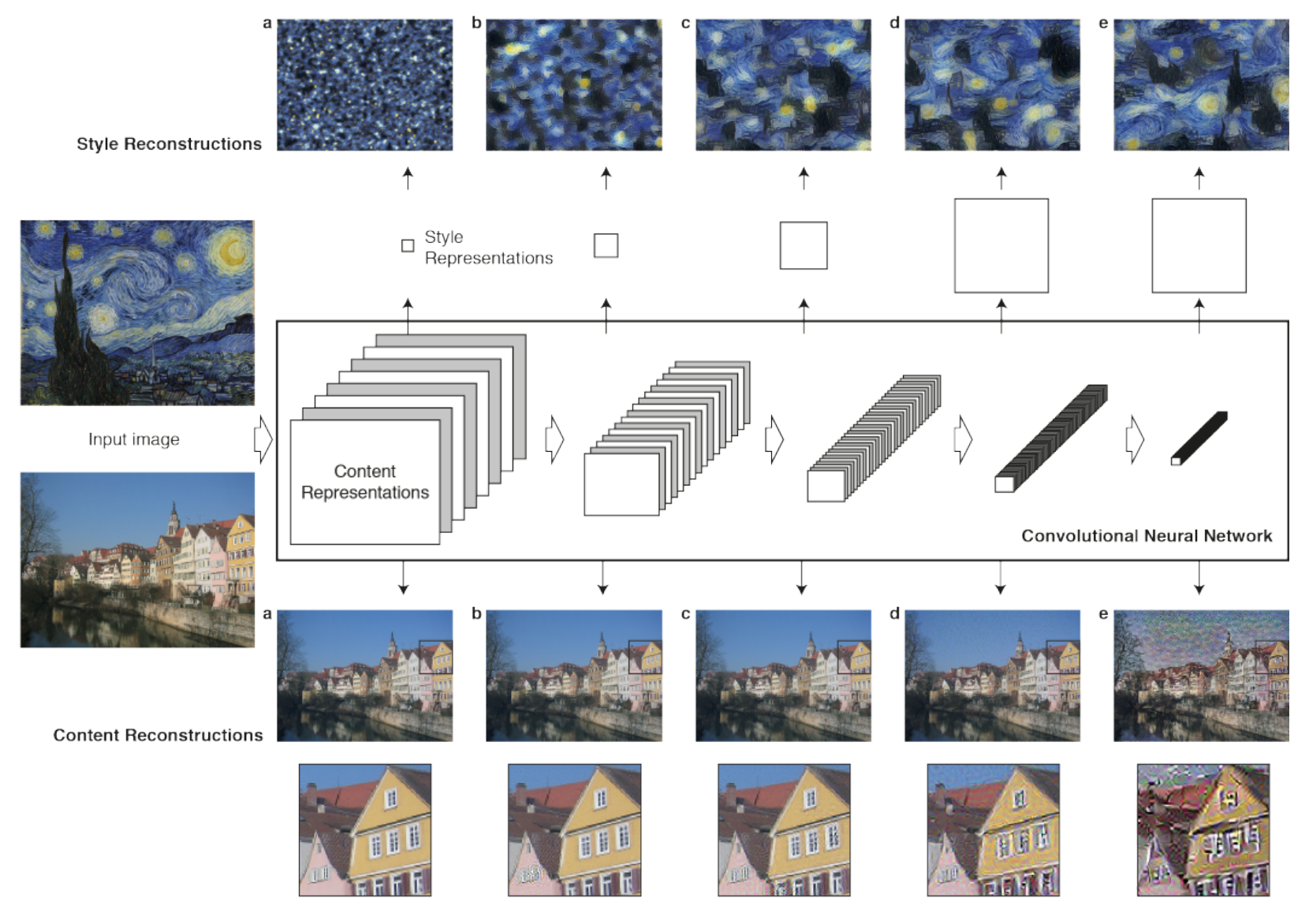
对于X\*Y的图片，若Y=X，直接缩放即可。则不妨设Y>X，若Y为X的整数倍，将Y拆为Y/X份即可。若Y不为X的整数倍，则可以将原图分为份，由于，则利用Y中部分段重叠使用即可。举个简单的例子，对于长为180，宽为100的矩形，要将其切成两个正方形，第一个正方形长的范围为0~100，而第二个正方形长的范围为80~180即可。利用图片的分割和缩放，预处理得到的数据集大小如下：

表格 2实验所用数据集规模

|  |  |  |
| --- | --- | --- |
| 数据名 | 总数量（张） | 大小(MB) |
| WLOP | 638 | 8.8 |
| Landscape | 2052 | 29.4 |
| Illustration | 4085 | 76.5 |

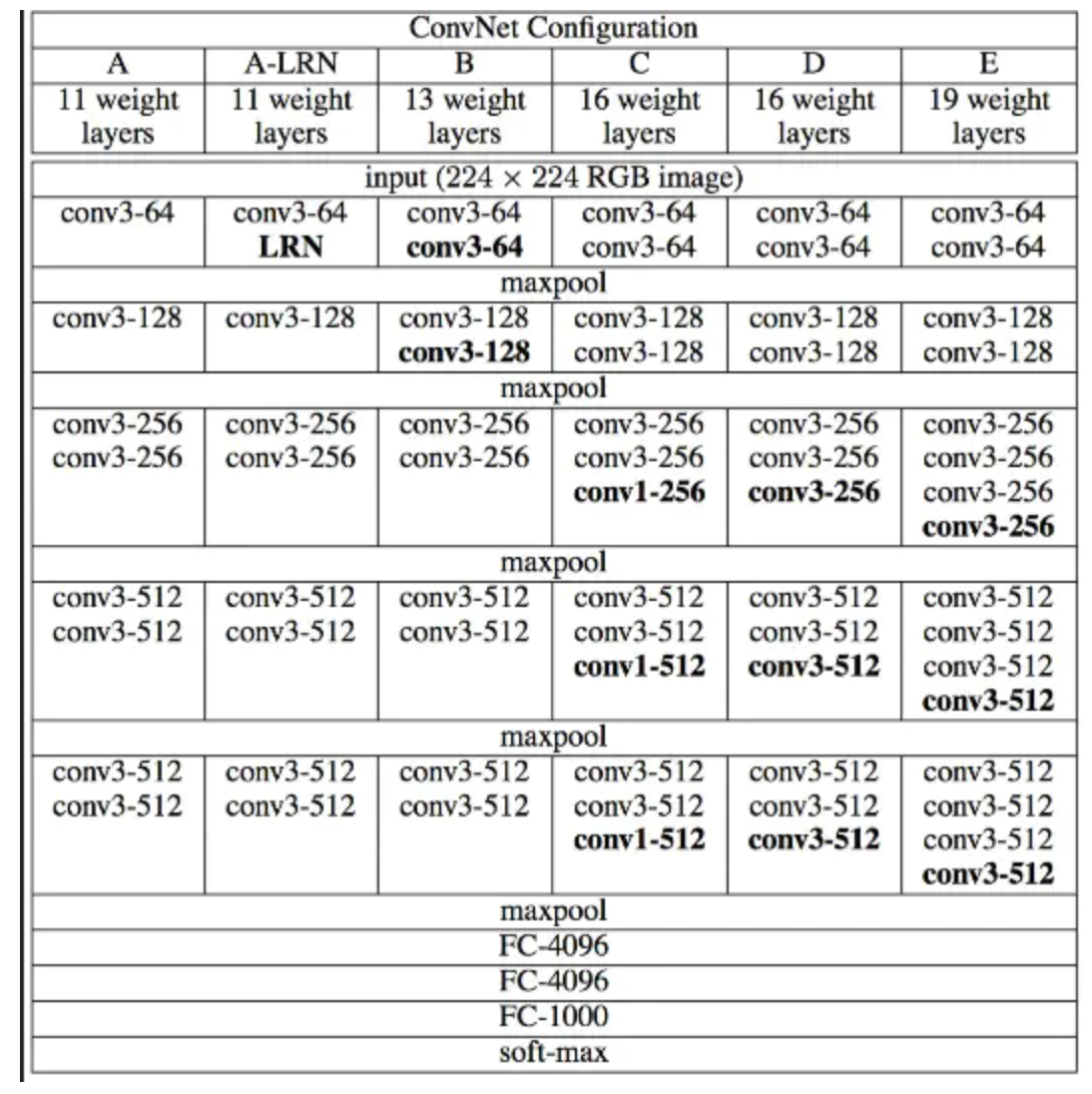
1. 模型简介
   1. Neural Style Transfer

在Gatys之前的风格研究主要有纹理生成算法和物体识别模型。Gatys的模型从本质上是将两种提取特征的方法分别进行研究，并定义损失函数。由于存在风格损失和内容损失，在调整模型时，可以通过控制两种损失的比例，训练得到既能保存内容也能有目标风格的图片。



图表 2Neural Style Transfer(CNN)

Gatys的NST使用网络主要是VGG，而本次实验中使用的是VGG19。VGG的特点是不断利用卷积提取特征的网络结构和图像的识别效率。



图表 3 VGG网络的结构

对于图像的风格迁移，目标是使得生成的图像需要具有原图片的内容特征，且有风格图片的纹理特征。由此，可以定义两个loss值：即内容loss和风格loss，之后只需要根据结构调整两种loss的比例即可。

内容损失为对两张图片直接求特征并进行损失计算，损失函数定义为：

风格损失由多个特征一起计算，首先需要计算一个Gram Matrix：

Gram Matrix可以看成是特征的偏心协方差矩阵，Gram Matrix代表着不同特征之间的一种综合能力，调整它的大小有助于把握图像整体的风格。有了表示风格的Gram Matrix，要度量两个图像风格的差异，只需比较他们Gram Matrix的差异即可。 故在计算损失的时候函数如下：

综上，NST模型的主要步骤为，粗略决定一个损失权重比例，训练生成图片的参数，使得loss达到最小，再根据需要调整权重比例不断重复即可。

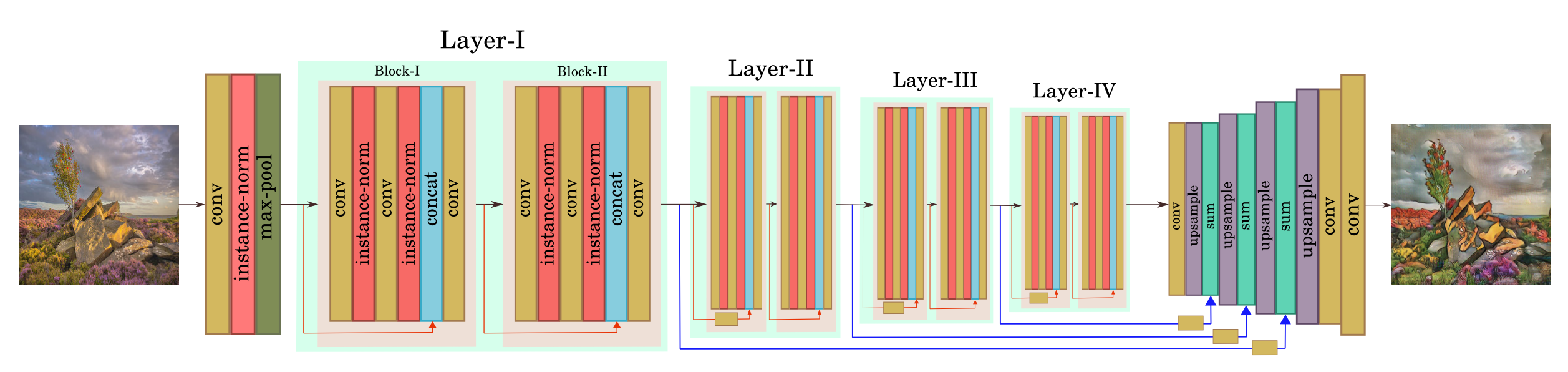
* 1. GANILLA

GANILLA是今年3月提出的模型，主要想解决的是之前GAN模型中不足的两个点：过分关注内容信息，导致风格没合理的转化到另一张图片。过分关注于风格，导致图像中出现奇怪的特征。文中举了以下的例子：



图表 4 从CycleGAN，DualGAN，GANILLA中生成的例子

可以从上图看出，CycleGAN过度关注于风格的转化，导致图像中出现了奇怪的天空，而DualGAN虽然看起来是比较自然的图片，但其风格基本没有任何变化。基于以上两种方法的局限性，GANILLA的模型则相当于一种综合，结合两者的优势并得到一种平衡。下图是其主要网络结构：



图表 5 GANILLA生成的网络

GANILLA基于现有效果最好的模型，提出两点改变。首先，作者对每个残差层的特征图做下采样。此外，作者提出用skip connections 和上采样来将低级特征和高级特征进行融合。低级特征指轮廓信息，这些信息能更好地帮助图片保留结构。总的说，本文没有对网络内部做太大的调整，而是提出了一种网络架构。希望能在风格迁移和内容保留中得到较好的平衡。

1. 实验

本次作业的实验都在有GPU的linux服务器上实现。实验代码主要参考了两篇论文在github上的实现，并对参数稍微进行了调整。由于风格迁移的结果具有主观性，无法有一个比较好的量化效果，所以本人设计了一个展示选择程序，并邀请5人，对NST和GANILLA生成的结果进行选择。每个人完成100组判断，根据5人判断的平均值，我们可以统计得到一个关于两种方法效果的对比结果如下表：

表格 3 风格迁移模型用户打分

|  |  |  |
| --- | --- | --- |
|  | NST得分 | GANILLA得分 |
| 风景照片转WLOP | 32.8(2.9) | 67.2(2.9) |
| 插画图片转WLOP | 8.6(1.1) | 91.4(1.1) |

为了让读者有一些直观对本次实验结果的感受，以下展示部分两种风格迁移方法的结果对比：

表格 4 插画转WLOP风格结果展示

|  |  |  |  |
| --- | --- | --- | --- |
| GANILLA | NST | 原图 | 参考图片风格 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

表格 5 风景照片转WLOP风格结果展示

|  |  |  |  |
| --- | --- | --- | --- |
| GANILLA | NST | 原图 | 风格图片 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

可以从以上展示的部分结果看出，总体上GANNILA转化的效果比较稳定，由于其训练过程是整个数据集进行训练，有效对抗了因为风格图片本身的噪声产生的偏差。在NST中，由于WLOP作品边缘不清晰，有时候会产生类似于白噪声的图片，限制了它的效果。GANLILLA生成的图片，虽然不一定都更接近原图，但整体上会更趋近于WLOP作画的风格，即善用暗色，并使用明亮的色调，突出整个画面的对比效果。

对比插图画和风景转化的效果来说，从画转化到画的效果，GANILLA的用户评分会稍微高一点，其原因可能是GANILLA更能保证两种损失都不太大，而NST则是一个固定权重，在识别插画边缘时，GANILLA能更大程度保证插画不变形，NST识别插画时，则更容易变形。所以最终的结果是在插画转化时，GANILLA能得到更高的分数。

1. 总结

本次实验通过对比2015年的NST，和2020年的GANILLA，探究了风格迁移的发展趋势，并通过用户打分的策略，得到了对模型的评价。可以看出，图像迁移领域已经取得比较明显的变化。GAN能更稳定地转化图片的风格，GANILLA提出的综合考虑内容和风格的框架，实际是对NST的一种延伸。用户打分的评价标准始终不是最好的量化手段，在未来也许会有人探索更加合适的风格迁移量化指标。

**参考资料**

1. Hicsonmez, S., Samet, N., Akbas, E., & Sahin, P.D. (2020). GANILLA: Generative Adversarial Networks for Image to Illustration Translation. *Image Vis. Comput., 95*, 103886.
2. Gatys, L. A. , Ecker, A. S. , & Bethge, M. . (2015). A neural algorithm of artistic style. *Computer Science*.
3. <http://fancyerii.github.io/books/neural-style-transfer>[/](http://fancyerii.github.io/books/neural-style-transfer/)
4. <https://github.com/giddyyupp/ganilla>
5. <https://www.jianshu.com/p/6e754f66a169>学习一个宫崎骏画风的图像风格转换GAN
6. <https://zhuanlan.zhihu.com/p/26746283> 图像风格迁移(Neural Style)简史

1. https://azz.net/wlop/illust [↑](#footnote-ref-1)
2. http://pic.netbian.com/4kfengjing/ [↑](#footnote-ref-2)