

DBSCAN聚类算法在图像风格迁移中的应用

王飞龙¹, 田江波¹, 蒋博蓝¹, 曾文广¹, 叶忠勇¹, 黄宏博^{1,2*}

¹北京信息科技大学计算机学院, 北京

²北京信息科技大学计算智能研究所, 北京

收稿日期: 2021年11月25日; 录用日期: 2021年12月21日; 发布日期: 2021年12月28日

摘要

尽管图像风格迁移技术在不断发展, 但现阶段对于图像风格迁移的效果而言仍有较大的提升空间。为了使图像在风格迁移后具有与目标风格图像更贴近的颜色特征, 本文基于GAN (Generative Adversarial Network)模型提出了一种使用DBSCAN (Density-Based Spatial Clustering of Application with Noise)聚类算法定义颜色数量的损失函数并用于约束模型的训练, 从而使生成的图像更趋近于目标风格图像。实验结果表明在引入由DBSCAN算法定义的颜色数量损失函数后图像风格迁移效果得到了更好的提升, 生成图像在颜色上更具有原风格图像的特征, 视觉上更具立体感。

关键词

图像风格迁移, DBSCAN, 聚类算法, 生成对抗网络

Application of DBSCAN Clustering Algorithm in Image Style Transfer

Feilong Wang¹, Jiangbo Tian¹, Bolan Jiang¹, Wenguang Zeng¹, Zhongyong Ye¹,
Hongbo Huang^{1,2*}

¹Computer School, Beijing Information Science and Technology University, Beijing

²Institute of Computational Intelligence, Beijing Information Science and Technology University, Beijing

Received: Nov. 25th, 2021; accepted: Dec. 21st, 2021; published: Dec. 28th, 2021

Abstract

Although the image style transfer technology is constantly developing, there is still much room for improvement in the effect of image style transfer. In order to make the image have similar color

*通讯作者。

文章引用: 王飞龙, 田江波, 蒋博蓝, 曾文广, 叶忠勇, 黄宏博. DBSCAN 聚类算法在图像风格迁移中的应用[J]. 计算机科学与应用, 2021, 11(12): 3051-3059. DOI: 10.12677/csa.2021.1112308

characteristics with the target style image after the style transfer, this paper introduces a color-quantity-criteria defined by the DBSCAN (Density-Based Spatial Clustering of Application with Noise) clustering algorithm. The transfer model is designed based on the GAN (Generative Adversarial Network). The color-quantity-loss-function is used to constrain the training of the model, so that the generated image is closer to the style image in the number of colors. The experimental results show that the image style transfer effect of the proposed algorithm is better improved, and the generated image not only has the characteristics of the original style image in color but also has more stereoscopic visual effect.

Keywords

Image Style Transfer, DBSCAN, Clustering Algorithm, Generative Adversarial Network

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着近年来深度学习技术的发展, 图像处理领域掀起了一波新的发展浪潮, 其中图像风格迁移也是研究热点之一。图像风格迁移技术的发展在计算机视觉和艺术创作等领域有着极其重要的作用, 如何提高和实现更好的图像风格迁移效果不论从技术上还是在应用上都有迫切的需求。目前主流的图像风格迁移技术大都是基于深度神经网络, 不同的技术涉及不同的网络模型, 而不同的模型使用了不同的损失函数进行训练, 实现的图像风格迁移效果也各有优缺点。由于图像可以映射成 R、G、B 三维空间中的像素点集合, 相同的颜色块在 R、G、B 三维空间中表现为一类邻近的像素点, 通过聚类算法可以对一个样本集中的样本点依据样本点的分布特性对样本点进行划分, 因此本文认为可以通过聚类算法对图像在三维空间中的像素点进行聚类从而实现图像风格迁移过程中图像颜色数量上的约束, 并在此基础上提出了用聚类算法定义的损失函数来约束模型训练的过程, 使得风格迁移后的图像在颜色数量上能够趋近于原风格图像, 从而实现更好的图像风格迁移效果。

2. 相关研究

2.1. 神经风格迁移

基于神经风格的图像风格迁移方法在对图像风格化的过程中, 主要分为对图像进行纹理建模和图像重建这两大步骤。这两种方法分别研究如何对纹理信息进行表示, 以及如何把输入的给定特征进行还原, 并输出一张具有特定内容、特定风格的图像。目前基于神经风格迁移主流的算法主要有基于最大均值差异、基于马尔科夫随机场、基于深度图像类比三类。

2.1.1. 基于最大均值差异的图像风格迁移

在 2015 年, Gatys 等人在前人已有的研究基础上发现了 VGG-19 网络所提供的良好的特征提取能力可以很好的匹配图像风格迁移[1]。Gatys 的关键发现在于: 神经网络在训练的过程中风格和内容是分离的, 而通过构造和使用 Gram 矩阵可以提取出任意图像的风格特征表示。

2.1.2. 基于马尔可夫随机场的图像风格迁移

Gatys 等人通过使用 VGG 网络来提取图像高层表征的方法, 只能使不同卷积层捕获图像每个像素之

间的联系,而在空间分布上缺少联系。基于以上问题, Li 等人提出了结合马尔科夫随机场和卷积神经网络的模型[2],他们将 Gatys 模型中的 Gram 矩阵匹配替换为马尔科夫正则化模型。马尔科夫模型可对具有同类特征信息的集合进行提取和描述。因此使用马尔科夫模型可将 CNN 的图像特征分别分割成许多区域块,并通过区域块的匹配提高合成图像在视觉上的合理性。

2.1.3. 基于深度图像类比的图像风格迁移

Liao 等人通过结合 VGG-19 和图像类比(PatchMatch)的方法进行风格迁移[3]。他们通过使用 VGG19 提取图像特征,并将 PatchMatch 的应用扩展到特征领域。使用这两种方法的结合可以让图片在进行外观上转换的同时,可以保留感知上相似的纹理结构。并在两个语义相似的输入图之间找到有关联的对应关系。基于深度图像类比的图像风格迁移可以对语义级的视觉属性迁移有很好的指导作用,但是这种方法缺乏泛化能力,当输入图片语义相差过大时则无法产生很好的效果。

2.2. 基于对抗生成网络的图像风格迁移

尽管基于神经网络的图像风格迁移研究方法在不断改进,但通过其训练方法生成的图像依旧只能体现出简单的纹理转变,而无法显示对应图片本身的风格特点。基于上述问题,研究者们发现使用生成对抗网络(GAN)能一定程度上解决这种问题。

2.2.1. 基于有监督模型训练的图像风格迁移

Mirza M 等人在 2014 年提出了基于对抗生成网络的 CGAN 模型[4],为图像风格迁移提供了新的研究方法。研究通过在生成网络和判别网络中额外的引入了一个条件 y ,来指导模型训练。通过添加条件 y ,在训练过程中,模型需要分辨图片真伪的同时还要结合输入条件 y 与图片的匹配情况进行双重判断。通过这种训练模式,可以让模型更好的学习风格图片与内容图片间对应关系。

2.2.2. 基于无监督模型训练下的图像风格迁移

CycleGAN 模型[5]为解决上述的困难,使用非成对数据集,进行图像风格迁移训练。它们通过将图片分为不同的域,通过实现不同的域之间的图像映射进行风格迁移训练。通过这种方法 CycleGAN 解决了模型训练过程中数据集的问题,但 CycleGAN 模型无法很好的捕捉漫画风格,依然存在输出图像无法充分保留输入图语义结构的问题。

基于上述问题清华大学 SubodhDeolekar 团队提出了 CartoonGAN 模型[6]。他们引用了一种新颖的网络架构和针对漫画风格进行设计的损失函数。最大程度地在使用不成对数据集训练的情况下保留原动画的特征。但是 CartoonGAN 在针对人像生成方面的效果并不理想,在人像的面部保留上会出现色块间相混淆的现象。

2019 年 Jie Chen 等人根据 CartoonGAN 暴露出的问题提出了 AnimeGAN 模型[7]。AnimeGAN 模型相较 CartoonGAN 的网络结构,采用了一种更加轻量级的生成器架构。并且使用了三个全新的损失函数用来解决 CartoonGAN 在生成人像时暴露出的问题,同时也使得动漫图像风格迁移的效果得到了更加明显的提升。并在 2021 年更新并开源了 AnimeGANv2 版本[8],解决了生成图像中的高频伪影问题,在充分保留原图像特征的前提下,又进一步减少了生成器网络的参数数量。

2.3. 聚类算法

2.3.1. 基于划分的聚类

基于划分的聚类算法是通过对于给定的 n 个数据对象的数据集 D ,将数据对象组织成 k ($k \leq n$) 个分区,每个分区代表一个簇。该聚类算法的聚类目标是使得类内的点足够近,类间的点足够远,常见的有 k-means

及其衍生算法。这类算法优点是速度快，缺点是必须提前知道 k 值，算法容易陷入局部最优，同时运用这类聚类算法的数据必须符合“数据之间的相似度可以使用欧式距离衡量”，否则分类效果会很差。

2.3.2. 基于密度的聚类

基于密度的聚类是一种非常直观的聚类方法，即把临近的密度高的区域聚成一片形成簇。该类算法的主要目标是使用更少的领域知识来确定输入参数，发现任意形状的簇并在大规模数据库上具有更高的效率。该类算法能够将足够高密度的区域划分成簇，并能在具有噪声的空间数据库中发现任意形状的簇。其优点是不需要指定簇的数目，可以发现任意形状的聚类簇，对噪声不敏感。不足之处是需要设置两个固有参数即半径 Eps 和 $MinPts$ ，空间聚类密度不均匀时难以设置参数。

2.3.3. 层次聚类

层次聚类算法是通过计算不同类别数据点间的相似度来创建一棵有层次的嵌套聚类树，包括合并的层次聚类和分裂的层次聚类。该类算法基于聚类稳定性，使用了提取平面聚类的技术，可以处理密度不同的聚类问题。其优点是不需要自行设置阈值，只需定义最小簇的数量同时计算消耗相对小，速度较快，而且参数敏感度较低。

3. 方法

3.1. DBSCAN 聚类算法

DBSCAN 聚类算法是一种基于密度的聚类算法。该算法用密度作为样本的分类标准，即同一类别样本之间是紧密相连的。通过将紧密相连的样本划分为一类就得到了一个聚类类别，也称一个簇，通过将各组紧密相连的样本划分为各个不同类别就得到了最终的聚类结果。

为描述邻域样本分布的紧密程度，DBSCAN 算法使用了两个参数 ϵ 和 $MinPts$ 。其中 ϵ 描述某一样本的邻域距离阈值， $MinPts$ 描述了某一样本的距离为 ϵ 的邻域内含有的样本个数的阈值。基于两个参数的定义，作者将样本数据点分成了三类并提出了三个概念用于描述聚类对象样本之间的关系。其中样本数据点包括核心对象点、边界点和噪音点。核心对象点指在指定两个参数 ϵ 和 $MinPts$ 的情况下，如果某一样本在其 ϵ 的邻域内含有的样本数超过 $MinPts$ 值则可视该样本为一个核心对象点。边界点指某一样本在其距离为 ϵ 的邻域内含有的样本数小于 $MinPts$ 值而该样本落在了某一核心对象点的邻域内，则称该样本点为边界点。噪音点指样本集中除核心对象点和边界点之外的点。描述样本之间的关系的三个概念分别是密度直达、密度可达和密度相连。其中密度直达是指当某一样本对象 x_i 位于某一核心对象样本 x_j 的邻域内，则称 x_i 由 x_j 密度直达。密度可达是指对于两个样本 x_i 和 x_j ，如果存在某一样本序列 $p_1, p_2, p_3, \dots, p_T$ ，满足 $p_1 = x_i$ ， $p_T = x_j$ ，且 p_i 由 p_{i-1} 密度直达，则称 x_j 由 x_i 密度可达。密度相连指对于样本 x_i 和 x_j ，如果存在某一核心对象样本 x_k ，使 x_i 和 x_j 均由 x_k 密度可达，则称 x_i 与 x_j 密度相连。

DBSCAN 算法将聚类定义为：由密度可达关系导出的最大密度相连的样本集合。DBSCAN 算法进行聚类的主要思想是每次任意选定一个核心对象样本并找到该核心对象样本所有密度可达的样本集合，此即构成一个聚类簇，接着继续任选一个没有类别的核心对象样本，寻找其所有密度可达的样本集合构成一个新的聚类簇，如此往复直至所有核心对象样本都归属于一个聚类簇则达到了对样本集进行聚类的目的。

3.2. 颜色数量损失函数

为了实现将 DBSCAN 聚类算法引入图像的风格迁移中，本文定义了一种新的损失函数用来约束模型的训练，称为颜色数量损失函数。损失函数的公式定义如下：

$$L_{color} = \|DB(P_i) - DB(P'_i)\| \quad (1)$$

其中 $DB(x)(x = p_i, p'_i)$ 表示调用 DBSCAN 算法对参数 x 进行基于密度的聚类计算, 返回后得到的簇的数量。 p_i 表示由模型生成的图像经转换后得到的张量表示, p'_i 表示生成的风格迁移图像的张量表示。通过分别对 p_i 和 p'_i 调用 DBSCAN 算法进行聚类计算可分别得到各自图像的颜色数量即簇的数量, 再将两个聚类返回值加入到 L1 损失函数中, 以此可以构建一个新的颜色数量损失函数用于约束模型的训练, 使模型生成的图像的颜色数量能够趋近于原风格图像。

3.3. 基于动漫风格的迁移

在基于上述定义的颜色数量损失函数的基础上, 本文认为该损失函数可以适用于所有带有丰富颜色的图像的风格迁移的训练, 使模型生成的图像在颜色数量上更趋近于原风格图像。

本文以基于 AnimeGAN 模型的动漫图像的风格迁移训练为基础, 使用成对的动漫人脸数据集进行训练。数据集包括一组普通的真人人脸图像和一组与真人人脸图像一一对应的动漫风格人脸图像。通过将定义的颜色数量损失函数加入到模型的生成器的损失函数中可以实现约束生成器训练的目的, 使生成器生成的图像的颜色数量能够趋近于原动漫风格图像的颜色数量。

在引入新的损失函数后, AnimeGAN 模型的生成器 G 的损失函数更改为如下所示:

$$L(G) = \omega_{adv} E_{p_i \sim S_{data}(p)} \left[(G(p_i) - 1)^2 \right] + \omega_{con} L_{con}(G, D) + \omega_{gra} L_{gra}(G, D) + \omega_{col} L_{col}(G, D) + \omega_{colcnt} L_{colcnt}(G, D) \quad (2)$$

模型的判别器 D 的损失函数不变, 如下所示:

$$L(D) = \omega_{adv} \left[E_{a_i \sim S_{data}(a)} \left[(D(a_i) - 1)^2 \right] + E_{p_i \sim S_{data}(p)} \left[(D(G(p_i)))^2 \right] + E_{x_i \sim S_{data}(x)} \left[(D(x_i))^2 \right] + 0.1 E_{y_i \sim S_{data}(y)} \left[(D(y_i))^2 \right] \right] \quad (3)$$

模型总损失函数更改为如下所示:

$$L(G, D) = \omega_{adv} L_{adv}(G, D) + \omega_{con} L_{con}(G, D) + \omega_{gra} L_{gra}(G, D) + \omega_{col} L_{col}(G, D) + \omega_{colcnt} L_{colcnt}(G, D) \quad (4)$$

3.4. SSIM 值衡量迁移效果

本实验采取了结构相似性指数 SSIM (Structural Similarity) 值作为图像风格迁移效果的衡量标准。SSIM 值是目前被广泛采取的衡量两幅图像相似度的指标之一, 相比于 MSE 等图像质量评价指标, SSIM 值对于图像质量的评判更符合人类的主观真实感知。

本实验的训练采用了成对的动漫人脸数据集[9], 该数据集提供了 10,000 对动漫人脸图片, 其中包括 10,000 张真人人脸图片和 10,000 张成对的动漫人脸图片。将由 AnimeGAN 模型生成器生成的图像与所采用数据集中对应的原动漫风格图像作为 SSIM 值计算的输入, 通过计算即可得出模型生成的图像与原风格图像之间的相似度。实验中分别采用两种训练方式对比训练模型, 一种是引入由 DBSCAN 算法定义的颜色数量损失函数训练模型, 另一种则是不引入该颜色数量损失函数即用原 AnimeGAN 模型进行训练, 分别对模型生成的图像和原动漫风格图像进行 SSIM 值的计算, 并对计算结果取均值作比较。

4. 实验和分析

4.1. 模型生成效果图对比

采用加 DBSCAN 算法和不加 DBSCAN 算法两种方式对成对的动漫人脸数据集进行模型训练, 其中

本次模型训练使用的 DBSCAN 算法参数如下：邻域半径 ϵ 设为 2、邻域内样本点数阈值设为 10。模型训练效果图如图 1 所示。

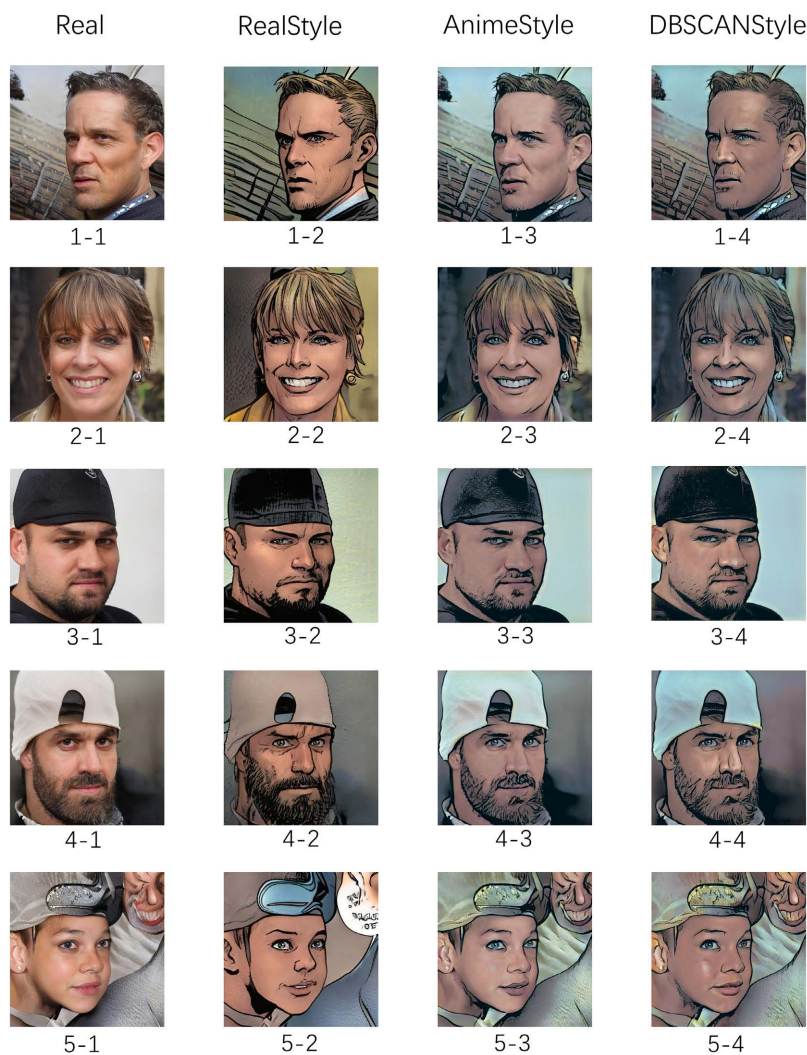


Figure 1. Model generation effect comparison chart

图 1. 模型生成效果对比图

在引入 DBSCAN 算法定义的颜色数量损失函数后模型训练生成的图像的颜色数量即簇的数量更趋近于原动漫风格图像，这一特点也直观体现在生成图像的明暗色块上。对比第三列未引入 DBSCAN 算法训练得到的生成图像，第四列引入 DBSCAN 算法训练得到的生成图像在面部细节上具有与原内容图像更贴近的明暗色块特征，即第四列生成图像的脸部明暗色块更真实丰富，图像更具有立体感，而第三列未引入 DBSCAN 算法得到的生成图像的面部颜色更显单一，没有明显的明暗色块区分，立体感较弱。

4.2. 像素分布分析

取图 1 中的五组模型训练效果图分别计算引入 DBSCAN 算法得到的生成图像和未引入 DBSCAN 算法得到的生成图像的像素分布,绘制成如图 2 所示对比图。该图横坐标是像素种类，纵坐标是每种像素类别的像素数量。

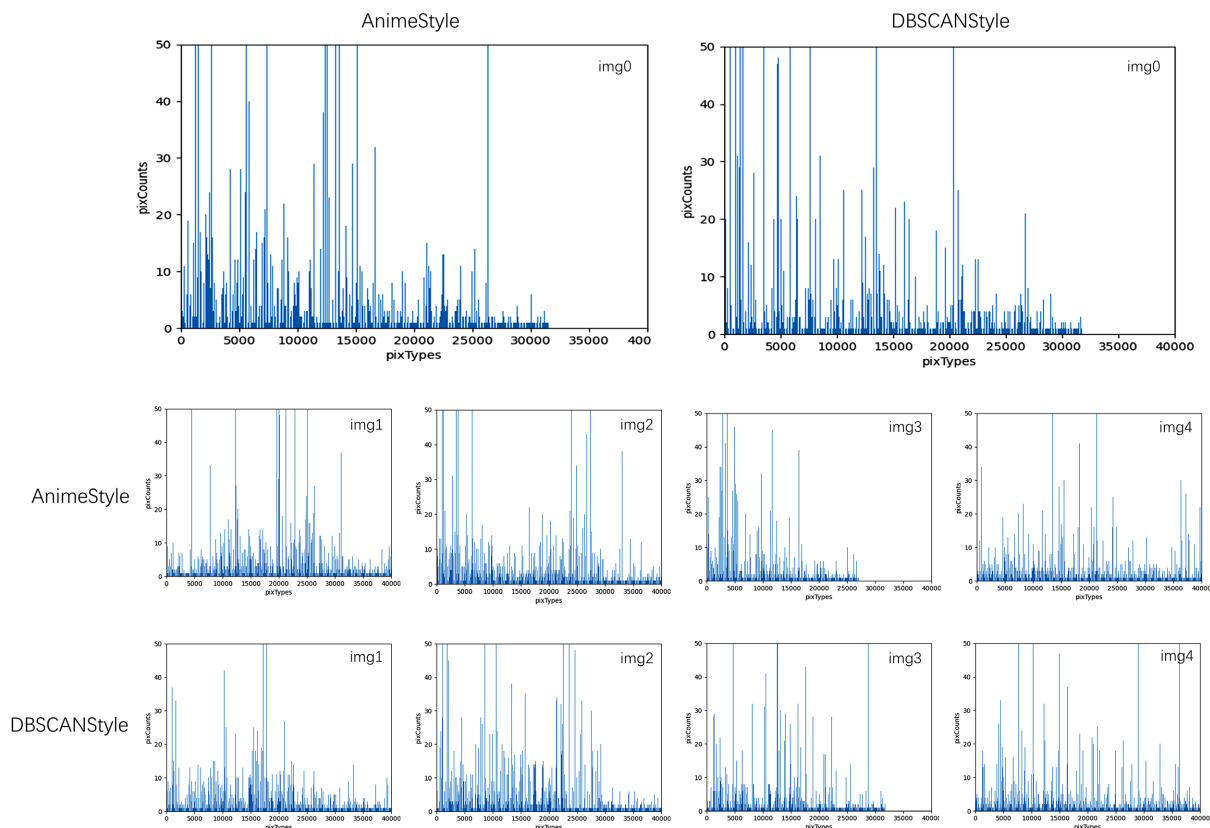


Figure 2. Pixel number distribution comparison chart

图 2. 像素数量分布对比图

在真实图像中，由于光线的原因，位于同一区域的像素值会有细微的差异，在颜色过渡区域的像素较多，而在动漫风格图像中同一色块内这样的差异很小，可以忽略不计，颜色过渡区域中的像素较少。这使得在内容大致相同、尺寸一致的情况下，真实图像中像素分布较为均匀连续，而动漫风格图像中像素分布更加离散，以不同颜色为区域成组出现。

通过对比图 2 中五组模型训练效果图的像素分布可以直观看出在引入 DBSCAN 算法定义的损失函数后模型更趋向于生成颜色区域明显的图像，这也进一步验证了 DBSCAN 算法在指导模型生成风格图像上的积极作用。

4.3. 聚类后簇的数量分布对比

使用 DBSCAN 算法对图像进行聚类的输出结果是一个二维的标签列表，其中的每一个值代表的是原图像中位于该位置的像素所属的类别。基于这个标签值对聚类结果进行可视化得到如图 3 所示效果图。

由于 DBSCAN 聚类算法是根据像素的 R, G, B 值在三维空间进行聚类，通过对比这几组图像和他们对应的聚类可视化结果可以得出这样的结论：在引入了 DBSCAN 聚类算法定义的损失函数后模型生成的图像更能准确反映出原始图像中的颜色和 content 分布。

4.4. SSIM 值对比

将图 1 中的两组生成图像分别与原动漫风格图像进行 SSIM 计算，将计算结果绘制成表格，如表 1 所示。

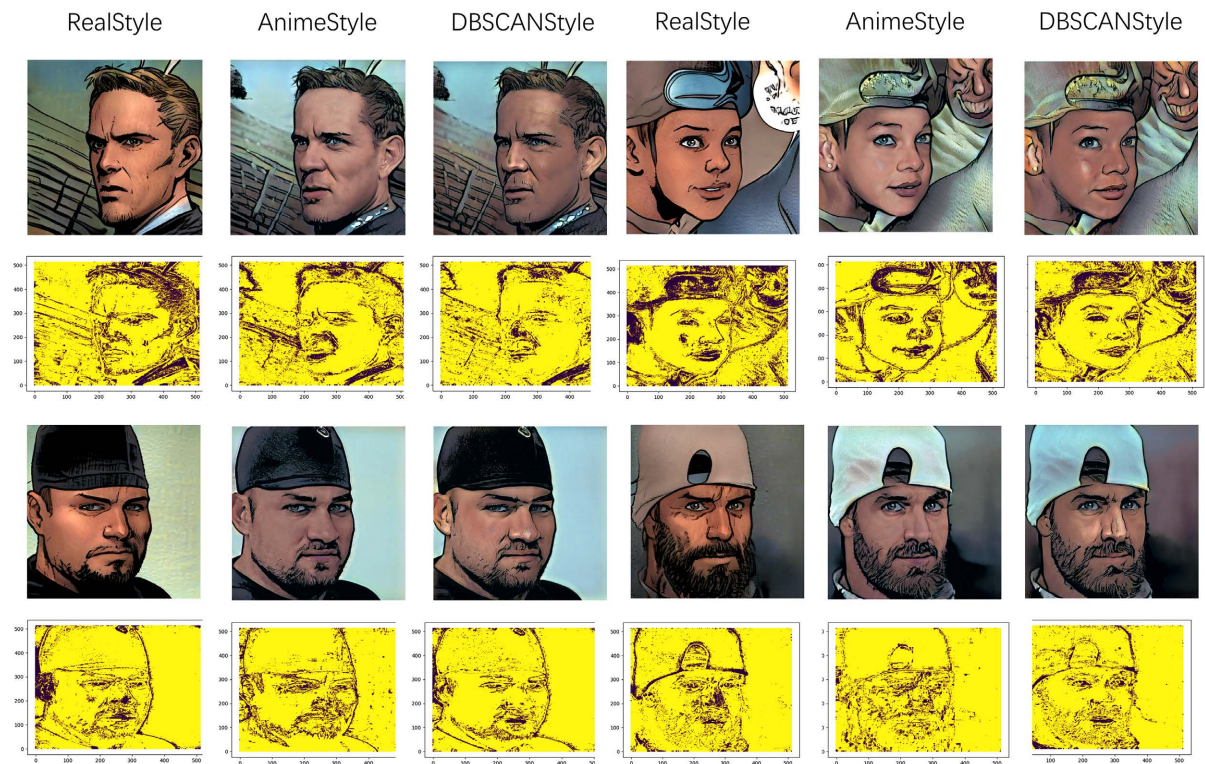


Figure 3. Comparison chart of cluster number distribution
图 3. 簇数量分布对比图

Table 1. SSIM value comparison of model generated renderings
表 1. 模型生成效果图的 SSIM 值对比

样本编号	不加 DBSCAN (SSIM)	加 DBSCAN (SSIM)
1	0.4169	0.4220
2	0.3984	0.3999
3	0.4741	0.4751
4	0.5377	0.5511
5	0.4103	0.4162
平均值	0.4475	0.45286

从表 1 可以看出在引入了 DBSCAN 算法训练后模型生成的图像的 SSIM 值普遍都要大于未引入 DBSCAN 算法训练得到的生成图像。这一客观评价标准进一步表明引入 DBSCAN 算法后模型训练生成的图像更具有原风格图像的特征，两者结构相似度也更高。

5. 结论

本文针对动漫风格图像自动生成问题进行了研究，基于 GAN 模型进行神经网络风格迁移的算法，提出了一种颜色数量损失函数用于约束模型的训练。原始图像和生成动漫图像的颜色数量分别使用 DBSCAN 聚类算法自动计算，从而使生成的图像能够在颜色数量上趋近于目标风格图像。实验结果表明在引入由 DBSCAN 算法定义的颜色数量损失函数后图像风格迁移效果得到了良好的提升，模型生成的图像在颜色上、明暗色块上更贴近于真实图像，图像在视觉上呈现出更好的立体感，在观感上具有更鲜活

的特征。

基金项目

北京信息科技大学 2021 年大学生创新创业训练计划项目资助(5102110805); 北京市教委科技计划一般项目(KM201811232024); 北京信息科技大学科研基金(2021XJJ30, 2021XJJ34); 北京信息科技大学高教研究重点项目(2019GJZD01)。

参考文献

- [1] Gatys, L.A., Ecker, A.S. and Bethge, M. (2015) A Neural Algorithm of Artistic Style. arXiv:1508. 06576.
- [2] Li, C. and Wand, M. (2016) Combining Markov Random Fields and Convolutional Neural Networks for Image Synthesis. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 27-30 June 2016, 2479-2486 <https://doi.org/10.1109/CVPR.2016.272>
- [3] Liao, J., Yao, Y., Yuan, L., *et al.* (2017) Visual Attribute Transfer through Deep Image Analogy. arXiv:1705. 01088.
- [4] Mirza, M. and Osindero, S. (2014) Conditional Generative Adversarial Nets. ArXiv:1411.1784.
- [5] Zhu, J.Y. and Park, T., *et al.* (2017) Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. <https://arxiv.org/abs/1703.10593>
- [6] Chen, Y., Lai, Y.K. and Liu, Y.J. (2018) CartoonGAN: Generative Adversarial Networks for Photo Cartoonization. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 18-23 June 2018, 9465-9474. <https://doi.org/10.1109/CVPR.2018.00986>
- [7] Chen, J., Liu, G. and Chen, X. (2019) AnimeGAN: A Novel Lightweight GAN for Photo Animation. In: Li, K.S., Li, W., Wang, H. and Liu, Y., Eds., *Artificial Intelligence Algorithms and Applications*, Springer, Singapore, 242-256. https://doi.org/10.1007/978-981-15-5577-0_18
- [8] <https://github.com/TachibanaYoshino/AnimeGANv2>
- [9] 彼佳. 动漫人脸数据集[EB/OL]. <https://www.kaggle.com/defileroff/comic-faces-paired-synthetic-v2>, 2021-05-14.