

硕 士 研 究 生 读 书 报 告



题目 Coarse-to-Fine Facial Structure Editing of Portrait Images读书报告

作者姓名 练云轩

作者学号 22151116

指导教师 李启雷

学科专业 电子信息（软件工程）

所在学院 软件学院

提交日期 二○二一 年 12月

Reading Report of Coarse-to-Fine: Facial Structure Editing of Portrait Images via Latent Space Classifications

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor: Qilei Li

By

Yunxuan Lian

Zhejiang University, P.R. China

2021

# 摘要

鉴于人脸的多样性、正确标注的缺乏、联合调整颜色和形状的必要性、以及去除视觉伪影的要求，肖像图像的面部结构编辑是一件极具挑战性的课题。在本论文中，作者以进行下巴编辑为例来研究面部结构编辑方法。文中提出了一种新的方法来自动消除肖像图像中的双下巴。其核心思想是在肖像图像的潜在空间中训练一个精细的分类边界，通过将肖像图像的潜在代码输入到分类边界中，就能实现保留原始肖像特征的同时又成功消除双下巴的效果。

**关键词**：面部编辑，潜在代码编辑，StyleGAN

# Abstract

Facial structure editing of portrait images is challenging given the facial variety, the lack of ground-truth, the necessity of jointly adjusting color and shape, and the requirement of no visual artifacts. In this paper, the author investigated how to perform chin editing as a case study of editing facial structures. They present a novel method that can automatically remove the double chin effect in portrait images. The core idea is to train a fine classification boundary in the latent space of the portrait images. The double chin effect will be removed together with the original portrait features preserved by inputting the latent code to the classification boundary above.

**Keywords：**Face editing, latent code manipulation, StyleGAN

# 1引言

随着社会和科技的发展、便携摄影设备的普及，拍摄人像照片变得越来越容易，人像照片（包括自拍和他拍）变得无处不在。日常生活中，人们往往倾向于使用照片来表达自己的状态或进行社交活动。因此，这些照片对于在人类社会生活中发挥日益重要作用的社交媒体应用（如Facebook、Twitter、Instagram等）来说非常重要。同时，人们对对美貌的追求也进一步增加了“美颜”的需求。于是，这促使研究人员研究人像图像编辑的计算方法，以实现肖像图像的优化。

## 1.1研究现状

近些年来，研究人员们全力推动着肖像图像的优化工作。基础层面上，已经使得阴影、噪音和扭曲等不完美之处很容易地得到修复或纠正[1]。细节层面上，针对肖像图像的编辑工作主要涉及调整图像颜色[2]或全局扭曲面部区域[3][4]。如采用风格转移方法（Style transfer methods）[5]将给定肖像照片的风格转移到新照片上等。

而如何有效地编辑面部结构的某一部分，甚至删除某一个具体的结构特征，在很大程度上仍处于尚未开拓的状态。相比较整体图像的调整而言，局部编辑的难度更大，因为面部结构的颜色和形状都需要部分更新，且编辑后仍需保持与周围区域兼容，否则将出现视觉上明显的不和谐的失真情况。

## 1.2研究动机

在肖像图像的面部编辑中，论文选取了编辑下巴的结构作为研究案例。众所周知，在拍摄人像照片时，下巴区域的形状和颜色很容易因头部姿势及光照条件而变化，这可能对面部结构产生影响，出现常见的“双下巴效应”。而移除双下巴并不是字面意义上的抹平那么简单，首先模型需要检测到颈部区域，然后对双下巴进行移除。同时，还需要保持不改变照片中的其他特征，以实现自然的效果。

虽然人们可以利用诸如PhotoShop之类的工具进行手动编辑，但这明显是个耗时耗力的行为，且对技术有一定的要求。

目前现有的一些编辑面部纹理的方法，如风格转移方法[5]、MaskGAN[6]、SEAN[7]、SC-FEGAN[8]等，在消除双下巴方面的表现效果均不尽人意。原因在于这些模型更多地只考虑了面部颜色和形状的编辑，而消除双下巴则需要考虑更多的几何因素。

论文作者受生成对抗网络（GANs）最近在高质量肖像图像合成方面的成功启发[9][10]，特别是受通过StyleGAN[11]来完成受控制生成与分解面部特征（如身份，姿势，发型，肤色等）的启发，将下巴编辑归类为肖像图像潜在空间（latent space）中的肖像图像合成问题。

总得来说，该研究做出的贡献如下：

* 首次提出了用于肖像图像的自动下巴编辑方法。它可以生成一个没有双下巴的新面部结构，同时保持其他区域不变。
* 介绍了一种新的基于从粗到细分离边界训练的结构级人脸编辑方法，该方法允许在肖像图像的潜在空间中进行直接编辑，并进行合理的语义操作和人脸身份保持
* 创建了第一个大规模下巴编辑数据集，以促进未来的研究。该数据集包含13990对具有或不具有双下巴的真实肖像图像。

# 2方法

## 2.1方法概述

论文中提出的方法的主要思想如下：为了将双下巴分离，首先训练一个分类器，根据双下巴的普遍度对不同图像的潜在代码进行评分，并由此生成一个粗糙的分离边界，用来合成没有双下巴的中间状态图像。然后，为了保持其他面部特征不变，引入一个扩散（Diffusion）过程，利用颈部遮罩将中间状态图像和原始图像进行混合，从而得到一个成对的肖像图像（有双下巴和无双下巴）及其各自对应的潜在代码。最后，根据上一步得到的潜在代码训练一个精细的分离边界，在保留原始肖像特征的同时，将输入的肖像图像的潜在代码编辑到没有双下巴的人脸的域中。整个算法的流程图如图2.1所示。

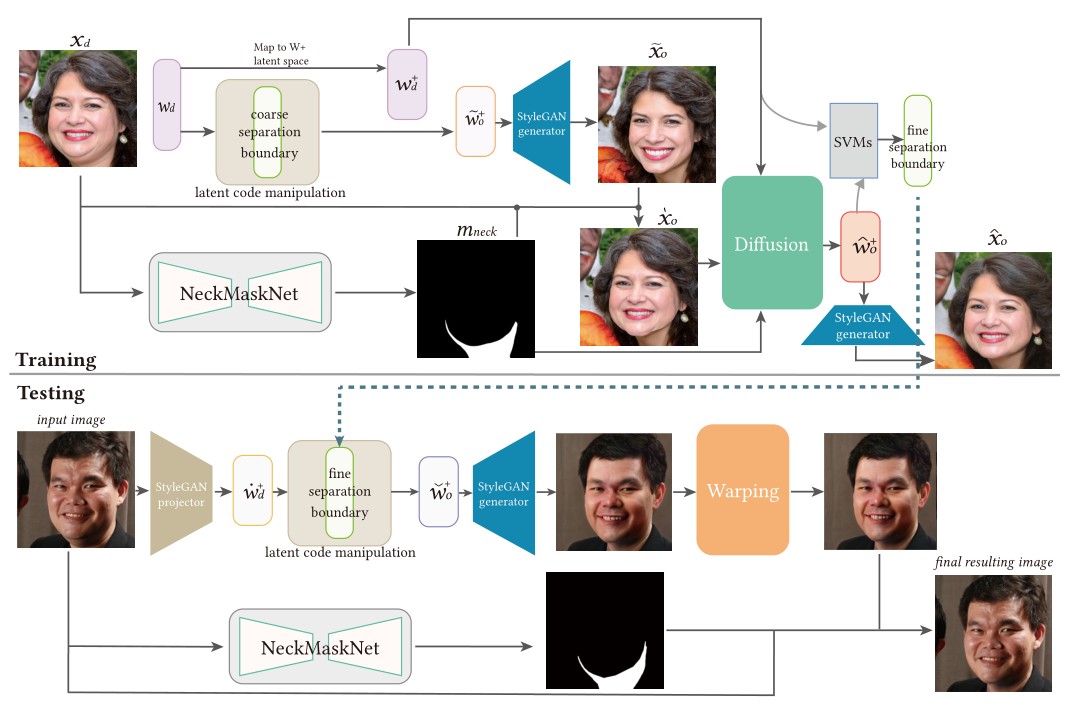


图2.1算法流程图

## 2.2粗略双下巴分离边界生成

移除双下巴的第一步便是在潜在空间*W*上搜索生成一个粗略的双下巴分离边界。首先，需要将潜在代码空间与双下巴空间联系起来。为此，作者编写了一个名为DoubleChinClassifier的分类器，用来对由潜在代码生成的肖像图像进行评分。此处使用了StyleGAN模型，利用其中的合成网络g将潜在代码映射到图像空间X中。

为了训练分类器，作者首先构造了数据集。所有的数据集可以表示为：

其中，表示分数集，有双下巴的图像将被评分为，否则；是含有双下巴图像的数据集，是不含有双下巴图像的数据集。

为了避免其他冗余信息干扰，作者为每个图像生成一个矩形遮罩，并仅仅将颈部区域输入到分类器中进行训练。

回到潜在空间W中，随机抽取50000个潜在代码，并生成50000幅肖像图像。随后利用DoubleChinClassifier来标记所有这些肖像图像，并构建出分离边界训练数据集如下：

其中，表示潜在代码，表示双下巴分数。

得到训练数据集后，作者利用支持向量机（SVM）对粗糙分离边界进行训练。分离边界的编辑公式如下：

即通过将原始图像的潜在代码输入到粗糙分离边界后，我们可以得到编辑后的潜在代码和一张没有双下巴的肖像图像。此时的的皮肤特征与原图像差别较大，故作者采取了StyleMixing方法来对和进行操作，获得了混合潜在代码，最终生成了一个新的肖像图像，它保留了的几何特征和的皮肤特征（见图2.2）。



图2.2 从左到右分别是：原始图像，粗分离边界生成的图像，StyleMixing调整后生成的图像

## 2.3颈部遮罩生成及扩散（Diffusion）过程

从图2.2可以看出，尽管调整后保留了皮肤特征，但面部特征并没有得到很好的保留。为了更好地保存面部特征，作者进一步使用了NeckMaskNet来生成颈部遮罩，以便允许从原始图像直接复制其他面部区域。

为了生成颈部遮罩，该算法将Faceparsing模型和颈部关键点（由人脸对齐方法[12]检测而得）结合起来，并在遮罩边界进行抗锯齿操作，以获得较为合理的颈部覆盖效果。NeckMaskNet的约束公式如下：

其中，表示由下巴的关键点组成的折线，和表示由Faceparsing模型预测的脸部遮罩和颈部遮罩，U是并集操作，表示仅将遮罩的一部分保留在下方的切割操作。

在得到颈部遮罩后，便可以将的颈部区域复制并粘贴到上，得到新图像。然而新图像往往存在视觉伪影，展现出肉眼可见的不协调部分。为此，需要优化并得到一个没有这类伪影的图像。

作者受域内GAN反演方法的启发，采用了一个Diffusion过程在的指导下对潜在代码（是将截断技巧和堆栈操作应用于后得到的潜在代码）执行优化。

为了保留面部特征，算法使用重建loss（reconstruction loss）来惩罚除了颈部区域之外的和的像素差异（表示优化后的潜在代码），此处的损失记作。同时，还采用了一个预训练的VGG16感知损失模型来惩罚和之间的高级（几何）特征差异，此处的损失记作。如此可得总体扩散目标函数如下：

其中，和是用来衡量不同损失的超参数。

于是，便可以通过最小化来优化潜在代码，以便获得更加自然的图像生成效果，如图2.3所示。

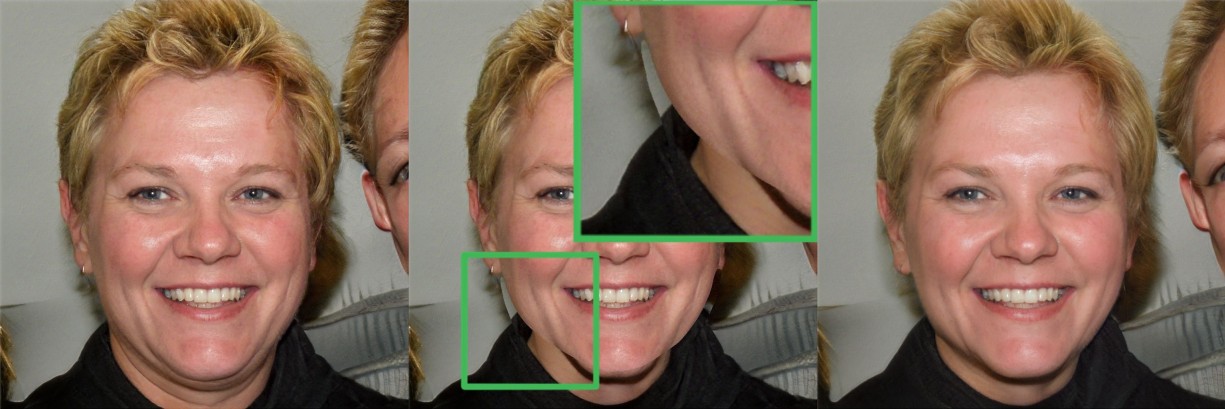


图2.3从左到右分别是：原始图像，粘贴面部区域得到的新图像和经过Diffusion过程优化的

## 2.4精细双下巴分离边界

通过上文训练粗分离边界的过程（2.2节）不难看出，存在着一些缺点。如分类器进行评分的时候可能会误判（俯拍视角容易忽略双下巴），不能很好地保持面部特征等。从上一步的扩散过程中，模型得到了潜在代码与。其中，与由“理想”分离边界（即既可以去除双下巴，同时又能保留面部特征的边界）编辑的潜在代码十分接近。由此，考虑到潜在空间的线性度，作者假设可以通过线性变换而非扩散操作转换为。因此，这一步便是利用中训练一个精细分离边界。

首先构建一个新的分离边界训练数据集：

随后在新数据集上训练精细分离边界，并将其应用于实际图像处理。与的对比如图2.4所示。



图2.4从左到右分别是：原始图像，经过生成的图像和经过生成的图像

## 2.5Warping

尽管精细分离边界已经得到了较好的效果，但沿面部边缘的地方仍然存在细微的错位现象。为此，算法引入Warping来减少偏差。

首先，算法使用FaceParsing来从（即经过生成的图像）和中提取面部皮肤遮罩和颈部皮肤遮罩，然后选取点如下：

其中，表示中面部边缘的点，表示中面部边缘的点，表示一个固定的点集。注意，此处的和有相同的y坐标。表示中颈部边缘的点。为了保证颈部形状不变，使用相同的颈部点集来构建和。

然后，算法使用点集来构建一个Delaunay三角剖分*D*，然后利用*D*来扭曲。也就是说，将拉到，使的面部边缘与的面部边缘对齐。

# 3算法评估

## 3.1优势与创新点

正如前文中所提到的那样，该论文的研究成果某种意义上可以称之为“开山之作”——因为在这之前几乎没有关于如何对肖像图像中面部结构的具体某一部分进行编辑的研究成果。除了应用层面的有效性外，该论文还为相关领域之后的研究提供了思路和方向。

另外，在论文中，作者还将该算法与现有的几个面部编辑方法（加之以手动的适应调整）进行比较，包括MaskGAN、SC-FEGAN和inpainting[13]算法。其中，对于MaskGAN，作者使用FaceParsing提取颈部遮罩和面部遮罩，然后手动编辑输出的遮罩以减少多余的下巴区域。对于SC-FEGAN，作者使用官方的SC-FEGAN demo，在双下巴上添加一个遮罩，然后手动绘制下巴草图以生成新下巴。对于inpainting，作者利用本算法中的颈部遮罩生成方法和ChinEditing数据集，通过将具有双下巴的图像和相应的颈部遮罩设置为训练数据集来训练修复模型。比较结果如图3.1所示。



图3.1 三种现有算法和论文所提出的算法的处理结果比较图

从比较图可以看出，该论文提出的自动消除双下巴的算法取得了较为显著的成果。

## 3.2局限性

虽然算法取得了不错的效果，但它仍然存在一些局限性。

首先，对双下巴的消除会导致颈部区域缩小，进而导致靠近颈部的衣服边缘错位或照片背景空缺等问题。而对于背景空缺，可以利用inpainting来进行填补。

其次，颈部遮罩会覆盖一些靠近脸部的饰品，造成物品边缘伪影，甚至会覆盖部分面部区域，使下巴缩短。对于这个问题，或许可以尝试使用更为精确的颈部遮罩加以解决。

最后，对于处于极端姿势的脸，该算法效果不佳，无法去除双下巴。其原因主要是原始的训练数据集中，大部分的肖像图像的姿势都接近中性。若能够在训练数据集中添加更多不同姿势的人脸以增强普适性，或许就能获得更好的效果。

## 3.3未来工作

本算法的目的是移除双下巴，是一个正向过程。同理，或许可以使用该算法进行反向下巴编辑，即添加下巴结构。在这种情况下，可以通过将超参数α设置为正值（本算法中为负值），沿着关于分离边界的相反方向更新输入肖像的潜在代码。

另外，该算法的思想可以引申应用到其他面部结构的编辑行为中，如头发操纵、面部妆容设计等，而不仅仅局限于下巴编辑。

# 4总结

在该论文中，作者提出了一种新的人脸结构编辑方法，并以人像图像为例首次介绍了自动下巴编辑方法。其关键思想是训练一个精细的分离边界，并通过编辑原始肖像的潜在代码生成一个包含新下巴的图像，并保证其他面部特征与原始肖像匹配。

# 参考文献

1. Arakawa K, Nomoto K. A system for beautifying face images using interactive evolutionary computing[C]//2005 International Symposium on Intelligent Signal Processing and Communication Systems. IEEE, 2005: 9-12.
2. Scherbaum K, Ritschel T, Hullin M, et al. Computer‐suggested facial makeup[C]//Computer Graphics Forum. Oxford, UK: Blackwell Publishing Ltd, 2011, 30(2): 485-492.
3. Shih Y C, Lai W S, Liang C K. Distortion-free wide-angle portraits on camera phones[J]. ACM Transactions on Graphics (TOG), 2019, 38(4): 1-12.
4. Xiao Q, Tang X, Wu Y, et al. Deep Shapely Portraits[C]//Proceedings of the 28th ACM International Conference on Multimedia. 2020: 1800-1808.
5. Chang H, Lu J, Yu F, et al. Pairedcyclegan: Asymmetric style transfer for applying and removing makeup[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 40-48.
6. Lee C H, Liu Z, Wu L, et al. Maskgan: Towards diverse and interactive facial image manipulation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 5549-5558.
7. Zhu P, Abdal R, Qin Y, et al. Sean: Image synthesis with semantic region-adaptive normalization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 5104-5113.
8. Jo Y, Park J. Sc-fegan: Face editing generative adversarial network with user's sketch and color[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 1745-1753.
9. Brock A, Donahue J, Simonyan K. Large scale GAN training for high fidelity natural image synthesis[J]. arXiv preprint arXiv:1809.11096, 2018.
10. Karras T, Aila T, Laine S, et al. Progressive growing of gans for improved quality, stability, and variation[J]. arXiv preprint arXiv:1710.10196, 2017.
11. 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.
12. Bulat A, Tzimiropoulos G. How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks)[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 1021-1030.
13. Yu J, Lin Z, Yang J, et al. Generative image inpainting with contextual attention[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 5505-5514.