

三 维 动 画 和 交 互 设 计 课 程 读 书 报 告



题目 基于深度学习的3D人脸与漫画建模草图系统

作者姓名 孙松涛

作者学号 21951094

指导教师 李启雷

学科专业 软件工程

所在学院 软件学院

提交日期 二○一九年十二月

A Deep Learning Based Sketching System for 3D Face and Caricature Modeling

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

Songtao Sun

Zhejiang University, P.R. China

2019

摘要

脸部建模在计算机图像和视觉领域一直很火。卡通人物建模、人脸艺术设计和人脸实时重构等场景，都对该领域有很大的需求。而人们需要通过简单的系统进行建模，交互式人脸建模也是近几年研究比较多的一个方向。但是想要通过简单的操作就获得表现力强的人脸模型，还是一件很有挑战的事情。

本文采用CNN神经网络构建了一个快速的、交互的、基于深度学习的人脸建模框架。通过简单勾勒人脸图画，本文的模型可以迅速生成对应的三维人脸模型，并且可以同时拟合面部轮廓和细节表情。同时我们也提供了多种方式进行快速的模型修改。本文的模型还适用于手势识别，用户可以通过手势对三维人脸模型进行快速的模型修改，便于用户操作。

**关键词**：人脸模型，面部数据库，深度学习，人脸刻画，手势，素描模型

Abstract

Face modeling has been paid much attention in the field of visual computing.There exist many scenaios, including cartoon characters, avatars for social media, 3D face caricatures as well as face-related art and design. But creating expressive face models with a minimal amount of labor is still a major challenge in interactive face modeling.

In this paper, a fast, interactive, deep learning-based face modeling framework is constructed using CNN neural network. By simply sketching the face picture, the model of this paper can quickly generate the corresponding 3D face model, and can fit the facial contour and the detail expression at the same time. At the same time, we also provide a variety of ways to make rapid model changes. The model in this paper is also applicable to gesture recognition. Users can quickly modify the 3D face model by gestures, which is convenient for users to operate. Experiments have shown that our results are highly accurate and fast.

**Keywords：**Face Modeling, Face Database, Deep Learning, Face Caricatures. Gestures, Sketch-Based Modeling

# 目录

[摘要 3](#_Toc28182427)

[Abstract 4](#_Toc28182428)

[目录 5](#_Toc28182429)

[一、 该文研究背景与问题描述 6](#_Toc28182430)

[二、 贡献及不足 6](#_Toc28182431)

[2.1 贡献 6](#_Toc28182432)

[2.2 不足 6](#_Toc28182433)

[三、 实现过程 7](#_Toc28182434)

[3.1 数据准备 7](#_Toc28182435)

[3.2 像素-输入层 7](#_Toc28182436)

[3.3 造型-输入层 8](#_Toc28182437)

[3.4 双线性输出层 8](#_Toc28182438)

[3.5 损失层 8](#_Toc28182439)

[四、 思考与扩展 9](#_Toc28182440)

[4.1 思考 9](#_Toc28182441)

[4.2 应用场景 9](#_Toc28182442)

[参考文献 9](#_Toc28182443)

# 该文研究背景与问题描述

脸部建模在计算机图像和视觉领域一直很火。卡通人物建模、人脸艺术设计和人脸实时重构等场景，都对该领域有很大的需求。而人们需要通过简单的系统进行建模，交互式人脸建模也是近几年研究比较多的一个方向。但是想要通过简单的操作就获得表现力强的人脸模型，还是一件很有挑战的事情。

# 贡献及不足

## 2.1 贡献

本文提出了一种新颖的框架，致力于创建一个通过手工草图创建三维人脸模型的框架，主要有以下贡献：

1. 本文提出了一种用于三维人脸建模和漫画造型的新颖框架，这个框架具有高效建模的功能，三维人脸模型能够依据二维人脸素描的特征关系自动生成，具体效果如图1所示。同时该系统支持手势识别，用户可以使用手势对初始面部模型进行完善与修改。
2. 本文采用了一种基于深度回归网络的新颖卷积神经网络，正是该网络能够从二维人脸素描自动推断出其三维模型。该网络融合了卷积神经网络和二维素描的形状特征，具有两个独立的全连接层分枝。正是该结构保证了三维人脸模型建模的效率与准确性。
3. 生成了一个扩展的三维及二维的人脸数据库，该人脸数据库具有更多的身份，表情及不同程度的夸张。该数据库可直接用于训练与测试，方便后续人脸模型研究者的工作。

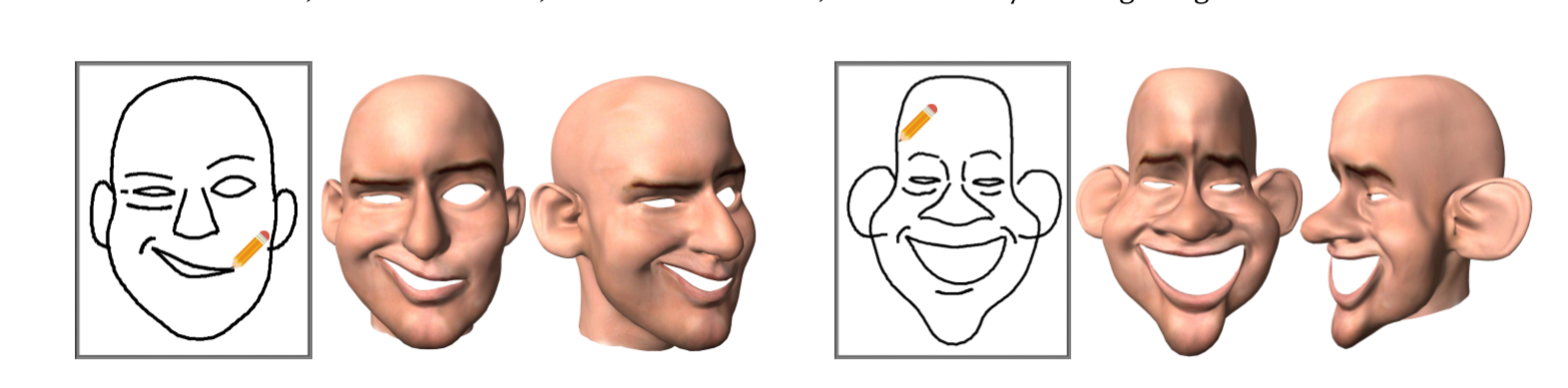


图1 该系统通过手工草图，创建了响应的三维人脸模型

## 2.2 不足

1. 作者在创建数据集的时候，对面部的不同部分做了一致性比较强的形状改变。所以当不同部位间的形变特别不一致的时候，该系统便会创建出特别不自然的结果，如图2的两个例子所示。解决数据集的该问题是作者下一步的工作。
2. 另外，形变模型无法创建类似于皱纹等几何细节，如何解决该问题也是作者进一步研究的方向。

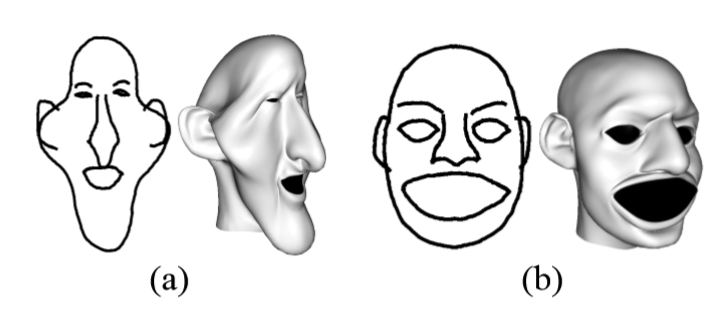


图2 该系统创建的不自然结果

# 实现过程

## 3.1 数据准备

1、 面部表情扩充：原人脸数据集（Cao et al.2014）针对每一个人脸提供了20个面部表情，但许多面部表情间仅有镜像差异，这使得2D素描的特征提取变得难以进行与解释。为了提高模型的效果，本文针对原数据集中的11个面部表情进行了变形转换操作，并邀请艺术家重新设计了14个新的面部表情，使得面部表情数据更为丰富且多样。其中人脸的表情包括了人类基本的表情，例如：开心、难过和恐惧等。

2、 造型夸张：提取原有的人脸造型数据的梯度尺度因子，通过改变梯度尺度因子的改变，重构人脸的3D模型。但由于眼睛和嘴部的梯度不稳定，本文仅针对脸颊，下巴，后脑勺，鼻子和耳朵等部位进行夸张。同时设置了4个不同的夸张程度，夸张程度为1时造型和原有造型相同，夸张程度随等级上升而增加。丰富了人脸造型数据集。

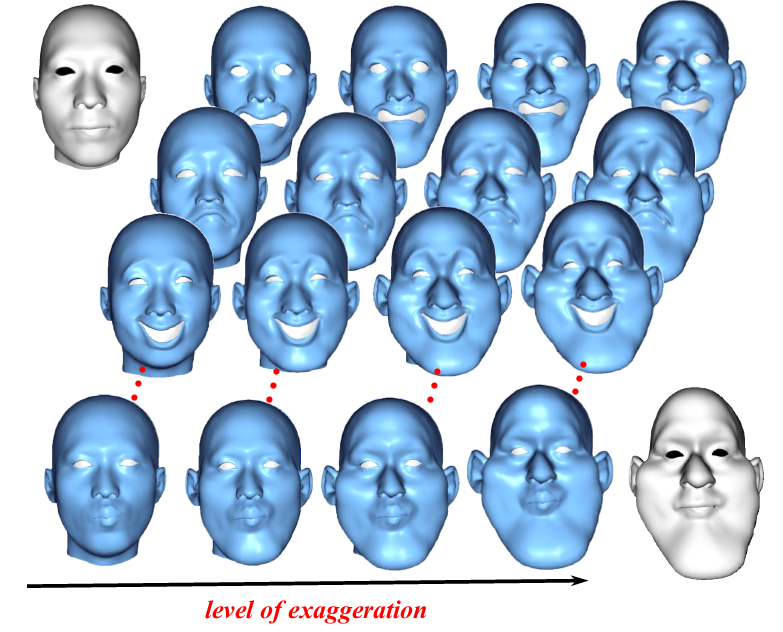


图3 面部表情夸张处理效果

## 3.2 像素-输入层

深度回归神经网络的输入是一个尺寸为的二值化图像，图像包含了轮廓，部位特征和线条特征等信息。该层的图像将会作为AlexNet模块的输入进行训练。最后获得一个1024维的向量作为下一步的输入。

## 3.3 造型-输入层

即使是最顶端的卷积层也只具备有限的视野而无法观察到整个图像，因此本文深层网络中的AlexNet模块不能够完全掌握输入草图中的全局上下文信息。为了充分利用草图中的信息，最大程度上得了解草图的特征，本层直接对轮廓和特征线上进行抽样，并抽取固定点数。通过另一个双线性表示这些采样点的2D位置模型，其中人脸身份和表情模式的系数仍然设置为50和16.这66个系数形成我们的造型-输入向量。在每一个分枝中，造型-输入层后都设置了一个具有512个神经元的全连接层，两个分支的输出最终作为输入，输入到双线性输出层中。

## 3.4 双线性输出层

本文深度回归网络的输出是一组双线性人脸造型的系数，是最终能够重建三维人脸模型的一组3D顶点。这部分的双线性输入是两个部分：人脸身份和人脸表情，这两个部分将作为独立的模式，并赋予不同的权重和。为了避免这两个部分之间产生相互影响，本文使用了两个不同的独立分枝，并使用不同数量的全连接层，来生成权重和。本文仍然借鉴了AlexNet的结构，生成权重μ的分枝包括个全连接层，而生成权重γ的分枝仅包含个全连接层。每个全连接层设置了个神经元，其数量是通过实际操作经验得到的，实验数据可表名该神经元数量效果最佳。

## 3.5 损失层

本文的深度回归网络是基于最小化顶点损失的训练模型。为了实现这一目标，本文将顶点损失定义为真实顶点与预测顶点之间的L2误差，所有的顶点都可以通过和矩阵得到。该L2误差定义为如下所示：

其中是对应于核心张量C的2D切片第个顶点，代表真实模型中第个顶点的位置，表示第个顶点的权重，是顶点的总数。可表示顶点作为某些顶点的相对重要性，例如那些在3D轮廓和特征线上的顶点具有更高的重要性。

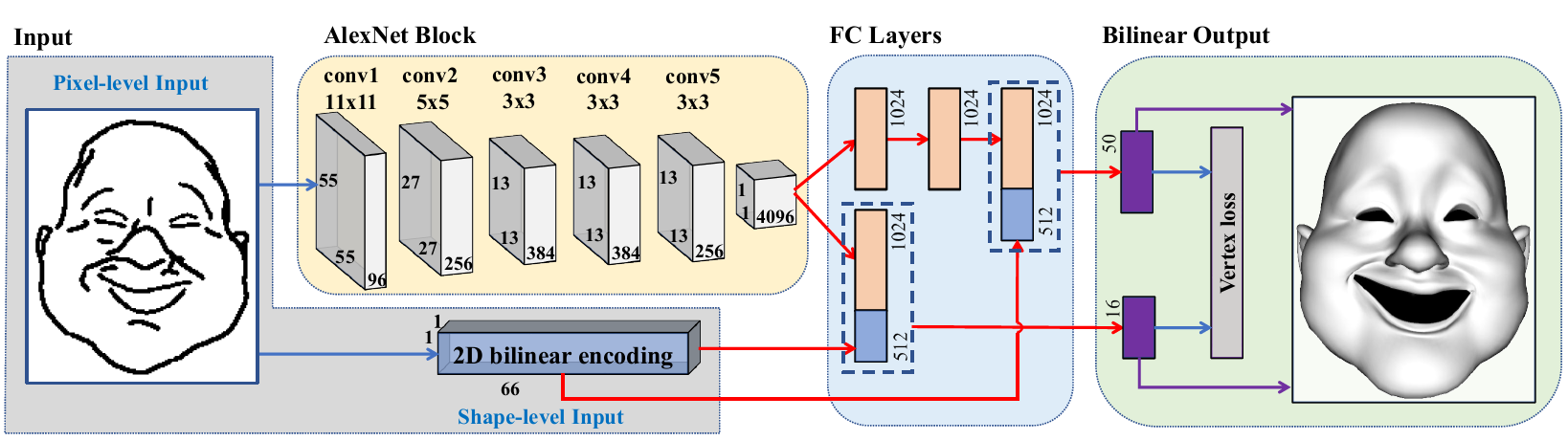


图4 本文所使用的网络框架

# 思考与扩展

## 4.1 思考

本文使用的backbone还是最初的alexnet，使用alexnet作为backbone当然可以获取到图像的特征信息。但是近几年出现了更好的backbone模型，如resnet。该模型在准确率和计算速度上都具有很大的优势，所以可以将特征提取的backbone替换成resnet的特征提取层。

## 4.2 应用场景

该论文提出的系统对于创作者具有很大的便利性。创作者有时候并不想要大量的时间进行3D的建模，或者对3D建模并不熟悉。所以通过该交互式系统，创作者只需要画出响应二维草图，在草图中体现人物的脸部特征，便可以轻松获得该角色的三维面部模型。对于不满意的部分还可以进行二次修改，使生成的模型具有灵活性。

# 参考文献

[1] Han X , Gao C , Yu Y . DeepSketch2Face: A Deep Learning Based Sketching System for 3D Face and Caricature Modeling[J]. Acm Transactions on Graphics, 2017, 36(4):126:1-12.