****



**本 科 毕 业 论 文**

院 系 计算机科学与技术系

专 业 计算机科学与技术

题 目 基于tensorflow的深度学习肖像抠图算法

年 级 本科四年级 学 号 141220080

学生姓名 浦溪

指导教师 顾荣 职 称 讲师

提交日期 2018.5.15

**南京大学本科生毕业论文（设计、作品）中文摘要**

题目：基于tensorflow的深度学习肖像抠图算法

院系：计算机科学与技术系

专业：计算机科学与技术

本科生姓名：浦溪

指导教师（姓名、职称）：顾荣 讲师

摘要：深度学习由于其强大的表达能力和拟合能力，近几年在机器学习，模式识别等领域表现出优秀的性能。抠图属于语义分割的范畴，对于人物肖像的抠图算法，即将人物肖像和背景分割，涉及的主要难点是背景的复杂性以及肖像细节部分的分割。通过深度学习的方法进行人物肖像抠图能够取得远超过传统算法的准确率。我采用标注好的公开肖像数据集，首先用全卷积神经网络模型进行训练，然后进一步提出加入Resnet结构，对边缘区域进行模糊处理等改进模型，在IOU的性能度量上获得提高。

我采用Tensorflow作为训练框架，利用其并行化的特点以及pipeline数据流技术，大幅提高训练效率。由于深度学习所需要的大量数据，数据经常存储在不同的存储系统中，而且深度学习上层的应用框架多样化，我采用Alluxio为数据访问统一接口，使得不同的应用框架可以透明地读取多种存储系统中的数据，并且利用Alluxio的内存加速数据的访问速度。

关键词：深度学习；语义分割；抠图；全卷积神经网络；边缘模糊化; Tensorflow; Alluxio

**南京大学本科生毕业论文（设计、作品）英文摘要**

THESIS：Deep Portrait Segmentation methods based on tensorflow

DEPARTMENT：Computer Science and Technology

SPECIALIZATION: Computer Science

UNDERGRADUATE: Xi Pu

MENTOR: Rong Gu

ABSTRACT：Deep learning has been the state of art in the field of machine learning and pattern recognition because of its strong and robust learning ability. Portrait segmentation is a branch of semantic segmentation. It aims to generate pixel-level prediction as portrait(foreground) or background. Due to the complexity of various backgrounds and the detail area such as clothes and hair, it is challenging to achieve high accuracy in this task. Deep learning methods have far exceeded the performance of traditional methods in semantic segmentation tasks. So I choose the fully convolutional networks and train the model on the labelled portrait dataset. Then I apply the Resnet structure to improve the performance and use a soft boundary to make the model more boundary-sensitive.

I use Tensorflow as the framework and take advantage of its parallel characteristics and pipeline data stream to improve the efficiency of training and predicting. Due to the large volume of data in deep learning, we usually have to store these data in different storage systems. Alluxio transparently connects to existing storage systems and presents them as a single interface to users. Also, it provides local caching of frequently used data which makes deep learning training more cost effective and take less time.

KEY WORDS: deep learning; semantic segmentation; portrait segmentation; FCN; soft boundary; Tensorflow; Alluxio

目录

[1 引言 5](#_Toc513040606)

[1.1 研究问题及研究目的 5](#_Toc513040607)

[1.2 研究背景和相关工作 6](#_Toc513040608)

[1.2.1 语义分割 6](#_Toc513040609)

[1.2.2 肖像抠图 6](#_Toc513040610)

[2 流程和方法 7](#_Toc513040611)

[2.1 全卷积神经网络 7](#_Toc513040612)

[2.2 加入Resnet结构 7](#_Toc513040613)

[2.3 边缘模糊化处理 7](#_Toc513040614)

[2.4 数据增强 7](#_Toc513040615)

[3 实验结果 7](#_Toc513040616)

[3.1 性能指标IOU 7](#_Toc513040617)

[3.2 实验对比 7](#_Toc513040618)

[4 分布式存储和训练加速 7](#_Toc513040619)

[4.1 Tensorflow加速训练和预测 7](#_Toc513040620)

[4.2 Alluxio统一接口 7](#_Toc513040621)

[5 结论和后续工作 7](#_Toc513040622)

[5.1 实验中的不足 7](#_Toc513040623)

[5.2 后续的工作 7](#_Toc513040624)

[参考文献 8](#_Toc513040625)

# 引言

1.1 研究问题及研究目的

图像的语义分割是计算机视觉领域一个很基础的问题，语义分割的目的是将图片中属于不同类别的物体以像素点的精度标注出来。语义分割是很多计算机视觉任务的第一步要求，高精度的分割使得一些应用能够得到实际价值，例如自动驾驶中的物体识别，图像搜索引擎，医疗图像的精确识别，增强现实等等。与通常的语义分割任务相比，人物肖像的分割对精度有更高的需求，比如人物的边缘，包括衣服，头发等细小的区域，而且一般来说，人物的背景具有多样性和复杂性，将肖像和背景区分开有较大的难度。人物肖像的分割提取可以理解为，对一张肖像图进行像素级别的预测，对每一个像素点判断其属于前景(人物)还是背景。在人物肖像的分割中，对边缘区域的高精度识别是最重要的也是最具挑战性的问题，对边缘的准确的，平滑的标注可以带来更好的视觉效果和更方便的后续处理操作。

对人物肖像进行分割后，可以很容易的进行一些简单的应用，如图1.1所示[1]，(a)为输入图像，(b)为进行分割后的结果，白色为识别出的人物肖像前景，黑色为背景。图(c)-(e)为对分割后的结果进行不同的后续处理的图像，分别为风格化的人物肖像，深度背景图，卡通化的人物肖像。如图1.2所示，在完成肖像分割后，可以很容易地替换背景，将人物肖像和不同的背景相互组合。

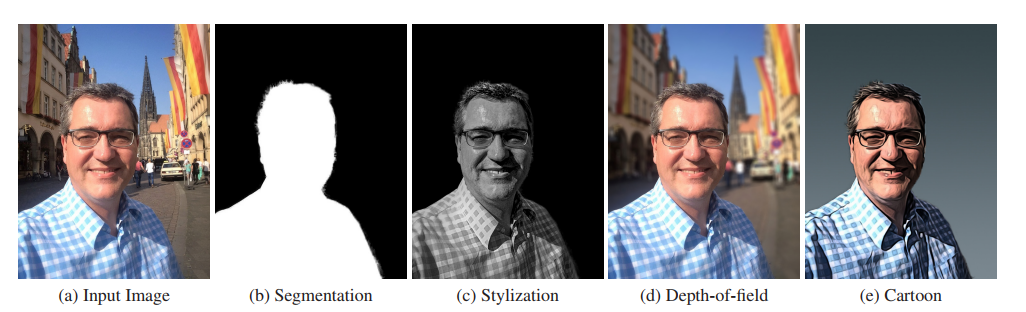


图1.1



图1.2

1.2 研究背景和相关工作

1.2.1 语义分割

目前进行语义分割的主要方法是基于深度学习的模型，深度学习模型往往是作为特征提取的工具，可以不需要先验知识地提取图像训练的高维特征，然后选取分类器进行分类。常见的深度学习模型也经常被用作语义分割的一个模块进行使用，所以我先介绍几种经典的深度网络结构。

**AlexNet [2]**

AlexNet是2012年提出的深度卷积神经网络模型，在ILSVRC-2012数据集上获得 84.6%的top-5准确率。AlexNet的网络结构较为简单，包含五层卷积层，一层池化层，3层全连接层，使用ReLu作为激活函数

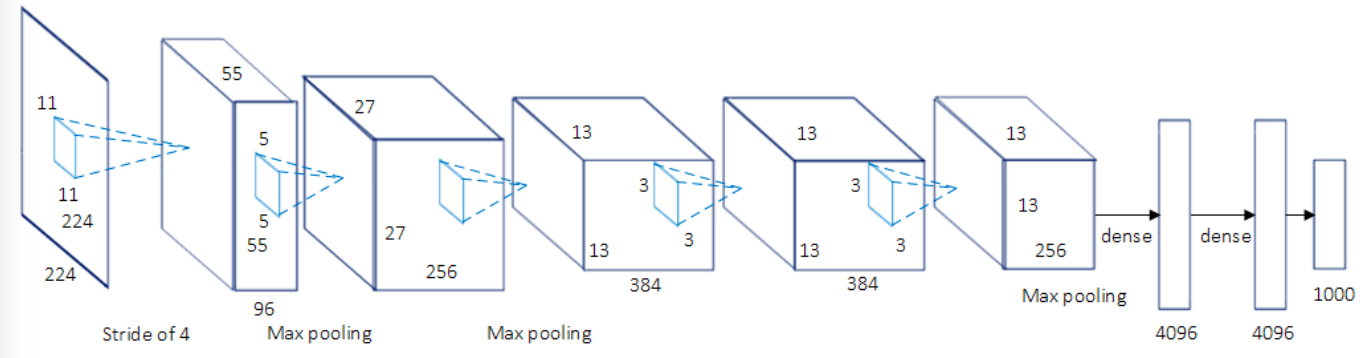


图1.3

**VGG [3]**

VGG将连续的几层卷积组成一个模块，每一个模块后连接池化层，VGG用更小的卷积核和更深的网络结构来提取特征，这样可以大幅减小参数数量，同时更容易拟合非线性函数。VGG-16包含16个卷积层，VGG-19包含19个卷积层。VGG-16在ILSVRC-2013数据集上获得了92.7%的top-5准确率。

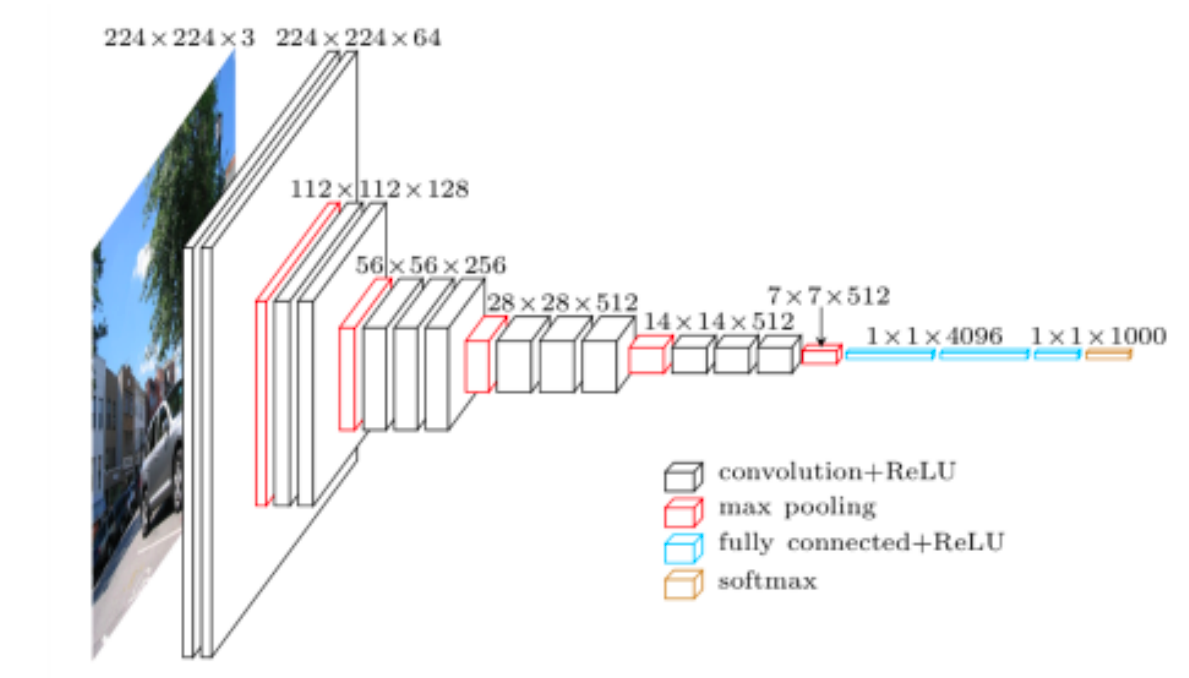


图1.4

**GoogleNet [4]**

GoogleNet是一个很复杂的深度网络模型，包含22层和一种感知模块的的部分。GoogleNet的主要特点是用不同的方式组合网络，包括一种网络中的网络(NiN)的结构，同时采用了1X1的卷积层来提取特征。GoogleNet在ILSVRC-2014的数据集上获得了93.3%的top-5准确率。

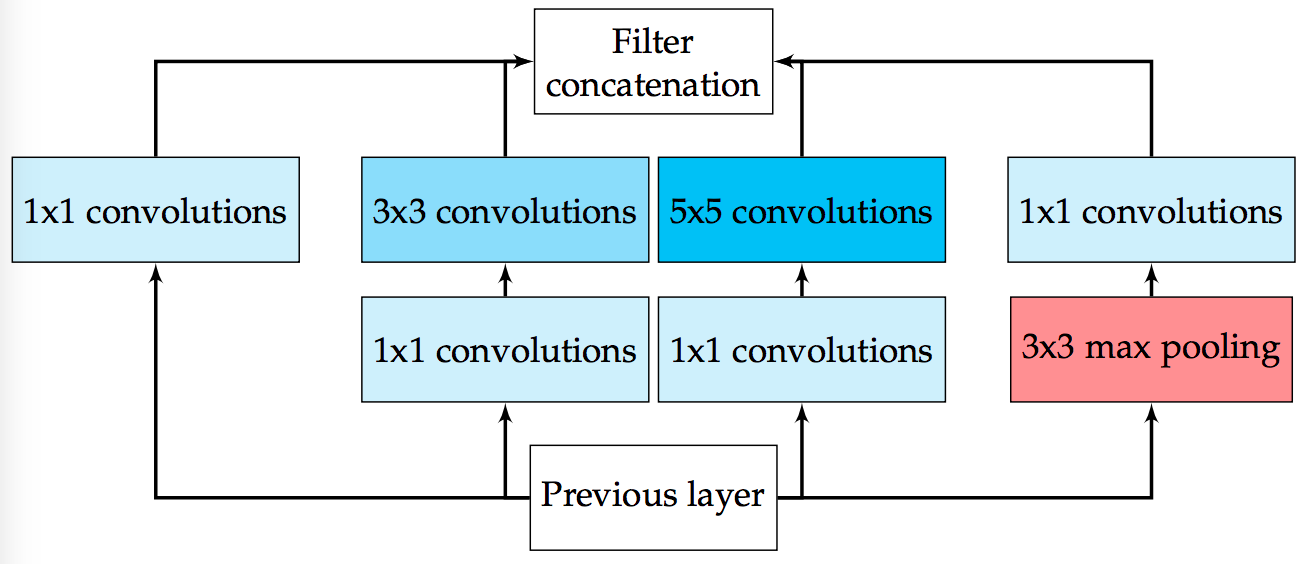


图1.5

ResNet [5]

1.2.2 肖像抠图

# 流程和方法

2.1 全卷积神经网络

2.2 加入Resnet结构

2.3 边缘模糊化处理

2.4 数据增强

# 实验结果

3.1 性能指标IOU

3.2 实验对比

# 分布式存储和训练加速

4.1 Tensorflow加速训练和预测

4.2 Alluxio统一接口

# 结论和后续工作

5.1 实验中的不足

5.2 后续的工作

参考文献

[1] Shen, X. , Hertzmann, A. , Jia, J. , Paris, S. , Price, B. , Shechtman, E. and Sachs, I. (2016), Automatic Portrait Segmentation for Image Stylization. Computer Graphics Forum, 35: 93-102.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097–1105.

[3] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

[4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. arXiv:1512.03385

[] Shen X., Tao X., Gao H., Zhou C., Jia J. (2016) Deep Automatic Portrait Matting. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9905. Springer, Cham

[] Xianzhi Du, Xiaolong Wang, Dawei Li, Jingwen Zhu, Serafettin Tasci, Cameron Upright, Stephen Walsh, Larry Davis. Boundary-sensitive Network for Portrait Segmentation. arXiv:1712.08675

[] Jonathan Long, Evan Shelhamer, Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. CVPR 2015.

[] A. Garcia-Garcia, S. Orts-Escolano, S.O. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez.

A Review on Deep Learning Techniques Applied to Semantic Segmentation. arXiv:1704.06857