

硕 士 研 究 生 读 书 报 告



题目 手绘角色自动着色的深度法向估计

作者姓名 刘振江

作者学号 21951027

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○ 19 年 12 月

Deep Normal Estimation for Automatic Shading of Hand-Drawn Characters

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor: Li Qilei

By

Liu Zhenjiang

Zhejiang University, P.R. China

2019. 12

**摘要**

本文提出了一种用于生成手绘图片阴影效果的全自动渲染的模型。 该模型的输入为任意分辨率的线稿，输出为一张适合渲染的法向贴图，且其中并不需要任何的人为输入。该方法的核心是一个深度残差的编码-解码卷积网络。首先，通过将输入的线稿采样成若干小的3通道图片，每张图片用以3种不同的比例的窗口捕获感兴趣的局部区域。然后，将每张图片经过训练好的神经网络结构以得到局部法向贴图。最后，我们将所有小的法向贴图整合在一起以得到一张与输入图片大小相等的法向贴图。同时，本文还提出了一种有效的方法来生成一组丰富的训练数据。实验结果显示，本文提出的网络模型能够毫不费力的将线稿转换为高质量的渲染图片。

**关键词**： 卡通， 非真实渲染，法向估计，深度学习

**Abstract**

This paper presents a fully automatic rendering model for generating shadow effect of hand-drawn pictures. The input of the model is a line draft of any resolution, and the output is a normal map suitable for rendering, and there is no need for any human input. The core of this method is a deep residual convolution network. First of all, by sampling the input sketch into several small 3-channel pictures, each picture is used to capture the local area of interest with three different scales of windows. Then, each image is trained by neural network structure to get local normal map. Finally, we integrate all the small normal maps to get a normal map of the same size as the input image. At the same time, this paper also proposes an effective method to generate a group of rich training data. The experimental results show that the network model proposed in this paper can easily convert line draft into high-quality rendered image.

**Keywords：**Cartoons, non-photorealistic rendering, Normal estimation, Deep learning

**1论文内容简介**

这篇论文来自于2018年ECCV，标题为《Deep Normal Estimation for Automatic Shading of Hand-Drawn Characters》，作者是Hudon，该篇文章利用卷积神经网络完成了计算机图形学中的法向估计，与本人的专业方向很相似，因此选为读书报告的文章。

**一、引言**

尽管三维动画和艺术作品的激增，二维绘画和手绘动画仍然是重要的艺术传播媒介。这主要是因为二维绘画不受任何约束工具的约束，给艺术家带来了最高的表达自由。艺术家创作通常分为三个步骤：首先，他们创建原始动画或绘图，包括找到正确的场景组成、角色姿态和表达。其次，他们对艺术品进行精炼、数字化，并清理线条艺术。最后，他们添加颜色或装饰纹理，灯光和阴影。在绘制大量图形时，其中一些步骤可能会变得相当乏味。为了帮助完成这些耗时且重复的任务，科学家们尝试将管道的某些部分自动化，例如，清洁管线图、扫描、着色、开发图像注册和中间技术。

本文提出了一种高质量和高分辨率的法线贴图的估计方法，该方法适合于在线稿和动画中添加合理和一致的阴影效果。与SOTA方法不同，本文提出的技术不依赖于几何假设或额外的用户输入，是在没有任何用户交互的情况下直接在输入线稿上工作。

本文的贡献可以总结如下：

* 本文提出了一种新的CNN模型，用于从线稿中预测高分辨率正态图；
* 本文提出了一种新的输入数据的分块多尺度表示方法，用于有效和定性的预测；
* 本文提出一个取样策略来产生高质量和高解析度的正态图，并与包括全卷积网路的最近的CNN作比较。

2 具体方法

作者在文中提取了一套全新的pipeline，通过这套方案能够较好的完成法向生成和卷积神经网络训练的要求。

2.1 输入数据格式

本文提出了一种新的数据表达形式。通过将多尺度的图片信息输入到卷积神经网络中可以帮助网络更好的抓住输入图像的局部特征，同样的也能保证网络在处理时丢失高尺寸的密度信息，具体输入细节如图1所示。

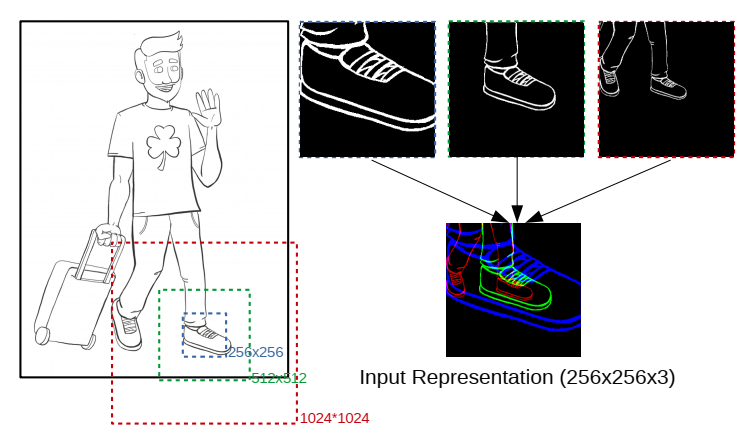


图1 图像输入格式

2.2 模型构建

本文需要的网络输出格式为与输入图像同等大小的图片，因此便采用了U-Net网络结构。在U-Net网络结构中，图像主要分成两个部分，即编码和解码两个分支，完整网络结构图如图2所示。

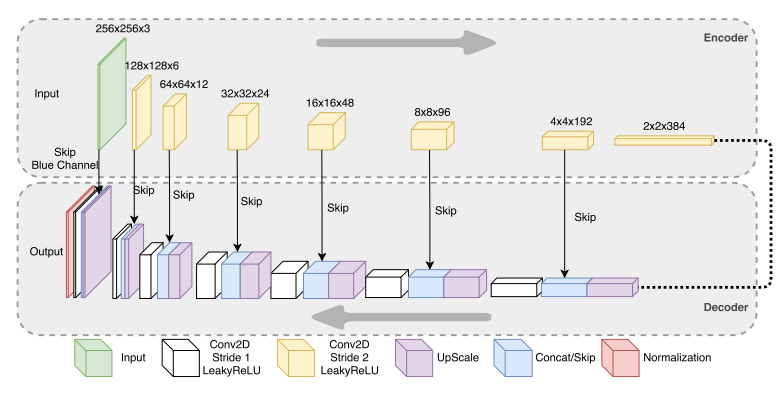


图2 神经网络结构图

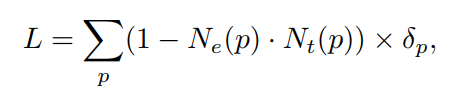
编码器：在编码器中，整个网络结构比较简单，主要由卷积操作组成，其中值得一提的是在本文中采用了步长为2的卷积来代替最大池化。同时因为网络结构最后期望得到的值为[-1, 1]，因此在本文中采用了Leaky ReLu作为激活函数，避免使用ReLu函数带来的一些问题。

解码器：解码器的主要作用是对特征向量进行上采样，除最后一层外，其余每一层的结构均有上采样，卷积操作和激活函数这三部分组成。此外，根据U-Net结构，本文还将对应层特征向量大小一致的进行skip-connection操作，通过这一结构能够有效防止梯度丢失的现象发生。每一层都接收大小一致的对应特征图，其中最后一层只接收输入特征向量的第一个通道。最后网络经过一个L2正则化以保证数据特征的一致性。

2.3 模型训练

众所周知，训练一个卷积神经网络模型需要大量的数据。在本文设计的模型中通过将Blender模型中的42个视角进行截图并获得其对应的法向贴图。然后将其分割成200张256\*256大小的图片，最后的图像pair对在84000张。

本文设计的损失函数如下：



其中，和分别表示预测的法向贴图和真实的法向贴图，其中表示该点是物体前景还是背景，前景表示1，背景表示0。本文选用Adam优化器作为训练的优化器，学习率设置为0.001。

2.4 网格重建

本文设计的网络结构在输入时，将图像大小拆分成若干200张256\*256\*3大小的图片，这些图片经过网络传播之后得到对应的法向贴图。但是网络得到的图像大小格式为256\*256\*3，如果单一的将这些图像组合在一起时会出现像素不连续的情况。因此为了能够比较好的求解输出法向贴图，本文提出了一种网格重建的方法，操作具体示意图如图3所示。

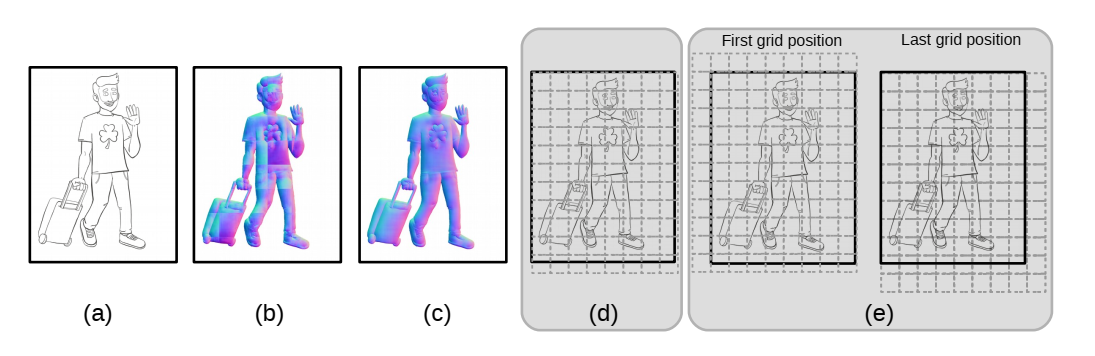


图3 网格重建图

通过将一张完整的输入图片经过一个简短的步长进行输入，最后将对应位置得到的输出值进行平均连续化便可以使得图像保持正常的一致性。通过RMSE方法对各种方法进行验证，作者发现利用10个网格重建既能保证良好的效率，又能使得像素断层不是特别严重。

2. 5 图像渲染

输出得到法向之后，单一的观看法向无法对整体的结果进行预测，因此该文章利用OpenGL对法向进行渲染，最终得到渲染后的图片。具体做法是将上色后的图像和法向进行计算，对其添加光源最后得到结果图，具体示意图如图4所示。

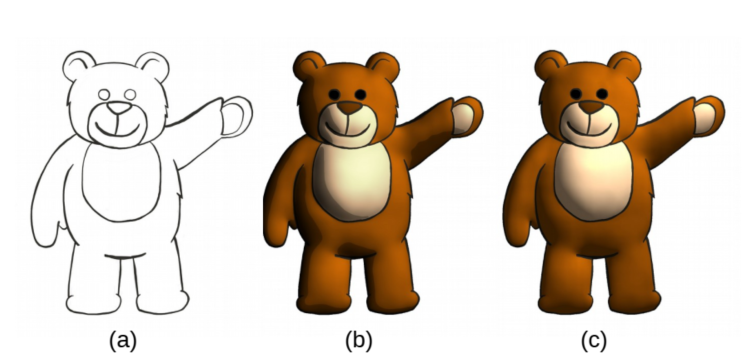
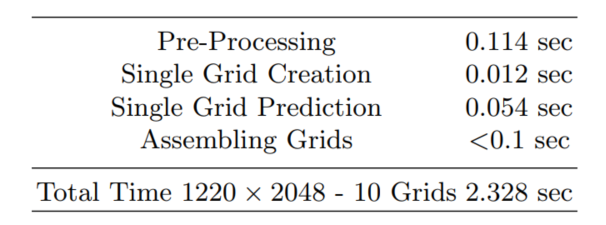


图4 渲染示意图

3 结果

该篇文章从定性和定量的角度对文章中提出的方法进行了验证，验证的环境为Python和TensorFlow，硬件配置为Intel Core I7、32G内存和NVIDIA TITAN Xp GPU。表1描述了该方法在每个步骤所消耗的时间。

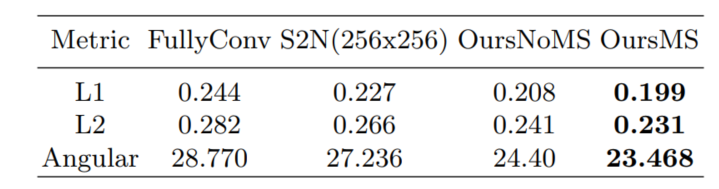
表1 各步骤时间消耗表



从上述实验结果可以看出，文章提出的方法时间效率较高，在每个步骤消耗的时间均较小，只需要2s即可得到一张完整的法向贴图。

此外，文章还定量的测试了方法的各个方法的平均误差，实验得到误差结果如表2所示。

表2 各方法平均误差统计表



从上表的实验结果可以看到，文章提出的方法在各个衡量指标上表现均比其余方法来的优越，说明这个方法的精度较高，且相比于之前的各种算法性能都有了较为明显的提升。

作者还从多个方法对方法的效果进行验证，通过不同角度的光照信息来对图像进行照射，最后得到了很多结果，具体的实验结果图如图5所示。



图5 不同角度的渲染结果图

在补充材料中，作者还展示了文章算法得到的其他效果图，具体的效果示意图如图6所示。

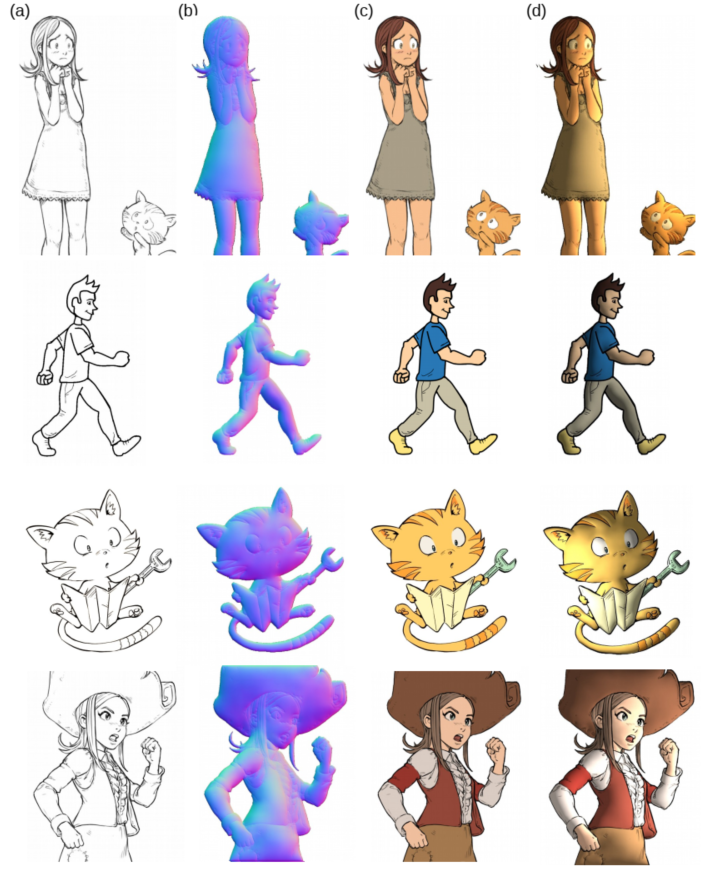


图6 实验结果图



图7 实验结果图

从上述结果可以看出，本文提出的方法能够有效的完成法向重建，并通过OpenGL的纹理贴图和光照渲染可以使得原有的二维图片变得更加生动形象，使得整张图片的生命力得到了提高。

4 小结

总体来说，本文提出的方法具有较好的泛化性，减少了在绘画中的重复工作，且效果优良，但是从另一方面来说整套pipeline略显复杂，若拿来应用落地还需要一定的时间。这篇论文表明，卷积神经网络也能够在3D重建上具有较好的表现，以后可以在这个方向上投入较多的精力，看看更多的文章，争取能在这个领域产出一定的结果。

参考文献

[1] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1125-1134.

[2] Hudon M, Grogan M, Smolic A. Deep Normal Estimation for Automatic Shading of Hand-Drawn Characters[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 0-0.

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

[4] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.