

硕 士 研 究 生 读 书 报 告



题目 深度提取漫画结构线

作者姓名 谢佳烽

作者学号 21851080

指导教师

学科专业 软件工程

所在学院 软件学院

提交日期 2018 年 12 月

深度提取漫画结构线

Deep Extraction of Manga Structural Lines

谢佳烽 21851080 刘二腾 软件工程

摘要：从富含图案的漫画中提取结构线时将传统漫画迁移至数字领域的关键步骤。本文提出了一种基于卷积神经网络的新的数据驱动方法来识别漫画中的结构线。为此，本文设计了一种深度网络模型来处理各种图案并提高输出精度。此外，本文还开发了一种高效的方法来生成大量的训练数据对。通过用各种风格的漫画进行评估后，本文的方法在视觉上基本优于现有的较为先进的方法。

**关键词**：数字漫画，结构线提取，卷积神经网络

Abstract：Extraction of structural lines from pattern-rich manga is a crucial step for migrating legacy mange to digital domain. In this paper, we present a novel data-driven approach based on convolutional neural networks to identify structural lines out of manga. To do so, we design a deep network model to handle various patterns and raise output accuracy. We also develop an efficient way to generate a rich set of training data pairs. After the test on several manga images of different styles, our method substantially outperforms present advanced methods in terms of visual quality.

**Keywords**：digital manga, structural lines extraction, convolutional neural network

1 引言

随着互联网的发展和便携式设备的广泛普及，越来越多的传统漫画转换为电子漫画。基于纸张的传统漫画大多采用黑白打印以节省成本，相比之下，电子漫画可以轻松引入更多视觉元素，甚至可以针对不同的设备调整到相应的分辨率，以实现最佳显示效果，因而具有更好的前景。

漫画的数字化过程通常从扫描开始，然后是各种语义处理过程，再对之进行修改。在语义处理过程中，往往都需要对漫画中的结构线进行识别和提取，它是漫画中用于描绘轮廓的线条。在漫画中，有各种类型和样式的图案，它们可以是规则的，也可以是不规则的；可以由线组成，也可以由点组成；可以是纯色的，也可以是颜色渐变的。因此，对于手工漫画而言，要准确提取出高质量的结构线依然是一个比较耗时的过程，而自动提取结构线的技术能够较好地解决这个问题。本文将在现有提取结构线技术的基础上，提出一种可用性较强、提取质量较高的提取方法。

2 现有相关技术

目前已有的线提取技术大致可以分为三类，即边缘检测方法、纹理去除方法和基于卷积神经网络（CNN）的技术。

2.1边缘检测方法

边缘检测是一个比较经典的研究内容，基于该问题已经提出了多种边缘检测器，比如Canny边缘检测器[1]、高斯拉普拉斯算子（LoG）和基于流的高斯差分滤波器[2]。它们在检测物体的整体结构时有较好的效果，而当应用于从漫画图像中提取精细结构线时则表现不佳。这些检测器都十分依赖于梯度变化，很容易把黑白高对比度的图案和结构线混淆，尤其是当线条处于漫画图案中时，很难识别出结构线。

尽管目前也提出了一些为提取漫画结构线而作出改进的边缘检测方法，如Zhang等人[3]通过结合Canny边缘检测器和Steger的链接算法来检测漫画中的线条。但这些方法依然存在相似的问题，它们往往只能应用于包含较少图案的漫画图像，而在黑白漫画中往往需要更多的图案来弥补它在色彩上的不足。另一方面，此类方法同样十分依赖于梯度变化，不能较好地应用于高对比图案。

2.2纹理去除方法

纹理去除方法多用于一些规则的图案，而在不规则的图案上没有太大效果。为了处理这些随机性较强的图案，提出了基于局部振荡的纹理平滑技术，例如双边滤波、总变差正则化、L0梯度优化和相对总变分。这些方法也能用于区分结构线和图案，但通常仅限于较小的图案，而漫画中往往包含大型的图案，如稀疏的网格或条纹。

同样，也有多种针对漫画的纹理去除方法，如Qu等人[4]提出通过分析Gabor小波特征，用水平集对屏蔽区域进行分割。虽然该方法可以识别出随机性较强的图案，但是当纹理特征在边界附近突然发生变化时，该方法不能精确得识别出边界，从而产生的边界会非常粗糙。另外，当内部结构线的纹理特征与周围图案的纹理特征非常相近时，也会难以识别出结构线。

2.3基于卷积神经网络（CNN）的技术

基于深度神经网络，也提出了几种边缘提取方法。如HED边缘检测[5]采用多尺度特征提取网络，并通过网络输出与不同接受场的融合生成整体边缘图。虽然这种方法在应用于普通图像时有不错的效果，但当应用于漫画时，容易将漫画中图案的纹理线误认为是结构线。这是因为漫画中的图案往往与结构线有相似的线条特征。Semo-Serra等[6]提出了一种基于CNN的草图简化方法。他们对网络结构进行了优化，并用相应的训练数据对其进行训练。这种方法的不足在于仅局限于简单的线条图，而难以应对相对复杂的漫画。本文设计了一种新颖的、更深入的CNN结构来克服漫画中相对拥挤的内容，并针对提取漫画结构线这一目的准备了训练数据，同时对处于图案之中的结构线也能实现较好的提取。

3 设计CNN模型

对CNN模型的设计主要依照两方面的要求。首先，提取得到的结构线要和原始图像的结构线在像素级别上对齐，并尽量避免线条粗细的改变。其次，由于输入图像处于灰阶（取决于扫描的分辨率），最好保留输出中结构线的像素强度，且不引入额外信息或移除原始信息。基于这两点，本文选择了像素级别的CNN模型。

典型的像素级别的CNN模型由编码网络和解码网络两部分组成。编码网络来自于分类网络，同时将输入图像压缩成特征向量。这个提取过程通常由一系列块构建，这些块由三类层组成，分别为卷积层、激活层和池化层。具有步幅的池化层经常用于减小输入的大小。解码网络在编码网络之后构建，作用为从编码的特征向量重建所需要的输出。同样，解码网络也由包含三种类型的层的块序列建模，包括卷积层、激活层和上池化层。其中，上池化层用于将特征向量重新采样为原始图像分辨率。

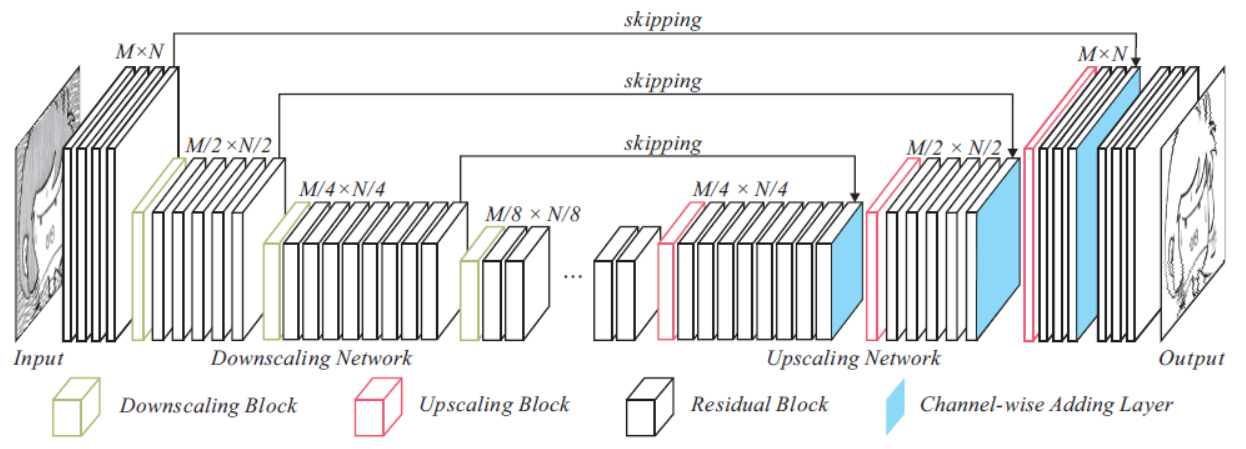


图1 网络整体结构

上图为本文提出的网络结构，它具有缩放结构，左半边为缩小网络，右半边为放大网络，放大块的数量和缩小块的数量必须相同，以确保输入和输出具有相同的分辨率。具有相同大小的块处于同一个级别，每个级别包含一个缩小块。缩小网络通过一系列缩小块和多个级别的残差块提取图像特征，残差块用以增加网络深度。本文发现，3级缩小足够滤除各种大小图像的大部分纹理成分。当减少缩小级别时，难以过滤掉纹理图案；而进一步增大缩小级别时，滤除效果没有提高，且会不必要地过滤掉重要成分从而导致输出中的结构线变得模糊。由缩小网络产生的结构线相对比较粗糙，而放大网络能够通过反卷积将输出从粗糙到精细进行重建。

经典的CNN模型常用普通块来增加网络深度，这些块可以是缩小、放大或者规则的，它们由卷积层、批量归一化层和修正线性单元层按序组成。但是这些块太过简单，且其本身的流线型可能会导致降级问题，因为梯度难以从较高水平传播到较低水平。当输入中包含较复杂的内容时，即使用更深的卷积，也难以得到高质量的输出。因此，本文在CNN模型中采用了残差块，其主要不同点在于它能从输入到输出建立一条捷径，从而使输入绕过流线处理，并直接按通道添加到输出通道，以保存输入中的重要信息。通过这种方式，可以在输入和输出之间直接传递信息，从而提高线提取的质量。

此外，本文在相同大小的层之间创建了跳跃层，以进一步减少由缩小引起的信息丢失。通过跳跃层，缩小网络中的某一级别的图像信息可以在不被压缩的情况下直接传递到放大网络中相同级别的层，从而更好地保留输入的细节，提高输出精度。为此，在放大网络中的每个级别的最后添加一层合并层，用通道式的方式对跳跃处理的图像和顺序处理的图像求和，然后将合并的结果用于下一更高级别的放大。

4 准备训练数据

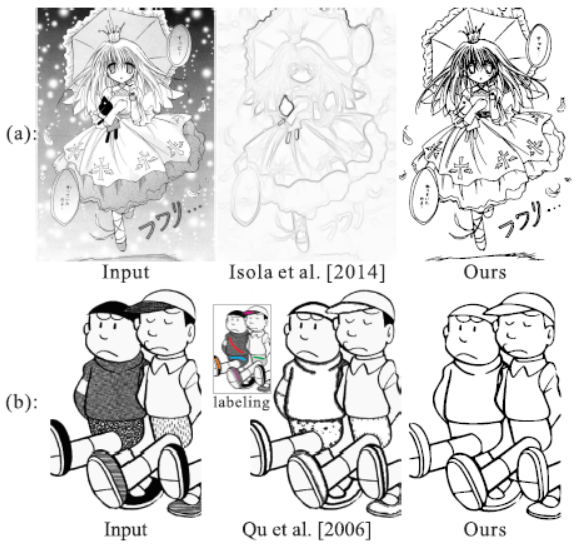
设计基于CNN的方法要求大量的训练数据进行训练，每个训练数据对应该包含一张多图案的漫画和其相应的无图案漫画。手动追踪结构线基本上属于漫画的重新绘制过程，对于人力和时间的耗费是比较严重的，且得到的两张漫画中结构线的位置和样式很难完全匹配。因此本文提出自动生成一个大规模的训练数据集。与繁琐地追踪已有漫画中的结构线条不同，本文采用了一个逆向过程，先创建一个丰富的图案库，再从无图案的线条图中合成图案得到包含多图案的漫画，从而避免了训练数据对不完全匹配以及繁琐的识别提取结构线问题。

本文收集了117种不同风格的黑白线条图以及110种常见图案。这些图案包括36种规则图案、24种不规则图案、27种色调变化图案和23种图片类图案，并通过改变图案的密度、方向和线宽来进一步收集和创建多个变体。接下来在线条图上选择某些区域铺设各种图案来生成包含图案的漫画的过程中，还需要设定一些规则来对线条图和图案进行匹配，来模拟现实中存在的合理的漫画图像。设定的规则大致为：填充区域和图案之间的大小需要进行匹配；含有色调变化的图案应填充到较大的区域，否则色调变化难以识别；包含有结构线的区域不能填充纯黑色，同时两个相邻区域也不能同时填充纯黑色图案。

此外，为了使CNN模型能够适应不同扫描分辨率的输入漫画，可以将训练数据中的图像进行不同程度的缩小，从而将每个训练数据对重新采样为多个分辨率。

5 结果和讨论

为了评估所设计的方法，本文在几个不同风格的漫画图像上进行测试，包括西方的漫画和日本的漫画，并将输出结果和其他已有的方法进行比较。



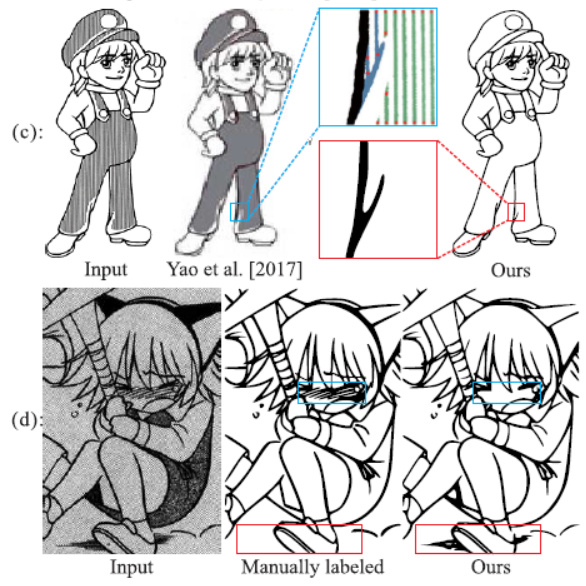


图2 输出结果和其他方法的比较

在图2（a）中，将本文方法和Isola等人[7]提出的清晰边界提取方法进行比较。由于Isola的方法是针对自然图像的，因此从漫画图像中提取高质量的结构线的效果不佳，尤其是当结构线位于图案之中时。此外，由于Isola的方法主要用于检测较大物体的边界，因此容易忽略一些精细的结构线（如头发和裙子上的褶皱）。相比之下，我们的方法可以在去除图案的同时提取出清晰细致的结构线条。

在图2（b）中，将本文方法和Qu等人的漫画着色方法[4]进行比较，漫画着色本质上是一种使用Gabor小波分析进行的图案模式分割方法。这两种方法都可以用于随机性较强的图案。由于纹理分析需要一个分析窗口，且要求位于边界附近的纹理特征和位于区域中心的纹理特征有明显不同，因此通过这种方法容易产生不清晰的、粗糙的边界（如左边男孩的裤子）。相比之下，本文的方法可以获得清晰的结构线。

在图2（c）中，将本文的方法和Yao等人[8]的方法进行了比较。他们提出分析和识别三种类型的图案模式：点、线和网格。如果输入图案的类型属于这三种之一，则他们的方法会对图案进行建模，如图中蓝色框所示，最终能够得到比较好的输出。但是这种方法的局限性太强，当图案不属于这三种类型，或者当点、线和网格的错位或破损都有可能会影响到建模过程。相比之下，本文的方法没有限定图案的类型，也不需要对图案进行建模，因此适用性更加广。

但是，本文的方法也存在一些不足。在图2（d）中，将本文的方法与手动标记结构线的结果进行了比较。例如，图2（d）中鼻子周围的阴影线（蓝色框）在语义上有表达情感的作用，应该进行保留，但本文的方法将线条视为图案而将其移除。此外，图中鞋子后方的线条表示阴影，没有描绘漫画对象的结构，应该去除，但本文的方法将其识别为结构并保留下来。这些差异是由于本文的方法缺乏对图像的更高层次的语义理解而导致的。

6 应用

通过提取清晰的结构线，可以将其应用到其他与数字漫画相关的领域上，比如漫画着色、漫画重定向。由于传统漫画主要为黑白色，漫画着色可能是数字漫画第一个有意思的扩展内容。通过提取得到的高质量结构线，可以轻松准确地分离筛选各个区域以进行着色。与基于纹理分析的漫画着色相比，本文的方法可以重现更加清晰、平滑的界限。此外，随着现代便携式设备的广泛普及，需要将漫画显示为不同的分辨率，不成熟的重采样技术可能会导致漫画在其他分辨率的显示器上变得模糊。通过将结构线和图案进行分离，就能更加方便、有效地对漫画进行重定向，从而在不同分辨率上都能得到清晰的图像。

7 结论

本文提出了一种基于CNN的方法，用于从富含图案的漫画中提取清晰光滑的结构线，并设计出了一种具有更加深入的网络结构的CNN模型，来处理各种各样的图案和提高输出结果的质量。此外还开发了一种自动生成大量训练数据集的有效方法。将本文的方法与目前已发表的其他方法进行比较后，发现用本文的方法得到的漫画结构线更清晰，更接近于原图。最后，还对本文的方法所潜在的应用领域进行了展望。

参考文献

[1] Canny J. A Computational Approach to Edge Detection[M]. 1986.

[2] Kang H, Lee S, Chui C K. Coherent line drawing[C] International Symposium on Non-photorealistic Animation & Rendering. 2007.

[3] Zhang S H, Chen T, Zhang Y F, et al. Vectorizing Cartoon Animations[J]. IEEE Trans Vis Comput Graph, 2009, 15(4):618-629.

[4] Qu Y, Wong T T, Heng P A. Manga colorization[J]. Acm Transactions on Graphics, 2006, 25(3):1214-1220.

[5] Xie S, Tu Z. Holistically-Nested Edge Detection[J]. International Journal of Computer Vision, 2015, 125(1-3):3-18.

[6] Simoserra E. Learning to simplify: fully convolutional networks for rough sketch cleanup[J]. Acm Transactions on Graphics, 2016, 35(4):121.

[7] Isola P , Zoran D , Krishnan D , et al. Crisp Boundary Detection Using Pointwise Mutual Information[C]// European Conference on Computer Vision. Springer, Cham, 2014.

[8] Yao C Y , Hung S H , Li G W , et al. Manga Vectorization and Manipulation with Procedural Simple Screentone[J]. IEEE Transactions on Visualization & Computer Graphics, 2016, PP(99):1-1.