

硕 士 研 究 生 读 书 报 告



题目 自动化图像调整方案研究

作者姓名 林友松

作者学号 21551154

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○ 一五 年 十二 月

Research on Automatic Image Adjustment Scheme

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

Lin YouSong

Zhejiang University, P.R. China

2015

摘要

现实生活中摄影师可以通过其自身的拍照技术来提升照片的整体视觉美感，然而，对于普通人来说这却是一项具有挑战性的工作。因此，如果存在一个可以自动对图像进行色彩和色调方面的调整的算法，将能够帮助摄影爱好者们更容易拍出好的照片。

本文重点探讨了照片的自动化调整方案，并在实验过程中与其他现有调整方案方案进行比较，以及分析各种方案在实际情况中的应用效果。并介绍了一种基于深度神经网络来建模，引入一个图像描述符来解释图像的语义的不仅仅依赖于图像内容的图像自动化调整算法。

**关键词**：图形处理，特征描述，神经网络

Abstract

In real life, the photographer can enhance the overall visual sense of beauty through his own skills, however, it is a challenging task for the average person. Therefore, using an automatic algorithm can adjust the color and tone of the image can help photographers more easily take a good picture.

This paper focuses on the automatic processing of photos, and compares the advantages and disadvantages of various schemes, and the application effect of various schemes in the actual situation. In this paper, a new method is introduced, which is based on the neural networks, and an image descriptor is introduced to explain the image's semantic.

**Keywords：**Graphics Processing, Feature Descriptors, Neural Networks

1引言

随着数码摄像机和社交网络变得越来越流行，通过社交媒体分享照片已经变得非常流行了。最常见的就是使用类似INS这样的照片分享社交平台，这类APP内置了很多对图片进行效果增强的功能。

在一般情况下，这些APP处理过的照片可以在一定程度上增强图片的美感，因为它不仅会尝试修正摄影缺陷（欠压/过曝光，对比度差，等等），而且还会通过夸张的颜色和色调调整来达到戏剧性的视觉效果。从传统上来说，高品质的图片是由一群拥有娴熟图片处理技术的摄像师们编辑出来的。

在这项工作中，作者研究的是如何从图片样本中学习艺术照片增强样式的问题。具体来说，给定一组图片样本，组中每个图片代表特定的风格，其包括像素级别的色调和色彩增强。我们需要寻找出一个计算模型，以便对新输入的照片自动应用所学习的计算模型来对照片风格进行修正。

学习一门高质量的艺术照片增强技术是很有挑战性的。首先，照片调整通常是一个需要经验积累和相当感性的过程，这要求能够熟练的通过某一种方式来调整图片的各方面属性以达到增强效果。此外，一个艺术家是不会关注图像的单个像素的；相反，他/她只会关注语义上的有意义的对象（例如人，汽车，动物等），并通过一定的手段修正这些对象的外观。例如，某个艺术家更注重改善人类的外貌，而不是去修改同一张照片上的天空区域。我们想把这种语义应用到在我们的学习模型上。其中一个挑战是如何在学习模型中正确的表示语义信息，以使得该学习模型可以像人类一样根据具体的图像内容对图像进行相应的调整。

作者提出了一种基于机器学习的自动照片增强方法。这种方法最近在计算机视觉和语音分析数据的语义等领域的研究中起着重要的作用。面对上述挑战，作者把基于样例的照片视作回归问题，并使用深层神经网络（DNN）与多个隐藏层来表示输入和增强图像之间的高度非线性和空间变化的颜色映射。深度神经网络（DNN）是可以代表任意复杂连续函数的一个通用逼近。同时，它也是一个易扩展的紧凑模型。

**2 相关研究成果**

传统的图像增强规则主要是操作人员根据经验来确定的。现在也已经有许多软件工具可以进行全自动色彩校正和色调的调整了，例如Adobe Photoshop，Google Auto Awesome，和Microsoft Office图片管理器。除了这些工具外，还有许多关于交互式或自动颜色和色调调整[1]的相关研究。自动化方法通常不会考虑图像得具体内容而是从全局的角度对图像进行操作。为了解决这个问题，Kaufman等人提出了一种自动化的方法，这种方法首先会检测图片的语义内容，包括脸、天空以及图中显著的阴影区域，然后应用一系列有效的步骤来调整图像的饱和度、对比度以及曝光等。然而，这种方法的最大的限制是输出样式是被硬编码到算法中的，不能容易地进行调整以实现所需的风格。

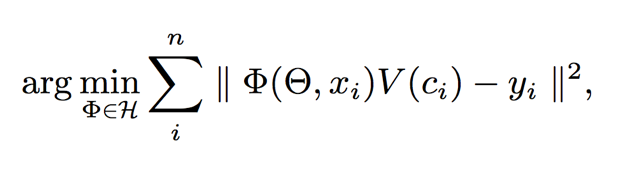
照片调整的另一个研究路线主要是基于数据驱动模型的研究。基于学习的图像增强[2]和图像恢复研究已得到可喜的成果，因此这方面研究逐渐受到了越来越广泛的关注。Kang等人发现，图像质量评估实际上是非常个性化的，这引申出一种基于全局图像调整角度来进行个人偏好学习的自动化处理方式。Bychkovsky等人提出了一种基于高斯过程的根据全局图像分析的色调映射方法。由于这些方法是基于全局图像的角度来处理的，所以它们不会考虑本地图像的背景，也不能进行局部的增强处理。Wang等人提出了一个基于分段函数逼近的颜色映射的方法。这个方法并不考虑语义或上下文信息。此外，它不是完全自动的，并一定程度上需要依赖于软分割。相比之下，论文的作者提出了一种可自动学习用户自定义的一些复杂增强模式的可扩展框架。它可以分析出图像的确切语义，并对语义分析的结果使用某种特征描述来构建图像增强模型。

Hwang等人提出了一种基于上下文感知的局部图像增强技术。该技术首先从数据库中搜索最相似的图像，然后再搜索其中最相似的像素，最后将那些最相似的像素通过某种参数的组合来进行合并。当有一个足够大的图像数据库时，这种方法的效果是很好的。但在实践中，邻近搜索需要一个相当大的训练集，而这是一个非常有挑战性和缓慢的搜索过程，因此，这极大的限制了这种方法的可扩展性。

**3 一个深度学习模型**

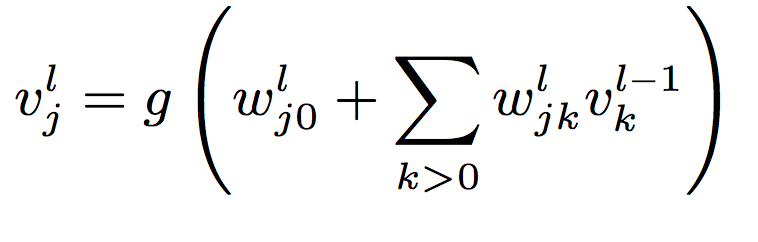
一种照片增强样式可以用一个典型的图像对Λ={Ik,Jk}mk=1来表示，Ik和Jk分别表示增强前后的图像。我们假设存在一个固定的颜色映射函数F使得Ik中的每一个像素的颜色都可以映射为任意k值的Jk中的一个像素的颜色。我们的目标就是通过Λ来找出一个逼近函数F~，这样我们就可以通过将这个映射函数应用到新的图像中实现相同的图像增强效果。对于图像Ik中的每一个像素pi，函数F~映射的值近似于Jk中的像素点pi。然而，函数F~的输入是非常复杂的，因为F~的输入不只是pi的颜色，也包含了Ik中获取的各种局部或全局信息。因此，我们函数定义为F~(Θ, xi) ，Θ表示参数，xi表示Ik中的像素点pi的特征向量以及额外的局部和全局信息。

频繁的像素间颜色变化的比较是很难通过建模来解决的，因为这需要我们找出一个对频繁的细节变化很敏感的映射函数。然而这样的映射函数通常会在相对平缓的区域产生一些意外的结果。为了解决这个问题，我们使用基于像素点Pi的向量V(ci)去重写F~为F~=(Θ,xi)V(ci)。V(ci)是一个向量函数，当他被用来处理不同的颜色变换时会有不同形式。作者在论文中是基于CIE Lab颜色空间的，所以如果我们使用3\*4的仿射颜色变换的话，Pi的颜色表示为ci=[Li ai bi]T，V(ci)= [Li ai bi 1]T。如果我们使用3\*10的平方颜色变换，则V(ci)=[L2i a2i b2i Liai Libi aibi Li ai bi 1]. 然后得到如下公式：



**3.1 神经网络结构与训练**

为了更完整的描述，我们的神经网络将会遵循接下来将要描述的标准体系结构。多层深神经网络已被证明是能够代表任意复杂的连续函数。每一个网络都是一个非循环图，其中的每个节点都是一个神经元。神经元被组织在多个层中，包括一个输入层，一个或多个隐藏层，以及一个输出层。输入层直接映射到输入特征向量。输出层被映射为颜色变换(Θ,xi)的元素。隐藏层或输出层中每一个神经元作为前面层的输入而得到的响应。每一对神经元之间的连接都与一个权值相关。然后把vlj看作是l-th层中的神经元j-th的输出。因此可以把vlj表达为：



请注意，在输出层中没有任何神经元的非线性激活函数。输出层中的一个神经元的输出仅是其输入的线性组合。图1显示了整体架构，其中有2层（黄色和紫色的神经元）以上的输出层用于计算产品之间的颜色变换和颜色的基础矢量。一旦整个网络的体系结构被确定，对于给定的训练数据集，我们可以使用经典的误差反向传播算法得出权重。此外，我们还采用了丢弃输出的训练策略，它已被证明对于提高网络的泛化能力是非常有效的。

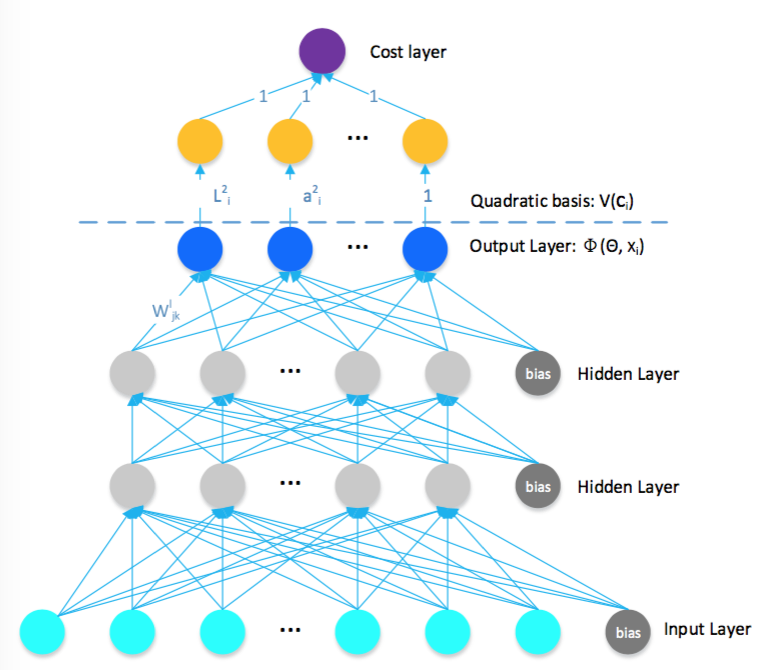


图1 深度神经网络的体系结构

我们设置的每个隐藏层中神经元的输出概率为0到0.5。那些被“抛弃”的神经元将不会进行正向传递，也不会参与误差反向传播。实验表明，在训练过程中进行“抛弃”通常会使得出错的几率降低2.1%，这实际上对图像的增强效果的差异起非常重要的作用。

**4 特征描述符**

每个样本中的像素pi对应的特征描述符被当做神经网络的输入层。其主要有三个部分，xi=(xpi, xci,xgi),xpi表示基于像素的特征，xci表示从全局区域pi计算而得出的上下文特征，xgi表示从整个图像的像素pi计算而得的全局特征。三个部分详情如下：

**4.1 像素特征**

像素特征反映的是高分辨率图片的变化，对于多样化的图像增强模型的学习是不可或缺的。它们被定义为xpi=(ci, pi)，ci表示3\*3邻近区域基于CIELab颜色空间下的平均颜色，pi=(xi, yi)表示图像中样本的位置。

**4.2 全局特征**

在摄影实践中，图像的平均强度,整体的属性和整体的印象，在艺术家决定如何提高图像的时候都会产生一定的影响。因此，我们将全局图像特征加入到特征表示中。具体而言，我们提出六种类型的全局特征，包括强度分布，场景亮度，均衡曲线，详细加权均衡曲线，剪裁和空间分布，这些聚集起来可以产生一个207维的向量。

**4.3 上下文特征**

我们的上下文特征尝试描述不同语义类别的分布，例如天空，建筑，汽车，人和树。这些特征是在本地区域范围内的样本像素点中提取出来的语义分析的结果。典型的图像语义分析算法包括场景解析[3]和目标检测[4]。场景解析试图将每一个像素的图像标记为其语义类别。另一方面，目标检测是一个用于不同场景的高度专业化的检测器。场景分析擅长于对那些没有特征形状但有相对一致的纹理的物体进行分类（如草，道路和天空）。这些类别的规模会很大，通常会构成一张图片的整个背景。目标探测器可以更好的识别对象的类别（如人和汽车），这使得它可以更好的描述对象的整体形状而不是局部性状。这些类别有一个较小的规模，通常指前景的图像。因为这两种技术是相辅相成的，所以我们使用一个组合的场景解析和目标检测算法来进行语义分析。图2显示了一个场景分析和检测结果的融合实例。

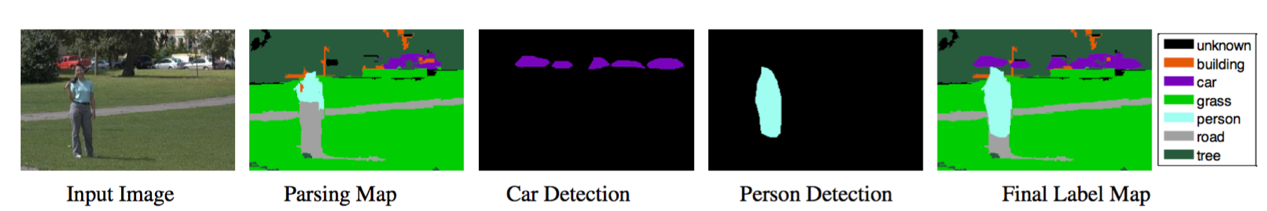


图2 使用组合算法得出的语义检测结果

**5 实验的建立**

**神经网络的建立。**纵观本文所有的实验，我们使用的是具有一个输入层，两个隐含层和一个输出层的固定DNN。隐藏层的神经元数目是根据经验而设置为192，并且输出层的神经元的数量被设定为等于预测的可能要进行色彩变换的数目。我们的实验已经证实，二次颜色转换比仿射颜色变换可以更好地再现色彩调节后的图像。因此，本实验的输出层有30个神经元，其中三个颜色信道每个拥有10个神经元。

**数据采样。**由于我们的实验是基于像素级的颜色映射的，所以图像中每个像素都是一个潜在的样本。在实践中，我们把每张图像划分为大约7000个子像素块，然后每10个像素块随机选择10个像素作为样本。例如，即使只有70样本图像来用于学习一个特定局部效果，也可以得到大约490万个学习样本。如此大规模的学习集很大程度上可以消除过度拟合的风险。但是它通常需要花费几个小时才能完成一个中等大小学习数据集的神经网络。尽管如此，一个经过学习的神经网络只需要0.4秒就可以完成一个512像素宽的测试图像的增强操作了。

**图像增强与学习颜色映射。**一旦得知了神经网络的参数（权重），那么在图像增强阶段，就可以在学习阶段对输入的想象应用相同的特征提取管道。也就是说，首先进行场景分析和目标检测，然后应用将图形进行分割以获得子像素块。接下来提取每个子像素快的质心对应的特征向量，并对子像素块中的每一个像素都应用神经网络中返回的颜色变换来进行操作。具体而言，像素Pi调整后的颜色的计算公式为yi=(Θ,xvi)V(ci)，其中vi是覆盖Pi的子像素块。

**5 结论与讨论**

该本文展示了基于深度学习的自动照片调整的成效。作者抛出的这个问题，即通过把打包好的特征集作为输入层的深度神经网络来学习一个高度非线性映射。那些打包好的特征集包括像素特征描述符，全局特征描述符，以及一种创新的上下文描述符，这些产物都是通过场景解析和目标检测获取的。他们进行了大量的实验，包括了常规的作品或者是和艺术家们的作品进行对比。最终实验表明，该方法是一个能够有效地进行自动图像调整的计算模型。

**限制。**作者的方法某种程度上来说比较依赖于场景解析和目标检测来构建上下文特征。然而，在一般情况下，这些都还是计算机视觉和图案识别方面比较具有挑战性的问题。更可怕的是，一些解析错误的语义可能会被传递到上下文特征中，最后对照片的调整产生不利影响。

作者的系统采用了深度的全连接神经网络来处理空间变化的颜色转换。在DNN的体系结构中还存在很多别的设计方案，包括隐藏层的数目，每个层的神经元的数量以及神经激活函数的类型等。它们可能需要一个非常耗时的实验和出错的过程，才能查找出一个对于给定任务最适合的DNN体系架构。此外，DNN的行为就像一个黑盒子，它并不完全清楚如何合并不同规模和特征，无法准确预测最终的颜色变换。事实上，解释深层神经网络的内部的表示仍是一个正在进行的研究课题[5]。

参考文献

[1] LISCHINSKI, D., FARBMAN, Z., UYTTENDAELE, M., AND SZELISKI, R. 2006. Interactive local adjustment of tonal values. ACM Trans. Graph. 25, 3, 646–653.

[2] BYCHKOVSKY, V. PARIS, S. CHAN, E., AND DURAND, F. 2011. Learning photographic global tonal adjustment with a database of input/output image pairs. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. CVPR ’11. 97–104.

[3] LIU, C., YUEN, J., AND TORRALBA, A. 2011. Nonparametric scene parsing via label transfer. IEEE Transactions on Pattern Analysis and Machine Intelligence 33, 12, 2368?382.

[4] WANG, X., YANG, M., ZHU, S., AND LIN, Y. 2013. Regionlets for generic object detection. In ICCV’13: Proc. IEEE 14th International Conf. on Computer Vision.

[5] ZEILER, M. D. AND FERGUS, R. 2013. Visualizing and understanding convolutional neural networks. arXiv preprint arXiv:1311.2901.