

计算机视觉大作业

21821121 徐 东

阅读《DeepContour: A Deep Convolutional Feature Learned by Positive-sharing Loss for Contour Detection》

一、该文献的核心思想

轮廓检测是视觉的基础任务。解决这个问题的做法是着重于设计工程梯度特征。在这项工作，作者表明，轮廓检测的准确性可以通过使用卷积神经网络（CNNs）学习的深度网络得到提高。而不仅仅是使用深度网络作为黑盒特征提取器，作者通过划分轮廓（正）子类和子类来定制培训策略通过不同的模型参数。定义了一个新的损失函数（*positive-sharing loss*），其中每个亚类分担损失对于整个阳性类，提出了学习参数。与 softmax 损失函数相比较，新的损失函数引入了强调阳性和负性类的损失，这有利于更好的产生可区分的特征。实验结果表明，深层特征可以达到顶峰 Berkeley Segmentation Dataset 和 Benchmark (BSDS500) 的性能和 NYUD 数据集的结果。

二、该文献的基本方法

(1) 对图像先进行形状处理。根据像素的内在结构，对轮廓区块先进行一个聚类。聚类的过程中，产生一个中等形状的代表。这种形式的聚类，称之为形状（shape），每个形状会被标准一个标签（k 值）。然后再用不同的模型拟合不同形状数据这样可以有效的降低模型训练的难度，并提高数据的多样性。

(2) 在 (1) 的基础上，轮廓检测问题，已经转化为一个多类的问题，但是由于 softmax 函数，对所有的类的损失惩罚是相同的，而在轮廓检测中，轮廓区块和背景区块并非是同等重要的，因此作者在 softmax 损失函数的基础上加入了是否为轮廓观测函数，这样能够带来更好的轮廓特征学习，并在实验中达到了最优的结果。

三、实验结果

该文通过固定数据集的 best F-measure、aggregate F-measure 和平均查准率于其他方法进行了对比。

在 BSDS500 数据集上，该文的方法达到了最好的差准和查全率，比之前较好的方法 Structured Edge 和 SE-Var，以及 deepnet 的方法都要好。在 ODS 和 OIS 方面要高出 1 个百分点的情况下，有着较好的平均查准率。对于某些使用最近邻特征的 CNN，我们使用随机森林的方

法，这有效的提高了对抗噪声的鲁棒性。

通过在不同数据集的比较，可以证明，作者的方法具备较好的通用性的轮廓识别方法。

阅读《Holistically-Nested Edge Detection》

一、该文献的核心思想

该文中，作者介绍了一种新的基于卷积神经网络的边缘检测系统，该系统能够在了一幅图像上展示自然图像的最新性能。实际相关的速度（例如，使用 GPU 的 0.4 秒和 12 秒钟使用 CPU）。作者的算法建立在顶层完全卷积神经网络和超深度网络的思想。作者通过初始化的网络结构并通过采用预先训练的修剪 VGGNet 参数。该文认为即使在没有明确的上下文和高层次的信息的情况下，通过对尺度、多层次的可视化反应完成图像到图像的学习是可行的。

二、该文献的基本方法该文主要

针对两个问题进行工作：

（1）整体形象训练和预测，受完全卷积神经网络启发，为了图像到图像的分类（系统需要一个图像作为输入，并直接产生边缘地图图像作为输出）。

（2）嵌套的多尺度特征学习，深受监督的网络启发，通过深度网络“指导”前期分类结果。

三、突出亮点

该文的突出在于在已有的网络架构上，提出了自己的网络结构，并充分利用了已有网络（VGGnet）的特性，文献比较细致的列出了常用的卷积网络框架。

阅读《DeepEdge: A Multi-Scale Bifurcated Deep Network for Top-Down Contour Detection》

一、该文献的核心思想

轮廓检测是典型的低级特征问题，常用来解决物体检测等更高级别任务。然而，可以说物体检测和轮廓预测是密切相关的。因此，在给出轮廓的图像中，很容易推断出有什么样的物体。同样，如果我们给定物体的准确位置，我们也很容易完成轮廓预测。在计算机视觉中常见的处理方法，从低级的轮廓检测开始，然后转向更高层的物体检测。该文利用对象特征信息作为轮廓检测的高层线索，代替使用轮廓作为物体识别的低层线索。因此，某种意义上，该文的方法可以视作是一种利用对象层线索完成低层轮廓检测过程的自上而下方法。

在这个工作中，文献提出了一个统一的多规模深度学习方法，该方法利用高层的对象信息进行轮廓检测。具体地说，作者提出了一个前端到后端的卷积结构，在这个结构中轮廓直接从原始像素中被识别。作者提出的深度学习架构重用了第一个 Krizhevsky 等的网络组成的五层卷积网络的计算特征。作者将这个网络称之为 KNet。因为 KNet 已经被训练进行对象分类，重用它的特征能够体现作者提出的轮廓检测方法使用了对象级信息。此外，作者将定义的结构同时运行在多个规模上，并结合图像的本地和全局信息，这显著地提高了轮廓的检测准确率。将由 KNet 卷积层在四种不同级别输入上计算出来的特征与一个已学习的子网络连接在一起，该子网络分为两个分支。其中一支的权重是利用边缘分类对象进行优化，而另一支利用回归标准训练来预测与人类标签的符合分数。

二、该文献的基本方法

该文，提出了一种自上而下多规模二分叉深度网络的轮廓检测方法。过去，轮廓检测都是自下而上达成的，首先获得低层特征，然后再进行轮廓检测，最后轮廓被用做对象识别的对象。文献展示了如何利用高级对象线索来检测轮廓，并利用高级对象特征带来的轮廓检测准确率的大量提高。

另外，该文还展示了多规模结构在轮廓检测的优势。通过使用多个规模，我们的方法结合候选点周围的局部和全局信息，这导致了显著提高的检测结果。进一步，展示了独立优化的轮廓分类和回归提高了轮廓检测的准确率。正如我们实验展示的，深度边缘算法比之前和当下的算法达到了更高的平均精度。

总而言之，该文献的实验结果显示，单纯的 CNN 系统可以成功用来完成边缘检测，有很大的适用性对于其低层视觉任务。

三、突出亮点

一是反转了从低级特征到高级特征的一般做法，提高了基本视觉任务的完成方法。

二是多规模的卷积层结构，有效的增加了特征向量的维度，显著低提高了 F-score 和平均精度，各方面的性能也要由于单规模卷积层的结构。

三是二分子网络的方法，从文献的及格过中可以观察到单独使用分类子网络产生高的 F-score，而回归分支子网络产生高的平均精度。两个分支的预测值混合后提高了两个维度的值，因此这验证了文献提出独立优化的边缘分类和回归对象对轮廓检测是有效果的。

四是该文避免了手动特征的提取工程，而是直接通过原始数据进行识别。有效减轻了人为参与的工作量。

阅读《Attribute-Enhanced Face Recognition with Neural Tensor Fusion Networks》

一、文献的研究动机

在很多困难场景下, 如大姿态, 基于深度学习的人脸识别性能会下降很多。在这些困难场景下, 如大姿态, 人脸的属性信息(facial attribute)如性别、人种、发色、眉毛形状等是保持不变的。很自然地, 深度人脸识别的特征(FRF)和属性识别的特征(FAF)具备互补性。融合这两种特征可以提高人脸识别率。我们尝试了现有的融合策略, 表现令人难以满意。作者提出了一种基于张量(tensor)的特征融合方式来融合 FRF 和 FAF。

二、主要方法

两个特征用 tensor 连接可以用于特征融合。融合的效果取决于 tensor 参数优化的效果。但 tensor 的维度很高, 难以优化, 作者使用 TUCKER 张量分解的方式降低待优化的 tensor 的维度, 即使这样优化仍然面临着求导及正则化设计的困难。作者认为这个 tensor 优化问题可以与一个 gated 的人工神经网络进行等价。基于这个这个等价的效果, 作者进行了如下操作 (1) 使用现有深度学习的优化工具进行优化, 如 tensorflow, 而不需要手动设计复杂的 tensor 的优化。由于在 tensorflow 优化时可以用 mini-batch, 这样我们的优化天然就是可伸缩的 (scalable) 的; (2) 融合部分可以与特征学习的神经网络进行端到端的学习。这篇论文提供的一个创新是, 将 tensor 和深度学习进行等价。这样 tensor 领域的知识可以用来理解神经网络, 反之亦然。

三、实验结果

该文主要是在 3 个数据库进行了对比实验

(1) 在 MPIE 库上, 作者测试了一下属性特征在大姿态, 表情和光照情框下对识别率的提高, 其中对大姿态的提高最显著。

(2) 在近红外-可见光库(CASIA NIR-VIS 2.0)上取得了目前最高的 99.94%的识别率。NIR(近红外图片)和 VIS(可见光图片)来自不同的 domain, 通常大家认为应该使用 domain adaptatino 技术来 NIR-VIS 识别。但文献的训练集全部来自 VIS, 这说明只要训练集足够大(文献用了 7M 可见光图片), domain shift 可以自然被解决, 因为大的 VIS 训练集能抓住足够多的细节, 这些细节 NIR 和 VIS 是共享的。

(3) 在 LFW 上文献取得了 99.65% 的识别率, 超过了 google 的 facenet (99.63%)。

四、和其他方法的比较作者的方

法有几个明显的优势:

(1) 高阶非线性。这使得该方法相比较去一些线性的聚合方法而言, 具备了更加强的表达能力, 另外由于采用了低秩的模型, 使得该方法可以具备相对少的参数和较好的过拟合鲁棒性。

(2) 可扩展性。由于作者在张亮和 gated 的人工神经网络进行了等价, 使得该方法可以通过 mini-batch SGD-based learning 扩展到大数据。

(3) 可移植性。作者 GTNN 天然的可以延伸到更深的网络架构, 因此该方法可以转换成一个端到端的框架。