机器学习实验报告

————图像识别与分割

姓名：林聪　　学号：16339026　　日期：2019/5/25

摘要： 对图像中的特定目标进行识别与分割在工厂车间等实际生产环境中有着非常重要的作用，该任务通常要求后端系统在无人参与的情况下自动分析前端拍摄设备回传的图像，从而识别生产环境中的异常情况并进行定位。具体而言，本次任务聚焦于对生产环境图像中的人手进行识别与分割。对于识别任务，我们根据训练数据的特性采用了 KNN 分类模型，该方法在验证过程及测试过程中获得了良好的效果。对于分割任务，我们采用了基于 Autoencoder 框架的分割网络模型 SegNet，该方法在训练过程中得到了很好的 Dice 指标。

1. 导言

a) 问题背景

图像识别与分割是图像处理领域研究最多的课题之一，相关技术在生产实践，例如医学图像分析、卫星遥感成像、无人汽车的街景识别等方面有着广泛而重要的应用。从研究目标角度而言，图像识别与分割的意图是根据观测到的图像，分辨其中物体的类别，并标记出特定物体在图像中的分布轮廓。这一过程的数学本质是一个映射问题，其中图像识别要求构建某种从模式空间到类别空间的映射，从而将图像内容的分布转化为一个或若干个分类属性，而图像分割要求构建某种从模式空间到带类别信息的模式空间的映射，从而将图像内容的分布转化为一个或若干个分类属性的分布。

b) 相关方法

从图像识别提取的特征对象来看，可以将图像识别方法划分为基于形状特征的识别技术，基于色彩特征的识别技术，以及基于纹理特征的识别技术。而从特征的选择及判别决策方法的角度来看，则可以将图像识别方法大致归纳为统计模式识别方法、句法模式识别方法、模糊模式识别方法和神经网络模式识别四类。在这里我们比较关注统计模式识别方法，具体而言它可以细分为直接优化代价函数的几何分类法和和基于密度估计的概率密度法，前者以距离分类法 KNN、判别函数法 SVM 等为代表，后者则以贝叶斯方法为代表。

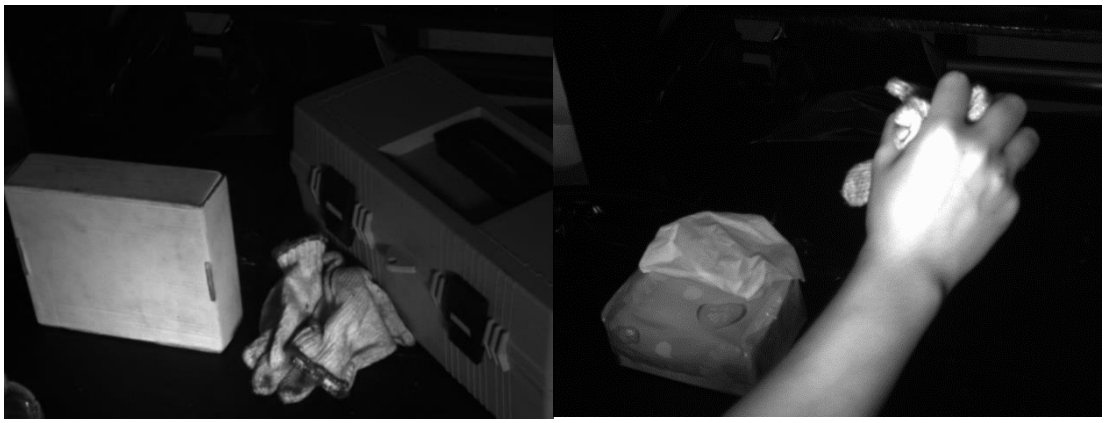
图像分割领域我们则着重关注近年来兴盛的图像语义分割神经网络，其思想在于对图像进行像素级分类。该领域的开山之作为全卷积网络 Fully Convolutional Network (FCN)，该框架将传统用于图像分类的深度神经网络中的全连接层替换为了卷积网络，同时引入了反卷积层，使得图像在被映射为较小的特征图后能够被还原回原始大小，从而能进行像素级分类。

在此基础上衍生出来的代表性框架有 DeepLab、U-Net、SegNet 等等。这其中 DeepLab 为了解决 FCN 的噪声问题引入了带洞卷积，并且使得卷积核能够掌握到更多图像信息；U-Net 则是一个与 FCN 非常形似的框架，其突破点在于将 FCN 的特征融合模式——相加，改进为拼接，使得网络中间层能够保留更多特征信息，它的训练速度通常也因此比 FCN 更快；而 SegNet 是一个 Autoencoder 框架的网络，它将 FCN 中的反卷积操作全部替换为带 pooling indice 的反池化操作，网络的结构简单明了，训练也相对容易。

c) 实验分析

本次实验的要求对工厂实际生产环境中采集的真实图片数据进行以人手为目标的图像识别与分割作业，其中人手识别任务要求使用非深度学习方法，人手分割任务要求给出原图像分辨率的二值掩模图像。

所给的训练数据集为以黑色为背景色的灰度图像，共 4384 张，分辨率为 480*640，占用空间 1.27GB，用于分割任务的掩模 Ground Truth 图像并未提供。大部分图像为操作工作台的图像，其中包含人手的图像有 3358 张，其余 1026 张不包含人手。图像中的人手可能手持诸如瓶子、盒子之类的物品，也可能空手张开、握拳或其它手势，手背、手面及手侧都可能为拍摄角度，通常部分手臂也会入镜，手臂可能有衣物覆盖。此外手可能从图像的任一方向伸出，光照也有强有弱。除了人手之外，图片中可能有工件、瓶子、手套等其它物品入镜，位置和光照强度不一。下图为随机选取的无人手图和有人手图各一张：



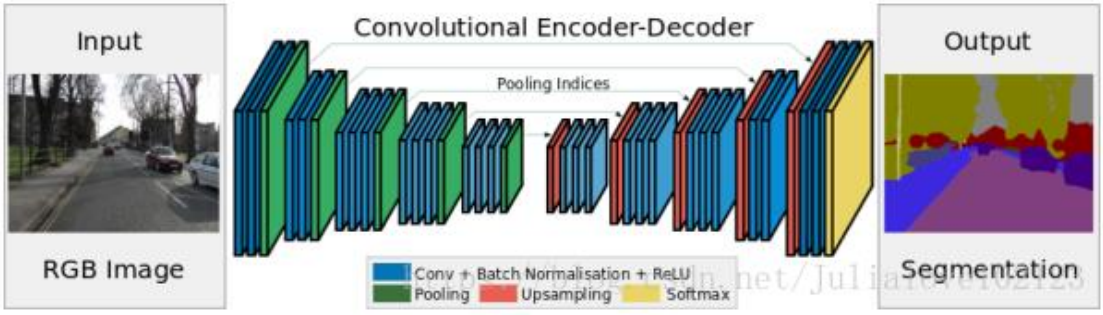
值得注意的是，虽然图像内容纷繁，但是其模式有迹可循。通常模式分布相近的图片会出现数次，并且以下图为例，大致相同的环境中入镜方向、拍摄角度及光照强度不同的图片数据也会多次出现：

d) 模型选择

基于相关方法的研究以及实验数据的分析，本次实验中为各任务选择的核心模型如下：

对于人手识别任务，我们采用统计模式识别中从属于几何分类法的距离分类法 K 最近邻分类算法（KNN），原因在于其简洁朴素，而且鉴于图像数据的模式分布较为稳定，各种模式实例虽然称不上多，但也能够反映模式的特征，故有理由相信 KNN 能够有很好的效果。

对于人手分割任务，综合实现以及训练的难度，我们采用基于 Autoencoder 框架的语义分割网络 SegNet 进行图像人手分割，其结构如下：



2．实验过程

a) 人手识别

实验环境：Windows 10 系统上的 Python 3.6.4，主要的库有 numpy、scipy、sklearn；

- 预处理：基于对 KNN 算法时间复杂度和空间复杂度的考虑，我们先对图像进行降采样，将图像分辨率由原先的 480*640 降低至 80*80，采样率为(1/6, 1/8)，对这一超参数我们不做探讨。基于对图像强度不一的考虑，我们先逐图像地对图像像素值进行归一化，而后选择采用 gamma 矫正或直方图均衡对图像进行增强；
- 模型训练：KNN 算法没有显式的训练过程，我们每次将 200 个样本从训练集中除去并划为验证集，在固定的验证集上对超参数邻近数 k 以及投票权重（全取 1 或取距离的倒数）进行调优；

b) 人手分割

实验环境：Google Colab, Ubuntu 18.04.2, 16GB Tesla T4, Python 3.6.7, 主要的库为 tensorflow 1.13.1, keras 2.2.4；

- 数据准备：利用 Labelme 3.14.1 (Windows 10, Python 3.6.4)对原始数据中的一部分进行人工掩模标注，共得到 2292 张掩模图像，其中约 1400 张图像对应有人手的原图，其余约 800 张图像为对应无人手图像的全黑图；
- 预处理：由于本次实验我们将 SegNet 部署在 Google 的服务器 Colab 上借助 GPU 进行训练，考虑到神经网络训练的难度以及数据上传的速度，我们对 2292 张原始图像、2292 张掩模图像以及 200 张测试图像进行降采样，将图像分辨率由原先的 480*640 降低至 256*256，并且对真实的灰度图像做归一化处理，将掩模图像背景值设为 0、掩模值设为 1。最终得到训练图像文件 1.11GB，训练掩模文件 143MB，测试图像文件 100MB；
- 模型构建：本次实验的 SegNet 中的编码器和解码器分别包含 5 个模块，其中编码器将输入大小为(256,256,1)的图像压缩为(8,8,512)的特征图，解码器则对称地将其还原成(256,256,2)的掩模图像，详情参见代码中的实现或给出的 summary()；
- 模型训练：模型采用多标签交叉熵损失函数，优化器为 Adadelta 或 ADAM（按照建议地采用默认学习率），batch 为 16，epoch 分别取 1、2、5、10，评价指标为重合率 Dice；

3．结果分析

a) 人手识别

进行 gamma = 1 的 gamma 矫正后在验证集上所得的指标（五次最低/平均/最高）：

Weights	n = 1	n = 3	n = 5
Uniform	0.935/0.947/0.970	0.920/0.928/0.935	0.915/0.915/0.916
1/Distance	0.935/0.947/0.970	0.925/0.940/0.950	0.920/0.928/0.940

进行 gamma = 1.25 的 gamma 矫正后在验证集上所得的指标（五次最低/平均/最高）：

Weights	n = 1	n = 3	n = 5
Uniform	0.935/0.948/0.965	0.895/0.916/0.930	0.915/0.922/0.930
1/Distance	0.935/0.948/0.965	0.930/0.932/0.935	0.925/0.937/0.945

进行 gamma = 0.75 的 gamma 矫正后在验证集上所得的指标（五次最低/平均/最高）：

Weights	n = 1	n = 3	n = 5
Uniform	0.955/0.958/0.960	0.910/0.918/0.925	0.920/0.930/0.940
1/Distance	0.955/0.958/0.960	0.920/0.937/0.940	0.935/0.947/0.960

进行直方图均衡后后在验证集上所得的指标（五次最低/平均/最高）：

Weights	n = 1	n = 3	n = 5
Uniform	0.970/0.982/0.990	0.930/0.945/0.960	0.930/0.940/0.950
1/Distance	0.970/0.982/0.990	0.960/0.968/0.975	0.955/0.960/0.965

根据以上结果，理论上应该先对训练集和测试集进行直方图均衡之后进行分类，但在实际测试过程中发现直方图均衡处理过后的分类效果并不乐观，最终发现反倒是验证效果一般的 gamma = 1.25, n = 5, Weights = Uniform 的 KNN 效果看上去比较好，故将之作为测试结果，具体分类结果参看附件。

b) 人手分割

重合率 Dice 与优化算法及训练轮数的关系如下：

	epoch = 1	epoch = 2	epoch = 5	epoch = 10
Adadelta	80.32%	93.71%	96.64%	98.91%
ADAM	87.91%	92.62%	95.34%	96.64%

实际测试过程中发现用 Adadelta 训练 10 个 epoch 所得结果较好，并未出现过拟合的情况，但部分结果在带阴影的边缘处比较破碎。相比之下给阶段的 ADAM 训练结果或是过拟合，或是欠拟合，均不太理想。故最终取 Adadelta 在 10 个 epoch 的训练后所获得的模型，以及其对应的结果，模型参数及掩模结果见附件。

4．结论

本次实验事实上完成了一次比较完整的图像处理工业生产过程。从获取数据、标记数据。数据处理、模型建立、模型训练、生成结果再到检验评估，这中间的每一个阶段在本次实验中都有体现，可以说本次实验试一次非常好的实践过程。不仅如此，这次实验既涉及到了图像识别，也涉及到了图像分割，它对知识储备、学习能力和实践能力显然是有一定要求的。我在准备这次实验的过程中就收获了很多新的知识和思想，并且将模型思想转化为实际代码的能力也得到了锻炼。

这样有意义的学习实践过程自然伴随着很多可改进的问题。在这次实验中我所获得的结果并不太好，预处理的方式比较简单，图像分类模型的鲁棒性并不太好，其表达能力也有限。应当采用更有效的预处理方法提取特征，并建立表达能力更好的模型，过度依赖数据本身会降低模型的泛化能力和鲁棒性。图像分割模型在实现过程中绕了很多弯路，最终结果才有所改观，但不知道问题出在哪，掩模结果仍停滞在了仍有很大提升空间的阶段。主要问题可能出在预处理和优化器上，有待进一步探讨。这些不足之处启示我去继续学习与尝试机器学习，去探索不同的方法和技巧，

主要参考文献：

[1] Keras Documentation
https://keras.io/
[2] Vijay Badrinarayanan, Alex Kendall and Roberto Cipolla "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation." PAMI, 2017.
[3] 【深度学习】语义分割网络介绍对比-FCN，SegNet，U-net DeconvNet
https://blog.csdn.net/qq_34106574/article/details/82453615
[4] 图像分割综述【深度学习方法】
https://blog.csdn.net/weixin_41923961/article/details/80946586
[5] 图像模式识别的方法
https://blog.csdn.net/gdut2015go/article/details/46762323