|  |  |
| --- | --- |
| dlib库人脸检测算法解构报告 | 文件类型：技术文档 |
| 文件编号： |
| 面向的部门： 研发中心 |
| 保密等级：高 |
| 作者： 刘凯 |
| 日期：2018-09-6 |
| 版本：1.0V |

dlib库人脸检测算法解构报告



目 录

[1 dlib库人脸检测算法 3](#_Toc524097031)

[1.1 算法框架 3](#_Toc524097032)

[1.2 候选框选择 3](#_Toc524097033)

[1.3 特征提取 4](#_Toc524097034)

[1.4 NMS 4](#_Toc524097035)

[1.5 MMOD训练 6](#_Toc524097036)

[2 算法实现 8](#_Toc524097037)

[参考文献 1](#_Toc524097038)

# dlib库人脸检测算法

* 1. 算法框架

dlib库是一个开源的机器学习库，采用C++编写，并提供python接口。dlib库里也有人脸检测的实现，本文将对dlib库里的人脸识别算法进行解构。算法的流程图如图1所示。其中，候选框选择是从图像中选择一组人脸图像的候选框，选择算法在1.2节介绍；特征提取是计算候选框图像的特征，具体参考1.3节；算法的核心步骤是用最大边缘目标检测(MMOD, Maximum Margin Object Detection)[1]来训练打分函数的参数，具体的算法解析参考1.5节；非极大抑制(NMS, Non-Maximum Suppression)是目标检测中常用的去除重复框的算法，将在1.4节介绍。

图片输入

候选框选择

特征提取

MMOD

NMS

画图输出

图1：dlib库人脸检测算法流程图

* 1. 候选框选择

dlib中是采用固定大小的人脸框在图片上滑动遍历得到所有的候选框，对图片做图像金字塔(image pyramid)则可以得到不同尺度的候选框。在dlib库中，图像金字塔参数为6，即每次的缩放因子为5/6，最小人脸框为。

* 1. 特征提取

文献[1]中的特征提取采用的是HOG特征[2]，dlib库里也有基于HOG特征的人脸检测类，但随着CNN(Convolutional Neural Networks)在人脸检测和识别中的成功应用，dlib库之后更新了基于CNN特征提取的人脸检测类，本文主要解析基于CNN的特征提取。

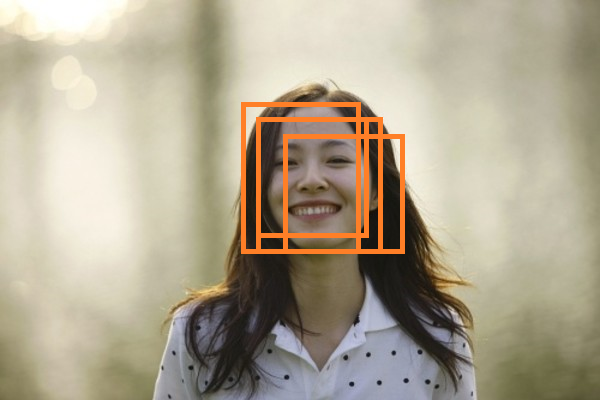
CNN特征提取网络包括7个卷积层，具体的参数维度如表1所示。其中前三个卷积层均采用的卷积核，且步幅Stride=2，相当于对图像进行2倍下采样，经过三次2倍下采样，输入图片一共达到8倍下采样。第4个到第6个卷积层均采用的卷积核，步幅Stride=1，无下采样。前面六个卷积层出来的结果均要经过bn层(Batch Normalization)和relu层(Rectified Linear Unit)。最后一个卷积层只有一个通道，该输出越大，表面此处有人脸的可能性越大，实际上这一层卷积相当于对人脸特征进行打分，也是MMOD算法要训练的参数。

表1：CNN特征提取网络参数

|  |  |  |
| --- | --- | --- |
| 层 | 权重 | Stride |
| **Conv1** |  | 2 |
| **Conv2** |  | 2 |
| **Conv3** |  | 2 |
| **Conv4** |  | 1 |
| **Conv5** |  | 1 |
| **Conv6** |  | 1 |
| **Conv7** |  | 1 |

* 1. NMS

NMS算法主要用于目标检测中，消除交叉重复的窗口，找到最佳物体检测位置，如图2所示。NMS的算法流程如图3所示。可以看出NMS算法实际上属于贪婪算法，本质是搜索局部极大，抑制非极大值，因此在一些情况下，NMS可能无法得到正确的值，如图4所示，当正确的结果是图4两边得分为6的两个框时，NMS算法可能会搜索失败。因此1.5节的MMOD算法是为了避免这种情况提出的一种打分参数的学习算法。



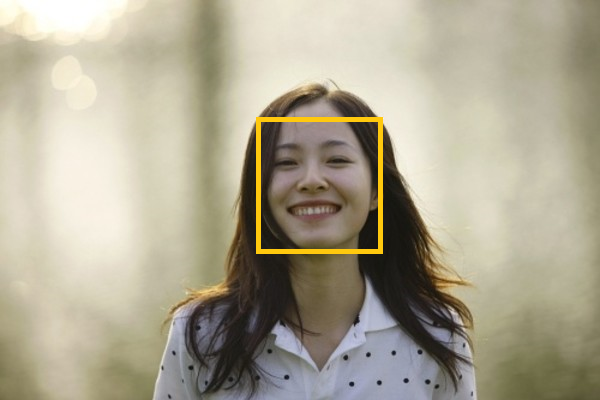


图2：NMS消除冗余的框



图3：NMS算法流程



图4：NMS搜索失败示例

* 1. MMOD训练

从1.4节知道，NMS有可能无法搜索到最优的候选框，本节的MMOD算法旨在训练一组打分参数(scoring parameters)，以提高NMS的性能。

设打分函数为：



其中x为图片，r为图片x中的一个矩形框，为根据x和r计算的特征矢量，可以用HOG特征或CNN特征，为待学习的打分参数。于是目标检测过程可以表示为：



其中，y为一组“互不交叠”的矩形框组成的集合，为所有的集合y组成的全集。这里的“互不交叠”是指两个矩形框和满足：



设：



则式等价于：



表示选取所有“互不交叠”的框中得分最高的一组。

现有一组图片和对应标号作为打分参数的训练数据，则MMOD优化问题为：



其中，为中间变量，这里加入是因为实际系统中存在噪声引入的软边界(soft margin)变量，当时，上述问题就变成硬边界(hard margin)的MMOD问题了。C为契合训练数据的参数，C越大，对训练数据的契合度越大，但越容易过拟合。为预测标号为，而真实标号为时的loss函数，即：



其中和分别表示高召回率(recall)或高精度(precision)的权重。

问题可以等价地转化为如下无约束问题：



其中：



注意，问题是关于的凸函数，可以用文献[3]的cutting plane method求解，其算法流程如图5所示。



图5：MMOD算法流程

# 算法实现

参考文献

1. Davis E King. Max-Margin Object Detection. arXiv:1502.00046, 2015.
2. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition (CVPR), 2005.
3. T. Joachims, T. Finley, and C. J. Yu. Cutting-plane training of structural svms. Machine Learning, 77(1):27–59, 2009.