**使用HoG+SVM进行目标检测**

Daniel ChenJH

Partner 1

**摘要**：本课题针对RAF-DB人脸数据集的部分子集，在生成负例后提取HOG特征，并利用SVM检测人脸存在性。接着对于存在人脸的候选框进行候选框筛选，即采用NMS、soft-NMS、WBF等算法输出对于人脸的最终检测结果。

**关键词**：人脸检测，HOG特征，SVM算法，NMS算法，WBF算法

# 1 引言

在日常生活中，我们常常会面临需要利用采集的摄像头进行人脸检测的场景，如智能家居、监控场所等。对人脸的准确识别可以使管理者及时得知环境中的人物存在信息，以便做出下一步决策。在我们的实验中，我们取RAF-DB人脸数据集的部分子集，在生成负例后提取HOG特征，并利用SVM检测人脸存在性。接着对于存在人脸的候选框进行候选框筛选，即采用NMS、soft-NMS、WBF等算法输出对于人脸的最终检测结果。

# 2 相关理论介绍

## 2.1 方向梯度直方图HOG

### 2.1.1 HOG特征的简介

方向梯度直方图（Histogram of Oriented Gradient, HOG）特征是一种在计算机视觉和图像处理中用来进行物体检测的特征描述子。HOG特征通过计算和统计图像局部区域的梯度方向直方图来构成特征。

HOG的核心思想是所检测的局部物体外形能够被光强梯度或边缘方向的分布所描述。通过将整幅图像分割成小的连接区域（称为cells），每个cell生成一个方向梯度直方图或者cell中pixel的边缘方向，这些直方图的组合可表示出（所检测目标的目标）描述子。为改善准确率，局部直方图可以通过计算图像中一个较大区域(称为block)的光强作为measure被对比标准化，然后用这个值(measure)归一化这个block中的所有cells.这个归一化过程完成了更好的照射/阴影不变性。与其他描述子相比，HOG得到的描述子保持了几何和光学转化不变性（除非物体方向改变）。因此HOG描述子尤其适合人的检测。

### 2.1.2 HOG特征的实现方法

首先将图像分成小的连通区域，这些连通区域被叫做细胞单元。然后采集细胞单元中各像素点的梯度的或边缘的方向直方图。最后把这些直方图组合起来，就可以构成特征描述符。

将局部直方图在图像的更大的范围内（叫做区间）进行对比度归一化，可以提高该算法的性能，所采用的方法是：先计算各直方图在这个区间中的密度，然后根据这个密度对区间中的各个细胞单元做归一化。通过这个归一化后，能对光照变化和阴影获得更好的效果。

如图2.1.2所示，黑色表示窗口的划分，蓝色表示块的划分，黄色表示细胞的划分。其中每个cell静止不动计算每个像素的方向梯度，block和window滑窗移动，实现不同区域的检测。通过滑窗，将提取到的HOG特征首尾相连，形成一个大的一维向量，就是最后得到的特征向量。

粉色头发的女人

描述已自动生成

图2.1.2 HOG划分示意图（黑色表示窗口的划分，蓝色表示块的划分，黄色表示细胞的划分）

### 2.1.3 HOG的优点

1.核心思想是所检测的局部物体外形能够被梯度或边缘方向的分布所描述，HOG能较好地捕捉局部形状信息，对几何和光学变化都有很好的不变性；

2.HOG是在密集采样的图像块中求取的，在计算得到的HOG特征向量中隐含了该块与检测窗口之间的空间位置关系。

### 2.1.4 HOG的缺点

1.很难处理遮挡问题，人体姿势动作幅度过大或物体方向改变也不易检测（这个问题后来在DPM中采用可变形部件模型的方法得到了改善）；

2. HOG没有选取主方向，也没有旋转梯度方向直方图，因而本身不具有旋转不变性（较大的方向变化），其旋转不变性是通过采用不同旋转方向的训练样本来实现的；

3. HOG本身不具有尺度不变性，其尺度不变性是通过缩放检测窗口图像的大小来实现的；

4.由于梯度的性质，HOG对噪点相当敏感，在实际应用中，在Block和Cell划分之后，对于得到各个像区域中，有时候还会做一次高斯平滑去除噪点。

## 2.2 SVM

### 2.2.1 SVM的简介

SVM，全称是support vector machine，中文名叫支持向量机。SVM是一个面向数据的分类算法，它的目标是为确定一个分类超平面，从而将不同的数据分隔开。

支持向量机的基本模型是在特征空间上找到最佳的分离超平面使得训练集上正负样本间隔最大。SVM是用来解决二分类问题的有监督学习算法（实际上还有多分类），在引入了核方法之后SVM也可以用来解决非线性问题。它的基本学习策略是间隔最大化，可形式化一个求解凸二次优化问题，他的学习算法是求解凸二次优化的最优化问题，基本思想是能够正确划分训练数据集且几何间隔最大化的分离超平面。在决定分离超平面时只有支持向量（sv）起作用，其它实例点并不起作用，线性可分情况下，训练集的样本点与分离超平面最近的点的实例称为支持向量（sv）。

如图2.2.1所示，SVM的算法是首先假设数据在高维特征空间中大致线性可分，然后在此空间中寻找一个超平面，也可以理解为一个最优的决策边界。这个边界使得距离两个类别最近的样本最远，即最大化支撑向量直接的margin。SVM主要应用于二分类问题，在本实验中用于判断滑窗中是否存在有效人脸。

一般SVM有下面三种：

硬间隔支持向量机（线性可分支持向量机）：当训练数据线性可分时，可通过硬间隔最大化学得一个线性可分支持向量机。

软间隔支持向量机：当训练数据近似线性可分时，可通过软间隔最大化学得一个线性支持向量机。

非线性支持向量机：当训练数据线性不可分时，可通过核方法以及软间隔最大化学得一个非线性支持向量机。

图表, 散点图

描述已自动生成

图2.2.1支持向量机的原理示意图

### 2.2.2 SVM的优点

SVM 是一种有坚实理论基础的新颖的小样本学习方法。它基本上不涉及概率测度及大数定律等,因此不同于现有的统计方法。从本质上看,它避开了从归纳到演绎的传统过程,实现了高效的从训练样本到预报样本的“转导推理”,大大简化了通常的分类和回归等问题。

SVM 的最终决策函数只由少数的支持向量所确定,计算的复杂性取决于支持向量的数目,而不是样本空间的维数,这在某种意义上避免了“维数灾难”。

此外，依靠少数支持向量决定了最终结果,这不但可以帮助我们抓住关键样本、“剔除”大量冗余样本,而且注定了该方法不但算法简单,而且具有较好的鲁棒性。这种鲁棒性主要体现在：

1. 增加或删减非支持向量样本对模型没有影响;
2. 支持向量样本集具有一定的鲁棒性;
3. 在已有的某些成功应用中，SVM 方法对核的选取不敏感。

### 2.2.3 SVM的缺点

SVM算法主要存在两大缺点。

一是SVM对大规模训练样本难以实施。由于SVM是借助二次规划来求解支持向量，而求解二次规划将涉及m阶矩阵的计算（m为样本的个数），当m数目很大时该矩阵的存储和计算将耗费大量的机器内存和运算时间。

二是用SVM解决多分类问题存在困难。经典的支持向量机算法只给出了二类分类的算法，而在数据挖掘的实际应用中，一般要解决多类的分类问题。可以通过多个二类支持向量机的组合来解决。主要有一对多组合模式、一对一组合模式和SVM决策树；再就是通过构造多个分类器的组合来解决。主要原理是克服SVM固有的缺点，结合其他算法的优势，解决多类问题的分类精度。如：与粗集理论结合，形成一种优势互补的多类问题的组合分类器。

## 2.3 候选框筛选

在完成HOG特征提取、SVM人脸检测推理的过程后，我们得到了很多的候选检测框，每个检测框都有一个分数score值代表置信度。然而，这里面很多检测框都是检测同一个目标，导致候选框之间的重叠部分过多，这时我们必须对众多的候选框进行筛选或合并等操作，从而输出相对准确的预测结果。

在本次实验中，我们分别采用MSC、NMS、soft-NMS、WBF四种方法对候选框进行筛选，并对其结果进行比较分析。下面将对各方法进行详细介绍说明。

图片包含 户外, 人, 草, 体育

描述已自动生成 人拿着手机

低可信度描述已自动生成

（a）筛选前的效果 （b）使用筛选后的效果

图2.3 候选框筛选的作用示意图

### 2.3.1 最大分数筛选（MSC）

Max Score Choosing，即MSC，该方法对于每张图片的所有候选检测框的得分进行排序，取分数最高的一个框作为该图片的检测结果输出，抛弃其余所有检测框。

可以预想到的是，无论图片中有几张人脸，此方法检测出的人脸数量都不会超过1个。

### 2.3.2 交并比（IoU）

交并比，即两个框之间交集的面积与并集的面积之比，用于描述两个框之间的重叠关系。在判断过程中，我们会对交并比设置一个阈值，如果交并比的值大于这个阈值，我们就认为这两个框是重叠的。式(1)给出了交并比的计算方法。

其中，area of overlap表示两个框相交部分的面积；area of union表示两个框并集的面积，这个面积可以通过计算两个框的面积之和再减去两个框交集的面积得到。

### 2.3.3 非极大值抑制（NMS）

在计算机视觉目标检测的任务中，模型总是会依据一些先验知识对物体的大小进行一些假设，之后，再在图片上把这些假设的框“铺设”出来，进行回归任务，在最终的模型输出上，就会回归出大量的、重叠的框，而它们往往在标注着同一个物体。非极大值抑制(NMS)算法的作用就是去除这些重复、冗余的框。其作用效果示意图如图2.3.3所示。



（a）NMS前的效果 （b）使用NMS后的效果

图2.3.3 使用NMS前后效果对比图

对于传统的NMS算法，其往往通过计算交并比的方式来判断两个框是否重叠。如果重叠则保留置信度较高的一个，删除置信度较低的一个，其具体做法如式(2)所示。

其中，表示这个框的置信度，置信度为0表示删除这个框。

NMS的具体的实现步骤如下：

(1)设定目标框的置信度阈值，常用的阈值是0.5左右；

(2)根据置信度降序排列候选框列表；

(3)选取置信度最高的框A添加到输出列表，并将其从候选框列表中删除；

(4)计算A与候选框列表中的所有框的IOU值，删除大于阈值的候选框；

(5)重复上述过程，直到候选框列表为空，返回输出列表。

### 2.3.4 软非极大值抑制（soft-NMS）

在上述对NMS的介绍中不难发现，经典NMS算法存在着一些问题：对于重叠物体无法很好的检测。当图像中存在两个重叠度很高的物体时，经典NMS会过滤掉其中置信度较低的一个。经典NMS算法的做法是直接删除Iou大于阈值的Bounding box；而Soft-NMS则是使用一个基于IoU的衰减函数（一般为高斯函数），IoU越大，衰减程度越大。直到置信度经过几次衰减变为0后，再删除置信度很低的候选框。这种方法的好处是可以充分利用各候选框的分数与IoU值之间的关系，而不是一次性使用之后就直接丢弃这种关系信息。式(3)给出了使用高斯函数减小置信度的计算式。

其中，表示这个识别框的置信度，表示衰减系数，M表示置信度最高的框，表示与M有交集的框，表示二者的交并比。

NMS和soft-NMS的伪代码如图2.3.4所示。

文本, 信件

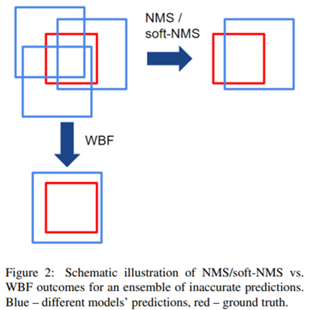
描述已自动生成 

图2.3.4 NMS与Soft-NMS的伪代码 图2.3.5 WBF与NMS的选择特征示意图

### 2.3.5 加权框融合（WBF）

无论是NMS还是soft-NMS，都是在已有的候选框列表中挑选某些框作为输出。但这种思路没有考虑到候选框密度与人脸存在可能性之间的关系。加权框融合Weighted Boxes Fusion，即WBF，认为最终的输出框应该更多地考虑候选框列表中有效框对于输出框的影响，即最终的输出应该由候选框列表中每个有效框根据置信度值加权形成。WBF算法的流程如下：

假设有N个模型。

List B：每个模型的预测box都放到该list里，按可信度分数C降序排序。

List L：box的簇（cluster）的集合；List F： 融合（fused）box的集合。在 list L的每一个位置包含一个box集合（也可能只有一个box），其形成一个簇（cluster）。在List F的每个位置只有一个融合box，其对应着L中的簇（cluster）。

循环迭代List B，尝试在F中找到一个匹配的box。如果IoU阈值超过某个值就认为找到匹配的。

如果没匹配到，就将来自B的这个box添加到list L和List F的末端。接着处理B中下一个box。

如果匹配到了，就将来自B的这个box添加到list L对应的位置的簇（cluster）。

list L对应的位置的簇（cluster）重新计算list F中box的坐标和可信度得分，簇（cluster）中有T个box。计算式如式(4)，(5)和（6）所示：

如公式(4)，(5)和（6）所述，拥有较大可信度得分C的box对融合（fused）box的贡献度更大。

当B中的所有box都处理完后，重新缩放list F中的可信度得分C。乘以簇（cluster）中box的数量T，除以模型数量N。如果一个簇（cluster）中的box数量降低，可能意味着只有很少的模型预测到它，所以我们需要降低其可信度得分。WBF最终输出List B及可信度得分列表C作为框选结果。得分C的计算式由式(7)给出。

# 3 实现效果

此处只展示部分关键代码及实现效果。完整代码见附录/附件。

## 3.1 构建负例数据集

本次实验使用的数据集是RAF-DB的子集，共3954张图片。RAF-DB提供了原图、人脸裁剪图、检测边界框，但无负例图片，因此需要我们根据原图和边界框构建非人脸图片。

我们采用的构建策略是：在原图的人脸框左侧/上侧框选与人脸框相同大小的框作为负例图片并保存。如果图像尺寸受限，则在整个图像的左上角框选与人脸框相同大小的框作为负例图片并保存。负例图片效果如图3.1所示。

图片包含 PowerPoint

描述已自动生成

图3.1 负例图片效果示意图

## 3.2 提取特征并训练模型

此处我们使用skimage库中的hog函数对3954张正、负例图像分别提取hog特征，并将提取到的特征向量存储为pkl文件，完成图片特征的提取。

接着我们调用sklearn库的SVC类，我们分别使用了‘linear’、‘sigmoid’、‘poly’和‘rbf’核函数进行准确率的测试，得到的结果如表3.2所示。

由于‘poly’的准确率最好，最终我们选取‘poly’作为SVM的核函数，结合前一步提取的hog特征，训练SVM分类器检测人脸，并在测试集上验证分类器的效果，最后将训练好的SVM模型存储为SVM.model文件，便于后续预测使用。

表3.2 不同核函数及其准确率

|  |  |  |  |
| --- | --- | --- | --- |
| 核函数 | 测试数量 | 检测正确的数量 | 准确率 |
| ‘linear’ | 1584 | 1560 | 0.9848 |
| ‘sigmoid’ | 1584 | 1565 | 0.9880 |
| ‘poly’ | 1584 | 1572 | 0.9924 |
| ‘rbf’ | 1584 | 1570 | 0.9912 |

## 3.3检测人脸并输出

对于每张输入图片，通过滑窗式地截取部分图片、提取特征后输入训练好的SVM分类器，即可得到滑窗中是否存在人脸信息。我们将存在人脸信息的滑窗位置保存作为候选列表，然后对这些候选框列表进行2.3中介绍的候选框筛选。完成筛选后在原图上标记出人脸框定结果并保存到本地。

### 3.3.1 最大分数筛选（MSC）

MSC得到效果如图3.3.1所示。图中可见，模型可以较好、较准确地框定出一个人脸，该模型对于单个人脸的识别效果较好，但无法处理一张图片中存在多个人脸的情况。

为了更加清晰地反映出MSC算法的缺陷，选取了有着多张人脸的照片作为展示，如图3.3.2-2(a)所示，可见即使照片中有着两张人脸，MSC算法只能识别出置信度最高的一张人脸而不会识别出其他的人脸。

图片包含 游戏机, 建筑, 公寓

描述已自动生成

图3.3.1 MSC算法识别结果示意图

### 3.3.2 非极大值抑制（NMS）

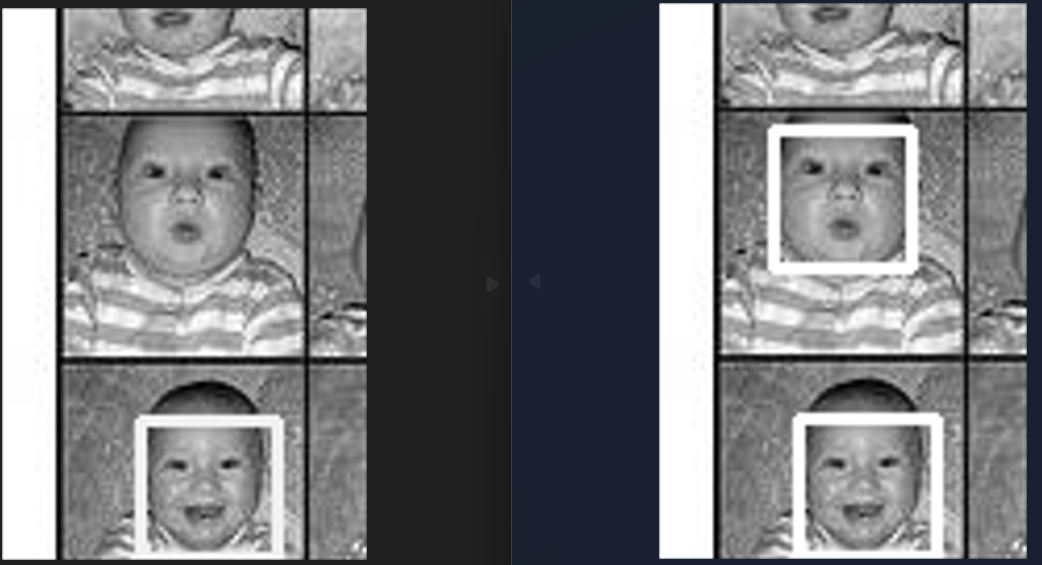
NMS得到的结果如图3.3.2-1所示。图中可见模型可以完成多个人脸的识别，但虚警率较高，主要体现在容易将得分相对较低的人体手臂识别为人脸，或者对于同一张人脸框定多个输出框。

图片包含 游戏机

描述已自动生成

图3.3.2-1 使用NMS算法识别结果示意图

为了更加清晰地反映出NMS算法的效果，选取了有着多张人脸的照片作为展示，如图3.3.2-2(b)所示，可见照片中有着两张人脸，且NMS算法能够正确地识别出这两张人脸。但同时，当一张照片中的两张脸有遮挡或重叠时，NMS很容易将两张脸的识别框识别为重复框进行，进而导致删除了置信度较小的识别框，如图3.3.3-2(a)所示。



(a)使用MSC算法的识别结果 (b)使用NMS算法的识别结果

图3.3.2-2 使用MSC算法和NMS算法识别结果对比图

### 3.3.3 软非极大值抑制（soft-NMS）

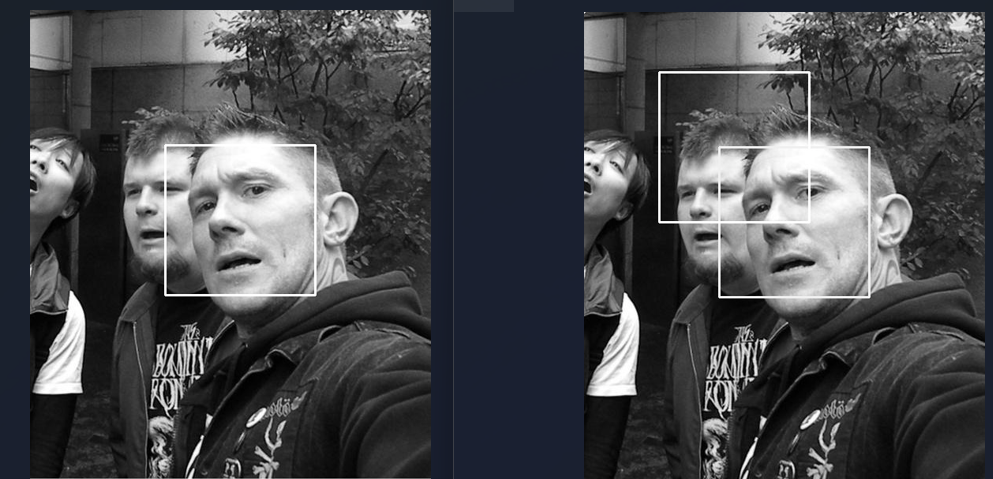
Soft-NMS得到的结果如图3.3.3-1所示。图中可见模型可以完成多个人脸识别，相对于NMS算法来说，Soft-NMS算法确实能够识别出更多重叠部分的人脸，但同时由于仅仅只是对重叠部分的处理方法进行了改进，因此并没有降低虚警率，同时，由于计算方法的改变与对部分图像的保留，Soft-NMS方法的虚警率甚至略高于NMS算法。

建筑的摆设布局

低可信度描述已自动生成

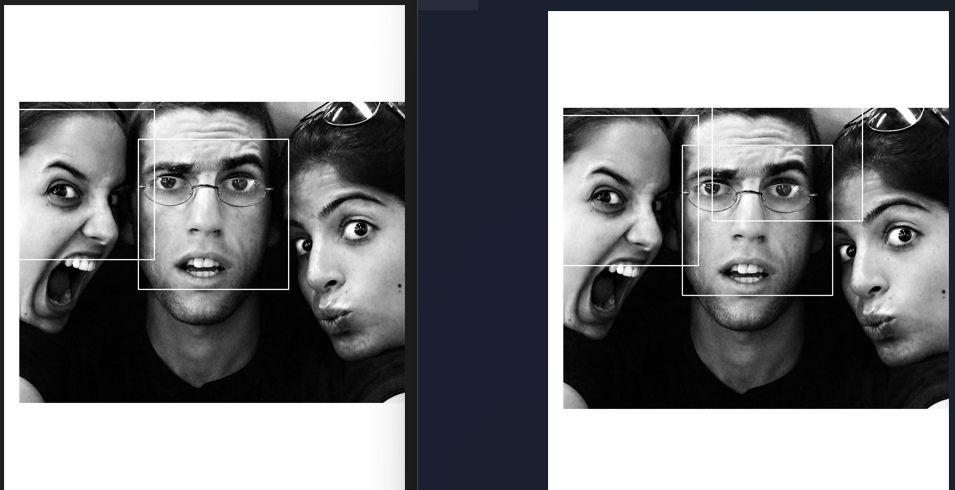
图3.3.3-1 使用Soft-NMS算法识别结果示意图

为了更加清晰地反映出Soft-NMS算法的效果，此处选择了一张有人脸遮挡人脸特征的照片与没有人脸遮挡人脸特征的照片，可以得到如图3.3.3-2和图3.3.3-3所示的结果。可以发现，在有遮挡的情况下，soft-NMS确实能够有效识别出被遮挡部分的人脸；但在没有遮挡或者人脸方向不一致时，由于对重叠框的权重的更改后再进行删除，Soft-NMS算法的可能会保留一些意想不到的特征框。如图3.3.3-3所示，Soft-NMS就将男子的右半边脸“误认”为一张人脸，这可能是选取框的大小太小导致的。



(a)使用NMS算法的识别结果（阈值0.3） (b)使用Soft-NMS算法的识别结果（阈值0.1）

图3.3.3-2 使用Soft-NMS算法和NMS算法识别结果对比图



(a)使用NMS算法的识别结果（阈值0.1） (b)使用Soft-NMS算法的识别结果（阈值0.3）

图3.3.3-3 使用Soft-NMS算法和NMS算法识别结果对比图

### 3.3.4 加权框融合（WBF）

WBF得到的结果如图3.3.4-1。图中可见模型可以完成多个人脸的识别，即使是侧脸的识别结果也相当不错，并在一定程度上解决了NMS模型的一个人脸多个输出框的问题。虚警率比NMS也有明显降低。整体来说WBF效果比NMS得到了较大提升。

然而我们也注意到，WBF对于某些图片的输出相对人脸的真实位置存在一定偏差。这种偏差是因为WBF的结果框并不是SVM滑窗的结果，而是多个滑窗的加权框。有时候这种加权会导致输出框位置偏离中心位置。

建筑的摆设布局

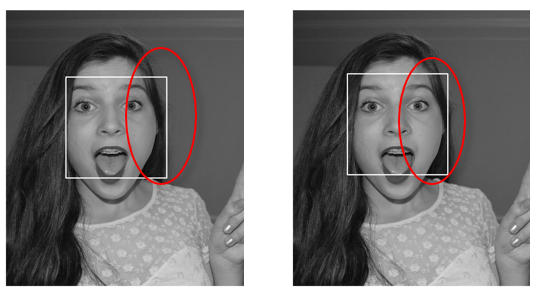
低可信度描述已自动生成

图3.3.4-1 使用WBF算法识别结果示意图

为了使效果更加明显，选择其中的几张照片作为对比，如图3.3.4-2所示。可以发现相比于NMS，WBF有更低的虚警率，且对于一个目标人脸也只框定出一个结果框。同时，图中WBF的结果框为候选框加权位移得到的，这也体现出了我们算法的有效性。



(a1)使用NMS算法处理后的图片效果 (a2) 使用WBF算法处理后的图片效果



(b1) 使用NMS算法处理后的图片效果 (b2) 使用WBF算法处理后的图片效果

图3.3.4-2 使用WBF算法(a2)(b2)与NMS算法(a1)(b1)效果对比

### 3.3.5 交并比比较

在本次实验中，各方法筛选候选框后与真值的交并比平均值如表3.3.5所示。在表中我们可以看出，由于MSC只取候选框中分数最大的作为结果输出，因此对SVM结果中分数的依赖性最强也最直接，反而导致了其交并比最低。NMS由于可以完成多个人脸的识别检测，因此交并比有所提高。Soft-NMS改善了传统NMS对于位置较近的人脸的漏检问题，交并比进一步提高。最后，WBF采用了全新的加权融合框的思路，对SVM中输出框的结果进行位置修正，同时还降低了虚警率，使得WBF输出结果与真值的交并比在0.7以上，为四种方法中最高。

受限于数据集大小以及SVM的效果，我们认为soft-NMS算法以及WBF算法都对候选框筛选工作做出了值得肯定的提升。

表3.3.5 不同筛选方法结果的交并比

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 筛选方法 | MSC | NMS | Soft-NMS | WBF |
| 交并比均值 | 0.612 | 0.652 | 0.683 | 0.704 |

# 4 结论

本课题较好的完成了利用HOG和SVM算法实现图像的人脸检测。在几种候选框筛选算法效果的对比方面，MSC算法简单高效，对于单个人脸的识别效果较好，但无法处理一张图片中存在多个人脸的情况；NMS算法可以完成多个人类的识别，但虚警率较高，无法通过筛选排除掉得分较低的人体手臂，也无法很好地处理多个目标时存在的遮挡问题；soft-NMS算法对NMS算法进行了优化，主要体现在对多目标遮挡识别的性能提升，但无法改善NMS算法存在的虚警率高的问题；WBF算法则是一种全新的思路，依据候选框得分对最终输出框的位置进行加权调整，能较好地完成多个人脸的有遮挡/侧脸识别，虚警率较低，但存在一定的输出框偏离问题。受限于数据集大小以及SVM的效果，我们认为soft-NMS算法以及WBF算法都对候选框筛选工作做出了值得肯定的提升。

# 5 总结

本课题较好的完成了利用HOG和SVM算法实现图像的人脸检测，提高了使用python解决实际工程问题的能力，使笔者深入了解了HOG和SVM等算法以及相关候选框筛选算法，如NMS与WBF算法等。

在解决问题的过程中，提高了我们的自学能力与知识迁移能力，学会了依靠查阅相关资料解决所面临的问题，提高了科研能力。

# 6 任务分工

共同完成：知识调研、报告撰写、PPT制作、汇报展示。

Daniel-ChenJH：负例生成实现、HOG特征提取、MSC候选框筛选实现、WBF候选框筛选实现。

Partner 1：SVM检测器实现与参数调整、NMS选框筛选实现、soft-NMS候选框筛选实现。

# 7 致谢

感谢徐老师向我们讲解图像处理与内容分析课程，介绍多种图像相关算法及其应用场景。感谢老师、助教的辛勤付出。感谢小组两位同学的合作。

# 8 参考文献

[1] Roman Solovyev, Weimin Wang, Tatiana Gabruseva, Weighted boxes fusion: Ensembling boxes from different object detection models, Image and Vision Computing, Volume 107, 2021, 104117, ISSN 0262-8856, https://doi.org/10.1016/j.imavis.2021.104117.

[2] Github, Weighted boxes fusion, https://github.com/ZFTurbo/Weighted-Boxes-Fusion

[3] CSDN社区论坛. Weighted Boxes Fusion 源码解析

https://blog.csdn.net/m0\_46179553/article/details/118759655

[4] CSDN社区论坛. WBF(Weighted boxes fusion)-模型融合与NMS、soft-NMS

https://blog.csdn.net/zephyr\_wang/article/details/123609925

[5] CSDN社区论坛. 使用HOG+SVM进行人脸检测

https://blog.csdn.net/YunduanBeichen/article/details/124717748

[6] CSDN社区论坛. 目标检测之—非极大抑制(NMS)综述

https://blog.csdn.net/qq\_25344301/article/details/120094530

[7] Bodla N , Singh B , Chellappa R , et al. Improving Object Detection With One Line of Code:, 10.48550/arXiv.1704.04503[P]. 2017.