

基于位置敏感模型的深度跟踪算法研究

查宇飞¹, 吴敏², 库涛³, 陈兵³, 张国强³

(1. 西北工业大学计算机学院, 陕西西安 710072; 2. 95894 部队, 北京 10200;
3. 空军工程大学航空工程学院, 陕西西安 710038)

摘要: 视觉目标跟踪旨在寻找与跟踪目标具有相同语义信息的样本,并在视频中精确定位样本的位置.最近,深度分类模型被用来提取跟踪目标的深度嵌入式特征,然而,由于深度分类模型给予相同类别的样本一样的标签,这样容易导致跟踪模糊,甚至失败.为了解决这个问题,本文将样本的空间位置信息加入深度分类模型中,提出位置敏感损失函数.本文所提出的损失函数不仅继承了分类损失函数的特性,并根据样本的空间位置信息对相同标签的样本进行了排序.也就是说,本文的损失函数可以同时实现类间可分和类内排序.相比于分类损失函数,本文的损失函数更适合目标跟踪任务.本文在 OTB100^[1]和 VOT2016^[2]上进行了测试,结果表明本文算法可以实现较好的跟踪性能.

关键词: 目标跟踪; 深度学习; 位置敏感模型; 类内排序

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2019)10-2076-07

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2019.10.008

Deep Tracking Algorithm Research Based on Location-Sensitive Model

ZHA Yu-fei¹, WU Min², KU Tao³, CHEN Bin³, ZHANG Yuan-qiang³

(1. School of Computer Science, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China;
2. 95894 Troops, Beijing 10200, China;
3. Aeronautics Engineering College, Air Force Engineering University, Xi'an, Shaanxi 710038, China)

Abstract: Visual target tracking is to find samples that have the same semantic information as the tracking target and pinpoint the position of the sample in the video. Recently, the deep classification model is used to extract the deep embedded features of the tracking target. However, since the deep classification model gives the same class of sample labels, it can easily lead to tracking and even failure. In order to solve this problem, we add the spatial location information of the sample to the deep classification model and propose a location-sensitive loss function. The proposed loss function not only inherits the characteristics of classification loss, but also sorts samples of the same label according to the spatial location information of samples. In other words, the loss function in this paper can also encourage the classification between classes and classes. Compared with the classification loss function, the loss function in this paper is more suitable for the task of target tracking. In this paper, OTB100 and VOT2016 were tested, the results show that this algorithm can achieve better tracking performance.

Key words: target tracking; deep learning; location-sensitive model; intra-class ordering

1 引言

视觉目标跟踪^[3]是计算机视觉领域的一个重要研究课题,它旨在寻找跟踪目标的位置,为高级别的分析(如视频监控,视觉导航等)提供基础.视觉跟踪过程可以被描述为:给定一个视频序列,在得知初始帧目标状态的前提下,如目标的位置、尺度大小等,在后续帧中实现对目标运动状态的有效估计.视觉跟踪任务具有较强的实际应用价值,在智能视频监控,人机交互,机器人

等领域得到了广泛的应用.

近来,与传统特征相比,深度学习^[4]被成功地应用于各种计算机视觉任务中.然而,目标跟踪的任务与其他计算机视觉^[5]任务不同,卷积神经网络的传统结构虽然可以很好的应用于图像识别和目标检测中,但却不适用于跟踪任务.首先,深度网络提取的语义信息随着网络深度的增长而变得更加丰富,但同时目标的空间分辨率会下降,不利于定位跟踪目标的空间位置^[6].此外,具有相同类别的样本被分配一样的标签,并被鼓

励得到相似的语义信息,这样最终会损害类内样本的空间差异.跟踪任务只关注一个特定的跟踪目标,着眼于跟踪目标与背景的区别.然而,分类模型对于靠近目标的样本是无能为力的,这会降低定位精度,甚至会导致跟踪失败.

当前的目标跟踪任务主要有两种体系结构来学习判别特征:直接分类体系结构、相似度网络体系结构.这些体系结构充分利用了目标的类别信息或相似性信息来获得良好的性能.在实时跟踪过程中,存在的关键问题是如何在跟踪任务中选择适当的候选样本.尽管分类网络或相似度网络很容易将目标从背景中分离出来,但由于候选样本具有相同的分类属性,因此它们不能将拥有相似语义信息的候选样本区分开来,容易将不太适合的候选样本作为跟踪结果,这会降低跟踪的准确性,甚至导致跟踪失败.目前的大部分深度学习跟踪方法的损失函数是由 softmax 函数^[7]或 contrastive 函数构成的.据我们所知,softmax 函数只能保证可分性,contrastive 函数则考虑到了类内的相似和类间的差异.然而,在训练过程中具有相似语义信息的样本分配到了相同的类别标签,忽略类内的排序,这将降低定位精度,甚至导致跟踪失败.相应地,这些损失函数能够训练网络学习更好的特征来区分背景与目标,但是不能区分与跟踪目标属于同一类别的候选样本.

为了提高深度特征对空间位置信息的敏感性,得到高位置精度的判别特征,本文将目标的位置信息代入目标损失函数中,提出位置敏感损失函数.如图 1 所示,在图 1(b)中,样本 x_1 和 x_2 拥有相似的语义信息,将其与目标模板 w 进行卷积时,得到相同的响应.由图 1(a)可知,样本 x_2 与跟踪目标的覆盖率要大于 x_1 ,更符合当前帧目标的要求,而目前的方法将其同等对待,不利于目标跟踪任务.为了解决这个问题,本文将目标的空间位置信息代入目标损失函数中,提出位置敏感损失函数,不仅考虑目标的语义信息,同时考虑目标的空间位置信息.如图 1(c)所示,利用本文的损失函数训练的网络提取出来的目标的深度特征,与目标模板进行卷积后,样本 x_2 的响应明显大于 x_1 ,符合目标跟踪任务的要求,可以实现目标的精确定位.

本文将位置敏感损失函数引入到深度神经网络中来学习跟踪任务的判别特征.它可以将跟踪目标与背景区分开来,并且用语义信息对样本进行排序.与将空间位置信息作为标签进行编码的回归模型不同,本文将空间位置信息引入到分类模型中,将目标的语义信息和空间位置信息同时考虑在内以学习目标的判别特征.在离线训练阶段将样本的空间位置信息加入深度分类模型中.通过将大量训练数据代入深度网络中,提取训练样本的深度特征,将训练样本的深度特征与核卷

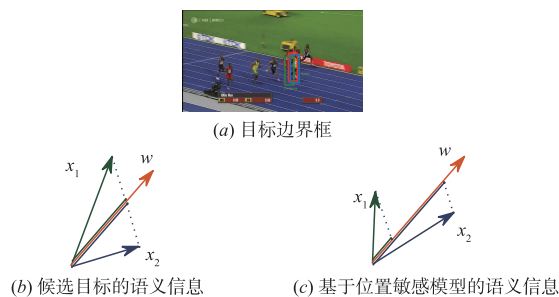


图1 样本的语义信息

积得到响应值,然后将训练样本的响应值与其所对应的空间位置信息代入位置敏感损失函数中,实现对深度网络的参数训练.通过大量数据训练网络参数,使得训练样本的响应值与其空间位置信息相对应,即空间位置信息越大的训练样本所得到的响应值越大.在跟踪阶段,测试样本是未知的,因此测试样本的空间位置信息也是未知的,但是深度网络是由大量训练样本通过离线训练得到的,利用离线训练好的深度网络来提取测试样本的深度特征,深度网络提取的测试样本的深度特征可以反映测试样本的空间位置信息,因此,深度特征与核卷积得到的响应值大的测试样本是目标的概率比较大,进而提高了跟踪的准确度.

2 位置敏感损失函数

2.1 位置敏感损失函数

分类模型直接学习非线性函数,并在最后一层之后使用交叉熵损失.然后,通过判别特征之间的距离度量来计算跟踪目标和候选样本之间的相似度.而相似度模型采用一对图像作为弱标签的输入,判断候选样本是否属于跟踪目标.上述方法的主要缺点在于只利用了目标的语义信息来进行分类,忽略了对于视觉追踪至关重要的空间位置信息.

为了提高学习特征的判别能力,本文利用目标的空间位置信息作为监督信号来构建位置敏感损失函数.给定训练数据集 $\{I_i, y_i, m_i | i = 1, 2, \dots, N\}$, 其中 $I_i \in R^n$ 为输入样本,对应的标签为 $y_i \in \{1, 2, \dots, C\}$, 空间位置信息为 $m_i \in [0, 1]$ 是根据样本 x_i 和跟踪目标计算的.

Softmax 函数已被证明是一个简单而有效的损失函数,在卷积神经网络 (Convolutional Neural Network, CNN)^[8] 中训练深层模型,在最后一个全连接层中输入 $\{x_i \in R^d, i = 1, 2, \dots, N\}$ 以及所对应的权重 $\{w_i \in R^d, i = 1, 2, \dots, C\}$:

$$L_{\text{softmax}} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{w_i \cdot x_i}}{\sum_{j=1}^c e^{w_j \cdot x_i}} \quad (1)$$

其中 C 是类别的数量.通过上述损失学习到的深度特征,这些样本可以被归类为不同的类别.事实上,softmax

函数重视类间可分,而忽略了类内排序.

与分类任务相比,目标跟踪问题不仅要考虑目标的语义标签,还考虑目标的空间位置.我们的目标是区分目标与背景,同时将拥有相同类别的样本根据其不同的空间位置给予不同的分数.为了同时实现类内排序和类间可分,在 softmax 函数中引入空间位置信息 m :

$$L_i = -\log \frac{e^{w_i^T x_i - m_i}}{e^{w_i^T x_i - m_i} + \sum_{j=1, j \neq y_i}^c e^{w_j^T x_i}} \quad (2)$$

其中 L_i 表示第 i 个样本的损失值.空间位置信息 m_i 是第 i 个样本与跟踪目标之间的位置关系.

2.2 位置-语义信息

分类模型只利用了目标的语义信息,不能精确定位目标的空间位置.本文在分类模型的基础上增加目标的空间位置信息,在实现类间可分的同时实现类内排序.为了更好的说明本文的损失函数,考虑 $C=2$ 的简单情况,此时本文的损失函数成为一个二分类函数.

2.2.1 类间语义可分

将 x_i 作为模型的输入,带入本文的损失函数中,分别得到两个类别的概率:

$$p_1 = \frac{e^{w_1^T x_i - m_i}}{e^{w_1^T x_i - m_i} + e^{w_2^T x_i}} \quad (3)$$

$$p_2 = 1 - p_1 = \frac{e^{w_2^T x_i}}{e^{w_1^T x_i - m_i} + e^{w_2^T x_i}} \quad (4)$$

其中 p_1 为 x_i 属于第 1 类的概率, p_2 表示 x_i 属于第 2 类的概率.

如果 x_i 为第一类,我们则希望 $p_1 > p_2$,也就是,

$$e^{w_1^T x_i} > e^{w_2^T x_i + m_i} > e^{w_2^T x_i} \quad (5)$$

如不等式(5)所示,增加目标的空间位置信息 m_i 可以确保学习到的模型 w_1 比传统 softmax 学习到的模型 w_2 更强.通过 softmax 函数训练的模型可以区分不同类别的样本,但是,当目标出现在两个类别的边界线时,softmax 函数训练的模型无法很好地进行区分.本文在 softmax 函数的基础上增加目标的空间位置信息 m_i ,可以在两个类别之间增加一个距离,能够更好地区分不同类别的目标.通过增加空间位置信息可以实现目标与背景之间的可分.在训练过程中,通过公式 $p_1 = \frac{e^{w_1^T x_i - m_i}}{e^{w_1^T x_i - m_i} + e^{w_2^T x_i}}$, $p_2 = 1 - p_1 = \frac{e^{w_2^T x_i}}{e^{w_1^T x_i - m_i} + e^{w_2^T x_i}}$,如果训练样本 x_i 为目标,令 $p_1 > p_2$ 来训练网络参数,得到深度特征 $w_1^T x_i$;在跟踪过程中,如果测试样本与训练样本类似,此时,将测试样本的深度特征 $w_1^T x_i$ 代入公式 $p_1 = \frac{e^{w_1^T x_i}}{e^{w_1^T x_i} + e^{w_2^T x_i}}$, $p_2 = 1 - p_1 = \frac{e^{w_2^T x_i}}{e^{w_1^T x_i} + e^{w_2^T x_i}}$ 中,可以很容易得到 $p_1 > p_2$ 的结论.

2.2.2 类内空间位置排序

对于二分类问题,softmax 函数可以退化为逻辑回归函数.此时,本文将属于第一类的两个样本 x_1 和 x_2 作为模型的输入,可以形成 x_1, x_2 的概率 p_1 和 p_2 .

$$p_1 = \frac{1}{1 + e^{-(w_1^T x_1 - m_1)}} \quad (6)$$

$$p_2 = \frac{1}{1 + e^{-(w_1^T x_2 - m_2)}} \quad (7)$$

当空间位置信息 $m_1 > m_2$ 时,我们希望样本 x_1 的概率大于样本 x_2 的概率,即 $p_1 > p_2$:

$$e^{-w_1^T x_1} < e^{-w_1^T x_2 - \Delta m} < e^{-w_1^T x_2} \quad (8)$$

其中 $\Delta m = m_1 - m_2$.根据式(8)所示,如果样本 x_1 的空间位置信息 m_1 大于样本 x_2 的空间位置信息 m_2 ,则样本 x_1 和模板 w 的内积大于样本 x_2 和模板 w 的内积,并使他们的内积之间增加了一个 Δm 大小的间隔.也就是说,通过增加样本的空间位置信息训练得到的网络,可以

使样本的概率 $p_1 = \frac{1}{1 + e^{-w_1^T x_1}}$, $p_2 = \frac{1}{1 + e^{-w_1^T x_2}}$ 根据其所对应的空间位置间隔 Δm 产生一个概率上的间隔,使得空间位置信息大的样本更容易排在前面,而排在后面的样本拥有较小的空间位置信息.图 2(a) 显示 softmax 函数训练的模型所提取的样本的深度特征分布情况,softmax 函数训练的模型实现了类间的可分,但是没有将类内的样本进行排序.本文的损失函数通过增加目标的空间位置信息,将样本按照其空间位置的大小情况进行排序,使得空间位置信息大的样本离边界线更远,从而离跟踪目标越近,实现类内排序.

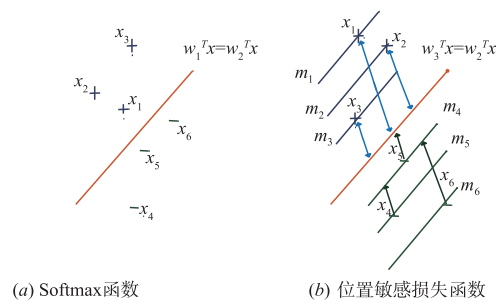


图2 不同函数的特征对比

2.3 度量方式

当分类网络运用到目标跟踪任务中时,它们在跟踪目标和背景之间具有较好的可分性,但是由于候选样本在分类任务中与跟踪目标拥有相同的类别标签,所以跟踪结果的位置精度会被附近的候选样本混淆.为了缓解这种模糊性,本文利用样本的空间位置信息训练深度神经网络来提取判别性特征.

在跟踪评估中,通常使用精确度和成功率作为度量方式来评估序列中的跟踪结果,其中精确度是中心

位置误差为基础,成功率是以目标框覆盖率为基础.本文采用上述两个度量方式来确定每一帧目标的位置.

中心位置误差是样本与跟踪目标的中心位置之间的欧式距离.使用一个序列的所有帧上的平均中心位置误差来总结该序列的总体性能,并作为跟踪结果的评估方式.在我们的损失函数中,基于中心位置误差的度量方式 m 可以表示为:

$$m(P_q, P_g) = e^{-\|P_q - P_g\|^2} \quad (9)$$

其中 P_q 和 P_g 分别是样本和跟踪目标的中心位置.

边界框覆盖率是视觉追踪的另一个评估指标.给定样本边界框 q 和跟踪目标边界框 g ,基于边界框覆盖率的度量方式 m 可以定义为:

$$m(p, g) = \frac{|q \cap g|}{|q \cup g|} \quad (10)$$

其中 \cap 和 \cup 分别是两个边界框的交集和联合. $|\cdot|$ 是相对应区域的面积.

2.4 优化

本文的位置敏感损失函数是基于传统的 softmax 函数提出的,将样本的语义信息和空间位置信息合并.其反向传播的推导如下所示:

$$\begin{aligned} \frac{\partial L_i}{\partial w_k} &= -\frac{\partial \log e^{w_i^T x_i - m_i}}{\partial w_k} + \frac{\partial \log (e^{w_i^T x_i - m_i} + \sum e^{w_j^T x_i})}{\partial w_k} \\ &= -\frac{\delta(y_i, k) x_i e^{w_i^T x_i - m_i}}{e^{w_i^T x_i - m_i}} \\ &\quad + \frac{\delta(y_i, k) x_i e^{w_i^T x_i - m_i} + (1 - \delta(y_i, k)) x_i e^{w_i^T x_i}}{e^{w_i^T x_i - m_i} + \sum e^{w_j^T x_i}} \\ &= \left[-\delta(y_i, k) + \frac{\delta(y_i, k) e^{w_i^T x_i - m_i} + (1 - \delta(y_i, k)) e^{w_i^T x_i}}{e^{w_i^T x_i - m_i} + \sum e^{w_j^T x_i}} \right] x_i \end{aligned} \quad (11)$$

其中, $\delta(\cdot)$ 为指示函数.输入的推导为:

$$\frac{\partial L_i^s}{\partial x_i} = \frac{\sum_{j=1, j \neq y_i}^c (w_j - w_{y_i}) e^{w_j^T x_i}}{e^{w_i^T x_i - m_i} + \sum_{j=1, j \neq y_i}^c e^{w_j^T x_i}} \quad (12)$$

从式(11)和式(12)可以看出,本文的损失函数得到的计算结果与传统的 softmax 函数类似,不同之处在于正确类别的得分为样本 x_i 减去相对 m_i 得到的.显然,当 m_i 为 0 时,本文的损失函数的推导与传统的 softmax 函数的推导一致.

3 目标跟踪

通过设计位置敏感损失函数,我们能够训练网络学到适合跟踪任务的判别特征.将训练好的网络嵌入到跟踪框架中,跟踪过程如图 3 所示.将候选样本和目标模板投影到嵌入空间中,然后用距离度量来评估目

标模板与候选样本之间的相似性,将得分最高的候选样本作为当前帧的跟踪结果.

具体来说,给定一个跟踪目标 t ,判别特征 $f(t)$ 被视为目标模板,在图 3 中是一个红色的五角星.以前一帧的跟踪结果为中心提取出当前帧的候选样本,并表示为 $S = \{z_1, z_2, \dots, z_N\}$,其中包括背景和相似目标等.将候选样本输入到网络中,得到的嵌入式特征被表示为 $\epsilon = \{x_1, x_2, \dots, x_N\}$.当在嵌入空间中获得目标模板和候选样本的特征时,目标模板和候选样本之间的欧氏距离可以表示为:

$$D_i = \|f(t) - f(x_i)\|_2^2 \quad (13)$$

然后,离目标模板最近的候选样本被认为是当前帧中的跟踪结果.

$$i^* = \arg \min_{i=1,2,\dots,N} D_i \quad (14)$$

在图 3 中,绿色四角星表示的候选样本被认为是跟踪结果.在实际跟踪过程中,为了提高跟踪结果的准确性,我们从样本集 S 中选择最相似的 k 个样本,并将这些样本边界框的平均值估计为当前帧的目标状态.

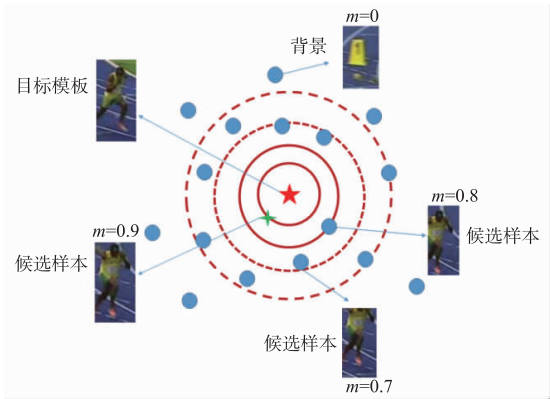


图3 深度嵌入空间的候选样本

4 实验

本文算法在 MATLAB 2016a 中实现的,所采用的工具箱为 MatConvNet^[9].测试视频来自于当前目标跟踪领域最常用的 OTB100 (Online Object Tracking 100 Benchmark) 和 VOT2016 (Visual Object Tracking 2016),在 Nvidia GeForce GTX 1080Ti GPU 和 3.6GHz Intel Core i7-6850K 上运行速度约为 5fps.接下来,我们将研究在训练中加入位置敏感损失函数的效果.

本文的网络架构基于 MDNet^[8].在网络中,利用三个卷积层来提取特征,并且通过两个全连接层学习判别特征.训练数据来自 ALOV300 + ^[10] 数据库,验证数据来自 OTB100 和 VOT2016.在这些数据库中,每个序列被认为是一个类,即拥有相同语义信息的目标,例如,汽车,在不同的序列中被分配不同的标签.此外,使用手动标记的目标边界框来提取训练样本.在跟踪目标周

围提取重叠度大于 0.7 的样本作为正样本. 此外, 重叠度低于 0.5 的候选样本被作为负样本, 在训练过程中正样本和负样本被集合为一个类别. 这样, 我们构建了多类样本进行网络训练.

4.1 算法验证

与关注分类错误率的分类任务不同, 目标跟踪任务更重视定位的准确性. 为了验证位置敏感损失函数可以学习更好的判别特征以完成目标跟踪任务, 本文采用准确率和成功率作为算法的评估标准, 在跟踪任务中讨论本文算法进行位置排序和语义分离的能力.

4.1.1 空间位置信息与语义信息

Softmax 函数采用深度网络来学习语义分离性, 通过语义信息来区分不同类别的目标; 但目标跟踪任务

不仅需要区分不同类目标, 更重要的是精确定位目标的空间位置, 而单一的 softmax 函数无法达到这个要求. 通过增加每个样本的空间位置信息, 根据空间位置的不同来对样本进行排序, 可以精确定位目标的位置. 结果如图 4 所示. 显然, 样本空间位置信息的引入可以有效提高精确度和成功率, 得分分别为 0.758 和 0.647.

如图 4 所示, softmax 函数的精确度和成功率^[11]为 0.708 和 0.606. 也就是说, 本文算法所得出的结果和 softmax 函数相比相比, 实现 5% 左右的提高. 实验结果说明, 本文所提出的位置敏感损失函数可以使候选样本按照其所对应的空间位置信息进行排序, 离目标模板越近的候选样本的得分越高, 选取得分最高的候选样本作为跟踪结果, 可以提高跟踪的准确性.

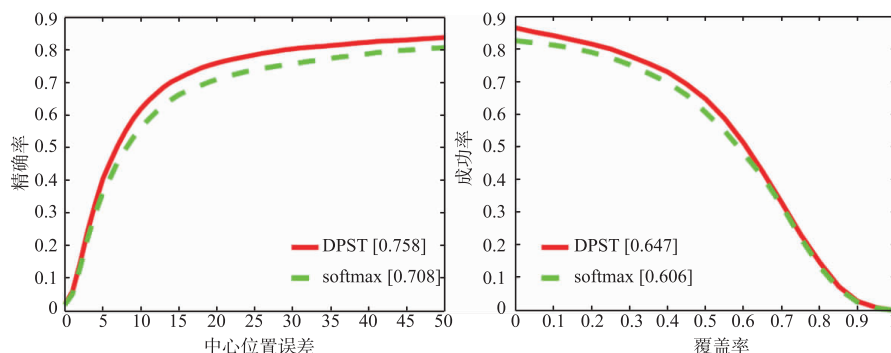


图4 位置敏感损失函数与softmax函数的性能对比

4.1.2 中心位置误差与覆盖率

为了更好地实现本文的位置敏感损失函数的性能, 准确的将候选样本按照其所对应的空间位置信息进行排序, 本文选择了两个不同的度量方式来进行实验: 中心位置误差 (CLE) 和边界框覆盖率 (BBO). 如表 1 所示, 两种度量方式得到相似的得分, 这表明两种度量方式都能够实现类内排序的功能. 此外, 中心位置误差的精确度比覆盖率稍好, 但在成功率上稍差. 也就是说, 通过中心位置误差学习到的性能更多的体现在中心坐标上, 而覆盖率则希望候选样本和目标模板之间有更多的重叠. 但对于跟踪任务来说, 覆盖率更能反应跟踪准确度, 因此, 本文选择了边界框覆盖率为本文采用的空间位置信息.

表 1 不同度量方式下的精确度和成功率

度量方式	精确度	成功率
中心位置误差 (CLE)	0.767	0.660
边界框覆盖率 (BBO)	0.763	0.674

4.2 整体性能比较

4.2.1 OTB100 数据库

为了证明本文算法的先进性, 本文在 OTB100 数据

库上与目前最先进的算法进行了对比. 对比算法包括 TCNN^[12], CCOT^[13], MDNet^[8], Staple^[14], DeepSRDCF^[15], EBT^[16], DAT^[17], DSST2014^[18], TGPR^[19], DFT^[20].

图 5 为各算法在 OTB100 数据库上的跟踪结果. 为了更好的说明本文算法的性能, 本文在精确度对比图上列出中心位置误差门限为 20 像素时算法的精确度, 在覆盖率图中列出覆盖率大于 0.5 时算法的成功率. 结果显示, 本文算法 (DPST) 的精确度为 0.918, 超过了排名第二的 MDNet 算法, 本文算法的成功率为 0.686, 超过了所有的对比算法. 实验表明, 通过增加目标的空间位置信息提出的位置敏感损失函数可以实现有效跟踪.

4.2.2 VOT2016 数据库

为了更好地说明本文算法的性能, 除了 OTB100 数据库外, 还在 VOT2016 数据库上进行了实验, 并与 TCNN、CCOT 等先进算法进行对比. 实验结果如表 2 所示.

如表 2 所示, 本文不仅将算法精确度和鲁棒性的得分和排名进行了比较, 还将平均覆盖率作为评价指标之一. 红色代表性能最好的算法, 蓝色次之, 绿色排在第三. 由表 2 可知, 本文算法 DPST 除了在鲁棒性得分中位于第二外, 在其他的评价指标中都取得了最好

的效果. 实验结果表明, 位置敏感损失函数在分类模型中合并目标的语义信息和空间位置信息, 一方面, 强大的分类能力能使目标与背景分离, 另一方面, 它的类内排序特性保证了离目标较远的候选样本的得分比离目标近的候选样本的得分低, 这利于区分好的候选样本与差的候选样本. 因此, 候选样本投影到一个特征空

间, 不同类别的候选样本距离较远, 同一类别的候选样本根据它们空间位置信息的不同进行排序. 由于该特征是深度网络学习的, 所以找到一个简单的距离度量如欧几里德距离来选择最近的候选样本作为特征空间中的跟踪结果是合理的. 这是简单而有效的, 并已得到我们的证明.

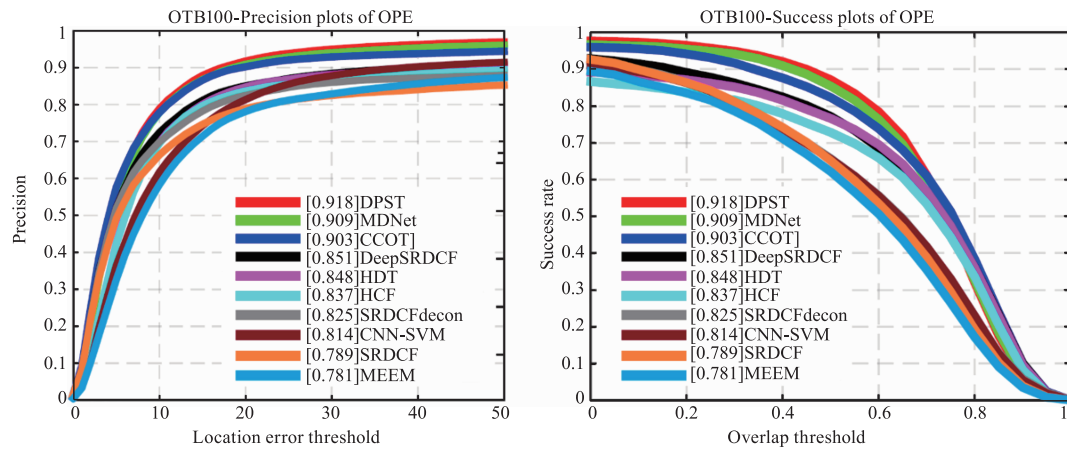


图5 OTB100数据库的整体效果图

表2 VOT2016 数据库的实验结果

跟踪算法	Acc. Score	Acc. Rank	Rob. Score	Rob. Rank	EAO
TCNN	0.5391	5.4931	0.9589	5.4000	0.3249
CCOT	0.5226	5.9651	0.8500	5.0167	0.3310
MDNet	0.5331	5.6254	1.2044	5.7333	0.2572
Staple	0.5382	5.6945	1.3500	6.3667	0.2952
DeepSRDCF	0.5134	6.2119	1.1667	6.0750	0.2763
SRDCF	0.5235	6.1801	1.5000	6.5833	0.2471
EBT	0.4407	7.7926	0.9000	5.0583	0.2913
DAT	0.4626	7.3101	1.7167	6.7250	0.2167
DSST2014	0.4841	6.7834	2.5167	8.1694	0.1814
TGPR	0.4499	7.8484	2.2500	8.3250	0.1811
DFT	0.4327	8.0454	3.5833	9.7194	0.1395
DPST	0.5610	4.9420	0.9000	4.8000	0.3602

5 总结

本文通过在分类模型的基础上增加空间位置信息, 建立位置敏感损失函数, 采用深度网络学习位置敏感判别特征进行跟踪. 本文用类别标签构造位置敏感损失函数, 根据目标的空间位置信息对类内样本进行排序. 因此, 位置敏感损失函数可以同时实现类内排序和类间可分. 在跟踪过程中, 本文利用训练好的网络提取目标的位置敏感判别特征, 通过对嵌入空间中的特征进行排序来跟踪目标. 本文在 OTB100 和 VOT2016 上测试了所提出的方法, 结果表明本文取得了较好的跟踪效果, 可以实现有效跟踪.

参考文献

- [1] WU Y, LIM J, YANG M H. Object tracking benchmark [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834 - 1848.
- [2] CEHOVIN L, LEONARDIS A, KRISTAN M. Visual object tracking performance measures revisited [J]. IEEE Transactions on Image Processing, 2016, 25(3): 1 - 1.
- [3] 孙航, 等. 基于多阶段学习的相关滤波目标跟踪 [J]. 电子学报, 2017, 45(10): 2337 - 2342.
SUN Hang, et al. Correlation filtering target tracking based on online multi-lifespan learning [J]. Acta Electronica Sinica, 2017, 45(10): 2337 - 2342. (in Chinese)

- [4] 李康,李亚敏,等. 基于卷积神经网络的鲁棒高精度目标跟踪算法[J]. 电子学报,2018,46(9):2087-2093.
LI Kang, LI Ya-min, et al. A robust and accurate object tracking algorithm based on convolutional neural network [J]. Acta Electronica Sinica, 2018, 46(9): 2087-2093.
- [5] REN S, HE K, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Image Processing, 2017, 39(6): 1137-1149.
- [6] MA C, HUANG J B, YANG X, et al. Hierarchical convolutional features for visual tracking [A]. The 2015 IEEE International Conference on Computer Vision (ICCV) [C]. US: IEEE, 2015. 3074-3082.
- [7] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: Convolutional architecture for fast feature embedding [A]. Proceedings of the 22nd ACM International Conference on Multimedia [C]. US: ACM, 2014. 675-678.
- [8] NAM H, HAN B. Learning multi-domain convolutional neural networks for visual tracking [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. US: IEEE, 2016. 4293-4302.
- [9] VEDALDI A, LENC K. Matconvnet: Convolutional neural networks for matlab [A]. Proceedings of the 23rd ACM International Conference on Multimedia [C]. US: ACM, 2015. 689-692.
- [10] SMEULDERS A W M, CHU D M, CUCCHIARA R, et al. Visual tracking: An experimental survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(7): 1442-1468.
- [11] WU Y, LIM J, YANG M H. Object tracking benchmark [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(9): 1834-1848.
- [12] NAM H, BAEK M, HAN B. Modeling and propagating cnns in a tree structure for visual tracking [OL]. arXiv Preprint, 2016. arXiv: 1608. 07242.
- [13] MARTIN D, ANDREAS R, FAHAD S, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking [A]. Proceedings of the European Conference on Computer Vision [C]. Netherlands: Springer, 2016. 472-488.
- [14] BERTINETTO L, VALMADRE J, et al. Staple: Complementary learners for real-time tracking [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. US: IEEE, 2016. 1401-1409.
- [15] DANELLJAN M, HAGER G, SHAHBAZ KHAN F, et al. Adaptive decontamination of the training set: A unified formulation for discriminative visual tracking [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. US: IEEE, 2016. 1430-1438.
- [16] ZHU G, PORIKLI F, LI H. Beyond local search: Tracking objects everywhere with instance-specific proposals [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. US: IEEE, 2016. 943-951.
- [17] POSSEGGGER H, MAUTHNER T, BISCHOF H. In defense of color-based model-free tracking [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. US: IEEE, 2015. 2113-2120.
- [18] DANELLJAN M, HÄGER G, KHAN F S, et al. Discriminative scale space tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8): 1561-1575.
- [19] JIN G, HAIBIN L, et al. Transfer learning based visual tracking with gaussian process regression [A]. Proceedings of the European Conference on Computer Vision [C]. Zurich, Switzerland: Springer, 2014. 188-203.
- [20] SEVILLA-LARA L, LEARNED-MILLER E. Distribution fields for tracking [A]. The 2012 IEEE Conference on Computer Vision and Pattern Recognition [C]. US: IEEE, 2012. 1910-1917.

作者简介



查宇飞 男, 1979 年出生, 湖北人. 副教授, 硕士生导师, 主要从事机器学习、计算机视觉、目标跟踪、目标检测及人工智能等方面的研究.



吴敏 (通信作者) 男, 1994 年 11 月出生, 湖南常德人. 2016 年毕业于空军工程大学电子工程系获学士学位, 2019 年毕业于空军空城大学信息与通信工程专业获硕士学位, 主要从事目标跟踪、机器学习及计算机视觉方面的研究.
E-mail: kj2671231@163.com