doi:10. 16055/j. issn. 1672-058X. 2020. 0001. 004

基于相关滤波和卷积神经网络的目标跟踪算法

王雪丽, 李 昕*
(安徽理工大学 电气与信息工程学院,安徽 淮南 232001)

摘 要:在目标跟踪系统中,获得目标的良好表征是确定目标跟踪性能的关键,因此提出一种基于相关滤波和卷积神经网络的目标跟踪算法;该算法首先在各视频场景内预先选定可清晰区分目标外观的参考区域块用以构造训练样本,并构建了两路不完全对称但权值共享的卷积神经网络;该卷积神经网络使得参考区域外目标的输出特征尽可能与参考区域内目标的输出特征相似,以便于获得参考区域内目标的良好表征,并在其中一路加入了相关滤波模块,实现了卷积网络与相关滤波的结合;实验结果验证了该算法的可行性。

关键词:相关滤波;卷积神经网络;目标跟踪;傅里叶
中图分类号:TP391 **文献标志码**:A **文章编号**:1672-058X(2020)01-0019-06

0 引 言

通常来说,在计算机视觉领域,目标跟踪是基本问题之一。其基本任务是在给定目标初始状态的情况下,在接下来的连续帧中估算目标图像序列的任务轨迹。目前,目标跟踪在众多实时视觉领域中扮演着至关重要的作用,例如智能监测系统、自动驾驶、无人机监控、智能交通管制和人机界面等。

目前,大体上可把目标跟踪模型分成两类:生成式和判别式。生成模型是指使用生成模型来描述目标的表现特征,然后以此为模板在搜索区域内进行最小化模式匹配,以寻找最佳匹配窗口。典型的生成式模型跟踪算法有 L1 APG(Accelerated Proximal Gradient,APG)[1],其假设每一个候选目标都可以由字典稀疏表示,在所有候选目标中,选择具有系数最稀疏同时具有重构误差最小的目标作为跟踪结果。所谓的判别式跟踪模型则是把跟踪问题视为二元分类问题,主要通过训练分类器从背景中区分目标。其中相对来说 KCF(Kernelized Correlation Filters)跟踪算法是一种比较经典的跟踪算法[2]。该算法使用了岭回归模型,巧妙地引入了具有循环结构的模板对其进行傅里叶变换,避免了岭回归中矩阵求逆问题,大大地提升了跟踪的速度和效率。

近年来,基于 CF(Correlation Filter,相关滤波)的目标跟踪算法在计算效率和竞争效果上的显著优势引起了广泛的关注。CF 引入了傅里叶变换,降低了计算量。由此衍生出了一系列的跟踪算法,比较值得一提的有带有多通道特征的 KCF 跟踪算法,该算法的核心思想是循环移动跟踪目标区域,用以构造大量的样本来训练分类器。此外,2016 年 Martin Danelljian 在相关滤波算法的基础上,采用 CNN(Convolution Neural Netuork,卷积神经网络)+

收稿日期:2019-05-13;**修回日期**:2019-06-28.

作者简介:王雪丽(1994—),女,安徽滁州人,硕士研究生,从事图像处理、信号处理研究.
* **通讯作者**:李昕(1981—),女,安徽淮南人,副教授,硕士生导师,从事智能信息处理、信号处理、图像处理等研究,Email:417345060@qq.com.

HOG+CN 的特征组合,同时为了降低特征的维度,在特征提取上做了简化,用原来特征的子集,在 n 维特征中选取了其中的 m 维,大大减小了特征维数过多导致滤波器冗余的问题[3],通过引入空间正则化度量抑制背景延伸了 CF 的训练域。然而,基于 CF 的目标跟踪有两个主要的缺点:(1)手动抽取特征无法捕捉目标的语义信息;(2)缺乏训练数据。为了克服 CF 手动抽取特征性能的不足,将深度卷积特征引入 CF,成功地优化了此种缺陷。尽管基于 CNN 特征的 CF 目标跟踪[4]已经在一定程度上克服了几何和环境变化,提高了鲁棒性。但是该方法抽取每一帧的 CNN 特征和训练/更新 CF 跟踪器仍然需要更大的计算量。因此,无法实现实时的目标跟踪。

针对抽取 CNN 特征计算量大的问题,在改进的卷积神经网络的基础上,结合相关滤波器,加入了快速傅里叶变换(Fast Fourier Transform,FFT),降低了计算量。CF 可以通过高效的解岭回归问题,将一个块图像从周围块中区分出来,并且由于采用了 FFT 和 element-wise 操作,比随机梯度下降(Stochasic Gradient Descent,SGD)算法更高效,相比于嵌入的方法,判别器更适应于特定的场景目标。上述的许多研究工作,都仅仅是将 CF 应用到了预训练的 CNN 特征上,而没有任何两种方法的集成,笔者实现了端到端的训练 CNN-CF 的结合,其关键点是将 CF 作为可微的 CNN 层,以便误差可以通过 CF 反传 CNN 特征,并且所有的网络结构都是轻量级网络,仅需要几个参数,就可以有很好的性能表达。

1 相关滤波

通常的相关滤波器是一种学习判别分类器,通过搜索相关场景图的最大响应值来估计目标对象。即在场景中,对每个感兴趣的目标产生高响应,对于背景则产生低响应。可使用一种单通道信号生成多通道数据或图像的方式来简化符号。由于在实际过程中并不只处理一维单通道图像,更多的是处理像彩色图像(具有 R、G、B 3 个通道)和梯度方向直方图(HOG)这样具有多个通道的情况。所以定义了 f 作为尺寸为 $M\times N$ 的训练信号(当前帧),把全部的循环移位的 f 作为训练样本。每一个移位样本 $f_{m,n}\in\{0,1,\cdots,M-1\}\times\{0,1,\cdots,N-1\}$,高斯函数为 $g(m,n)=e^{-\frac{(m-M/2)^2+(n-N/2)^2}{2\sigma^2}}$,$\delta$ 是核大小。然后,带有相同尺寸大小 f 的相关滤波器 h 可以通过式(1),最小优化值来求解。

$$\min\|h\otimes f-g\|_2+\lambda\|h\|_2 \tag{1}$$

其中,⊗是循环卷积符号,λ 是正则化参数。可以通过傅里叶变换在傅里叶域中求出目标位置。得到如下公式:

$$H^*=\frac{G\cdot F^*}{F\cdot F^*+\lambda} \tag{2}$$

其中,* 是共轭符号,· 表示两个向量的对应元素相乘操作。$H=\mathcal{F}(h)$,$F=\mathcal{F}(f)$,$G=\mathcal{F}(g)$。当一个搜索图像 y(下一帧)到来,反应图 z 可以描述如下:

$$z=\mathcal{F}^{-1}(H\cdot Z) \tag{3}$$

联立式(2)和式(3)可以得到:

$$z=\mathcal{F}^{-1}\left(\frac{G\cdot F}{F\cdot F+\lambda}\cdot Z\right)=\mathcal{F}^{-1}\left(\frac{G\cdot(F\cdot Z)}{F\cdot F+\lambda}\right) \tag{4}$$

大体说来,相关滤波跟踪算法巧妙地通过循环矩阵的偏移得到分类器的训练样本,使得许多样本矩阵具有了循环矩阵性质。然后,由于循环矩阵的特点,便于把矩阵问题的求解变换到傅里叶域内计算量很低的向量点积操作,从而消除了矩阵求逆过程,大大降低了算法计算量。并且通过训练获得的分类器分类效果较好,并将最大响应值处的位置作为目标新位置,从而实现了快速检测目标的目的,并且可以使用新的目标位置来更新分类器。

然而相关滤波目标跟踪采用固定目标尺度,因而当目标尺度发生变化、目标被遮挡以及目标丢失时,该模块没有采用相应的处理措施,因而存在优化空间。在此提出了一种带有相关滤波器模块的对称网络结构,采用了一种基于 CNN 的多域学习框架,从某一特定的域抓取到共享有用的表达,分离与域无关的表达[5]。系统主要框架图如图 1 所示。

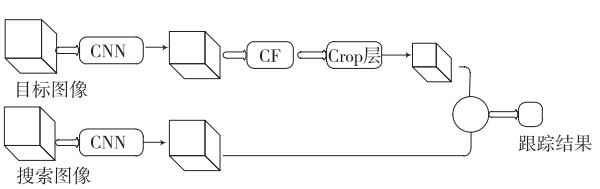


图 1 算法整体框架图

Fig. 1 Algorithm overall frame

2　相关滤波网络

采用了一种基于多域学习的框架网络,并加入了一种 x 和互相关运算操作的相关滤波器模块,框架如图 1 所示。这种变化可以公式化如下:

$$h_{p,s,p}(x',y') = sw(f_p(x')) \quad f_p(y')+b \qquad (5)$$

其中,x'是目标区域,y'是搜索区域,f_p 是学习率为 p 的 CNN,CF 模块 $\omega=\omega(x)$的计算可从训练的特征图 x 中计算出标准的 CF 模板,解决了傅里叶域中的岭回归问题[2]。其效果可以理解成制作的一个鲁棒性转换过程的判别性模板。s(权重)和 b(偏差)是 2 个标量参数,在逻辑回归中使分数范围更加合适。

在训练过程中提供一个大规模的上下文区域的图像的相关滤波器是很重要的。由于增添了最小二乘的思想,所以浅层也具有不错的性能,但是这样做的问题是把 CF 的边界问题也带入到了网络中,因此增加了 Crop 层[6],只保留中间的一部分,这就降低了边界效应。该网络前向传播加入了一个 CNN 特征的 CF 跟踪器,但是此前的算法并不能端到端训练 CF,本文提出了一种可以端到端训练 CF 的方法,即 CF 中的模板对输入的导数使得 CF 也能够被端到端训练。

3　卷积神经网络

在计算机视觉领域中,CNN 已经被广泛应用并取得了很好的效果。Krizhevsky 等[7]通过大规模的数据集训练深度 CNN 和高效的 GPU 进行图像分类,显著提高了性能。2018 年由哈尔滨工业大学牵头提出 STRCF[8](Spatial-Temporal Regularizen Correlation Filters)算法,在相关滤波跟踪领域取得优异成绩。通过在 DCF(Discriminative Correlation Filters)框架中加入空间和时间正则化,提出 STRCF 模型,基于在线 PA 不仅可以合理地逼近多幅训练图像上的 SRDCF(Spatial Regularizen Discriminative Correlation Filters)形式,而且在较大的外观变化情况下更具鲁棒性。

尽管 CNN 已经取得了巨大的成功,但是由于缺少大规模的训练数据使得跟踪算法的性能始终得不到很大的提升。早期基于 CNN 的跟踪算法只能处理预先定义的目标对象类,例如人类。自从 CNN 可以进行离线训练,再进行预测以后,许多算法应运而生。如文献[9]提出了一种基于 CNN 池的学习方法,但是它缺乏训练数据深度网络并且其准确性与基于手动抽取特征的方法相比较并不特别好。近来提出的一些方法[10]为图像分类构建的大型数据集上转移了预训练的 CNN,但是在分类和跟踪任务之间的基本区别表现形式可能不是很明显。与现有方法不同,本文的算法利用大规模视觉跟踪数据进行预训练 CNN 并获得有效的表现。典型的卷积神经网络一般包括卷积层、池化层、全连接层以及 Softmax 回归层。卷积层是卷积神经网络中最重要的层结构,通过卷积层提取图像特征图的好坏直接影响后续层的处理。卷积层通过卷积核与前一层的特征图进行局部连接,得出图像的局部特征,并通过共享权值的方式得出新的特征图。池化层又称为下采样层,主要承接卷积层,对卷积层卷积后的特征图进行特征降维,以减少网络的复杂度,减少计算量。全连接层实际上是特殊的卷积层,但与卷积层不同的是,全连接层中每个神经元与前一层中所有神经元相连接。全连接层的作用是维度变换,把前一层的高维矩阵数据变换成低维矩阵,提取和整合有鉴别能力的特征[11]。

4　多域学习网络

为预训练深度 CNN,采用了多域训练的网络结构(Multi-Domain Network,MDNet)[5],指的是一种训练数据来自多个领域和域信息包含在学习过程中的学习方法。多域学习在自然语言处理中广泛应用。在计算机视觉社区,多领域学习进行的讨论仅仅只是少数域适应的方法。比如 Duan 等[12],引入了域加权组合用于视频概念检测的 SVM 和 Hoffman 等,提出了对象的混合变换模型分类。

MDNet 分为共享层和特定域层。共享层和检测网络类似,用于学习通用的物体表征。特定域层是对每一类物体都有一个针对它的二分类层,用于区分前景和背景。网络基本上包含了接收 RGB 图像的输入,还有 5 个隐藏层,其中包括了 3 个卷积层和 2 个全连接层。此外,最后一个全连接层通过 K 分支($fc6^1$—$fc6^K$),相对应 K 域,换句话说就是训练序列。卷积层完全对应相应 VGG-M 网络部分,除

了特征图尺寸通过输入尺寸进行调整。后面的2个全连接层都有512个输出单元。每一个K支都包含一个带有叉熵损失分类器的二进制分类层,用于区分每个域的目标和背景。并将 $fc6^1$—$fc6^K$ 特定领域层和前面的所有层作为共享层。

因此,MDNet网络有以下优点:

(1) 网络体系规模远小于通常的识别网络,例如,AlexNet和VGG-Nets等。

(2) 应用了专门的跟踪数据来训练;

(3) 对于同一类物体进行的特定域分类,可以有效区分目标和背景。

5 跟踪算法

网络本身仅衡量两个图像块的相似性,并且通过评估网络前向传播来进行在线跟踪。为了将此网络应用到目标跟踪中,必须要和跟踪器逻辑程序结合起来。该算法使用一个简单的跟踪算法评估相似函数的实用性。

在线跟踪算法的评估通过简单的向前模式的网络来评估。简单说来,就是以新的一帧的前一帧估计的目标位置为中心提取一个搜索区域,将目标的特征和搜索区域相比较,目标新的位置就是得分最高的位置。

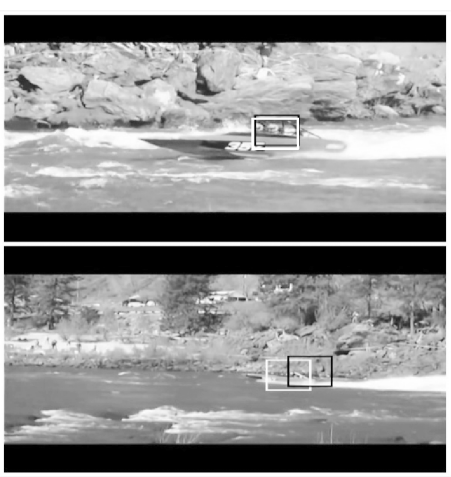
6 实验结果与分析

通过预训练数据集,并按照本文算法中给出的跟踪算法,利用Matlab编程对数据集中的视频序列进行仿真分析。实验结果验证了该算法的可行性,并与KCF算法进行了测试序列视频上的对比分析,鉴于篇幅限制,这里仅给出部分视频序列的仿真结果,如图2所示,其中白色实线为本文算法,黑色实线为KCF算法。

就成功率而言,本文算法对不同场景的视频序列的跟踪效果都还不错,对于视觉而言,可从图2一窥究竟,可以观察到跟踪器的执行效果相对来说还是比较好的。实验结果验证了该算法在场景短期闭塞、快速运动、规模变化和场景混乱这些方面还是具有一定的鲁棒性。实验结果表明,该算法在一定程度上改善了对不同跟踪场景适应性的问题,也验证了该算法的可行性。

(a) 机场

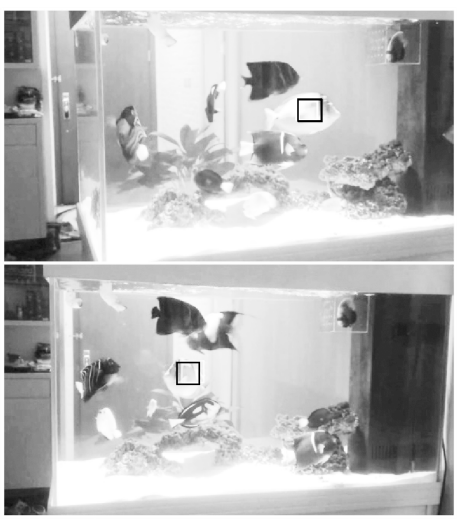
(b) 自行车



(c) 船

(d) 公共汽车

(e) 面部



(f) 鱼

图 2　算法在部分测试视频上的跟踪结果

Fig. 2　The tracking results of algorithms on the part of the test video

表 1 为本文算法与传统 KCF 算法在精确度中值与运行速度上的对比结果。从中可以看出本文算法在精确度和运行速度上都稍优于 KCF 算法。

表 1　两种算法精确度中值和运行速度对比

Table 1　Two kinds of algorithm precision value and running speed

算　法	精确度中值	运行速度/fps
实验组	85	75
KCF	84.8	72.9

7　结束语

采用了一种通过在线学习算法优化浅层特征的反向传播梯度不完全对称的相关滤波器网络。验证了通过建立高效的反向传播图来解决循环系统方程是可行的。实验结果表明:高效的多域学习卷积神经网络加上相关滤波层并不能显著提高跟踪精度。但是,该算法在训练数据时将相关滤波器和相似性网络合并可使浅层网络提取深层特征。基于这种深度特征的算法在不同视频序列上都表现出了良好的鲁棒性和准确性,在某些场景复杂的背景下,也有不错的跟踪效果,在一定程度上,相比传统算法具有更好的性能。但是预训练数据的不足也影响了跟踪效果。

参考文献(References):

[1] BAO C, WU Y, LING H, et al. Real Time Robust L1 Tracker Using Accelerated Proximal Gradient Approach [J]. Proc Institute of Electrical and Electronics Engineers. Comput Soc Conf Comput Vis Pattern Recognit, 2012, 157(10): 830—1837

[2] 罗海波,许凌云,惠斌,等. 基于深度学习的目标跟踪方法研究现状与展望[J]. 红外与激光工程,2017,46(5):0502002—0502003
LUO H B, XU L Y, HUI B, et al. Target Tracking Method Based on the Deep Learning Research Present Situation and Prospect[J]. The Infrared and Laser Engineering, 2017, 46(5):0502002—0502003(in Chinese)

[3] DANELLJAN M, ROBINSO A, KHAN F S, et al. Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking[C]//European Conference on Computer Vision, 2016

[4] ZHANG P H, ZHEN D, JANG C, et al. Fast Fourier Transform Networks for Object Tracking Based on Correlation Filter [J]. Institute of Electrical and Electronics Engineers Access, 2017, 34(4):2169—2171

[5] NAM H, HAN B. Learning Multi-Domain Convolutional

Neural Networks for Visual Tracking[J]. Computer Vision and Pattern Recognition,2016:4293—4302

[6] VALMADRE J, SRIDHARAN S, LUCEY S. Learning Detectors Quickly with Stationary Statistics[C]//Asian Conference on Computer Vision, 2014

[7] KRIZHEVSKY A, SUTSKEVER L, HINTON G E. Imagenet Classification with Deep Convolutional Neural Networks[C]//Conference and Workshop on Neural Information Processing Systems. 2012

[8] LI F, TIAN C, ZUO W, et al. Learning Spatial-Temporal Regularized Correlation Filters for VisualTracking[EB/OL]. http://arxiv.org/abs/1863.08679

[9] LI H, LI Y, PORIKLI F. DeepTrack: Learning Discriminative Feature Representations by Convolutional Neural Networks for Visual Tracking[C]//British Machine Vision Conference. Nottingham,2014

[10] Kristan M, Leonardis A, Matas J, et al. The Visual Object Tracking Vot 2016 Challenge Results[C]//European Conference on Computer Vision Workshop. Amsterdam, Netherlands,2016

[11] 杨文斌,杨会成,鲁春,等. 基于肤色特征和卷积神经网络的手势识别方法[J] 重庆工商大学学报(自然科学版),2018,35(4):77—78
YANG W B, YANG H C, LU C, et al. Recognition Method of Neural Network Based on Color Feature and Convolution Gesture[J]. Journal of Chongqing Techology and Business University(Natural Science Edition),2018, 35(4):77—78(in Chinese)

[12] HOFFMAN J, KULIS B, DARRELL T, et al. Discovering Latent Domains for Multisource Domain Adaptation[C]// European Conference on Computer Vision. Berlin, Heidlberg,2012

Target Tracking Algorithm Based on Correlated Filters and Convolutional Neural Network

WANG Xue-li, LI Xin

(School of Electrical and Information Engineering, Anhui University of Science and Technology, Anhui Huainan 232001, China)

Abstract: In target tracking system, the key to obtaining good characterization is to determine target tracking performance, therefore, this paper proposes a target tracking algorithm about the correlated filters and convolutional neural networks. This algorithm firstly pre-selects reference blocks which can distinguish the target appearance in each video scene to construct the training samples and then build the two-way convolutional neural network which is not completely symmetric and which shares the weights. This convolutional neural network makes the target output characteristics outside the reference area as similar as possible to the target output characteristics in the reference area so as to get good characterization of the target in the reference area. The correlated filter module is added into a way to realize the combination of convolutional network and the correlated filter. Experimental results verified the feasibility of the algorithm.

Key words: correlated filter; convolutional neural network; target tracking; Fourier

责任编辑:罗姗姗

引用本文/Cite this paper:

王雪丽,李昕. 基于相关滤波和卷积神经网络的目标跟踪算法[J]. 重庆工商大学学报(自然科学版),2020,37(1):19—24

WANG X L, LI X. Target Tracking Algorithm Based on Correlated Filters and Convolutional Neural Network[J]. Journal of Chongqing Technology and Business University (Natural Science Edition),2020,37(1):19—24