

ECO 跟踪算法中 CNN 分层插值及加权策略改进^①

陈志旺^② 王昌蒙^③ 王莹 宋娟 彭勇

(燕山大学工业计算机控制工程河北省重点实验室 秦皇岛 066004)

(燕山大学国家冷轧板带装备及工艺工程技术研究中心 秦皇岛 066004)

摘 要 本文是在深度特征与相关滤波相结合的高效卷积运算符(ECO)目标跟踪算法基础上进行的改进。首先,为了提高跟踪速度,提出“浅层特征不插值,深层特征插值”的卷积神经网络(CNN)分层插值处理方法,对具有较高分辨率的浅层特征不插值,对分辨率低的深层特征进行插值计算来提高分辨率;其次,改进了样本空间分类策略,给 CNN 特征层分配不同的权重,突出不同特征层对样本间距离的影响,并且将所有样本信息都保留在训练样本集中;最后,应用判别尺度空间跟踪(DSST)算法提出的对目标尺度估计的方法,增加了目标尺度的候选数量,使尺度估计更加准确。实验结果验证了所设计算法的有效性。

关键词 目标跟踪; 高效卷积运算符(ECO); 卷积神经网络(CNN); 相关滤波; 尺度估计

0 引言

目标跟踪的主要任务是给定视频序列中的第一帧目标的初始状态(目标位置、目标框尺度),在后续的视频序列中去估计目标新的运动轨迹或状态,它是计算机视觉中的一个基础问题。目标跟踪技术已经广泛应用在安防监控、人机交互、机器人、军事侦查、智能交通^[1]等领域,其实用价值日益增强。经过多年的研究,目标跟踪技术已经有了长足的发展和进步,但依旧面临着很多复杂环境下的挑战,如目标外形变化、光照变化、遮挡、背景复杂等,这些因素都将导致目标漂移甚至跟踪失败。

跟踪方法可以分为基于生成模型的跟踪和基于判别模型的跟踪。生成模型跟踪方法应用特征空间来描述目标,学习一个有代表性的目标模型,利用该模型去搜索图像区域,按照模式匹配的方法,在图像中找到和该模型最匹配的区域即为目标。比较典型

的例子有均值漂移 Meanshift^[2]、粒子滤波 Particle Filter^[3]和 L-K 光流算法^[4]。判别模型将跟踪看成一个二分类问题,由构建的正负样本训练最优分类面,把目标和背景在图片中区分开来,将分类器置信度最大的位置作为检测到的目标位置。基于支持向量机^[5](support vector machines, SVM)的方法,基于随机森林^[6](random forest, RF)分类器的方法都是判别模型跟踪方法的代表。近几年,相关滤波理论被应用于判别模型跟踪算法中,取得了较好的跟踪效果。

2010 年, Bolme 等人^[7]提出了最小均方误差输出和(minimum output sum of squared error, MOSSE)相关滤波器,首次将相关滤波应用到目标跟踪算法上,应用灰度等级图像提取单通道特征图,在训练和检测过程中利用相关滤波的性质,将时域的卷积操作转化为频域的点乘,使计算效率大大提高。2014 年, Henriques 等人^[8]把核技巧引入到脊回归中,提出了核相关滤波器(kernelized correlation filters, KCF)

① 国家自然科学基金(61573305)资助项目。
② 男,1978 年生,博士,副教授;研究方向:多旋翼飞行控制,目标跟踪;E-mail: czwaaron@ ysu. edu. cn
③ 通信作者, E-mail: wangchangmeng123@ 163. com
(收稿日期:2019-05-25)

目标跟踪算法,应用核函数把低维空间不可分地映射到高维核空间变为线性可分,应用循环矩阵进行正负样本采集,由于循环矩阵在傅里叶空间可对角化,所以可将复杂的矩阵运算转化为元素的点乘,提高了运算速度,同时利用脊回归训练目标检测器,进一步提升了跟踪算法的性能。此后,很多研究者基于 KCF 跟踪算法进行了改进,大量论文相继发表出来。胥方等人^[9]针对核相关滤波(KCF)算法对快速移动目标跟踪中由于边界效应导致误差跳动与跟踪丢失的问题,提出了基于核相关滤波的自适应跟踪算法。Danelljan 等人^[10]利用颜色(color-names, CN)特征来构建目标外观特征,将 RGB 三通道的颜色转换成 11 种颜色特征,使用主成分分析^[11](principal components analysis, PCA)将 11 维特征降至 2 维,得到目标低维颜色特征,然后选择最为明显的颜色作为目标特征进行跟踪。文献[12]设计了 2 个相关滤波器,一个 2 维的位置滤波器和一个 1 维的尺度滤波器,分别实现目标的位置跟踪和目标尺度变换,STAPLE 算法^[13]提出一种融合算法,用方向梯度方图(HOG)特征和颜色直方图特征 2 种互补的特征因子对目标进行学习,融合跟踪算法,在跟踪速度有所提高的同时,也保证了跟踪效果。张立国等人^[14]考虑到应用环境复杂,跟踪算法对光照变化鲁棒性不强,导致目标跟踪时数据丢失、实时跟踪时检测速度不够等问题,结合哈希指纹特征对光照变化鲁棒性强的特点,提出一种多尺度(multi-scale)感知哈希(Hash)特征的目标跟踪算法,即 Mhash 算法。

随着深度学习在视觉分类和检测领域获得突破性的进展,由于其强大的特征学习能力使其在跟踪领域也逐渐得到重视。DLT^[15](deep learning tracking)是第一个在目标跟踪算法中引入深度学习的方法。随后,越来越多的深度神经网络模型,如自动编码器^[16]、卷积神经网络^[17](convolutional neural networks, CNN)、循环神经网络^[18](recurrent neural network, RNN)、孪生(siamese)网络^[19]等都被应用到了目标跟踪领域,并在跟踪效果上取得了较好的绩。用于跟踪的高效卷积运算符(efficient convolution operators for tracking, ECO)^[20]是 C-COT^[21]的改进版本,C-COT 在 OTB100^[22]上获得了较好的效果,并且

在 VOT2016^[23]获得了第 1 名,ECO 在 C-COT 的框架上旨在通过解决计算复杂度和过拟合问题来同时提高速度和跟踪效果。文献[24]在 ECO 的基础上,只用 CNN 提取的特征,将 CNN 的浅层特征和深层特征进行融合,使追踪器具有较好的性能。Bhat 等人^[25]通过分析浅层特征和深层特征的特征差异,提出两种特征分而治之,深层特征部分加入了数据增强,增加训练样本数量,两部分响应图自适应融合,得到最优的目标定位结果。STRCF 算法^[26]将空间正则化和时间正则化应用到判别相关滤波(discriminative correlation filter, DCF)框架中,使算法对目标遮挡具有很强的鲁棒性,并且能很好地适应目标较大的外观变化,应用交替方向乘子法(alternating direction method of multipliers, ADMM),在不降低效率的情况下提高了跟踪的准确性和速度。

本文在 ECO 算法的基础上,为了提高算法的跟踪速度,CNN 提取的特征图进行插值处理时采取分层插值,分辨率较高的浅层特征不插值,分辨率低的深层特征插值,以此降低插值操作带来的复杂运算,提高跟踪速度;在生成样本空间策略上,保留了所有的样本信息,使训练样本集有更丰富的多样性,并且考虑到每个特征层的权重不同对样本的分类会产生影响从而影响跟踪效果,为不同特征层分配了不同的权重;为了能更好地应对目标尺度的变换,应用文献[12]提出的 1 维尺度滤波器对目标尺度变换进行估计,把尺度的候选数从 ECO 算法中的 5 个增加到 33 个,使尺度估计更加鲁棒。

1 ECO 算法介绍

1.1 连续卷积算子

在 C-COT^[21]中提出学习连续卷积算子的理论,在训练样本中引入一个内插模型,把特征图转换到连续的空间域,假设 $\mathbf{x} = [x_1, x_2, \dots, x_j, \dots, x_M]$, 第 j 个样本 x_j 包含 D 个特征通道 $x_j^1, x_j^2, \dots, x_j^d, \dots, x_j^D$, 特征通道 $x_j^d \in \mathbb{R}^{N_d}$ 被看作由离散空间变量 $n \in \{0, \dots, N_d - 1\}$ 索引的函数 $x_j^d[n]$, N_d 表示 x_j^d 中的空间样本数,对于每一个特征通道 d , 定义插值操作 $J_d\{x_j^d\}(t)$ 有如下形式:

$$J_d\{x_j^d\}(t) = \sum_{n=0}^{N_d-1} x_j^d[n] b_d\left(t - \frac{T}{N_d}n\right) \quad (1)$$

$J_d\{x_j^d\}(t)$ 是由内插函数 $b_d(t)$ 的各个平移叠加构造出来的, $d \in \{0, 1, 2, \dots, D\}$, $x_j^d[n]$ 是每个平移后的函数权值。特征图的空间支持是连续域 $[0, T) \subset \mathbb{R}$, $t \in [0, T)$ 。

通过学习一个线性卷积算子 $S_f\{x_j\}$, 该算子通过寻找图片区域中最大置信分数来估计新一帧目标位置, 在连续域方程中, 算子 $S_f\{x_j\}$ 是由一系列卷积滤波器 $f = (f^1, f^2, \dots, f^d, \dots, f^D)$ 构造出的, 定义卷积算子为

$$S_f\{x_j\} = \sum_{d=1}^D f^d * J_d\{x_j^d\} \quad (2)$$

这里 $*$ 表示卷积运算, 下同。每个特征通道首先用式(1)进行插值操作, 然后和对应的滤波器卷积, 最后所有卷积响应加起来产生最终的置信函数 $S_f\{x_j\}$ 。

每个训练样本 x_j 由卷积算子的期望输出 y_j 所标定, 所以相关滤波器的损失函数公式为

$$E(f) = \sum_{j=1}^M \alpha_j \|S_f\{x_j\} - y_j\|^2 + \sum_{d=1}^D \|w f^d\|^2 \quad (3)$$

式中, α_j 表示每个训练样本的权重, w 是空间正则化惩罚项, M 表示训练样本的个数, 通过最小化函数式(3)来训练滤波器 f 。

虽然 C-COT 算法在跟踪效果上取得了非常好的效果, 但是该算法模型参数多, 需更新的参数达 800 000 个, 这么多参数除了速度慢, 还容易引起过拟合; C-COT 生成训练样本集的方法是把每一帧的跟踪结果放在一个训练集中, 这样随着视频越来越长, 训练集就会越大, 在更新模型时还是容易过拟合。

1.2 因式分解卷积操作

ECO 算法中, 因式分解的引入主要是用来解决模型参数多的问题, 在 C-COT 中学习的滤波器 f^d 中, 很多滤波器在卷积过程中对目标定位所起的作用并不大, 所以用了一个更小的滤波器集合 $f^1, \dots, f^c, \dots, f^C$, $C < D$, 对于特征层 d 的滤波器由滤波器 f^c 和学习系数 p_{dc} 的线性组合 $\sum_{c=1}^C p_{dc} f^c$ 组成, 这个系

数可以看作是 $D \times C$ 的矩阵 $P = (p_{dc})$, 新的多通道滤波器可以被表示为矩阵的矢量积 Pf , 可以得到因式分解卷积操作公式为

$$\begin{aligned} S_{pf}\{x_j\} &= Pf * J\{x_j\} = \sum_{cd} p_{dc} f^c * J_d\{x_j^d\} \\ &= f * P^T J\{x_j\} \end{aligned} \quad (4)$$

式(4)可以看作 2 步, 首先把插值的特征图乘以 P^T , 得到 C 维的特征图, 然后再和相应的滤波器 f 进行卷积。因为 $C < D$, 因此因式分解卷积操作大大降低了特征图的维度, 而且权值矩阵 P 只需在第 1 帧中学习得到, 在后续帧中保持不变。

1.3 样本空间模型生成

在 ECO 中, 作者应用高斯混合模型来生成训练样本集, 目的在于消除冗余和增加训练样本集的多样性。通过样本 x 和目标输出 y 的联合概率分布 $p(x, y)$, 应用高斯混合模型将目标损失函数式(3)进一步完善, 用下式来训练滤波器:

$$E(f) = \sum_{l=1}^L \theta_l \|S_f\{u_l\} - y_0\|^2 + \sum_{d=1}^D \|w f^d\|^2 \quad (5)$$

式中, L 是高斯分组的个数, θ_l 是第 l 个分组的权重, u_l 是期望。从式(5)可以发现, 高斯期望 u_l 和权重 θ_l 直接替换了式(3)中的 x_j 和 α_j , 在 L 个分组中, 每个组内样本高度相似, 代表某种特定场景, 组与组间差异较大, 增加了训练样本的多样性, 减小了训练样本集的大小, 用式(5)代替式(3)来训练滤波器。

2 改进方法

2.1 CNN 特征层分层插值

从第 1 节的介绍可以知道, C-COT 和 ECO 跟踪算法提出一种用于学习连续空间域中的卷积算子的新公式, 利用插值模型从预训练好的深度网络中提取出多分辨率特征图。式(1)中利用了双立方插值对提取的特征图进行插值处理。

双立方插值的本质是把图像某个像素点最近的 16 个像素点权重卷积之和作为新的像素值。双立方插值能得到精确的插补图形, 但也计算复杂, 所以速度也很慢。在 ECO 算法中, 应用了 CNN 卷积神经网络提取的特征和人工设计的 HOG 特征进行融

合, CNN 卷积神经网络选用的是 VGGNet 模型。图 1 是应用 VGGNet 网络对输入的图片通过不同的卷积层提取的特征图的可视化过程,本文里使用的是 16 层的 VGGNet,应用第 3 个卷积层(对应图 1 中

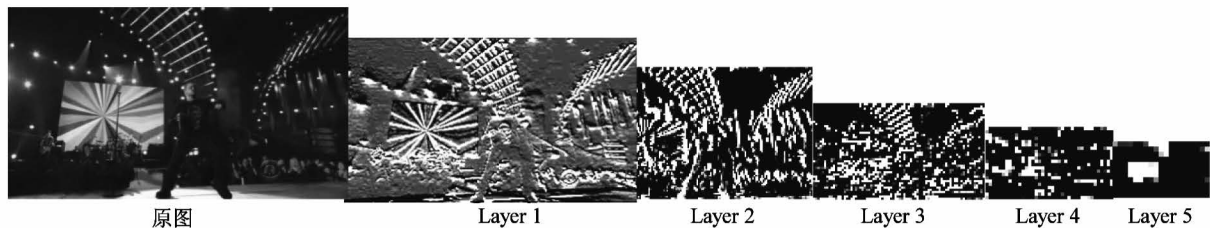


图 1 VGGNet 各层卷积特征图

ECO 应用双立方插值把特征图变为连续域来提高分辨率,由于双立方插值计算比较复杂,如果对每层特征图像都进行插值无疑增加了计算的复杂度,从而影响跟踪的速度。对于卷积神经网络提取的特征来说,浅层特征更有利于目标的精确定位,而且浅层特征包含了较复杂的视觉信息,分辨率比较高,深层特征更能表达图像的语义信息,能够提高跟踪器的鲁棒性,但是分辨率会比较低。所以,为了提高跟踪算法的速度,提出了对提取的特征进行分层插值的改进方法,对于分辨率较高的浅层特征不进行插值,对包含较多语义信息的深层特征进行插值。从图 1 可以看出,随着卷积层的增加,提取的特征越抽象,而且特征图的分辨率也越来越低,对深层特征进行插值来提高其分辨率非常有必要,图 2 是对深层特征图插值与不插值进行的对比。

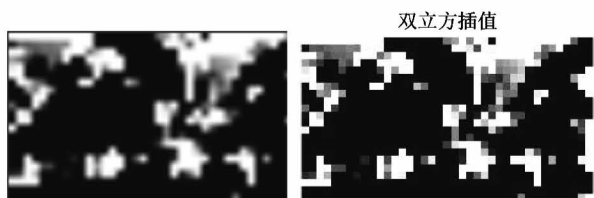


图 2 插值对比

图 2 中,左图是提取的未进行插值处理的深层特征图,右图是对左图进行双立方插值处理后的特征图,可以看到直接提取的深层特征图的分辨率非常低,在经过插值处理后,深层特征图的分辨率大大提高,对提高跟踪器的鲁棒性有很大的作用。插值的特征图的公式可以用下式表示:

的 Layer1) 提取的特征作为浅层特征和第 14 个卷积层(对应图 1 中的 Layer5) 提取的特征作为深层特征。

$$J_{c2}\{x_j^{c2}\}(t) = \sum_{n=0}^{N_{c2}-1} x_j^{c2}[n] b_d(t - \frac{T}{N_{c2}}n) + \sum_{n=0}^{N_{ch}-1} x_j^{ch}[n] b_d(t - \frac{T}{N_{ch}}n) \quad (6)$$

式(6)是对特征层进行插值的公式,和式(1)一样, $b_d(t)$ 是插值函数, $c2 \in \{1, \dots, c2, \dots, C_2\}$ 是深层特征第 $c2$ 个通道, $ch \in \{1, \dots, ch, \dots, C_{hog}\}$ 表示 HOG 特征的第 ch 个通道,那么卷积算子的公式为

$$S_f(x_j) = \sum_{c1=1}^{C_1} f^{c1} * J_{c1}\{x_j^{c1}\} + \sum_{c2=1}^{C_2} f^{c2} * J_{c2}\{x_j^{c2}\} + \sum_{ch=1}^{C_{hog}} f^{ch} * J_{ch}\{x_j^{ch}\} \quad (7)$$

式中, $J_{c1}\{x_j^{c1}\}$ 表示没有插值的浅层特征, $c1 \in \{1, \dots, c1, \dots, C_1\}$ 是浅层特征第 $c1$ 个通道, f^{c1} 表示浅层特征的第 $c1$ 个通道对应的滤波器, f^{c2} 表示深层特征的第 $c2$ 个通道对应的滤波器, f^{ch} 表示 HOG 特征的第 ch 个通道对应的滤波器,通过式(7)从每个特征图求取卷积响应和就可以获得目标最终的位置。

2.2 样本空间分类策略

第 1 节介绍了样本空间模型的生成方法,利用概率密度分布和高斯混合模型(GMM)把训练样本集的个数分成 L 个,更新 GMM 时采用 Declercq 和 Piater 提出的在线更新算法。给定一个新样本 x_j , 首先用 $\theta_m = \gamma, u_m = x_j$ 初始化一个新的分组 m , 如果所分的组数超过设定的值,则对 GMM 进行简化;如果某一分组的权重 θ_i 低于预先设置好的阈值,则用新样本替换该组,否则合并 2 个距离最近的组 k

和 l 为一组,合并后的权重和均值由下式得到。

$$\theta_n = \theta_k + \theta_l, \quad u_n = \frac{\theta_k u_k + \theta_l u_l}{\theta_k + \theta_l} \quad (8)$$

随着视频帧的不断读取,这种生成训练样本集合的策略会把一些图片从组中剔除掉,用新的一帧代替,这些被替换掉的图片与分组中的图片也存在着很高的相似性,这样直接替换掉会使在训练样本时可能会丢掉一些信息,使训练不充分,对后面进行模型更新时会产生影响。基于上面的分析和研究,本文在样本集合分组上进行了改进,把所有的图片都进行保留,不进行替换,只进行合并。后面的实验结果也验证了改进策略的有效性。

在应用高斯混合模型 (GMM) 进行训练样本集合分组时,要用到提取的所有特征层的信息进行计算来决定最终当前帧图片的分组情况。由于每种提取的特征在跟踪时对跟踪效果影响不同,简单的等权值融合在一起可能会导致不好的结果,所以应给不同的特征层分配不同的权重来提高分组的效果。不同特征层的权重分配可以用下式来表示:

$$\| \mathbf{x}_j \| = 2\delta \| \mathbf{x}_j^{sl} \| + 2(1 - \delta) \| \mathbf{x}_j^{dl} \| + 2 \| \mathbf{x}_j^{hog} \| \quad (9)$$

其中, $\| \mathbf{x}_j \|$ 是第 j 帧的 L_2 范数, $\| \mathbf{x}_j^{sl} \|$ 是第 j 帧浅层特征的 L_2 范数, $\| \mathbf{x}_j^{dl} \|$ 是第 j 帧深层特征的 L_2 范数, $\| \mathbf{x}_j^{hog} \|$ 是第 j 帧 HOG 特征的 L_2 范数,用来计算新样本与各分组之间的距离, δ 是浅层特征分配的权重。3.2 节实验将表明经不同特征层分配的权重不同确实会影响分组情况,从而影响跟踪精度。

2.3 多尺度估计滤波器

针对目标运动过程中出现的尺度变化问题,一个理想的尺度估计方法可以精确有效地对目标的大小进行检测提高跟踪准确率。判别尺度空间跟踪 (discriminative scale space tracker, DSST) 算法是一种精确有效的尺度估计方法,采用 1 维的相关滤波器进行尺度估计,称为尺度滤波器,它可以将所提出的尺度估计的方法移植到任意算法中去,以下对 DSST 尺度估计滤波器的原理进行介绍。

尺度滤波器是为适应目标尺度的变化设计的一维滤波器,用来估计目标的尺度变化,算法中尺度评估的原则为

$$a^n W \times a^n R, \quad n \in \left\{ \left[-\frac{S-1}{2} \right], \dots, \left[\frac{S-1}{2} \right] \right\} \quad (10)$$

其中, W 和 R 分别为前一帧目标的宽和高, a 为尺度因子,用来对目标大小的控制, S 为估计尺度的个数。

设输入样本 x_j 中某一图像块的维度为 D ,为了得到最佳尺度相关滤波器 f_{scale} , 最小化代价函数:

$$\varepsilon = \left\| \sum_{d=1}^D f_{scale}^d * x_j^d - g \right\|^2 + \varphi \sum_{d=1}^D \| f_{scale}^d \|^2 \quad (11)$$

其中, $*$ 表示卷积, g 为期望输出, d 表示特征的某一维度, φ 为权重系数。将其转换到复频域中采用帕斯瓦尔 (Parseval's formula) 定理求解可得:

$$F_{scale}^d = \frac{\overline{G} X_j^d}{\sum_{k=1}^D \overline{X}_j^k X_j^k + \varphi} = \frac{A_j^d}{B_j} \quad (12)$$

F_{scale}^d 、 X_j^d 、 G 为 f_{scale}^d 、 x_j^d 、 g 经过离散傅里叶变换 (DFT) 得到的对应变数, $\overline{G}$ 表示 G 的共轭转置。为了使结果更加鲁棒性,对式 (12) 中的 F_{scale}^d 的分子分母分别进行更新:

$$A_j^d = (1 - \eta) A_{j-1}^d + \eta \overline{G}_j X_j^d \quad (13)$$

$$B_j = (1 - \eta) B_{j-1} + \eta \sum_{k=1}^D \overline{X}_j^k X_j^k \quad (14)$$

其中, η 是学习率。

设 Z_j 为以上一帧预测的目标位置为中心的图像块, Z_j 为其经离散傅里叶变换得到的变量,即选定的正样本区域,在新的一帧中可通过求解离散傅里叶变换 (DFT):

$$\hat{y}_j = F^{-1} \left\{ \frac{\sum_{d=1}^D \overline{A_{j-1}^d} Z_j^d}{B_{j-1} + \varphi} \right\} \quad (15)$$

来确定尺度滤波器的响应,求得使 $\hat{y}_j$ 最大尺度响应得分作为下一帧目标尺度,并使用式 (13)、(14) 来更新尺度滤波器。

3 实验

3.1 结果分析

本文实验硬件环境为 Intel Core (TM) i7-8700K 3.70 GHz CPU, 内存 (RAM) 16 GB, 算法开发平台

Matlab 2016b。为了更好地评估跟踪方法的性能,在实验中,采用 OTB100 数据集对跟踪算法进行跟踪性能评估。首先对本文算法和 ECO 算法在跟踪速度上进行了实验对比,从 OTB100 中选择了 10 个具有代表性的视频序列对其跟踪速度进行了测试,测试结果如表 1 所示。从表 1 中可以看出,本文的方法在跟踪速度上比 ECO 算法有非常大的提高。这也说明在对提取的特征图进行分层插值的方法是可行的,可以提高跟踪速度。

表 1 部分视频跟踪速度对比结果

视频名称	本文算法 (fps/s)	ECO (fps/s)
Basketball	8.70	2.69
Bolt	10.81	3.51
Car24	11.50	3.51
Freeman1	11.29	3.61
Woman	11.50	3.47
Biker	10.88	3.50
Crossing	10.52	3.50
Panda	11.40	3.60
Football	11.01	3.52
Gym	9.14	2.72

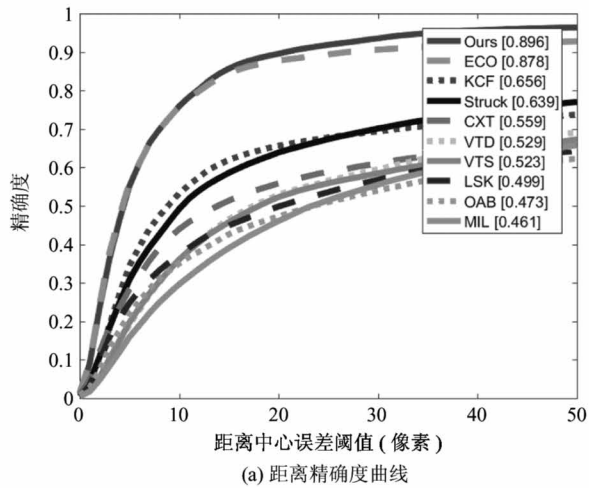


图 3 是在 OTB100 测试集上跟踪算法的成功率曲线图,其中,图 3(a)为距离精度曲线图,图 3(b)为重合度成功率曲线图。从图 3 中可以看出,在整体的跟踪效果上,本文算法(Ours)无论是在距离精度上还是在重合度成功率上都比 ECO 算法要好,而且还超过了其他目前比较流行的算法,比如核相关滤波(KCF)跟踪算法。在精确度上,本文算法(Ours)的结果为 89.6%,ECO 算法为 87.8%,提高了 2.05%,在成功率上,本文算法(Ours)的结果为 81.6%,ECO 算法为 81.2%。

本文中使用多尺度滤波器,把目标的尺度估计从 ECO 算法中的 5 个改进为 DSST 算法中提出的 33 个。图 4 是针对目标尺度变化进行的测试对比,从图 4 中可以看到,在 OTB100 测试集中,有 45 个视频序列目标的尺度是不断在发生变化的。曲线图的结果显示,本文的算法在面对目标尺度变化时跟踪的精确度为 85.2%,成功率为 78.0%,ECO 算法的跟踪精确度为 84.2%,成功率为 77.9%,可以看出,本文的算法在精确度和成功率上都比 ECO 中的尺度估计要好,也表明本文提出的尺度估计方法能较好地应对目标尺度的变化,在一定程度上同时提高了目标的精确度和成功率。

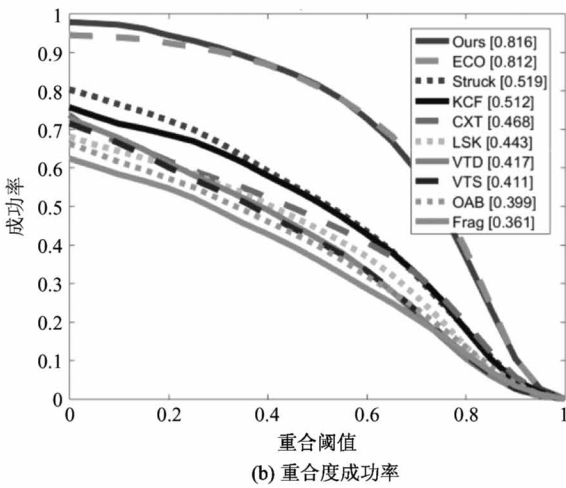


图 3 总体成功率

3.2 样本空间分类策略实验

在第 2.2 节中,对 ECO 的样本空间生成进行了改进,改进方法有 2 点:(1)不替换,所有样本都保留分到不同组中;(2)在分组时给不同特征层分配

不同的权重。针对不替换策略来生成样本空间模型,在 OTB100 上做了单独测试,测试结果如表 2 所示。为了找到最合适的权重分配,在 OTB100 上也做了大量实验,实验结果如表 3 所示。

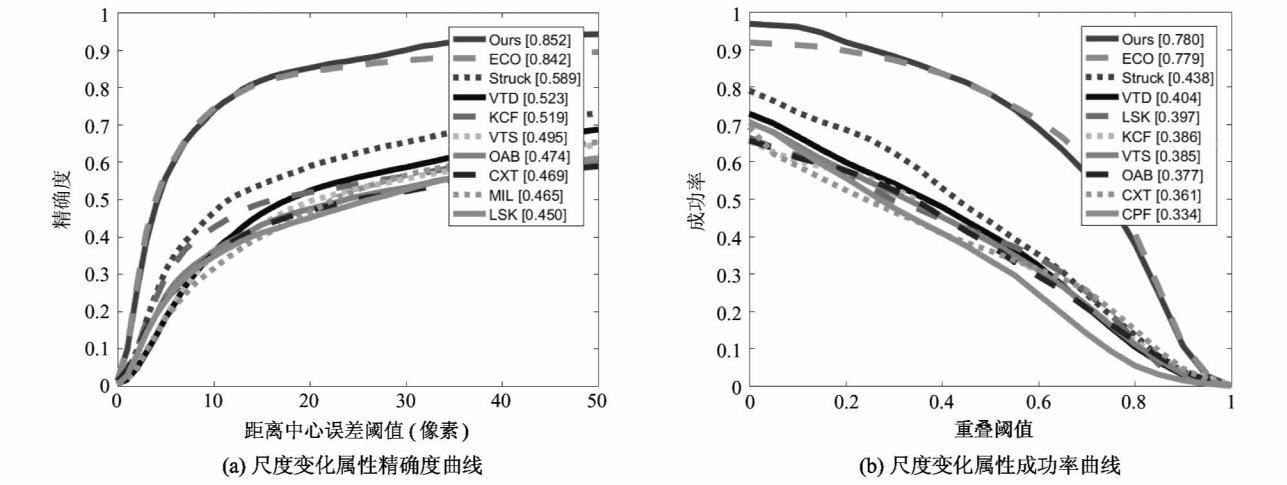


图 4 尺度变化曲线

表 2 中加粗的数据是 2 种算法中表现最佳的数据,从表 2 中可以看出,当对生成样本空间采用不替换策略时,本文的算法在快速移动、运动模糊、平面内旋转、低分辨率、遮挡、尺度变化属性下精确度和成功率都超过了 ECO 算法;在形变和平面外旋转属性下,精确度超过了 ECO,成功率略低或持平;在出视野属性下,精确度和 ECO 持平,成功率高于 ECO;但在背景复杂和光照变化下,略低于 ECO。但是从

整体的效果上看,在成功率上虽然与 ECO 持平,但是在精确度上好于 ECO 算法,说明本文提出的不替换策略在一定程度上是能够提升跟踪效果的。

从表 3 中给 CNN 提取的浅层和深层特征分配权重不同所得到的结果可以看到,当浅层特征分配的权重为 0.8,深层特征权重为 0.2 时,所得到的结果最好,其中成功率为 81.6%,高于 ECO,在精确度上为 87.6%,略低于 ECO。

表 2 不替换策略精确度与成功率对比

环境属性		快速移动	背景复杂	运动模糊	形变	光照	平面内旋转	低分辨率	遮挡	平面外旋转	出视野	尺度变化	总体
精确度	本文算法	0.886	0.850	0.826	0.907	0.866	0.851	0.851	0.887	0.885	0.883	0.851	0.880
	ECO	0.876	0.856	0.812	0.897	0.871	0.844	0.844	0.883	0.882	0.883	0.842	0.878
成功率	本文算法	0.783	0.768	0.789	0.876	0.824	0.770	0.748	0.797	0.802	0.792	0.785	0.812
	ECO	0.780	0.768	0.785	0.879	0.825	0.766	0.740	0.796	0.802	0.788	0.779	0.812

表 3 不同特征层权重分配精确度与成功率对比

浅层权重		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
本文算法	精确度	0.872	0.874	0.871	0.873	0.863	0.876	0.866	0.868	0.876	0.874	0.866
	成功率	0.807	0.812	0.811	0.812	0.803	0.811	0.804	0.810	0.816	0.812	0.809
ECO	精确度	0.878	0.878	0.878	0.878	0.878	0.878	0.878	0.878	0.878	0.878	0.878
	成功率	0.812	0.812	0.812	0.812	0.812	0.812	0.812	0.812	0.812	0.812	0.812

4 结 论

在基于深度特征加人工特征与相关滤波相结合 ECO 目标跟踪算法的基础上,对其进行了 3 方面的

改进。在跟踪速度上,采取对 CNN 提取的特征层进行分层插值的方法,对其进行了改进,改进后算法的跟踪速度比原来的速度有了很大的提高;在生成样本空间时,为了能保留训练样本更多的信息,保留了

所有图片,并且在分组时,考虑到各层权重不同会影响分组情况,即训练样本集,从而影响跟踪效果,所以给浅层和深层特征分配了不同的权重;为了更好地估计目标的尺度变化,对 ECO 的尺度估计进行了改进,把 DSST 提出的尺度估计方法引入到本文中。在 OTB100 上的实验结果表明,本文算法无论是在跟踪速度上还是在跟踪效果上,都比 ECO 算法有所提升。

参考文献

- [1] 胡硕,赵银妹,孙翔,等. 基于卷积神经网络的目标跟踪算法综述[J]. 高技术通讯,2018,28(3):207-213
- [2] Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(5): 603-619
- [3] Merwe R V D, Doucet A, Freitas N D, et al. The unscented particle filter[C]//International Conference on Neural Information Processing Systems, Massachusetts, USA, 2000:563-569
- [4] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision[C]//International Joint Conference on Artificial Intelligence, San Francisco, USA, 1981:121-130
- [5] Avidan S. Support vector tracking[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(8):1064-1072
- [6] Saffari A, Leistner C, Santner J, et al. On-line random forests[C]//IEEE International Conference on Computer Vision Workshops, Kyoto, Japan, 2009:1393-1400
- [7] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters[C]//The 23rd IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010: 2544-2550
- [8] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3):583-596
- [9] 胥方,丁信斌,占红武,等. 高精度移动目标位姿测量方法[J]. 高技术通讯,2019,29(2):119-133
- [10] Danelljan M, Khan F, Felsberg M, et al. Adaptive color attributes for real-time visual tracking[C]//IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014:1090-1097
- [11] Schölkopf B. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond[M]. Massachusetts: MIT Press, 2003: 7-61
- [12] Danelljan M, Häger G, Khan F S, et al. Discriminative scale space tracking[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 39(8):1561-1575
- [13] Bertinetto L, Valmadre J, Golodetz S, et al. Staple: complementary learners for real-time tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 1401-1409
- [14] 张立国,王松,金梅,等. 基于多尺度感知哈希特征的目标跟踪算法研究[J]. 高技术通讯,2018,28(3): 219-226
- [15] Wang N, Yeung D Y. Learning a deep compact image representation for visual tracking[C]//International Conference on Neural Information Processing Systems, Red Hook, USA, 2013: 809-817
- [16] Vincent P, Larochelle H, Bengio Y, et al. Extracting and composing robust features with denoising autoencoders[C]//Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 2008:1096-1103
- [17] Lecun Y L, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11):2278-2324
- [18] Jozefowicz R, Zaremba W, Sutskever I. An empirical exploration of recurrent network architectures[C]//International Conference on International Conference on Machine Learning, Lille, France, 2015:2342-2350
- [19] Bromley J, Bentz J W, Léon B, et al. Signature verification using a “siamese” time delay neural network[C]//International Conference on Neural Information Processing Systems, Denver, USA, 1993:669-688
- [20] Danelljan M, Bhat G, Khan F S, et al. ECO: efficient convolution operators for tracking[J]. *arXiv*: 1611.09224, 2016
- [21] Danelljan M, Robinson A, Khan F S, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking[C]//The 14th European Conference of Computer Vision, Amsterdam, Netherlands, 2016:472-

488

[22]

Wu Y, Lim J, Yang M H. Object tracking benchmark [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9):1834-1848

[23]

Battistone F, Santopietro V, Petrosino A. The visual object tracking VOT2016 challenge results [C] //The 14th European Conference on Computer Vision, Amsterdam, Netherlands,2018: 777-823

[24]

He Z, Fan Y, Zhuang J, et al. Correlation filters with weighted convolution responses [C] // IEEE International

Conference on Computer Vision Workshop, Venice, Italy, 2017:1992-2000

[25]

Bhat G, Johnander J, Danelljan M, et al. Unveiling the power of deep tracking [C] // the 15th European Conference on Computer Vision, Munich, Germany, 2018:1-22

[26]

Li F, Tian C, Zuo W, et al. Learning spatial-temporal regularized correlation filters for visual tracking [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA,2018:1-10

Improvement of CNN hierarchical interpolation and weighted strategy for ECO tracking algorithm

Chen Zhiwang, Wang Changmeng, Wang Ying, Song Juan, Peng Yong
(Key Laboratory of Industrial Computer Control Engineering of Hebei Province,
Yanshan University, Qinhuangdao 066004)
(National Engineering Research Center for Equipment and Technology of Cold Strip Rolling,
Yanshan University, Qinhuangdao 066004)

Abstract

This paper presents improvement of efficient convdution operator (ECO) which combinates depth features and correlation filters. Firstly, in order to improve the speed of tracker, ‘shallow feature no interpolation, deep feature interpolation’ is proposed for convolutional neural networks (CNN) hierarchical interpolation, i. e. the shallow features with higher resolution are not interpolated, but the deep features with low resolution are interpolated to promote resolution. Secondly, the sample space classification strategy is improved, different weights is assigned for CNN feature layer, the influence of different feature layers on the distance between samples is highlighted, and all sample information is retained in the training sample set. Finally, the discriminative scale space tracker (DSST) is used to estimate the target scale, and the number of reference target scale is increased, so the target scale is estimated accurately. The experimental results verified the effectiveness of the designed algorithm.

Key words: target tracking, efficient convdution operator (ECO), convolutional neural network (CNN), correlation filter, scale estimate