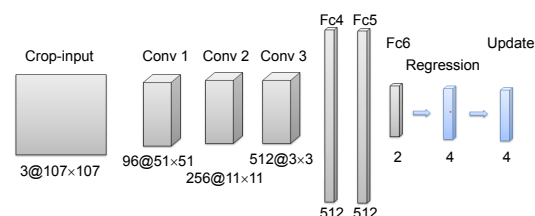


DOI: 10.12086/oe.2019.180261

基于深度学习的飞机目标跟踪应用研究

赵春梅^{1,2}, 陈忠碧^{1*}, 张建林¹¹中国科学院光电技术研究所, 四川 成都 610209;²中国科学院大学, 北京 100049

摘要: 本文针对飞机目标, 提出了基于多域网络(MDNet)的改进网络用于飞机跟踪的快速深度学习(FDLAT)跟踪网络, 使用迁移学习弥补目标跟踪的小样本集缺陷。卷积层作为特征提取层, 全连接层作为目标和背景的分类层, 采用特定的飞机数据集来更新网络参数。训练完成之后, 结合回归模型, 采用简单的线性更新对飞机进行跟踪, 算法实现了飞机旋转、相似目标、模糊目标、复杂环境、尺度变换、目标遮挡以及形态变换等复杂状态的鲁棒跟踪, 速度达到平均 20.36 f/s, 在 ILSVRC2015 飞机检测数据集上成功率均值达到 0.592, 基本满足飞机实时跟踪。

关键词: FDLAT; 迁移学习; 飞机目标; 鲁棒跟踪; 实时跟踪

中图分类号: TB872

文献标志码: A

引用格式: 赵春梅, 陈忠碧, 张建林. 基于深度学习的飞机目标跟踪应用研究[J]. 光电工程, 2019, 46(9): 180261

Application of aircraft target tracking based on deep learning

Zhao Chunmei^{1,2}, Chen Zhongbi^{1*}, Zhang Jianlin¹¹Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, Sichuan 610209, China;²University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: In this paper, based on multi-domain network (MDNet), fast deep learning for aircraft tracking (FDLAT) algorithm is proposed to track aircraft target. This algorithm uses feature-based transfer learning to make up the inferiority of small sample sets, uses specific data sets to update parameters of convolutional layers and fully connected layers, and use it to distinguish aircraft from background. After building the training model, we put the aircraft video sets into the model and tracked the aircraft using regression model and a simple line on-line update, to increase the speed while ensuring the accuracy. This algorithm achieves robust tracking for aircraft in rotation, similar targets, fuzzy targets, complex environment, scale transformation, target occlusion, morphological transformation and other complex states, and runs at a speed of 20.36 frames with the overlap reached 0.592 in the ILSVRC2015 detection sets of aircraft, basically meets the real-time application requirement of aircraft tracking.

Keywords: FDLAT; feature-based transfer learning; aircraft target; robust tracking; real-time tracking

Citation: Zhao C M, Chen Z B, Zhang J L. Application of aircraft target tracking based on deep learning[J]. *Opto-Electronic Engineering*, 2019, 46(9): 180261

收稿日期: 2018-05-17; 收到修改稿日期: 2018-10-15

基金项目: 重大专项基金资助项目(G158207)

作者简介: 赵春梅(1993-), 女, 硕士, 主要从事深度学习, 机器学习, 目标跟踪。E-mail: 841143386@qq.com

通信作者: 陈忠碧(1975-), 女, 博士, 副研究员, 主要从事运动目标检测与跟踪。E-mail: ioeyoyo@126.com

1 引言

目标跟踪是计算机领域重要的研究热点之一,应用领域广泛,包括无人机监察,空域目标跟踪,行人与车辆监控等^[1-3]。目标跟踪面临目标旋转、变化、多目标、目标模糊、环境复杂、尺度变换、目标遮挡、光照变化等问题非常困难,而确保目标跟踪的鲁棒性、准确性以及实时性则极具挑战性。传统的目标跟踪方法大多需要人工提取特征,如采用核方法结合直方图特征^[4]具有较好的准确性,使用均值法结合金字塔特征^[5]具有较好的鲁棒性,而核相关滤波(kernel correlation filter, KCF)^[6]采用相关特征,实时性很好,但是在目标发生尺度变化以及遮挡等情况时会跟丢目标。

传统手工特征的弊病在于无法提取更高层次的语义信息,对于手工提取的特定场景特征无法适应普遍场景,泛化能力较差^[7-8]。由此引入深度学习来解决特征提取遇到的问题,深度学习在图像分类^[9-11]、语义分割^[12]、目标检测与识别^[13]等方面取得显著成效,但是在目标跟踪方面,却因样本集较小等受到限制。鉴于此,提出基于深度学习的飞机目标跟踪算法,通过多层网络特征的提取与特定目标的迁移学习,使算法在实时性和准确性上都取得了较好的效果,同时对于尺度变换、遮挡和模糊等状态具有较好的鲁棒性,有着较好的应用价值。

2 深度学习研究现状

基于深度学习的目标跟踪网络研究越来越多,一般采用的是离线训练结合在线更新模板,如孪生网络(Siamese-fc)^[14],通过提取跟踪对象的模板与跟踪当前帧的候选区域,并且通过卷积层提取目标,再将模板和候选区的特征进行相关操作得到一个目标的得分图,将得分图进行简单的线性插值得到跟踪目标所在帧中的位置,而对于模板的在线更新则采用当前帧位置与前一帧位置进行简单的线性处理。而端到端网络相关滤波网络(correlation filter network, CFNet)^[15]则是基于 Siamese-fc 的改进网络,网络还是采用模板特征和候选区域特征进行相关,差别在于相关操作是在傅里叶域进行,有效地提高了实时性,但在精度方面效果不是很显著,同时也存在遮挡状态跟丢目标的问题。

而在精度上面表现优异的多域网络(multi-domain network, MDNet)^[16],也是采用离线训练在线更新,训练过程主要是更新能提取多类目标特征的卷积网络,对于不同视频目标采用相同的卷积层,不同的全

连接层,而在跟踪时则只更新全连接层,固定卷积层参数,更新策略较复杂,网络设置了长时更新和短时更新两种更新模式,精度高,实时性很差。采用离线模型进行跟踪的典型网络就是 GOTURN^[17],将上一帧结果作为模板与候选区域进行操作,模板和候选区的卷积特征进行简单的拼接操作送入全连接网络,网络输出即为目标位置,实时性好,达到 100 f/s,但是精度相较于 MDNet 相差很多,对于目标运动速度较快或者尺度变化较大的情况跟踪效果不好。

本文主要针对特定的应用研究,提出了基于 MDNet 改进的用于飞机跟踪的快速深度学习(fast deep learning for aircraft tracking, FDLAT),将目标跟踪方法应用在飞机目标跟踪领域。FDLAT 通过迁移视觉几何组网络(visual geometry group networks, VGGNets)^[18]的部分卷积层用于提取飞机类目标的特征,通过全连接层作为分类层,区分目标和背景,最后通过一个回归网络进行位置拟合,对于拟合后的候选框进行一个简单的线性更新操作。该方法利用迁移学习解决了跟踪的小数据集缺陷,提高了跟踪的鲁棒性,同时基本满足实时性要求,精度不低于 MDNet,具有较强的应用价值。

3 FDLAT 跟踪网络

MDNet 跟踪网络在准确性上表现非常好,但是在速度上却是一大缺陷。基于深度学习的目标跟踪网络需要在实时性和准确性上做出权衡。FDLAT 则是基于 MDNet 改进的飞机目标跟踪应用网络,在保证精度的情况下基本满足实时跟踪是 FDLAT 最大的优点。

3.1 基础网络 MDNet

MDNet^[14]跟踪结构简单,由 Conv1~Conv3 三个卷积层提取目标特征, Fc4 和 Fc5 两个全连接层加上 ReLUs 和 Dropouts,最后的二分类层 Fc6 是 MDNet 的特色之处:在训练过程中,不同的视频序列输入,得到不同的 Fc6,即 Fc6 层更新只对应相应的视频序列,这样做的原因是提高了卷积层提取目标特征的统一性。跟踪时则固定卷积层参数,更新 Fc4~Fc6 层参数,更新分为长时更新和短时更新;Fc6 层的输出对应目标和背景的概率,根据得分概率最高的五个图片候选框对应的回归框确定为目标。

3.2 FDLAT 改进网络

对于本文飞机目标跟踪应用,提取的特征就是飞机类目标的特征。本文提出的 FDLAT 网络的前馈网

络通过训练得到,在训练过程中使用 VGGNets 的 3 个卷积层 Conv1~Conv3 用于提取飞机类目标特征,修改全连接层 Fc6 为单一层,使用 Fc4~Fc6 用于飞机与背景的二分类,并输出飞机与背景的概率。对于卷积层的选择,考虑因素有计算量和特征充分性,低层卷积层(1~2 层)提取图像的线特征和边缘特征,属于底层信息,而高层卷积层提取图像的语义信息,属于高层特征,在目标检测和目标识别等操作中一般卷积层较深,用于提取足够的特征,而对于卷积层越深的网络,计算量越大,实时性则越差。对于本文的目标跟踪课题,由于只需要判断目标和背景,即对于目标的特征要求不高。本文选择三层卷积层,前两层用于提取目标边缘信息,第三层用于提取目标全局信息,增加特征的感受野。当然如果选择更深层次的卷积层,目标的特征表达能力将更强,在精度上还可以提升,但是增加了计算复杂度,本文卷积层数的选择考虑到了特征的充分性和计算复杂度。

对于跟踪过程,则采用训练好的网络作为前馈网络,将网络输出的最大得分对应的候选框经过回归网络得到目标位置,而在线更新则采用简单的线性操作完成。改进总结为以下四点:

- 1) 修改全连接层 Fc6 为单一域,使用特定的飞机视频来更新网络参数,卷积层 Conv1~Conv3 提取飞机类目标特征,全连接层进行飞机和背景二分类操作。这个改进提升了训练速度,以及收敛性,不影响精度。
- 2) 在线更新的时候不更新全连接层参数,保持卷积层和全连接层参数固定,不占用计算内存,只进行前馈网络计算,这个改进提高了跟踪速度。
- 3) 添加新的线性更新规则,最终位置由回归位置和上一帧位置加权决定,添加一个超参数,经过跟踪过程测试,最终确定为 0.59,这个改进措施提高了跟

踪精度,但对于中心稳定性有一定影响。

- 4) 跟踪候选框的选取改进,限定跟踪框选取位置为上一帧目标框 2x2 倍,缩小选取框的数量,减少计算量,提升跟踪速度。

3.2.1 FDLAT 网络结构

前馈网络结构如图 1 所示,灰色表示前馈网络,蓝色为跟踪时增加的更新相关操作,表示对网络输出进行回归和更新。

训练过程为灰色的前馈网络,训练得到卷积层和全连接层参数。网络的输入为预处理后的图片,通过对原始图片进行候选框裁剪以及尺寸变换后得到 107x107x3 的彩色图像,经过卷积层特征提取,最后经过全连接层得到网络输出维度为 2。跟踪过程需要添加蓝色的回归网络和线性更新操作,具体实施将在 3.3.2 节叙述。

FDLAT 前馈网络每一层的操作参数以及操作结果如表 1 所示,包括输入,卷积尺度,步进以及输出。

在全连接层后加入 Dropout 层,防止过拟合现象,而 LRN 局部响应归一化操作是为了增加网络的泛化能力。

3.2.2 FDLAT 算法

FDLAT 网络的作用是将输入分为两类:飞机和背景。对于二分类问题,飞机类目标特征提取网络不用太复杂,三层卷积网络足够^[14,17,19-21],同时深层的卷积网络导致位置信息丢失,这是在跟踪领域不愿意看到的。本文通过迁移学习获取初始卷积层,使用预训练网络 VGGNets 的 Conv1~Conv3 层,在跟踪视频序列中,目标一般不会太大,故输入的图片尺寸设定为 107x107,这样可以得到一个 3x3 的特征图,同时降低了网络的深度。在测试其他网络结构时发现精度和实时性变得差一些,同时出现飞机目标漂移和丢失现象。

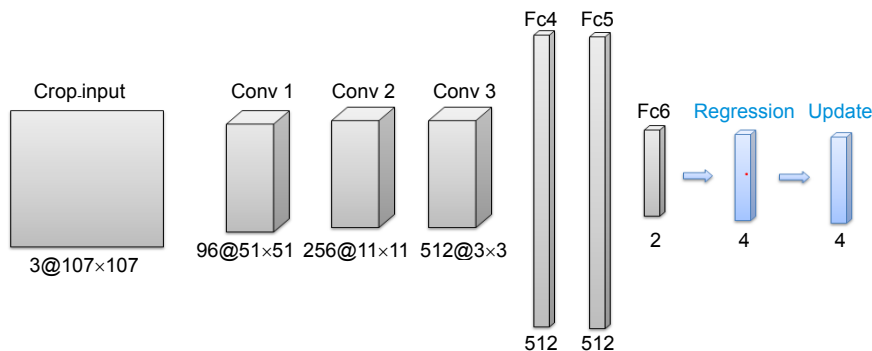


图 1 FDLAT 前馈网络

Fig. 1 Feedforward network of FDLAT

表 1 FDLAT 网络操作以及结果
Table 1 Operation and results of FDLAT network

Operation	Input	Fiter_size	Strides	Output
Conv1	3@107×107	96@7×7	2	96@51×51
ReLU, LRN				
Max_pooling	96@51×51	96@3×3	2	96@25×25
Conv2	96@25×25	256@5×5	2	256@11×11
ReLU, LRN				
Max_pooling	256@11×11	256@3×3	2	256@5×5
Conv3	256@5×5	512@3×3	1	512@3×3
ReLU				
Fc4, Dropout	512×3×3			512
ReLU				
Fc5, Dropout	512			512
ReLU				
Fc6	512			2

为了节省参数的调试时间，本文网络训练参数初始值采用 MDNet 的训练参数，经过参数调试，最终确定使用随机梯度下降法(stochastic gradient descent, SGD)^[16,22-23]进行训练，动量(momentum)为 0.9，权值衰减(weight decay)为 0.0005，Conv1~Conv3 的学习速率(learning_rate)为 0.0001，Fc4~Fc6 的学习速率为 0.001。

使用判别法来获得正负样本的逻辑损失(Loss, 用 S_{loss} 表示)， y 为负样本的真实标签， f_n 表示负样本对应的负样本得分， f_p 表示正样本对应的正样本得分，通过 SGD 进行优化从而得到网络参数。

$$l(y, f_n) = \log(1 + \exp(-y \cdot f_n)) \quad (1)$$

$$l(y, f_p) = \log(1 + \exp(-y \cdot f_p)) \quad (2)$$

$$S_{loss} = \sum_{\chi^{neg}} \log(1 + \exp(-y \cdot f_n)) + \sum_{\chi^{pos}} \log(1 + \exp(-y \cdot f_p)) \quad (3)$$

$$\arg \min_{(W)} (S_{loss}) \quad (4)$$

训练集来自于 ILSVRC2015 的检测数据集当中的飞机类，共有 134 个训练视频，训练次数为 100 次。对于每一帧图片，选取交并比(IOU)大于 0.7 正样本 50 个，通过均匀分布选取 IOU 小于 0.3 的负样本 200 个，每个视频选取 8 张图片进行训练，从 400 个正样本中选取 32 个正样本，从 1600 个负样本中选取 96 个负样本，每批次的输入样本为 128。输入经过卷积层之后再经过全连接层，得到一个表示飞机和背景分类结果的二维输出。

算法 1 给出了 FDLAT 算法的前馈网络训练过程，采用伪代码方式给出了网络的结构，用于训练的数据，损失计算参数，卷积层以及全连接需要更新的权值(weight)以及更新速率(update rate)，循环次数(cycle times)，循环过程计算总次数(times)以及网络的输出结果等。

算法 1 FDLAT 训练过程算法流程

Net: pretrained Conv1~Conv3 filters $\{w_1, w_2, w_3\}$, Fc4~Fc6 filters $\{w_4, w_5, w_6\}$
Data: negative samples with neg=1, positive samples with pos=1
Mini-batch: 128 samples with negative 96 and positive 32
Loss: SGD with momentum=0.9 and weight_dacay=0.0005
Udataaweght: $\{w_1, w_2, w_3\}$, with learning_rate =0.0001, $\{w_4, w_5, w_6\}$ with learning_rate=0.001
Loop: loop_time=100, video_number=134, running_time=13400
Output: neg= f_n and pos= f_p

3.3 跟踪

3.3.1 图片候选框

将训练好的网络固定网络参数，作为前馈网络用于飞机类目标跟踪。跟踪过程中输入的视频序列，以上一帧的目标位置 $g_t[i-1]$ 为中心，采用多维高斯在长宽以及尺度三个维度选取 N 个交并比(IOU)大于 0.6 的候选框 x_1, \dots, x_n ，并且输入图片修改尺寸为 $107 \times 107 \times 3$ 输入 FDLAT 网络，网络的输出为 2 维向量 $f_p(x_i)$ 和 $f_n(x_i)$ ， $f_p(x_i)$ 代表飞机的概率， $f_n(x_i)$ 代表背景概率，选取飞机概率最高的候选框为回归网络的输入框。

$$x_{\max} = \max_{x_i} f_p(x_i) \quad (5)$$

3.3.2 回归网络与在线更新

候选框并不是直接作为目标，还要经过回归网络，做法与 R-CNN^[11] 一样，是常规算法。对于每一个跟踪视频，第一帧目标所在位置已知，将第一帧的候选框、第一帧对应的网络输出以及第一帧目标所在的实际位置作为回归网络训练的数据，训练出来的回归网络用于预测视频中目标所在的框。在跟踪测试视频序列的第一帧图片上选取 800 个样本，样本选取规则和跟踪候选框的选取规则一致，将这 800 个样本作为回归网络的训练数据集，测试视频的第一帧的目标真实位置作为回归训练的标签，训练好的回归器可直接用于该视频后续帧的回归定位操作。

回归算法的具体操作的训练流程如算法 2 所示。

算法 2 回归算法流程

Regression net: liner-regression
 Input of train: convolution feature of FDLAT as X , 800 candidate boxes as $bbox$ and ground-truth as g_t (come from the first frame)
 Error: computing the center error and width-height from g_t and $bbox$ error as Y
 Train: using feature X and error Y to train the liner-regression net
 Input of predict: convolution feature of FDLAT as X , candidate boxes as $bbox$ of objects
 Output of predict: ground-truth of objects

算法 3 FDLAT 跟踪过程算法流程

Net: fixed $\{W_1, W_2, W_3, W_4, W_5, W_6\}$
 Data: 32 candidate boxes with Gaussian distribution for every frame
 Regression: 800 candidate boxes with Gaussian distribution for first frame and output 1 box as plane position for training. Put x_{\max} into regression net and get a regression_box
 Online updata: $g_t[i]=0.59 \times regression_box + 0.41 \times g_t[i-1]$
 Output : the position of plane g_t

候选框经过回归网络得到飞机的回归位置 r_box ，对回归框进行线性更新。其目的为：1) 替代 MDNet 中的网络参数在线微调；2) 起到时序上的记忆作用，即当前帧的框具有历史帧框的作用，而历史帧框对当前帧框的影响按照帧数往前依次指数下降。更新系数 $(1-m)$ 为超参数，本文设置记忆参数 $m=0.59$ ，当前帧的目标位置为 $g_t[i]$ ，那么在线更新公式可表示为

$$g_t[i] = (1-m) \times r_box + m \times g_t[i-1] \quad (6)$$

这样的更新对于遮挡目标有一个更好的预测，即不进行在线微调网络参数时，能够根据历史帧的记忆估计出一个当前被遮挡目标的大致位置，当遮挡结束，能够找到目标的大致位置。

算法 3 给出了跟踪过程的算法流程，包括前馈网络，回归网络的训练和使用，以及跟踪过程的在线更新，输出飞机目标的位置。

4 实验

硬件实验环境：Intel Core i7-6700 CPU@3.40 GHz×8，GeForce GTX 1080GPU 软件实验环境：Linux Ubuntu 16.04，python 3.5，Pytorch3.0 为检验本文算法的实时性和准确性，将对多个飞机视频进行测试，测试集为 ILSVRC2015 检测数据集里面的 17 个飞机数据集，其中测试视频序列中包含飞机的尺度变换，旋转，

运动方向改变, 干扰因素, 隐形, 遮挡, 光照变化等多个复杂情况, 有利于验证 FDLAT 算法的实用价值。

4.1 评价参数

为了验证 FDLAT 算法的良好性能, 将测试集同时经过 FDLAT 算法和 MDNet 进行测试。本文使用三个通用的评价参数进行评价:

1) 重叠(Overlap)

Overlap 表示跟踪成功率, 其定义:

$$o_{\text{overlap}} = \frac{R \cap G}{R \cup G}, \quad (7)$$

其中: R 表示跟踪结果, G 表示标识的真实位置, 当 $\text{Overlap} > 0.34$ 可认为跟踪完成, $\text{Overlap} > 0.5$ 则表示跟踪成功^[24]。

2) 每秒帧数(FPS)

每秒帧数(Frame per second, FPS)表示跟踪速度, 而 12 FPS 是连贯图片的最低标准, 20 FPS 是 RPG 游戏运行的最低标准。FPS 的定义:

$$F = \frac{l}{t}, \quad (8)$$

其中: t 表示视频跟踪时间, l 表示视频中的帧数^[25-26]。

3) 跟踪中心误差(CLE)

跟踪中心误差(Central location error, CLE)的定义为

$$E = \sqrt{(x_R - x_G)^2 + (y_R - y_G)^2}. \quad (9)$$

其中: (x_R, y_R) 表示跟踪的中心坐标, (x_G, y_G) 表示标识的中心坐标^[24]。

4.2 定性分析

首先, 在不同状态下对飞机视频序列进行测试, 并将 FDLAT 算法的跟踪结果与 MDNet 进行比较。在定性分析图中, 红色框表示跟踪结果, 绿色框表示目标检测的标签。黄色图像命名是 ILVRC2015 中测试视频序列的测试方法和视频帧数。

测试是在具有挑战性的视频序列中进行, 例如尺度变换、快速运动、燃烧、模糊小目标、目标消失、复杂背景和目标遮挡, 结果如图 2 所示。定性分析了 FDLAT 的跟踪性能, 具有良好的准确性和鲁棒性。

图 2(a)中第 195 帧和第 849 帧处于不同尺度, FDLAT 在尺度适应性和跟踪中心误差上的效果优于 MDNet;

图 2(b)中第 44 帧和第 107 帧是飞机下降的不同位置, 在第 44 帧上, FDLAT 的中心距离误差优于

MDNet, 并且 FDLAT 的第 107 帧上重叠率也略优于 MDNet;

图 2(c)显示飞机燃烧变形的结果, 第 44 帧和第 107 帧处于正常和烧毁, 而 FDLAT 在两个场景中表现更好;

图 2(d)中飞机多次改变方向, 并且在第 147 帧和第 234 帧中改变了方向, FDLAT 在精度和稳定性方面表现良好, MDNet 效果略差于 FDLAT;

图 2(e)中出现了模糊的多个小目标, 两种方法都具有良好的性能, FDLAT 第 2 帧的效果略好于 MDNet;

图 2(f)中飞机隐形, 从第 37 帧消失到第 41 帧出现, 与 MDNet 相比, FDLAT 算法的中心距离误差具有更好的效果;

图 2(g)表明, 这两种方法在复杂环境下的飞机目标跟踪中具有良好的抗干扰性能;

图 2(h)显示的是对遮挡的效果, 图中飞机从第 498 帧开始出现遮挡, 再到完全遮挡, 最后第 513 帧遮挡消失, 可以看出 FDLAT 算法能够在遮挡情况下找到目标位置, 且优于 MDNet。

4.3 定量分析

这里采用 4.1 中定义的 Overlap、FPS、CLE 作为定量评价标准, 测试集 ILSVRC2015 检测数据集具有各种飞行环境的 17 个视频, 如图 3 所示为测试视频的定量评价指标柱状图。图 3(a)为跟踪准确率对比, FDLAT 与 MDNet 的 Overlap 效果相当, 整体效果 FDLAT 较 MDNet 略好; 图 3(b)为速度对比, FDLAT 相较于 MDNet 有较大提升, 跟踪速度提升接近 10 倍, 这就是对 MDNet 的最大改进, 提高跟踪速度, 基本达到实时跟踪要求; 图 3(c)稳定性比较, FDLAT 与 MDNet 整体差别不大, 大多数视频跟踪较为平稳, 但是少数视频出现跟踪不稳定现象。从图中分析可知, 对于同一个跟踪视频 FDLAT 与 MDNet 的跟踪准确率效果相当, 整体效果 FDLAT 较 MDNet 略好。

通过柱状图分析大致可以对比 FDLAT 和 MDNet 在指定评价参数上效果, 表 2 给出了测试视频序列在两种算法上的评价参数值。从表 2 中可以看出, 在跟踪精度上, FDLAT 的平均跟踪精度率高于 MDNet, 在跟踪中心错误率上, MDNet 稍低于 FDLAT, 但是在跟踪速度上, FDLAT 通过改进, 从 MDNet 的 2.08 帧提升到 20.36 帧, 基本满足实时跟踪。

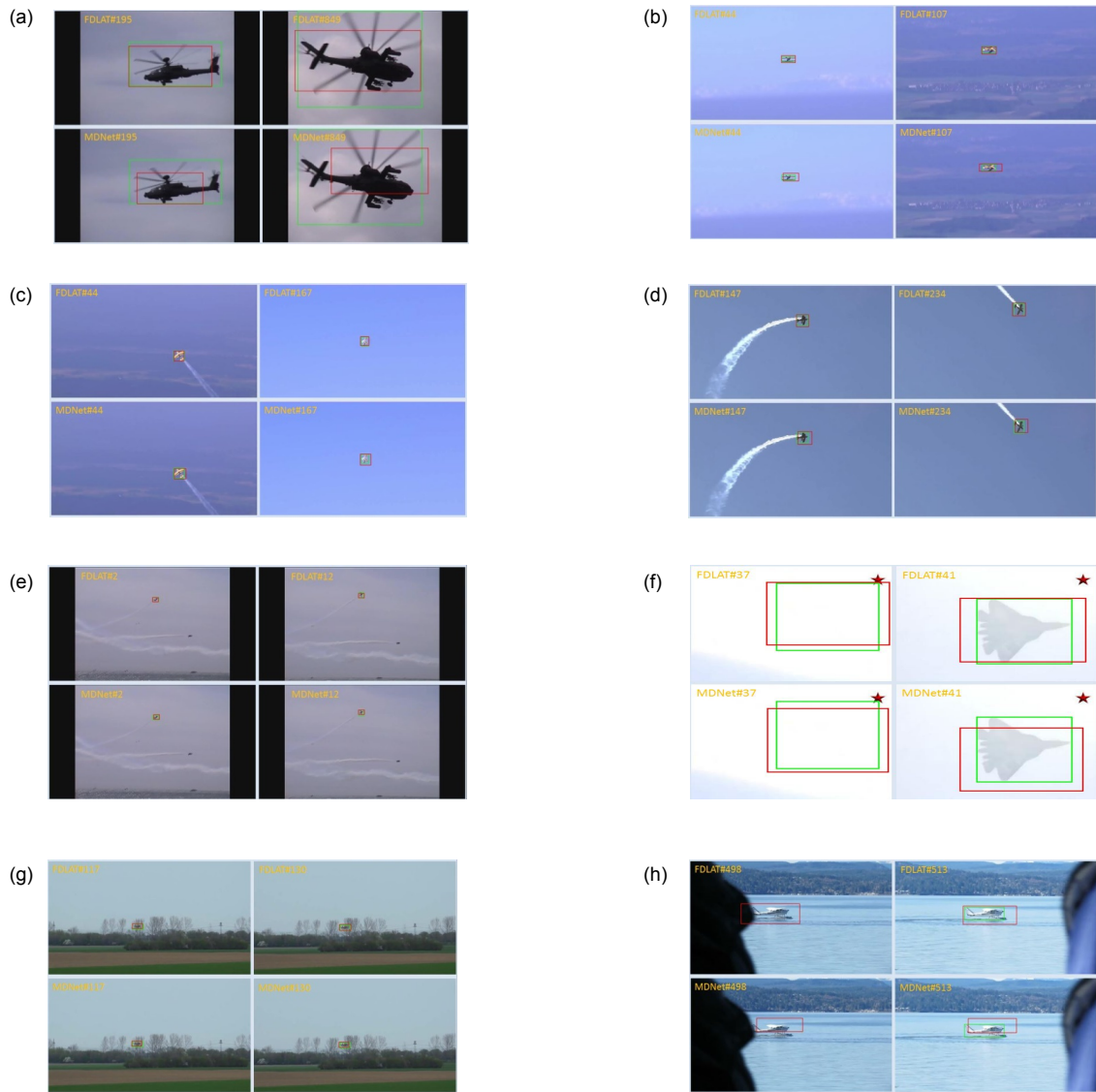


图 2 FDLAT 与 MDNet 定性对比。(a) 尺度变换; (b) 快速下降; (c) 形态变换; (d) 方向变换; (e) 模糊小目标; (f) 隐身消失; (g) 复杂背景; (h) 遮挡消失

Fig. 2 Qualitatively comparison between FDLAT and MDNet. (a) Scale transformation; (b) Rapid decline; (c) Morphological transformation; (d) Directional transformation; (e) Blurred small targets; (f) Stealth disappearance; (g) Complex background; (h) Occlusion disappearance

通过以上分析可知，对于特定的飞机目标跟踪，FDLAT 较 MDNet 在精度上略微提升，在速度上大幅度提升，在稳定性上略微下降，改进效果明显，具有较好的应用意义。

5 总结

本文提出的算法是在 MDNet 上进行改进的一种飞机目标跟踪应用研究，在多个环境下的测试集上的实验证明，该方法在飞机目标跟踪应用中具有较好的鲁棒性，基本满足实时性要求，并且有较高的准确性。

本算法采用了卷积层进行特征提取，全连接层进行飞机和背景分类，然后对网络输出进行回归和位置更新，在测试过程中对于尺度变化、遮挡、隐形、干扰等有良好的表现，同时弥补了 MDNet 速度上面的不足，达到 20.36 帧的速度，基本满足实时要求，可用于飞机目标跟踪应用。

本文所提出的方法还有进一步的优化和提升空间，如考虑利用相似性来确定跟踪目标，在保证精度的情况下进一步加快跟踪速度以及跟踪稳定性。

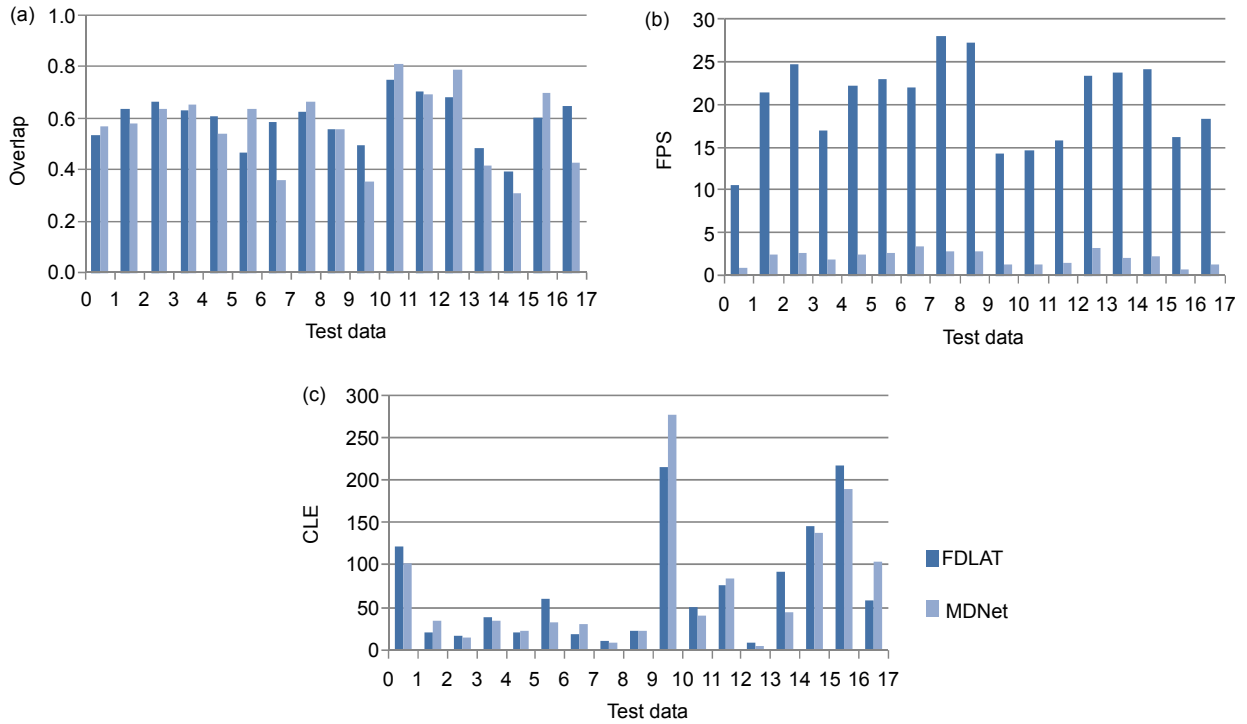


图3 定性评价指标分析。(a) 跟踪成功率; (b) 跟踪帧率; (c) 中心误差
 Fig. 3 Qualitative analysis of evaluation index. (a) Overlap; (b) FPS; (c) CLE

表2 FDLAT 与 MDNet 的定量分析结果

Table 2 Quantitative analysis results of FDLAT and MDNet

Sequences	MDNet			FDLAT		
	Overlap	FPS	CLE	Overlap	FPS	CLE
0/0030004	0.566	0.90	102.62	0.534	10.64	122.755
0/0034004	0.579	2.52	33.777	0.636	21.41	20.466
0/0034009	0.636	2.61	13.549	0.663	24.77	16.612
0/0034014	0.650	1.87	33.85	0.629	16.85	38.619
0/0034019	0.539	2.43	22.368	0.609	22.21	20.521
0/0034023	0.634	2.60	31.659	0.468	22.90	59.312
0/0117004	0.357	3.37	30.684	0.587	21.98	18.592
0/0117019	0.666	2.72	7.505	0.623	27.96	10.787
0/0117041	0.556	2.89	22.420	0.556	27.11	23.200
1/0259029	0.355	1.26	277.460	0.497	14.14	215.733
1/0321003	0.809	1.35	40.024	0.751	14.60	49.550
2/0473003	0.693	1.44	83.954	0.701	15.75	76.477
2/0555003	0.788	3.22	5.050	0.680	23.38	8.635
2/0743004	0.418	2.05	44.799	0.485	23.74	92.193
3/0939002	0.306	2.26	138.078	0.395	24.12	146.211
3/1054001	0.698	0.66	189.850	0.599	16.24	218.183
3/1099003	0.425	1.30	103.770	0.647	18.33	57.920
Mean	0.569	2.08	69.495	0.592	20.36	70.339

参考文献

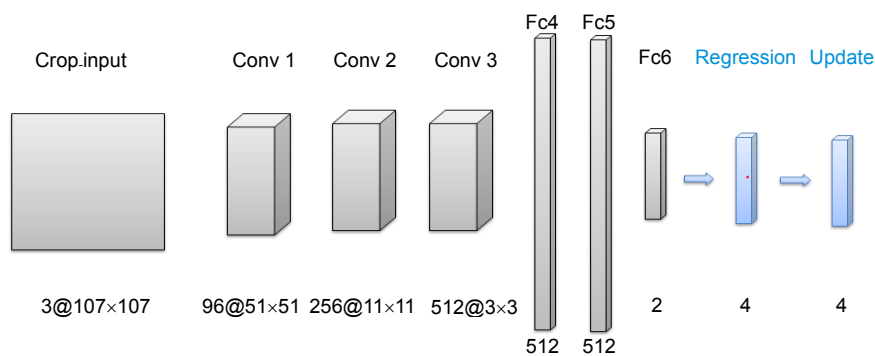
- [1] Sivanantham S, Paul N N, Iyer R S. Object tracking algorithm implementation for security applications[J]. *Far East Journal of Electronics and Communications*, 2016, **16**(1): 1–13.
- [2] Kwak S, Cho M, Laptev I, et al. Unsupervised object discovery and tracking in video collections[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*, 2015: 3173–3181.
- [3] Luo H B, Xu L Y, Hui B, et al. Status and prospect of target tracking based on deep learning[J]. *Infrared and Laser Engineering*, 2017, **46**(5): 502002.
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): 502002.
- [4] Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(5): 564–577.
- [5] Jia X, Lu H C, Yang M H. Visual tracking via adaptive structural local sparse appearance model[C]//*Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012: 1822–1829.
- [6] 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.
- [7] Fan X S, Xu Z Y, Zhang J L. Dim small target tracking based on improved particle filter[J]. *Opto-Electronic Engineering*, 2018, **45**(8): 170569.
樊香所, 徐智勇, 张建林. 改进粒子滤波的弱小目标跟踪[J]. *光电工程*, 2018, **45**(8): 170569.
- [8] Xi Y D, Yu Y, Ding Y Y, et al. An optoelectronic system for fast search of low slow small target in the air[J]. *Opto-Electronic Engineering*, 2018, **45**(4): 170654.
奚玉鼎, 于涌, 丁媛媛, 等. 一种快速搜索空中低慢小目标的光电系统[J]. *光电工程*, 2018, **45**(4): 170654.
- [9] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//*Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012: 1097–1105.
- [10] Chatfield K, Simonyan K, Vedaldi A, et al. Return of the devil in the details: delving deep into convolutional nets[J]. arXiv:1405.3531 [cs.CV], 2014.
- [11] Hyeonseob N, Mooyeol B, Bohyung H. Modeling and Propagating CNNs in a Tree Structure for Visual Tracking[J]. arXiv:1608.07242v1[cs.CV], 2016:1–10.
- [12] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(4): 640–651.
- [13] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//*Proceedings of 2014 Conference on Computer Vision and Pattern Recognition*, 2014: 580–587.
- [14] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional Siamese networks for object tracking[C]//*Proceedings of 2016 European Conference on Computer Vision*, 2016: 850–865.
- [15] Valmadre J, Bertinetto L, Henriques J F, et al. End-to-end representation learning for Correlation Filter based tracking[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [16] Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 4293–4302.
- [17] Held D, Thrun S, Savarese S. Learning to track at 100 FPS with deep regression networks[C]//*Proceedings of the 14th European Conference on Computer Vision*, 2016: 745–765.
- [18] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv:1409.1556[cs.CV], 2014.
- [19] Chen K, Tao W B. Once for all: a two-flow convolutional neural network for visual tracking[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, **28**(12): 3377–3386.
- [20] Leal-Taixé L, Canton-Ferrer C, Schindler C. Learning by tracking: Siamese CNN for robust target association[C]//*Proceedings of 2016 Computer Vision and Pattern Recognition Workshops*, 2016: 418–425.
- [21] Tao R, Gavves E, Smeulders A W M. Siamese instance search for tracking[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 1420–1429.
- [22] Wang N Y, Li S Y, Gupta A, et al. Transferring rich feature hierarchies for robust visual tracking[J]. arXiv:1501.04587 [cs.CV], 2015.
- [23] Zhai M Y, Roshtkhari M J, Mori G. Deep learning of appearance models for online object tracking[J]. arXiv:1607.02568 [cs.CV], 2016.
- [24] Wang H Y, Yang Y T, Zhang Z, et al. Deep-learning-aided multi-pedestrian tracking algorithm[J]. *Journal of Image and Graphics*, 2017, **22**(3): 349–357.
王慧燕, 杨宇涛, 张政, 等. 深度学习辅助的多行人跟踪算法[J]. *中国图象图形学报*, 2017, **22**(3): 349–357.
- [25] 王晓冬. 视觉角度对游戏可玩性的影响[J]. *河南科技*, 2014(7): 12.
- [26] Horikoshi K, Misawa K, Lang K. 20-fps motion capture of phase-controlled wave-packets for adaptive quantum control[C]//*Proceedings of the 15th International Conference on Ultrafast Phenomena XV*, 2006: 175–177.

Application of aircraft target tracking based on deep learning

Zhao Chunmei^{1,2}, Chen Zhongbi^{1*}, Zhang Jianlin¹

¹Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, Sichuan 610209, China;

²University of Chinese Academy of Sciences, Beijing 100049, China



Feedforward network of FDLAT

Overview: Deep learning has achieved good results in image classification, semantic segmentation, target detection and recognition. However, it is still restricted by small sample training sets on object tracking. Object tracking is one of the most important research topics in the field of computer vision, and has a wide range of applications. The challenge of object tracking lies in the complex states such as the target rotation, multi target, blur target, complex background, size change, target occlusion, fast moving and so on. In this paper, based on multi-domain network (MDNet), fast deep learning for aircraft tracking (FDLAT) algorithm is proposed to track aircraft target. This algorithm uses feature-based transfer learning to make up the inferiority of small sample sets, uses specific data sets to update parameters of convolutional layers and fully connected layers, and use it to distinguish aircraft from background. After building the training model, we put the aircraft video sets into the model and tracked the aircraft using regression model and a simple line on-line update, to increase the speed while ensuring the accuracy. This algorithm achieves robust tracking for aircraft in rotation, similar targets, fuzzy targets, complex environment, scale transformation, target occlusion, morphological transformation and other complex states. FDLAT is designed for lifting the speed while guaranteeing the precision of tracking. For the application of aircraft tracking, the FDLAT networks is trained by using 3 convolutional layers (Conv1~Conv3 of VGGNets) to extract the feature of aircraft target. Fc6 is a single layer and Fc4~Fc6 are used for the two classifications of aircraft and background, and the outputs are the probability of aircraft and background. In the process of tracking, the trained networks are used as feed-forward networks, and the candidate box of the maximum score of outputs is regressed to get the target location, while on-line updating is done by a simple linear operation. Our FDLAT algorithm is robust in aircraft target tracking, and basically meets the real-time requirements with high accuracy. This algorithm uses convolutional layers for feature extraction and fully connected layers for classification. Then the outputs of the networks perform a regression and location update operation in the testing process, which has good performance for scaling, occlusion, stealth, interference scene and covers the shortage of MDNet. A speed of 20.36 frames with the overlap reached 0.592 is achieved in the ILSVRC2015 detection sets of aircraft, basically meets the requirement of real-time for aircraft target tracking application.

Citation: Zhao C M, Chen Z B, Zhang J L. Application of aircraft target tracking based on deep learning[J]. *Opto-Electronic Engineering*, 2019, 46(9): 180261

Supported by Major Special Fund (G158207)

* E-mail: ioeyoyo@126.com