基于高效通道注意力机制的 FairMOT 多目标跟踪

张慧旺 [1]
ZHANG Huiwang

摘 要 多目标跟踪是计算机视觉领域中的一个重要且热门的任务。针对在真实复杂场景中目标的漏检以及 ID 匹配不准确的问题，提出一种基于 FairMOT 算法的改进算法。通过引入双分支高效注意力机制模块即 DMECA，分别加强检测与重识别分支的特征，以解决多任务训练平衡问题。优化分支头的结构，将传统卷积方式修改为深度可分离卷积，并采用 LeakyRelu 激活函数。在数据关联模块的第二阶段匹配中，使用距离交并比（DIOU）替代交并比（IOU）计算代价矩阵进行匹配。实验结果表明，在 MOT17 数据集上 IDS 下降了 625，此外 HOTA、IDF1 分别提高了 0.3%、0.4%。

关键词 多目标跟踪；通道注意力机制；深度可分离卷积；DIOU

doi：10.3969/j.issn.1672-9528.2024.01.019

0 引言

多目标跟踪可以按照跟踪框的初始化方式分为两类，一类是 Decetion-Based-TrackingDBT），另一类是 Decetion-Free-Tracking（DFT）。DBT 的方式是先检测出需要跟踪的目标，然后使用数据关联等方式形成轨迹，而 DFT 是当有新的目标出现的时候，采用人工操作的方式，告知哪一个是需要跟踪的对象，这样的缺点是存在过多的交互，当目标较多时是不可接受的，所以目前的主流仍然是 DBT 的跟踪方式，但是由于 DBT 方式依赖于检测框的准确性，直接影响着跟踪算法的准确性。

对于 TBD 的多目标跟踪方式，传统的方法是将目标检测与数据关联分为两个独立的任务，也可以称为 tracking by detection（TBD）。Bewley 等人 [1] 首先提出了 SORT 算法，它利用 Two-Stage 的 Faster-RCNN[2] 为检测器，使用卡尔曼滤波器预测下一帧的位置获得目标的运动信息以及使用匈牙利算法解决分配问题。随后，Bewley 等人在 SORT 算法基础上提出了 DeepSORT[3] 算法，利用额外的外观模型进一步提高匹配的准确性。Zhang 等人提出了 ByteTrack[4] 算法，将低分框应用于匹配，取得了良好的效果。

另一类方法是联合目标检测与跟踪的方式，如 Wang 等人提出了 JDE[5] 算法，使用 DarkNet53[6] 作为主干网络并与外观模型融合提升了整体算法的效率。紧接着，Zhang 等人提出了 FairMOT[7] 算法，使用基于关键点的 CenterNet[8] 检测方式，提高了 Re-ID 的准确性，同时取得了良好的精度与速度。本文提出一种基于 FairMOT 的改进算法，在原有 ECANet[9] 中的高效通道注意力机制基础上，加入最大池化并按一定量比例增加通道间交互，提出了 DMECA 模块，改进分支头的结构将传统卷积替换为深度可分离卷积 [10]，并使用 LeakeyReLU 作为激活函数，将匹配过程中的 IOU 匹配改进为 DIOU[11] 匹配，提高多目标跟踪算法的准确度。

1 基于 FairMot 的改进

FairMot 对于跟踪任务已经取得了不错的结果，但由于其主干网络提取出的不同维度的特征融合通道之间缺乏交互，所以这里提出了加入高效通道注意力机制 DMECA 模块来解决这个问题。针对分支头中，使用深度可分离卷积代替传统卷积降低冗余的参数量，并且加入批归一化操作提高模型的鲁棒性。在匹配阶段，IOU 匹配仅考虑到边界框之间的重叠程度，而 DIOU 考虑到了检测框与预测框的中心点之间的距离，当目标形状、尺度变化较大时效果更好，所以采用匹配更为准确的 DIOU 作为匹配方法。

1.1 高效注意力机制 ECANet

ECANet 注意力机制在不降维的情况下，使用一维卷积进行一定范围的跨通道交互来学习每个通道的权重，从而提高特征的表现能力。图 1 展示了 ECA 模块的结构图。ECA 模块需要计算出每一个通道的权重来捕捉通道之间的关系，它仅仅需要让每一个通道的相邻 k 个通道进行交互，可以分为以下三部分。

1. 太原师范学院计算机科学与技术学院 山西晋中 030619

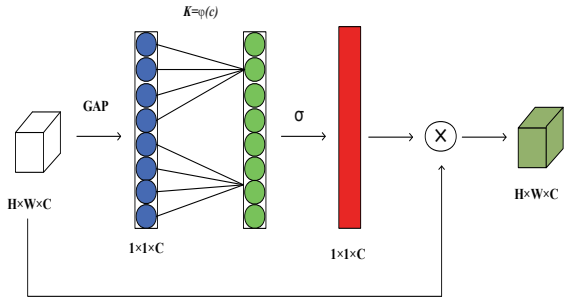

图 1 ECA 高效通道注意力机制

（1）首先通过池化操作，将每一个通道做全局平均池化。

（2）然后进行卷积核大小为 k 的一维卷积操作，再经过 Sigmoid 函数获取每个通道的权重 w。

（3）最后将权重与原始输入特征图上对应的特征相乘，输出最终的特征图。获取权重的公式可以简化为：

$$\omega_i = \sigma\left(\sum_{j=1}^{k} w^j y_i^j\right), y_i^j \in \Omega_i^k \tag{1}$$

式中：ω_i 代表最终通道的权重，w 代表卷积核参数的位置，y 代表与它相邻的 k 个通道。

由于 ECA 模块需要与它相邻的 k 个通道进行交互，但需要在适当的范围内进行局部的通道交互，因此需要选择适当的交互范围确定卷积核 k 的大小。由于手动交互需要耗费大量的时间精力，所以 ECA 模块提供了映射函数可以直接计算 k 的值。

$$k = \phi(c) = \left|\frac{\log_2^{(C)}}{\gamma} + \frac{b}{\gamma}\right|_{odd} \tag{2}$$

式中：k 是卷积核的大小，C 是给定的通道数，$gamma$ 与 b 为指定的参数这里分别设置为 2、1，$|t|_{odd}$ 代表与 t 最相近的奇数。当输入通道数越多时 k 越大、交互的范围越广，反之越小。

1.2 高效注意力机制 DMECA

DMECA 模块针对 ECANet 中池化仅仅考虑全局平均池化以及通道交互范围较小的问题提出了 DMECA 模块，图 2 展示了 DMECA 模块的结构。其中与 ECANet 的不同点如下。

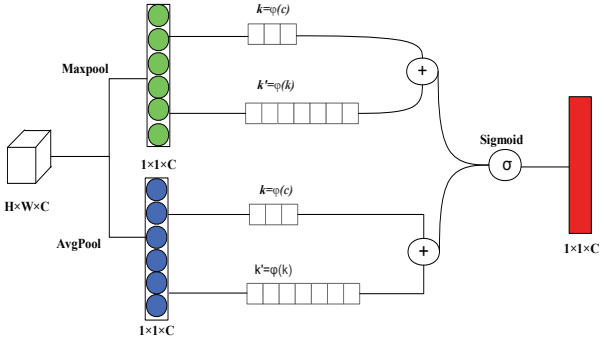

图 2 DMECA 模块图

（1）在全局平均池化基础上，增加全局最大池化操作，保留原有的全局平均池化分支，增加全局最大池化分支，将两个分支的信息相加融合，再经过 Sigmoid 函数计算出最终的权重。

（2）ECANet 中选取的一维卷积大小 k 值一般是比较小的，这样虽然模型需要的参数较小，但跨通道交互的范围也相应较小。为了解决跨通道范围局限的问题，在原有的 k 值交互卷积外，按照一定比例增加交互通道的 k 值，公式为：

$$k' = (2 \cdot k) + 1 \tag{3}$$

增加一维卷积和为 k' 的模块作为长跨通道交互模块，再将两部分的结果相加作为 Sigmoid 函数的输入值。

1.3 改进分支头

在 FairMOT 中，三个并行的分支回归头与身份嵌入头都通过对主干网络特征图应用 256 个 3*3 的卷积核，生成具有 256 个通道的特征图后送入每个分支对应的回归头中得到每个分支对应的结果。由于加入了 DMECA 通道注意力机制已经考虑了通道之间的相关性，所以在分支头部分可以进一步减少头部需要的参数，将二维卷积操作替换为深度可分离卷积。深度可分离卷积由两个部分组成，分别是深度卷积（DW）与逐点卷积（PW）。改进前后的结构如图 3 所示。对提取的特征图分别经过 DW 卷积与 PW 卷积并在每个卷积后增加批归一化操作，最后将原始的 Relu 激活函数替换为 LeakRelu 激活函数。

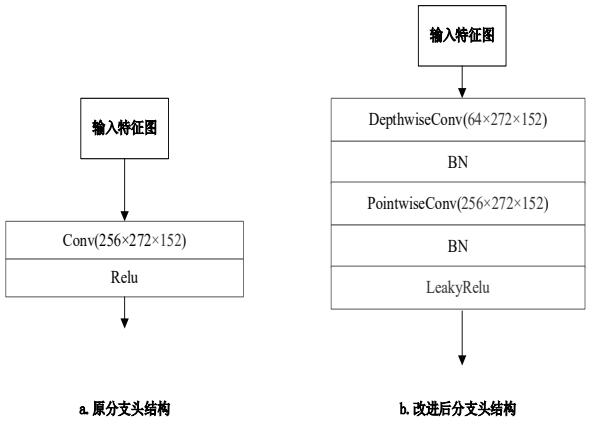

图 3 改进前后分支头结构图

深度可分离卷积是一种轻量级的卷积，在分支头中应用深度可分离卷积可以显著减少模型的参数。在深度卷积中，首先对输入的数据的每一个通道分别进行卷积操作，这样的操作对比传统卷积操作，可以以更少的参数量对每一个特征通道进行特征提取。然后进行逐点卷积操作，将深度卷积的输出与大小为 1 的卷积核进行卷积操作，将通道维度提升至所需维度，这样可以将不同通道之间的信息进行交互生成

所需特征并增加模型的非线性能力。最后，在 DW 与 PW 卷积之后均加入归一化操作，将 Relu 激活函数替换为了 LeakyRelu，使其可以更好地缓解梯度消失与梯度爆炸的问题。

1.4 DIOU 匹配

在目标跟踪过程中，在二次匹配中需要通过计算预测框与检测框之间的 IOU 得到 IOU 距离代价矩阵，再使用匈牙利算法完成匹配，将在第一阶段没有匹配上的轨迹与检测进行匹配。IOU 计算的是两个边界框的之间的交并比，即两个边界框之间的交集与并集的比值，它可以用来衡量两个边界框的相交的程度。但是 IOU 考虑的只是两个框之间的重叠区域，并没有考虑到两个框的中心点距离，所以引入了 DIOU（distance intersection over union）。DIOU 考虑到了边界框中心点之间的欧式距离，提供了更全面的度量方式，可以更好适应目标因遮挡或检测带来的目标尺度变化。DIOU 示意图如图 4 所示。

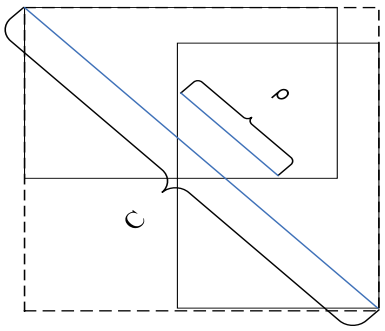

图 4 DIOU 图示

其公式为：

$$DIoU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} \tag{4}$$

式中：ρ 是检测框与预测框之间的中心点的距离，b 与 b^{gt} 分别是检测框与标注框的中心点位置，c 为最小包闭区的对角线的长度。DIOU 在原有 IOU 的基础上增加了惩罚项，即两边界框之间距离的平方与包闭区对角线平方的比值，使得最终匹配的结果更加准确。

2 实验与分析

2.1 实验环境与参数配置

本算法实验环境为操作系统 Ubuntu20.04，CPU 为 Intel 至强 Gold 6418，GPU 为 Nvidia Tesla V100（16 GB 显存），内存为 16 GB，使用 Python3.9 版本进行编译与运行，采用百度自研的 paddlepaddle 深度学习框架实现本算法模型，版本号为 2.4.0。实验参数配置中优化器使用 Adam 优化器，batchsize 设置为 6，初始学习率为 1e-4，实验设置为共训练 30 轮。

2.2 数据集与评价指标

本文使用的数据集为 VisDrone-2019 数据集以及 MOT16、MOT17 数据集。VisDrone 数据集是由天津大学机器学习与数据挖掘实验室 AISKYEYE 收集，通过各种无人机摄像头采集自中国的 14 个城市，涵盖了多种不同的场景，包括城市、乡村、高速公路等场景下的视频数据。MOT 系列数据集是多目标跟踪领域最为权威的基准数据集之一，包含了不同时间、地点下，因遮挡、变形、人群密集等因素造成的各种高难度跟踪场景。模型消融实验在 VisDrone-2019 数据集上进行，对比实验在 MOT17 数据集上进行实验，对比实验均使用 CrowdedHuman 数据集进行预训练，在 MOT 对应训练集进行训练，并使用对应数据集测试集进行测试，评估结果均提交至 MOTChallenge 官网得到测试结果。

本论文采用的多目标跟踪评价指标包括 CLEAR MOT Metric 中定义的 MOTA、MOTP、ID Switch、IDF1、MT、ML、HOTA 和 FPS。

2.3 实验与结果分析

2.3.1 消融实验

为了验证改进方法的有效性，本实验将各改进点加入模型后使用 VisDrone-2019 数据集中的训练集进行训练，实验结果在 VisDrone-2019 验证集上进行比较，结果如表 1 所示。

表 1 各改进方法对多目标跟踪算法影响结果

算法	HOTA	MOTA	MOTP	IDF1	IDS	MT	ML	FPS
FairMOT	43.5	40.2	76.1	55.0	583	36.2	36.6	16.4
FairMOT-E	43.1	40.3	76.4	54.4	392	33.2	35.8	16.4
FairMOT-ES	44.9	41.0	74.8	58.3	384	37.1	34.5	16.7
FairMOT-ESD	45.2	41.3	74.8	58.9	238	38.8	34.9	16.6

通过表 1 中的结果，可以看到加入 DMECA 模块（FairMOT-E）后 MOTA 与 MOTP 均有小幅度提升，IDS 下降了 191，说明了在加入高效通道注意力模块后，模型可以提取出更适应于检测分支与重识别分支的特征。FPS 基本没有变化，这是由于高效通道注意力机制的实现仅仅需要很少的参数。在改进分支头（FairMOT-ES）后，HOTA 提升 1.8%，MOTA 提升了 0.7%，IDF1 提升了 3.9%，IDS 有小幅下降，MT 提升了 3.9%，ML 下降了 0.8%，FPS 提升了 2.2，说明了在改进分支头后可以提取出更准确的特征。在加入 DIOU 匹配后，即（FairMOT-ESD），可以从表中看到改进算法与原算法的

对比，改进后的算法HOTA提升了1.7%，MOTA提升了1.1%，IDF1提升了3.9%，IDS也下降较为明显，MT上升了2.6%，ML下降了0.7%，说明了其在目标跟踪的准确度上有所提升且跟踪匹配上更为精确，但在FPS上并没有显著变化，证明了方法的有效性。

2.3.2 对比实验

为了进一步验证模型的准确性，将本模型算法与主流算法进行对比，在MOT17数据集的测试集上进行对比实验，MOT17测试集对比结果如表2所示。

表2 MOT17测试集与主流模型对比结果

算法	HOTA	MOTA	IDF1	IDS	MT	ML	FPS
LMOT_Tracker[12]	56.7	72.0	70.3	3071	45.4	17.3	28.6
TransCenter[13]	52.1	70.0	62.1	4647	38.4	24.9	11.8
FairMOT	59.3	72.7	72.3	3303	43.2	17.3	25.9
本文算法	59.6	72.3	72.7	2678	40.4	19.4	26.0

在MOT17测试集上，本模型算法指标优于绝大多数算法，HOTA指标取得了最高的结果，相较于FairMOT算法HOTA、IDF1均有小幅度提升，IDS下降明显。LMOT_tracker在速度上优于本模型，是由于其算法采用简化版的主干网络，舍弃了1/32的特征图，但在MOTA、HOTA、IDS等跟踪指标上均不如本模型算法。

3 结论

本文通过改进FairMOT算法，引入改进的DMECA高效注意力机制模块，在参数代价很小的情况下对通道之间的关系加权，分别对目标检测分支与重识别分支的信息进行调整，产生更准确的特征。在分支头结构上，使用深度可分离卷积的结构减少参数量并加入批归一化操作，使得整体算法更加高效。在匹配时使用DIOU代替IOU增加二次匹配的准确性。实验表明，在MOT数据集上HOTA、IDF1指标上都可以有一定的提升，IDS较原始模型下降明显，可以有效减少目标在复杂场景下的ID切换次数，从而更准确地跟踪目标。

参考文献：

[1] BEWLEY A, GE Z, OTT L, et al. Simple online and realtime tracking[C]//2016 IEEE international conference on image processing (ICIP). Piscataway:IEEE, 2016:3464-3468.

[2] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 39(6): 1137-1149.

[3] WOJKE N, BEWLEY A,PAULUS D. Simpleonline and realtime tracking with a deep association metric[C]//2017 IEEE international conference on image processing (ICIP). Piscataway: IEEE, 2017: 3645-3649.

[4] ZHANG Y, SUN P, JIANG Y, et al. Bytetrack: multi-object tracking by associating every detection box[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 1-21.

[5] WANG Z, ZHENG L, LIU Y, et al. Towards real-time multi-object tracking[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow. Cham: Springer Nature Switzerland, 2020:107-122.

[6] REDMON J , FARHADI A .YOLOv3: an incremental improvement[J].arXiv e-prints, 2018.DOI:10.48550/arXiv.1804.02767.

[7] ZHANG Y, WANG C, WANG X, et al. Fairmot: on the fairness of detection and re-identification in multiple object tracking[J]. International journal of computer vision, 2021, 129: 3069-3087.

[8] ZHOU X, WANG D, KRAHENBUHL P. Objects as points[EB/OL].(2019-04-25)[2023-06-23].https://arxiv.org/abs/1904.07850.

[9] WANG Q, WU B, ZHU P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.Piscataway:IEEE,2020: 11534-11542.

[10] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications[EB/OL].(2017-04-27)[2023-05-29]. https://arxiv.org/abs/1704.04861.

[11] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//Proceedings of the AAAI conference on artificial intelligence. Palo Alto: AAAI, 2020: 12993-13000.

[12] XU Y, BAN Y, DELMORME G, et al. Transcenter: transformers with dense queries for multiple-object tracking[EB/OL].(2022-09-30)[2023-06-01].https://arxiv.org/abs/2103. 15145.

[13] MOSTAFA R, BARAKA H, BAYOUMI A E M. LMOT: efficient light-weight detection and tracking in crowds[J]. IEEE access, 2022, 10: 83085-83095.

【作者简介】

张慧旺（1997—），山西太原人，硕士研究生，研究方向：图像处理、目标跟踪。

（收稿日期：2023-09-12）