



智能系统学报

CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

一种基于深度学习目标检测的长时目标跟踪算法

邵江南, 葛洪伟

引用本文:

邵江南, 葛洪伟. 一种基于深度学习目标检测的长时目标跟踪算法[J]. 智能系统学报, 2021, 16(3): 433–441.

SHAO Jiangnan, GE Hongwei. A long-term object tracking algorithm based on deep learning and object detection[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(3): 433–441.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.201910029>

您可能感兴趣的其他文章

区域损失函数的孪生网络目标跟踪

Regional loss function based siamese network for object tracking

智能系统学报. 2020, 15(4): 722–731 <https://dx.doi.org/10.11992/tis.201910005>

基于特征融合及自适应模型更新的相关滤波目标跟踪算法

Correlation filter target tracking algorithm based on feature fusion and adaptive model updating

智能系统学报. 2020, 15(4): 714–721 <https://dx.doi.org/10.11992/tis.201803036>

模糊直方图模型的运动目标跟踪

Target tracking based on the fuzzy histogram model

智能系统学报. 2019, 14(5): 939–946 <https://dx.doi.org/10.11992/tis.201807033>

联合外形响应的深度目标追踪器

A deep object tracker with outline response map

智能系统学报. 2019, 14(4): 725–732 <https://dx.doi.org/10.11992/tis.201807029>

一种自适应模板更新的判别式KCF跟踪方法

Adaptive template update of discriminant KCF for visual tracking

智能系统学报. 2019, 14(1): 121–126 <https://dx.doi.org/10.11992/tis.201806038>

高斯核函数卷积神经网络跟踪算法

Convolutional neural network tracking algorithm accelerated by Gaussian kernel function

智能系统学报. 2018, 13(3): 388–394 <https://dx.doi.org/10.11992/tis.201612040>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.201910029

一种基于深度学习目标检测的长时目标跟踪算法

邵江南^{1,2}, 葛洪伟^{1,2}

(1. 江南大学 江苏省模式识别与计算智能工程实验室, 江苏 无锡 214122; 2. 江南大学 物联网工程学院, 江苏 无锡 214122)

摘要: 针对长时目标跟踪所面临的目标被遮挡、出视野等常常会导致跟踪漂移或丢失的问题, 基于 MDNet 提出一种深度长时目标跟踪算法 (long-term object tracking based on MDNet, LT-MDNet)。首先, 引入了一种改进的收缩损失函数, 以解决模型训练时正负样本不均衡的问题; 其次, 设计了一种高置信度保留样本池, 对在线跟踪时的每一帧的有效并且置信度最高结果进行保留, 并在池满时替换最低置信度的保留样本; 最后, 在模型检测到跟踪失败或连续跟踪帧数达到特定阈值时, 利用保留样本池进行在线训练更新模型, 从而使模型在应对长时跟踪时保持鲁棒和高效。实验结果表明, LT-MDNet 在跟踪精度和成功率上都展现了极强的竞争力, 并且在目标被遮挡、出视野等情况下保持了优越的跟踪性能和可靠性。

关键词: 目标跟踪; 长时跟踪; 神经网络; 卷积特征; 类不平衡问题; 损失函数; 特征提取; 深度学习

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2021)03-0433-09

中文引用格式: 邵江南, 葛洪伟. 一种基于深度学习目标检测的长时目标跟踪算法 [J]. 智能系统学报, 2021, 16(3): 433-441.
英文引用格式: SHAO Jiangnan, GE Hongwei. A long-term object tracking algorithm based on deep learning and object detection[J]. CAAI transactions on intelligent systems, 2021, 16(3): 433-441.

A long-term object tracking algorithm based on deep learning and object detection

SHAO Jiangnan^{1,2}, GE Hongwei^{1,2}

(1. Jiangsu Provincial Engineering Laboratory of Pattern Recognition and Computational Intelligence, Jiangnan University, Wuxi 214122, China; 2. School of Internet of Things, Jiangnan University, Wuxi 214122, China)

Abstract: Aiming at the problem of tracking drift or loss caused by the occlusion and the out-of-view of the target in long-term tracking, this paper proposes a new deep, long-term tracking algorithm based on MDNet (LT-MDNet). First, an improved shrinkage loss function is introduced to solve the problem of the positive-negative class imbalance in the model training. Second, a high confidence retention sample pool is designed to retain the valid and highest confidence results of each frame during online tracking and to replace the lowest confidence retention samples when the pool is full. Finally, when the model detects a tracking failure or when the continuous tracking frame number reaches a specific threshold, the reserved sample pool is used for online training to update the model to make the model robust and efficient in dealing with long-term tracking. Experimental results show that LT-MDNet is highly competitive in its tracking accuracy and success rate and maintains excellent tracking performance and reliability in the case of target occlusion and out-of-view.

Keywords: object tracking; long-term tracking; neural network; convolutional features; class imbalance problem; loss function; feature extraction; deep learning

随着计算机视觉领域的发展, 目标跟踪在人机交互、视频监控、自动驾驶和机器人等领域得

到越来越多的重视和应用。早期的跟踪模型常见的有粒子滤波^[1]、Mean shift^[2]、相关滤波^[3]及其衍生模型^[4-8]等。虽然这些传统的跟踪器经过近年来的发展在精度和速度上都有了明显提升, 但是对目标特征的提取仍以光流、外观形状、颜色等

收稿日期: 2019-10-24.

基金项目: 江苏省研究生创新计划项目 (KYLX16_0781); 江苏高校优势学科建设工程项目 (PAPD).

通信作者: 葛洪伟. E-mail: ghw8601@163.com.

浅层特征为主,不能捕捉目标语义特征,难以在面对长时跟踪时目标外观形变、被遮挡或出视野等情况下保持鲁棒跟踪。而随着深度学习在跟踪领域的发展,这些问题正逐渐得到改善。

基于深度学习的跟踪器,能有效利用目标的深度特征对目标进行语义级抽象,并拟合目标的运动过程,这大大提高了跟踪精度;且模型的特征提取器在离线时得到了大量标记图片的端到端预训练,这进一步提高了跟踪器在线跟踪速度,使深度学习在目标跟踪领域的应用成为可能。由于目标检测与目标跟踪具有相似性,检测领域的很多算法思想都逐渐被应用于跟踪:MDNet^[9]在前一帧的目标周围高斯采样出大量候选框,再利用预训练的二分类器区分候选框中的目标和背景,从而定位目标位置;SiamFC孪生网络模型^[10]利用同样结构的CNN(convolutional neural network)网络对初始帧目标和当前帧搜索域提取深度特征,然后通过卷积操作进行相似性计算,得到目标在搜索域位置的响应;RasNet^[11]将传统的相关滤波引入孪生网络中,再把检测领域的通道注意力、残差注意力和通用注意力结合,提高了跟踪器的泛化能力,增大了对首帧目标信息的利用率。

然而,基于深度学习的目标跟踪算法,需要收集含有目标信息的正样本,以及含有背景信息的负样本用于模型的迭代训练。但在实际取样中,正样本的可采样数量远远小于负样本,且这些负样本中绝大多数是无意义的简单负样本,对分类器的训练贡献很少。不仅如此,虽然简单负样本的损失值低,但数量庞大,这些值的累积和很可能掩盖了正样本和重要负样本损失值的变化,从而影响模型训练的收敛速度以及跟踪器性能。针对这种不平衡问题,MDNet采用检测领域的难分样本挖掘^[12]技术筛选难分负样本,使模型的训练免受大多数简单样本的干扰,但其对难分负样本的采样数量固定,在很多应用场景下难以避免地对简单样本过多采集;FocalLoss^[13]对所有的负样本损失值进行抑制,但同时降低了有价值的难分样本对训练的贡献。

此外,长时跟踪时目标不可避免地会出现形变、遮挡和出视野等情况,未采用模型在线更新机制的深度跟踪器往往由于过度依赖首帧特征而导致跟踪漂移或目标丢失。MDNet保留最近 T 帧的跟踪结果用于更新模型,但这可能导致有价值的保留样本被新出现的无效样本替代,同时其忽视首帧目标信息,也使得模型在跟踪失败后难以通过有效更新重新定位目标位置;CFNet^[14]对当前帧之前的所有跟踪结果计算平均值,更新目标模板,同样,随着跟踪的持续,目标模板不断被

污染,最终导致跟踪漂移,而在目标被遮挡时,这种污染的影响将更加明显。近年来,随着深度跟踪器的骨干网络层数逐渐加深,模型在线更新对跟踪效率的影响逐渐加大,因此大多数深度跟踪器都未引入在线更新策略,但模型更新仍是维持长时跟踪鲁棒性的重要途径。

针对上述问题,本文结合MDNet提出了一种深度长时目标跟踪算法LT-MDNet。通过改进收缩损失函数和提出一种全新的高置信度保留样本池,进一步解决类不平衡和模型更新问题。在标准跟踪数据集OTB2015^[15]和标准长时数据集UAV20L^[16]上的实验结果表明,LT-MDNet有着优越的跟踪性能并能在长时跟踪时的遮挡和目标出视野等复杂情况下保持鲁棒性。

1 MDNet 算法

MDNet将每一个视频帧序列视为一个单独的域,基于VGG-M^[17]架构,提出多域卷积网络。如图1所示,conv1、conv2、conv3、fc4、fc5为域无关的共享层,初始参数通过离线训练得出;fc6(fc6^{*i*}, $i=1,2,\dots,k$)为特定域层,参数在每次迭代训练或跟踪特定域时均随机初始化,灰色框和白色框分别表示每个域中的正样本(目标)和负样本(背景); $1,2,\dots,k$ 为离线训练时所迭代训练模型的 k 个视频序列。MDNet在共享层中对域无关的信息建模,从而获得通用的特征表示,并在fc6层训练一个二分类器,区分目标与背景,通过将目标跟踪抽象为一个二分类任务,最终达到跟踪运动目标的目的。

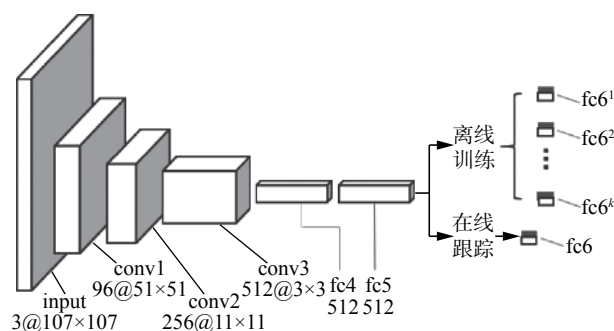


图1 MDNet模型主要结构

Fig. 1 Main architecture of MDNet

MDNet将目标搜索域采样出的样本,统一裁剪成 107×107 大小,输入图1的网络中,通过网络fc6层输出样本属于目标的置信度 p ,将模型难以判别的负样本归为难分样本。为降低正负样本不平衡问题对模型在线训练的干扰,MDNet采取了难分样本挖掘策略,取 p 值最高的前96个负样本,再将这组难分样本和在目标周围高斯采样出的32

个正样本结合,通过式(1)损失函数和随机梯度下降算法(stochastic gradient descent, SGD)对模型参数进行迭代训练更新:

$$L(p, y) = -(y \cdot \ln(p) + (1 - y) \cdot \ln(1 - p)) \quad (1)$$

式中: $y \in \{0, 1\}$ 为样本类别标签; $p \in \{x | 0 \leq x \leq 1\}$ 为样本属于目标 ($y=1$) 的模型估计概率(置信度)。

2 LT-MDNet 算法

本文提出的 LT-MDNet 算法在 MDNet 基础上通过改进收缩损失函数解决采样时的类不平衡问题,并提出了高置信度保留样本池,以鲁棒更新模型,使模型能够有效应对长时跟踪下的遮挡、出视野等复杂因素。

2.1 损失函数

从式(1)可以看出,当样本实际为负时, $y=0$, 原式可简化为式(2),其损失值变化曲线如图2中虚线所示。

$$L(p, 0) = -\ln(1 - p) \quad (2)$$

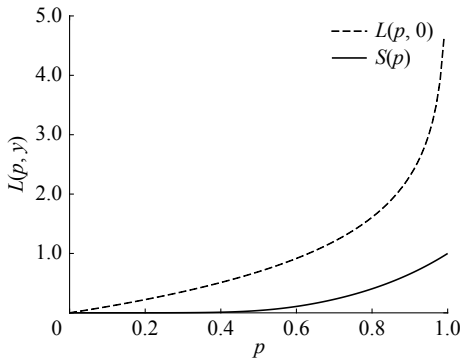


图2 $L(p, 0)$ 和 $S(p)$ 损失值对比曲线

Fig. 2 Loss comparison curve of $L(p, 0)$ and $S(p)$

可见,模型对样本的预测输出越接近真实样本标签0,损失值 L 越小;预测越接近1, L 越大。但在处理简单负样本时,由于负样本的采样数量远高于正样本,尽管采样的负样本经过难分样本挖掘,简单负样本占比仍较大,依然存在类不平衡问题。而图2中虚线在样本置信度 $p \in [0, 0.5]$ 的损失值未能被有效抑制,使得大量的较简单负样本损失值累加,仍可以掩盖正样本和难分负样本损失值的变化,从而影响模型训练。

为进一步解决上述类不平衡问题,本文结合文献[18]的思想,改进了收缩损失函数,在式(1)的基础上增加了收缩抑制项(见式(3)),通过抑制损失值曲线在 $p \in [0, 0.5]$ 内的梯度,降低简单负样本的损失值,以突出难分样本和正样本损失值的变化对模型训练的贡献率。

$$S(p) = \frac{p^4}{1 + \exp(a(c - p^2))} \quad (3)$$

式中: $\exp(\cdot)$ 能随着模型输出的样本置信度 p 的变化非线性改变其损失值 $S(p)$; a, c 为超参数,分别控制损失收缩速度(曲线梯度变化率)和损失值调节敏感范围($S(p)$ 值趋于0时的 p 区间)。

式(3)损失值变化曲线如图2中实线所示,当 $p < 0.5$ 时,损失值几乎为零,而当 p 接近1时,损失值迅速增大,比式(2)更有效地非线性抑制了简单负样本的损失值,且负样本越简单抑制比越大。

在线训练与离线训练存在本质区别:在线训练迭代次数少,涉及网络层不多,且已经预训练的模型在线更新时只需微调部分参数即可快速收敛,因此,虽然抑制大多数负样本会降低模型的泛化能力,但这样做允许模型对当前视频域存在一定程度过拟合,这在提高跟踪器性能的同时有效降低了训练成本;对 p 的平方操作,使得简单负样本在损失计算时得到进一步抑制。修正后的模型损失函数如式(4)所示:

$$L_s(p, y) = L(p, y) + \delta S(p) \quad (4)$$

式中: δ 为超参数,控制损失收缩比。

2.2 模型更新

为了尽可能地保留有效跟踪结果使模型更新更为有效,本文设计并实现了基于置信度的跟踪样本池,保留最近最高置信度的跟踪结果样本,并在池满时替换置信度最低的保留样本(第一帧目标永久保留在样本池中)。置信度评估标准,按照当前帧所采样的正样本在 fc6 层输出的5个最高分类分数的均值计算。样本池的具体工作流程如图3所示。

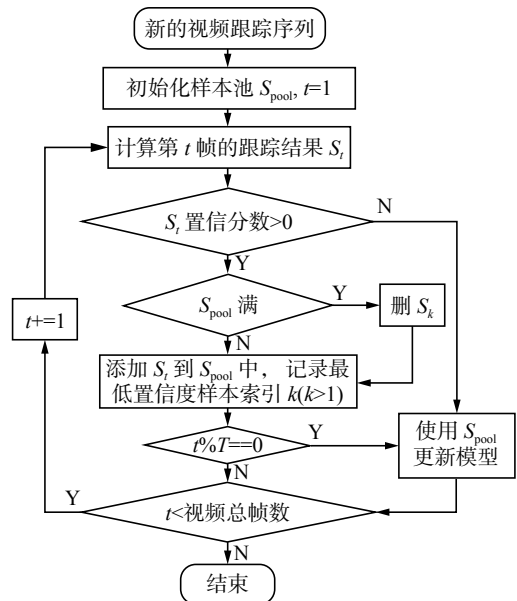


图3 高置信度保留样本池工作流程

Fig. 3 Workflow chart of the high-confidence retention sample pool

模型应用了短时更新和长时更新2种更新策略,前者当检测到跟踪失败(置信分数低于0)时利用 S_{pool} 在线更新模型,后者每隔 T 帧对模型进行更新以提高跟踪精度。2种更新策略的结合,以及高置信度样本池的应用,使 LT-MDNet 能够保持对视频帧环境变化的适应能力,并延长首帧目标特征的时域有效范围,从而可以很好地应对长时跟踪下的遮挡、出视野等复杂因素。

2.3 算法流程

LT-MDNet 在前帧目标位置周围高斯采样出多个候选正样本,通过预训练和在线更新后的网络模型输出这些候选样本的目标置信度 $f^+(\cdot)$,取置信度最高的5个正样本的坐标和长宽数据的平均值作为当前帧跟踪结果,该结果的置信度为这5个正样本置信度的均值。算法的主要步骤:

- 1) 输入视频帧序列和首帧的目标位置,随机初始化 fc6 层的权值参数 w_6 ;
- 2) 采样正样本 S_1^+ 和负样本 S_1^- ,初始化 S_{pool} ,训练边框回归器 BBR;
- 3) 使用 S_1^+ 和 S_1^- 迭代训练模型,通过式(4)和 SGD 更新 fc4、fc5、fc6 层的权值 w_4 、 w_5 、 w_6 ;
- 4) 对于第 t 帧图像 N_t ,在 N_{t-1} 目标位置周围高斯采样候选样本集合 X_t ;
- 5) 取 X_t 中5个最高置信度候选位置的均值 x_t^* ,通过 BBR 调整后的 x_t^* 值作为第 t 帧目标估计位置;
- 6) 若 $f^+(x_t^*) > 0$,则按照流程图3所示维护高置信度保留样本池 S_{pool} ;
- 7) 若 $f^+(x_t^*) \leq 0$ 或 $t \% 10 = 0$,则使用 S_{pool} 更新模型;
- 8) 重复步骤4)~7)直至目标跟踪结束。

3 实验结果与分析

本文 LT-MDNet 算法基于 PyTorch 1.2.0 编程语言和 CUDA 10.0 深度学习架构实现,实验平台操作系统为 Windows 10,处理器为 AMD R5-2600 3.4 GHz, GPU 为 NVIDIA RTX2070,内存 16 GB。

为验证模型的有效性、泛化能力和长时跟踪性能,分别在 OTB2015 和 UAV20L 数据集上与 MDNet^[9]、DSL^[18]、SRDCF^[6]、TADT^[19]、SiamRPN^[20]、CIResNet22-FC^[21]、MUSTer^[22] 和 BACF^[7] 等主流跟踪模型进行对比实验,将统计的各数据集上所有图像序列跟踪结果平均值作为最终结果。

3.1 实验设置

模型离线训练于 IMAGENET-VID^[23] 目标检

测数据集,每间隔 $T=10$ 帧时更新模型参数;首帧模型更新训练迭代 50 次,学习率为 0.0005;非首帧更新迭代 15 次,学习率为 0.001;式(3)中的 a 、 c 分别设置为 10、0.2,式(4)中 δ 取 1.3。为保证实验的公平性,实验中所对比的 MDNet 模型为作者开源的 PyTorch 版,LT-MDNet 模型参数与 MDNet 保持一致;其余对比跟踪器的实验结果均来自于作者论文所给出的实验结果或使用作者提供的开源代码和参数实际运行所得。

3.2 评价指标

本文采用跟踪成功率 (tracking success rate) 和跟踪精度 (tracking precision) 来比较各跟踪器的性能。

1) 跟踪成功率^[24]。通过计算跟踪器对目标位置边界的预测框 (R_t) 和目标真实边界框 (R_a) 的重叠率 (IoU) 来计算:

$$\text{IoU} = \frac{|R_t \cap R_a|}{|R_t \cup R_a|} \quad (5)$$

IoU 值越大表示跟踪器的成功率越高,当 $\text{IoU} > 0.5$ 时可认为该帧目标被成功跟踪,通过计算成功率图中的曲线下面积 (area under curve, AUC) 得到模型的跟踪成功率。

2) 跟踪精度^[24]。通过计算 R_t 的中心位置 (x_t, y_t) 与 R_a 的中心位置 (x_a, y_a) 间的欧氏距离来衡量:

$$\varepsilon = \sqrt{(x_t - x_a)^2 + (y_t - y_a)^2} \quad (6)$$

两者的欧氏距离 ε 越小表示跟踪精度越高。

3.3 OTB2015 数据集的测试

OTB2015 包含 100 个跟踪视频帧序列,含目标遮挡、尺度变化、出视野等 11 种常见的复杂跟踪因素,可有效评估跟踪器的综合性能。

图4(a)、(b)为 LT-MDNet 和对比跟踪器在 OTB2015 数据集上对所有视频帧序列的跟踪结果的综合统计平均曲线图。可以看出,LT-MDNet 的综合跟踪精度为 91.6%,成功率为 68.2%,分别较对比实验中的最优算法提高 0.8% 和 1.1%。

图4(c)、(d)、(e)、(f)分别为各跟踪器在 OTB2015 上关于目标遮挡、目标出视野的平均跟踪精度和跟踪成功率,在这些复杂因素下,LT-MDNet 跟踪效果均为最优,尤其是对比 MDNet,在遮挡因素下均提高了 1.8%,在目标出视野因素下则分别提高了 2.2% 和 1.2%。可见,除了综合性能的领先,LT-MDNet 在应对目标遮挡、出视野等复杂因素同样保持了良好的跟踪能力和稳定性。

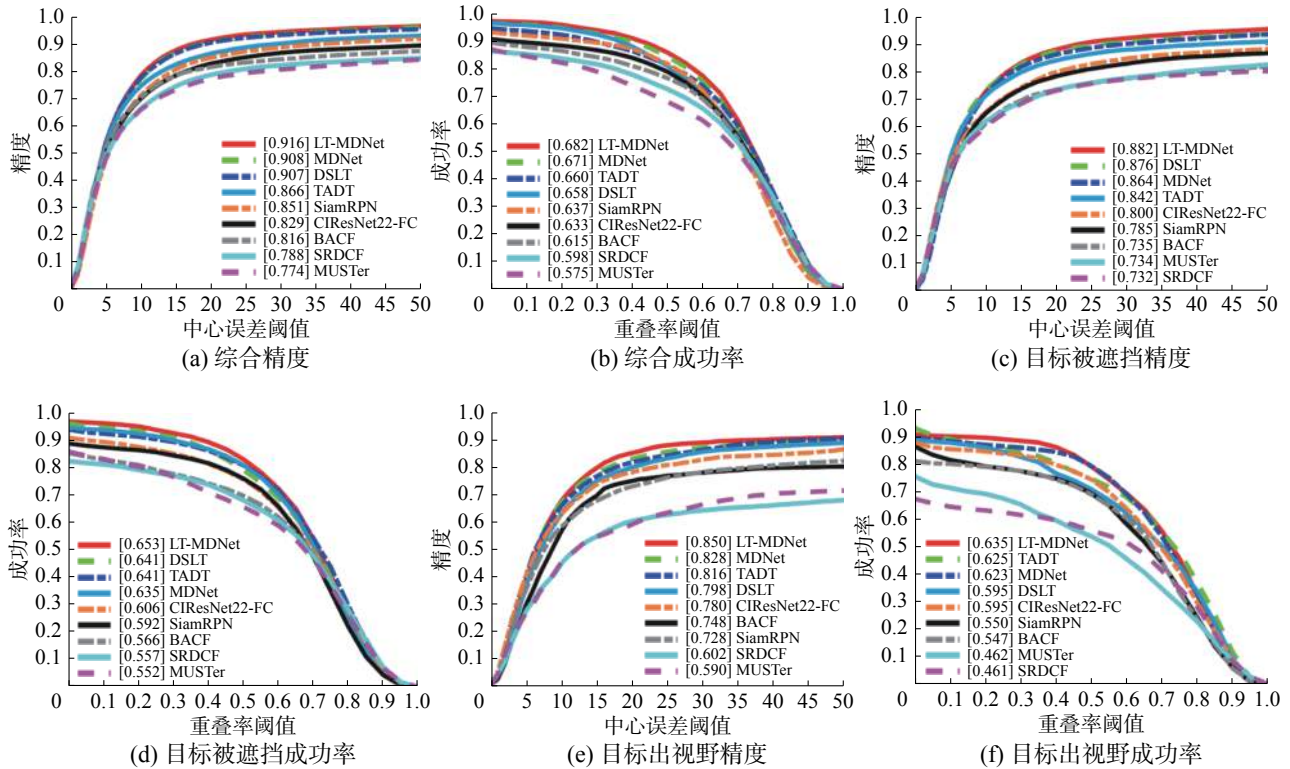


图4 LT-MDNet 和对比算法在 OTB2015 上的跟踪结果评估曲线

Fig. 4 Evaluation plots of the tracking results on OTB2015 of contrast algorithms and LT-MDNet

3.4 UAV20L 长时数据集的测试

UAV20L 包含 20 个无人机拍摄的视频帧序列, 这些序列时间跨度大, 目标视野变化剧烈, 目标形变也更大, 有利于评估跟踪器在长时跟踪环境下的有效性和鲁棒性。本文基于 UAV20L 数据集与 MDNet、TADT、SiamRPN、CIResNet22-FC、BACF、MUSTer 和 SRDCF 等主流长短时跟踪算法进行了对比实验, 并通过定量分析和定性分

析, 进一步验证 LT-MDNet 在长时跟踪下的跟踪性能。

3.4.1 定量分析

图 5(a)、(b) 分别为各跟踪器在 UAV20L 的平均跟踪精度和成功率曲线对比图, 可以看出 LT-MDNet 在长时跟踪数据集上性能大幅领先, 较最优对比模型的精度和成功率分别提升 3.7%、2.9%, 较 MDNet 则分别提升 7.3%、4.4%。

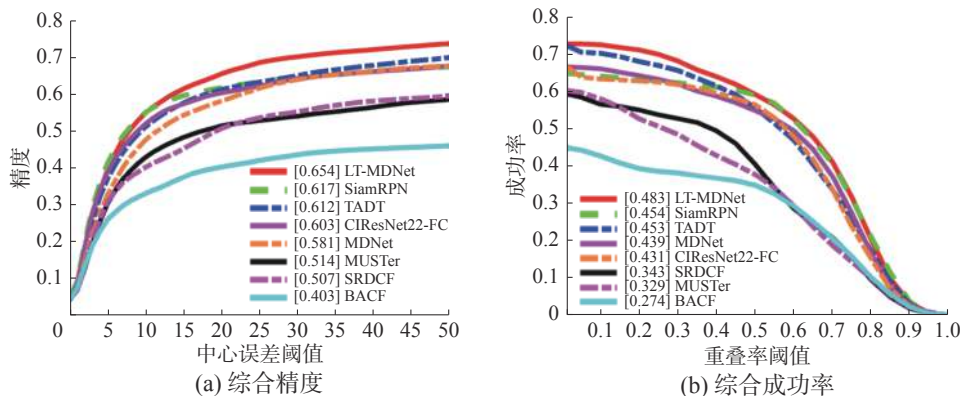


图5 LT-MDNet 和对比算法在 UAV20L 上跟踪结果评估曲线

Fig. 5 Evaluation plots of the tracking results on UAV20L of contrast algorithms and LT-MDNet

针对 UAV20L 所包含的 12 个不同的视频属性, 即相机运动 (CM)、快速运动 (FM)、纵横比变化 (ARC)、相似目标干扰 (SO)、光照变化 (IV)、背景杂乱 (BC)、视角变化 (VC)、尺度变化 (SV)、低分

辨率 (LR)、部分遮挡 (PO)、全部遮挡 (FO) 和目标出视野 (OV), 表 1 定量展示了 LT-MDNet 和各对比跟踪器在应对这些复杂跟踪因素下的平均 AUC 数据。从表 1 中可以看出, LT-MDNet 在 12 个视频

属性中的10个均保持了最优或次优性能,除在CM、ARC下稍稍落后于MDNet外,其余复杂跟踪环境下均优于MDNet。而究其原因,主要是因为训练集中缺少CM和ARC复杂变化因素,且在线训练时为使模型快速收敛而抑制了大部分简单背

景负样本对参数的影响,一定程度上弱化了模型在应对未包含在训练集中的复杂因素时的泛化能力;而在应对已知变化因素时,模型则可以通过有效的离线学习和鲁棒的在线更新机制,迅速适应跟踪视频域内的目标变化并准确定位目标位置。

表1 UAV20L数据集上对比算法与本文LT-MDNet在12个视频属性上AUC的定量对比结果

Table 1 Quantitative comparison results of the AUC of contrast algorithms and the proposed LT-MDNet algorithm on the 12 video attributes in the UAV20L database

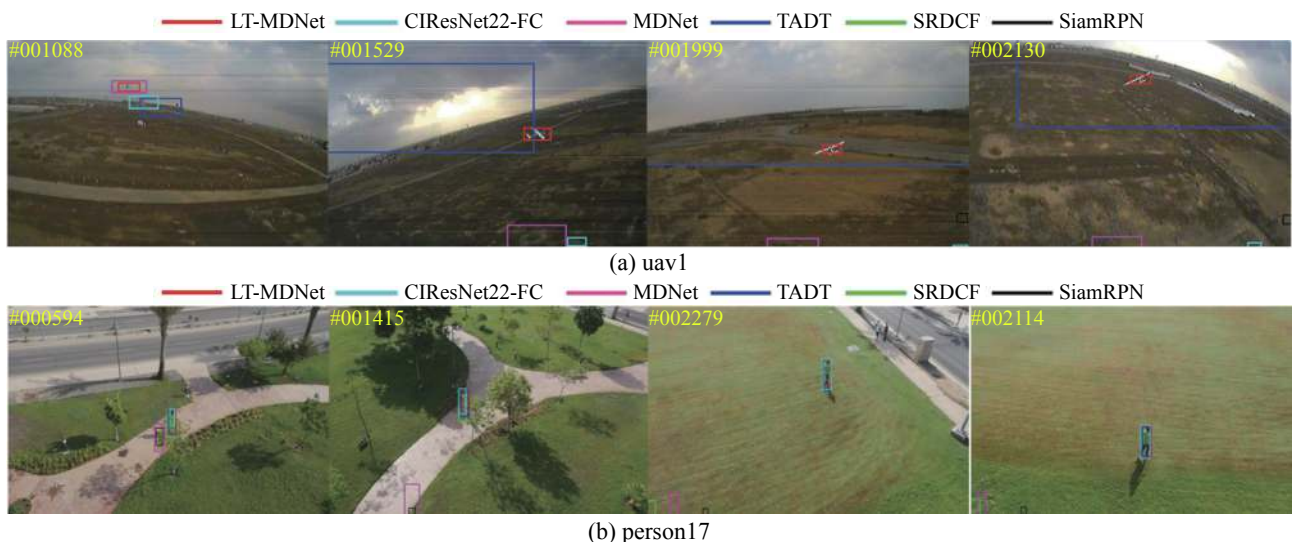
视频属性	视频总数	SRDCF	MUSTer	BACF	TADT	SiamRPN	CiResNet22-FC	MDNet	LT-MDNet
IV	19	0.332	0.314	0.251	0.439	0.443	0.417	0.425	0.470
CM	12	0.329	0.309	0.293	0.411	0.479	0.385	0.469	0.448
OV	7	0.170	0.200	0.121	0.277	0.192	0.223	0.201	0.258
FM	18	0.320	0.305	0.276	0.433	0.423	0.399	0.422	0.460
ARC	9	0.303	0.318	0.174	0.406	0.450	0.368	0.444	0.436
SO	7	0.295	0.242	0.154	0.420	0.343	0.425	0.371	0.469
BC	4	0.156	0.230	0.105	0.297	0.104	0.249	0.148	0.275
VC	19	0.327	0.307	0.271	0.436	0.436	0.414	0.422	0.468
SV	16	0.270	0.275	0.191	0.390	0.388	0.361	0.363	0.417
LR	14	0.397	0.342	0.335	0.478	0.532	0.488	0.479	0.524
PO	12	0.228	0.278	0.205	0.341	0.335	0.273	0.327	0.351
FO	5	0.197	0.206	0.174	0.355	0.422	0.303	0.324	0.367
UAV20L	20	0.343	0.329	0.274	0.453	0.454	0.431	0.439	0.483

3.4.2 定性分析

为了更为直观地对比跟踪器在应对长时跟踪复杂因素下的实际跟踪性能,本文选取了UAV20L中较为典型的4组图像序列,即uav1、person17、car1、bike1,包含全部12个复杂跟踪因

素,并选取了在UAV20L上综合性能最好的5个对比算法(SiamRPN、TADT、MDNet、CiResNet22-FC和SRDCF)与LT-MDNet进行定性实验对比。

图6着重对跟踪器在目标快速移动、目标遮挡、目标出视野和目标形变下的实际表现做了展示。



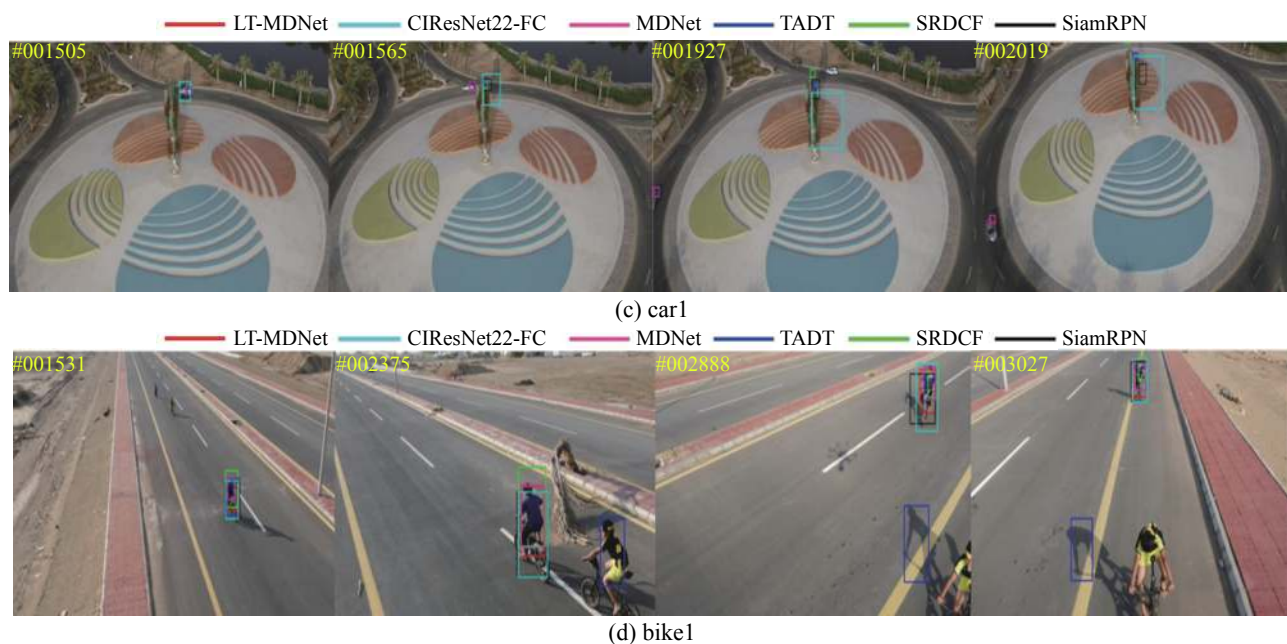


图6 LT-MDNet与对比算法在uav1、person17、car1和bike1下的定性对比结果

Fig. 6 Qualitative comparison results of contrast algorithms and the proposed LT-MDNet algorithm on uav1, person17, car1 and bike1 in the UAV20L database

图6(a)为对小型飞机这种快速移动的小目标进行跟踪,在第1088帧,飞机平稳飞行,此时只有LT-MDNet和MDNet可以锁定目标位置;而在第1529帧和之后的帧序列里,由于飞机航向、目标背景和航拍角度均剧烈变化,除LT-MDNet能由于模型有效更新操作精确锁定目标位置外,其余跟踪器均跟踪失败。

图6(b)中目标行人在第594、1415帧被树叶遮挡,此时LT-MDNet较其他对比跟踪器更能精确捕捉目标轮廓信息,并在随后持续稳定跟踪;SRDCF、MDNet和SiamRPN均在目标重现后失去跟踪能力。

图6(c)中目标汽车在第1505帧被遮挡、在第1927帧出视野,从图中可以看出,除LT-MDNet和MDNet外其余跟踪器均在目标被遮挡和出视野时发生了跟踪漂移。

图6(d)中目标小男孩在第2375帧掉转头,目标外观由男孩正面变成背影,而随后航拍镜头加速向前,目标迅速被甩在后面,其在图像帧中轮廓尺度也剧烈变化。此时,LT-MDNet仍可准确定位目标位置,并保持对目标纵横比信息的有效估计,而其余跟踪器均难以检测目标的准确轮廓,TADT则彻底跟踪失败。

从上述分析可见,LT-MDNet能够在长时跟踪下很好地应对这些复杂因素,在目标遮挡、出视野重现时仍能迅速定位目标;对快速位移、剧烈形变下的目标,在经过模型更新后可以维持鲁

棒跟踪,并且较对比算法能更精确地捕捉目标的轮廓信息。

3.5 消融实验

为进一步验证LT-MDNet改进的有效性,本文在OTB2015上对模型进行了消融实验,分别测试在MDNet模型基础上所做出的2个主要改进:收缩损失(SK)和高置信度保留样本池(HU),实验环境与模型主要超参数均一致。MDNet+HU、MDNet+SK和LT-MDNet在线跟踪时速度约为3 f/s,与MDNet持平,跟踪成功率(Succ)与跟踪精度(Prec)对比数据如表2所示。实验结果表明,高置信度保留样本池和收缩损失函数均能在不影响跟踪速度的情况下有效提高跟踪器的定位精度和成功率,且两者结合能够协同提高跟踪器性能。

表2 LT-MDNet模型在OTB2015数据集上的消融实验结果

Table 2 Ablation study results of the LT-MDNet algorithm in the OTB2015 database

模型	Prec/%	Succ/%
MDNet	90.8	67.1
MDNet+HU	91.0	67.2
MDNet+SK	91.4	67.6
LT-MDNet	91.6	68.2

4 结束语

针对长时跟踪时的目标遮挡和出视野问题,

本文提出了一种能够有效在线更新的深度跟踪算法 LT-MDNet。在难分样本挖掘的基础上,引入改进的收缩损失函数,进一步抑制简单负样本的损失值,以缓和类不平衡问题;在模型在线追踪运动目标时,设计了一种全新的高置信度保留样本池,对在线跟踪时的高置信度正样本进行保留,并在池满时替换最低置信度的保留样本,首帧样本则长期保存在池中;当模型检测到跟踪失败以及跟踪达到特定帧间隔时,利用保留样本池进行在线训练更新模型,以维持跟踪器的长期鲁棒性。在 OTB2015 标准数据集和 UAV20L 长时跟踪数据集上与主流跟踪算法对比, LT-MDNet 在跟踪精度和成功率上均保持了优越的性能,并能很好地应对目标遮挡和出视野问题。接下来的工作将探索在不同跟踪环境下自适应收缩比 δ 和其他收缩损失超参数的值来进一步提高模型的泛化能力。

参考文献:

- [1] ARULAMPALAM M S, MASKELL S, GORDON N, et al. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking[J]. IEEE transactions on signal processing, 2002, 50(2): 174–188.
- [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] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA, 2010: 2544–2550.
- [4] 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.
- [5] MUELLER M, SMITH N, GHANEM B. Context-aware correlation filter tracking[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 1396–1404.
- [6] DANELLJAN M, HÄGER G, SHAHBAZ KHAN F, et al. Learning spatially regularized correlation filters for visual tracking[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 4310–4318.
- [7] KIANI GALOOGAHI H, FAGG A, LUCEY S. Learning background-aware correlation filters for visual tracking[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy, 2017: 1135–1143.
- [8] 宁欣, 李卫军, 田伟娟, 等. 一种自适应模板更新的判别式 KCF 跟踪方法 [J]. 智能系统学报, 2019, 14(1): 121–126.
NING Xin, LI Weijun, TIAN Weijuan, et al. Adaptive template update of discriminant KCF for visual tracking[J]. CAAI transactions on intelligent systems, 2019, 14(1): 121–126.
- [9] 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. Las Vegas, USA, 2016: 4293–4302.
- [10] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional Siamese networks for object tracking[C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 850–865.
- [11] WANG Qiang, TENG Zhu, XING Junliang, et al. Learning attentions: residual attentional Siamese network for high performance online visual tracking[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 4854–4863.
- [12] SUNG K K, POGGIO T. Example-based learning for view-based human face detection[J]. IEEE transactions on pattern analysis and machine intelligence, 1998, 20(1): 39–51.
- [13] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy, 2017: 2980–2988.
- [14] VALMADRE J, BERTINETTO L, HENRIQUES J, et al. End-to-end representation learning for correlation filter based tracking[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 2805–2813.
- [15] WU Yi, LIM J, YANG M H. Object tracking benchmark[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1834–1848.
- [16] MUELLER M, SMITH N, GHANEM B. A benchmark and simulator for uav tracking[C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 445–461.
- [17] CHATFIELD K, SIMONYAN K, VEDALDI A, et al. Return of the devil in the details: Delving deep into convolutional nets[EB/OL]. (2014-11-05)[2019-10-01]<https://arxiv.org/abs/1405.3531>.
- [18] LU Xiankai, MA Chao, NI Bingbing, et al. Deep regres-

- sion tracking with shrinkage loss[C]//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany, 2018: 353–369.
- [19] LI Xin, MA Chao, WU Baoyuan, et al. Target-aware deep tracking[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 1369–1378.
- [20] LI Bo, YAN Junjie, WU Wei, et al. High performance visual tracking with Siamese region proposal network[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 8971–8980.
- [21] ZHANG Zhipeng, PENG Houwen. Deeper and wider Siamese networks for real-time visual tracking[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 4591–4600.
- [22] HONG Zhibin, CHEN Zhe, WANG Chaohui, et al. Multi-store tracker (muster): a cognitive psychology inspired approach to object tracking[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 749–758.
- [23] RUSSAKOVSKY O, DENG Jia, SU Hao, et al. ImageNet large scale visual recognition challenge[J]. International journal of computer vision, 2015, 115(3): 211–252.
- [24] ZHANG Shunli, LU Wei, XING Weiwei, et al. Using fuzzy least squares support vector machine with metric learning for object tracking[J]. Pattern recognition, 2018, 84: 112–125.

作者简介:



邵江南, 硕士研究生, 主要研究方向为目标跟踪、深度学习。



葛洪伟, 教授, 博士生导师, 主要研究方向为人工智能与模式识别、机器学习、图像处理与分析。主持和参与国家自然科学基金等国家级项目5项, 省部级重点项目4项。发表学术论文100余篇。

2021 中国人工智能大会

Chinese Congress on Artificial Intelligence 2021

2021 中国人工智能大会(CCAI 2021)以“智启非凡”为主题, 将聚焦万物智能互联时代的学术前沿与技术重点, 围绕“强化自立、硬化创新、深化融合、优化协同”四大办会目标, 邀请政府部门、科研院所、高校、科技企业的领导、专家学者、企业家、技术领军者等各界代表进行最新成果交流, 以产业共智启迪非凡未来。

作为我国人工智能领域的顶级盛会, 中国人工智能大会自2015年创办以来, 以落实国家科技创新战略为方向, 以引领智能科技发展为目标, 以打造行业交流平台为抓手, 横向聚合生态圈, 纵向打通产业链, 在成果分享和实践论道中, 不断为产学研用各方协同发展梳理主线、把脉大势。

天府之国, 智者之约。地处“一带一路”重要枢纽与战略支撑点, 成都已成为我国西南地区的人工智能创新高地, 伴随国家人工智能创新应用先导区的确立, 成都正式进入国家新一代人工智能创新发展试验区和国家人工智能创新应用先导区“双轮驱动”的新发展阶段。相信, 2021 中国人工智能大会的举办, 将为成都以及川渝地区带来“非凡”的智力资源与合作机遇。八月蓉城, CCAI 2021 期待与您同谱华章、共启非凡。

会议举办日期: 2021 年 8 月 7—8 日

会议举办地点: 中国·成都

指导单位: 中国科学技术协会、中国科学院

主办单位: 中国人工智能学会