

改进的 DeepLabV3+指针式仪表图像分割算法^{*}

杨 武 胡 敏 常 鑫 赵昕宇 余华云

(长江大学计算机科学学院 荆州 434023)

摘要:针对现有的仪表自动化读数算法占用空间大、推理速度较慢以及不能有效分割图像中密集细小目标的问题,提出改进的 DeepLabV3+指针式仪表分割算法。首先以轻量化的 MobileNetV2 来构建网络主干达到降低参数量和推理权重、提高检测速度的目的。其次通过分块并归策略设计 CSP-ASPP 结构,在保证网络性能的同时降低参数量。之后使用改进后的 SKFF 模块通过自注意力机制以非线性方式融合多尺度特征,将原网络解码器中的二尺度特征融合变为四尺度特征融合。最后使用交叉熵损失联合加权的 Dice 损失作为网络的总损失函数,解决仪表分割中各类别像素分布不均的问题。最后通过实验证明,改进后的 DeepLabV3+算法在仪表分割数据集上的平均交并比(mIoU)和平均像素准确率(mPA)达到了 89.3% 和 94.8%,相对原网络分别提高了 0.7%、0.6%,参数量和推理权重却仅有原网络的约 7%,同时在 GPU 和 CPU 上的推理速度分别达到 91 和 16 fps,解决了嵌入式设备部署困难的问题,达到了实时检测的要求,提高了仪表自动化读数的效率。

关键词:指针式仪表图像分割;DeepLabV3+;轻量化;分块并归;多尺度特征融合;Dice Loss

中图分类号: TP 391 文献标识码: A 国家标准学科分类代码: 520.2

Improved image segmentation algorithm of DeepLabV3+ pointer meter

Yang Wu Hu Ming Chang Xin Zhao Xinyu Yu Huayun

(School of Computer Science, Yangtze University, Jingzhou 434023, China)

Abstract: Aiming at the problems that the existing automatic instrument reading algorithm occupies a large space, the reasoning speed is slow, and it cannot effectively segment the dense and small objects in the image, an improved DeepLabV3+ pointer instrument segmentation algorithm is proposed. Firstly, MobileNetV2 is used to build the network backbone to reduce the amount of parameters and inference weight, and improve the detection speed. Secondly, the CSP-ASPP structure is designed through the block merge strategy to reduce the amount of parameters while ensuring the network performance. Then, the improved SKFF module is used to fuse multi-scale features in a non-linear manner through the self-attention mechanism, and the two-scale feature fusion in the original network decoder is changed to four-scale feature fusion. Finally, the Dice Loss jointly weighted by cross-entropy loss is used as the total loss function of the network to solve the problem of uneven distribution of pixels in each category in instrument segmentation. Finally, it is proved by experiments that the improved DeepLabV3+ average intersection ratio (mIoU) and mean pixel accuracy (mPA) reached 89.3% and 94.8%, respectively, increased by 0.7% and 0.6% compared with the original network, but the amount of parameters and inference weight is only about 7% of the original network, while the inference speed on GPU and CPU reaches 91 and 16 frames/s, respectively. Meet the requirements of real-time detection, which solves the problem of difficult deployment of embedded devices and improves the efficiency of automatic instrument reading.

Keywords: semantic segmentation of pointer meters; DeepLabV3+; lightweight; block-wise aggregation; multi-scale feature fusion; Dice Loss

收稿日期:2023-08-05

*基金项目:国家自然科学基金(61440023)项目资助

0 引言

指针式仪表是一种常见的测量和监测工具,广泛应用于工业、医疗、交通等领域。虽然数字化仪表在某些应用领域中已经取代了指针式仪表,但指针式仪表仍然具有直观性好、可靠性高、成本低的特点。因此,指针式仪表仍然是一种重要的测量和监测工具,具有广泛的应用前景。然而,传统的指针式仪表需要手动读数,操作繁琐、容易出错,而且无法实现实时监测和数据分析。深度学习技术可以通过图像处理技术将指针式仪表的指针位置进行自动检测和识别,从而实现自动读数,这种方法可以大大提高工作效率和准确度,因此前景非常广阔。

仪表自动化读数通常分为仪表检测、图像分割、仪表读数3部分,其中仪表检测和仪表读数的方法已经比较成熟,但是图像分割却面临许多挑战,例如细小且密集的刻度线难以准确分割、分割结果易受到环境光照强度变化等因素的影响。这些困难限制了图像分割算法的自动化程度和准确性,也是目前需要解决的问题之一。传统的图像分割方法多利用阈值分割来提取表盘图像,董海等^[1]提出一种无先验知识的仪表自动检测与判读算法,通过结合椭圆检测方法和非极大值抑制完成了仪表判读。沈卫东等^[2]采用最大类间方差进行二值化操作并基于双阈值法以达到有效的分割效果。但是这些方法不够智能化,基于深度学习的分割算法相较于传统的图像分割具有自适应性强、精度高、可扩展性强等优势,例如经典的语义分割网络U-Net多被改进用于仪表图像分割^[3-4],取得了较好的识别效果。杨诗琪等^[5]提出了一种基于YOLOv3和DeepLabv3+的指针式仪表读数识别方法提高了读数方法的有效性。陶金等^[6]通过在YOLOV4-tiny网络结构基础上添加残差模块来提高模型的鲁棒性,并对hourglass网络结构改进,达到精准识别指针式仪表读数的目的。但是在处理具有细小且密集刻度线的指针式仪表图像时,现有的语义分割算法仍然存在一些困难,同时算法也不够轻量。

DeepLabv3+是一种用于语义分割任务的深度学习模型^[7],拥有较高的准确性、多尺度信息融合能力以及高可扩展性和灵活性以适应不同的计算资源和任务需求,因此,本文在DeepLabv3+网络的基础上,对模型进行轻量化处理并融合更多尺度的特征以解决仪表图像分割的问题。

- 1) 以MobileNetV2^[8]为主干,构建一个高效而相对轻量化的特征提取网络。
- 2) 依据CSPNet^[9]的思想改进空洞金字塔池化(atrous spatial pyramid pooling, ASPP),使模型更加轻量。
- 3) 针对刻度线细小且密集的特点,在解码器阶段加入中、低层细节信息,并使用改进后的选择性卷积特征融合模块(selective kernel feature fusion, SKFF)^[10]融合多尺度特征,避免感受野过大导致细节信息丢失。

4) 使用CrossEntropy Loss联合加权的Dice Loss作为总损失缓解样本中前景背景不平衡带来的消极影响。

1 基于DeepLabV3+网络的改进

DeepLabv3+是一种用于语义分割任务的卷积神经网络模型,主要由主干网络、ASPP、解码器等部分组成,其中主干网络采用了Xception深度卷积神经网络^[11]作为特征提取器,以提取输入图像的高层次特征表示。Xception的网络结构较深,适合处理大尺寸、大规模的图像语义分割任务,而对于较小尺寸和分辨率的图像可能不适合,因此本文没有采用传统的主干网络作为特征提取器,而是以轻量级的MobileNetV2来构建骨干网络,避免细节信息丢失。

同时在解码器阶段,原网络使用了一个简单的上采样融合策略,将ASPP模块的输出上采样与低层次的特征图进行融合,以产生语义分割结果。

针对仪表分割的特点,本文在原网络解码器融合低层次特征图的基础上额外增加了一个更小层次和一个中层次特征图,使模型更关注区域内的中小目标,同时将原来简单的融合策略变为使用自注意力机制以非线性方式融合多分辨率特征。改进后的DeepLabV3+网络结构如图1所示。

1.1 主干网络的改进

Xception网络具有许多分支和较深的网络结构,这些特性使得其在训练和推理时需要大量的参数,在小规模数据集上使用可能会导致过拟合的问题,同时推理速度也较慢。目前神经网络发展迅速,其中一个重要趋势是追求具有低延迟和高吞吐量的快速神经网络,这是因为现代计算机应用需要实时性能,而快速神经网络能够为这些应用提供高效的解决方案。因此,设计低延迟、高吞吐量的快速神经网络已经成为了深度学习领域的一个研究热点,例如MobileNet等利用深度卷积(DWConv)来提取空间特征,以减少FLOPs,并采取适当的优化策略来保证网络的高效性和精度。ShuffleNet^[12-13]主要特点则是使用了channel shuffle技术,以减少计算和存储需求,提高模型的效率。GhostNet^[14]使用了分组卷积和通道拆分技术,其中通道拆分将输入通道分为两部分,幽灵通道被用于学习输入的低频信息,而主要通道则被用于学习输入的高频信息。FasterNet^[15]则提出了一种名为PConv(partial convolution)的新型卷积方法,它能够更加有效地提取图像中的空间特征。

在工业仪表自动化读数系统中,轻量化的模型具有更快的推理速度、更低的计算资源需求、更好的可移植性、更容易训练和更容易解释和调试等优点,因此选择合适的轻量级主干网络具有重要意义。本文通过将目前流行的轻量级网络移植到DeepLabV3+模型中,并通过一系列实验验证,选择出最适合的轻量级主干。实验结果表明,以MobileNetV2作为主干网络时,DeepLabV3+达到精确度

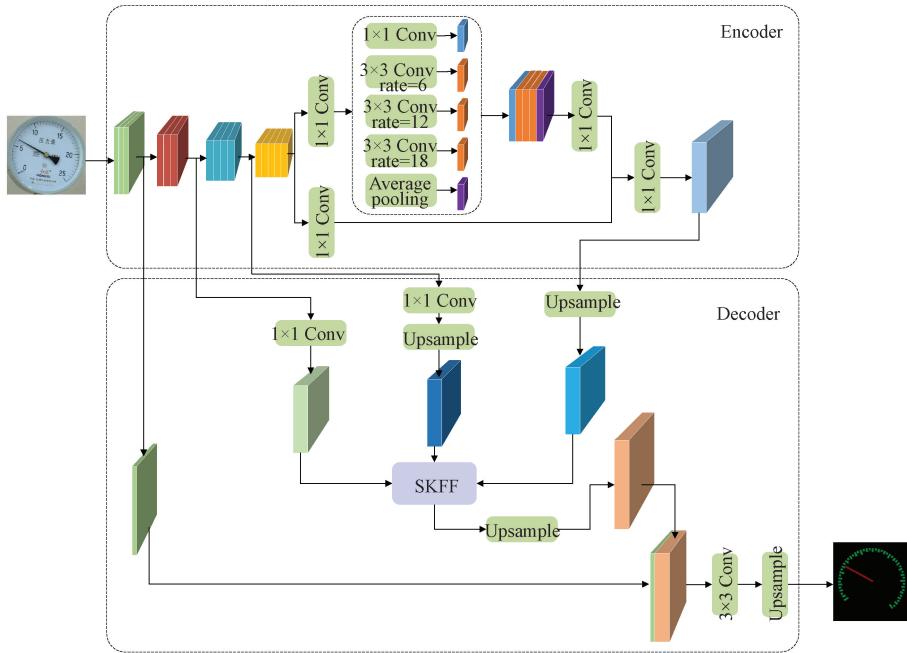


图 1 改进后的 DeepLabV3+ 结构

和推理速度的最佳平衡。

MobileNetV2 采用了倒残差的结构,该结构首先对输入进行升维,扩充通道数,然后进行深度卷积,最后再对特征图进行高维降维,由此构建一个中间较大、两端较小的梭形结构,如图 2 所示,倒残差结构的设计可以减少模型的计算量和参数数量,提高网络的性能和效率。

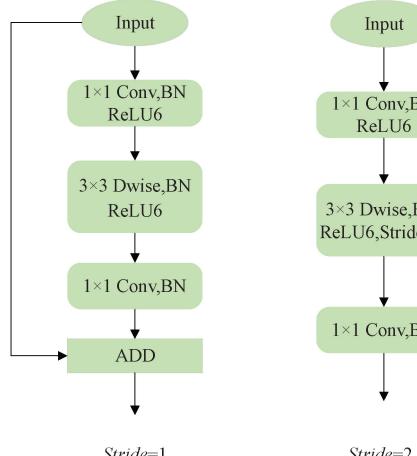


图 2 瓶颈结构

通过堆叠瓶颈结构来构建主干网络是一种高效、灵活和可扩展的方法。本文对 MobileNetV2 网络进行改进,调整下采样的次数为 4 次,从而得到了改进后的主干网络,如表 1 所示。

1.2 ASPP 结构的改进

受到 SPP(spatial pyramid pooling)结构^[16]的启发,DeepLab 网络提出了 ASPP 特征提取模块,包括一个全局

表 1 主干网络结构

输入尺寸	算子	升维倍率	通道深度	重复次数	步长
$256^2 \times 3$	Convd2d	—	32	1	2
$128^2 \times 32$	bottleneck	1	16	1	1
$128^2 \times 16$	bottleneck	6	24	2	2
$64^2 \times 24$	bottleneck	6	32	3	2
$32^2 \times 32$	bottleneck	6	64	4	2
$16^2 \times 64$	bottleneck	6	96	3	1
$16^2 \times 96$	bottleneck	6	160	3	1
$16^2 \times 160$	bottleneck	6	320	1	1

池化分支,多个不同采样率的空洞卷积分支以及一个 1×1 卷积分支。

分支之间通过连接层进行连接,从而将不同分支提取的特征进行融合。通过这些关键组件的组合,ASPP 模块可以在不同尺度上捕获丰富的信息,并将这些信息进行有效的聚合,从而提高图像分割的准确性。

但是 ASPP 庞大的分支结构带来了大量的参数和较大的推理权重,参考 CSPNet 截断梯度流防止过多重复梯度信息的思想,本文选择对 ASPP 进行轻量化处理,设计了 CSP-ASPP 结构,如图 3 所示。

与 ASPP 结构不同的是,CSP-ASPP 首先使用两个 1×1 的卷积对输入通道进行减半处理,左侧分支形成残差边,右侧分支则使用 ASPP 结构来进行多特征提取,最后两个分支再进行融合。通过分块并归策略,参与密集层卷积操作的通道数仅占原始结构的 $1/2$,解决了近 $1/2$ 的

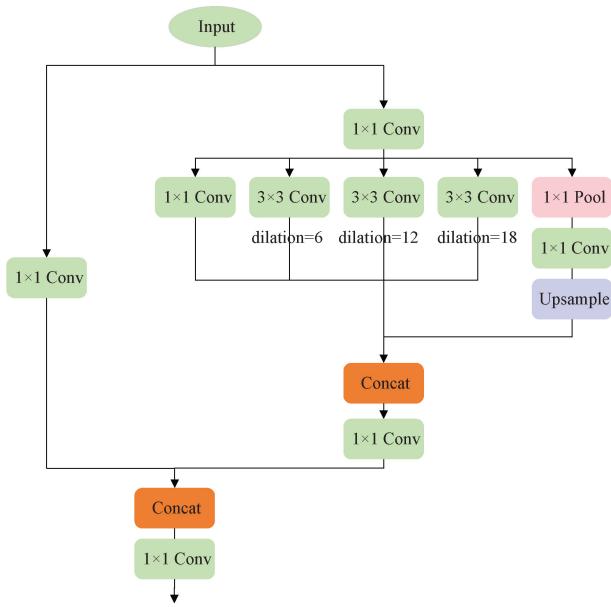


图3 CSP-ASPP结构

计算瓶颈,但是使梯度路径的数量增加了1倍,让梯度组合的差异最大化,防止不同的层学习重复的梯度信息,通过后续实验证明,改进的CSP-ASPP结构在保持准确性的同时也极大减少了参数量和推理权重。

1.3 解码器结构的改进

解码器模块主要是为了增加分辨率,原网络将低分辨率的特征图上采样到原始图像1/4分辨率,并与1/4高分辨率的特征图融合,以提高分割的精度。其中低级特征通常指的是浅层的卷积特征图,由于它们的分辨率较高,因此具有更好的位置信息。高级特征则是深层次的卷积特征图,分辨率较低,但是具有更好的语义信息。

本文为更好的对中小目标进行分割,首先在解码器中增加原始图像1/8分辨率的中层特征图输入,并将其上采样到CSP-ASPP输出相同的分辨率,为特征融合的时候提供更多的层次信息。其次原网络中特征融合的方式为通道上的串联(concat)操作,融合能力比较有限,本文利用坐

标自注意机制^[17]的思想改进多分辨率的特征融合模块SKFF,对1/4、1/8、1/16不同分辨率的特征图进行非线性融合,增强网络的表达能力。最后为进一步提高小目标分割能力,为特征图增加更多目标的位置和边界信息,将经过SKFF融合后的特征图继续上采样,并与1/2分辨率大小的特征图进行堆叠融合。最终解码器综合了来自4个不同尺度的特征,充分考虑了多个层级的特征信息。解码器的具体情况可见改进后的DeepLabV3+结构。

解码器中SKFF结构如图4所示,SKFF模块利用多分辨率流的信息,通过融合和选择两个操作动态调整感受野。融合操作将来自不同分辨率流的信息组合,生成注意力激活;选择操作利用这些注意力激活重新校准并聚合特征图。

针对高、中、低3个层次的特征图 L_1, L_2, L_3 ,融合操作首先将其按元素求和的方式组合在一起, $L = L_1 + L_2 + L_3$,考虑到仪表分割中目标位置信息的重要性,本文选择对位置信息进行编码,在 L_1 上对宽度和高度两个方向分别使用全局平均池化获得两个方向的特征图,具体来说,先使用 $(H, 1)$ 和 $(1, W)$ 的池化核沿着水平坐标和垂直坐标方向对每个通道进行编码,则高度为 h 的第 c 个通道:

$$z^h = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (1)$$

同理宽度为 w 的第 c 个通道:

$$z^w = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (2)$$

并以此得到两个不同方向的特征图,之后级联两个特征图并使用一个Conv进行降维处理,再经过卷积变换函数进行 F_1 变换:

$$f = F_1([z^h, z^w]) \quad (3)$$

之后进行不同方向上的通道分离,得到水平、垂直两个方向上的特征向量,再分别穿过3个并行的卷积恢复通道数,最后使用Sigmoid函数生成对应的注意力激活 $h_1, h_2, h_3, w_1, w_2, w_3$:

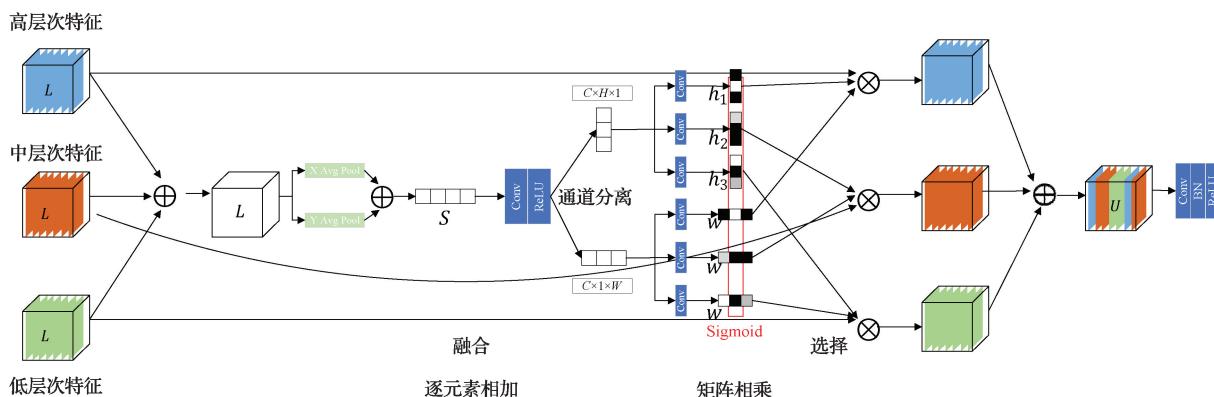


图4 SKFF结构

理论与方法

$$\begin{aligned} h_i &= \text{Sigmoid}(F_1(f^h)) \\ w_j &= \text{Sigmoid}(F_1(f^w)) \end{aligned} \quad (4)$$

选择操作使用生成的注意力激活重新校准多尺度特征图 L_1, L_2, L_3 , 特征重新校准和聚合的过程定义为:

$$U = h_1 \cdot w_1 \cdot L_1 + h_2 \cdot w_2 \cdot L_2 + h_3 \cdot w_3 \cdot L_3 \quad (5)$$

因此,SKFF 模块能够以自注意力的方式汇聚来自不同层级的特征信息,并将其用于优化特征图。

1.4 损失函数的改进

交叉熵(cross entropy, CE)损失函数可以用来衡量预测结果和真实结果之间的差异,其值越小表示模型的预测结果越接近真实结果,是语义分割中最常用的损失函数之一,也被用于 DeepLabV3+ 网络原始模型中。交叉熵的公式如下:

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log(p_{ij}) \quad (6)$$

式中: N 表示样本数量; C 是类别数; y_{ij} 表示第 i 个样本的第 j 个类别的真实标签; p_{ij} 表示第 i 个样本的第 j 个类别的预测概率。

传统的交叉熵损失函数在分割任务中有一定的局限性,在交叉熵损失函数中,每个像素的损失权重是相同的,但在实际任务中,不同的类别可能具有不同数量的像素,这意味着,如果直接使用交叉熵损失函数进行训练,模型可能会倾向于预测像素较多的类别,从而忽略掉其他类别。例如在仪表分割任务中,背景、刻度线和指针所属的像素具有很大差异,如果使用传统的交叉熵损失函数,模型可能会倾向于预测占主导地位的背景,而忽略掉其他类别的预测。

Dice Loss 使用 Dice 系数作为相似度度量,该系数的值会根据两个集合之间的相似度而变化,使得每个类别的预测在损失函数中都有不同的权重,从而可以更好地训练模型。Dice Loss 公式如下:

$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^N \sum_{j=1}^C y_{ij} p_{ij}}{\sum_{i=1}^N \sum_{j=1}^C y_{ij}^2 + \sum_{i=1}^N \sum_{j=1}^C p_{ij}^2} \quad (7)$$

联合使用交叉熵损失函数和 Dice 损失函数,则总的损失函数可以表示为:

$$L_{Total} = L_{CE} + \lambda L_{Dice} \quad (8)$$

式中: λ 为超参数,用于设置 Dice 损失函数的权重,后续将针对 λ 进行对比实验,选择出最合适的权重系数。

2 数据集准备及处理

本文数据来源于飞桨软件设计大赛提供的工业仪表检测数据集,共 414 张图片,其中包括 3 类分割目标,背景、指针和刻度线。考虑到数据集过小,容易造成过拟合导致模型的泛化性能下降,因此在预处理过程中对图像样本通过旋转、翻转、裁剪的方式扩充数据集,再人工过滤掉裁剪过程中产生的无目标图片,最终图片扩充至 3 714

张,部分数据集如图 5 所示。

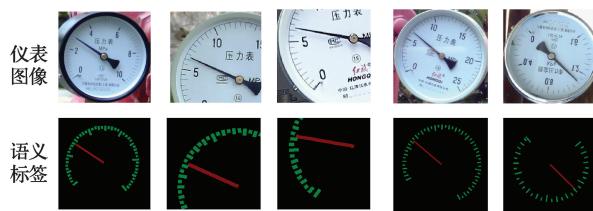


图 5 数据集示例

通过数据增强,不仅可以增加图像数据的多样性,还有助于提升模型的鲁棒性,从而获得更好的实验结果。

3 实验与分析

3.1 实验环境与评价标准

实验所用的服务器环境为 CPU 为 Intel(R) Xeon(R) Platinum 8255C, GPU 为 NVIDIA RTX 3080, 显存为 10 GB, 操作系统为 ubuntu20.04, 编译环境 Python3.8+PyTorch1.10.0+Cuda11.3。实验阶段将本文第 2 节处理后的 3 714 张图像按照 8 : 1 : 1 的比例随机划分为训练集、验证集和测试集,最后得到训练集 2 972 张,验证集和测试集各 371 张。训练参数中,选择使用 Adam 优化器,初始学习率设置为 5×10^{-4} , 循环学习率为 0.01, 动量为 0.9, 同时批尺寸设置为 8, 输入图片大小统一调整为 256×256 pixels, 采用余弦退火作为学习率的下降方式。

本文实验使用平均交并比(mIoU)、平均像素精度(mPA)作为评价算法分割效果的指标,其中平均交并比可以衡量模型对于每个类别的分割精度:

$$mIoU = \frac{TP}{TP + FN + FP} \quad (9)$$

平均像素精度可以衡量模型对于每个类别的像素分类精度:

$$mPA = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

式中: TP 代表真阳性,表示正确预测的正类别像素数量; TN 代表真阴性,表示正确预测的负类别像素数量; FP 代表假阳性,表示预测为正类别但真实标签为负类别的像素数量; FN 代表假阴性,表示预测为负类别但真实标签为正类别的像素数量。

3.2 主干网络实验分析

现如今流行的轻量化主干特征提取网络类型较多,为选择出最合适的,本文分别移植了模型大小在同量级的 MobileNetV2、MobileNetV3^[18]、ShuffleNetV2、GhostNet、G-GhostNet^[19] 和 Fasternet 作为 DeepLabV3+ 的主干网络,同时为保证实验结果的可靠性,在数据集和实验环境相同的情况下,各网络均不使用预训练权重各重新训练 200 个 Epoch,训练结果如表 2 所示。

表2 不同主干网络实验结果

网络名称	主干网络	mIoU/%	mPA/%	速度/fps		参数量/ $(\times 10^6)$	权重/MB
				GPU	CPU		
DeepLabV3+	Xception	87.2	93.7	65	7	54.7	210
DeepLabV3+	MobileNetV2×1.0	86.6	92.5	120	27	5.8	22.4
DeepLabV3+	MobileNetV3Large	86.9	92.6	82	27	5.3	20.5
DeepLabV3+	ShufflenetV2×1.5	85.9	92.8	85	29	5.4	21.0
DeepLabV3+	Ghostnet×1.0	86.7	92.9	63	23	5.3	20.3
DeepLabV3+	G-Ghostnet 002	85.7	92.1	94	36	5.8	22.3
DeepLabV3+	Fasternet T0	85.8	91.3	119	20	6.3	23.8

由表2可以看出,Xception主干在精确度方面有明显优势,但是其检测速度最慢,在CPU上甚至只有7fps,参数量庞大,权重也最大,达到了210MB,不适合部署在内存和算力较低的边缘智能设备。

MobileNetV2是MobileNet系列的里程碑,从表2可以看出,以它为主干网络具有适中的参数量、权重大小和精确度,并且在GPU上检测速度达到了120fps,CPU检测速度也较快。

MobileNetV3则是在MobileNetV2的基础上,使用了NAS网络搜索选择合适的网络结构提高资源利用,引入SENet网络并使用Hard-Swish激活函数提高精度和降低延迟,但从实验结果上看,MobileNetV3在精确度上较于MobileNetV2确实有微弱优势,且参数量和权重也较小,但是在GPU上的检测速度相比竟降低了32%,推测原因是SENet增加了模型的复杂度,且为嵌入式移动设备设计的Hard-Swish激活函数也降低了GPU检测速度。

ShufflenetV2是旷视公司针对轻量级、低延迟而推出的网络,从实验可以看出其在CPU上确实有不错的检测的速度,但是其精确度较低,综合性能相比MobileNetV2没有优势。

Ghostnet和G-Ghostnet均为华为诺亚方舟实验室在轻量化网络架构设计方面的力作,Ghostnet从特征图冗余问题出发,设计出幻影卷积及其对应模块,使用廉价操作达到了生成信息量较大的特征,移植后的网络具有较高的精确度、最少的参数量和最低的权重,但是其基本模块复杂度太高导致检测速度几乎是轻量级网络中最慢的,而它的升级版G-Ghostnet研究跨Block的冗余,利用复杂路径中的中间特征来增强廉价操作降低计算成本,确实提高了检测速度,但是精确度综合来看不是很理想。

最后移植的是Fasternet,其提出的部分卷积仅对特征图的一部分进行常规卷积操作,减少冗余计算和内存访问,提高检测速度,实验发现移植后的网络在GPU上确实具有较快的速度,而在CPU上效果不明显,分析原因是常规卷积在CPU上的速度受限,且该主干的精确度也不如其他轻量型网络。因此综合考虑下选择MobileNetV2为本文改进网络的主干。

3.3 消融实验结果分析

为验证各改进点的有效性,在数据集和实验环境相同的情况下,将各改进点的迭代次数设置为200,均不使用预训练权重,重新开始训练,具体情况如表3所示。

表3 消融实验结果

改进名称	主干网 络	CSP- ASPP	解码器	联合损失 函数	mIoU/%	mPA/%	速度/fps		参数量/ $(\times 10^6)$	权重/MB
							GPU	CPU		
DeepLabV3+	×	×	×	×	87.2	93.7	65	7	54.7	210
改进1	√	×	×	×	86.6	92.5	120	27	5.8	22.4
改进2	√	√	×	×	86.6	92.2	110	29	3.8	14.8
改进3	√	√	√	×	88.5	93.4	91	16	3.8	14.8
改进4	√	√	√	√	88.4	94.2	91	16	3.8	14.8

由表3可以看出,将主干网络更换为MobileNetV2后,mIoU和mPA指标均有不同程度的减少,但检测速度在GPU和CPU上分别提高了85%和280%,参数量大幅度减少,权重降至22.4MB,仅有原来的11%,证明了主干网络轻量化的可行性。

之后将原始的ASPP结构替换为改进后的CSP-AS-

PP后,参数量进一步减少,权重降至14.8MB,并且精确度没有明显降低,表明CSPNet网络分块并归策略应用在ASPP结构上的有效性。虽然检测速度在GPU上的有所降低,但在CPU上有少许提高,分析原因是GPU的优势在于并行计算,计算效率通常较高,CSP-ASPP额外增加的分支结构使得GPU需要额外开销保存分支结果,降低

了利用效率。而 CPU 的检测瓶颈往往和计算量有很大关系,CSP-ASPP 中每个分支中仅仅只有原来 1/2 的通道数,极大降低了参数量和推理权重,提高了 CPU 的检测速度。对解码器 Decoder 的改进使得 mIoU 和 mPA 指标相较于上一步分别提高了 1.9% 和 1.2%,这说明改进后的 SKFF 模块以及四尺度特征融合结构产生了较好的效果,使得不同尺度的特征得到了有效的融合,提高了模型对于仪表分割的精确度。为更好地分析解码器改进后的效果,此处原网络以及改进后网络的解码器中,对特征融合前后,仪表分割中最为重要的刻度线热力图进行可视化处理,如图 6 所示。

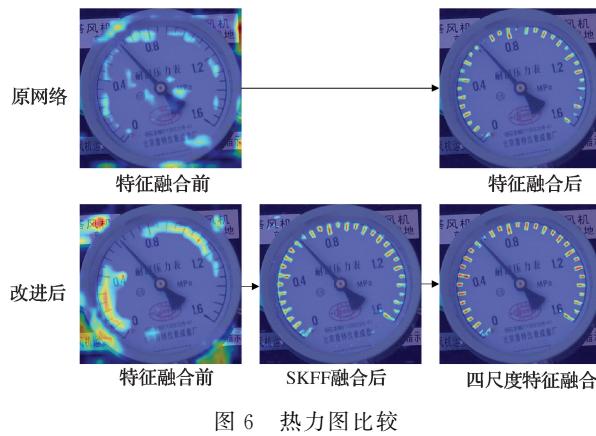


图 6 热力图比较

可以看出原网络在特征融合前关注点多为刻度线,但关注度不高。改进后的网络在特征融合前感兴趣的目标较为模糊,甚至对于背景中的黑色文字也发生了较多的关注。

经过改进后的 SKFF 模块以自注意力机制融合不同尺度的特征后,模型关注的区域发生了较大的变化,基本上侧重于刻度线,同时相比原网络特征融合后,改进后网络对于刻度线的形态也产生了较好的关注。融合第 4 个尺度特征后,模型对于刻度线形态的处理能力进一步得到了提升,已经产生了比较好的效果。

另外添加解码器的改进后,参数量和权重没有发生变化,但是检测速度下降幅度较大,分析原因是改进后的结构产生了较多的分支,提高了模型的复杂程度,极大降低了检测速度,但是尚在可接受的范围内,基本达到了实时分割的要求。

对于最后一步改进,由于改进后的损失函数存在一个超参数 λ ,本文后续对该参数进行选择实验,这里选取的是最佳超参数 λ 的实验结果,可以发现 mIoU 指标没有明显变化,mPA 则相比改进 3 提高了 0.8%。总体来看,改进后的网络相对原 DeepLabV3+ 的 mIoU 和 mPA 指标分别提高了 1.2% 和 0.5%。

3.4 损失函数实验

本文选择交叉熵损失联合加权的 Dice 损失作为总损失函数:

$$L_{\text{Total}} = L_{\text{CE}} + \lambda L_{\text{Dice}} \quad (11)$$

为选择最合适的加权系数,针对改进后的网络对超参数 λ 进行实验对比分析,在数据集和实验环境相同的情况下,仅改变 λ 的取值并充分训练 200 个 Epoch,实验结果如表 4 所示。

表 4 损失函数实验结果

λ	mIoU/%	mPA/%
0.5	88.3	93.4
1.0	88.4	94.2
1.5	88.2	94.2
2.0	88.2	93.7

当 λ 小于或大于 1.0 时,精确度均有不同程度的衰退,由此可以看出 Dice 损失的加权系数对模型的精确度也有较大影响,且当 $\lambda=1.0$ 的时候,模型拥有最佳的精确度。

3.5 预训练与迁移学习

预训练是迁移学习中的一种具体策略,通常指的是在大规模数据集上预先训练一个模型,然后将该模型的参数或特征用于特定任务的训练。

预训练的目标是通过学习大规模数据集中的通用特征,使模型具有更好的初始表示能力。因此为进一步提高模型性能,本文使用改进后的 DeepLabV3+ 网络在训练之前加载 MobileNetV2 的分类权重,然后在增强版的 Pascal VOC2012 语义分割数据集上进行训练^[20],其中增强版的 Pascal VOC2012 数据集有 21 类分割目标,训练集 10 582 张,验证集 1 449 张,训练完成后得到在该数据集上的预训练权重。

为分析预训练权重对训练结果的影响,在本文指针式仪表分割数据集的基础上,原 DeepLabV3+ 网络加载官方提供的预训练权重,改进后的网络则加载本文在 Pascal VOC2012 数据集上得到的权重,各自训练 200 个 Epoch,实验结果如表 5 所示。

表 5 预训练实验结果

网络	有无预训练	mIoU/%	mPA/%
DeepLabV3+	×	87.2	93.7
DeepLabV3+	√	88.6	94.2
本文	×	88.4	94.2
本文	√	89.3	94.8

改进后的网络在是否使用预训练权重的条件下损失值变化的曲线如图 7 所示。

从实验结果可以看出,在使用预训练权重之后模型参数不再是随机初始化,原网络和改进后的网络在仪表图像分割上的精确度都得到了较大的提升,mIoU 指标分别提升 1.4% 和 0.9%,mPA 指标则分别提升 0.5% 和 0.6%,

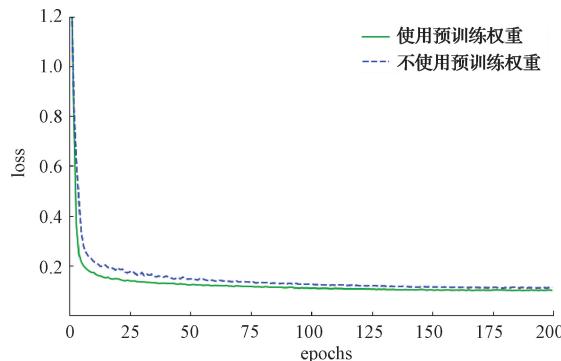


图 7 损失值对比

因此将预训练的模型作为源模型,可以在特定任务上获得更好的性能,同时从损失值变化曲线可以看出,使用预训练权重后也能更快地收敛。

除此之外,在均使用预训练权重的条件下,改进后的网络相比原始网络,mIoU 和 mPA 指标分别提高了 0.7% 和 0.6%,可见改进后网络的有效性。

3.6 对比实验分析

为进一步验证改进后的算法在仪表图像分割上的优势,本文将其与经典语义分割模型 U-Net^[21]、PSPNet^[22]、HRNetV2^[23]与 DeepLabV3+进行实验对比,在数据集和实验环境相同的情况下,均加载各自对应的预训练权重训练 200 个 Epoch,实验结果如表 6 所示。

从表 6 可以看出,经典的语义分割网络 U-Net 在 mIoU 指标上取得了最高的准确率,达到 89.7%,同时在 GPU 上的检测速度也最快,这得益于它没有额外分支的 VGG 式网络主干,但是其在 CPU 上的检测速度则较慢,仅仅只有 7 fps。综合来看 U-Net 适合特征较少、对浅层

表 6 网络对比结果

网络	mIoU /%	mPA /%	速度/fps		权重 /MB
			GPU	CPU	
U-Net	89.7	93.7	125	7	95
PSPNet	58.9	64.3	96	12	178.5
DeepLabV3+	88.6	94.2	65	7	210
HRNetV2_w18	88.9	93.1	22	7	37.5
本文	89.3	94.8	91	16	14.8

特征要求较高的数据集,比较适合本文指针式仪表分割的情况,但是其权重较大,针对算力和内存较低的嵌入式设备实际使用不是特别友好。

PSPNet 网络在实验结果上准确率最差且差距较大,推测其使用的 ResNet 这样的深层主干网络可能会导致感受野过大,从而导致特征图中的边缘信息严重损失,甚至可能完全丢失目标。因此,对于本文数据集来说,ResNet 等深层次主干网络可能不太适合。

原 DeepLabV3+ 网络的 mPA 为 94.2%,取得了较高的结果,但其推理权重最大,达到了 210 MB,检测速度上也没有优势。

HRNetV2 网络与本文改进后的算法在模型大小上属于同一量级,准确率也较高,但本文改进的算法相对 HRNetV2,mIoU 和 mPA 分别高 0.4% 和 1.7%,检测速度也较快。综合来看本文改进后的算法以最低的权重取得了较高的准确率,特别是在平均像素准确率和 CPU 上的推理速度取得了较为明显的优势。

为更全面的评估本文改进的网络,对比实验中的 5 种方法在本文测试集上进行分割结果,如图 8 所示。

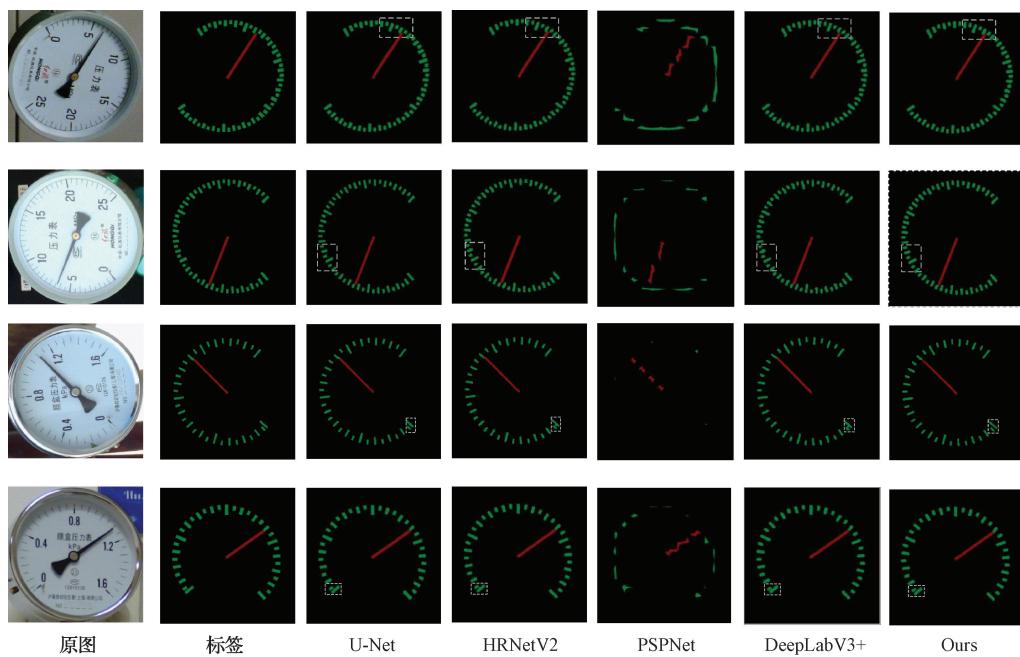


图 8 分割效果对比

由分割效果图中圈出的区域可以看出, U-Net 和本文改进后的网络分割效果最好, 对刻度线的形态处理得较好, 尤其是边缘基本上没有出现粘连融合的现象, 而其他网络都产生了刻度线边缘在不同程度上融合的现象。特别是与原 DeepLabV3+ 网络相比, 本文改进后的算法对于刻度线和指针分割得更准确。在使用 SKFF 模块融合了四尺度信息之后对于目标的边缘和位置信息处理的更好, 克服了原 DeepLabV3+ 算法对于刻度线密集导致边缘无法精确分割的问题。

4 结 论

本文针对指针式仪表细小密集的刻度线难以准确分割、检测设备内存和性能受限的问题提出了一种改进的 DeepLabV3+ 指针式仪表分割算法。在网络结构中采用 MobilenetV2 为主干, 并将原始 ASPP 替换为改进后的 CSP-ASPP 模块, 之后在解码器阶段使用改进后的 SKFF 模块以非线性自注意力机制融合不同尺度的特征, 将原网络二尺度特征融合变为四尺度特征融合, 提供了新的特征融合策略, 最后在原网络损失函数的基础上, 联合使用加权的 Dice 损失解决样本不平衡问题, 提高模型的鲁棒性。

经过实验表明, 改进后的网络在本文仪表分割数据集上的 mIoU 和 mPA 分别达到了 89.3%、94.8%, 相对原网络分别提高了 0.7%、0.6%, 但参数量和推理权重却仅有原网络的约 7%, 同时在 GPU 和 CPU 上的推理速度分别达到 91 和 16 fps, 基本解决了低内存、低算力设备应用困难、无法实时检测的问题。最后经过对比实验证明, 改进后的网络相对其他经典分割算法在本文仪表分割任务中有较大的改进, 但分割效果图上来看, 改进后的网络对于刻度线和指针分割的形态处理得还不是特别好, 因此在未来的工作中需要进一步针对仪表分割的特点优化模型的性能, 进一步提高分割的精确度。

参 考 文 献

- [1] 董海, 徐晓鹏, 谢英红, 等. 无先验知识的仪表自动检测与判读算法[J]. 仪器仪表学报, 2020, 41(2): 205-217.
- [2] 沈卫东, 李文韬, 刘娟, 等. 基于改进 Canny 边缘检测的指针式仪表自动读数算法研究[J]. 国外电子测量技术, 2021, 40(2): 60-66.
- [3] 耿磊, 史瑞资, 刘彦北, 等. 基于多重感受野 UNet 的仪表图像分割方法[J]. 计算机工程与设计, 2022, 43(3): 771-777.
- [4] 金爱萍, 袁亮, 周德勤, 等. 基于 YOLOv5 和 U-net 的指针式仪表读数识别方法[J]. 仪表技术与传感器, 2022 (11): 29-33.
- [5] 杨诗琪, 吴佳仪, 陈墨楠, 等. 基于深度学习的指针式仪表自动读数方法[J]. 电子测量技术, 2023, 46(5): 149-156.
- [6] 陶金, 林文伟, 曾亮, 等. 基于 YOLOv4-tiny 和 Hourglass 的指针式仪表读数识别[J]. 电子测量与仪器学报, 2023, 37(5): 1-10.
- [7] CHEN L C, PAPANDREOU G, SCHROFF F, et al. DeepLabv3+: Encoder-decoder with atrous separable convolution for semantic image segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018: 123-129.
- [8] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018: 4510-4520.
- [9] WANG N, LI J, ZHANG Z, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 6929-6938.
- [10] ZHANG K, ZUO W, ZHANG L, et al. MIRNet: A multi-scale information refinement network for image super-resolution[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2021: 13048-13057.
- [11] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 1251-1258.
- [12] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6848-6856.
- [13] MA N, ZHANG X, ZHENG H T, et al. ShuffleNet V2: Practical guidelines for efficient CNN architecture design [C]. Proceedings of the European Conference on Computer Vision (ECCV). ECCV, 2018: 116-131.
- [14] HAN K, WANG Y, TIAN Q, et al. GhostNet: More features from cheap operations[C]. Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 1580-1589.
- [15] CHEN J, KAO S H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 12021-12031.
- [16] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual

- recognition[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015: 346-361.
- [17] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [18] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3[C]. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019: 1314-1324.
- [19] HAN K, WANG Y, XU C, et al. GhostNets on heterogeneous devices via cheap operations [C]. International Journal of Computer Vision, 2022, 130(4): 1050-1069.
- [20] EVERINGHAM M, VAN GOOL L, WILLIAMS C K, et al. The pascal visual object classes (VOC) challenge [C]. International Journal of Computer Vision (IJCV), 2010, 88(2): 303-338.
- [21] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C]. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, 2015: 234-241.
- [22] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 2881-2890.
- [23] SUN K, ZHAO B, JIANG Y, et al. Deep high-resolution representation learning for human pose estimation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019: 5693-5703.

作者简介

杨武,硕士研究生,主要研究方向为目标检测与计算机视觉。

E-mail:201603285@yangtzeu.edu.cn

胡敏,硕士研究生,主要研究方向为计算机视觉与图像处理。

常鑫,硕士研究生,主要研究方向为计算机视觉与图像处理。

赵昕宇,硕士研究生,主要研究方向为计算机视觉与无线传感网络。

余华云(通信作者),硕士,副教授,主要研究方向为计算机视觉与无线传感网络。

E-mail:yhuayun88@sina.com