DOI: 10.13382/j.jemi.B2205553

基于 Transformer 的车道线分割算法研究

丁志江 李 丹 马志程 张宝龙
(天津科技大学电子信息与自动化学院 天津 300222)

摘 要: 车道线检测任务包含道路磨损、阴影遮挡和弯道等困难样本,这些样本中线条信息均有不同程度的缺失,使检测结果出现漏检或误检现象。基于深度学习的检测方案通过卷积操作提取特征信息。卷积操作摒弃人工设计滤波器等一系列传统图像处理的繁琐操作,得益于权重共享和归纳偏置大大减少了特征提取的工作量。该操作在缩小图像分辨率的同时获取长距离的信息,导致小分辨率的特征图损失区域边缘等细节,影响检测结果的质量。深度学习中分割模型比检测模型处理的信息更细致,本文在分割模型的基础上引入 Transformer 改进采样方式,改善卷积操作在获取全局信息上的不足。模型改进后在 Tusimple 上测试准确率提高 0.4%,像素精准度提高 0.3,乘法累加运算量增加 36.09 G。结果表明 Transformer 特有的采样方式可以改善卷积操作采样的不足,改善语义分割网络对车道线困难样本识别漏检的情况。

关键词: 车道线检测;语义分割;Transformer;U-Net
中图分类号: TN9 **文献标识码**: A **国家标准学科分类代码**: 520.20

Research on transformer-based lane segmentation algorithm

Ding Zhijiang Li Dan Ma Zhicheng Zhang Baolong
(School of Electronics and Automation, Tianjin University of Science and Technology, Tianjin 300222, China)

Abstract: The task of lane line detection includes difficult samples such as road wear, shadow occlusion and curves. The line information in these samples can be missing with different levels, which results in missed or false detection of the detection results. The detection scheme based on deep learning extracts feature information through convolution operation. Convolution operation discards a series of tedious operations of traditional image processing, such as manually designing filters, and benefits from weight sharing and inductive bias, which greatly reduces the workload of feature extraction. This operation not only reduces the image resolution, but also obtains long-distance information, resulting in the loss of regional edge and other details of the small resolution feature map, which affects the quality of the detection results. In deep learning, the segmentation model processes more detailed information than the detection model. Based on the segmentation model, this paper introduces transformer to improve the sampling method and improve the lack of convolution operation in obtaining global information. After the model is improved, the test accuracy on Tusimple is improved by 0.4%, the pixel accuracy is improved by 0.3, and the amount of multiplication and accumulation operation is increased by 36.09 G. The results show that the transformer's unique sampling method can improve the lack of convolution operation sampling, and improve the situation of missing detection of lane line difficult samples in semantic segmentation network.

Keywords: lane line detection; semantic segmentation; Transformer; U-Net

0 引 言

车道线检测作为汽车智能系统后处理的基础,具体的应用场景和服务的功能包括车道偏移预警、车道保持、路线规划、防碰撞检测和避障。一方面车道线检测的质量影响后操作的评判,质量差的检测结果导致不能及时刹车和随意变换车道的问题,另一方面车道线检测的速度影响整个系统的响应速度,高速行驶状态下的汽车无法达到毫秒级的响应影响智能系统性能的上限。车道线本身细长的外观结构就是任务难点之一,这需要方案能够在提取车道线关键点的同时,理解关键点之间的联系,

收稿日期: 2022-06-03 Received Date: 2022-06-03

即融合高层次语义信息和低层次空间结构信息。此外，生活中的车道线形状不固定，不仅会因为受到遮挡，磨损和恶劣天气等因素的影响使线条信息缺失，道路拐弯也会干扰模型预测结果。因此，整体系统的准确性、响应速度和鲁棒性需要在实际中平衡。

基于深度学习的车道线检测网络按任务类型可以分成语义分割和目标检测。目标检测模型以 RCNN[1-2] 系列模型和 YOLO[3-5] 系列模型为代表，得益于如今大量的公开数据集，车道线的识别精度和效率不断提升。其中 YOLO 系列模型更是以 95%的平均检测准确率和 48 FPS 的处理速度将车道线检测推向实际应用的阶段。但是目标检测方案采用边界框框选的方法识别目标区域，当车道线出现倾斜弯曲等情况时，网络模型难以获取完整的信息，这往往需要结合传统的图像处理方法和语义分割方法对框选出来的区域进行更细致的识别。

语义分割是将图像中的各个像素标注所属的类别，将输入图像分成若干个具有一定语义信息的部分。基于语义分割的车道线检测能够对图像进行像素级的分类，与传统的方法相比检测效果比较理想，同时很好地改善目标检测算法检测不细致的缺点。比较有代表性语义分割模型有 LaneNet[6]，该模型将图像进行逐像素的黑白二值分类的同时，将图像按像素区域分成不同的车道线实例，然后对分割结果进行处理得到车道线的数量和类别，最终输出分割的车道线图像。在 Nvidia 1080Ti 显卡上运行 LaneNet，512×256 的图像处理图像可以满足 30 FPS 的实时性要求。SCNN(spatial CNN)[7] 将传统的逐层卷积推广到特征映射中的逐层卷积进行车道线检测，增强了像素之间的信息传递，特别适合车道线这种长距离连续形状的目标检测，该算法在 tusimple 基准车道检测数据集上准确率达到 97.53%。用空洞卷积[8] 代替普通卷积可以是 SCNN 变成更轻量的网络，提高了模型处理时间上的性能，但是计算精度较低。也正因为语义分割算法处理过于细致，模型推理速度比目标检测算法稍微为慢些，并且当车道线磨损严重或者被遮挡时，导致分割结果出现断裂漏检的情况。

为了改善模型预测结果出现的断裂漏检现象，本文引入 Transformer 结构改变采样方式。在自然语言处理中计算相关性时，输入矩阵和输出矩阵的形状相同。处理图像信息时需要缩小特征图的分辨率，本文通过修改 Transformer 结构，使其实现下采样的功能，再优化语义分割模型，使采样特征兼具局部信息和全局信息，改善对困难样本的分割结果。

1 Transformer 与 SRA 模块

1.1 Transformer 结构组成

Transformer[9-10] 最初应用在自然语言处理(natural language processing，NLP)[11-12] 领域，并在该领域获得了巨大的成功。在机器翻译方向引入了基于注意力机制的 Transformer，使模型可以将所有信息并行计算，使每一层的计算复杂度更优，节省训练时间。Venugopalan 等[13] 在语言表示方向引入基于 Transformer 的 BERT 模型，该模型将文本中部分信息自动转换成掩码以供训练，以此为特征提取器的网络在当时 11 个自然语言任务上均取得了第一名。黄聪等[14] 基于巨型 Transformer 的 GPT-3 模型训练了 45 TB 的文本量，巨大的训练文本和模型参数使模型在面对不同任务时，无需调整超参数就能获得出色的表现。近年来，基于 Transformer 已经在 NLP 领域大放异彩。

Transformer 在 NLP 领域的成功得益于其可以将整个文本并行训练，这是区别于循环神经网络(recurrent neural network，RNN)[15-16] 的，而这种处理信息的方式于深度神经网络又有相似之处。因此，研究人员考虑将 Transformer 引入机器视觉(computer version，CV)领域。陈港等[17] 提出 Sequence Transformer 在图像分类任务上对比 Transformer 与 CNN 提取特征的能力，为在 CV 领域应用纯 Transformer 结构打下基础。Dosovitskiy 等[18] 最近提出 ViT(visual transformer)模型证明了纯 Transformer 可以应用在机器视觉领域。ViT 将图像差分成 16×16 的块状序列，以此模拟句子的序列顺序，使模型更好地对图像进行了特征提取。一般 Transformer 模型有注意力机制、多头注意力机制、位置编码和前向传播网络结构 4 个重要的组件。

1)自注意力机制

计算自注意力机制(self-attention mechanism，SA)需要先将输入向量转换成 3 个不同的向量，即查询向量 $\boldsymbol{Q}$、键向量 $\boldsymbol{K}$ 和维度为 $d_q = d_k = d_v$ 的值向量 $\boldsymbol{V}$。计算如式(1)所示：

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = softmax\left(\frac{\boldsymbol{Q}\cdot\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_K}}\right)\boldsymbol{V} \tag{1}$$

该公式计算 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 相似度，结果经过 softmax 归一化处理后当作权重进行赋值运算。引入 SA 后会更容易捕获全局中长距离的相互依赖的特征，而 CNN 或者 RNN 中只能计算短距离特征之间的关系，再通过若干次的缩小距离实现全局信息获取。

2)多头注意力机制

多头注意力(multi-head attention，MHA)分别计算图像中像素点或像素集合的相关性，然后融合矩阵信息，通过降低维度减少损耗。融合过程通过拼接操作实现，将分开计算的矩阵在通道维度叠加，达到统筹信息的目的。

3)位置编码

Transformer 中不用通过不断地卷积和池化压缩位置信息，早在 NLP 中简单样本已经通过限定字数等方式为

文本安排顺序。对于图像信息,Transformer 也是直接或者间接为像素或者切分好的像素集合编码。位置编码(position embedding)的方式有 3 种:1)绝对位置编码,该方法在 ViT 等论文中使用;2)用三角函数构造相对位置编码;3)使用随机值初始化的位置向量,在 BERT 等模型中使用。最常用的是三角函数构建相对位置编码如式(2)所示:

$$\begin{cases} PE(pos,2i) = \sin\left(\dfrac{pos}{1\ 000\ \dfrac{2i}{d_{model}}}\right) \\ PE(pos,2i+1) = cos\left(\dfrac{pos}{1\ 000\ \dfrac{2i}{d_{model}}}\right) \end{cases} \tag{2}$$

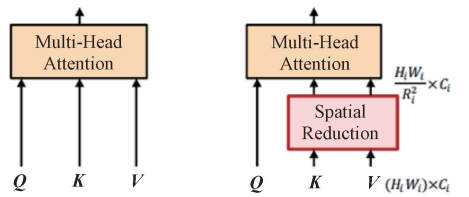
式中:pos 表示输入的位置,i 表示计算对象所在的维度,d_{modle} 表示输入向量的维度。$x = pos$ 与编码函数相交 d_{modle} 个交点作为输入图像每个像素的位置信息。

4)前向传播网络

Transformer 模型的前馈运算均以模块的形式出现,当任务复杂需要的信息量大时通过堆叠模块来获取更多的信息。因此该系列模型显的更灵活。此外模块化的设计使搭建模型时不仅可以在模型大小上做出响应调整,更可以将其只当作编码器或者解码器为 CNN 或者 RNN 服务。

1.2　SRA 采样模块

ViT 在计算机视觉领域具有很好的泛化性,但难以应用于密集预测的下游任务中。此外,受限于显存占用和模型的表示方法,ViT 在语义分割和目标追踪领域不能满足工业现场要求。出于上述考虑,PVT 模型通过 SRA 模块将 ViT 中的直通结构替换成金字塔结构。一来使特征图的分辨率不在依赖切分图像的大小,换而言之,模型的运算效率与输入图像的质量相关;同时,金字塔结构使每一层像素大大减少,从而可以灵活地调整现存开销,降低了推理速度。图 1 为实现下采样的改进方案。



(a) 多头注意力示意图　(b) 空间注意力示意图
(a) MHA sketch map　(b) SRA sketch map

图 1　MHA 与 SRA 对比图

Fig. 1　MHA vs. SRA comparison chart

0SRA 计算如式(3)和(4)所示:

$$SRA(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = Concat(Attention_1,\cdots,Attention_n)\cdot w \tag{3}$$

$Concat$ 操作通过权重 w 将分开计算的特征矩阵进行信息融合。

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = softmax\left(\frac{(\boldsymbol{Q}\cdot w^Q)\cdot(SR(\boldsymbol{K}^{\mathrm{T}})\cdot w^k)}{\sqrt{d_k}}\right)\cdot(SR(\boldsymbol{V})\cdot w^v) \tag{4}$$

式中:w^v 和 w^k 维度相同,与并行计算的分支数相关,w^q 的维度与输入图像相同,$SR(\cdot)$ 为空间缩减操作,运算如式(5)所示,该操作将输入图像分辨率从原本的 $hw \times 1$ 重新排列组合成 $\dfrac{hw}{R_i^2} \times R_i^2$。

$$SR(X) = Norm(Reshape(x,R_i))\cdot \boldsymbol{w}^s \tag{5}$$

式中:$\boldsymbol{W}^s$ 为 $\boldsymbol{K}$ 和 $\boldsymbol{V}$ 的调整矩阵,调整后的矩阵每个维度上的信息分开运算,使计算复杂度和内存消耗降低到 MHA 的 $\dfrac{1}{R_i^2}$,因此该模块可以在有限的资源下处理更大的特征图。

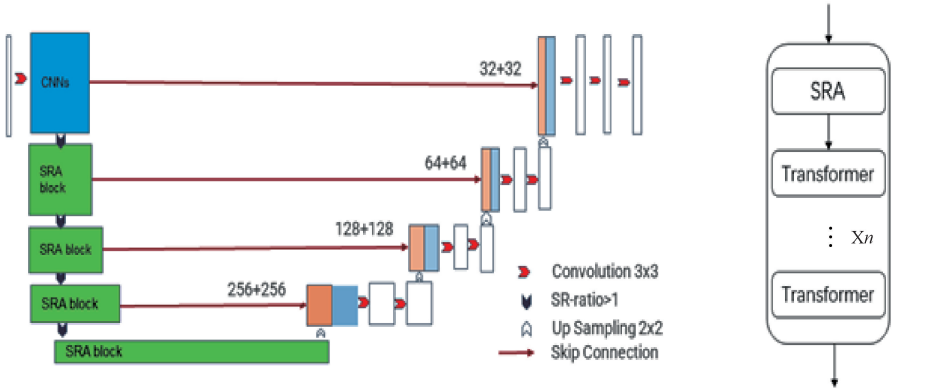
线性空间注意力(spatial reduction attention, SRA)模块模块使 Transformer 模型可以作为下游密集预测的骨干网络使用。大量实验表明,该结构比设计良好的 CNN 骨干网络效果更好。尽管 Transformer 可以作为替代 CNN 骨干网络的一个选择,但是专门针对 CNN 设计的一些模块却不能用到 PVT 中,如 SE、SK、空洞卷积、NAS 等。此外,经过多年发展,CNN 骨干网络也有很多优秀的设计,如 Res2Net 和 EfficientNet。与之相反,基于 Transformer 的模型在 CV 领域大放异彩还需要未来持续不断的研究。

2　模型改进

本文在“U”型拓扑结构基础上引入 Transformer,通过改变采样方式以获取更全面的信息。改进模型结合 U-Net[19-20] 和 Transformer,前者在特征金字塔的基础上进行多尺度融合,后者可以在每次采样过程中引入全局视野避免细节特征消失。

如图 2(a)为本实验所设计网络的结构图。结构上参照 U-Net 的骨干模块 VGG-16 进行 4 次下采样,每次下采样倍数为 2。为了在解码部分不同分辨率的特征图均能进行多尺度融合,图像输入先通过卷积模块增加特征图的维度,将特征信息投影到多维空间,再通过图 2(b) SRA block 进行下采样,减小模型计算量的同时特征提取更全面。从特征提取的方式上来看,主要由 Transformer 模块组成。图像输入网络前先进行卷积操作,不仅可以增加维度上的信息,也为解码器提供相同通道数的特征

图。相对于 DCNNs,本模型更好的把握全局信息。此外,考虑到分割对象为细长结构,不易进行大倍数的下采样,相较于 PVT 模型中下采样(4,8,16,32),在本实验选取的下采样率为(2,4,8,16)。



(a) 改进模型结构
(a) Improved model structure diagram

(b) SRA堆叠结构
(b) SRA stacking structure diagram

图 2　模型总体框架图

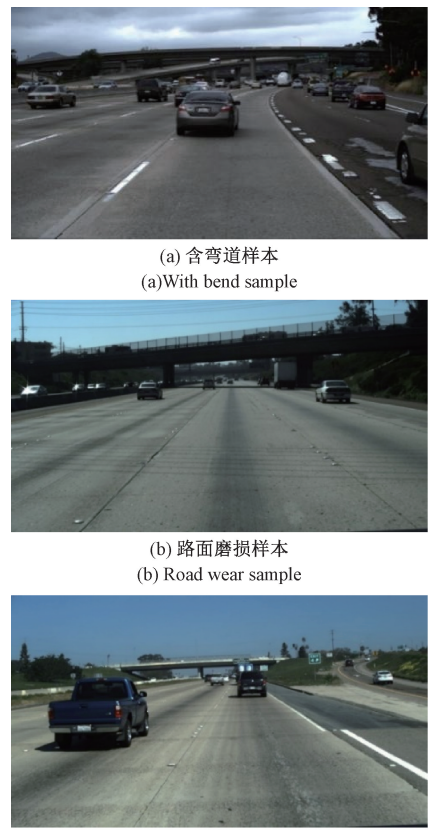
Fig. 2　General framework diagram of the model

3　实验与讨论

3.1　构建数据集

本文选择 tusimple[21] 数据集作为模型的训练材料,经统计该数据集中包含 844 张弯道图片和 2 782 张直道图片,2 736 张有遮挡图片和 890 张无遮挡图片,所有图像中的车道线均含有磨损的情况。tusimple 数据集提供足够数量的困难样本,同时为使用者公开了标注数据集的流程,方便研究人员更好地丰富数据集的多样性。经过对视频的抽帧得到样本如图 3 所示。

经过亮度变换和几何变换的数据增强操作后,共得到了 14 504 张图像,其中训练样本 11 000,测试样本 3 504 张。



(a) 含弯道样本
(a)With bend sample

(b) 路面磨损样本
(b) Road wear sample

(c) 有遮挡样本
(c) Obscured sample

图 3　Tusimple 数据集样本图

Fig. 3　Sample graph of Tusimple dataset

3.2　评估指标

在评价语义分割算法的效果时,通常会用相应的指标来衡量。这些指标与 TP(true positive)、FP(false positive)、TN(true negative)和 FN(false negative)4 个统计量相关。TP 表示预测对象为实例并且结果为实例,FP 表示预测对象为背景并且结果为实例,TN 表示预测对象为实例并且结果为背景,FN 表示预测对象为背景并且结果为背景。精度是评价图像分割网络最主要也是最流行的技术指标,这些精度估算方法各种不同,但是主要可以分为两类,分别是基于像素精度(pixel accuracy,PA)和基于交并比(intersection over union,IoU)。当前最流行的语义分割方法评估都是基于像素标记为基础完成的。PA 表示预测类别正确的像素数占总像素数的比例,重点判断标准为是否检测到了物体。其计算如式(6)所示:

$$PA = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

IoU 即交并比，在语义分割中作为标准度量一直被人使用。交并比不仅仅在语义分割中使用，在目标检测等方向也是常用的指标之一。计算如式(7)所示：

$$IoU = \frac{TP}{TP + FP + FN} \tag{7}$$

MACC(multiply-accumulate operation)表示乘加运算。模型中矩阵运算居多，矩阵运算基本都是乘加。因此，统计 MACC 可以减少由硬件系统对模型推理速度影响。

3.3　实验结果

实验针对采样深度和特征图通道两个与参数量相关的配置上做了对比试验，分析参数量变化对不同采样方式下模型性能的影响。

1)减少特征图通道对比实验

采样过程中通道数越多，对应尺寸的特征图的信息量越大，提取特征的计算量越大，系统需要更多内存空间存储更多的中间变量并进行信息交互，导致减缓模型推理速度。通过改变特征图的通道数，对比减半通道数量前后模型推理的准确度和内存占用量，分析其对模型性能的影响。

表 1　通道数变化的模型性能表

Table 1　Model performance table for channel number variation

	IoU	PA	Acc/%	Strg/MB	Pars/M
原 U-Net	0.57	0.79	98.1	124.3	31.04
改 U-Net	0.54	0.80	97.8	111.6	7.77
原 U-Net-T	0.58	0.80	98.5	101.4	56.08
改 U-Net-T	0.54	0.82	98.5	89.0	20.03

表 1 中“原 U-Net”表示 U-Net 原网络模型，“改 U-Net”表示模型在采样时的通道数量上做了修改。“原 U-Net-T”表示引入 Transformer 结构改进的 U-Net 模型，“改 U-Net-T”表示模型改进后在采样通道数上做了修改。对比表格中的数据，可以得到如下结论：1)特征图通道数的变化大大减少了参数量，且在针对具体任务时识别准确率变化不大；2)引入 Transformer 结构增加模型的参数量，当特征图通道数变化时，预测准确度的变化幅度小于 CNN 模型；3)对比权重文件可以发现，引入 Transformer 结构虽然增加了推理过程中的中间变量，但是减少了最终存储的权重和偏置参数量。

2)采样深度变化模型性能对比

CNN 模型在提取特征时，采样深度严重影响模型拟合效果。一般来说，增加网络深度能使 CNN 模型获取更抽象、更高层次且更全局的信息，而 Transformer 凭借自身 SA 机制给模型提供了浅层化采样的可能。

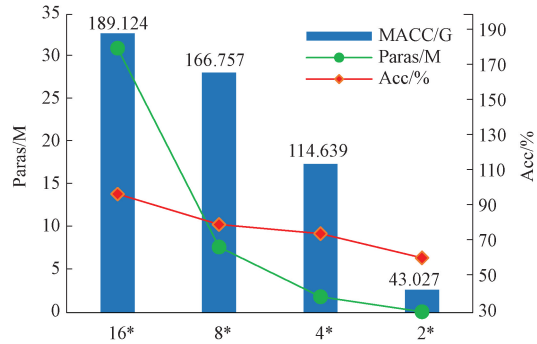


图 4　U-Net 性能变化图

Fig. 4　U-Net performance change chart

从图 4 中可以发现：(1)采样深度减少，参数量的变化趋势是先陡后缓，与内存响应时间的变化趋势刚好相反。由此可知内存被占用的越来越多会影响后续数据处理的速度，进而减慢模型的响应；(2)采样率每减少 2 倍，识别的准确率随之下降 10%左右，验证了 CNN 网络模型需要通过下采样来获取更全面的信息，采样率不够意味着模型难以拟合出好的预测结果。

基于 Transformer 模型在采样深度变化时各指标的变化，从图 5 中可以发现：(1)从曲线变化的趋势可以看出，引入 Transformer 后各项指标变化平缓，这得益于该结构本身每次采样都对全局特征和局部特征有整体的把控。但与 CNN 中相同的是采样深度增加，参数量会几何倍数增加；(2)从 MACC 可以看出 Transformer 相对于 CNN 会有更多的参数堆积在内存中，这样影响了硬件以最好的状态运行，因此模型运行速度相对较慢；(3)模型改进后即使采样深度不够，准确率不会剧烈下跌。考虑到对于对比的模型结合的是配置最小的 Transformer 结构，如果在每个特征图尺寸叠加更多该结构，识别率会有更好的表现。

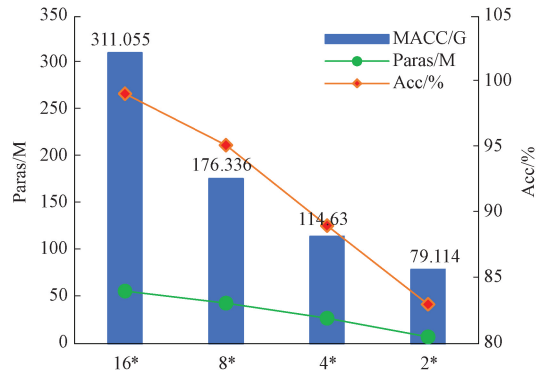


图 5　改进模型性能变化图

Fig. 5　Improved model performance change graph

3)图像测试结果对比

第1列为待分割原图,第2列为U-Net分割结果图,第3列为模型改进后的分割结果图。从图6(a)中可以看出,当路面因为磨损导致车道线模糊时,U-Net模型的分割结果断断续续,图像边缘处分割的精准度会低于改进后的模型效果。纵向比较U-Net的分割结果,分割的不完整情况会随机出现在各个车道线上,鲁棒性不高。模型改进后预测结果更加平滑,只有当车道线被严重磨损时,结果才会出现断裂现象。

在图6(b)含弯道的困难样本测试中,改进前后模型的分割准确度相差不大。当弯道处于图像边缘时,U-Net模型的分割效果比较差。模型改进后,Transformer结构可以对上下文信息进行预测,边缘信息得以延申,体现出了Transformer从周边信息预测目标区域的优势。

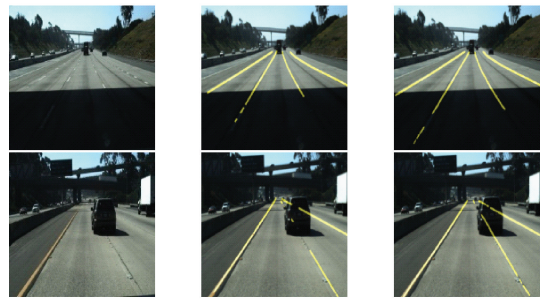
从图6(c)中可以看出,含遮挡的样本分为两类,一种是前方行驶的汽车压在车道线上,另一种为桥和车等阴影遮住车道线。U-Net对以上情况均难以保证好的分割准确率,尤其是有桥的阴影遮挡时,分割的结果会出现更多的漏检或断裂。相对而言,改进后的模型能很好地改善两种情况。当车道线被遮挡的长度比较短时,被遮挡部分的信息可以通过前后文信息进行预测,补足信息使预测结果更连贯;当被遮挡长度比较长时仍需要更多的信息进行预测,造成车道线断裂。因此,Transformer弥补了CNN模型过度依赖标签的缺点,体现了该结构能更好地把握全局信息。

(a) 路面磨损样本分割图

(a) Segmentation diagram of road surface wear sample

(b) 含弯道样本分割图

(b) Split chart of bend sample



(c) 含遮挡样本分割图

(c) Split chart of masked sample

图6　困难样本分割图

Fig. 6　Difficult sample segmentation diagram

4　结　论

车道线分割是高级驾驶员辅助系统的重要组成部分,分割结果直接影响车道保持和车辆启动等任务。分割算法应该有很好的鲁棒性,在实践中,该模型应该能够在不同的道路条件下准确地进行识别。一个好的模型还应该克服车道标记本身的模糊和弯曲等情况。此外,模型结构化等要求也是易于部署的先决条件。本文通过引入Transformer,验证了该结构在采样过程中能更好地融合局部信息和全局信息,并在保证了识别准确性的基础上通过减少参数数量减少了模型文件的内存占用。

未来,网络在非理想条件下的分割效果将进一步优化,同时需要从克服Transformer中自注意力机制计算的复杂度大的问题,朝着实际应用的方向发展。

参考文献

[1] MITTAL P, SHARMA A, SINGH R, et al. Dilated convolution based RCNN using feature fusion for low-altitude aerial objects [J]. Expert Systems with Applications, 2022, 199(199): 1-3.

[2] 杨浩杰, 王璐, 杨省伟. 一种基于特征融合的道路目标检测方法[J]. 长沙大学学报, 2022, 36(2): 1-6.
YANG H J, WANG L, YANG SH W. A road target detection method based on feature fusion[J]. Journal of Changsha University, 2022, 36(2): 1-6.

[3] SHIN D J, KIM J J. A deep learning framework performance evaluation to use YOLO in nvidia jetson platform[J]. Applied Sciences, 2022, 12(8): 3734.

[4] PENG Z L, CHEN G P, ZHOU Y C, et al. Millimeter wave image detection based on YOLOv3-tiny [J]. Scientific Journal of Intelligent Systems Research, 2022, 4.

[5] JIANG P, ERGU D, LIU F, et al. A review of YOLO algorithm developments[J]. Procedia Computer Science, 2022, 199: 1066-1073.

[6] SEYED M A, PETER F, MARCO K, et al. Aerial laneNet: Lane-marking semantic segmentation in aerial

imagery using wavelet-enhanced cost-sensitive symmetric fully convolutional neural networks [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(5): 2920-2938.

[7] 单兆晨, 黄丹丹, 耿振野, 等. 结合 Spatial CNN 的端到端自动驾驶研究[J]. 长春理工大学学报(自然科学版), 2021, 44(3):102-108.
SHAN ZH CH, HUANG D D, GENG ZH Y, et al. Research on end-to-end automatic driving combined with spatial CNN [J]. Journal of Changchun University of Technology (Natural Science Edition), 2021, 44(3): 102-108.

[8] ZHANG C J, XU S H, JIANG T, et al. Integrating normal vector features into an atrous convolution residual network for LiDAR point cloud classification[J]. Remote Sensing, 2021, 13(17):3427.

[9] XU Y F, WEI H P, LIN M X, et al. Transformers in computational visual media: A survey[J]. Computational Visual Media, 2022, 8(1): 33-62.

[10] HUANG Q, HUANG C, WANG X, et al. Facial expression recognition with grid-wise attention and visual transformer [J]. Information Sciences, 2021, 580: 35-54.

[11] 李琳, 董璐璐, 马洪超. 基于 BERT 的汉语作文自动评分研究[J]. 中国考试, 2022, (5): 73-80.
LI L, DONG L L, MA H CH. Research on automatic scoring of Chinese composition based on BERT [J]. Chinese Examination, 2022, (5): 73-80.

[12] ZIRIKLY A, DESMET B, NEWMAN-GRIFFIS D, et al. Information extraction framework for disability determination using a mental functioning use-case [J]. Jmir Medical Informatics, 2022, 10(3): E32245.

[13] VENUGOPALAN M, GUPTA D. An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis [J]. Knowledge-Based Systems, 2022, 246:108668.

[14] 黄聪, 郭杭. 基于 GPT3 模型的 ZTD 及 PWV 反演精度分析[J]. 大地测量与地球动力学, 2022, 42(5): 489-493.
HUANG C, GUO H. ZTD and PWV inversion accuracy analysis based on GPT3 model [J]. Geodesy and Geodynamics, 2022, 42 (5): 489-493.

[15] 李华旭. 基于 RNN 和 Transformer 模型的自然语言处理研究综述[J]. 信息记录材料, 2021, 22(12):7-10.
LI H X. A review of natural language processing research based on RNN and transformer model [J]. Information Recording Materials, 2021, 22 (12): 7-10.

[16] AYUB S, KANNAN R J, ALSINI R, et al. LSTM-based RNN framework to remove motion artifacts in dynamic multicontrast MR images with registration model [J]. Wireless Communications and Mobile Computing, 2022, 2022.

[17] 陈港,张石清,赵小明. 采用 Transformer 网络的视频序列表情识别[J]. 中国图象图形学报,2022,27(10): 3022-3030.
CHEN G, ZHANG SH Q, ZHAO X M. Expression recognition of video sequences using transformer network[J]. Journal of Image and Graphics, 2022,27(10):3022-3030.

[18] DOSOVITSKIY A, FISCHER P, SPRINGENBERG J T, et al. Discriminative unsupervised feature learning with exemplar convolutional neural networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2016,38(9).

[19] HEN C, LEI J, CHANG Y W, et al. Deep U-Net architecture with curriculum learning for myocardial pathology segmentation in multi-sequence cardiac magnetic resonance images [J]. Knowledge-Based Systems, 2022, 249:108942.

[20] 李晨曦, 李健. 基于 GAN 和 U-Net 的低光照图像增强算法[J]. 计算机系统应用, 2022, 31(5): 174-183.
LI CH X, LI J. Low illumination image enhancement algorithm based on GaN and U-Net [J]. Computer System Application, 2022,31(5): 174-183.

[21] PETE B. TuSimple embracing self-driving challenges[J]. Automotive News, 2022,96(7021):58-69.

作者简介

丁志江,2022 年于天津科技大学获得硕士学位,现为苏州镁伽科技 AI 算法工程师。主要研究方向为测试计算技术与仪器与人工智能包括图像分割、关键点检测与目标检测。
E-mai: dzj19960727@ 163. com

Ding Zhijiang received M. Sc. degree from Tianjin University of Science and Technology in 2022. Now he is an AI algorithm engineer of Suzhou MGGA Technology. His main research interests include testing and computing technology and instruments and artificial intelligence, including image segmentation, key point detection and target detection.

李丹,1999 年于南开大学获得学士学位,2002 年于南开大学获得硕士学位,2007 年于香港大学获得博士学位,现为天津科技大学副教授,硕士生导师。主要研究方向为光电检测与计算机视觉检测。
E-mail: lidan@ tust. edu. cn

Li Dan received a B. Sc. degree in 1999 from Nankai University, received a M. Sc. degree in 2002 from Nankai

University and received a Ph. D. degree in 2007 from Hong Kong University. Now she is an associate professor and a M. Sc. advisor at Tianjin University of Science and Technology. Her main research interests include photoelectric detection and computer vision detection.

马志程,2020 年于辽宁石油化工大学获得学士学位,现为天津科技大学硕士研究生。主要研究方向为光电检测与计算机视觉检测。

E-mail: 1826831029@ qq. com

Ma Zhicheng received his B. Sc. degree in 2020 from Liaoning Shihua University. Now he is a M. Sc. candidate at Tianjin University of Science and Technology. His main research interests include photoelectric detection and computer vision detection.

张宝龙(通信作者),1999 年于南开大学获得学士学位,2001 年于香港科技大学获得硕士学位,2006 年于香港科技大学获得博士学位,现为天津科技大学教授,硕士生导师,主要研究方向为集成电路设计及半导体制备工艺的研究。

E-mail: eezbl@ tust. edu. cn

Zhang Baolong(Corresponding author) received his B. Sc. degree in 1999 from Nankai University, received his M. Sc. degree in 2001 from Hong Kong University of Science & Technology and received his Ph. D. degree in 2006 from Hong Kong University of Science & Technology. Now he is a professor and a M. Sc. advisor at Tianjin University of Science and Technology. His main research interests include IC design and semiconductor fabrication process.