 DOI: https://doi.org/10.12349/iser.v7i1.9027

Road Scene Semantic Segmentation Algorithm for Complex Weather Condition

JinLong Chen

College of Artificial Intelligence, Gansu University of Political Science and Law, Lanzhou, Gansu, 730070, China

Abstract

With the widespread application of intelligent unmanned driving technologies, road drivable area detection has become increasingly important. In complex climatic environments, various weather-related interfering factors can lead to low detection accuracy and high false positive rates. To address this, a road semantic segmentation algorithm for complex weather conditions is proposed. First, a Cascaded Attention Encoder module is designed. It employs a cross-stage efficient attention mechanism to achieve global road boundary modeling and collaborative perception for segmenting various objects. Images are split into patches via overlapping embedding blocks and fed into a Transformer encoder to obtain multi-level features at the original image resolution.. Next, a lightweight decoder called Light ALL-MLPs is introduced. Its progressively stacked modules gradually generate the output sequence, Finally, a Gram matrix impurity filter is used to extract factors related to fog, rain, snow, and sandstorms. This reduces the discrepancy between images with different impurity levels within the segmentation model's style space. Combined with feature fusion, it achieves adaptive filtering of fog, rain, snow, and dust impurities.

Keywords

Drivable Area Detection; Semantic Segmentation; Impurity-pass Filtering; Cascade attention

复杂天气下自动驾驶道路语义分割识别算法

陈晋龙

甘肃政法大学人工智能学院，中国 · 甘肃 兰州 730070

摘　要

随着智能无人驾驶技术的广泛应用，道路可行驶区域检测变得越来越重要。针对复杂气候环境中，各种天气的干扰因素均可能导致检测精度低下和高误检率。为此，提出了面向复杂天气环境下的道路语义分割算法。首先，设计了级联注意力编码器模块，通过跨阶段高效注意力机制实现全局道路边界建模和分割各类物体的协同感知，通过重叠嵌入块分割成补丁输入给Transformer编码器以获得原始图像分辨率的多级特征多分支自注意力；接着，引入了一个Light ALL-MLPs轻量级的解码器，逐层堆叠的模块逐步生成输出序列。最后，通过Gram矩阵杂质滤波器提取雾、雨、雪、风沙等相关因素，减小不同杂质度图像在分割模型风格空间中的差异，结合特征融合实现雾、雨、雪、沙尘杂质的自适应过滤。

关键词

可行驶区域检测；语义分割；杂质过滤；级联注意力

1 引言

随着人工智能技术的快速发展，语义分割和目标跟踪技术的结合促进了自动驾驶技术迅速发展，目前的摄像头可以获取道路场景的可行驶区域，为安全驾驶提供了安全保障。虽然激光雷达技术更为先进，在考虑成本太情况下，基于单目摄像头的语义分割技术具有更好的应用价值。目前主流分割技术网络有 OpenMMlab 社区的 MMsegmentation、SegFormer[1]（Xie 等，2021）、DeepLabV3+[2]（Chen 等，2018）、ResNet[3]（He 等，2016），这些主流的分割技术拥有很高的精确度，但是在强光或者雷雨特殊天气下会出现明显的道路分割误差，算法无法适应复杂交通道路环境，在实际使用中局限性较大。

【作者简介】陈晋龙（2000–），男，中国河北保定人，硕士，从事计算机视觉研究。

2 相关工作

2.1 总体网络结构

随着视觉 Transformer 模型的提出，众多研究人员开始专注于语义分割模型的优化工作。通过持续的模型迭代与改进，模型在提升分割精度和效率方面表现出显著优势。本文提出的级联注意力分割模型是由骨干网络、多分支自注意力机制模块、级联注意力编码器和 Light ALL-MLPs 轻量级解码器四个部分组成。主体框架的算法是基于混合型 Transformer 编码器，采用主干网络对输入图像进行深层语义特征提取。

首先，通过重叠块嵌入对图像进行编码时将图形划分为多个具有重叠块的局部块，每个局部块为固定的大小矩形区域，相邻块之间通过设定一定的步长实现重叠，这种架构设计提升了模型在细节处理与全局信息捕捉上的能力，使其在复杂图像场景下表现出更高的分割精度。再引入杂质过滤模块用于提取复杂天气图像中的低光、雨痕、雾霾等杂质特征，并通过嵌入向量精确量化，辅助模型学习和表征，从而达到过滤杂质的目的，学习风格中提取的与杂质相关的因素。最后，采用交叉熵损失函数，综合考虑了语义分割中边界模糊、前后空间关系的因素。总体设计为编码 - 解码器的架构，使用级联分割网络先进行粗略分割任务，再细化具体边缘的分割细节，通过逐层优化的方式逐渐提高质量。再引入补丁级判别器，以关注图像中更细小的区域特征，从而提升对局部区域一致性及细节质量的检测能力。随后，通过纹理判别器进一步细化处理，将关注范围缩小至更小的图像区域，以增强对局部细节的分割精度。

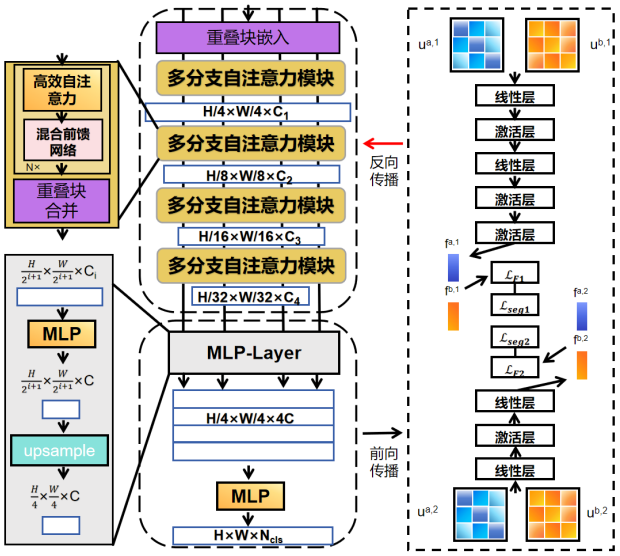


图 1 级联自注意力机制总体模型结构

2.2 级联注意力编码器模块

该模型通过多分支结构构建全局和局部自注意力窗口，以提高特征建模的灵活性和表达能力。每个输入的嵌入维度（Embedding Dim）、头数（Head）和维度（Dim）都经过精心定义，以确保输入能够有效划分为窗口块和数字序列（Sequence）。通过这种结构设计，模型能够在全局范围内捕捉长距离依赖，同时在局部范围内聚焦细节，提升了特征的多尺度表达能力，进一步增强了自注意力机制的效果。这些划分好的序列被转化为感受野（Receptive Field）分辨率，从而为模型提供了对图像不同局部和全局关系的深入理解。通过这种方式，模型能够更精确地捕捉图像中复杂的空间依赖和语义联系，使得每个序列在感受野内的表现更加细致，进一步增强了图像特征的表达能力和模型的推理能力。

在复杂的城市交通环境中，行人、非机动车和机动车的高密度共存通常会导致车道线出现遮挡或缺失现象。此外，恶劣天气条件及夜间光照不足等因素也对复杂道路的检测与语义分割带来了极大挑战。在研究现有道路感知算法[15]方法的基础上，设计出一种改进的 Transformer 结构。该结构针对道路环境中的特定挑战进行了优化，能够更有效地捕捉多尺度空间特征和语义信息，从而提升复杂场景下的感知精度和鲁棒性。

在多分支自注意力模块中，首先对图像进行块级序列化处理，将其转换为序列格式后输入到双分支通道中。每个通道接着通过 1×1 卷积层，实现通道数的调整并引入非线性变换，以增强特征的表达能力和模型的灵活性。此过程不仅有效地适配了通道维度，还显著增强了特征表示的非线性表达能力，从而提升了模型对复杂模式的捕捉和适应能力。在卷积操作后引入批归一化（BN）层，以确保图像块中的各像素维持在一致的数值范围内，从而加速训练过程并提升模型的稳定性与收敛性。为降低计算负担，模型引入了 Transformer 的多头自注意力机制以完成整体运作。此机制能够在不同头之间并行计算注意力权重，从而高效捕捉特征间的关联，显著提升模型的计算效率与特征表达能力。表示为：

$$Attention(Q、K、V)= (head_0 \oplus ,...,\oplus head_n)W^0 \quad (1)$$

通过使用爱因斯坦求和约定（einsum）[14]计算键值对的矩阵积，并得到注意力得分矩阵。该模型采用高效的多头注意力机制，每一个 head 都包含查询（Query, Q）、键（Key, K）、值（Value, V）三个关键元素。各个 head 可以并行地捕捉不同的特征关系，增强了模型对输入特征的全面理解。多层注意力结果最终汇聚为总体输出注意力，使得模型能够从不同视角整合信息，从而更精确地建模输入特征间的复杂依赖关系。使用掩码（Mask）将其注意力用到注意力得分矩阵上，再采用 softmax 函数对其得分进行归一化处理。

$$Attention(Q、K、V)=\sum_{i=1}^{n} Softmax(\frac{QK^T}{\sqrt{d}}) \quad (2)$$

在该模型中，K 和 V 分别表示通过特征映射生成的键（Key）和值（Value），而 Q 表示对应的查询（Query）。这些映射通过模型能够有效计算输入特征之间的关系，对特征进行精细化处理，从而得到更具判别力的输出特征。在多头注意力机制中 $head_0$ 到 $head_n$ 中，一组 $head_0$ 到 $head_{n/2}$ 用于高频特征从而精确捕捉细节；另一组 $head_{n/2}$ 到 $head_n$ 用于低频特征以抓取整体结构信息，并在最终进行拼接（Concat）提供全局信息。*W0* 表示输出的注意力权重矩阵，用于整合多头注意力机制的各个头的结果，从而生成最终的特征表示。双通道特征通过矩阵逐元素相加实现拼接融合，随后引入 ReLU 激活函数层以增加非线性，从而有效缓解梯度消失问题，提升了模型对复杂特征的捕捉能力，使其能够更准确地表达多层次信息。特征图经过一个 5×5 的最大池化层处理，最大池化操作通过选择局部区域中的最大值，既能有效

减少特征图的尺寸，又能够保留图像中的重要空间信息。这种方法使得模型对显著特征更加敏感，同时增强了对细节信息的捕捉能力，有助于提高模型的鲁棒性和泛化能力。接下来，引入批归一化（BN）层以进一步标准化数据特征。批归一化层通过调节数据的均值和方差，确保各特征维持在相对稳定的数值范围内，从而加速模型的训练过程，提升模型的稳定性和泛化能力。

$$F(x_i,x_j)=\frac{x_i * x_j}{|x_i||x_j|} * g(x_i,x_j) \quad (3)$$

输入的张量 X 是由 xi,xj 两个不同的矢量组成，$g(xi,xj)$ 表示张量的拉伸，并通过一层前馈神经网络逼近输入和输出之间的复杂映射关系。模型能够根据不同层次的特征来提升特征信息的提取能力，从而更好地捕捉输入数据的深层语义。最终，通过 Sigmoid 激活函数层将输出压缩到 (0, 1) 的范围内，从而生成归一化的图像特征。这一步确保了输出特征的数值稳定性，并为后续任务提供了标准化的特征表示，有助于提升模型的泛化能力和预测效果。

2.3 轻量化 Light ALL–MLP 解码器

2.3.1 解码器结构

多分支自注意力模块的输出（特征序列）形成四个通道，直接作为解码器模块的输入，确保信息在编码器和解码器之间顺畅传递。轻量级 MLP 解码器，仅仅由 MLP 层组成，为了避免其他方法中通常使用的计算量巨大和人工设计的组件模块。四个不同分辨率的通道特征，经过 MLP-Layer 层内部的上采样层拼接为 4 个相同分辨率的通道数，统一为一样的分辨率，通过编码器的多级特征进行多维度统一。接着经过 MLP 模块，负责对提取的特征进行解码和非线性变换，通过样条函数和多层结构进一步提升特征的表达能力，提取出语义的掩码 M 所需要的数量 N_c，上采样后的特征被进行拼接。最后经过与杂质过滤模块进行相互反向和正向传播，消除雾霾、雨痕、低光等杂质，完成最后的优化实现良好的语义分割。该解码器通过减少卷积操作，仅依赖轻量级 MLP 结构实现高效计算，在降低参数量的同时保持高精度分割性能，适用于资源受限场景下的实时语义分割任务。

2.3.2 MLP-Layer

给定分辨率为 $H\times W\times 3$ 的输入图像进行补丁合并操作后，获得分辨率 $H/2^{i+1}\times W/2^{i+1}\times C^i$，i 是属于集合 {1,2,3,4}，补丁操作是从 F1（$H/4\times W/4\times C_1$）缩小到 F2（$H/8\times W/8\times C_2$），然后对层次结构中的任何其他特征图进行迭代，用于合并非重叠的图像或特征补丁），设置 K（补丁大小）=7，S（相邻补丁之间的步长）=4，P（填充大小）=3，K=3，S=2，P=1，执行重叠补丁合并，已生成与非重叠过程相同大小的特征。输入至 MLP-Layer 中将属于集合 {1,2,3,4} 的四层分辨率 F_i 进行通道维度统一被特征上采样（Upsample）成同样大小即原来的 1/4，并进行拼接成 $H/4\times W/4\times 4C$，接着采用 MLP 层来融合拼接后的特征 F。最后，外部的 MLP 层将融合的特征输入，预测分割掩码 M，分辨率为分辨率为 $H/4\times W/4\times Ncls$，cls 是类别数量，达到既能有效保持较大的感受野又能有效捕捉上下文信息，且设计不复杂模型较为简洁。

2.3.3 杂质过滤器模块

引入一个杂质过滤器模块，其关键是将图像的杂质条件视为其风格，并缩小分割模型的神经风格空间中不同杂质条件图像之间的差异，特别是由于图像神经网络中的风格通常受其他因素和杂质的影响不断地学习风格中的因素，通过对杂质过滤器模块和分割模型交替优化，逐渐缩小了不同环境下的风格差异。解码器模块输出的数据 $H\times W\times N_c$，完成雨痕、雾霾等杂质特征的过滤。首先，本模块完成了复杂天气条件下的图像中提取，并通过嵌入杂质向量因子的形式实现精准量化，以辅助模型学习和表征各种杂质特征。线性层层是对输入特征执行了一次线性变化，激活层层在处理负输入时不会完全抑制信号传递。在该模块中，采用 Gram 矩阵对特征通道之间的相关性进行建模，其中矩阵元素 G_{ij} 表示第 i 通道和第 j 通道之间的相关性，计算公式为 $G_{ij}=a_i^T a_j$。这种方法提供了丰富的风格信息，有助于捕捉杂质特征的局部和全局相关性。通过 ALL-MLP 模块处理后，该模块生成包含杂质相关信息的嵌入向量，从而为下游任务提供更准确的特征描述。

3 结语

这种创新的级联注意力分割模型，通过融合多分支自注意力机制、混合型 Transformer 编码器和轻量级解码器等先进技术，实现了在复杂视觉场景下的高效语义分割。模型不仅能够在动态环境中精确捕捉多尺度的图像特征，还通过引入杂质过滤器技术，通过分割模型和杂质过滤器交替优化，显著提升了对不同环境（低光、雨天、雾霾等）条件下的处理能力。未来的研究可以进一步探讨如何结合更加高效的噪声标签处理与自监督学习方法，以实现更加稳健和泛化的分割效果。

参考文献

[1] Xie E, Wang W, Yu Z, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers[J]. Advances in neural information processing systems, 2021, 34: 12077-12090.

[2] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]// Proceedings of the European conference on computer vision (ECCV). 2018: 801-818.

[3] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Jun 27-30, 2016. Washington: IEEE Computer Society, 2016: 770-778.