李忠旭. 基于 Transformer-CNN 的图像实时语义分割方法[J]. 智能计算机与应用,2025,15(7):155-161. DOI:10.20169/j.issn.2095-2163.24101103

基于 Transformer-CNN 的图像实时语义分割方法

李忠旭
(中国烟草总公司 北京市公司, 北京 100021)

摘 要: 为提高图像处理任务的完成质量和实时性,研究设计了一种结合 Transformer 技术的图像语义分割技术。过程中利用深度特征的空间和通道权重提升图像分割精度,通过子区域划分确定特征图的关键部分;加入通道敏感性权重进行信息优化,将不同尺度的特征图融合为一个统一的特征表示。在 Visual Object Classes Challenge 2012 数据集和 Pascal-Person-Part 数据集上,实验结果表明,该方法在高对比度图像中训练迭代次数达到 29 次时,损失值下降到接近 0 的位置并基本保持稳定;在公开视频数据集上的实验结果表明,研究方法在 1 080 p 分辨率和 30 FPS 的视频数据上,实现了每帧约 9 ms 的处理速度,整体处理延迟控制在 0.3 s 以内。实验表明,研究方法具有更强的图像特征实时提取性能,能够更准确合理地进行图像语义分割。

关键词: Transformer; 图像处理; 语义分割; 卷积神经网络; 特征提取
中图分类号: TP751 **文献标志码**: A **文章编号**: 2095-2163(2025)07-0155-07

Real time semantic segmentation method for images based on Transformer-CNN

LI Zhongxu
(China National Tobacco Corporation Beijing Company, Beijing 100021, China)

Abstract: To improve the quality and real-time performance of image processing tasks, a semantic segmentation technique combining Transformer technology has been studied and designed. During the process, the spatial and channel weights of deep features are utilized to improve image segmentation accuracy. The key parts of the feature map are determined through sub region partitioning, and channel sensitivity weights are added for information optimization. Different scales of feature maps are fused into a unified feature representation. On the Visual Object Classes Challenge 2012 dataset and Pascal Person Part dataset, the experimental results show that the research method reduces the loss value to near 0 and remains stable when the training iterations reach 29 in high contrast images; The experimental results on public video datasets show that the research method achieved a processing speed of approximately 9 milliseconds per frame on 1 080 p resolution and 30 FPS video data, with an overall processing delay controlled within 0.3 seconds. This indicates that the research method has stronger real-time image feature extraction performance and can perform more accurate and reasonable image semantic segmentation.

Key words: transformer; image processing; semantic segmentation; Convolutional Neural Network; feature extraction

0 引 言

随着人工智能技术的不断进步,图像语义分割开始占据越来越重要的技术地位[1-2],图像语义分割技术为多种计算机智能分析任务提供了强大的技术支持[3-4]。早期的图像分割方法依赖于图像的纹理、颜色、边缘等低级视觉特征,这些方法在处理复杂场景时往往效果不佳。高阶条件随机场技术在图像语义分割中被广泛应用。该技术通过在像素级别建立全局信息约束,将相邻像素之间的关系进行建模,从而提高分割精度。跨尺度融合技术则是另一种提升图像分割精度的方法。该技术通过提取图像的多尺度特征,并将这些特征在不同尺度上进行融合,从而实现对图像局部细节和全局上下文信息的综合建模。随着深度学习技术的兴起,卷积神经网络(Convolutional Neural Network, CNN)因其强大的特征提取能力而成为图像分割领域的主流技术[5]。但 CNN 在处理长距离依赖和全局上下文信息方面

作者简介: 李忠旭(1975—),男,硕士,高级工程师,主要研究方向:信息化项目建设,人工智能,图像识别等。Email:w184922@163.com。
收稿日期: 2024-10-11

存在局限,限制了其在复杂场景下的分割性能。因此,许多学者针对 CNN 的图像分割性能进行了优化。如:王维等[6]将 CNN 进行了多目标自适应优化,实验结果表明,所提方法具有良好的图像语义分割效果。Tayal A[7]使用具有多层的 CNN 进行医学图像分割任务,证明研究方法具有良好的图像语义特征提取准确性。Transformer 模块具有全局上下文建模能力,能够通过自注意力机制有效地捕捉图像中的长距离依赖关系[8]。在这样的背景下,研究尝试创新性地将 Transformer 技术和 CNN 进行结合,构建出 Transformer-CNN 技术,并利用多维聚合向量和信息聚合模块加强对图像的信息提取能力,以期为图像处理提供一定的技术参考。

1 结合 CNN 和 Transformer 模型的图像实时语义分割技术

1.1 基于 CNN 的图像语义分割技术

在多个领域的图像处理任务中,图像语义分割都是重要的处理环节之一[9-10]。图像语义分割作为计算机视觉的一个重要技术内容,涉及将图像中的每个像素分配到对应的语义类别中,实现对图像的细粒度理解[11-12]。在进行图像语义分割后更方便进行图像内容理解,进而为后续的图像分析、再处理环节打下基础[13-14]。但在进行图像分割时,涉及的图像内容尺度、图像光照、像素均衡度等问题,导致图像分割涉及大量特征需要提取[15-16]。CNN 在图像分割任务中能够捕获图像的多尺度信息[17-18],本研究选用 CNN 搭建图像的语义分割技术基础。对图像进行多区域特征聚合时,定义若干个区域,并对每个区域求聚合特征向量,公式如下:

$$f_{\Re}=[f_{\Re,1},\cdots,f_{\Re,\ell},\cdots,f_{\Re,\Im}]^{\mathrm{T}} \tag{1}$$

式中:$f_{\Re}$ 代表区域聚合特征向量,T 代表转置矩阵,$f_{\Re,\ell}$ 代表第 ℓ 个特征图处于 $\Re$ 区域时的最大值。求出所有区域的特征向量后,将归一化特征进行求和,形成多维聚合特征向量:

$$\boldsymbol{F}_{R\text{-}MAC}=[\sum_{\Re\subseteq\Omega}f_{\Re,1},\cdots,\sum_{\Re\subseteq\Omega}f_{\Re,\ell},\cdots,\sum_{\Re\subseteq\Omega}f_{\Re,L}]^{\mathrm{T}} \tag{2}$$

式中:$\boldsymbol{F}_{R\text{-}MAC}$ 代表多维聚合特征向量,L 代表向量维数。通过深度特征的空间权重和通道权重进行多维度特征聚合时,生成特征向量:

$$F_{crow}=[f_1,\cdots,f_\ell,\cdots,f_{\Im}],$$
$$f_\ell=\sum_{y=1}^{H}\sum_{x=1}^{W}\alpha_{xy}\beta_\ell S_\ell(x,y) \tag{3}$$

式中:F_{crow} 代表深度特征聚合向量,β_ℓ 代表通道权重,α_{xy} 代表空间权重。研究建立的特征聚合过程如图 1 所示。

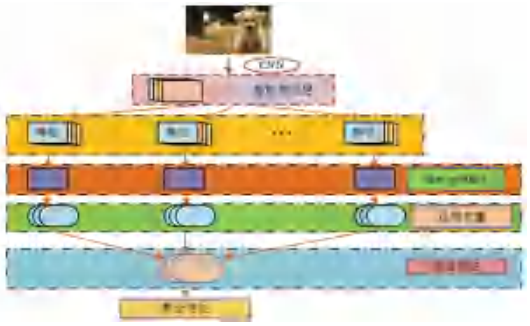


图 1 特征聚合过程

Fig. 1 Feature aggregation process

由图 1 可见,在基于 CNN 的图像语义分割的特征聚合中,通过子区域划分确定特征图的关键部分,并为这些区域的通道分配敏感性权重。通过 sum-pooling 方法聚合这些特征向量,并进行最后一次归一化,形成用于分割任务的聚合特征向量,从而提高分割的准确性和鲁棒性。为了更有效地提取出特征图中的有效信息,研究加入通道敏感性权重进行优化。通道敏感性:

$$\mathbb{N}=1-\gamma_\ell \tag{4}$$

式中:$\mathbb{N}$ 代表通道敏感性,γ_ℓ 代表每个通道上所有非 0 响应值的强度幅值与正响应值比例的和。

在进行图像分割时,图像中不同内容的特征显著性也存在区别,研究对区域的显著性权重进行定义,公式如下:

$$\beta_r=\frac{\sqrt{\sum_{i\in R_r}\sum_{j\in R_r}S'(i,j)^2}}{\sum\sum_{ij}S'\times A_r} \tag{5}$$

式中:β_r 代表区域的显著性权重,S' 代表所有特征图在每个位置上的值之和,A_r 代表目标区域占整个特征图的大小比例,(i,j) 代表所处的特征图位置。某个特征区域占整个特征图的大小比例计算如下:

$$A_r=\frac{w_r\times h_r}{W\times H} \tag{6}$$

式中:$W\times H$ 代表特征图尺寸,$w_r\times h_r$ 代表特征区域尺寸。对区域特征向量进行生成,公式如下:

$$\hat{f}_r^{\ell}=(\sum_{i\in R_r}\sum_{j\in R_r}\lambda_\ell S_\ell(i,j))^{\alpha} \tag{7}$$

式中:$\hat{f}_r^{\ell}$ 代表区域特征向量在第 ℓ 个特征图上所体现出的数值,S_ℓ 代表特征图在所处位置上的数值和,λ_ℓ 代表通道敏感性权重。其中,通道敏感性权重计算如下:

$$\lambda_{\ell} = \log\left(\frac{L\mathbb{C} + \sum_{\ell=1}^{L}\gamma_{\ell}}{\mathbb{C} + \gamma_{\ell}}\right) \tag{8}$$

式中：$\mathbb{C}$ 是为了确保计算过程中数值稳定的小常数。通过特征语义相似性对图像分割的区域范围进行划分，相似度计算如下：

$$s_{i,j} = \sum_{k=1}^{K}(\mu_{\xi_{P_i}}(P_j^k) + \mu_{\xi_{P_j}}(P_i^k)) \tag{9}$$

式中：$s_{i,j}$ 代表相似度，$\mu_{\xi_{P_i}}$ 代表词条的模糊语义，P_j^k 代表词条样本近邻中的第 k 个最近邻样本。由语义相似度输出图像分割的区域范围划分结果，完成图像语义分割中的图像特征提取。

1.2　融合 Transformer 和信息聚合模块的实时分割技术

CNN 标准卷积操作的感受野有限，对于长距离依赖关系的空间信息捕捉能力有限，Transformer 通过自注意力机制能够捕捉长距离的依赖关系。研究将 Transformer 模块加入图像语义分割技术中，构建 Transformer 模块编码的结构如图 2 所示。

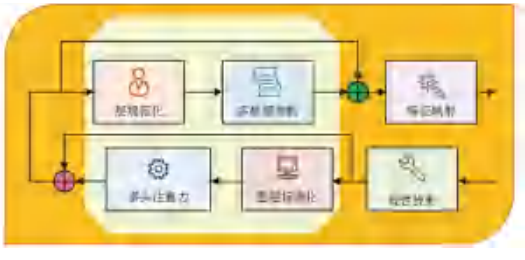


图 2　Transformer 编码框架

Fig. 2　Transformer coding framework

由图 2 可见，在进行图像实时语义分割时，将 CNN 所提取到的图像特征输入 Transformer 模块。Transformer 编码模块首先通过一个线性层对输入序列进行映射，随后映射输出传递至高效变压器模块，高效变压器模块内部集成了一个高效的多头注意力机制。在运行时，使用投射矩阵转化全局特征为 3 组向量：

$$\begin{cases} Query_i = \boldsymbol{Feat}_{\text{global}} \times \boldsymbol{W}_i^{Q} \\ \boldsymbol{Key}_i = \boldsymbol{Feat}_{\text{global}} \times \boldsymbol{W}_i^{K} \\ Value_i = \boldsymbol{Feat}_{\text{global}} \times \boldsymbol{W}_i^{V} \end{cases} \tag{10}$$

式中：$\boldsymbol{Key}$ 代表键向量，$\boldsymbol{W}$ 代表投射矩阵，$\boldsymbol{Feat}_{\text{global}}$ 代表全局矩阵。使用 *Sigmoid* 函数计算注意力值，并激活神经元，公式如下：

$$\text{Attention}(Key_i, Query_i, Value_i) = Sigmoid\left(\frac{Query_i Key_i^{\mathrm{T}}}{\sqrt{d_k}}\right) \times Value_i \tag{11}$$

式中：$\sqrt{d_k}$ 代表注意力收缩参数，$Sigmoid(\cdot)$ 代表 *Sigmoid* 函数。拼接时序注意力特征后，导入投射矩阵计算最终图像特征提取结果，公式如下：

$$\boldsymbol{Feat}_{\text{att}} = \boldsymbol{W}^{o} \times \text{Concat}(\text{Attention}_1, \text{Attention}_2, \cdots, \text{Attention}_n) \tag{12}$$

式中：$\boldsymbol{Feat}_{\text{att}}$ 代表最终图像特征提取结果，$\boldsymbol{W}^{o}$ 代表投射矩阵，Concat(·)代表拼接操作。注意力模块中的多头注意力模块参数约束如下：

$$d_k = d_v = d_q = \frac{d_{\text{model}}}{n_{\text{head}}} \tag{13}$$

式中：d_k 代表键向量维度，d_v 代表值向量维度，d_q 代表查询向量维度，d_{model} 代表全局特征向量维度。

研究在 Transformer 模块之后插入一个信息聚合模块进行特征融合。使用 Add 运算替换原信息聚合模块中的 Concat 运算步骤。在解码器中加入注意力机制，以保持图像特征的明确性，其上采样特征如下：

$$F_{\text{up}} = \text{Upsample}(F_{\text{high}}) \tag{14}$$

式中：F_{up} 代表上采样后的特征，F_{high} 代表信息聚合模块输出的特征图像。特征权重计算如下：

$$\beta = \text{Channel/Spatialattention}(F_{\text{high}} + F_{\text{low}}) \tag{15}$$

式中：β 代表特征权重，F_{low} 代表 CNN 编码器的输出特征图像。对图像特征进行权重求和，以强化图像特征的坐标信息，输出特征图像，公式如下：

$$F_{\text{out}} = \text{Coodinateattention}(F_{\text{up}} \cdot \beta + F_{\text{low}} \cdot (1-\beta)) \tag{16}$$

式中：F_{out} 代表输出的特征图像。研究构建基于 Transformer-CNN 的图像实时语义分割技术如图 3 所示。

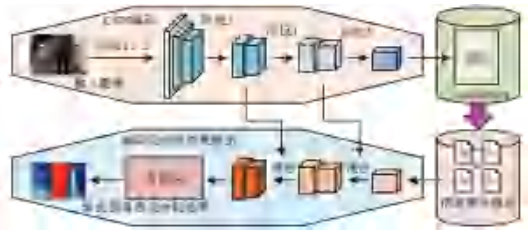


图 3　基于 Transformer-CNN 的图像实时语义分割技术

Fig. 3　Transformer-CNN based real-time semantic segmentation of images

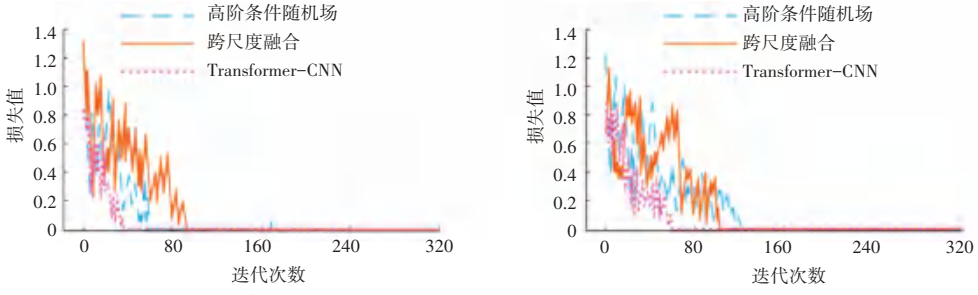
由图 3 可见，研究构建基于 Transformer-CNN 的图像实时语义分割技术在运行编码阶段，利用 CNN 结构提取图像的局部特征，然后通过 Transformer 模块进一步捕获全局上下文信息和像素

间的长距离依赖关系;再通过一个特别设计的信息聚合模块,对编码器输出的多尺度特征进行整合;最后,在解码阶段,采用基于注意力机制的特征融合模块对不同层级的特征进行有效融合,增强特征的表达力并输出最终的图像语义分割结果。

2 图像实时语义分割法有效性分析

2.1 Transformer-CNN 模型性能测试

在对基于 Transformer-CNN 模型的图像实时语义分割方法进行性能测试时,使用 Visual Object Classes Challenge 2012 数据集和 Pascal-Person-Part 数据集作为测试数据集。将研究方法与目前主流的高阶条件随机场技术和跨尺度融合技术进行对比[19-20]。将两个数据集混合后,划分高对比度组和低对比度组,分析不同方法的训练损失情况,如图 4 所示。

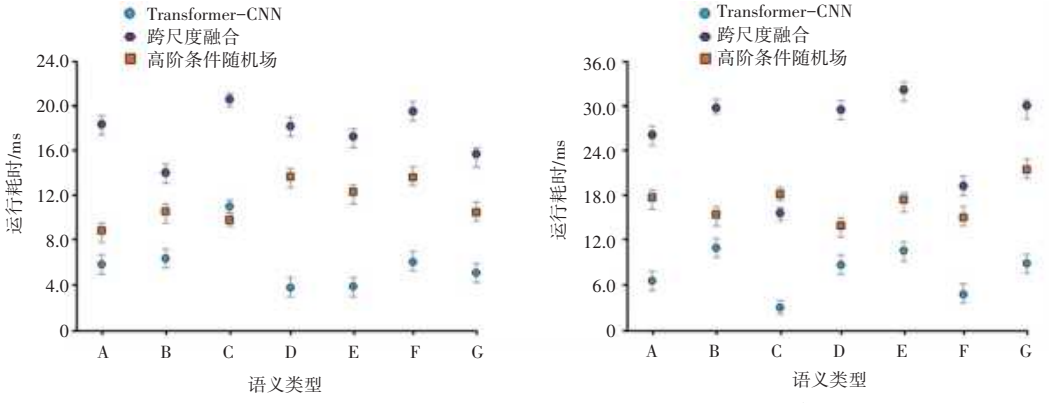


(a) 高度对比 (b) 低度对比

图 4 训练损失情况分析

Fig. 4 Analysis of training loss scenarios

由图 4(a)可见,跨尺度融合技术在迭代次数为 63 次时,损失值下降到接近 0 的位置,并基本保持稳定。Transformer-CNN 在高对比度图像中训练时,初始的损失值为 0.82;在迭代次数达到 29 次时损失值下降到接近 0 的位置,并基本保持稳定。图 4(b)显示,Transformer-CNN 在低对比度图像中训练时,初始的损失值为 0.90;在迭代次数达到 63 次时损失值下降到接近 0 的位置并基本保持稳定。说明研究方法在训练时的迭代效率更高。设置 7 种语义类型,对研究方法在特征提取和图像语义分割两个阶段的处理耗时结果进行分析,如图 5 所示。



(a) 特征提取 (b) 图像语义分割

图 5 任务处理耗时分析

Fig. 5 Analysis of task processing elapsed time

图 5(a)显示,在进行特征提取时,高阶条件随机场在 7 种语义类型上的处理耗时都保持在 12 ms 以上。Transformer-CNN 在 7 种语义类型上的处理耗时处在 2~12 ms 区间。由图 5(b)可见,在进行图像语义分割时,高阶条件随机场在 7 种语义类型上的处理耗时都保持在 12 ms 以上;跨尺度融合技术在 7 种语义类型上的处理耗时处在 12~20 ms 区间内;Transformer-CNN 在 7 种语义类型上的处理耗时处在 2~12 ms 区间内,在 7 种类型上皆低于其他方法。说明本文方法具有更快的模型运行效率以及在不同阶段的处理速度,具有更强的实时性。

2.2　图像实时语义分割方法应用效果分析

为了对研究方法在实际应用中效果进行分析，研究收集 100 幅真实图片和 100 幅虚拟图片，对研究方法进行实际应用，图片的分辨率都保持为 1 080 p，应用时从收集的图片中进行随机抽取。对不同方法的图像特征提取准确率进行分析，结果如图 6 所示。

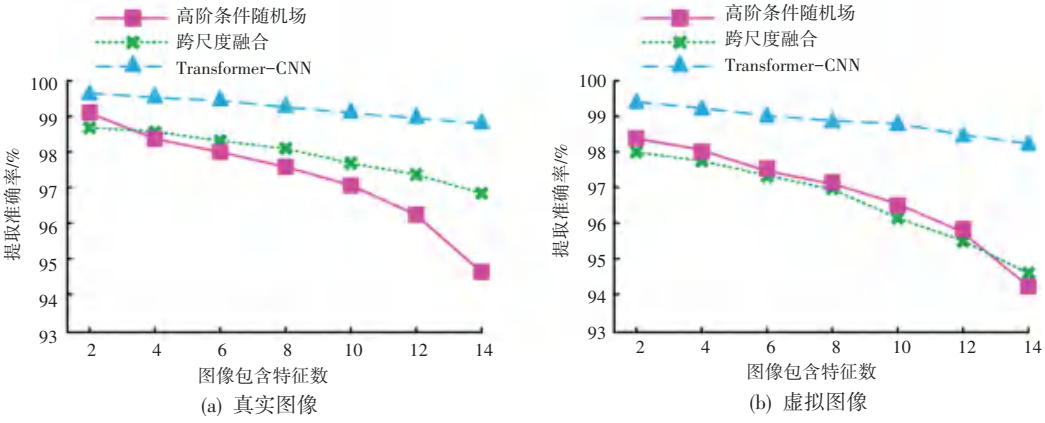


图 6　图像特征提取准确率分析

Fig. 6　Image feature extraction accuracy analysis

由图 6(a)真实图像中可见，Transformer-CNN 在图像包含 2 个特征时，特征提取准确率为 99.7%；在图像包含 14 个特征时，特征提取准确率下降到 98.8%。在图 6(b)虚拟图像中可见，Transformer-CNN 在图像包含 2 个特征时的特征提取准确率为 99.4%；在图像包含 14 个特征时的特征提取准确率下降到 98.2%。说明研究方法能够更准确地提取到图像中的特征信息。随机抽取一幅图像进行图像语义分割，其结果如图 7 所示。

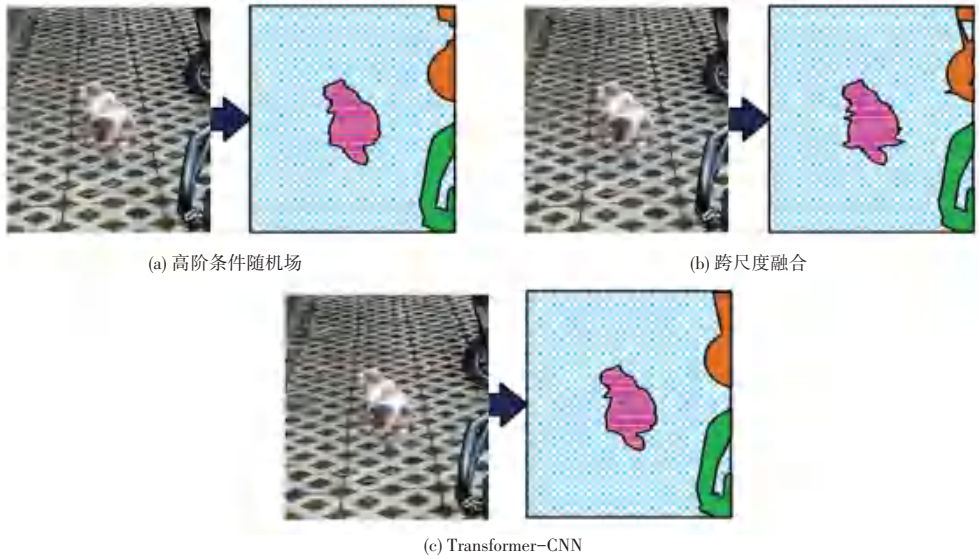
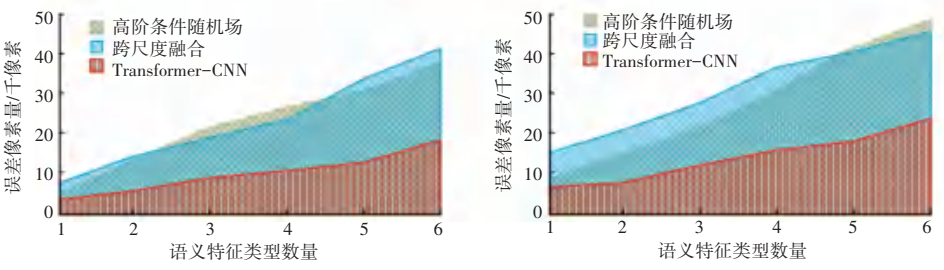


图 7　图像语义分割效果分析

Fig. 7　Analysis of image semantic segmentation effect

图 7(a)显示，高阶条件随机场对于不同语义的边缘以及语义混杂部分存在较为明显的划分误差；由图 7(b)可见，跨尺度融合的方法能够划分出主要的具有不同语义的内容，但在不同语义之间的边缘位置存在较为明显的划分误差；由图 7(c)可见，Transformer-CNN 在图像包含不同类型语义内容时，保持了良好的边缘准确度和对不同语义理解的准确性，不存在明显的影响划分效果的误差。说明研究方法具有更好的图像语义理解和分割能力，能够有效保证关键信息处理的质量。对研究方法在进行图像语义分割时出现的误差像素量进行分析，其结果如图 8 所示。



（a）真实图像 （b）虚拟图像

图 8 图像语义分割时出现的误差像素量

Fig. 8 Amount of error pixels occurring during semantic segmentation of images

图 8(a)在处理真实图像时显示,Transformer-CNN 在图像语义特征类型数量为 1 时的误差像素量为 4 千像素;在语义特征类型数量增加到 6 时的误差像素量增加到 18 千像素;由图 8(b)可见,在处理虚拟图像时,跨尺度融合在图像语义特征类型数量增加到 6 时的误差像素量增加到 45 千像素,而 Transformer-CNN 的表现相对稳定,与在真实图像中的表现没有明显差别,在图像语义特征类型数量增加到 6 时的误差像素量保持在 25 千像素以下。由此说明,本文研究方法能够进行更准确的图像语义及内容分割,且在真实图像和虚拟图像中都能良好保持性能。为验证研究方法在视频数据中的实时性,研究从公开视频数据集中随机抽取 10 段视频,每段视频长度为 10 s,分辨率为 1 080 p,帧率为 30 FPS。对每段视频的每帧图像进行逐帧语义分割,并记录每帧的处理时间,以评估模型在视频处理时的实时性表现,结果见表 1。

表 1 视频处理实时性测试

Table 1 Video processing real-time tests

视频序列编号	平均帧处理时间/ms	视频处理总延迟/s	分割准确率/%
视频 1	9.5	0.29	98.2
视频 2	8.7	0.26	97.9
视频 3	9.0	0.27	98.4
视频 4	8.8	0.27	98.1
视频 5	9.3	0.28	97.7

由表 1 可见,研究方法在 1 080 p 分辨率和 30 FPS 的视频数据上实现了每帧约 9 ms 的处理速度,整体处理延迟控制在 0.3 s 以内,相较于静态图像处理,视频处理更能体现模型的实时性和连续性。在视频分割过程中,研究方法保持了较高的分割准确率,达到了 98%以上,证明了其在连续帧中有效捕捉并处理图像特征的能力。

3 结束语

研究提出了一种结合 CNN 和 Transformer 模型的图像实时语义分割方法,以改善图像分割和处理的效果。过程中使用 CNN 搭建图像的语义分割技术基础,定义若干个区域,并对每个区域求聚合特征向量,对区域的显著性权重进行定义,由语义相似度输出图像分割的区域范围划分结果,将多层感知机用于进一步提取和学习数据的非线性特征,采用基于注意力机制的特征融合模块对不同层级的特征进行融合。实验结果表明,在任务处理耗时分析中,研究方法在进行图像语义分割时,在 7 种语义类型上的处理耗时在 2~12 ms 区间内;进行图像语义分割效果分析时,研究方法保持了良好的边缘准确度和对于不同语义理解的准确性;在分割时出现的误差像素量分析中,研究方法在虚拟图像中语义特征类型数量增加到 6 时的误差像素量保持在 25 千像素以下。说明研究方法具有更好的图像语义分割准确性,且具有良好的实时处理性能。

由于研究仅对静态图像进行了技术设计,对于目前逐渐增加的动态图像尚不具备足够的应用能力。后续将专门针对动态图像等特殊类型图像进行分析,并引入其他技术结构进行优化,以扩大研究方法的适用范围。

参考文献

[1] 王海鹏, 丁卫平, 黄嘉爽, 等. FTransCNN:基于模糊融合的 Transformer-CNN 不确定性医学图像分割模型[J]. 小型微型计算机系统, 2024, 45(6):1426-1435.

[2] 丁才富, 杨晨, 纪秋浪, 等. MCA-Net:多尺度综合注意力 CNN 在医学图像分割中的应用[J]. 微电子学与计算机, 2022, 39(3):71-77.

[3] 蔡超丽, 李纯纯, 黄琳, 等. ED-NAS:基于神经网络架构搜索的陶瓷晶粒 SEM 图像分割方法[J]. 电子学报, 2022, 50(2):461-469.

[4] 李赵春, 周永照, 冯卫奔, 等. 基于 Transformer 模型的手势脑

电信号分类识别[J]. 科学技术与工程, 2023, 23(5):2044-2050.
[5] 刘肇隆, 范馨月. 基于全尺度跳跃连接的 TransUNet 医学图像分割网络[J]. 国外电子测量技术, 2023, 42(11):42-48.
[6] 王维, 王显鹏, 宋相满. 基于自适应多目标进化 CNN 的图像分割方法[J]. 控制与决策, 2024, 39(4):1185-1193.
[7] TAYAL A, GUPTA J, SOLANKI A, et al. DL-CNN-based approach with image processing techniques for diagnosis of retinal diseases[J]. Multimedia Systems, 2022, 28(4): 1417-1438.
[8] 李轩, 杨舟, 陶新宇, 等. 基于 Mask R-CNN 结合边缘分割的颗粒物图像检测[J]. 应用光学, 2023, 44(1):93-103.
[9] 吴惠思, 陈文杰, 黄晓婷, 等. 基于视觉 Transformer 内在归纳优化的齐白石虾画真假鉴定[J]. 计算机辅助设计与图形学学报, 2023, 35(11):1654-1663.
[10]李顺平, 彭成. 基于高效通道注意力机制和图像分割的轻量级表情识别算法[J]. 现代电子技术, 2022, 45(20):149-156.
[11]纪建兵, 陈纾, 杨媛媛. 调整窗宽/窗位对卷积神经网络模型自动筛选胰腺肿瘤 CT 图像性能的影响[J]. 中国医学影像技术, 2023, 39(2):270-275.
[12]濮子俊, 张寿明. 基于特征融合与 Transformer 模型的声音事件定位与检测算法研究[J]. 计算机工程与科学, 2023, 45(6):1097-1105.
[13]李雷垚, 张惊雷, 文彪, 等. 基于多元空洞特征金字塔的电气设备图像实例分割方法[J]. 天津理工大学学报, 2023, 39(6):14-19.
[14]辛紫麒, 李忠伟, 王雷全, 等. 基于光谱-空间联合 Transformer 模型的黄河三角洲湿地高光谱影像分类[J]. 海洋科学, 2023, 47(5):90-101.
[15]王静, 李沛橦, 赵容锋, 等. 融合卷积注意力和 Transformer 架构的行人重识别方法[J]. 北京航空航天大学学报, 2022, 50(2):466-476.
[16]YAO Xujing, WANG Xinyue, WANG Shuihua, et al. A comprehensive survey on convolutional neural network in medical image analysis[J]. Multimedia Tools and Applications, 2022, 81(29): 41361-41405.
[17]张小勇, 张洪, 高清源, 等. 室内动态场景下基于稀疏光流与实例分割的视觉 SLAM 算法[J]. 东华大学学报(自然科学版), 2023, 49(6):111-119.
[18]李洋, 朱春山, 张建亮, 等. 基于改进 Transformer 的变电站复杂场景下电力设备分割[J]. 太原理工大学学报, 2024, 55(1):57-65.
[19]张帆, 闫敏超, 倪军, 等. 高阶条件随机场引导的多分支极化 SAR 图像分类[J]. 中国图象图形学报, 2023, 28(10):3267-3280.
[20]李记恒, 褚霄杨, 王涛, 等. 基于跨尺度特征融合的泵站安全帽检测方法[J]. 测控技术, 2023, 42(7):16-21.