



基于双流融合网络的输送带跑偏检测方法

杨志方 张立亚 郝博南 刘渊 赵青

引用本文：

杨志方, 张立亚, 郝博南, 等. 基于双流融合网络的输送带跑偏检测方法[J]. 煤炭科学技术, 2023, 51(S2): 259–267.
YANG Zhifang, ZHANG Liya, HAO Bonan. Conveyor belt deviation detection method based on dual flow network[J]. Coal Science and Technology, 2023, 51(S2): 259–267.

在线阅读 View online: <https://doi.org/10.13199/j.cnki.est.2023-0215>

您可能感兴趣的其他文章

Articles you may be interested in

基于轻量化的输送带块煤实时监测方法

A lightweight-based method for real-time monitoring of lump coal on conveyor belts
煤炭科学技术. 2023, 51(S2): 285–293 <https://doi.org/10.12438/cst.2023-1217>

基于激光测距矿用带式输送机多参数检测方法研究

Research on multi-parameters detection method of mine belt conveyor based on laser ranging
煤炭科学技术. 2020, 48(8): 131–138 <http://www.mtkxjs.com.cn/article/id/77cc2e01-881c-4fff-990b-ac99b657cf4c>

基于特征增强与Transformer的煤矿输送带异物检测

Coal mine conveyor belt foreign object detection based on feature enhancement and Transformer
煤炭科学技术. 2024, 52(7): 199–208 <https://doi.org/10.12438/cst.2023-1336>

基于深度神经网络的煤矿井下低光照图像增强算法

Deep neural network-based image enhancement algorithm for low-illumination images underground coal mines
煤炭科学技术. 2023, 51(9): 231–241 <https://doi.org/10.12438/cst.2022-1626>

基于深度神经网络的煤岩显微图像划痕检测与修复

Scratch detection and restoration of coal photomicrograph via deep neural network
煤炭科学技术. 2023, 51(S2): 275–284 <https://doi.org/10.13199/j.cnki.est.2023-0058>

基于Adaboost的改进BP神经网络地表沉陷预测

Prediction of surface subsidence with improved BP neural network based on Adaboost
煤炭科学技术. 2019(2) <http://www.mtkxjs.com.cn/article/id/5e547320-c2e8-4921-9e49-77876f8c6ea4>



关注微信公众号，获得更多资讯信息



移动扫码阅读

杨志方, 张立亚, 郝博南, 等. 基于双流融合网络的输送带跑偏检测方法[J]. 煤炭科学技术, 2023, 51(S2): 259–267.

YANG Zhifang, ZHANG Liya, HAO Bonan, et al. Conveyor belt deviation detection method based on dual flow network[J]. Coal Science and Technology, 2023, 51(S2): 259–267.

基于双流融合网络的输送带跑偏检测方法

杨志方^{1,2,3}, 张立亚^{1,2,3}, 郝博南^{1,2,3}, 刘渊^{1,2,3}, 赵青^{1,2,3}

(1. 煤炭科学技术研究院有限公司, 北京 100013; 2. 煤矿应急避险技术装备工程研究中心, 北京 100013;

3. 北京市煤矿安全工程技术研究中心, 北京 100013)

摘要: 传统输送带跑偏检测方法中, 接触式检测技术成本高, 非接触式检测技术则精度低。随着人工智能技术的发展, 虽然基于卷积神经网络的方法可以有效提高检测精度, 但受限于卷积操作本身局部运算特性的限制, 仍存在对长距离、全局信息感知不足等问题, 很难再提升在输送带边缘检测上的精度。为解决上述问题, ①通过将传统卷积神经网络的卷积对局部特征的提取能力与 Transformer 结构对全局、长距离信息感知能力相结合, 提出了一种全局与局部信息相互融合的双流输送带边缘检测网络模型(Dual-Flow Transformer Network, DFTNet), 能够较好地提高输送带边缘检测精度并抑制输送带图像噪声和背景的干扰; ②通过设计卷积神经网络(Convolutional Neural Network, CNN)和转换器 Transformer 特征融合模块, 形成双流编码器-解码器结构, 利用结构上的巧妙设计, 可以更好地融合全局上下文信息, 避免了 Transformer 结构在大规模数据集上预训练, 可以灵活调节网络结构; ③通过从实际工业场景中所采集到多场景的运输机输送带图片, 构建了包含 5 种不同场景下多角度、不同位置的输送带输送带数据集。研究结果表明, 双流融合网络 DFTNet 综合性能最佳, 均交并比 mIoU 达 91.08%, 准确率 ACC 达 99.48%, 平均精确率 mPrecision 达 91.88%, 平均召回率 mRecall 达 96.22%, 相比纯卷积神经网络 HRNet 分别提升了 25.36%、0.29%、17.70% 与 29.46%, 相比全卷积神经网络(Fully Convolutional Networks, FCN)分别提升了 29.5%、0.32%、24.77% 与 34.13%, 在参数量、计算速度上均有较大提升。同时, 处理图像帧率达 53.07 fps, 满足工业中实时性的要求, 具有较大实用价值。

关键词: 输送带跑偏; 边缘检测; 神经网络; 编码器-解码器; 图像分割

中图分类号: TP181 文献标志码: A 文章编号: 0253-2336(2023)S2-0259-09

Conveyor belt deviation detection method based on dual flow network

YANG Zhifang^{1,2,3}, ZHANG Liya^{1,2,3}, HAO Bonan^{1,2,3}, LIU Yuan^{1,2,3}, ZHAO Qing^{1,2,3}

(1. China Coal Research Institute, Beijing 100013, China; 2. Engineering Research Center for Technology Equipment of Emergency Refuge in Coal Mine, Beijing 100013, China; 3. Beijing Mine Safety Engineering Technology Research Center, Beijing 100013, China)

Abstract: Among the traditional belt edge detection methods, the contact detection technology has high cost and the non-contact detection technology has low precision. With the development of artificial intelligence technology, although the method based on convolutional neural network can effectively improve the detection accuracy, but limited by the local operation characteristics of the convolutional operation itself, there are still problems such as insufficient perception of long-distance and global information, it is difficult to improve the accuracy of the belt edge detection. In order to solve the above problems, ① by combining the traditional convolutional neural network's ability to extract local features and the Transformer structure's ability to perceive global and long-distance information, a dual-flow transformer network (DFTNet) which integrates global and local information is proposed. The edge detection network model can better im-

收稿日期: 2023-03-19 责任编辑: 周子博 DOI: 10.13199/j.cnki.cst.2023-0215

基金项目: 天地科技股份有限公司科技创新创业资金专项资助项目(2023-TD-ZD005-005, 2022-TD-ZD001); 煤炭科学技术研究院有限公司新产品新工艺开发资助项目(2023CG-ZB-12)

作者简介: 杨志方(1994—), 男, 河南洛阳人, 研究实习员, 硕士。E-mail: hnezzsf@163.com

prove the belt edge detection accuracy and suppress the interference of belt image noise and background; ② By designing the CNN and Transformer feature fusion modules, a dual-flow encoder-decoder structure is formed. The clever design can better integrate the global context information, avoid the pre-training of the Transformer structure on large-scale data sets and be flexibly adjusted; ③ By Through the multi-scene conveyor belt pictures collected from the actual industrial scene, a belt conveyor belt dataset containing five different scenes, various angles and different positions is constructed. Through experimental verification, the DFTNet proposed in this paper has the best comprehensive performance with mIoU 91.08%, ACC 99.48%, mPrecision 91.88% and mRecall 96.22%. which are 25.36%, 0.29%, 17.70% and 29.46% respectively compared to the pure convolutional neural network HRNet, and 29.5%, 0.32%, 24.77% and 34.13% respectively compared to FCN. At the same time, the frame rate of DFTNet processing images reaches 53.07 fps, which meets the real-time requirements in the industry and has great practical value.

Key words: belt deviation; edge detection; neural network; encoder-decoder; image segmentation

0 引言

输送带运输机跑偏事故会对企业造成巨大经济损失,严重时甚至会造成人员伤亡。及时准确地检测出传送带跑偏情况并采取有效措施,可以有效避免事故发生,对安全生产意义重大^[1]。现有的传送带跑偏检测主要依靠人工检查及传统接触式检测方法,存在成本高、不稳定、易错检或漏检等问题,对于长距离、复杂环境下的工业输送带场景往往并不适用^[2]。随着工业中摄像头的普及,工业输送带图像数据量日益增大,兴起了对运输机输送带图像数据相关应用技术的研究,其中包括利用机器视觉技术实现对传送带跑偏检测,其优点是可以对传送带边缘进行连续的检测^[3]。

基于机器视觉检测技术的传统跑偏检测算法主要根据为当输送带跑偏时,图像灰度值会在输送带边缘附近发生变化,并采用基于微分的方法进行边缘的提取^[4-5]。此种方法虽然有检测速度快、实现方便等优点,但是在合适的梯度阈值选取方面比较困难,结果受到阈值选取的影响较大。为了进一步提高基于机器视觉技术的检测精度,随着大数据和人工智能技术的发展,基于深度学习的检测技术开始应用于传送带跑偏检测方面。LIU 等^[6] 分别将全卷积神经网络(Fully Convolutional Networks, FCN^[7]),深度实验室网络模型 DeepLab^[8],整体嵌套边缘检测网络(Holistically-Nested Edge Detection, HED^[9])应用于传送带跑偏检测任务。POMA 等^[10] 在 HED 网络的基础上压缩了模型结构并简化了最后的网络输出,但以上网络仍存在以下不足:①预测的边缘效果较粗糙且处理速度较慢;②由于卷积操作本身仅进行局部运算,很难构建像素间长距离的依赖关系;③目前应用于输送带跑偏检测的网络模型的检测速度没有达到 25 帧/s 的实时性要求。

Transformer^[11] 结构不受限于局部运算,能够建模全局上下文信息,在自然语言处理任务上有着出色的表现。DOSOVITSKIY 等^[12] 提出了视觉 Transformer(Vision Transformer, ViT),该网络首次将 Transformer 结构应用于图像任务中,在大规模的数据上进行训练后,在精度上超越了基于卷积神经网络(Convolutional Neural Network, CNN) 的方法,但 ViT 结构输出特征图的分辨率低且单一,会导致局部信息的丢失; TOUVRON 等^[13] 提出了数据高效的图像变换器(Data-Efficient image Transformer, DEiT),这表明了 Transformer 结构能在中等规模的数据集上进行训练,并取得了较好的效果; LIU 等^[14] 提出了 Swin 变换器(Swin Transformer, SwinT) 结构,该网络在图像分类、目标检测及语义分割中均取得了当时学术界深度学习模型中最好的效果。ViT、DEiT、SwinT 的出现证明了 Transformer 结构在机器视觉任务中的应用潜力,但与纯卷积神经网络相比,Transformer 结构往往参数量大、计算复杂度较高。

结合 Transformer 结构的优势,针对卷积神经网络限于局部运算不能有效获得长距离依赖关系的问题,本文提出了一种全局与局部信息相互融合的双流输送带跑偏检测网络模型(Dual-Flow Transformer Network, DFTNet),通过将 Transformer 和 CNN 网络进行融合,构建了双流解码器-编码器结构,提高了网络模型对全局、长距离信息感知能力,和其他基于 CNN 的方法相比, DFTNet 能在较少的参数下,实现对输送带边缘的实时性准确检测。

1 双流融合网络的搭建

1.1 DFTNet 双流网络总体结构

为了增强网络上下文信息感知能力以及保留丰富的细节信息,针对输送带边缘模糊、背景类别复杂、干扰信息较多等特点,设计了基于 Transformer 和 CNN 双流融合网络的输送带边缘检测方法。通过设

计 CNN 和 Transformer 特征融合模块, 形成双流编码器-解码器结构, 可以更好地融合全局上下文信息, 避免了 Transformer 结构需要在大规模数据集上预训练的缺点, 同时可以灵活调节网络结构, 如图 1 所示。其中, C 为输送带特征图像的通道数, H 和 W 分别为输送带特征图像的高度和宽度。

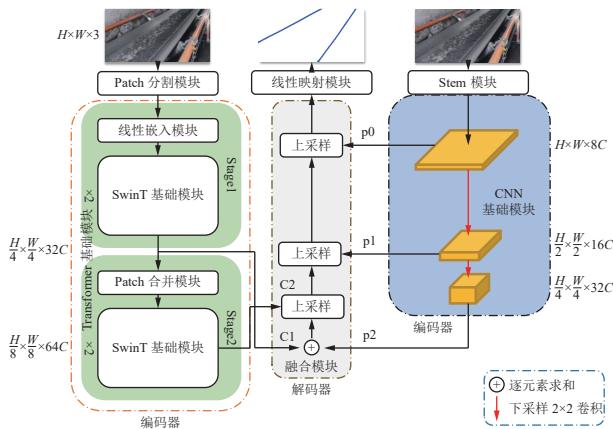


图 1 双流融合网络整体网络结构

Fig.1 Overall network structure diagram of dual-flow fusion network

编码器由 Transformer 基础模块与 CNN 基础模块 2 个分支构成。在 Transformer 基础模块中, 输送带图像首先通过 Patch 分割模块层将输送带特征图像转换为相互不重叠的 4×4 的序列小块, 同时特征通道数变成了 $4\times 4\times 3=48$ 。之后通过线性嵌入模块层, 将特征图像转换成任意维度, 统一表示为 C , 通过第一个 SwinT 基础模块后生成大小为 $H/4\times W/4\times 32C$ 的特征图像, 然后经过一个 Patch 合并模块层和第 2 个 SwinT 基础模块后生成 $H/8\times W/8\times 64C$ 的特征图像。编码器中 CNN 基础模块包含一个 Stem 层以及 2 个下采样层, 其输出特征图的大小为 $H\times W\times 8C$; 经过 Stem 层后图像经过 2 个下采样操作, 且每次下采样后特征通道数加倍, 最后共产生 $H/2\times W/2\times 16C$ 、 $H/4\times W/4\times 32C$ 等 2 个不同的特征图像输入到融合模块解码器当中。

解码器由 3 个上采样组成, 每个上采样由一个可以使图像特征通道数量减半的卷积操作构成。特征图像在经历每个上采样步骤之后, 通过将提取到的上下文特征与编码器的多尺度特征融合在一起, 进而抵消由于 CNN 卷积操作带来的空间信息的损失。经过 3 个上采样操作之后, 图像的分辨率恢复到 $H\times W\times 2$, 最后通过线性映射模块将这些上采样特征进行像素级的分割预测, 最终实现对输送带边缘的准确检测。

1.2 SwinT 基础模块

编码器中的 SwinT 基础模块主要由基于 2 个连续的多头自注意力模块构成, 其主要基于窗口的多头自注意力机制(Window Based Multi-head Self Attention, W-MSA), 有别于卷积的多头注意力机制(Multi-head Self Attention, MSA), 其核心模块如图 2 所示。每个 SwinT 模块均包含 2 个连续的多头自注意力模块, 每个模块均由层归一化(Layer Normalization, LN)、多头自注意力机制、残差连接以及多层次感知机(Multi-Layer Perceptron, MLP)组成, 多层次感知机由两层线性层及 GELU(Gaussian Error Linear Unit)激活函数组成。其中第一个模块应用了基于窗口的多头自注意力机制, 第二个模块在 W-MSA 基础上加入了 Shift-Window 操作形成基于滑动窗口的多头自注意力机制 SW-MSA。

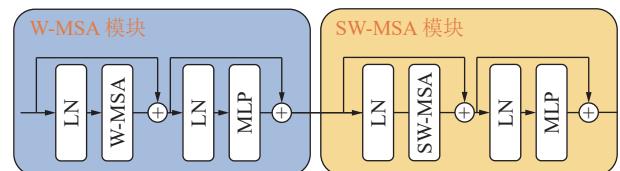


图 2 SwinT 基础模块

Fig.2 SwinT basic block

SwinT 模块中基于窗口的自注意力机制降低了常用自注意力机制的计算复杂度, 整体计算过程如式(1)—式(4), 其中 z^l 表示第 l 个模块经 MLP 和残差连接后输出的特征, \hat{z}^l 表示经 W-MSA 或者 SW-MSA 和残差连接后输出的特征, \hat{z}^{l-1} 、 \hat{z}^{l+1} 分别代表 SwinT 输出模块的输入与输出。

$$\hat{z}^l = \text{W-MSA}[\text{LN}(\hat{z}^{l-1})] + z^{l-1} \quad (1)$$

$$z^l = \text{MLP}[\text{LN}(\hat{z}^l)] + \hat{z}^l \quad (2)$$

$$\hat{z}^{l+1} = \text{SW-MSA}[\text{LN}(\hat{z}^l)] + z^l \quad (3)$$

$$z^{l+1} = \text{MLP}[\text{LN}(\hat{z}^{l+1})] + \hat{z}^{l+1} \quad (4)$$

这里 \hat{z}^l 和 z^l 分别作为 SW-MSA 模块和 MLP 模块的输出。参照文献 [15-16], 在计算多头自注意力机制时, 每个 head 的计算方式如下:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{SoftMax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}} + \mathbf{B}\right)\mathbf{V} \quad (5)$$

其中 $\mathbf{Q}, \mathbf{K}, \mathbf{V} \in \mathbb{R}^{M^2 \times d}$ 分别代表 Query、Key、Value 矩阵, M^2 和 d 分别代表一个窗口里图像块数量和 Query 或 Key 的特征维度数, B 的值来源于偏置矩阵 $\hat{\mathbf{B}} \in \mathbb{R}^{(2M-1) \times (2M+1)}$ 。

1.3 CNN 基础模块

编码器中的 CNN 基础模块主要由一个词干提取 stem 模块与 2 个下采样操作构成, 如图 3 所示。stem 模块由 2 个大小为 3×3 、步幅为 1、填充数为 1 的卷积、批量标准化和修正线性单元(Rectified Lin-

ear Unit, ReLU)激活函数组成, 通过 stem 模块后, 输入特征图像的大小变为 $H \times W \times 8C$, 后经过 2 个由大小为 2×2 、步幅为 2 的卷积操作构成的下采样层后分别输出大小为 $H/2 \times W/2 \times 16C$ 、 $H/4 \times W/4 \times 32C$ 的特征图像。

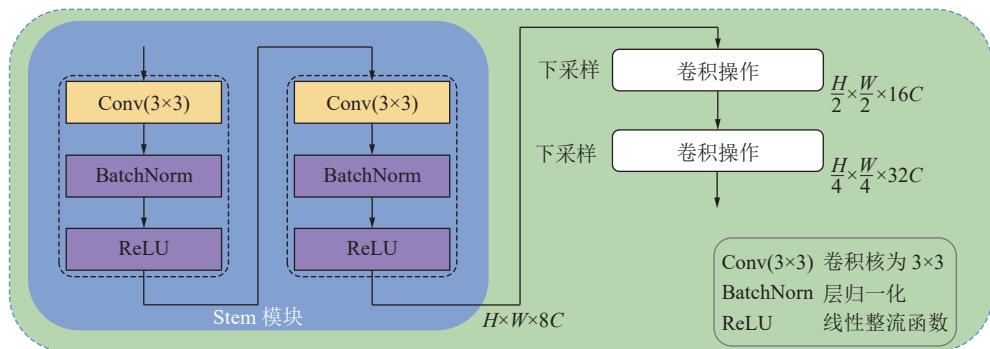


图 3 CNN 基础模块

Fig.3 CNN basic block

1.4 融合模块

解码器中的融合模块主要由双线性差值与 1×1 卷积组成的上采样操作、逐元素求和、线性映射模块组成。 c_1 和 p_2 通过逐元素求和操作进行融合, 输出特征图像大小不变, 之后通过连续 3 次上采样操作分别与 c_2 、 p_1 、 p_0 进行融合, 其中, c_1 、 c_2 、 p_2 、 p_1 、 p_0 代表图 1 中不同阶段输出的结果。特征图像通过上采样操作后, 特征图像通道数量减半, 输出特征图像尺

寸按照输入图像的最大尺寸作为输出, 最终通过线性映射模块后输出 $H \times W \times 2$ 的输送带边缘图像, 具体结构如图 4 所示。本文提出的融合 CNN 及 Transformer 结构的模块能够有效利用 CNN 和 Transformer 的各自优势分别提取局部和全局特征, 并通过通道加法和上采样操作, 达到交互融合的状态, 既构建了输送带特征图像上下文依赖关系, 又丰富了局部细节信息, 增强了网络提取特征的能力。

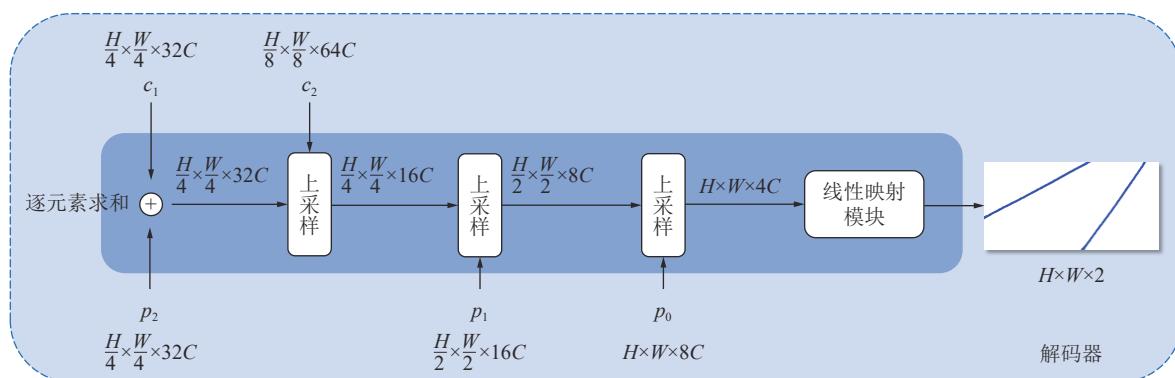


图 4 融合模块

Fig.4 Fusion module

1.5 损失函数

本网络使用的损失函数为交叉熵损失函数^[17], 该损失函数用于描述检测实际预测值与标签值两个概率之间的关系。当交叉熵损失函数的值越小时, 表明预测的结果越接近于目标结果。具体计算公式如下:

$$\text{loss}(x, \text{class}) = -\log \left(\frac{\exp(x[\text{class}])}{\sum_j \exp(x[j])} \right) = -x[\text{class}] + \log \left(\sum_j \exp(x[j]) \right) \quad (6)$$

其中, x 为样本; class 为类别; j 为样本 x 的标签值, 正类为 1, 负类为 0。

1.6 输送带跑偏检测算法

通过计算预测结果图中边缘中心线相对标签图中心线的偏移量, 当偏移量波动超出某一阈值, 视作输送带跑偏。偏移量具体计算流程按如下规则进行。假设 X_{as} 为后处理结果中左侧边缘起始点横坐标, X_{bs} 为后处理结果中右侧边缘起始点横坐标, X_{ae} 为后处理结果中左侧边缘终点横坐标, X_{be} 为后处理结果中

右侧边缘终点横坐标, X_s 为两侧起始点横坐标的平均值, X_e 为两侧起始点横坐标的平均值, 具体示意如图 5 所示。

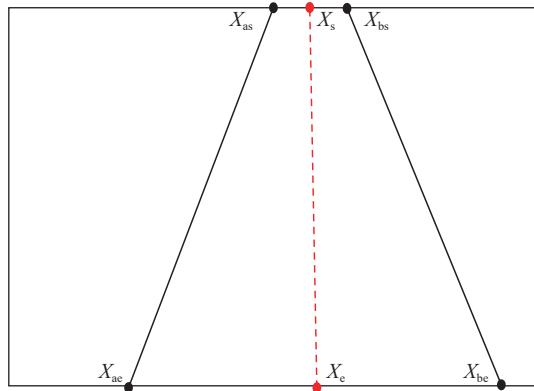


图 5 输送带跑偏检测示意

Fig.5 Schematic diagram of belt deviation detection

基于以上描述有如下关系式:

$$X_s = \frac{X_{as} + X_{bs}}{2} \quad (7)$$

$$X_e = \frac{X_{ae} + X_{be}}{2} \quad (8)$$

传送带偏移值计算公式:

$$\delta = (X_s - \bar{X}_s)^2 + (X_e - \bar{X}_e)^2 \quad (9)$$

其中 \bar{X}_s 和 \bar{X}_e 分别为标签中左右边缘起始点和终点横坐标的平均值, 当 δ 波动大于给定阈值时认为传送带已有跑偏迹象。

2 评价指标

本文网络模型的最终目的是在输送带数据集中检测出测试集中所有的输送带边缘, 以输送带跑偏检测为例, GT(Ground Truth)代表真实标签; TP(True Positive)表示标签为输送带边缘, 预测为输送带边缘; TN(True Negative)表示标签为输送带边缘, 预测为其他; FP(False Positive)表示标签为其他, 预测为输送带边缘; FN(False Negative)表示标签为其他, 预测为其他。指标混淆矩阵见表 1。

表 1 指标混淆矩阵

Table 1 Indicator confusion matrix

真实结果	预测结果	
	真实标签=True	真实标签=False
预测=True	真阳性TP	假阴性FP
预测=False	假阳性FN	真阴性TN

基于以上的概念, 精确度(Precision)表示检测类别是输送带边缘并且标签也是输送带边缘的部分占

所有检测器检测为输送带边缘的比例, 也叫查准率。平均检测精度 mAP 可表示为

$$P' = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (10)$$

$$\text{mAP} = \frac{1}{N} \sum_{n=1}^N \text{AP}(n) \quad (11)$$

$$\text{AP} = \sum_{k=1}^m (r_{i+1} - r_i) P_{\text{inter}}(r_i + 1) \quad (12)$$

其中, r_1, r_2, \dots, r_m 是按升序排列的 Precision 插值段第一个插值处对应的 Recall 值, P_{inter} 为插值操作。其中, 召回率也叫查全率, 表示检测到的是输送带边缘且实际标签也是输送带边缘的数量占检测器检测正确的比例, 召回率能衡量检测器检测出所需类别的能力。用 R' 代表召回率, 则可表示为

$$R' = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

计算得到召回率和精确率之后, 可以根据召回率和精确率计算调和平均值, 其计算数学公式为

$$F_1 = \frac{2 \times \text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (14)$$

准确度(Accuracy)表示为

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (15)$$

均交并比(Mean Intersection over Union, MIoU): 为语义分割的标准度量。其计算 2 个集合的交并比, 在语义分割的问题中, 这两个集合为真实值(ground truth)和预测值(predicted segmentation)。计算公式如下: i 为真实值, j 为预测值, P_{ij} 为预测值 j 预测为真实值 i 的概率。

$$\text{MIoU} = \frac{1}{k+1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}} \quad (16)$$

式(16)等价于:

$$\text{MIoU} = \frac{1}{k+1} \sum_{i=0}^k \frac{\text{TP}}{\text{FN} + \text{FP} + \text{TP}} \quad (17)$$

3 试验与分析

3.1 数据集采集与制作

本文试验使用的数据集是从实际工业场景中摄像仪所采集到多场景的输送带录像中截取相关输送带图片, 共包含五种不同场景下多角度、不同位置的输送带运行状态, 并制作相关数据集, 具体如图 6 所示。

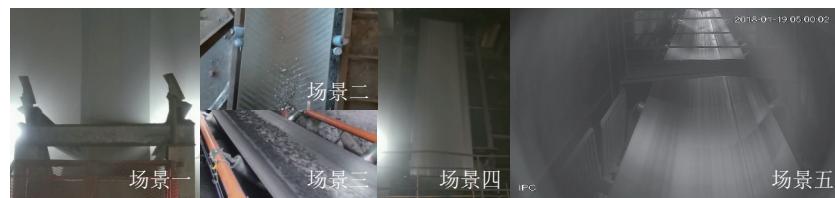


图 6 5 种工业场景下的输送带图片数据集
Fig.6 Dataset of belt images in five industrial scenarios

由于采集到的实际工业环境中的输送带图像大小不统一,且工业场景中固定机位的输送带图像角度单一,截取的图像帧与帧之间差距不大。因此在采集完数据后,通过扩大训练数据量、翻转操作来增加数据集的多样性和泛化性,以达到数据增强的目的,进而使检测模型具有更强的鲁棒性,其中部分数据预处理后的样例如图 7 所示。



图 7 图片预处理后相应结果

Fig.7 The corresponding results after image preprocessing

根据所采集到的图像,在经过数据预处理后,最终形成了包含 5 个场景的图片训练集 1 476 张、测试集 369 张以及对应的标签集,数据集总体情况见表 2。训练集和测试集图像格式为 jpg 格式,为彩色三通道图像;标签集图像格式为 png,为单通道图像,保留了标签真值和背景信息方便后续图像处理。

表 2 本文所制作的数据集介绍

Table 2 Introduction to the datasets produced in this paper

数据集	场景1	场景2	场景3	场景4	场景5	合计
训练集	744	141	135	324	132	1 476
训练标签集	744	141	135	324	132	1 476
测试集	186	33	36	81	33	369
测试标签集	186	33	36	81	33	369

数据集的标注是利用 Photoshop 软件平台对输送带图像进行像素级标注,使用像素宽度为 1 的直线对传送带边缘进行标注,剩余部分为背景,标注结果可视化后的图像如图 8 所示。

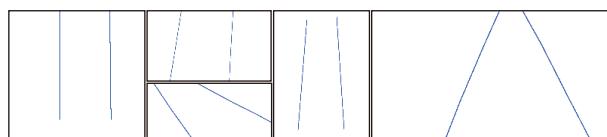


图 8 5 种工业场景下的输送带图片标签数据集

Fig.8 Belt image label datasets in five industrial scenarios

3.2 试验环境和参数

试验在 i9-9820X CPU、两块 NVIDIA RTX 2080 Ti GPU 的设备上进行,操作系统为 Ubuntu 20.04, cuda 版本 10.2, cudnn 版本 7.5, pytorch 版本 1.6, 并在 PyTorch 框架下实现网络模型。试验的总轮数设置为 100 轮,批数大小设为 4,初始的学习率设为 0.001,使用 warming up 预热 1 轮后采取 poly 学习率衰减策略,衰减率为 0.9。采用 AdamW 作为模型的优化器,动量为 0.9,权重衰减项设置为 1×10^{-4} 。

3.3 消融试验

为了验证 DFTNet 模型结构的最优性,从模型深度、输入图片大小以及模型大小 3 个方面设计了消融试验,并在同一数据集上进行试验验证。

模型深度的影响:根据 SwinT 有 4 个连续的模块分别称为“Stage1”、“Stage2”、“Stage3”和“Stage4”,分别对应着不同的模型深度;每个模块相应输出分别表示为“ c_1 ”、“ c_2 ”、“ c_3 ”和“ c_4 ”,详细见参考文献 [14],本文结构主要使用 SWinT 的“ c_1 ”、“ c_2 ”。“ c_1 ”、“ c_2 ”、“ c_3 ”和“ c_4 ”,对应输出特征图像的大小为 $H/4 \times W/4$ 、 $H/8 \times W/8$ 、 $H/16 \times W/16$ 、 $H/32 \times W/32$,意味着模型对图像语义信息提取的深度。本文分别设计构建了“ c_1, c_2 ”、“ c_1, c_2, c_3 ”、“ c_1, c_2, c_3, c_4 ”等 3 种不同深度的网络结构,相应的 CNN 基础模块也随之加深,试验结果见表 3。由试验结果可知模型的检测性能并不是简单地随着网络深度的增加而增加,而模型的计算负载则随着模型深度显着增加。因此,本文模型设置网络深度为“ c_1, c_2 ”,并由此构建相应的融合网络模型。

表 3 模型深度消融试验结果

Table 3 Ablation experimental results of module depth

网络模型深度	均交并比/%	准确度/%	平均调和平均精度/%	平均召回率/%	浮点运算量/ 10^9	参数量/ 10^6
c_1, c_2	91.08	99.48	88.40	91.88	96.22	5.71
c_1, c_2, c_3	87.10	99.33	85.25	89.31	90.37	7.53
c_1, c_2, c_3, c_4	86.68	99.35	84.91	88.61	90.66	9.78

输入图片尺寸的影响:本文分别测试了 $64 \text{ px} \times 64 \text{ px}$, $128 \text{ px} \times 128 \text{ px}$, $256 \text{ px} \times 256 \text{ px}$, $512 \text{ px} \times 512 \text{ px}$

等3种不同输入图片的尺寸,试验结果见表4。随着输入图片尺寸大小从 64×64 增加到 256×256 时,Transformer输入图片的特定序列尺寸也随之变大,模型的分割效果也随之变好,但是输入图片尺寸 512×512 不如 256×256 ,故本文输入图片尺寸大小设置为 256×256 。

表4 输入图片尺寸消融试验结果

Table 4 Ablation experimental results of input size

输入图片尺寸	均交并比/%		准确度/%		平均调和平均值/%		平均精度/%		平均召回率/%
	模型	并比/%	度/%	平均值/%	确度/%	平均值/%	精度/%	平均值/%	
64×64	HRNet	88.02	91.68	88.07	90.37	88.40	91.88	88.40	95.32
128×128	FCN	89.13	96.77	88.03	90.39	88.40	91.88	88.40	95.43
256×256	UNet	91.08	99.48	88.40	91.88	88.40	91.88	88.40	96.22
512×512	SegNet	91.08	99.48	87.50	91.00	87.50	91.00	87.50	94.23

模型大小的影响:根据SwinT的模型类型,一共包含tiny, small, base等3种不同参数量的模型,本文相应设计了3种模型参数量的DFTNet,试验结果见表5。随着模型大小的增大,模型效果并没有较大的提升,但是模型参数量增加,计算量增加,模型运行效率下降。因此,本文模型大小设置为tiny。

表5 模型大小消融试验结果

Table 5 Ablation experimental results of model scale

网络模型大小	均交并比/%		准确度/%		平均调和平均值/%		平均精度/%		浮点运算量/ 10^9	参数量/ 10^6
	模型	并比/%	度/%	平均值/%	确度/%	平均值/%	精度/%	平均值/%		
tiny	DFTNet	91.08	99.48	88.40	91.88	96.22	91.88	91.88	5.71	28.45
small	DeepLab V3	90.10	98.91	87.85	89.31	96.37	89.31	89.31	6.60	49.77
base	PSPNet	91.25	99.35	87.21	89.61	96.16	89.61	89.61	10.78	88.40

3.4 对比试验

为了验证算法方法的有效性及优越性,本文选取大量具有代表性的语义分割网络模型进行对比试验,包括纯深度神经卷积网络架构的FCN(2015)、UNet^[18](2015)、SegNet^[19](2016)、DeepLab v3(2017)、PSPNet^[20](2017)、DanNet^[21](2019)、BiSeNet V2^[22](2021)。本章所有对照试验方法均在同一数据集的训练样本、同一优化器、损失函数、参数的条件下进行试验。由表6、图9、图10、图11可知,在相同数据集下,提出的DFTNet的均交并比mIoU为91.08%、准确度ACC为99.48%、平均调和平均值mF1为88.40%、平均精确度为91.88%,平均召回率mRecall为96.22%为所有模型最高,高于所有纯卷积神经网络。同时,均交并比mIoU、准确度ACC、平均精度mPrecision,相比纯卷积神经网络HRNet分别提升了25.36%、0.29%与17.70%,相比FCN分别提升了29.5%、0.32%与24.77%,在参数量、计算速度

表6 各模型对比试验结果

Table 6 Comparing the experimental results of each model

网络模型	均交并比/%	准确度/%	平均调和平均值/%	平均精度/%	平均召回率/%	浮点运算量/ 10^9	参数量/ 10^6
HRNet	65.72	99.19	70.7	87.02	66.76	5.15	10.42
FCN	61.58	99.16	63.63	81.23	62.09	30.79	15.31
UNet	86.01	99.52	84.31	86.7	82.22	23.73	13.40
SegNet	79.06	99.41	87.76	86.55	82.42	26.50	35.00
DeepLab V3	86.87	99.44	85.05	81.41	89.68	31.33	41.68
PSPNet	87.46	99.49	85.53	83.51	87.83	33.96	48.63
DanNet	81.2	99.43	89.92	85.49	85.89	37.86	49.48
BiSeNet V2	66.43	99.22	71.81	80.16	67.45	2.46	3.62
DFTNet	91.08	99.48	88.4	91.88	96.22	5.71	28.45

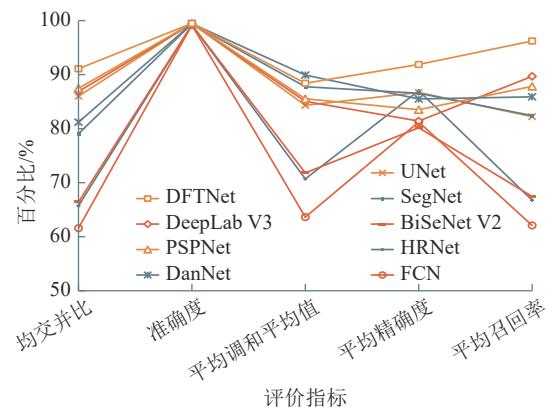


图9 各模型评价指标对比试验结果

Fig.9 The Comparative experiment results of each model's evaluation indicators

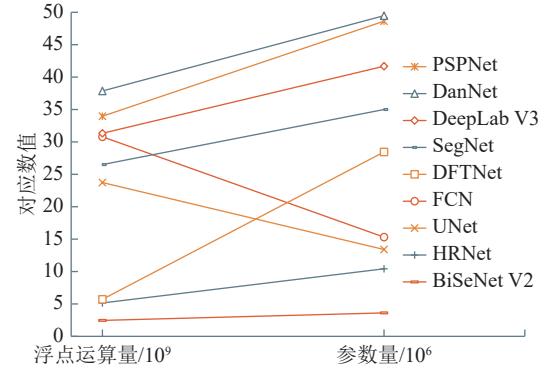


图10 各模型参数量对比试验结果

Fig.10 The Comparative experiment results of each model's parameters

尤其是均交并比等指标上均有较大提升。

从评价指标上来看,DFTNet的各项评价指标为对比试验中最优;从检测结果来看,如图11所示,可以清楚地看到,DFTNet的误分类像素最少;从综合性能上,DFTNet相比其他方法为最优,且在保证检测精度的同时,浮点运算和参数量与纯卷积网络相

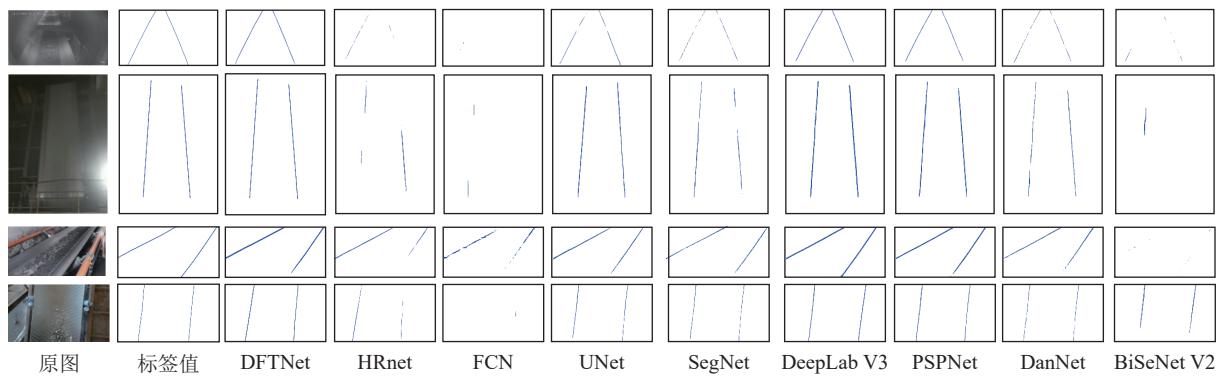


图 11 各模型结果

Fig.11 Results of each model

比也颇具优势,有更好的轻量化和更快的运算速度,同时,模型在批数大小设为4的情况下,处理图像帧率达53.07 fps,更加适用于工业实际场景,可以实现多种场景下输送带边缘的实时性检测,进而实现多场景下的输送带跑偏检测。

4 实际运行结果

在现场环境中搭建了如图12所示输送带跑偏监测系统,并应用了DFTNet网络,由于本模型网络

处理图像帧率可以达53.07帧/s,实际使用中,我们将摄像头采集到的图像实时传输到我们的服务器中进行逐帧处理,处理完毕后通过图像拼接程序还原成实时的跑偏检测图像并显示在前端软件界面,显示结果如图13所示。通过显示界面可以实时显示输送带偏移情况,误差为 ± 0.1 cm,通过实地验证,检测结果与输送带实际跑偏结果一致,证明了DFTNet网络模型的有效性,检测准确率 $\geq 90\%$,具有较好的应用价值。

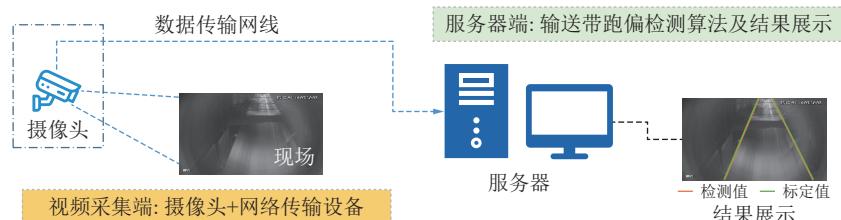


图 12 输送带跑偏监测系统示意

Fig.12 Conveyor belt deviation monitoring system

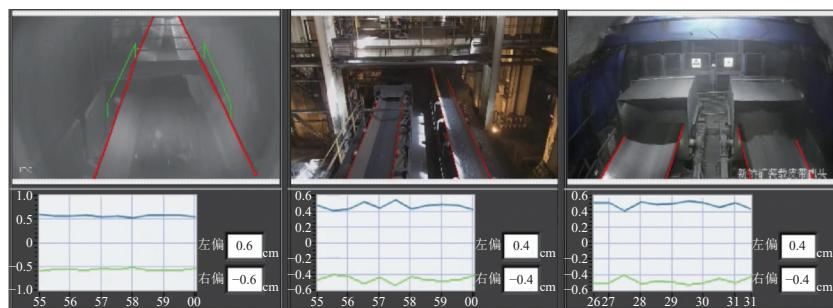


图 13 输送带跑偏检测结果

Fig.13 Conveyor belt deviation detection results

5 结 论

1)提出了一种全局与局部信息双流融合的输送带边缘检测网络模型DFTNet,该模型通过将传统卷积神经网络的卷积对局部特征的提取能力与Transformer结构对全局、长距离信息感知能力相结合,能

够较好地提高输送带边缘检测精度并抑制输送带图像噪声和背景的干扰,相比纯卷积神经网络HR-Net分别提升了25.36%、0.29%与17.70%,相比FCN分别提升了29.5%、0.32%与24.77%,在参数量、计算速度尤其是均交并比等指标上均有较大提升。

2)通过设计CNN和Transformer特征融合模块,

形成双流编码器-解码器结构，利用结构的巧妙设计，可以更好地融合全局上下信息，避免了Transformer结构在大规模数据集上预训练，可以灵活调节网络结构，构建了工业输送带数据集，并在数据集上通过大量试验验证，DFTNet取得了最好的模型性能，在保证检测精度的同时，浮点运算和参数量与纯卷积网络相比也颇具优势，有更好的轻量化和更快的运算速度，同时，模型在批数大小设为4的情况下，处理图像帧率达53.07 fps，更加适用于工业实际场景。

3)在后续研究中，将对该模型进一步改进及优化，同时对数据集进行扩充，使模型其更加具有泛化性，可以更加有效地应用到背景更加复杂的工业场景下输送带边缘检测中。

参考文献(References):

- [1] 张佳,尹君驰,王宏等.输送带输煤采样技术研究现状及发展趋势[J].煤炭科学技术,2022,50(9):200–206.
ZHANG Jia, YIN Junchi, WANG Hong, et al. Research status and development trend of conveyor belt coal transport sampling technology[J]. Coal Science and Technology, 2022, 50(9): 200–206.
- [2] 王海军,王洪磊.基于参数化对数图像处理模型的光照不均匀图像的边缘检测算法[J].煤炭科学技术,2022,50(12):225–239.
WANG Haijun, WANG Honglei. Status and prospect of intelligent key technologies of belt conveyorStatus and prospect of intelligent key technologies of belt conveyor[J]. Coal Science and Technology, 2022, 50(12): 225–239.
- [3] 谭恒,张红娟,靳宝全等.基于机器视觉的煤矿带式输送机跑偏检测方法[J].煤炭技术,2021,40(5):152–156.
TAN Heng, ZHANG Hongjuan, JIN Baoquan, et al. Machine vision-based coal mines band-type conveyor running partial detection method[J]. Coal Technology, 2021, 40(5): 152–156.
- [4] 徐欢,李振璧,姜媛媛等.基于OpenCV的输送带跑偏自动检测算法研究[J].工矿自动化,2014,40(9):48–52.
XU Huan, LI Zhenyu, JIANG Yuanyuan, et al. Research on automatic detection algorithm based on OpenCV-based conveyor belt running partial test algorithm[J]. Industry and Mining Automation, 2014, 40(9): 48–52.
- [5] 韩涛,黄友锐,张立志等.基于图像识别的带式输送机输煤量和跑偏检测方法[J].工矿自动化,2020,46(4):17–22.
HAN Tao, HUANG Yirui, ZHANG Lizhi, et al. The coal transmission volume and running test method based on image recognition [J]. Industry and Mining Automation, 2014, 40 (9): 48–52.
- [6] LIU Y, Wang Y, Zeng C, et al. Edge detection for conveyor belt based on the deep convolutional network[A]. Proceedings of 2018 Chinese Intelligent Systems Conference[C]. Springer, Singapore, 2019: 275–283.
- [7] LONG J, SHELHAMER E, & DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(4): 474–482.
- [8] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834–848.
- [9] XIE S, TU Z. Holistically-nested edge detection[A]//Proceedings of the IEEE international conference on computer vision[C]. 2015: 1395–1403.
- [10] POMA X S, SAPPA A D. Improving edge detection in RGB images by adding NIR channel [A]//2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). IEEE, [C]. Las Palmas de Gran Canaria, Spain, 2018: 266–273.
- [11] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[A]//Advances in neural information processing systems[C]. Long Beach, CA, United states: MIT, 2017: 5998.
- [12] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[A]. International Conference on Learning Representations [C] Vienna: Springer, 2021.
- [13] TOUVRON H, CORD M, DOUZE M, et al. Training data-efficient image transformers & distillation through attention[A]//International Conference on Machine Learning. PMLR[C], 2021: 10347–10357.
- [14] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[A]//Proceedings of the IEEE/CVF International Conference on Computer Vision[C]. Jeju, Korea: IEEE, 2021: 10012–10022.
- [15] HU H, GU J, ZHANG Z, et al. Relation networks for object detection[A]//Proceedings of the IEEE conference on computer vision and pattern recognition[C]. 2018: 3588–3597.
- [16] HU H, ZHANG Z, XIE Z, et al. Local relation networks for image recognition[A]. Proceedings of the IEEE/CVF International Conference on Computer Vision[C] 2019: 3464–3473.
- [17] ZHANG Z, SABUNCU M. Generalized cross entropy loss for training deep neural networks with noisy labels[J]. Advances in neural information processing systems, 2018, 31: 2241–2252.
- [18] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[A]. International Conference on Medical image computing and computer-assisted intervention[C] Springer, Cham, 2015: 234–241.
- [19] BADRINARAYANAN V, KENDALL A, CIOPPOLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(12): 2481–2495.
- [20] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[A]//Proceedings of the IEEE conference on computer vision and pattern recognition[C] Hawaii, 2017: 2881–2890.
- [21] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation [A]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition [C]. Tennessee, 2019: 3146–3154.
- [22] YU C, GAO C, WANG J, et al. Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation[J]. *International Journal of Computer Vision*, 2021, 129(11): 3051–3068.