Universe Scientific Publishing
Modern Chemical Industry, 现代化工 (4)2024,6
ISSN: 2661-3670(Print) 2661-3689(Online)

基于多帧图像信息与 Transformer 的弓网燃弧检测算法研究

张 雄[1] 吴 卓[2]
（1 北京轨道交通技术装备集团有限公司 北京 100071；2 北京京投轨道交通技术研究院有限公司 北京 100071）

摘 要：弓网燃弧的发生在相机成像上存在较明显的特征，因此可基于计算机视觉进行燃弧检测。但燃弧的判断易与环境中的其他干扰项混淆。传统的图像处理技术对燃弧的亮度与边缘形态等有较多的假设通用性差。基于深度学习模型的燃弧检测算法虽可通过大量样本进行训练但仅基于单帧图像作为输入，无效利用多帧图像信息。本文采用人工特征提取的方式对连续多帧图像进行降维，将预处理信息输入至 Transformer 模型中学习燃弧特征，利用单帧图像的亮度信息与连续多帧图像亮度的动态变化信息进行燃弧判断。且通过测试可知，相较于基于单帧图像作为输入，本算法采用多帧图像输入时效果更佳。

关键词：燃弧；深度学习；单帧；多帧；人工特征提取；Transformer

引言

随着计算机视觉技术的快速发展，基于图像处理的检测方法被广泛应用于弓网燃弧检测，存在传统图像处理方法与基于深度学习的方法。但目前研究大多基于单帧图像进行燃弧检测，考虑到燃弧发生场景的复杂性，仅凭单帧图像在一些场景下很难判断目标是否为燃弧。本文提出了一种利用多帧的图像信息与 Transformer 的处理方法，将多帧图像经预处理降低维度后输入给 Transformer 模型，通过学习燃弧发生时视频前后帧的上下文信息达到提升燃弧检测效果的目的。

1.基于图像信息的弓网燃弧检测

1.1 现有算法概述

现有的燃弧图像检测方法可分为基于传统图像处理与基于深度学习模型两大类。

基于传统图像处理的方法需对图像进行滤波、二值化、边缘提取、形态学操作等处理。此类研究成果大多对其中的一项或多项进行优化或改进。如张振琛等[1]改进了 canny 边缘提取算法提高图像边缘提取的质量。杨恒[2]将图像进行二值化，基于面积信息来判断燃弧。吴琛等[3]在边缘提取后经形态学处理得到接触线与受电弓的接触区域，在区域内通过合适的阈值对图像进行二值化来判断是否存在燃弧。

传统图像处理因依赖阈值选取存在局限性，基于深度学习的方法逐渐受到重视。而注意力机制自提出以来，也被广泛用于各类图像检测任务[4]。权伟[5,6]等将注意力机制引入深度学习模型用于燃弧检测。张雯[7]将注意力机制引入经典的目标检测算法（FasterR-CNN 和 YOLOv5）取得了比原始模型更好的效果。本文的算法也基于注意力机制，但在模型的输入上针对燃弧图像处理的难点和现有算法的缺陷做了针对性的优化。

1.2 燃弧图像处理算法难点

目前不论传统图像处理方法还是基于深度学习模型的方法大都基于单帧图像作为输入，在一些复杂的光影条件中仍在存在误判，这些复杂场景有以下几类情况。

1.2.1 参照物成像不佳

很多燃弧检测算法依赖于图像中受电弓的检测，但在一些场景下（如由于雨雾天气或镜头脏污的影响）虽然可观察到燃弧，但受电弓不一定能检测成功。

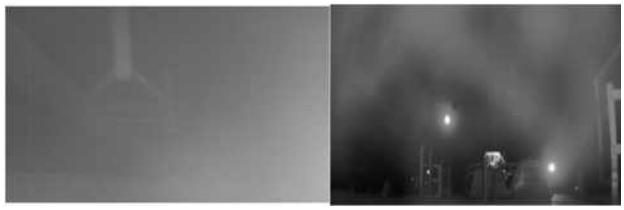

图 1 参照物成像不佳示例

1.2.2 燃弧亮斑形态的多样性

很多传统图像处理算法默认燃弧亮斑呈椭圆状。但存在一些其他形态的亮斑，如放射状亮斑，空洞式亮斑。

图 2　放射状燃弧示例

如图 2 所示的放射状亮斑多见于隧道线路，图像中的亮斑本身其实不属于燃弧，而是被燃弧照亮的周边金属物体。

图 3　空洞状燃弧示例

如图 3 所示的空洞状亮斑多见于露天线路，很多传统图像处理算法假设燃弧亮斑处的像素灰度值是较高区域，二值化后可被保留下来，但实际场景中存在不少非典型燃弧成像，燃弧发生位置存在极小的亮斑，其周围存在一圈暗影，而再外层则是另一圈逐渐变暗的辉光。

1.2.3 干扰多样性

燃弧的产生可导致图像中出现亮斑，但实际场景中也存在多种其他亮斑易对算法产生影响。干扰亮斑可能来自于日月光(如图 4)、灯光(如图 5，图 6)，有些甚至是线路附近飞行器的闪光(如图 7)。

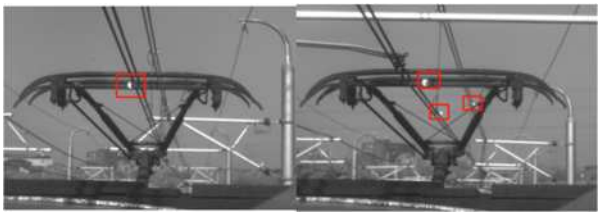

图 4　露天阳光反光干扰示例

图 5　灯光干扰示例

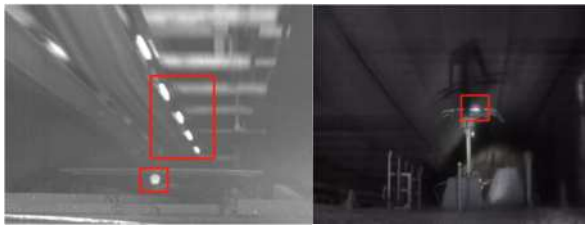

图 6　隧道内补光灯反光干扰示例

图 7　飞行器干扰示例

1.3 基于多帧图像信息的燃弧检测

深度学习模型可从单帧图像样本中学习到燃弧像素的灰度、燃弧目标的形态等信息，但这对燃弧判断是不够的。如 1.2 节中所述的场景，在仅提供单帧图像情况下，深度学习模型甚至是人眼有时也无法判断亮斑是否为燃弧。

燃弧所属的亮斑在单帧图像的亮度、形态可能与干扰亮斑极为类似，但在多帧图像的上下文背景下它们存在显著不同，即燃弧发生处的亮斑在视频前后帧存在较大的动态变化。

设 I_i 为第i帧图像，其与前一帧的像素差可表示为 $\Delta I_i = I_i - I_{i-1}$，其与后一帧的像素差记为 $\Delta I_{i+1} = I_{i+1} - I_i$。

2.基于多帧图像信息与 Transformer 的燃弧检测

2.1 用于图像检测任务的 Transformer

Transformer 是一种最初设计用于自然语言处理任务的深度学习模型，它也被成功地应用于图像检测任务中。其核心是自注意力机制 (Self-Attention Mechanism) [4]，其允许模型在处理序列数据时捕捉到长距离的依赖关系，从而提高检测的准确性，有时甚至超过了传统的 CNN 模型。ViT (Vision Transformer) [8]与 Detection Transformer (DETR) [9]是其中的典型代表。ViT 模型结构简单，但其无法用于目标检测任务，DETR 虽可用于目标检测任务但结构更为复杂。本文算法结合 ViT 与 DETR，将 DETR 的 CNN backbone 替换为 ViT 的前端处理，并预定义一些特征做数据降维降低处理复杂度。

2.2 基于多帧图像信息的 Transformer 模型结构

算法模型可以分为前处理、Transformer、后处理三个部分，流程

如图所示。

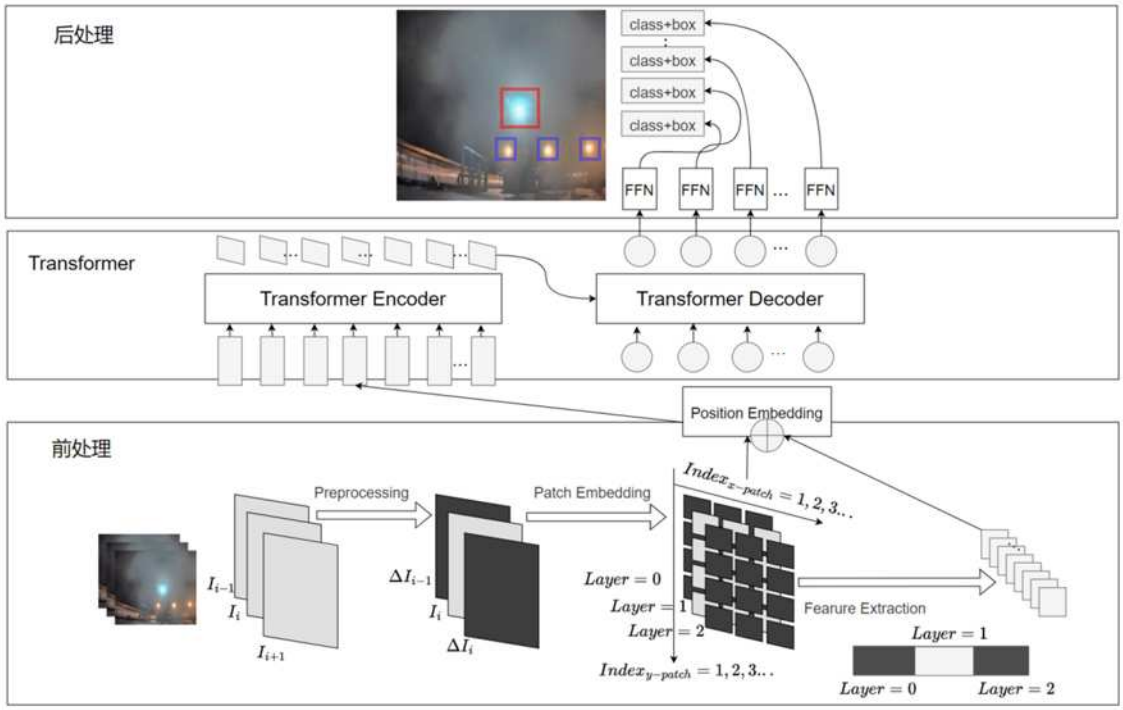

图 8　算法流程

2.2.1 前处理

因本模型输入为连续多帧，考虑连续的三帧图像，记为 I_{i-1}, I_i, I_{i+1}.

原始 DETR 模型经 CNN 进行特征提取后，得到维度为 2048，经降采样后为 256 个维度。与 DETR 不同的是，这里不采用端到端的处理方法直接把原始图像送入 CNN，而是利用先验信息人为对多帧图像提取特征降维后输入到 Transformer 中，以降低计算量。

对于单帧图像难以判别的燃弧，可利用多帧图像相关位置的明暗突变情况排除灯光等非燃弧目标的干扰。这里计算图像帧 i 与前后各一帧的帧间差，可得 $\Delta I_{i-1} = I_i - I_{i-1}, \Delta I_i = I_{i+1} - I_i$.由此可得到三帧图像，即第 i 帧图像本身 I 、与前后各一帧的帧间差 $\Delta I_{i-1}, \Delta I_i$. $\Delta I_{i-1}, I_i, \Delta I_i$ 这三层图像帧的位置编码分别记为 $Layer = 1,2,3$.

对于 $\Delta I_{i-1}, \Delta I_i$ 的每一帧采用与 ViT 中的分块方式对各帧进行分块，高度与宽度方向上的分块参数分别为 H_{block}, W_{block} ，由此可以将一帧分为 $H / H_{block} * W / W_{block}$ 个图像块。对于每个小块，仅考虑亮度信息，若输入为三通道的彩色图像，则采用 RGB 三通道转亮度进行降维，亮度值 $Y = 0.299R + 0.587G + 0.114B$. 得到亮度后，计算该小块内所有像素的灰度直方图，对于单通道图像而言是统计块内 0~255 各个灰度上的像素个数，共计 256 个数值。这里将直方图的灰度数从 256 个降采样为 H_{span} 个，即将灰度直方图的输出维度是 $1 \times H_{span}$ ，各灰度区间的像素个数分别为 $[Hist_1, Hist_2, ..., Hist_{H_{span}}]$.

除灰度直方图外，再求取该区域内的像素最大值 $gray_{\max}$ ，最小值 $gray_{\min}$ ，平均值 $gray_{avg}$ ，标准差 $gray_{std}$.这里将每个小块的直方图信息结合灰度最大值、最小值、平均值、标准差作为表征该区域内灰度信息的特征，特征维度个数为 $4 + H_{span}$.对三帧图像都进行上述的分块与块内提取，因此可以得到的特征的维度为 $\left(3, H / H_{block} * W / W_{block}, 4 + H_{span}\right)$.

在送入 Transformer 之前，对上述特征进行维度变换，首先将图像块展平，共计 $H / H_{block} * W / W_{block}$ 个图像块。每个图像块可以视为一个 token. 对每个图像块，每个帧提取得到的 $4 + H_{span}$ 个特征按照 $\Delta I_{i-1}, I_i, \Delta I_i$ 依次进行拼接， 因此每个 token 对应的特征维度为 $3\left(4 + H_{span}\right)$. 得到上述 token 后，在输入 Transformer 前需加入位置编码，这里编码方式与 DETR 保持一致，使用正弦余弦函数来生成绝对位置编码，与原特征进行相加。

2.2.2Transformer

Transformer 在结构上与 DETR 一致，分为 Encoder 与 Decoder。因燃弧检测的目标较少，远小于原始 DETR 模型要检测的目标，因此本模型对结构进行了简化。

其中，在 Encoder 端每个 token 的特征维度不再是 256，而是 $3\left(4 + H_{span}\right)$.

DETR 的 Decoder 端，object queries 为图像中目标检测框的个数

N，原始 DETR 中为 100，因为燃弧检测的场景目标个数较少，因此可以取较小的数值。

在燃弧检测任务中的待检测目标仅为燃弧一种，但燃弧极易与灯光，阳光等干扰物混淆，因此在训练上，存在两类标签，即“燃弧”，与“非燃弧发光物”，考虑到相机视野内可能同时存在包括燃弧在内的多个发光物体，这算法取 $N=16$.

2.2.3 后处理

后处理方法与 DETR 一致，基于 FFN（feed forward network）得到每个预测的目标类型和矩形框位置信息。与原始 DETR 不同的是，这里忽略其他目标信息，仅提取类型为“燃弧”的目标。

3.算法验证

为验证算法效果，在测试集上对算法的精度（AP）与召回率（R）进行计算。因业务场景仅需考虑燃弧，故仅评估对燃弧这一个目标类型，其他的目标类型不计入评估。

为验证多帧图像作为输入对燃弧判断的效果，基于前述模型，考虑仅采用当前帧图像、采用当前帧及前一帧图像、采用当前帧及前后各一帧三种输入方式，分别将算法记为 single-img, multi-img-2, multi-img-3，分别各算法在测试集上的效果。这三种算法仅前端输入不同，选取的超参数、特征提取方式与 Transformer 的结构完全一致，在同样的训练集上训练完成后分别测试其效果。图像采集自帧率为 20fps 的监控视频，图像前后帧的间隔时长为 50ms。对比结果如表 1 及图 1 所示。

表 1　各算法的精度与召回率测试结果

算法	精度 AP(%)	召回率 R(%)
single-img	84.83	87.07
multi-img-2	93.10	92.74
multi-img-3	94.33	93.21

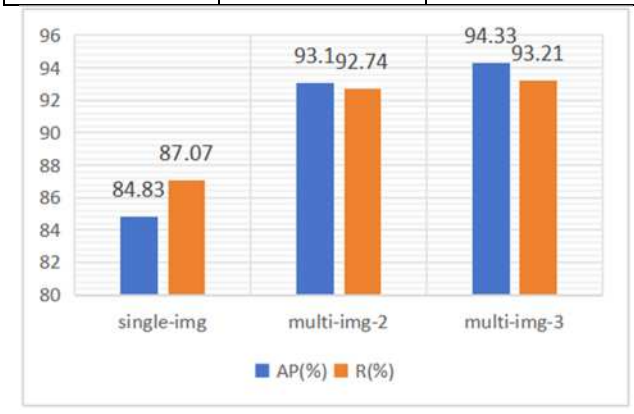

图 3　各算法的精度与召回率测试结果

采用单帧图像的预处理信息作为 Transformer 模型的输入时，燃弧判断的精度与召回率仅为 84.83%与 87.07%。当引入前一帧的图像信息后的 multi-img-2 算法表现显著提升，精度提高至 93.1%，召回率也提升至 92.74%。可见仅凭单帧图像信息进行燃弧检测是不够的，引入多帧信息后可有效提高判断准确性。

基于 multi-img-2 再引入后一帧图像，共计 3 帧图像的信息后，multi-img-3 算法精度与召回率进一步提高至 94.73%与 93.21%。多帧图像理论可为燃弧判断带来更丰富的信息。但相比 multi-img-2 较 single-img 的大幅提升，multi-img-3 较 multi-img-2 的提升幅度有限。

4.结论

图像的亮度信息与多帧图像的亮度动态变化情况可有效用于燃弧判断。旨在提取亮度信息的人工特征提取方法结合 Transformer 可有效用于图像燃弧检测。但在输入仅为单帧图像时，算法效果有限，借助连续多帧的图像信息可有效降低召回率并提高判断精度。

参考文献：

[1]张振琛,顾桂梅,李占斌.基于图像处理的弓网燃弧检测方法[J].兰州交通大学学报，2020，39(2):7.DOI:CNKI:SUN:LZTX.0.2020-02-008.

[2]杨恒,伍川辉,吴琛.基于图像处理弓网燃弧检测研究[J].铁道科学与工程学报，2018，15(4):6.DOI:10.3969/j.issn.1672-7029.2018.04.028.

[3]吴琛,伍川辉,杨恒,等.基于 LabVIEW 图像处理的弓网拉弧在线监测研究[J].铁道标准设计，2018，62(9):4.DOI:10.13238/j.issn.1004-2954.201711210005.

[4]Mnih V，Heess N，Graves A，et al.Recurrent Models of Visual Attention[J].Advances in Neural Information Processing Systems, 2014, 3.DOI:10.48550/arXiv.1406.6247.

[5]权伟，刘洋.一种受电弓与接触网燃弧视觉检测方法:CN202211373615.X[P].CN115861870A[2024-05-03].

[6]权伟,郭少鹏,周宁,等.一种电气化铁路弓网燃弧视觉检测方法.CN202110102075.0[2024-05-04].

[7]张雯. 基于注意力机制的弓网燃弧检测算法研究[D]. 北京:北京交通大学,2022.

[8]Dosovitskiy A，Beyer L，Kolesnikov A，et al.An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[C]// International Conference on Learning Representations.2021.

[9]Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//European conference on computer vision. Cham: Springer International Publishing, 2020: 213-229.