

# 基于卷积神经网络的原烟仓库虫情检测算法研究

姜福焕<sup>1</sup>, 严祥辉<sup>2</sup>, 范明登<sup>1</sup>

(1. 福建省龙岩金叶复烤有限责任公司, 福建 龙岩 364102;

2. 厦门烟草工业有限责任公司, 福建 厦门 361026)

**摘要:** 随着精细化管理的不断深入, 传统的仓储养护管理模式也面临着巨大的挑战, 原有的人工养护管理模式已经无法满足精细化和智能化仓储管理的需要; 为此, 以原烟仓储环节中虫情监测为探究应用方向, 基于深度学习技术与目标检测技术, 设计适用于烟草粉螟、烟草甲虫检测的检测网络, 该网络采用轻量化结构设计, 同时基于 Transformer 机制构建了主干网络, 以实现快速、精准的检测目的; 经过实验证明, 该网络模型在 GPU 上的检测速度达 50 FPS 可实现实时检测、且检测精度达 96.7%; 通过实时获取的检测数据结合虫情检测系统能够快速评估出原烟仓储环节中的虫情实况, 实现原烟仓储虫情监测管理的信息化和智能化, 进而为打叶复烤行业仓储虫情监测管理提供可行参考方案。

**关键词:** 卷积神经网络; 原烟仓库; 虫情监测; 深度学习; 目标检测

## Research on Pest Monitoring Detection Algorithm in Raw Tobacco Storage Based on Convolutional Neural Network

JIANG Fuhuan<sup>1</sup>, YAN Xianghui<sup>2</sup>, FAN Mingdeng<sup>1</sup>

(1. Fujian Longyan Gold Leaf Redrying Co., Ltd., Longyan 364102, China;

2. Xiamen Tobacco Industrial Co., Ltd., Xiamen 361026, China)

**Abstract:** With the deepening of fine management, traditional storage maintenance management models are also facing great challenges. Original artificial maintenance management models cannot meet the needs of fine and intelligent storage management. Therefore, insect pest monitoring in the storage of raw tobacco is taken as a research and application direction, a detection network based on deep learning and target detection technology is designed for the detection of tobacco meal borer and tobacco beetle. The detection network is designed with lightweight structure. Meanwhile, the main network based on Transformer mechanism is built to achieve the rapid and accurate detection. The experimental results show that the network can realize the real-time detection speed of 50 FPS and detection accuracy of 96.7% on the GPUs. Combined with the pest detection system, the insect pest situation in the raw tobacco storage link can be rapidly evaluated by the real-time detection data, which realizes the information and intelligence of pest monitoring management in the raw tobacco storage. Then, it provides a feasible reference scheme for the monitoring and management of pest situation in storage of threshing and rebaking industry.

**Keywords:** convolutional neural network; raw tobacco storage; pest monitoring; deep learning; target detection

## 0 引言

在原烟仓储中除了需要对温度、湿度有极高的要求外, 对于仓库中的虫情也需要进行严格的控制。以最常见的烟草粉螟、烟草甲虫为例, 该类害虫通过啃食烟叶为生, 除了对烟叶的品相、重量造成影响外, 其排泄物也会对烟叶的品质造成影响<sup>[1]</sup>。而在对于烟草害虫的防治上, 需要根据虫口数量、密度选择不同的治理方式, 这需要通过监测虫情以获得相关数据<sup>[2]</sup>。目前主要的监测方式为人工记录监测, 受限于仓库数量、人工数量等客观因素影响, 虫情监测流程工作效率低下, 且随着记录员的工作时间增加, 记录的精度也受到影响<sup>[3]</sup>。为此, 如何提高虫情监测效率是目前行业中主要研究方向。

经过多年的研究, 国内外学者已提出了很多种粮虫种类信息识别的先进技术方法, 其中最具代表性的主要有以下几种:

1) 声音检测法: 声音检测法是一种通过对粮堆中储粮害虫发出的声音进行数据采样, 随后采用技术手段对声音数据进行检测、分析, 特征提取与匹配, 据此来实现粮虫种类识别的方法<sup>[4]</sup>。

2) 近红外法: 近红外法是一种通过将近红外光向不同的储粮害虫进行照射, 然后对所得到的近红外光频谱进行图像分析, 来区分不同种类粮虫的方法。由于不同类别的粮虫身体成分有差异, 其反射和吸收的近红外光谱也有差异, 根据这些差异化表现便可以分辨出不同种类的储粮害虫<sup>[5]</sup>。

收稿日期: 2023-04-13; 修回日期: 2023-05-22。

作者简介: 姜福焕(1988-), 男, 大学本科, 工程师。

通讯作者: 范明登(1984-), 男, 大学本科, 工程师。

引用格式: 姜福焕, 严祥辉, 范明登. 基于卷积神经网络的原烟仓库虫情检测算法研究[J]. 计算机测量与控制, 2023, 31(10): 76-82, 89.

3) 电导率测量法: 电导率测量法是基于电导、电压的物理特性, 根据所测导体水分对电导率的影响, 来判断粮食中的害虫情况。通常, 受到害虫侵害的粮食湿度要比无虫害粮食大, 湿度越高则电导率越小<sup>[6]</sup>。

4) 图像识别法: 图像识别法是指利用摄像手段获取粮虫图像, 通过图像处理、机器学习等技术提取储粮害虫的图像特征, 然后进行图像特征的特征向量分类, 从而识别不同储粮害虫的类别的方法<sup>[7]</sup>。

以上各种方法虽然技术手段先进、丰富, 对粮虫种类的检测精度在特定环境下也能达到要求, 但都存在着一定的问题, 这些问题阻碍了各方法的进一步应用与推广<sup>[8]</sup>。具体问题如下:

1) 声音检测法对采样设备周围环境的安静程度要求很高, 在有环境噪声的情况下, 声音检测法难以达到理想的检测效果。

2) 近红外法对个体体积差异较大的粮虫分辨效果较好, 但对体积差异小的粮虫而言, 其反射的红外光谱图差异化较小, 区分难度大。

3) 由于粮堆电导率的数值大小与虫量成正比, 所以该法只能检测粮堆内是否有虫以及虫量的多少, 不能对粮虫的种类进行区分, 实际意义不大, 应用场景不多。

4) 关于图像识别法, 虽然取得了一些进展, 但大部分研究中使用的储粮害虫图像过于理想化, 粮虫图像背景简单, 背景色与粮虫颜色区分度大, 所采用的训练集皆为理想角度、完整、高清晰度的图像, 甚至一张图像中仅有一只粮虫, 这与实际的储粮环境不符。

得益于人工智能热潮, 计算机视觉技术在近年有了快速发展。相比于传统目标检测算法基于人工先验知识设计特定目标的特征检测器, 基于卷积神经网络的目标检测算法则可通过学习样本数据由网络自动提取目标特征值, 在泛化性及鲁棒性上相比于传统算法都有了大幅度的提升进步<sup>[9]</sup>。

综上, 为解决目前烟叶虫情管理模式的弊端以及将深度学习算法更好地应用于烟虫监测任务中, 本文基于深度学习算法设计了专门适用于原烟仓库的虫害检测网络, 通过对网络结构进行优化, 使其在保证高精度的同时兼具有轻量化的特性, 通过对虫情数据的实时采集, 实现原烟仓储虫情的预警与分析, 有效提升原烟仓储烟叶养护水平, 进而为仓储养护管理供数据化的支撑, 使其成为烟叶质量保障的新亮点。

## 1 理论分析

为了提高对小目标的检测精度, 本文主要采用了 Transformer 模块搭建特征提取网络, 依靠 Transformer 的全局自注意力机制, 实现对小目标特征的加权提取。Transformer 最初应用于自然语言领域中, 而通过对其输入端进行结构改造后, 其可应用于计算机视觉领域<sup>[10]</sup>。标准的 Vision Transformer 模块结构及对数据特征的处理流程如图 1

所示。

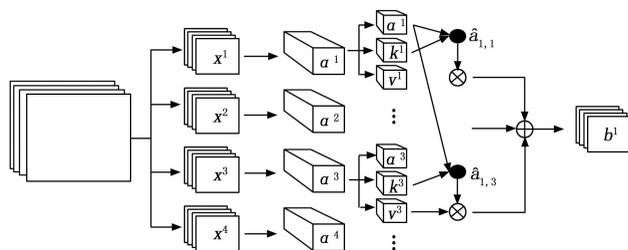


图 1 Vision Transformer 模块结构

不同于卷积模块对整张特征图进行特征提取, Vision Transformer 为了获取特征图中各个区域的关联特征, 模块的初始操作会将输入的特征图进行分块, 之后将每个特征图块由  $1 \times 1$  大小的卷积核转化为向量。如图 1 所示, 特征图先分割得出  $x^1$ 、 $x^2$ 、 $x^3$ 、 $x^4$  四个特征块, 再由卷积核得到  $\alpha^1$ 、 $\alpha^2$ 、 $\alpha^3$ 、 $\alpha^4$  四个向量。之后对于每个特征向量会分别由  $q$  (query)、 $k$  (key)、 $v$  (value) 三个向量进行计算。以  $\alpha^1$  向量为例,  $q$  向量的作用主要为计算  $\alpha^1$  向量与其他向量之间的关联匹配度; 而  $k$  向量的作用为当其他向量的  $q$  向量进行关联度计算时自身与其的关联匹配度;  $v$  向量则为特征表达计算。而  $\alpha^1$  在计算与其他输入向量的关系特征时, 其  $q^1$  会分别与  $k^1$ 、 $k^2$ 、 $k^3$ 、 $k^4$  进行查询计算并得到  $[\tilde{a}_{1,1}, \tilde{a}_{1,2}, \tilde{a}_{1,3}, \tilde{a}_{1,4}]$  这样一组权重向量, 之后每个权重向量与输入特征进行重组, 即可得到特征图块  $x^1$  与其他块之间的关系表达特征图  $b^1$ 。最终各个特征图块完成自注意力计算后, 将输出的各个关系表达特征图  $b^i$  合并后即可得到完整的全局注意力特征图。因此采用 vision Transformer 进行卷积计算使得网络除了能够获取到每个图片块中的目标特征信息外, 还能够对特征图中的目标进行提权, 同时可提取到特征图各个块区域中目标与目标在空间位置上的关联特征信息<sup>[11]</sup>。

## 2 网络设计

结合网络模型的应用场景, 模型需要尽可能地轻量化, 以尽可能满足在多种设备上嵌入运行, 其次为检测的目标具有小、密等特点, 不易于检测<sup>[12]</sup>。从第一方面考虑本文选用了单阶段的检测网络, 通过精简的检测流程达到快速的检测的目的<sup>[13]</sup>。而对于目标小而密的检测难点, 单阶段网络具有一定的不足性<sup>[14]</sup>。综上, 对于虫害检测网络的设计本文主要围绕模型轻量化与提升密集目标检测精度两方面进行。在网络的设计中, 本文将 Transformer 模块与标准的卷积模块进行组合, 构建了鲁棒性更强的特征提取网络, 网络结构如图 2 所示。

本文所设计的结构为“Conv+Transformer+Conv”的组合形式, 该结构的优势在于整合了卷积网络的局部感知性与 Transformer 的全局感知型, 使得网络整体对于目标具有更高的敏感性, 降低漏检、误检的概率。

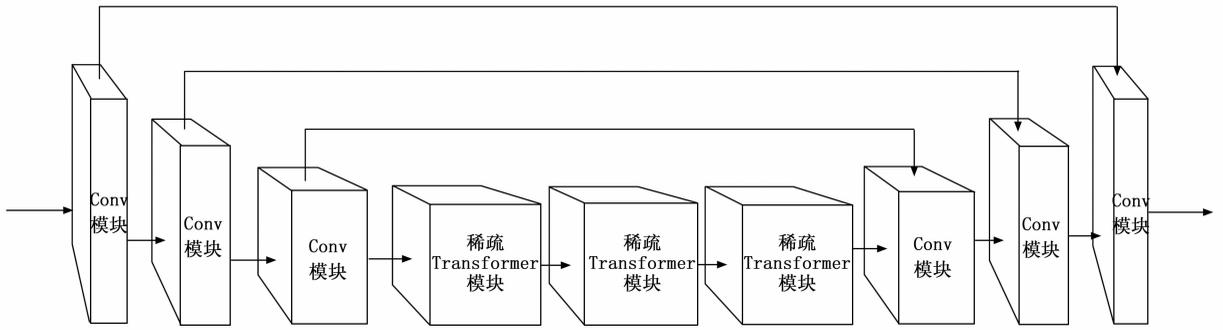


图 2 特征提取网络结构

### 2.1 卷积网络模块设计

在主干网络的首尾部本文采用了标准卷积层进行构建，由于卷积对于图片像素更加敏感，因此在网络的输入端部分采用标准卷积能够将图片快速转换为包含有低级形态的特征的特征图，使得后续的 Transformer 模块不需要进行特征转换可直接进行全局特征的提取<sup>[15]</sup>；而在网络的尾部，为了衔接检测器的输入仍采用了卷积层对特征图进行处理<sup>[16]</sup>。在首尾模块的设计中本文基于特征图金字塔结构进行优化在原基础上加入通道注意力机制，利用了 FPN 网络编解码及特征传递的特性，以提升特征图的丰富性。特征金字塔网络（FPN, feature pyramid network）是一种用于计算机视觉任务的神经网络结构。如图 3 所示，FPN 可以从多个尺度的特征图中提取特征信息，从而提高模型的性能和鲁棒性。FPN 的主要组成部分包括：1) 自底向上的特征提取网络，从输入图像中提取多个尺度的特征图，每个特征图对应不同的物体大小和尺度；2) 自顶向下的特征融合网络，将高层特征图与低层特征图进行融合，以便在不同尺度上提取更加丰富和准确的特征信息；3) 横向连接，在特征融合的过程中，FPN 还通过横向连接将相邻尺度的特征图进行连接，从而进一步提高特征的准确性和丰富性。FPN 在计算机视觉任务中的应用非常广泛，例如目标检测、语义分割和实例分割等任务。通过使用 FPN，可以使模型在不同尺度上都能够有效地提取特征信息，从而提高模型的性能和鲁棒性。

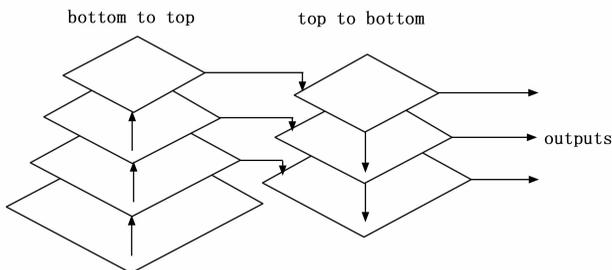


图 3 特征金字塔网络示意图

优化后的网络结构如图 4 所示。

在网络的首部卷积层以编码的形式通过下采样的形式进行组合逐层缩小特征图，而在尾部则是通过上采样以解

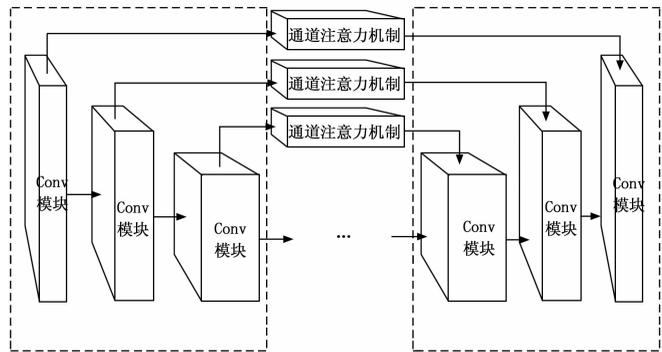


图 4 通道注意力 FPN 网络

码的形式逐层放大特征图。由于本文所检测的目标较小，特征图尺寸越大目标，所能在特征图上体现的特征信息也越多，也利于检测，因此在主干网络的尾部采用上采样卷积将特征图扩大。同时，在网络的首部与尾部本文设计了特征图通道注意力机制映射模块，将浅层网络输出的特征图映射输入到深层网络所输出的特征图中，对特征图种的信息进行补充。这种传递方式提高了对于特征信息的复用，但同样也会将背景等无效信息传输到深层网络，引入一定的噪声干扰。为了消除背景信息的噪声干扰，本文在链接通道中加入了通道注意力机制模块，通过对特征图的通道额外增加权值，以表示不同的通道与特征图中目标信息的相关性，权值越大说明相关性越强，通过加权的方式对噪声特征的通道进行降权抑制，对包含目标特征的通达进行提权增强，所设计的通道注意力机制如图 5 所示。

在通道注意力模块的设计中本文考虑了轻量化的设计因素，将对每个通道均进行权值计算改为对每组道通统一进行权值计算，以减少网络的参数量。整个加权流程分为两个部分，在第一个部分中首先将需要映射的特征图通道进行分组映射，以上图为例首先将  $n \times n \times c$  大小的特征图根据通道数量均等分为三组，重新分组后的特征图参数为  $n \times n \times \frac{c}{3}$ ，之后采用  $k \times k \times 1$  大小的卷积核将特征图的通道数压缩为 1，最后将压缩得到的特征图进行合并，得到参数大小为  $n \times n \times 3$  的特征图。而之后在第二部分中主要进行权值计

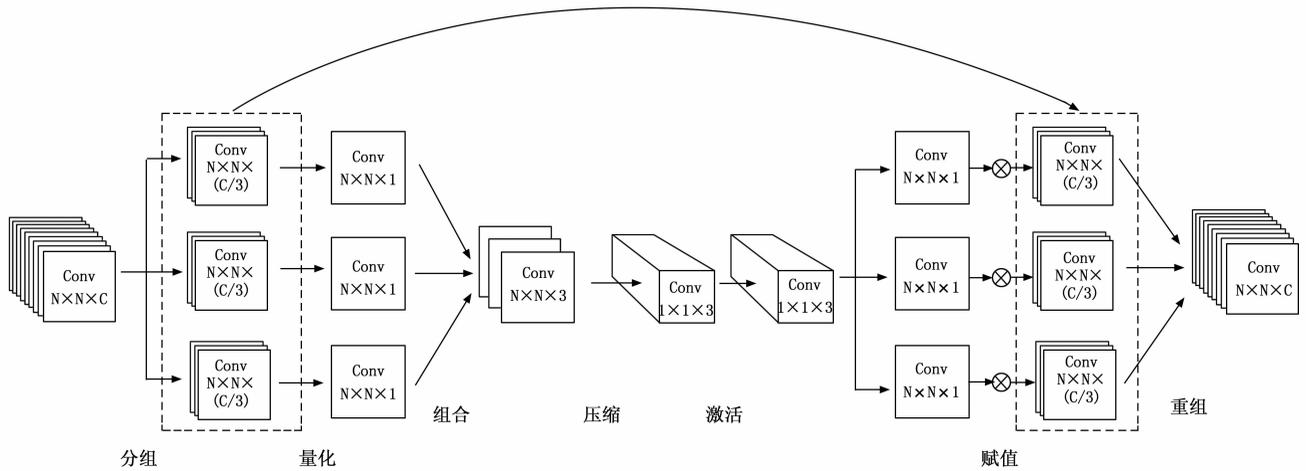


图 5 通道注意力机制映射模块

算工作, 首先将  $n \times n \times 3$  大小的特征图经全连接层压缩为  $1 \times 1 \times 3$ , 并采用 sigmoid 激活函数计算每个通道的权值, 之后将计算后得到的权值分别与第一部分中分组后特征图相乘赋值, 最后再将三组特征图重新组合, 得到通道注意力加权的特征图并进行传递。

### 2.2 Transformer 模块设计

Transformer 模块的引入主要为补充网络对于全局信息不够敏感的弊端。卷积模块侧重于对局部特征进行提取, 如果依靠卷积模块获取全局特征需要通过多层网络迭代缩小特征图, 使得在相同感受野大小下能够覆盖更多特征图区域<sup>[16]</sup>。而采用 Transformer 模块, 如第一小结所述其能够采用少量的层结构构建特征图的全局特征信息。但由于标准的 Transformer 模块需要对特征图划分的每一块特征区域进行关联性计算, 需要消耗较大的计算资源, 不利于网络的轻量化构建<sup>[17]</sup>。为解决标准 Transformer 模块计算复杂, 参数量多的问题, 本文采用了稀疏 Transformer 模块进行构建。

稀疏 Transformer 模块的功能实现主要由“局部—全局—局部”三个子模块组成实现, 先将局部相似特征进行整合, 之后将整合的特征块进行全局关联性计算, 最后在将计算结果重新分布的局部特征块中。每个稀疏模块的结构如图 6 所示。

如图 6 所示, 为稀疏 Transformer 模块的结构及特征图在其对数据的计算过程: 首先输入的特征图先进行 Patch Embed 块分割操作, 将特征图分为多个大小为  $K \times K$  的块区域, 完成分割后会由局部特征聚合模块对每个块区域的特征进行聚合, 该过程如图中 1 号虚线框所示, 特征聚合模块采用了  $K \times K$  大小的深度可分离卷积进行构建, 其将特征块聚合为一个像素点进行表示, 即该点代表了块特征(以上操作将特征图的大小进行  $K \times K$  倍的缩小, 能够减小后续自注意力计算的参数量及计算量)。之后由前馈网络将这些特征点进行组合, 并再次进行 Patch Embed 操作, 将每个特征点重新分割并通过编码转换为一维向量以进行

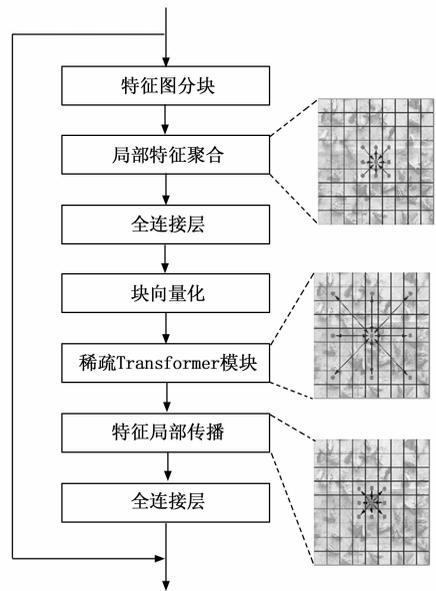


图 6 稀疏 Transformer 模块

Transformer 自注意力计算。对于每个全局注意力权值的计算, 本文采用了稀疏注意力的计算方式, 该方式如图中 2 号虚线框所示, 原特征图在每个  $K \times K$  大小的区域内由一个特征点进行稀疏表示, 之后仅对这些特征点进行自注意力计算, 获取点与点之间的特征关联度。最后通过局部传播模块将获取到的注意力权值进行转发, 使得该权值覆盖整张特征图。局部传播的实现为通过转置卷积将具有注意力权值的特征点转换为特征块, 即每个特征块共享一个注意力权值, 转置后即可获得每个块与块之间的关联度, 在通过映射的方式将输入稀疏 Transformer 模块的特征图与转置后的特征图进行融合以获得空间层面上的注意力权值。通过这种方式能够在保证获取全局特征的同时, 有效降低网络的计算量, 并且可以通过控制分割特征图区域大小的  $K$  值决定参数的减小倍数, 使其具有灵活性。

综上, 网络整体的检测流程为输入的图片首先由卷积

模块的编码器进行目标形态特征的提取及特征图的快速构建；之后由稀疏 Transformer 模块组对特征图进行空间层面上的全局自注意力计算，获取目标在特征图中的分布关系。完成之后输入到卷积模块的解码器中，将特征图进行还原，同时在卷积模块的编码器会将其输出的特征图通过通道注意力机制映射至解码器中，将低级形态特征与高级语义特征进行融合，提高了特征图的丰富性。最后，本文采用了目前单阶段网络中检测级精度较高的 YOLO 检测器对输出的特征图进行检测，获取目标在图中的数量及位置。由于网络的设计方案以轻量化为目的，且结合当下热门的 Transformer 多头自注意力机制，并采用 YOLO 检测器进行检测，因此将优化后的网络命名为 YOLO TLite 以便后续实验进行区分。

### 3 实验及数据分析

#### 3.1 数据采集

为验证本文所构建网络适用于烟草仓储间的虫情检测任务，本文通过对仓储间内的真实虫情环境进行拍照采样，制作了烟草粉螟、烟草甲两类虫害的数据集，所采样图片如图 7 所示。

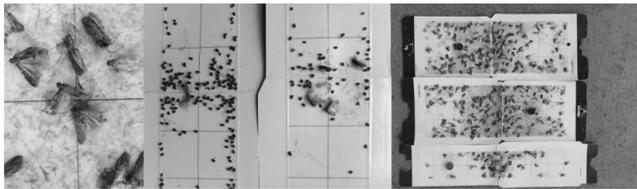


图 7 采样示例图

对于数据集的样本制作，本文通过拍摄诱虫板上的虫害为目标样本，为提升模型的泛化性与增强其鲁棒性，在拍摄的过程中，本文选择了不同虫害密度的诱虫板已经通过控制拍摄的距离获取不同大小的样本目标。经统计，本文共采集图片 1 638 张，烟草粉螟样本数 8 718 个、烟草甲虫样本数 6 539 个。

#### 3.2 评价指标

对于模型性能的评估，本文结合混淆矩阵采用以下量化指标进行客观评估：

混淆矩阵中共有  $TP$ 、 $TN$ 、 $FP$ 、 $FN$  四个参数，每个参数所代表的含义分别如下：

- $TP$ ：对于正样本，模型预测正确的个数；
- $TN$ ：对于负样本，模型预测正确的个数；
- $FP$ ：对于负样本，模型错误预测为正样本的个数；
- $FN$ ：对于正样本，模型错误预测为负样本的个数。

精确率 (Precision)：评估模型对于目标类别的分类精度；

$$Precision = \frac{TP}{TP + FP}$$

召回率 (Recall)：评估模型对于目标总数的查全能力；

$$Recall = \frac{TP}{TP + FN}$$

$F_1$  综合得分：评估模型在均衡精确率与召回率情况下的综合性能。

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

检测精度均值 (average precision)：评估某一类别的检测置信度；

$$AP = \int_0^1 P_{smooth}(r) dr$$

平均检测精度均值 (mean average precision)：评估模型的综合检测置信度；

$$MAP = \frac{\sum_{i=1}^n AP_i}{n}$$

模型复杂度 (FLOPs)：模型浮点运算计算量，评估模型复杂度；

检测速度 (FPS)：帧率，评估模型每秒能够检测图片的数量。

#### 3.3 数据分析

对于模型性能的分析，本文以目前在轻量化网络领域中应用较为常用的 YOLO v5s 为参照对比模型，并将本文对于主干网络的优化方法分别移植至 YOLO v5s 上进行消融实验；并且为验证本文所设计稀疏 Transform 模块在该任务中的适用性，本文选取了标准 Transformer 模块与之进行对比。首先对网络的基础性能进行比较分析，对比结果如表 1 所示。

表 1 基础性能对比

	FPN_Attention	Sparse Transformer	Vision Transformer	Precision	Recall	$F_1$ -Score
模型 1				0.835	0.945	0.887
模型 2	✓			0.862	0.963	0.910
模型 3		✓		0.868	0.958	0.911
模型 4			✓	0.862	0.956	0.907
模型 5	✓	✓		0.883	0.981	0.929
模型 6	✓		✓	0.880	0.979	0.927

注：模型 1 为 YOLO v5s 网络；模型 2 为在 YOLO v5s 的网络结构上加入了通道注意力映射；模型 3 为将 YOLO v5s 的主干网络原卷积模块采用稀疏 Transformer 模块替代；模型 4 为将 YOLO v5s 的主干网络原卷积模块采用标准 Transformer 模块替代；模型 5 为采用了通道注意力映射与稀疏 Transformer 模块像结合的 YOLO TLite 网络；模型 6 为在模型 5 基础上替换其中稀疏 Transformer 模块为标准 Transformer 模块网络。

如表 1 所示，在模型 1 的测试结果中，YOLO v5s 模型对于目标辨别的精确率较低，分析其原因主要为网络结构较为精简，无法对目标特征记性充分提取造成。而模型 2、模型 3 与模型 4 相比较与模型 1 YOLO v5s 均有一定的提升，分析其原因为模型 2 采用了 FPN 结合通道注意力机制进行高低特征融合，在网络深层融入低级形态特征；模型 3 与模型 4 采用多头自注意力机制 Transformer 模型能够对特征图的全局特征进行提取，与卷积模块所提取到的局部特征进行互补。但对于模型 4 采用的标准 Transformer 模块与

模型 3 所采用本文所设计的稀疏 Transformer 模块相比, 在检测精度与召回率赏要略低于模型 3, 分析其原因主要由于标准 Transformer 模块对于特征图的块区域分割为等分, 各个块区域无关联, 对于每个块区域的特征关联性相较于稀疏 Transformer 模块的计算方式更低, 使得所提取到的特征信息在全局性上与模型 3 相比有所欠缺。模型 5 与模型 6 采用了本文所设计的网络结构, 两个模型的不同主要在于所采用的 Transformer 模块不同, 结合模型 3 与模型 4 的对比分析, 模型 6 采用标准 Transformer 模块与模型 5 相比, 在检测性能上也略低于模型 5。以上这 5 个网络均能提升网络特征的丰富度以提升分类精度。而对于召回率, 由于诱虫板的背景为白色, 虫害为黑色有较大的色差易于区分前景与背景, 故各网络均有出色表现。综上模型 5 YOLO TLite 网络结合了通道注意力机制与稀疏 Transformer 模块, 在分类精度上相比于其他网络都有明显提升, 在  $F_1$  综合得分指标也反映了 YOLO TLite 的检测性能更优。

在表 2 的检测精度对比中, 同表 1 的数据相关联, 特征的丰富性与全面性同样对检测精度有一定的影响。对比结果验证了本文所设计的模型 5 YOLO TLite 网络在检测精度上同样优于其他网络模型。

表 2 检测精度对比

	FPN_Attention	Sparse Transformer	Vision Transformer	AP (粉螟)	AP (甲虫)	mAP
模型 1				0.896	0.935	0.915
模型 2	✓			0.920	0.951	0.936
模型 3		✓		0.916	0.962	0.939
模型 4			✓	0.915	0.960	0.937
模型 5	✓	✓		0.936	0.978	0.957
模型 6	✓		✓	0.932	0.974	0.953

在表 3 中, 主要对模型的轻量化性能进行比较。FLOPs 指标为客观评估指标, FPS 数据则与模型搭载的硬件参数相关, 本文选用了两种硬件环境进行对比测试。基于 FLOPs 指标, 模型 2 由于在原网络的基础上引入了轻量化通道注意力机制, 因此其计算量相比于模型 1 YOLO v5s 有一定的增加; 而模型 3 采用稀疏 Transformer 模块对原网络残差卷积模块进行替换, 且模块使用数量与原先相比有所减少, 因此在计算量上低于 YOLO v5s。但模型 4 采用标准 Transformer 模块, 由于 Transformer 模块计算的复杂度, 使得模型的计算量要高于模型 1, 在计算速度同样低于模型 1, 结合表 1 与表 2 的检测性能对比结果, 虽然标准 Transformer 模块与稀疏 Transformer 模块精度相似, 但在计算量与检测速度上有明显差距。在模型 5 YOLO TLite 网络结构中, 其另外将检测器的数量由 3 个改为 1 个, 同时结合轻量化通道注意力机制与稀疏 Transformer 模块, 因此其计算量整体也要低于 YOLO v5s。反之模型 6 采用了标准 Transformer 模块, 虽然网络做了一定的精简, 但由于模块计算的复杂度使得计算量仍要高于 YOLO v5s。实际检测

速度对比, 以上网络在搭载由 GPU 的硬件设备中均能达到实时检测效果, 而在 CPU 上也能满足实时检测。

表 3 检测速度对比

	FPN_Attention	Sparse Transformer	Vision Transformer	FLOPs	FPS (CPU)	FPS (GPU)
模型 1				7 015 519	7	46
模型 2	✓			8 237 843	5	37
模型 3		✓		5 973 115	8	48
模型 4			✓	9 625 482	3	31
模型 5	✓	✓		6 375 403	8	50
模型 6	✓		✓	1 125 358	2	28

注: CPU 硬件环境为 Intel i5-11400F; GPU 硬件环境为 Intel i5-11400F+RTX 2070。

以上消融实验证明了相比于优化前的 YOLO 网络, 本文所设计的 YOLO TLite 网络在仓储烟虫检测中更具有优势。并且所设计的稀疏 Transformer 模块与标准 Transformer 模块相比也更适用于该场景, 在具有高检测精度的同时, 兼具了轻量化的特性。

为进一步验证 YOLO TLite 在该任务中的通用性, 本文选用了目前主流的目标检测算法 PANet、Cascade RCNN、Detection Transformer 及 Swin Transformer 在本文所制作数据集上进行训练测试, 并与 YOLO TLite 进行对比, 对比结果如表 4 所示。

表 4 多类模型对比测试

	$F_1$ -Score	mAP	FPS (CPU)	FPS (GPU)
YOLO TLite	0.929	0.957	8.8	50.2
SENet	0.898	0.924	6.4	37.6
CBAM Net	0.917	0.942	4.7	32.4
Detection Transformer	0.884	0.920	1.2	21.0
Swin Transformer	0.912	0.944	0.95	19.4

在表 4 中, SENet 与 CBAM Net 为卷积神经网络, Detection Transformer 与 Swin Transformer 为 Transformer 模型网络。测试结果 CBAM Net 因同时采用了空间加权注意力机制与通道加权注意力机制提高了模型对于目标的感知能力、Swin Transformer 网络通过滑动分割计算的方式提高了 Transformer 模型中每个 token 的交互性, 这两个网络模型在该任务中同样具有较好的鲁棒性, 但与本文的 YOLO TLite 相比, 在检测性能与检测速度上能仍有一定的差距, 特别是在检测速度上, 由于 Transformer 模型的计算较为复杂整体的计算量高, 使得模型在检测速度上较慢。而对于 SENet 与 Detection Transformer 网络在网络性能上相比与其他网络均较有一定差距, 主要体现在 SENet 仅采用了通道加权注意力机制, 对于目标在特征图空间上的分布定位感知能力较弱, 而 Detection Transformer 则缺乏了特征图分割后 token 与 token 间的关联性。综上, YOLO TLite 综合了卷积模型与 Transformer 模型的特点, 在保证

检测精度的同时，仍具有较快的检测速度。同时 YOLO TLite 模型对于不同距离、密度场景下的实际检测结果如图 8 所示，均能够将图中的虫害目标正确框选。

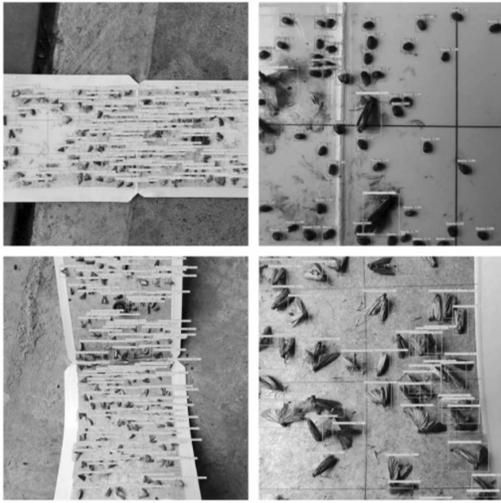


图 8 实际检测效果图

综上实验，验证了本文所设计的 YOLO TLite 在烟草仓储间虫害检测任务中兼具有较高的检测精度与速度，适用于虫情监测模型中。对于监测模型的实现方案，本文主要通过计算虫害密度进行预警模型构建，通过在仓储间内铺设多个 0.5 的诱捕区，使用 YOLO TLite 模型计算区域内虫害个数，以得出相应密度等级。为验证模型的实用性，本文在区域内设置了不同虫害个数以及在不同的外界光照环境下进行检测实验，测试结果分别如图 9 所示。

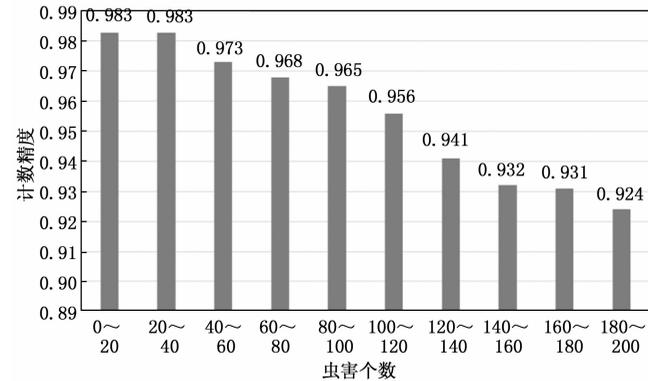


图 9 标准光照条件计数准确性检测

首先图 9 的测试结果来源于正常光照条件下的检测结果。图中，对相同区域面积大小下，不同虫害个数的平均计数精度对比结果上看，当虫害个数低于 50 时，模型的计数精度较高；而虫害个数在 120 只以上时，密度较高，使得模型的计数精度有一定幅度的下降。但虫害个数在 200 只左右时，模型的计数精度仍有 0.924，网络模型在整体鲁棒性强，能够用于统计虫数。

其次，本文模拟了在夜间或照明条件不充足情况下的检测结果，以证明模型在不同环境下对于不同密度的虫害

仍具有出色的检测性能，如图 10 所示。由于当光照条件不够充足时，画面会出现噪点，且虫害自身呈现灰黑色，因此检测结果相比于标准光照条件下略微有点下降，其检测精度的“拐点”出现于虫害个数为 80 只左右，但当虫害个数大于 180 只时检测精度仍达 0.90 以上，说明模型能够应付光照条件不足等复杂环境下的检测任务。

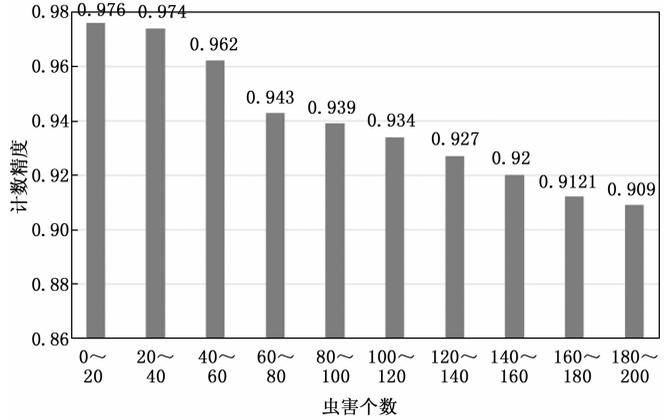


图 10 弱光条件计数准确性检测

#### 4 结束语

本文针对于烟草行业仓储环节中虫情检测难的问题，提出了一种基于深度学习的解决方案。该方案主要依靠深度学习网络计算单位面积内的虫害个数，进而换算出密度以得到相应虫情数据。为此本文主要基于 Transformer 模块优化设计了 YOLO TLite 轻量化检测网络，通过轻量化的结构设计，同时增强主干网络的特征提取能力，以提高检测精度。经实验验证，本文所设计的网络具有较强的鲁棒性，在不同密度环境下均能以较高的精度计算出虫害个数，具有一定的使用性能。

#### 参考文献：

- [1] 罗云, 陈斌. 烟叶仓储害虫综合治理研究进展 [J]. 昆明学院学报, 2021, 43 (6): 19-23.
- [2] 梁伟, 刘鸿. 烟叶仓储害虫绿色防控技术现状与发展趋势 [J]. 天津农业科学, 2021, 27 (7): 52-55, 60.
- [3] 张竞超, 翟乃琦, 王一博, 等. 基于物联网技术的仓储烟叶霉变状态智能监测方法研究 [J]. 湖北农业科学, 2021, 60 (20): 167-170, 178.
- [4] 李子茂, 刘恋冬, 夏梦, 等. 基于深度学习的月季多叶片病虫害检测研究 [J]. 中国农机化学报, 2021, 42 (8): 169-176.
- [5] 王金, 李颜娥, 冯海林, 等. 基于改进的 Faster R-CNN 的小目标储粮害虫检测研究 [J]. 中国粮油学报, 2021 (9): 164-171.
- [6] 曾伟辉, 唐欣, 胡根生, 等. 基于卷积块注意力胶囊网络的小样本水稻害虫识别 [J]. 中国农业大学学报, 2022, 27 (3): 63-74.

(下转第 89 页)