doi:10.6041/j.issn.1000-1298.2024.04.015

基于 FE－P2Pnet 的无人机小麦图像麦穗计数方法

鲍文霞　苏彪彪　胡根生　黄承沛　梁　栋

（安徽大学农业生态大数据分析与应用技术国家地方联合工程研究中心，合肥 230601）

摘要：针对无人机图像背景复杂、小麦密集、麦穗目标较小以及麦穗尺寸不一等问题，提出了一种基于 FE－P2Pnet（Feature enhance－point to point）的无人机小麦图像麦穗自动计数方法。对无人机图像进行亮度和对比度增强，增大麦穗目标与背景之间的差异度，减少叶、秆等复杂背景因素的影响。引入了基于点标注的网络 P2Pnet 作为基线网络，以解决麦穗密集的问题。同时，针对麦穗目标小引起的特征信息较少的问题，在 P2Pnet 的主干网络 VGG16 中添加了 Triplet 模块，将 C（通道）、H（高度）和 W（宽度）3 个维度的信息交互，使得主干网络可以提取更多与目标相关的特征信息；针对麦穗尺寸不一的问题，在 FPN（Feature pyramid networks）上增加了 FEM（Feature enhancement module）和 SE（Squeeze excitation）模块，使得该模块能够更好地处理特征信息和融合多尺度信息；为了更好地对目标进行分类，使用 Focal Loss 损失函数代替交叉熵损失函数，该损失函数可以对背景和目标的特征信息进行不同的权重加权，进一步突出特征。实验结果表明，在本文所构建的无人机小麦图像数据集（Wheat－ZWF）上，麦穗计数的平均绝对误差（MAE）、均方误差（MSE）和平均精确度（ACC）分别达到 3.77、5.13 和 90.87%，相较于其他目标计数回归方法如 MCNN（Multi-column convolutional neural network）、CSRnet（Congested scene recognition network）和 WHCNETs（Wheat head counting networks）等，表现最佳。与基线网络 P2Pnet 相比，MAE 和 MSE 分别降低 23.2% 和 16.6%，ACC 提高 2.67 个百分点。为了进一步验证本文算法的有效性，对采集的其它 4 种不同品种的小麦（AK1009、AK1401、AK1706 和 YKM222）进行了实验，实验结果显示，麦穗计数 MAE 和 MSE 平均为 5.10 和 6.17，ACC 也达到 89.69%，表明本文提出的模型具有较好的泛化性能。

关键词：麦穗计数；无人机图像；FE－P2Pnet；FEM；Wheat－ZWF 数据集

中图分类号：S127　　**文献标识码**：A　　**文章编号**：1000-1298(2024)04-0155-10　　**OSID**：

Method for Counting Wheat Ears in UAV Images Based on FE－P2Pnet

BAO Wenxia　SU Biaobiao　HU Gensheng　HUANG Chengpei　LIANG Dong

（*National and Local Joint Engineering Research Center for Agricultural Ecological Big Data Analysis and Application Technology, Anhui University, Hefei* 230601, *China*）

Abstract: Ear count is the committed step of wheat yield estimation. With the rapid development of unmanned aerial vehicle (UAV) and computer vision technology, the problem of automatic counting of wheat ears can be solved more quickly and efficiently. An automatic counting method for UAV wheat ear images was proposed based on feature enhance－point to point (FE－P2Pnet) to address issues such as complex background, dense wheat, small wheat ear targets, and varying wheat ear sizes. Firstly, the brightness and contrast of the UAV image were enhanced to increase the difference between the wheat ear target and the background, and the influence of complex background factors such as leaves and stems were reduced. Secondly, a point annotated network P2Pnet was introduced as the baseline network to address the problem of dense wheat ears. At the same time, in response to the problem of limited feature information caused by small wheat ear targets, a Triplet module was added to the backbone network VGG16 of P2Pnet, which interacted with the information of C (channel), H (height), and W (width) dimensions, allowing the backbone network to extract more feature information related to the target. In

收稿日期：2023－08－16　修回日期：2023－09－16

基金项目：安徽省自然科学基金项目（2208085MC60）、安徽省科学技术厅高校科研计划项目（2023AH050084）和国家自然科学基金项目（62273001、32372632）

作者简介：鲍文霞（1980—），女，教授，博士，主要从事农业生态视觉分析与模式识别研究，E-mail：bwxia@ahu.edu.cn

通信作者：梁栋（1963—），男，教授，博士，主要从事农业生态大数据分析与应用研究，E-mail：dliang@ahu.edu.cn

response to the issue of varying wheat ear sizes, feature enhancement module (FEM) and squeeze excitation (SE) modules were added to feature pyramid networks (FPN), enabling this module to better process feature information and fuse multi-scale information. In order to better classify targets, Focal Loss function instead of cross entropy loss function was used. This loss function can carry out different weights on the background and target feature information to further highlight features. The experimental results showed that the mean absolute error (MAE), mean square error (MSE), and accuracy (ACC) indicators of wheat ear counting on the constructed unmanned aerial vehicle wheat image dataset (Wheat-ZWF) achieved 3.77, 5.13, and 90.87%, respectively. Compared with other target counting regression methods such as MCNN, CSRnet, and WHCNETs, the performance was the best. Compared with the baseline network P2Pnet, the MAE and MSE values were decreased by 23.2% and 16.6% respectively, and the ACC value was increased by 2.67 percentage points. In order to further validate the effectiveness of the algorithm proposed, experiments were conducted on four other different wheat varieties (AK1009, AK1401, AK1706, and YKM222) collected. The experimental results showed that the average MAE and MSE values of wheat ear counting were 5.10 and 6.17, with ACC of 89.69%. This indicated that the proposed model had good generalization performance. The research can provide certain support and assistance for related studies on wheat ear counting.

Key words: ear count of wheat; UAV images; FE-P2Pnet; FEM; Wheat-ZWF dataset

0 引言

小麦是我国传统主食之一,也是世界上种植最广泛、种植面积最大的粮食作物之一[1]。因此小麦的产量预估问题成为了人们关注的焦点之一。田间小麦产量主要与单位面积麦穗数量、单位面积谷物数量和千粒质量有关[2]。在这些特征中,单位面积内的麦穗数量是最重要的指标[3]。单位面积内的麦穗数和穗粒数通过人工得到,然后计算出小麦产量。麦穗传统人工计数方法存在主观性强、效率低和缺乏统一统计标准等缺点。

随着计算机视觉技术的发展,基于图像处理的计数方法逐渐被用到小麦麦穗自动计数中,如 COINTAULT 等[4]对麦穗图像进行同现矩阵统计分析和傅里叶滤波处理,提取图像中麦穗的纹理和颜色特征最后获得麦穗数量;ALKHUDAYDI 等[5]提出了一种基于天顶彩色数字图像的麦穗计数算法,预测精度超过 90%;SADEGHI-TEHRAN 等[6]通过使用简单线性迭代聚类(SLIC)将麦穗图像分割成超像素,导出冠层的相关特征,解决了小麦麦穗的量化问题;ZHOU 等[7]利用半自主传感器的田间表型平台获取小麦图像,使用最大熵分割方法进行粗分割处理;FERNANDEZ-GALLEGO 等[8]采用了寻找最大值法以及滤波法对麦穗进行检测操作。近些年来深度学习逐渐被应用于麦穗计数领域中,如 MADEC 等[9]利用 R-CNN 对麦穗进行检测,从而实现麦穗计数;YANG 等[10]提出了一种用于麦穗检测的 CBAM-YOLO v4 计数方法,该方法结合了卷积神经网络和注意力机制技术;MISRA 等[11]提出一种 SpikeSegNet 网络,将局部补丁提取网络和全局掩码细化网络相结合,实现麦穗的识别与计数;鲍文霞等[12]提出了 WEDEnet(Wheat ears density estimation network),利用密度图估计的方法,提出了一种基于单幅地面图像小麦麦穗计数的方法;GUO[13]提出了 WHCNets,构建了高质量的密度图用于展示麦穗在图像中的分布,提高了麦穗计数效果。

虽然国内外学者在基于计算机视觉技术的小麦自动计数方面进行了广泛探索,并取得了一定研究进展。然而,大多数研究聚焦于地面图像的分析。随着无人机技术的成熟和广泛应用,利用无人机进行数据采集具有便捷高效和覆盖范围广的优势,有利于提升整体小麦产量预估精度。

目前关于无人机麦穗计数方面的研究主要集中在利用深度学习中目标检测和回归的方法实现麦穗计数。在基于检测的麦穗计数方面,鲍文霞等[14]提出了一种 TPH-YOLO(YOLO with transformer prediction heads)的麦穗检测模型,提高了无人机图像麦穗计数的精度;高姻燕等[15]提出了一种基于 YOLO v3 的麦穗计数方法,并且构建了单位面积麦穗估测模型。基于回归的麦穗计数方法由于采用点标注,相比于基于检测的方法更加适合背景复杂和目标密集的无人机小麦图像的麦穗计数,如 MA 等[16]提出了一种基于地面的全卷积网络(EarDensityNet)迁移学习方法,实现了无人机麦穗计数;孙俊等[17]在 CSRnet 基础上根据小麦图像特点构建了普适麦穗计数网络,可以将目标特征与上下文语义特征进行融合,充分提取麦穗的特征信息,更好地完成了麦穗计数任务;为了追求进一步网络结构简化,YANG 等[18]提出了一个轻量化的密度图计数网络 LWDNet(Lightweight wheat ear

density map regression counting network)，采用协同注意力机制 CA(Collaborative attention)改进了轻量化网络，并且利用上采样模块 CARAFE 进一步融合了特征，最后在不同的数据集上都取得了良好的精度。以上这些基于回归的无人机图像麦穗计数方法在一定程度上提升了计数的精度，但由于这些回归方法采用的是通过密度图方式实现计数的，而真值密度图的生成过程存在假设麦穗相对于图像平面是独立存在的。事实上，由于透视畸变的存在，不同位置麦穗对应着不同尺寸的像素区域，因此，要想生成精确的麦穗密度图像，就要考虑单应性引起的畸变，但是畸变参数是不容易得到的。该问题在人群计数中同样存在，SONG 等[19]提出了一种基于点标注的 P2Pnet 网络用于直接回归出人群的数目从而减小该问题的影响。因此，将人群计数的 P2Pnet 网络引入到无人机图像麦穗计数问题中，并且针对无人机图像背景复杂、麦穗目标小及麦穗尺寸不一等问题，设计一种 FE－P2Pnet 网络实现麦穗的精确计数。

针对小麦麦穗目标较小问题，本文在 P2Pnet 的主干网络中添加 Triplet[20]模块，使通道、高度和宽度 3 个维度的信息相互交互，进一步提取目标的特征信息，解决目标小带来的特征信息不足的问题。为了解决不同麦穗尺度的问题，设计 SE－FPN 模块。该模块将 SE[21]模块和 FEM 模块添加到 FPN 中，FEM 模块通过全局和局部信息的融合保留更多有效特征信息，而 SE 模块通过对不同通道权重的选择和加权增强各层的特征信息。FEM 和 SE 模块结合进一步增强 FPN 在解决多尺度问题方面的能力。在分类损失部分，采用 Focal Loss[22]代替交叉熵损失。该损失函数对简单样本和困难样本进行加权处理，使得困难样本更容易被分类，从而在预测阶段对目标进行更好的分类。

针对该环境下背景复杂的问题，在数据预处理阶段对数据进行亮度和对比度增强。针对麦穗目标密集、小及尺寸不一等问题，设计 FE－P2Pnet 网络实现无人机图像麦穗计数。

1　数据获取和预处理

1.1　数据获取

本文研究区域为安徽省合肥市农业科学研究院(31°09′N，117°25′E)的小麦种植区域，该种植区域小麦品种为中原丰 1 号(ZWF1)，数据采集区域如图 1 所示。数据采集时间是 2022 年 5 月 9 日 09:00—14:00，小麦在这个时间段处于灌浆期，麦穗比较明显，利于数据采集。



图 1　数据采集区域

Fig. 1　Data collection area

实验的采集设备为御 Mavic Air 2 型大疆无人机，见图 2，该无人机参数如表 1 所示。



图 2　采集设备

Fig. 2　Collecting device

表 1　无人机数据参数

Tab. 1　UAV data parameters

参数	数值
摄像头尺寸/寸	1/2
有效像素数	4.8×10^7
等效焦距/mm	24
最大图像分辨率/(像素×像素)	8 000×6 000
最大飞行时间/min	34
最大飞行速度/(km·h^{-1})	68.4
最大飞行高度/m	5 000
最大飞行距离/km	18.5

实验采集到的无人机图像分辨率为 4 000 像素×3 000 像素，无人机高度为 2 m，摄像头拍摄视角垂直向下。获取的无人机图像如图 3 所示。



图 3　无人机采集小麦麦穗示例图像

Fig. 3　Example of UAV wheat ear image

1.2　数据处理

1.2.1　数据裁剪

为了满足后续模型以及数据集数据多样化需要，将无人机小麦图像分别裁剪为 800 像素×

600 像素、500 像素 × 500 像素和 400 像素 × 300 像素的小图,最终构建 Wheat - ZWF 数据集。该数据集共计包含 865 幅图像,并按照比例 7∶1.5∶1.5 划分为训练集、验证集和测试集。裁剪后图像如图 4 所示。

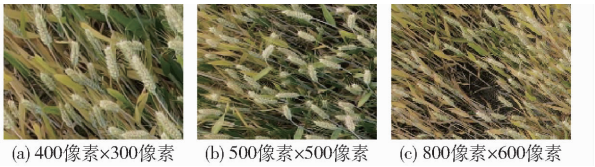


图 4 裁剪图像示例

Fig. 4 Crop image example

1.2.2 数据标注

本文的数据集是基于回归方法的点标注,在每个小麦麦穗的位置上标记一个点,每个点仅占据 6 个像素。相比于框标注方法,点标注方法不仅可以降低标注成本,还可以解决一部分小麦麦穗粘连和重叠的问题。图 5 展示了数据集标注的示例。

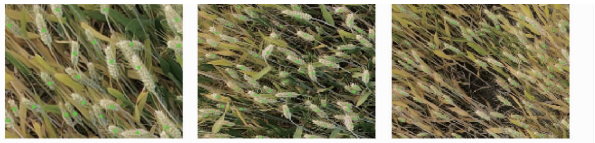


图 5 数据集标注示例

Fig. 5 Example of data annotation

1.2.3 数据增强

通过对图 4 中裁剪的小麦麦穗图像进行观察,可以发现小麦麦穗的颜色与小麦叶片和茎秆非常相似。为了减少背景对图像的影响,对小麦麦穗图像进行了亮度和对比度增强,以增加小麦麦穗与叶片、茎秆和其他背景之间的差异,从而使模型能够更好地提取目标特征。

本文采用的数据增强方法首先将输入的 RGB 彩色图像转换到 YCbCr 颜色空间下。其中,Y 表示亮度分量,采用自适应的对比度增强算法来增强 Y,最后将其从 YCbCr 颜色空间转换回 RGB 图像。自适应增强算法[23]的公式为

$$Y_{\text{enhanced}} = a(Y - Y_{\text{mean}})/\max(Y_{\text{std}}, \sigma) + \beta Y + (1 - a - \beta) Y_{\text{mean}} \tag{1}$$

式中 Y_{enhanced}——增强后分量

Y_{mean}——Y 分量平均值

Y_{std}——Y 分量标准差

a——对比度增益

σ——控制放大的程度

β——亮度偏移

式(1)的第 1 项可以看作是一种对亮度进行归一化操作,将亮度减去均值并除以标准差或 sigma,以消除亮度的不同尺度和范围,使得后续的增强操作更加稳定和准确。第 2 项是一个线性变换,用于将归一化后的亮度进行线性拉伸,增强亮度之间的差异。第 3 项是一种加权平均操作,用于保留原始亮度的部分信息和纹理细节。增强后图像如图 6 所示。

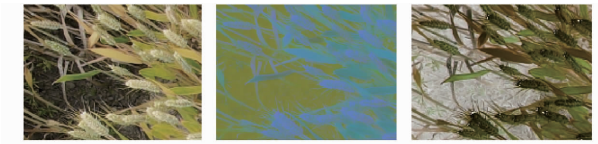


图 6 数据增强示例

Fig. 6 Data enhancement example

2 研究方法

2.1 FE - P2Pnet 模型

FE - P2Pnet 由 3 个主要模块组成:特征提取模块 VGG16 - Triplet、多尺度融合模块 SE - FPN 和预测模块。

特征提取模块 VGG16 - Triplet 主要功能是进行特征提取并将提取的特征传递给下个模块。针对无人机小麦麦穗图像中麦穗目标较小导致特征信息不足的问题,在 VGG16 中引入 Triplet 模块,该模块对通道(C)、高度(H)和宽度(W)3 个维度的信息进行交互,建立了维度之间的依赖关系。这样,在提取目标特征时可以考虑到更多的信息,对于解决小目标问题非常有帮助。

SE - FPN 模块主要任务是对前一模块中的特征进行多尺度融合,并将融合后的特征传递给预测模块。在 FPN 的基础上添加自己设计的 FEM 模块,并引入 SE 模块来解决无人机小麦图像中麦穗尺寸不一的问题。FEM 通过全局部分和局部部分的融合进一步增加了 FPN 各层信息的融合,而 SE 则有助于 FPN 各层获取更丰富的特征信息。SE - FPN 模块使得模型能够进一步提取更有效的特征,更加关注小麦麦穗,抑制和减小背景因素如麦秆和叶子的影响。

预测模块在分类部分采用 Focal Loss 来替代交叉熵损失。Focal Loss 降低了易分类样本的权重,同时增加了困难样本权重,使得模型能够更好地对目标进行分类,从而实现更好的结果。在回归预测点部分,使用均方差损失,进一步提高了小麦麦穗计数的准确性。

当图像输入到网络中后,首先进入特征提取模块 VGG16 - Triplet,提取到相应的特征后,将其送入 SE - FPN 模块中,经过进一步提取特征,最后送入预测模块进行小麦麦穗图像的预测。图 7 为网络结构整体图。

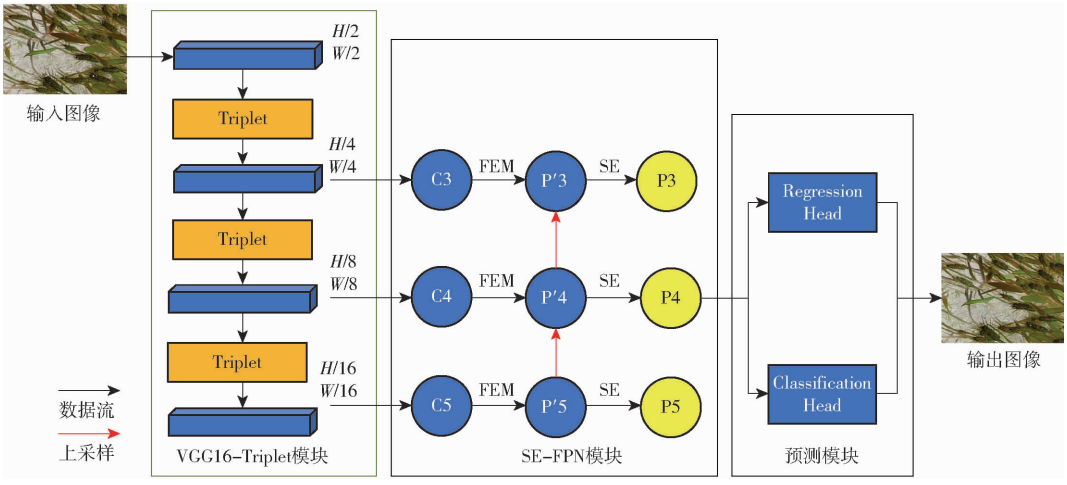


图 7　FE－P2Pnet 网络结构图

Fig. 7　FE－P2Pnet network architecture diagram

2.1.1　VGG16－Triplet 模块

本文提出的 VGG16－Triplet 模块主要针对无人机小麦麦穗目标较小的问题。通过在主干网络 VGG16 中添加 Triplet 模块，将 C、H 和 W 3 个维度的信息进行交互，使得主干网络能够提取更多相关的特征信息。Triplet 模块主要添加在 VGG16 中通道数变化的位置，例如 64～128、128～256 和 256～512 之间，如图 8 所示。这样添加的主要原因为：在通道数变化时，进一步关注特征的各个维度信息，从而确保通道变化前后的特征信息保持一致。

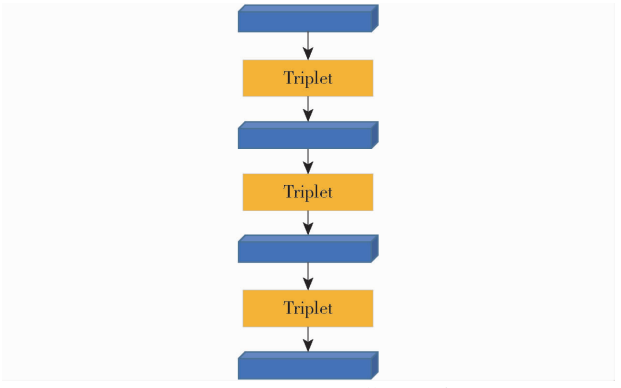


图 8　VGG16－Triplet 模块示意图

Fig. 8　VGG16－Triplet module diagram

如图 9 所示，Triplet 模块主要由 3 个平行的分支组成，其中 2 个负责捕获 C 和 H 或 W 之间的跨维交互。最后 1 个分支类似于 CBAM（Convolutional block attention module）在 H 和 W 之间建立交互。例如对于第 1 个分支，输入张量 $\boldsymbol{X}(C\times H\times W)$ 沿 H 轴逆时针旋转 90°。这个旋转张量 $\boldsymbol{X}$ 表示形状为 $W\times H\times C$ 再经过 Z－Pool 后的张量形状为 $2\times H\times C$；其次，通过内核大小为 7×7 的标准卷积层，再通过批处理归一化层，提供维数 $1\times H\times C$ 的中间输出。然后，将张量通过 Sigmoid 来生成注意力权值。最后再将得到的输出沿 H 轴进行顺时针旋转 90°保持和输入的形状一致。最终 3 个分支的输出使用平均加权进行聚合。Z－Pool 层负责将第 1 个维度的 Tensor 缩减到 2 维，将该维上的平均池化特征和最大池化特征连接起来。这使得该层能够保留实际张量的丰富表示，同时缩小其深度以使计算量更少。计算公式为

$$\boldsymbol{V}_{\text{Z-Pool}}(\boldsymbol{X})=[\boldsymbol{V}_{\text{MaxPool}}(\boldsymbol{X}),\boldsymbol{V}_{\text{AvgPool}}(\boldsymbol{X})] \tag{2}$$

式中　$\boldsymbol{V}_{\text{Z-Pool}}$——Z－Pool 后的向量输出

$\boldsymbol{V}_{\text{MaxPool}}$——最大池化特征

$\boldsymbol{V}_{\text{AvgPool}}$——平均池化特征

Triplet 模块，可以将 C、H 和 W 3 个维度的信息进行相互交互，增加一部分小目标的特征信息，可以使得特征提取模块更好地提取目标特征。

2.1.2　SE－FPN 模块

本文设计的 FEM 模块主要分为全局部分和局部部分。全局部分在 SE 模块的基础上增加了 BN（批量归一化）模块，以增强特征信息。SE 模块如图 10 所示。局部部分去除了平均池化模块，保留更多原始特征信息。该模块将全局部分和局部部分的输出进行加权相乘，得到增强后的特征。在输出之前，还引入了残差块，进一步抑制梯度消失和模型退化问题。最后，将输入与经过残差块处理的输出相加，得到经过 FEM 模块增强后的特征输出。网络结构如图 11 所示。

针对无人机图像中麦穗尺寸不一的问题，主要通过金字塔网络 FPN 进行改进。网络结构改进图如图 12 所示。在 FPN 的各层中添加了 FEM 模块和 SE 模块。FEM 模块主要添加在上采样之前，旨在保留原始特征信息的同时进一步选择和加权重要特征，并通过残差块平滑梯度损失。这样，在上采样过程中可以有更多特征信息进行交互，从

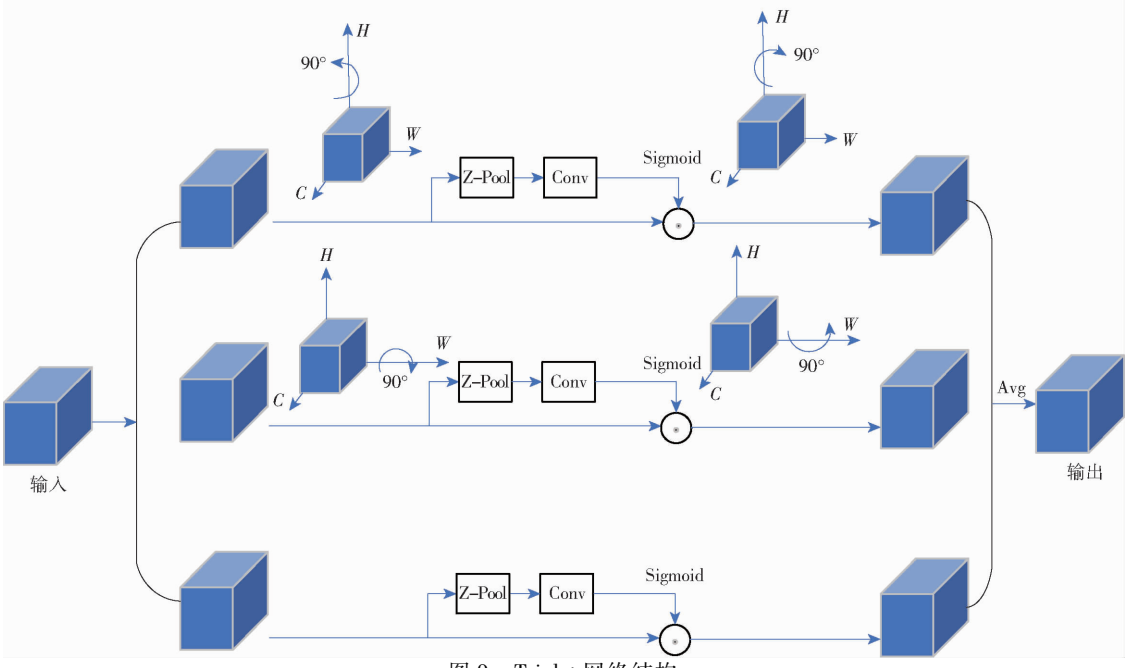


图 9 Triplet 网络结构

Fig. 9 Triplet module

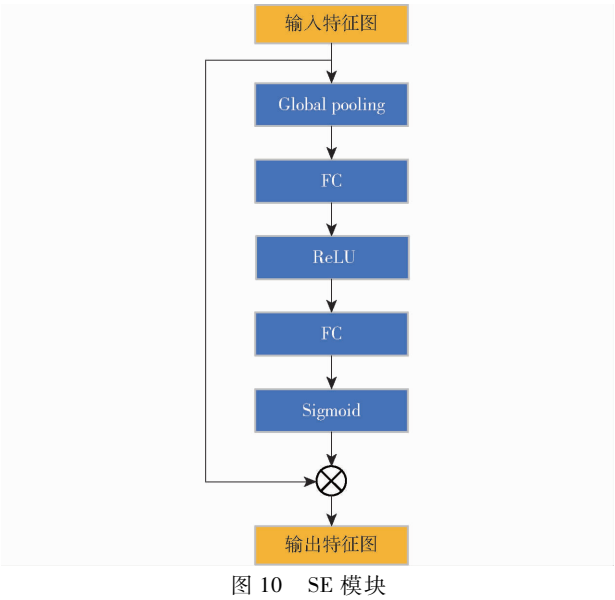


图 10 SE 模块

Fig. 10 SE module

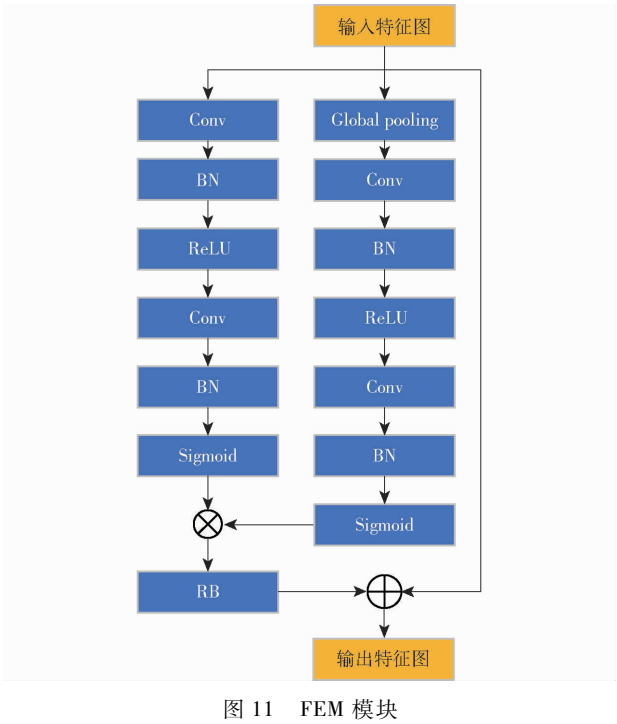


图 11 FEM 模块

Fig. 11 FEM module

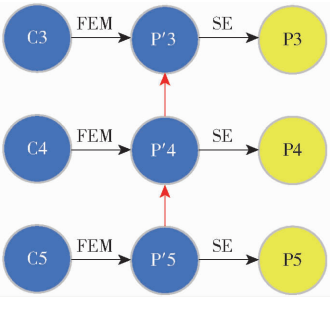


图 12 SE－FPN 模块

Fig. 12 SE－FPN module

而提高整个网络性能。SE 模块主要用于增强各层在输出之前的特征。它可以自适应地选择和加权不同通道的特征,从而提高网络的泛化能力和准确性。通过 SE－FPN 模块,可以更好地提取不同尺度麦穗的特征信息,从而增强网络的预测计数能力。

2.1.3 损失函数

本文预测模块主要使用 2 个损失函数,使用均方差损失 L_{res}来监督点回归,并使用 Focal Loss L_{cls}来训练建议分类。

均方差损失 L_{res}公式为

$$L_{res}=\frac{1}{M}\sum_{i=1}^{M}|x-x'|^2 \tag{3}$$

式中　x——真值　　x'——预测值

M——样本数

本文的目标就是麦穗，剩下的都是负样本，麦穗与其余背景的数量差别比较大，使得识别和计数比较困难，本文使用 Focal Loss，主要是为了解决正负样本数量不平衡以及难易分类样本不均衡的问题。Focal Loss 公式为

$$L_{\mathrm{cls}} = -\alpha(1-p_t)^{\gamma}\ln(p_t) \tag{4}$$

式中　p_t——模型预测某类别的概率

α——平衡正负样本数量的参数，正样本时 α 为数组的第 1 项，负样本时 α 为数组的第 2 项，取{0.45，0.55}

γ——调节难易样本的参数，取 1

总损失函数为

$$L = L_{\mathrm{res}} + L_{\mathrm{cls}} \tag{5}$$

2.2　方法流程

本文所提出的基于 FE－P2Pnet 的无人机小麦麦穗检测方法流程图如图 13 所示，具体为：

（1）在实际大田田间环境下获取无人机图像。

（2）对图像进行裁剪，分别裁剪为 400 像素×300 像素、500 像素×500 像素和 800 像素×600 像素 3 种不同尺寸的图像。

（3）对图像进行对比度增强处理。

（4）将数据集划分为训练集、验证集和测试集。

（5）将训练集图像输入到网络中进行模型训练，每经过 5 轮训练进行一次验证，并保存最佳模型权重。

（6）使用保存的模型权重对测试集图像进行准确度测试，并进行麦穗个数的预测。

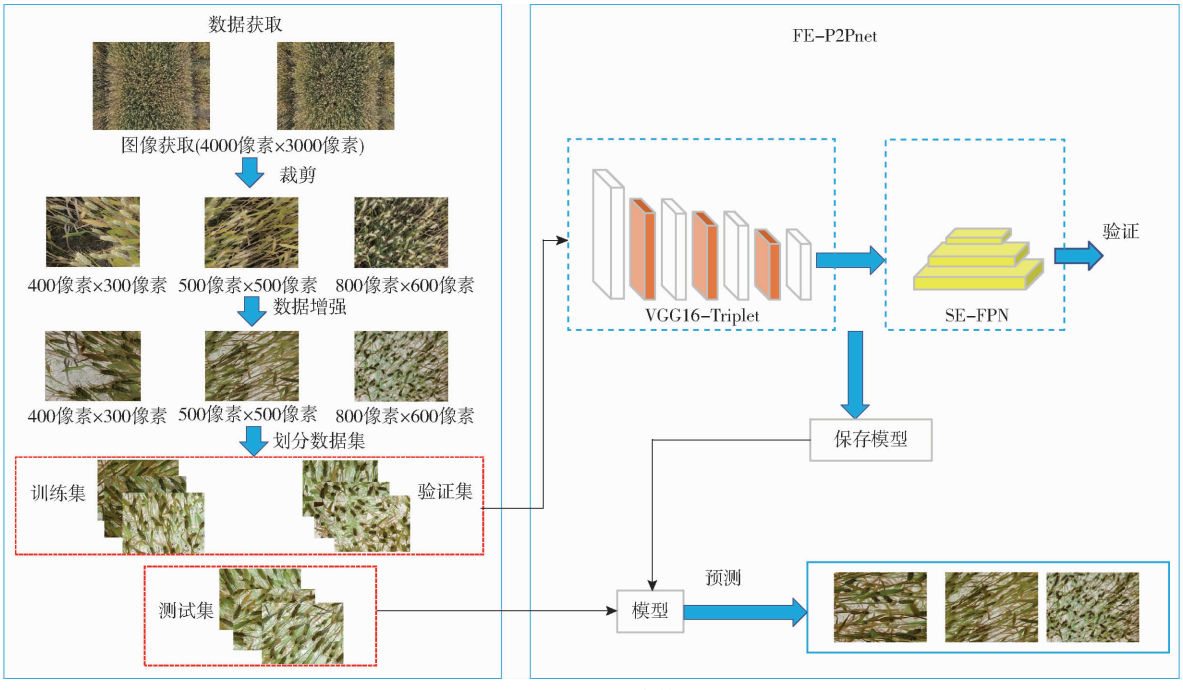


图 13　无人机小麦麦穗计数方法流程图

Fig. 13　Flowchart of UAV wheat ear counting method

2.3　评价指标

平均精确度（Accuracy，ACC）和平均绝对误差（Mean absolute error，MAE）能反映计数的准确性，均方误差（Mean squared error，MSE）可以反映计数方法的鲁棒性，常被用来量化计数性能[24]，MAE 和 MSE 越低，ACC 越高表示计数性能越好。本文使用这 3 个评价指标进行评判。假设有 N 幅测试图像，则 ACC 计算公式为

$$V_{\mathrm{ACC}} = 1 - \frac{1}{N}\sum_{i=1}^{N}\frac{|C_i^{GT} - C_i^{P}|}{C_i^{GT}} \times 100\% \tag{6}$$

式中　C_i^P——第 i 幅图像的输出麦穗估计个数

C_i^{GT}——第 i 幅图像的麦穗真实个数

V_{ACC}——平均精确度

3　实验结果及分析

本文的模型训练参数设置为：学习率为 0.000 1，迭代周期为 500，批量大小为 8，权重衰减 0.000 1，梯度裁剪参数为 0.1。实验的硬件条件为 Intel（R）Core（TM）i7－8700K CPU，NVIDIA GeForce GTX 1080，软件方面使用 Python 中的 pytorch 环境进行实验。

3.1　不同计数模型实验结果对比

主要选择目标计数回归领域的几种典型模型，包括 MCNN[25]、CSRnet[26]、WEDEnet、WHCNETs、

P2Pnet 模型以及本文提出的模型，进行比较评估。表 2 展示了不同模型在 Wheat－ZWF 数据集上的实验结果。可以观察到，MCNN、CSRnet、WEDEnet 和 WHCNETs 模型在该数据集上也取得了不错的计数效果，ACC 约 80% 左右。而本文的基线网络 P2Pnet 在该数据集上比上述模型更好，计数 ACC 也达到 88.90%，说明了该模型在无人机图像小麦麦穗计数方面具有良好的效果。本文提出的 FE－P2Pnet 取得了最佳性能，其 MAE 为 3.77，MSE 为 5.13，计数准确率达到 90.87%。对比实验结果验证了本文提出模型在无人机图像上进行小麦计数的有效性。

表 2 不同计数模型实验结果比较

Tab. 2 Comparison of experimental results using different counting models

模型	平均绝对误差	均方误差	平均精确度/%
MCNN	7.82	10.54	78.20
CSRnet	5.64	7.92	84.14
WEDEnet	5.25	7.28	87.53
WHCNETs	7.45	10.06	77.16
P2Pnet	4.50	5.88	88.90
本文模型	3.77	5.13	90.87

图 14 展示了不同计数模型的预测计数结果。基于密度图回归的模型如 MCNN、CSRnet、WEDEnet 和 WHCNETs 模型可以通过密度图大致显示出麦穗的位置，但是在计数方面的预测值与真实值存在一定差距，其主要原因是基于密度图的计数方法首先需要通过点标注图像生成真值密度图，将其作为标签进行训练。在这个过程中，不仅是透视畸变的问题，而且由于相邻麦穗距离太近，在进行高斯卷积生成密度图时部分区域可能会产生重叠，从而产生一定的误差，影响最后的计数结果，所以预测结果与真值有一定的偏差。而 P2Pnet 模型和本文提出的模型直接将点标注作为真值训练，最后直接回归预测点，避免了上述问题，所以预测值与真值几乎相同或非常接近，而且能够直观地预测和显示麦穗的位置。本文提出的 FE－P2Pnet 模型相较于 P2Pnet 模型，在网络结构方面进行了改进，解决了一部分误检和漏检的问题，取得了更好的计数结果。

3.2 消融实验

为了进一步验证本文提出模型的有效性，进行了消融实验，结果如表 3 所示。首先，对比使用数据增强和不使用数据增强的效果。通过数据增强，各项指标（如 MAE、MSE）都有提升，ACC 也提高 0.7 个百分点。这表明数据增强减少了背景对目标的影响。在其他消融实验中，对 FEM、SE、Focal Loss 和 Triplet 这 4 个模块分别进行了单独测试。结果显示，这些模块对模型性能都有一定提升，其中 Triplet 模块的提升效果较为显著，进一步说明了该模块对于增强主干网络的特征提取能力是有益的。当将 FEM、SE 和 Focal Loss 这 3 个模块结合使用时，提升

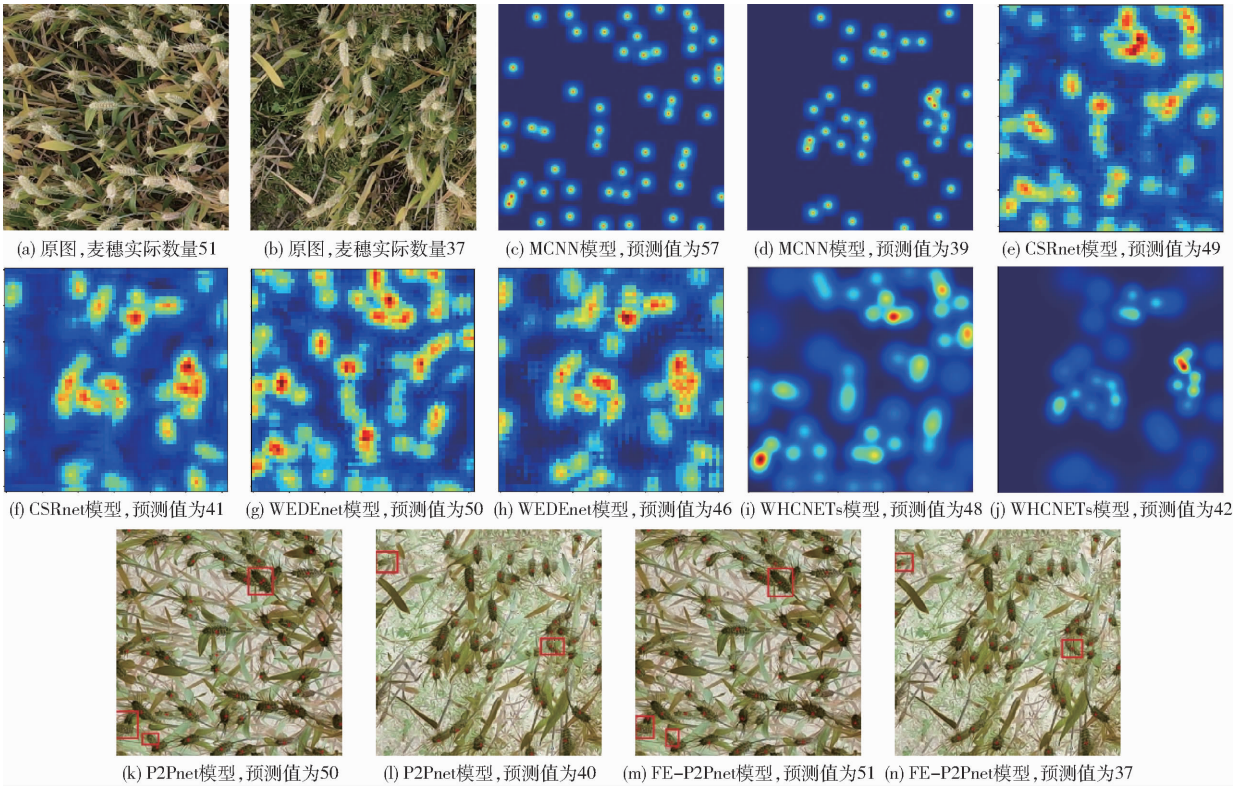


图 14 不同计数模型的预测计数结果

Fig. 14 Predictive counting results of different counting model

效果更加明显。MAE 降低 18.6%，相应地 MSE 也有所降低，ACC 也提升 1 个百分点。这证明了 SE－FPN 模块的有效性，并且表明 Focal Loss 在分类损失方面具有良好的效果。当将 4 个模块共同使用时，实验结果达到了在该实验条件下的最佳效果。MAE 降低 23.2%，MSE 降低 16.6%，ACC 上升 2.67 个百分点。

表 3　消融实验结果

Tab.3　Results of ablation experiment

FEM 模块	SE 模块	Focal Loss	Triplet 模块	数据增强	平均绝对误差	均方误差	平均精确度/%
－	－	－	－	－	4.93	6.15	88.20
－	－	－	－	√	4.50	5.88	88.90
√	－	－	－	√	4.19	5.70	89.20
－	√	－	－	√	4.33	5.93	89.05
－	－	√	－	√	4.27	5.61	88.70
－	－	－	√	√	4.10	5.45	89.34
√	√	√	－	√	4.01	5.40	89.20
√	√	√	√	√	3.77	5.13	90.87

注：“√”表示有此模块，“－”表示无此模块。

3.3　Focal Loss 中超参数选取实验

在 Focal Loss 中，有 2 个超参数 γ、α 会对其性能产生重大影响。Focal Loss 的推荐参考值为 $\gamma=2$，$\alpha\in\{0.25,0.75\}$。因此，本文以这些值作为基准进行预实验。例如，尝试取大于 2 或小于 1 的 γ 值，或取 α 为 $\{0.1,0.9\}$ ~ $\{0.9,0.1\}$。最终决定从 $\gamma=2$，$\alpha\in\{0.4,\ 0.6\}$ 开始进行。该实验目的是通过选择不同的 γ 和 α，找到最佳的实验结果，并主要通过 MAE 和 MSE 对结果进行评价，实验结果如表 4 所示。最终通过实验选择 $\gamma=1$，$\alpha\in\{0.45,0.55\}$ 作为最佳组合。

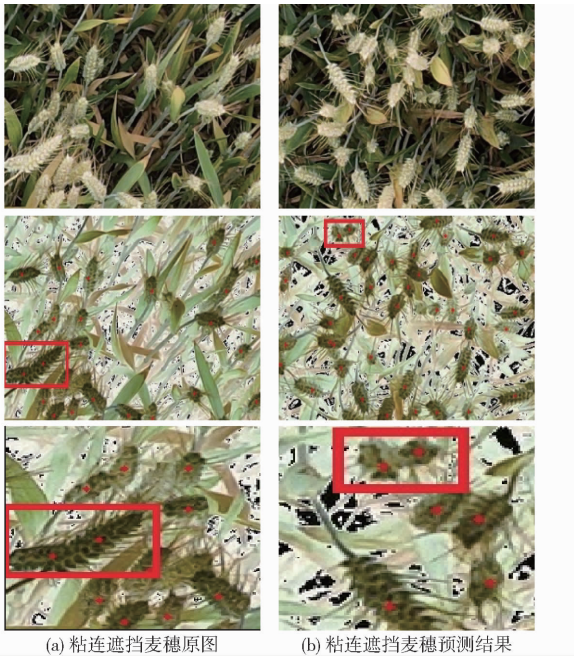
表 4　不同 γ、α 对计数结果的影响

Tab.4　Different γ、α impact on counting results

γ	α	平均绝对误差	均方误差
2	{0.4,0.6}	5.69	6.70
1.5	{0.4,0.6}	5.43	6.78
1.25	{0.4,0.6}	5.16	6.36
1	{0.4,0.6}	4.56	6.32
1	{0.5,0.5}	4.41	6.00
1	{0.45,0.55}	3.77	5.13

3.4　粘连遮挡问题的可视化

本文提出的模型在处理无人机小麦麦穗图像中的粘连和遮挡问题方面表现出一定的效果。根据图 15 的观察，原始图像中的小麦麦穗存在部分粘连和遮挡情况，而本文提出的模型能够预测出部分粘连和遮挡的麦穗，并回归出相应的预测点。这表明本文模型在处理无人机小麦麦穗图像中粘连和遮挡问题方面比较有效。



(a) 粘连遮挡麦穗原图　　(b) 粘连遮挡麦穗预测结果

图 15　粘连遮挡可视化结果

Fig.15　Visualization results of adhesive occlusion

3.5　不同品种小麦预测结果实验

为了验证本文所提出模型的泛化能力，构建一个包含 4 种不同小麦品种（AK1401、YKM222、AK1009 和 AK1706）的数据集，并进行了实验。实验结果如表 5 所示。对于所有品种的测试，平均绝对误差为 5.10，均方误差为 6.17，计数平均精确度达到 89.69%。在针对不同品种的单独测试中，4 个品种的计数平均精确度都在 90% 左右，这表明本文所提出的模型具有较强的泛化性能，可以对不同小麦品种进行麦穗的预测计数。

表 5　不同小麦品种实验结果对比

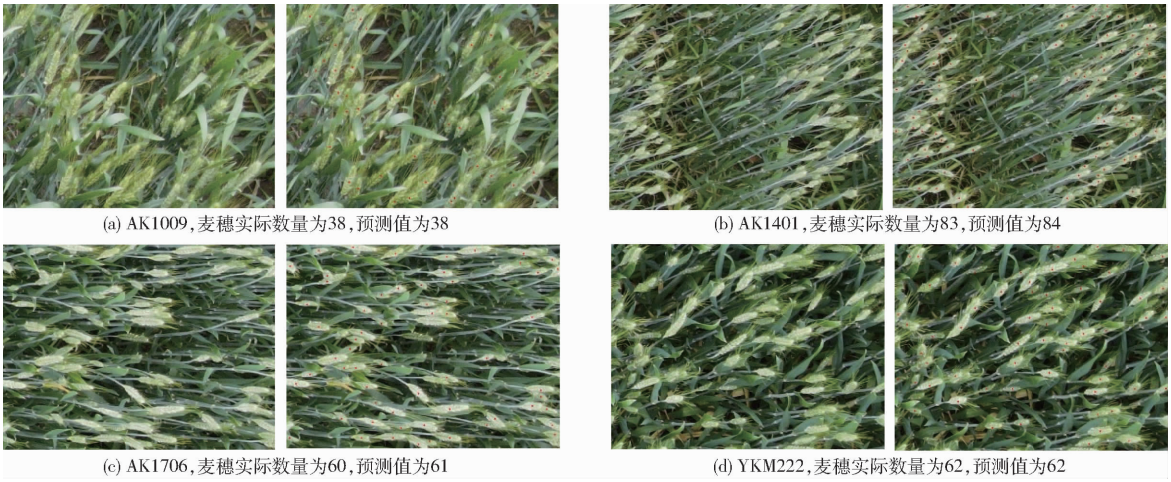
Tab.5　Comparison of experimental results of different wheat varieties

小麦品种	平均绝对误差	均方误差	平均精确度/%
AK1009	5.23	6.31	87.04
AK1401	3.84	5.14	93.47
AK1706	5.15	6.05	90.00
YKM222	5.69	7.47	88.25
所有	5.10	6.17	89.69

图 16 为不同小麦品种预测结果，由图 16 可以看出该模型对这 4 种不同品种小麦的预测计数比较准确，漏检、误检比较少，可以达到比较好的计数效果。

4　结论

（1）针对大田环境下的无人机小麦图像进行麦穗计数，构建了 Wheat－ZWF 数据集。为了解决小麦麦穗存在的尺寸不一、目标小、背景复杂等问题，



(a) AK1009,麦穗实际数量为38,预测值为38 (b) AK1401,麦穗实际数量为83,预测值为84

(c) AK1706,麦穗实际数量为60,预测值为61 (d) YKM222,麦穗实际数量为62,预测值为62

图 16 不同小麦品种的预测计数结果

Fig. 16 Predictive counting results of different wheat varieties

提出了一种基于 FE－P2Pnet 的无人机小麦麦穗计数方法。

(2)通过超参数的相关实验,确定了 Focal Loss 的最佳参数设置为 $\gamma=1,\alpha\in\{0.45,0.55\}$。实验结果表明,本文提出的 FE－P2Pnet 模型在无人机小麦麦穗计数任务中取得了较好的性能。MAE 和 MSE 分别达到 3.77 和 5.13,优于其他相关的回归目标计数模型,并且计数 ACC 达到 90.87%。这表明该模型在无人机小麦麦穗计数方面具有有效性。

(3)为验证模型的泛化性能,使用该模型对不同品种小麦进行了训练和测试。结果显示,平均计数 ACC 也达到 89.69%,各个品种的测试中,计数 ACC 基本在 90% 左右。证明了本文提出的模型具有较好的泛化性能。

参 考 文 献

[1] LIU H, WANG Z, YU R, et al. Optimal nitrogen input for higher efficiency and lower environmental impacts of winter wheat production in China[J]. Agriculture Ecosystems & Environment, 2016, 224: 1－11.

[2] ZHANG Heping, NEIL C T, NARELLE S, et al. Growing-season rainfall, ear number and the water-limited potential yield of wheat in south-western Australia[J]. Crop & Pasture Science, 2010, 61(4): 296－303.

[3] MU H, ZHOU L, DANG X, et al. Winter wheat yield estimation from multitemporal remote sensing images based on convolutional neural networks[C]//2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), 2019.

[4] COINTAULT F, JOURANUX L, GOUTON P. Statistical methods for texture analysis applied to agronomical images[C]// Conference on Image Processing: Machine Vision Applications, 2008: 29－31.

[5] ALKHUDAYDI T, REYNOLDS D, GRIFFITHS S, et al. An exploration of deep-learning based phenotypic analysis to detect spike regions in field conditions for UK bread wheat[J]. Plant Phenomics, 2019, 1: 17.

[6] SADEGHI-TEHRAN P, VIRLET N, AMPE E M, et al. DeepCount: in-field automatic quantification of wheat spikes using simple linear iterative clustering and deep convolutional neural networks[J]. Frontiers in Plant Science, 2019, 10: 1176.

[7] ZHOU C, LIANG D, YANG X, et al. Recognition of wheat spike from field based phenotype platform using multi-sensor fusion and improved maximum entropy segmentation algorithms[J]. Remote Sensing, 2018, 10(2): 246－270.

[8] FERNANDEZ-GALLEGO J A, KEFAUVER S C, GUTIÉRREZ N, et al. Wheat ear counting in-field conditions: high throughput and low-cost approach using RGB images[J]. Plant Methods, 2018, 14(1): 22.

[9] MADEC S, JIN X, LU H, et al. Ear density estimation from high resolution RGB imagery using deep learning technique[J]. Agricultural and Forest Meteorology, 2019, 264: 225－234.

[10] YANG B, GAO Z, GAO Y, et al. Rapid detection and counting of wheat ears in the field using YOLO v4 with attention module[J]. Agronomy, 2021, 11(6): 1202.

[11] MISRA T, ARORA A, MARWAHA S, et al. SpikeSegNet—a deep learning approach utilizing encoder-decoder network with hourglass for spike segmentation and counting in wheat plant from visual imaging[J]. Plant Methods, 2020, 16(40): 1－20.

[12] 鲍文霞, 张鑫, 胡根生,等. 基于深度卷积神经网络的田间麦穗密度估计及计数[J]. 农业工程学报, 2020, 36(21): 186－193.
BAO Wenxia, ZHANG Xin, HU Gensheng, et al. Estimation and counting of wheat ears density in field based on deep convolutional neural network[J]. Transactions of the CSAE, 2020, 36(21): 186－193. (in Chinese)

[13] GUO H. Wheat head counting by estimating a density map with convolutional neural networks[EB/OL]. (2023－03－19)[2023－07－01]. https://doi.org/10.48550/arXiv.2303.10542.

(下转第 289 页)

729 – 742.
ZHANG Xuelin, WU Mei, HE Tangqing, et al. Effects of straw decomposition on inorganic nitrogen and nitrous oxide emissions from two types of soils[J]. Scientia Agricultura Sinica, 2022,55(4):729 – 742. (in Chinese)
[26] 俞湾青,杨士红,祁苏婷.水碳联合管理对稻田浅层土壤 N_2O 浓度剖面分布规律的影响[J].河海大学学报(自然科学版),2021,49(4):335 – 341.
YU Wanqing, YANG Shihong, QI Suting. Effects of combined water and carbon management on the profile distribution of N_2O concentration in shallow soil of paddy field[J]. Journal of Hohai University(Natural Sciences),2021,49(4):335 – 341. (in Chinese)
[27] 杜二小,李焕春,任永峰,等.减氮结合不同肥料配施对农田土壤温室气体排放强度及马铃薯产量的影响[J].中国马铃薯,2023,37(5):440 – 451.
DU Erxiao, LI Huanchun, REN Yongfeng, et al. Effects of nitrogen reduction combined with different fertilizers on greenhouse gas emission intensity of farmland soil and potato yield[J]. Chinese Potato Journal, 2023,37(5): 440 – 451. (in Chinese)
[28] 谢雪,鲁艳红,廖育林,等.紫云英与稻草还田替代部分化肥对双季稻产量和土壤活性有机碳的影响[J].中国农业科学,2023,56(18):3585 – 3598.
XIE Xue, LU Yanhong, LIAO Yulin, et al. Effects of returning Chinese milk vetch and rice straw to replace partial fertilizers on double season rice yield and soil labile organic carbon[J]. Scientia Agricultura Sinica,2023,56(18): 3585 – 3598. (in Chinese)
[29] EAGLE A J, BIRD J A, HORWATH W R, et al. Rice yield and nitrogen utilization efficiency under alternative straw management practices[J]. Agronomy Journal, 2000, 92(6): 1096 – 1103.
[30] 秦子元,张忠学,孙迪,等.水氮耦合对黑土稻作产量与氮素吸收利用的影响[J].农业机械学报,2021,52(12):324 – 335,357.
QIN Ziyuan, ZHANG Zhongxue, SUN Di, et al. Effects of water and nitrogen coupling on rice yield and nitrogen absorption and utilization in black soil[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021,52(12):324 – 335, 357. (in Chinese)

(上接第 164 页)

[14] 鲍文霞, 谢文杰, 胡根生,等. 基于 TPH – YOLO 的无人机图像麦穗计数方法[J]. 农业工程学报, 2023,39(1):155 – 161.
BAO Wenxia, XIE Wenjie, HU Gensheng, et al. Estimation and counting of wheat ears density in field based on deep convolutional neural network[J]. Transactions of the CSAE, 2023, 39(1):155 – 161. (in Chinese)
[15] 高姻燕,孙义,李葆春.基于无人机 RGB 影像估测田间小麦穗数[J]. 中国农业科技导报,2022,24(3):103 – 110.
GAO Yinyan, SUN Yi, LI Baochun. Estimating of wheat ears number in field based on RGB images using unmanned aerial vehicle[J]. Journal of Agricultural Science and Technology,2022,24(3):103 – 110. (in Chinese)
[16] MA J, LI Y, LIU H, et al. Towards improved accuracy of UAV-based wheat ears counting: a transfer learning method of the ground-based fully convolutional network[J]. Expert Systems with Applications, 2021,191: 116226.
[17] 孙俊,杨锴锋,罗元秋,等.基于无人机图像的多尺度感知麦穗计数方法[J]. 农业工程学报,2021,37(23):136 – 144.
SUN Jun, YANG Kaifeng, LUO Yuanqiu, et al. Method for the multiscale perceptual counting of wheat ears based on UAV images[J]. Transactions of the CSAE, 2021,37(23):136 – 144. (in Chinese)
[18] YANG B, PAN M, GAO Z, et al. Cross-platform wheat ear counting model using deep learning for UAV and ground systems [J]. Agronomy, 2023, 13(7): 1792.
[19] SONG Q, WANG C, JIANG Z, et al. Rethinking counting and localization in crowds: a purely point-based framework[C]// Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2021:3345 – 3354.
[20] MISRA D, NALAMADA T, ARASANIPALAI A U, et al. Rotate to attend: convolutional triplet attention module[C]// Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), 2021: 3138 – 3147.
[21] JIE H, LI S, GANG S. Squeeze-and-excitation networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2018: 7132 – 7141.
[22] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017:2999 – 3007.
[23] LANGIN C. Introduction to data mining[M]. Beijing:Posts & Telecom Press, 2006.
[24] WU J, YANG J, HAN L, et al. Automatic counting of in situ rice seedlings from UAV images based on a deep fully convolutional neural network[J]. Remote Sensing, 2019, 11(6): 691 – 710.
[25] ZHANG Y, ZHOU D, CHEN S, et al. Single-image crowd counting via multi-column convolutional neural network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2016:589 – 597.
[26] LI Y, ZHANG X, CHEN D. CSRNet: dilated convolutional neural networks for understanding the highly congested scenes[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 1091 – 1100.