doi:10.6041/j.issn.1000-1298.2024.05.022

# 基于 Shuffle – ZoeDepth 单目深度估计的苗期 玉米株高测量方法

赵永杰<sup>1,2</sup> 蒲六如<sup>1,2</sup> 宋 磊<sup>1,2</sup> 刘佳辉<sup>1,2</sup> 宋怀波<sup>1,2</sup>

(1. 西北农林科技大学机械与电子工程学院, 陕西杨凌 712100;

2. 农业农村部农业物联网重点实验室,陕西杨凌 712100)

摘要:株高是鉴别玉米种质性状及作物活力的重要表型指标,苗期玉米遗传特性表现明显,准确测量苗期玉米植株 高度对玉米遗传特性鉴别与田间管理具有重要意义。针对传统植株高度获取方法依赖人工测量,费时费力且存在 主观误差的问题,提出了一种融合混合注意力信息的改进 ZoeDepth 单目深度估计模型。改进后的模型将 Shuffle Attention 模块加入 Decoder 模块的 4 个阶段, 使 Decoder 模块在对低分辨率特征图信息提取过程中能更关注特征图 中的有效信息,提升了模型关键信息的提取能力,可生成更精确的深度图。为验证本研究方法的有效性,在 NYU-V2 深度数据集上进行了验证。结果表明,改进的 Shuffle - ZoeDepth 模型在 NYU - V2 深度数据集上绝对相对差、均 方根误差、对数均方根误差为0.083、0.301 mm、0.036,不同阈值下准确率分别为93.9%、99.1%、99.8%,均优于 ZoeDepth 模型。同时,利用 Shuffle - ZoeDepth 单目深度估计模型结合玉米植株高度测量模型实现了苗期玉米植株 高度的测量,采集不同距离下苗期玉米图像进行植株高度测量试验。当玉米高度在 15~25 cm、25~35 cm、35~ 45 cm 3 个区间时,平均测量绝对误差分别为 1.41、2.21、2.08 cm,平均测量百分比误差分别为 8.41%、7.54%、 4.98%。试验结果表明该方法可仅使用单个 RGB 相机完成复杂室外环境下苗期玉米植株高度的精确测量。 关键词: 苗期玉米; 株高; 单目深度估计; 测量方法; 混合注意力机制 OSID:

中图分类号: TP391.4 文章编号: 1000-1298(2024)05-0235-09 文献标识码:A

## Measurement Method of Seedling Stage Maize Height Based on Shuffle – ZoeDepth Monocular Depth Estimation

ZHAO Yongjie<sup>1,2</sup> PU Liuru<sup>1,2</sup> SONG Lei<sup>1,2</sup> LIU Jiahui<sup>1,2</sup> SONG Huaibo<sup>1,2</sup>

(1. College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China 2. Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China)

Abstract: Plant height is an important phenotypic indicator for identifying maize germplasm traits and crop vigor, and maize genetic characteristics are obvious at the seedling stage, so accurate measurement of plant height at the seedling stage is of great significance for maize genetic characteristics identification and field management. Aiming at the problem that traditional plant height acquisition methods rely on manual measurement, which is time-consuming and subjective error, an improved ZoeDepth monocular depth estimation model incorporating mixed attention information was proposed. The improved model added the Shuffle Attention module to the various stages in the Decoder module, so that the Decoder module was more able to pay attention to the effective information in all the feature maps in the process of extracting information from the low-resolution feature maps, which enhanced the model's ability of key information extraction, and could generate more accurate depth maps. In order to verify the effectiveness of the method, the validation was carried out on the NYU - V2 depth dataset, and the results showed that the ARE, RMSE, LG were 0.083, 0.301 mm and 0.036, and the accuracy  $\delta$  under different thresholds of the improved Shuffle – ZoeDepth model were 93.9%, 99.1% and 99.8%, respectively, all of which were better than those of the improved Shuffle - ZoeDepth model on NYU - V2 depth dataset. In addition,

作者简介:赵永杰(2002—),男,博士生,主要从事计算机视觉在植物表型检测中的应用研究,E-mail: zhaoyj@ nwafu.edu.cn

收稿日期: 2023-09-29 修回日期: 2023-12-06

基金项目: 国家自然科学基金项目(32272931)

通信作者:宋怀波(1980—),男,教授,博士生导师,主要从事图像处理和模式识别研究,E-mail: songyangfeifei@163.com

the Shuffle – ZoeDepth monocular depth estimation model combined with the maize plant height measurement model was used to complete the measurement of seedling maize plant height, and maize height measurement experiments were carried out by collecting images of seedling maize at different distances, and when the maize height was in the three height intervals of  $15 \sim 25$  cm,  $25 \sim 35$  cm, and  $35 \sim 45$  cm, the AE were respectively 1.41 cm, 2.21 cm, and 2.08 cm, and the PE were 8.41%, 7.54%, and 4.98%, respectively. The experimental results showed that this method can accomplish the accurate measurement of maize plant height at the seedling stage in complex environments using only a single RGB camera with a complex outdoor environment.

Key words: maize in seedling stage; plant height; monocular depth estimation; measurement method; hybrid attention mechanism

#### 0 引言

株高作为植物表型学参数中的重要指标,可以 用来判断作物活力、鉴别作物性状和加速育种进程。 株高测量通常采用人工方式进行,该方法效率低下 且存在人为误差,难以满足大面积、连续测量的要 求。在玉米育种进程中,苗期玉米的遗传特性表现 最为明显,如耐盐碱性、抗低温胁迫性和抗旱性 等<sup>[1-3]</sup>,因此,针对苗期玉米进行高度测量有利于筛 选出更具优势和适应性的玉米品种,加快育种进程, 对实现苗期玉米植株表型鉴别,具有重要意义。

基于相机参数标定的空间距离获取法是目前最 为广泛使用的植株高度测量方法。张慧春等<sup>[4]</sup>利 用光学相机与运动中恢复结构算法生成三维点云, 实现了单株拟南芥主茎长度测量,平均相对误差为 1.07%。邢浩然<sup>[5]</sup>利用摄像机视野中已知高度的蓝 色参照物结合数学模型,完成了摄像机前方单列玉 米株高测量,其中玉米拔节期株高测量最大误差为 3.50 cm,最小误差为0.30 cm,平均误差为2.20 cm。 LI 等<sup>[6]</sup>开发了一款智能手机应用程序,实现了单株 水稻株高测量,均方根误差为1.23 mm。基于相机 参数标定的植物表型获取方法虽检测精度高,但该 类方法一般需要根据植株类型架设不同的参照物, 空间局限性较大,无法满足群体植株表型参数测量 的需求。

随着深度传感器精度的提高,部分研究者利用 深度传感器得到的深度图像来获取植株表型数据。 仇瑞承等<sup>[7]</sup>以拔节期玉米为观测对象,利用 RGB – D 相机获取田间玉米的 RGB – D 图像,通过计算玉 米最高点与地面差值,实现了单株玉米株高测量,单 株玉米株高平均测量误差为 1.62 cm,均方根误差 (RMSE)为 1.86 cm。王纪章等<sup>[8]</sup>利用 Kinect 相机 完成了黄瓜穴盘苗生长监测,在发芽 5 d 内,株高和 实际株高之间拟合度 *R*<sup>2</sup>为 0.875, RMSE 为 1.39 mm。 徐胜勇等<sup>[9]</sup>提出了使用 RGB – D 相机的黄瓜苗表 型无损测量方法。对子叶时期、1 叶 1 心时期和 2 叶1 心时期的黄瓜苗关键表型测量平均绝对误差 均不高于8.59%、*R*<sup>2</sup>不低于0.83。基于深度相机的 植株表型信息获取方法具有检测表型信息种类多、 测量精确等特点,但目前常用的深度相机往往基于 飞行时间(Time-of-flight, TOF)原理,其在短距离场 景表现较差,且在室外环境中易受多因素干扰,不利 于大规模植株表型信息获取。

近年来,单目深度估计技术获得了极大关注。 YUAN 等<sup>[10]</sup>利用 Swin Transformer 结合条件随机场 实现了单幅图像的深度估计。WU 等<sup>[11]</sup>提出了一 种结构蒸馏方法完成了室内场景的绝对深度估计。 BHAT 等<sup>[12]</sup>利用改进后的 Local Bins 方法结合针对 干单目深度估计任务的解码器,完成了单幅图像相 对深度与绝对深度估计。宋磊等<sup>[13]</sup>以 ResNeXt101 网络作为深度估计网络主体,实现了室内多种植株 株高测量,在拍摄距离为105 cm 内番茄幼苗平均绝 对误差为 0.56 cm,均方根误差为 0.82 cm,平均植 株高度比例为1.005。辣椒、甘蓝幼苗平均绝对误 差为0.61、0.32 cm,均方根误差为0.67、0.38 cm。 龙燕等<sup>[14]</sup>利用改进的 HRNet 网络构建了苹果果树 的单目深度估计模型,该模型在果树深度数据集上 平均相对误差和均方根误差分别为 0.123、 0.547 cm,该方法提高了苹果果实等小物体深度估 计的准确性。相较于相机参数标定、传感器获取深 度图像这两种植株高度获取方式,基于单目深度估 计的植株高度获取方法对拍摄环境与拍摄设备要求 较低,空间自由度高,同时在传感器成本方面有较大 优势,更适合室外田间玉米植株高度测量需求。

本文提出一种基于单目深度估计的室外苗期玉 米植株高度获取方法,使用 RGB 相机获取玉米幼苗 图像,利用深度学习技术从单幅图像中恢复相对深 度与绝对深度并分割幼苗区域,获得幼苗最高点与 最高点附近地面深度后结合玉米植株高度测量模型 完成室外苗期玉米植株高度测量。建立一套高效、 准确的室外苗期玉米植株高度测量方法,以期为玉 米幼苗表型监测与育种表型分析提供技术支撑。

### 1 材料与方法

### 1.1 数据采集平台与实验数据集

### 1.1.1 数据采集平台

数据采集在西北农林科技大学某试验田中进 行,为完成室外多株玉米图像获取,搭建如图 1a 所 示的数据采集系统。该系统由小型云台相机、黑色 遮光幕布和可移动平台组成,小型云台相机选用深 圳市大疆创新科技有限公司生产的大疆 DJI Pocket2 小型云台相机作为数据采集设备,用于采集玉米彩 色图像。设备与主要参数如图 1b 所示,该相机内部 配备存储卡,可将采集的图像数据进行存储,图像分 辦率为9216 像素 ×6912 像素,长宽比为4:3,在输 入模型前需将其调整为 640 像素 × 480 像素。可移 动平台由欧标 40 × 40 铝型材和丝杆脚轮构成,平台 长为80 cm, 宽为40 cm, 高度为20~120 cm 可调节, 能够在多株玉米植株之间进行移动。同时,为了避 免在深度估计中光照不均与阴影对模型精度的影 响,在数据采集系统两侧加装黑色遮光幕布。如 图 1c 所示,根据玉米生长特性,分别在玉米植株高 度为15~25 cm、25~35 cm、35~45 cm 3 个生长期 内对 200 余株玉米进行图像数据的获取,共拍摄 626 幅图像,并利用水平塔尺(精度 0.1 cm)测量并 记录植株实际高度。



(c) 不回高度玉米幼苗图像米集 图 1 数据采集平台与图像采集



### 1.1.2 单目深度估计数据集

由于待测目标距相机较近且光照较均匀,选择 NYU-V2 深度数据集作为单目深度估计模型训练 数据集。NYU-V2 深度数据集广泛用于室内场景 三维重建,由纽约大学研究团队采集并发布。该数 据集包含大量室内环境的 RGB 图像、深度图和相机 内外参数,共包含1449 幅图像。通过镜像、平移、 尺度变换后将训练集扩展到 50688 幅,测试集和验 证集分别扩展到 654 幅。如图 2 所示,每幅图像均 提供对应的 RGB 图像、深度图和相机内外参数,这 些信息可用于室内场景的多个计算机视觉任务,如 3D 重建、姿态估计、物体识别和语义分割等。



Fig. 2 NYU – V2 dataset example

### 1.2 单目深度估计模型

单目深度估计可从单幅彩色图像中推断出场景 中各像素点的深度信息,而无需使用多个摄像头或 深度传感器。单目深度估计有两大分支,分别为公 制深度估计(Metric depth estimation, MDE)即绝对深 度估计,与相对深度估计(Relative depth estimation, RDE)。MDE 可以在输出图像中输出公制深度,可 为下游视觉任务提供精确信息,但该方法泛化性能 较差。RDE 只能输出仿射不变深度,即图像中各物 体的远近关系,虽无法为下游视觉任务提供准确信 息,但泛化能力较强。因此,选择 BHAT 等<sup>[12]</sup>提出 的 ZoeDepth 模型为基础模型并进行改进,该模型结 合 MDE 与 RDE 的优点,在输出绝对深度的情况下 仍可保持较好的泛化性能。同时,在该模型的 Decoder 模块中各阶段添加 Shuffle Attention 注意力 模块,该模块可提高模型对不同尺度特征图的理解 能力,提高模型性能。

#### **1.2.1** 基于 Shuffle – ZoeDepth 的单目深度估计模型

如图 3 所示,提出的 Shuffle – ZoeDepth 单目深 度估计模型采用编码-解码(Encoder – Decoder)结构,Encoder 模块采用微软亚洲研究院提出的 Swin Transformer V2 模型<sup>[15]</sup>。该模型以 Swin Transformer<sup>[16]</sup>作为基础模型并进行改进,在单目深 度估计任务中可增强对多尺度信息的提取,提高模 型特征提取能力,在下采样过程中避免信息丢失。 因此采用 Swin Transformer V2 作为 Encoder 模块。

基于 Swin Transformer V2 的 Encoder 模块对图 像操作共分为4个步骤,先对输入图像进行尺寸变 换,之后输入 Encoder 模块,该模块中每个步骤均对 输入进行分辨率减半,通道数提升两倍的操作,由于 采用 Swin - L 框架,初始通道数规定为 192。所以, 编码器最终输出4种尺度特征图,通道数分别为 192、384、576、768,特征图尺寸分别为原图的 1/4、 1/8、1/16、1/32。

Decoder 模块采用密集预测模型 (Dense





prediction transformer, DPT)<sup>[17]</sup> 框架,该模型可在 Encoder 模块中 4 个不同的特征图上进行重新组装 特征,最后使用基于 RefineNet 的特征融合模块结合 从连续阶段提取的多尺度特征图,在每个融合阶段 进行上采样。为了提高上采样过程中信息的完整 性,引入 Shuffle Attention<sup>[18]</sup>注意力机制,该机制可 将通道维度分为多个子特征,然后利用 Shuffle 单元 为每个子特征集成互补通道和空间注意模块。最终 生成 4 种尺度的特征图,通道数分别为 576、384、 192、96,特征图尺寸分别为原图的 1/16、1/8、1/4、 1/2,最后附加一个特定于单目深度估计任务的输出 Head 来输出相对深度图。

绝对深度的输出通过 Metric Bins Module<sup>[19]</sup>模 块实现,该模块将 Encoder 的多尺度特征作为输入, 将上采样与加法操作相结合,利用多层感知机 (Multilayer perceptron, MLP)将5种不同尺度下的特 征图转换为一维嵌入,最后输出绝对深度图,每个像 素点绝对深度可通过在绝对深度图中查询获得。

#### 1.2.2 Shuffle Attention 模块

注意力机制使网络能够准确地聚焦于输入的所 有相关元素。在单目深度估计任务中,注意力机制 可以捕捉像素级的成对关系和通道信息,提高单目 深度估计的精度,同时由于注意力机制出色的多尺 度特征图融合能力,常将其应用在 Decoder 模块中, 本研究将 Shuffle Attention 加入 Decoder 模块中的各 个阶段,使得 Decoder 模块在对低分辨率特征图信 息提取过程中更能关注到所有特征图中的有效信 息,Shuffle Attention 结构如图 4 所示。对于给定的 特征图  $X \in \mathbf{R}^{C \times H \times W}$ ,其中  $C \setminus H \setminus W$  分别表示通道数、 图像高度和图像宽度, Shuffle Attention 首先沿通道 维度将特征图划分为G组,修改后特征图通道数变 为 C/G, 划分后特征图为  $X_{\kappa} \in \mathbf{R}^{C/G \times H \times W}$ , 再将  $X_{\kappa}$ 分 为2个分支,通道数变为 C/(2G),特征图变为  $X_{\kappa_1} \in \mathbf{R}^{C/(2C) \times H \times W}$ ,分别送入通道注意力模块与空间 注意力模块,之后将2个模块输出连接,使得通道数 与输入通道数相等,即 $X_{KIO} \in \mathbf{R}^{C/G \times H \times W}$ ,最终对所用 的子特征进行聚合,再使用 Shuffle 算子<sup>[18]</sup>,使信息 沿通道维度进行跨组传输,最终 Shuffle Attention 模 块的输出与输入特征图 X 相同。

#### 1.2.3 模型评价指标

在单目深度估计任务中,对数均方根误差(LG) 主要用于测量模型估计深度与实际深度差值,可衡 量模型对深度估计问题中不同深度范围的适应能 力,较小的对数偏差通常表示模型在深度估计任务



(2)

中取得了较好性能,LG 计算式为

$$L_{g} = \sqrt{\frac{1}{N}\sum |\lg d_{i} - \lg d_{i}^{*}|^{2}}$$
(1)

式中 *d<sub>i</sub>*——真实深度 *d<sub>i</sub>*<sup>\*</sup>——预测深度 *N*——测试集中图像像素点数

准确率δ为单目深度估计任务中常见的评价指标,用于测量模型预测深度与真实深度间的相对误差是否在一定阈值范围内。具体而言,给定一个阈值δ,准确率δ表示有多少比例的像素深度预测值相对于真实深度的比例在1/δ和δ之间。在这些阈值下,准确率越高,表示模型对深度估计问题的适应能力越好,计算式为

$$\delta = \max\left(\frac{d_i}{d_i^*}, \frac{d_i^*}{d_i}\right) < thr$$
  
(thr = 1.25, 1.25<sup>2</sup>, 1.25<sup>3</sup>)

式(2)为统计所有 d<sub>i</sub> 中小于阈值 thr 的像素点 所占总体像素点的百分比。thr = 1.25,表示模型在 大多数情况下能够预测相对较准确的深度(偏差在 25%以内),该阈值常用于评估模型对小幅度深度 误差的敏感性。thr = 1.25<sup>2</sup>,用于考察模型对中等幅 度深度误差的表现,如果模型能够在更大范围内产 生准确的深度估计,这将更好地反映其性能。thr = 1.25<sup>3</sup>,该阈值进一步扩大了对深度误差的容忍度, 用于评估模型在大幅度深度误差下的性能。综合考 虑这 3 个相对差异的阈值,可以更全面地评估深度 估计模型性能。

在单目深度估计任务中,绝对相对误差 (Absolute relative error, ARE)具有明确的物理意 义,对异常值不敏感。均方根误差(Root mean square error, RMSE)通过平方项对小误差和大误差 施加不同权重,使其更全面地反映深度估计的整体 准确性。ARE 计算式为

$$A_{RE} = \frac{1}{N} \sum \frac{|d_i - d_i^*|}{d_i}$$
(3)

#### 1.3 基于单目深度估计的玉米植株高度测量方法

获得苗期玉米相对深度图与绝对深度图后,如 图5所示,获得玉米植株最高点真实高度与植株最 高点附近地面深度并结合植株高度测量原理,完成 玉米植株高度测量。



Fig. 5 Flowchart of maize plant height measurement

#### 1.3.1 玉米植株最高点获取

获得苗期玉米相对深度与绝对深度图后,利用图 像处理方法在相对深度图上定位苗期玉米植株的最 高点坐标。如图6所示,利用相对深度图中的像素坐 标,在绝对深度图上进行查询获得植株最高点深度。



图 6 玉米植株最高点获取 Fig. 6 Maize plants highest point

#### 1.3.2 植株最高点附近地面深度确定

获取植株最高点深度后,还需获得最高点附近 地面深度。为完成该任务,该方法先在相对深度图 中利用 Canny 边缘算法对相对深度图进行边缘检 测。如图 7 所示,可以分离出玉米植株与背景地面, 再利用获取的玉米植株最高点的坐标位置。得到植 株轮廓后,在绝对深度图中去除植株区域,利用二维 平均卷积块对附近 3 × 3 像素块内的地面深度求平 均值,提取最高点附近地面区域真实深度。

#### 1.3.3 玉米植株高度测量原理

在获得玉米植株最高点深度与最高点附近地面

1/9 1/9

1/9 1/9

1/9 1/9



图 7 玉米植株最高点附近地面深度

Fig. 7 Ground depth near the highest point of maize plant

深度后,结合玉米植株高度测量模型完成玉米植株高 度测量。已知云台摄像机安装高度为 H,玉米植株最 高点距地面高度为 h,如图 8 所示,同时玉米植株最高 点距云台摄像机距离为 x,云台摄像机与玉米植株最 高点连线周围的地面距云台摄像机距离为 X。



Fig. 8 Principle of height measurement of maize plants

根据深度标定的结果,在深度图中可以通过玉 米最高点像素值与最高点附近的地面深度得到 *x* 与 *X* 的比值 *I* 为

$$I = \frac{x}{X} \tag{4}$$

因云台安装高度 H 已知,故可根据相似三角形 原理得出玉米植株高度 h 为

$$h = H - IH \tag{5}$$

评价指标为玉米植株高度绝对误差(Absolute error, AE)与百分比误差(Percentage error, PE),计算 式为

$$A_E = X - X^* \tag{6}$$

$$P_E = \left| \frac{X - X^*}{X} \right| \times 100\% \tag{7}$$

式中 *X*——真实值 *X*<sup>\*</sup>——计算值

1.4 运行环境

本文模型基于 Pytorch 框架实现,在深度学习工 作站上进行训练, CPU 配置为 13th Gen Intel(R) Core(TM) i5 - 13490F, 内存 32 GB, GPU 配置为 NVIDIA GeForce RTX 4060 Ti 16G, 操作系统为 Windows 10, 安装 CUDA, Python 版本为 3.9, PyTorch 版本为 1.13.1。

#### 2 结果与分析

#### 2.1 不同单目深度估计模型性能比较

采用 NYU - V2 深度数据集进行训练,采用上

述 4 个评价指标对 BTS<sup>[20]</sup>、SC - DepthV3<sup>[21]</sup>、 DistDepth<sup>[11]</sup>、ZoeDepth<sup>[12]</sup>4种不同的单目深度估计 模型进行对比,其中 ARE、RMSE、LG 3 类评价指标 越低模型性能越好, $\delta$ 越高模型性能越好,测试结果 如表 1 所示。与 BTS、SC - DepthV3 和 DistDepth 深 度估计模型相比,ZoeDepth 在各个评价指标上均可 取得最优结果,ARE、RMSE、LG为0.098、 0.361 mm、0.042,不同阈值下准确率为91.0%、 98.6%、99.7%。经过分析对比各模型性能, ZoeDepth 综合性能优势明显,故选择该模型作为本 研究的基础模型。

表1 不同单目深度估计模型评价结果对比

 
 Tab. 1
 Comparison of evaluation results of different monocular depth estimation models

		DMCE /		准确率/%		
模型方法	ARE	RMSE/	LG	δ<	$\delta <$	$\delta <$
		m		1.25	1. 25 <sup>2</sup>	
BTS	0.110	0.392	0.047	88.5	97.8	99.4
SC - DepthV3	0. 123	0.486		84.8	96.3	99.1
DistDepth	0.130	0.517		83.2	96.3	99.0
ZoeDepth	0.098	0.361	0.042	91.0	98.6	99.7

#### 2.2 不同注意力机制下 ZoeDepth 模型训练结果

为了验证本文算法的有效性,对添加不同注意 力机制的 ZoeDepth 网络进行试验。数据集采用 NYU - V2 深度数据集,训练时批量设置为 2,训练 次数设置为 10,初始学习率设置为 0.000 16,训练时 间为 48 h。

共探究 6 种不同的注意力机制对模型精度的影 响, 6 种注意力机制分别为  $CA^{[22]}$  (Coordinate attention)、 $ECA^{[23]}$  (Efficient channel attention)、 $SSA^{[24]}$  (Separable self attention)、 $SA^{[18]}$  (Shuffle attention)、CBAM (Convolutional block attention module)<sup>[25]</sup>、SGE (Spatial group-wise enhance)<sup>[26]</sup>。将以上 6 种注意力分别加入 ZoeDepth 中,采用 ARE、RMSE、LG 和  $\delta$  这 4 个评价指标对 7 种深度估 计模型进行对比分析,其中 ARE、RMSE、LG 3 类评价指标越低模型性能越好,  $\delta$  越高模型性能越好。测试结果如表 2 所示。

#### 2.3 玉米植株高度测量模型测试结果

使用 AE 与 PE 作为玉米植株高度评价指标。 采用表 3 所示的 3 个拍摄高度下共 626 幅室外田间 苗期玉米图像进行玉米植株高度测量,测试结果如 表 3 所示。由表 3 可知,当相机距地面高度 40 cm 时,植株高度为 15~25 cm,平均测量绝对误差为 1.41 cm,平均测量百分比误差为 8.41%,当相机距 地面高度 55 cm 时,植株高度为 25~35 cm,平均测

### 表 2 不同注意力机制下评价结果对比

 
 Tab. 2
 Comparison of evaluation results under different attentions

模型		DMSE /		准确率/%		
	ARE	NIJSE/	LG	$\delta <$	$\delta <$	δ <
		m		1. 25 1. $25^2$ 1. $25^3$	1. 25 <sup>3</sup>	
ZoeDepth	0.098	0.361	0.042	91.0	98.6	99.7
Zoe + CA	0.087	0.314	0.037	93.3	99.0	99.8
Zoe + ECA	0.098	0.359	0.041	91.1	98.6	99.7
Zoe + SSA	0.084	0.302	0.036	93.9	99.1	99.8
Zoe + SA	0.083	0.301	0.036	93.9	99.1	99.8
Zoe + CBAM	0.086	0.309	0.037	93.7	99.1	99.8
Zoe + SGE	0.098	0.361	0.042	90. 9	98.6	99.7

#### 表 3 不同相机距离下玉米植株高度测量结果

#### Tab. 3 Results of maize plants heights at different

camera distances

相机高度/cm	玉米高度/cm	株数	AE/cm	PE/%
40	15 ~25	216	1.41	8.41
55	25 ~35	209	2.21	7.51
70	35 ~45	201	2.08	4.98

量绝对误差为 2.21 cm, 平均测量百分比误差为 7.51%, 当相机距地面高度 70 cm 时, 植株高度为 35~45 cm, 平均测量绝对误差为 2.08 cm, 平均测量 百分比误差为 4.98%。

#### 3 讨论

### 3.1 不同单目深度估计网络对玉米植株高度测量 影响

将 BTS<sup>[20]</sup>、SC - DepthV3<sup>[21]</sup>、DistDepth<sup>[11]</sup>、 ZoeDepth<sup>[12]</sup>4种不同的单目深度估计模型应用于本 研究数据集,将AE与PE作为玉米植株测量评价指标,分别在3种相机高度40、55、70 cm下进行测试, 结果如表4所示。试验结果表明,提升单目深度估 计模型性能是提高玉米植株高度测量精度的有效手 段。同时,通过测量结果也可得出 ZoeDepth 模型综 合性能优势明显,故选择该模型作为基础模型。

### 3.2 不同注意力机制对模型性能影响

比较 CA、ECA、SSA、SA、CBAM 与 SGE 6 种注 意力机制对模型性能的影响。不同注意力模块对单 目深度估计模型的影响如图 9 所示。

由图 9 可知,在对原始 ZoeDepth 模型添加不同注意力模块后,模型生成的深度图均有不同程度改善。其中,添加 SA 与 CBAM 注意力模块后的改进效果更为明显(白框区域)。由表 2 可知,在ZoeDepth 模型添加 Shuffle Attention 模块后,模型各项评价指标提升最为明显。ARE 从 0.098降为0.083, RMSE由0.361 mm降为0.301 mm, LG从

### 表 4 不同单目深度估计模型玉米植株高度测量结果

### Tab. 4 Maize plant height measurements from different

monocular	depth	estimation	model
-----------	-------	------------	-------

liffe west	相机高	玉米高	株数	AE/	PE/
模型	度/cm	度/cm		$\mathbf{cm}$	%
	40	15 ~ 25	216	2.13	12. 91
BTS	55	25~35	209	2.46	8.22
	70	35 ~45	201	3.06	7.38
SC – DepthV3	40	15 ~ 25	216	2.21	13. 22
	55	$25\sim 35$	209	3.57	11.67
	70	35 ~45	201	3.34	8.05
DistDepth	40	15~25	216	2.29	13.69
	55	$25\sim 35$	209	3.72	12.32
	70	35 ~45	201	3.54	8.54
ZoeDepth	40	15~25	216	1.41	8.41
	55	25 ~35	209	2.21	7.51
	70	35 ~45	201	2.08	4.98



modules on monocular depth estimation

0.042 降为 0.036, δ 在多阈值(1.25, 1.25<sup>2</sup>, 1.25<sup>3</sup>)
 下均高于 ZoeDepth 原始模型。

单目深度估计任务需同时关注通道与空间信息,与其他4类注意力机制相比,Shuffle Attention 与 Convolutional Block Attention Module 可以更好地结 合通道维度和空间维度信息完成深度估计任务。但 Shuffle Attention 与 Convolutional Block Attention Module 相比,前者更适合完成单目深度估计任务。因此,在 ZoeDepth 模型中加入 Shuffle Attention 可较 大限度提升网络性能,提高输出精度。

#### 3.3 不同相机高度对深度图的影响

在单目深度估计任务中,待测物体在空间中的 深度分布也会对深度估计产生影响。本研究的数据 采集平台高度可调,在对苗期玉米进行数据采集时, 需固定拍摄高度。为探究苗期玉米植株高度测量任 务中,待测物体的空间深度分布如何对深度测量结 果产生影响,先固定摄像机距地面高度 50 cm,用固 定大小的标定物,该标定物为黑色立方体,尺寸为 10 cm × 10 cm × 10 cm,如图 10 所示,分别在标定物 上表面与相机距离 15、20、25、30、35、40 cm 下拍摄 6 组图像,对比其深度差异。





由图 10 可知,当标定物与相机距离过远即距背 景较近时,在生成背景深度时使背景深度不均,当标 定物与相机距离为 15、20 cm 时,生成的地面背景深 度较为平均,当标定物与相机距离为 25、30、35、 40 cm 时,生成的地面背景深度分布较为离散。所 以对 15~25 cm、25~35 cm、35~45 cm 3 个生长期 的苗期玉米,分别设置云台相机高度为 40、55、70 cm 以避免生成地面深度不均,影响测量结果。

### 3.4 不同拍摄高度对苗期玉米植株高度测量精度 的影响

为验证不同拍摄高度对苗期玉米植株高度测量 精度的影响,设置相机高度为70、55、40 cm。由表3 可知,当相机距离地面70 cm,植株高度为35~45 cm 时,测量精度最高,平均测量绝对误差为2.08 cm, 平均测量百分比误差为4.98%。可得当相机距地 面较远时,获得的深度图中地面深度更为平滑,即背 景对待测物体的拉伸效果较弱,最终生成的植株高 度更为准确。

由表5可以发现,在测量时存在误差偏大的数 据,如40 cm组中的第1、2组数据测量误差偏大,测 量值远小于真实值,通过对其原始图像与相对深度 图的观察可以发现,当地面上存在较多杂草时,提取 的地面深度X偏小,而玉米最高点深度x不变,导致 比值 I 增大,在最后进行株高计算时,由于 I 增大,H 不变,导致计算株高 h 较实际值偏小。而 40 cm 组 中的第3、4组数据测量误差偏大,测量值远大于 真实值,通过对其原始图像与相对深度图的观察 可以发现,由于本试验装置是在大田中移动时进 行数据采集,其两侧安装的黑色幕布存在抖动,在 1、2组试验中拍摄的玉米图像中有部分光线透过 幕布与地面间隙形成光斑,当光斑打在玉米植株 上,植株最高点的深度提取值x减小,在X不变的 情况下,比值 I 减小,又因云台摄像机高度 H 不 变,所以在计算株高时会比实际株高偏大。可见, 在保证地面杂草较少与光照均匀的情况下,提出 的玉米植株高度测量方法可在室外大田中高效准 确地完成高度在 10~45 cm 内苗期玉米植株高度 测量。

	表 5	测量中部分异常数据
ab. 5	Some a	bnormal data in measureme

相机高度/cm	真实值/cm	测量株高/cm	AE/cm	PE/%
40	16.3	14.33	1.97	12.09
	17.4	14.33	3.07	17.64
	16.0	18.26	2.26	14.13
	17.6	19.56	1.96	11.14

#### 4 结论

(1)将注意力机制与 ZoeDepth 单目深度估计模 型相结合,优选 Shuufle Attentition 加入 ZoeDepth 模 型,使得该模型各项评价指标均获得高效优化,最终 Shuffle – ZoeDepth 模型 ARE、RMSE、LG 降为 0.083、 0.301 mm、0.036, $\delta$  在多阈值下均高于 ZoeDepth 原 始模型。因此,该模型可完成高精度的单目深度估 计任务。

(2)采用单一 RGB 相机进行数据采集,将 RGB 图像输入单目深度模型中,输出相对深度图与绝对 深度图,再将其输入构建的玉米植株测量模型中,在 不使用多相机与深度相机的条件下,可以完成室外 玉米植株高度的高精度测量。

(3)提出的方法在室外大田环境中进行试验, 当玉米高度为 15~25 cm、25~35 cm、35~45 cm

3个高度区间时,平均测量误差分别为1.41、2.21、 2.08 cm, 平均测量百分比误差分别为 8.41%、 7.54%、4.98%。试验结果表明该方法可在室外复 杂环境下完成对苗期玉米植株高度的精确测量。

#### 考 文 献

- 张春宵,袁英,刘文国,等.玉米杂交种苗期耐盐碱筛选与大田鉴定的比较分析[J].玉米科学,2010,18(5):14-18. [1] ZHANG Chunxiao, YUAN Ying, LIU Wenguo, et al. Comparative analysis between salt-alkali tolerance in seedling stage and production in the field of maize hybrid [J]. Journal of Maize Sciences, 2010, 18(5): 14-18. (in Chinese)
- 丁焓赟, 何序晨, 方芳, 等. 玉米自交系苗期对低温胁迫的响应机制研究[J]. 种子, 2022, 41(10): 1-10. [2] DING Hanyun, HE Xuchen, FANG Fang, et al. Response mechanism of maize in bred linesto cold stressat seedling stage [J]. Seed, 2022, 41(10): 1 – 10. (in Chinese)
- 孙建禄,曾湧,李世成,等.干旱胁迫对玉米生长、生理指标及品质的影响[J].玉米科学,2023,31(4):91-98 [3] SUN Jianlu, ZENG Yong, LI Shicheng, et al. Effects of drought stress on maize growth, physiological indexes and quality[J]. Maize Science, 2023, 31(4): 91-98. (in Chinese)
- 张慧春,王国苏,边黎明,等. 基于光学相机的植物表型测量系统与时序生长模型研究[J]. 农业机械学报, 2019, [4] 50(10): 197 - 207.
  - ZHANG Huichun, WANG Guosu, BIAN Liming, et al. Visible camera-based 3D phenotype measurement system and vimeseries visual growth model of plant [J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(10): 197-207. (in Chinese)
- 邢浩然. 玉米株高远程测量系统关键技术研究[D]. 合肥:安徽农业大学, 2020. [5] XING Haoran. Research on key technologies of remote measurement system of corn plant height [D]. Hefei: Anhui Agricultural University, 2020. (in Chinese)
- [6] LI Hongwei, DONG Wenhao, LI Zehua, et al. Smartphone application-based measurements of stem-base width and plant height in rice seedling [J]. Computers and Electronics in Agriculture, 2022, 198;107022.
- 仇瑞承, 苗艳龙, 季宇寒, 等. 基于 RGB D 相机的单株玉米株高测量方法[J]. 农业机械学报, 2017, 48(增刊): [7] 211 - 219.

QIU Ruicheng, MIAO Yanlong, JI Yuhan, et al. Measurement of individual maize height based on RGB - D camera [J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48 (Supp.); 211-219. (in Chinese)

- 王纪章,顾容榕,孙力,等. 基于 Kinect 相机的穴盘苗生长过程无损监测方法[J]. 农业机械学报,2021,52(2):227-235. [8] WANG Jizhang, GU Rongrong, SUN Li, et al. Non-destructive monitoring of plug seedling growth process based on Kinect camera [J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(2); 227 – 235. (in Chinese)
- [9] 徐胜勇,李磊,童辉,等. 基于 RGB – D 相机的黄瓜苗 3D 表型高通量测量系统研究. [J]. 农业机械学报, 2023, 54(7): 204 - 213,281.

XU Shengyong, LI Lei, TONG Hui, et al. Research on 3D phenotypic high-throughput measurement system of cucumber seedlings based on RGB – D camera [J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(7): 204 – 213,281. (in Chinese)

- [10] YUAN W, GU X, DAI Z, et al. NeW CRFs: neural window fully-connected crfs for monocular depth estimation [J]. ArXiv, 2022, abs/2203.01502.
- [11] WU C Y, WANG J, HALL M, et al. Toward practical monocular indoor depth estimation [C] // IEEE/CVF Conference on Computer Vision Pattern Recognition, 2021: 3804-3814.
- [12] BHAT S, BIRKL R, WOFK D, et al. ZoeDepth: zero-shot transfer by combining relative and metric depth[J]. ArXiv, 2023, abs/2302.12288.
- [13] 宋磊,李嵘,焦义涛,等. 基于 ResNeXt 单目深度估计的幼苗植株高度测量方法[J]. 农业工程学报, 2022, 38(3): 155 - 163. SONG Lei, LI Rong, JIAO Yitao, et al. Method for measuring seedling height based on ResNeXt monocular depth estimation [J]. Transactions of the CSAE, 2022, 38(3): 155 – 163. (in Chinese)
- [14] 龙燕,高研,张广犇.基于改进 HRNet 的单幅图像苹果果树深度估计方法[J].农业工程学报,2022,38(23):122-129. LONG YAN, GAO Yan, ZHANG Guangben. Depth estimation of apple tree in single image using improved HRNet[J]. Transactions of the CSAE, 2022, 38(23): 122 - 129. (in Chinese)
- [15] LIU Z, HU H, LIN Y, et al. Swin Transformer V2: scaling up capacity and resolution [C] // IEEE/CVF Conference on Computer Vision Pattern Recognition, 2021: 11999-12009.
- [16] LIU Z, LIN Y, CAO Y, et al. Swin Transformer; hierarchical vision transformer using shifted windows [C] // IEEE/CVF International Conference on Computer Vision, 2021: 9992 - 10002.
- RANFTL R, BOCHKOVSKIY A, KOLTUN V. Vision transformers for dense prediction [C] // IEEE/CVF International [17] Conference on Computer Vision, 2021: 12159 - 12168.

243

(下转第253页)

45(1): 87 – 110.

- [20] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021:13713-13722.
- [21] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C] // Computer Vision and Pattern Recognition. IEEE,2016: 779 - 788.
- [22] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C] // Computer Vision and Pattern Recognition. IEEE, 2017: 6517-6525.
- [23] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression [C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 1-7.
- [24] REZATOFIGHI S H, TSO I N. Generalized intersection over union: a metric and a loss for bounding box regression [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
- [25] GEVORGYAN Z. SIOU loss: more powerful learning for bounding box regression [J]. Computer Vision and Pattern Recognition, 2022(5):1-12.
- [26] 朱红春,李旭,孟炀,等. 基于 Faster R CNN 网络的茶叶嫩芽检测[J]. 农业机械学报,2022,53(5):217-224. ZHU Hongchun, LI Xu, MENG Yang, et al. Tea bud detection based on Faster R - CNN network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(5):217-224. (in Chinese)
- [27] 武星,齐泽宇,王龙军,等. 基于轻量化 YOLO v3 卷积神经网络的苹果检测方法[J]. 农业机械学报,2020,51(8):17-25.
   WU Xing, QI Zeyu, WANG Longjun, et al. Apple detection method based on light-YOLO v3 convolutional neural network
   [J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(8):17-25. (in Chinese)
- [28] 郭建军,何国煌,徐龙琴,等. 基于改进 YOLO v4 的肉鸽行为检测模型研究[J]. 农业机械学报,2023,54(4):347-355. GUO Jianjun, HE Guohuang, XU Longqin, et al. Pigeon behavior detection model based on improved YOLO v4 [J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(4):347-355. (in Chinese)
- [29] 李健源,柳春娜,卢晓春,等. 基于改进 YOLO v5s 和 TensorRT 部署的鱼道过鱼监测[J]. 农业机械学报,2022,53(12): 314-322.

LI Jiangyuan, LIU Chunna, LU Xiaochun, et al. Fish passage monitoring based on improved YOLO v5s and TensorRT deployment[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(12):314-322. (in Chinese)

- [30] 宋怀波,马宝玲,尚钰莹,等. 基于 YOLO v7 ECA 模型的苹果幼果检测[J]. 农业机械学报,2023,54(6):233 242.
   SONG Huaibo, MA Baoling, SHANG Yuying, et al. Detection of young apple fruits based on YOLO v7 ECA model[J].
   Transactions of the Chinese Society for Agricultural Machinery,2023,54(6):233 242. (in Chinese)
- [31] MARTIN D J, FAHAD S K, MICHAEL F, et al. Adaptive color attributes for real-time visual tracking[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2014, 143(1): 1090 – 1097.

#### (上接第243页)

- [18] ZHANG QL, YANG Y. SA Net: shuffle attention for deep convolutional neural networks [C] // ICASSP IEEE International Conference on Acoustics, 2021: 2235 - 2239.
- [19] BHAT S, ALHASHIM I, WONKA P. LocalBins: improving depth estimation by learning local distributions [J]. ArXiv, 2022, abs/2203.15132.
- [20] LEE J H, HAN M K, KO D W, et al. From big to small: multi-scale local planar guidance for monocular depth estimation [J]. ArXiv, 2019, abs/1907.10326.
- [21] SUN L, BIAN J, ZHAN H, et al. SC DepthV3: robust self-supervised monocular depth estimation for dynamic scenes[J]. ArXiv, 2022, abs/2211.03660.
- [22] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [C] // IEEE/CVF Conference on Computer Vision Pattern Recognition, 2021: 13708 - 13717.
- [23] WANG Q, WU B, ZHU P F, et al. ECA Net: efficient channel attention for deep convolutional neural networks [C] // IEEE/ CVF International Conference on Computer Vision, 2019; 11531 – 11539.
- [24] MEHTA S, RASTEGARI M. Separable self-attention for mobile vision transformers[J]. ArXiv, 2022, abs/2206.02680.
- [25] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [J]. ArXiv, 2018, abs/1807.06521.
- [26] LI X, HU X, YANG J. Spatial group-wise enhance: improving semantic feature learning in convolutional networks [J]. ArXiv, 2019, abs/1905.09646.