doi:10.6041/j.issn.1000-1298.2023.S1.030

基于双流跨模态特征融合模型的群养生猪体质量测定

何　威[1,2]　米　阳[1,2]　刘　刚[1,2]　丁向东[3,4]　李　涛[5]

(1.中国农业大学信息与电气工程学院,北京 100083;2.中国农业大学农业农村部农业信息获取技术重点实验室,北京 100083;
3.中国农业大学动物科学技术学院,北京 100193;4.中国农业大学农业农村部动物遗传育种与繁殖重点实验室,北京 100193;
5.河南丰源和普农牧有限公司,信阳 464000)

摘要:针对生猪体质量准确测定问题,提出了一种跨模态特征融合模型(Cross-modality feature fusion ResNet,CFF-ResNet),充分利用可见光图像的纹理轮廓信息与深度图像的空间结构信息的互补性,实现了群养环境中无接触的生猪体质量智能测定。首先,采集并配准俯视猪圈的可见光与深度图像,并通过 EdgeFlow 算法对每一只目标生猪个体进行由粗到细的像素级分割。然后,基于 ResNet50 网络构建双流架构模型,通过内部插入门控形成双向连接,有效地结合可见光流和深度流的特征,实现跨模态特征融合。最后,双流分别回归出生猪体质量预估值,通过均值合并得到最终的体质量测定值。在试验中,以某种公猪场群养生猪为数据采集对象,构建了拥有 9 842 对配准可见光和深度图像的数据集,包括 6 909 对训练数据和 2 933 对测试数据。本研究所提出模型在测试集上的平均绝对误差为 3.019 kg,平均准确率为 96.132%。与基于可见光和基于深度的单模态基准模型相比,该模型体质量测定精度更高,其在平均绝对误差上分别减少 18.095% 和 12.569%。同时,该模型体质量测定精度优于其他现有生猪体质量测定方法:常规图像处理模型、改进 EfficientNetV2 模型、改进 DenseNet201 模型和 BotNet + DBRB + PFC 模型,在平均绝对误差上分别减少 46.272%、14.403%、8.847% 和 11.414%。试验结果表明,该测定模型能够有效学习跨模态的特征,满足了生猪体质量测定的高精度要求,为群养环境中生猪体质量测定提供了技术支撑。

关键词:群养生猪;体质量测定;双流网络;特征融合;跨模态学习

中图分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1000-1298(2023)S1-0275-08

Estimation of Pig Weight Based on Cross-modal Feature Fusion Model

HE Wei[1,2]　MI Yang[1,2]　LIU Gang[1,2]　DING Xiangdong[3,4]　LI Tao[5]

(1. *College of Information and Electrical Engineering, China Agricultural University, Beijing* 100083, *China*
2. *Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture and Rural Affairs, China Agricultural University, Beijing* 100083, *China*
3. *College of Animal Science and Technology, China Agricultural University, Beijing* 100193, *China*
4. *Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture and Rural Affairs, China Agricultural University, Beijing* 100193, *China*
5. *Henan Fengyuan Hepu Agricultural and Animal Husbandry Co., Ltd., Xinyang* 464000, *China*)

Abstract: In recent years, with the increasing scale of pig farming in the world, farms are in urgent need of automated livestock information management systems to ensure animal welfare. As one of the significant growing information of pigs, the weight of pigs can help farmers to grasp the healthy status of pigs. The traditional methods manually measure pig weight, which are time-consuming and laborious. With the development of image processing technology, the estimation of pig weight by analyzing images has opened up a way for intelligent determination of pig weight. However, many recent studies usually considered only one image modality, either RGB or depth, which ignored the complementary information between the two modalities. To address the above issues, a cross-modality feature fusion model CFF-ResNet was proposed, which made full use of the complementary between texture contour information of RGB images and spatial structure information of depth images, for realizing the intelligent estimation of pig weight

收稿日期:2023-06-20　修回日期:2023-08-16

基金项目:财政部和农业农村部:国家现代农业产业技术体系项目(CARS-35)

作者简介:何威(1999—),男,硕士生,主要从事家畜健康信息智能感知研究,E-mail:hewei@cau.edu.cn

通信作者:米阳(1987—),男,讲师,博士,主要从事计算机技术在智慧农业中的应用研究,E-mail:miy@cau.edu.cn

without human contact in a group farming environment. Firstly, RGB and depth images of the piggery in top view were acquired, andthe correspondence between the pixel coordinates of the two different modalities were used to achieve alignment. Then the EdgeFlow algorithm was used to segment each target individual pig in the coarse-to-fine pixel level, while filtering out irrelevant background information. A two-stream architecture model was constructed based on the ResNet50 network, and a bidirectional connection was formed by inserting internal gates to effectively combine the features of RGB and depth streams for cross-modal feature fusion. Finally, the two streams were regressed separately to produce pig weight predictions, and the final weight estimation values were obtained by averaging. In the experiment, the data was collected from a commercial pig farm in Henan, and a dataset with 9 842 pairs of aligned RGB and depth images was constructed, including 6 909 pairs of training images and 2 933 pairs of test images. The experimental results showed that the mean absolute error of the proposed model on the test set was 3.019 kg, which was reduced by 18.095% and 12.569% compared with the RGB and depth-based single-stream benchmark models, respectively. The average accuracy of proposed method reached 96.132%, which was very promising. Noting that, the model did not add additional training parameters when compared with the direct use of two single-stream models to process RGB and depth images separately. The mean absolute error of the model was reduced by 46.272%, 14.403%, 8.847%, and 11.414% compared with other existing methods: the conventional method, the improved EfficientNetV2 model, the improved DenseNet201 model, and the BotNet + DBRB + PFC model, respectively. In addition, to verify the effectiveness of cross-modal feature fusion, a series of ablation experiments were also designed to explore different alternatives for two stream connections, including unidirectional or bidirectional additive or multiplicative connections. The experimental results showed that the model with a bidirectional additive connection obtained the best performance among all alternatives. All the above experimental results showed that the proposed model can effectively learn the cross-modal features and meet the requirements of accurate pig weight measurement, which can provide effective technical support for pig weight measurement in group farming environment.

Key words: group farming pigs; weight measurement; two-stream network; feature fusion; cross-modal learning

0 引言

根据美国农业部统计数据,2021 年世界猪肉产量为 $1.061\ 0 \times 10^8$ t,相比于 2020 年增长 10.81%[1]。全球猪肉需求的增加,促进了生猪养殖规模的增长,养殖场迫切需要智能化的管理技术。猪的体质量作为生猪重要的生长参数之一,可以帮助生产者控制饲料的投喂量,并了解生猪的健康状况。传统的生猪体质量测定需要人工将猪只驱赶到秤台上,这样不仅消耗了大量的人力资源,同时也容易造成猪只的应激[2]。而且,以这种方式称量猪只可能会减少采食量和采食频率[3]。此外,过度的人工接触容易导致猪瘟在猪群中广泛地传播。

随着机器视觉技术的发展,基于可见光图像的生猪体质量估计在农业领域引起了广泛的关注。利用图像处理技术,自动化的非接触式生猪体质量测量方法不仅避免了对猪只造成应激,同时也降低了人力资源成本。进而促进绿色、高效的生猪养殖业发展。在现有的相关研究中,最广泛采用的方法之一是在猪舍顶部架设相机俯视拍摄生猪图像,然后提取有用的特征回归生猪的体质量。传统的方法对可见光图像进行处理,然后利用手工提取的特征通过回归方程估测体质量[4]。然而,空间结构信息对生猪的体质量测定具有重要意义,仅使用可见光图像可能会丢失相关的特征。近年来,张建龙等[5]和 HE 等[6]使用卷积神经网络(CNNs)学习深度特征,利用深度图像估测生猪的体质量,而单一的深度模态数据缺乏可见光图像中的纹理外观信息。

为了实现可见光图像与深度图像的互补,从而提高生猪体质量测定精度,本文提出一种基于 RGB - D 图像的跨模态特征融合双流回归模型,以提升生猪体质量测定的准确性与鲁棒性,为群养环境中生猪体质量测定提供有效的技术支撑。

1 材料与方法

1.1 试验材料

1.1.1 数据图像采集

利用 RGB - D 相机(微软 Azure Kinect DK)收集一个群养环境下生猪的可见光与深度图像数据集。数据采集于河南省信阳市新县丰源和普农牧有限公司种猪养殖场。采集时间为 2021 年 11 月 11 日—12 月 10 日,采集对象为平均日龄 110 d、体质量范围 70 ~ 110 kg 的长白和大白种公猪。养殖

场单个猪舍长、宽、高分别为 5、4、2.4 m，每个猪舍包含 12～13 头生猪。本次数据采集过程共持续 30 d，最终采集的生猪对象体质量范围为 90～125 kg。在采集过程中，首先通过相机的可见光传感器和 3D 传感器分别从俯视角度拍摄生猪的可见光和深度图像；然后，使用可见光与深度图像配准的算法，通过获取两者的像素坐标并使用相机内部传感器参数计算映射关系，实现两种图像之间的配准。

频繁的人工干预容易引起生猪的应激，进而影响体质量数据的准确性，为此相机被安装在配备了耳标读识器和体质量秤的自动饲喂栏入口附近的顶部，角度与地面水平且高度约为 2.3 m。并通过编写好的采集程序不间断地自动采集并配准俯视生猪的可见光和深度图像。相比人工驱赶生猪到体质量测定栏进行数据采集的方式，直接在群养猪圈上方架设相机采集数据可使生猪处于一个自然放松状态，避免了生猪体质量测定过程中不必要的猪只应激反应和人为接触。采集到的示例图像如图 1 所示。

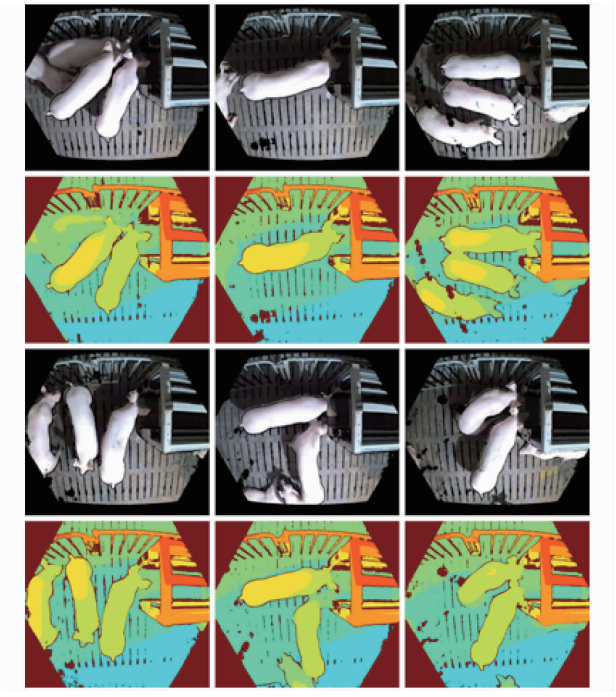

图 1　配准的 RGB－D 图像示例

Fig. 1　Example of registered RGB－D images

1.1.2　数据图像预处理与标定

对于采集后的数据图像，本文设计了一系列图像处理算法来构建本研究所需的数据集。首先当生猪进入饲喂栏后，饲喂系统自动记录其耳标号码、进入时间和体质量。虽然图像中同时出现多头猪只，但饲喂栏一次只能允许一头猪只进入和记录耳标号。因此，本文将采集的图像以时间戳命名，然后根据自动饲喂系统的记录筛选出对应时间段的图像帧。通过跟踪图像帧中进入饲喂栏的生猪便可确定其身份和体质量，从而手动标记加以区分。同时去除模糊的图像以及猪只之间严重粘连的图像。接着使用交互式分割框架 EdgeFlow[7] 基于 HRNetV2[8] 和 OCRNet[9]，从粗到细地像素级分割出可见光图像中目标猪只的前景图像，同时得到目标猪只的掩膜。最后，使用掩膜即可分割得到对应的深度图像中目标猪只的前景图像。通过以上步骤可以分割出图像中有身份和体质量真值标记的每一头生猪，并且过滤掉无关的背景信息。图 2 展示了图像分割过程的示意图。

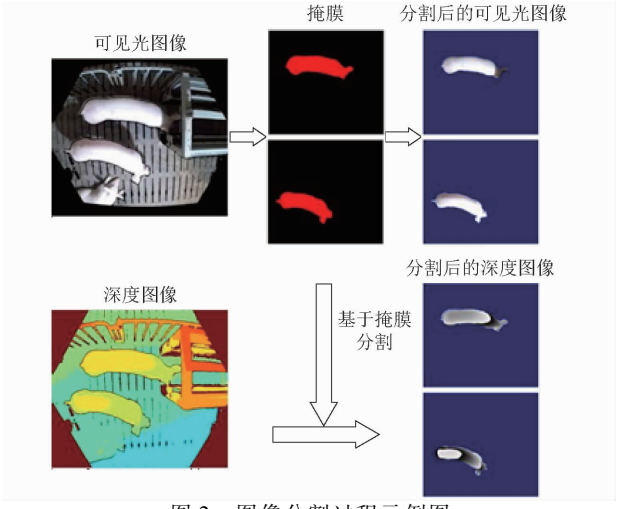

图 2　图像分割过程示例图

Fig. 2　Example diagram for image segmentation process

在深度图像中，存在一些像素点深度值远超相机离地高度的异常区域，比如：猪圈地面漏粪板缝隙。为此，本文通过循环遍历深度图像中每个像素点以定位像素值远大于相机到地面高度的点，并将这些像素点的深度值设置为相机对地的高度。之后，深度图像被归一化为 0～255 的范围，并且单通道的像素值被复制为三通道，以便后续使用迁移学习并有效地与可见光图像进行交互。

最终，本文构建一个包含 9 842 组 RGB－D 图像的数据集，其中用于训练的 6 909 组和用于测试的 2 933 组图像没有交叉。

1.2　方法

1.2.1　双流回归模型主干

本文所提出的双流跨模态特征融合回归模型将 ResNet50 作为主干网络，2 个流分别输入可见光图像和深度图像。为了解决深度神经网络随着层数加深，出现梯度消失或梯度爆炸以及退化问题，微软亚洲研究院提出使用大量残差结构的 ResNets[10－11]。残差结构的主要思想是在输入和输出之间建立一个直接连接，使得新增加的层仅需在原来输入层的基础上学习新的特征，定义表示为

$$y_l = h(x_l) + F(x_l, W_l) \tag{1}$$

其中 $$h(x_l) = x_l$$

式中 y_l——输出 l——第 l 个模块

x_l——输入

$h(x_l)$——恒等映射

$F(x_l, W_l)$——残差函数

W_l——残差结构中瓶颈卷积块的权重

虽然 ResNet 系列网络根据网络层数的不同有多种变体，但在工业界中使用最为广泛的是 ResNet50，其结构参数如表 1 所示。表中，stride 表示卷积的步长，max pool 表示最大池化，average pool 表示均值池化，fc 表示全连接层。输入的图像经过 7×7 的卷积核后，进行批归一化与 3×3 的最大池化，然后经过若干个残差模块完成深层特征的学习，最后是平均池化层和全连接层。与其他 ResNet 系列网络不同的是，ResNet50 的残差结构依次由一个 1×1 卷积核、一个 3×3 卷积核、一个 1×1 卷积核以及一个跳跃连接组成。整个网络中共包含 16 个这样的残差结构，以 3、4、6、3 的组合分配到 4 个阶段进行下采样。

表 1 ResNet50 网络详细结构

Tab. 1 Detailed structure of ResNet50 network

阶段	输出尺寸	模块
stage1	112×112	7×7, 64, stride 为 2
		3×3 max pool, stride 为 2
stage2	56×56	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
stage3	28×28	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$
stage4	14×14	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1\,024 \end{bmatrix}\times6$
stage5	7×7	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2\,048 \end{bmatrix}\times3$
	1×1	average pool, 1 维 fc

对于生猪体质量测定任务而言，这种网络结构简单高效。因此本试验以 ResNet50 网络作为基本网络构建双流回归模型。

1.2.2 跨模态特征融合

传统的双流网络模型[12-14]通常独立地处理不同模态的图像，然后网络末端将 2 个流的输出进行融合。然而对于可见光图像和深度图像而言，两者所包含的信息对生猪体质量测定是互补的，这类晚期融合的方式不一定利于两者之间的交互。因此，本文提出一种双流交互回归模型，通过 2 个网络内部模块的双向连接实现跨模态特征学习。模型框架和具体实现细节如图 3 所示。图 3 中，Conv 表示卷积，Max Pool 表示最大池化，Avg Pool 表示平均池化，FC 表示全连接层，“⊕”表示矩阵相加。首先，2 个流分别以可见光和深度图像作为输入，其中深度图像为处理后的三通道灰度图。之后将每个流中的特定残差块所计算得到的输出进行相加，分别作为下一个模块的输入。该过程的计算公式为

$$y_l^{rgb} = h(x_l^{rgb}) + F(x_l^{rgb} + h(x_l^d), W_l^{rgb}) \tag{2}$$

$$y_l^d = h(x_l^d) + F(x_l^d + h(x_l^{rgb}), W_l^d) \tag{3}$$

式中，函数 $h(x)$ 表示恒等映射，函数 $F(x)$ 表示残差。其中 x_l^{rgb} 和 x_l^d 分别表示可见光流和深度流的第 l 个模块的输入；W_l^{rgb} 和 W_l^d 分别为可见光流和深度流中瓶颈卷积块的权重，由对应阶段的最后一个瓶颈卷积块计算得到。通过这种方式，可见光图像和深度图像中的互补信息可以实现有效地融合。在反向传播过程中，可见光流的损失函数反向传播链式求导公式表示为

$$\frac{\partial L}{\partial x_l^{rgb}} = \frac{\partial L}{\partial y_l^{rgb}} \frac{\partial h(x_l^{rgb})}{\partial x_l^{rgb}} + \frac{\partial L}{\partial y_l^{rgb}} \frac{\partial}{\partial x_l^{rgb}} F(x_l^{rgb} + h(x_l^d), W_l^{rgb}) \tag{4}$$

式中 y_l^{rgb}——特征融合后深度流的输出

L——损失函数

同理，深度流的损失函数反向传播链式求导公式表示为

$$\frac{\partial L}{\partial x_l^d} = \frac{\partial L}{\partial y_l^d} \frac{\partial h(x_l^d)}{\partial x_l^d} + \frac{\partial L}{\partial y_l^d} \frac{\partial}{\partial x_l^d} F(x_l^d + h(x_l^{rgb}), W_l^d) \tag{5}$$

最后，对 2 个流的输出取平均值作为最终的体质量测定结果。

值得一提的是，整个模型的特征融合所插入的双向连接门控没有增加额外的训练参数。除了相加的方式，本文还提出其他的内部跨越连接方式进行消融试验。

1.3 模型训练与评价

1.3.1 试验平台

本文试验均在 64 位 Ubuntu 18.04 LTS 操作系统上进行。试验平台硬件配置为 2 块内存均为 24 GB的 NIVIDIA GeForce RTX3090 显卡和 1 块主频为 3.0 GHz 的 Intel Core i9－10980XE CPU。试验代码采用 Python 在 PyTorch[15]开源深度学习框架下编写，Python 版本为 3.8.8，PyTorch 版本为 1.9.0。

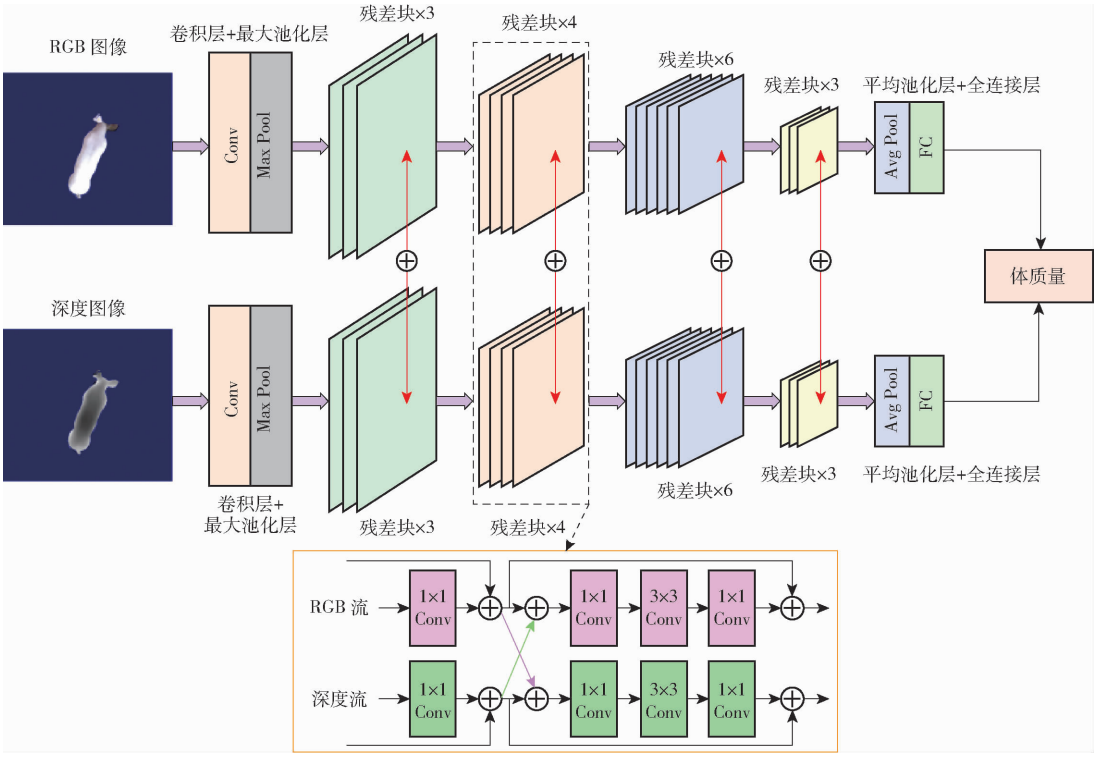

图 3　双流跨模态特征融合模型 CFF－ResNet 结构示意图

Fig. 3　Diagram of structure for two-stream cross-modal feature fusion model CFF－ResNet

1.3.2　试验参数设置

输入尺寸为 640 像素×576 像素的原始图像均被等比例调整为 249 像素×224 像素，然后随机裁剪为 224 像素×224 像素，以便输入到网络模型中。在训练过程中，采用随机翻转和旋转对训练集进行数据增强。优化器采用 Adam[16]并使用余弦退火学习率衰减策略[17]，初始学习率设置为 0.000 1，最小学习率设置为 0.000 01，迭代周期（epoch）为 10。动量因子（momentum）设置为 0.9，正则化权重衰减系数（weight_decay）设置为 0.05。批量大小设置为 64，所有模型均迭代 150 次，同时使用 ImageNet 预训练模型权重进行迁移学习。

1.3.3　试验评价指标

采用平均绝对误差（Mean absolute error，MAE）作为评价标准，并辅以平均百分比误差（Mean absolute percentage error，MAPE）和均方根误差（Root mean square error，RMSE）用于评估方法鲁棒性。平均绝对误差是一种经常用于回归模型的损失函数，可以很好地反映预测值误差的实际情况。因此可用于衡量生猪体质量测定的误差。平均绝对百分比误差可以用于衡量模型的拟合效果。MAPE 越小，说明预测模型拟合效果越好，具有更好的精确度。均方根误差表示预测值和样本真实值之差的样本标准差，可以用来反映体质量测定误差的波动程度。

2　结果与分析

2.1　与单模态基准模型的比较

多模态数据相比于单模态数据可以为模型提供更多的特征信息。为了验证跨模态特征融合模型 CFF－ResNet 在生猪体质量测定方面的优越性，本文首先选取基于单模态的 ResNet50[11]基准模型进行比较试验，结果如表 2 所示。

表 2　与单模态基准模型对比试验结果

Tab. 2　Comparison of experimental results with single-modality based methods

方法	可见光模态	深度模态	平均绝对误差/kg	平均百分比误差/%	均方根误差/kg
ResNet50		√	3.453	4.383	6.207
ResNet50	√		3.686	4.632	6.503
CFF－ResNet	√	√	3.019	3.868	5.896

注：“√”表示采用对应的模态数据。下同。

在使用可见光和深度两种数据后，本文所提出的方法相比于单模态基准方法达到了更高的准确度，其中 MAE 为 3.019 kg，MAPE 为 3.868%，即平均准确率为 96.132%，RMSE 为 5.896 kg。与单模态基准网络模型相比，MAE 比基于可见光的 ResNet50 减少 18.095%，比基于深度的 ResNet50 减少 12.569%。

图 4 展示了特征融合模型 CFF－ResNet 以及 2 个单模态基准模型测试结果的绝对误差分布图。误差区间 1～15 分别对应 0～0.5 kg、0.5～1.0 kg、1.0～

1.5 kg、1.5 ~ 2.0 kg、2.0 ~ 2.5 kg、2.5 ~ 3.0 kg、3.0 ~ 3.5 kg、3.5 ~ 4.0 kg、4.0 ~ 4.5 kg、4.5 ~ 5.0 kg、5.0 ~ 5.5 kg、5.5 ~ 6.0 kg、6.0 ~ 6.5 kg、6.5 ~ 7.0 kg、7.0 ~ 7.5 kg。

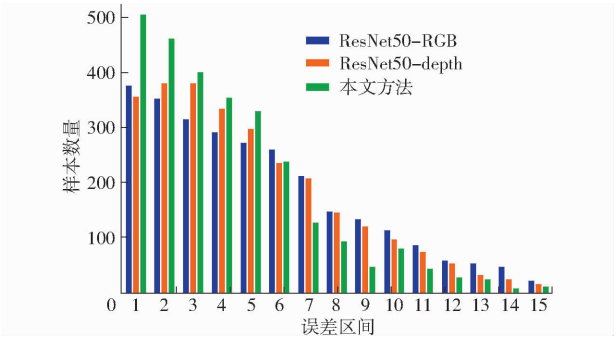

图 4 绝对误差分布

Fig. 4 Absolute error distribution

从图 4 中可见，与基于单流可见光模态和基于单流深度模态的基准模型相比，本文提出的跨模态特征融合方法的绝对误差分布更集中在较小值的区间内。

2.2 不同融合方式的影响

在 2.1 节中，为了融合 2 个模态的信息，本文设计了插入双向连接与加法门控的双流跨模态特征融合模型。本节将通过消融试验比较不同融合方式的跨模态特征融合模型在生猪 RGB - D 数据集上测试的结果。试验共设计 7 种双流跨模态特征融合方式，包括使用 3 种不同加法门控连接的方式、使用 3 种不同乘法门控连接的方式和仅对双流输出结果进行加权平均的方式。图 5 展示了 3 种加法门控连接方式的示意图，并在表 3 中展示了这 7 种模型在测试集上的试验结果。其中，模型 1 为将深度流单向连接到可见光流并插入加法门控；模型 2 为将可见光流单向连接到深度流并插入加法门控；模型 3 则为插入双向连接与加法门控的双流回归模型；模型 4、模型 5 和模型 6 的连接方式分别与模型 1、模型 2 和模型 3 相同，但将加法门控替换为乘法门控；模型 7 仅做晚期融合，即对双流最终输出结果进行加权平均。

模型 1 ~ 3 中 2 个流的连接门控为加法。模

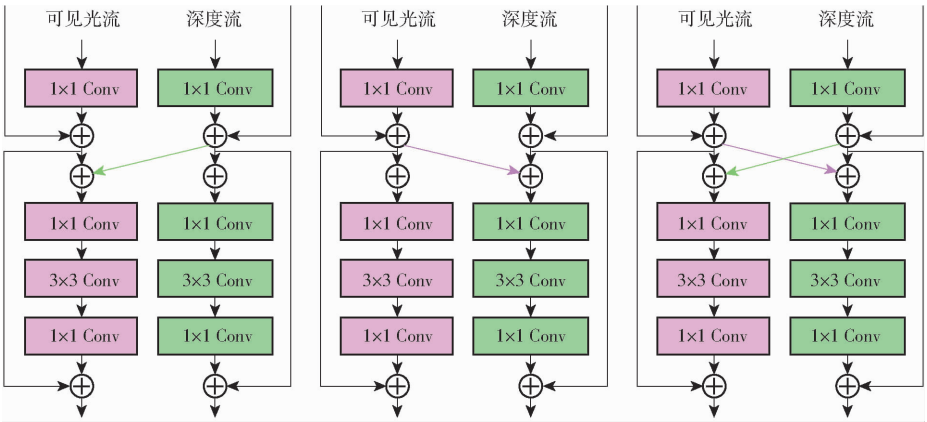

图 5 不同的双流融合方式

Fig. 5 Different ways of two-stream fusion

表 3 采用不同双流融合方式模型的试验结果

Tab. 3 Experimental results of models with different two-stream fusion ways

模型	连接方式	平均绝对误差/kg	平均百分比误差/%	均方根误差/kg
CFF - ResNetV1	Depth→RGB(⊕)	3.463	4.349	6.199
CFF - ResNetV2	Depth←RGB(⊕)	3.309	4.173	6.075
CFF - ResNetV3	Depth↔RGB(⊕)	3.019	3.868	5.896
CFF - ResNetV4	Depth→RGB(⊙)	3.459	4.378	6.239
CFF - ResNetV5	Depth←RGB(⊙)	3.293	4.211	6.223
CFF - ResNetV6	Depth↔RGB(⊙)	3.154	4.053	6.087
CFF - ResNetV7		3.475	4.413	6.435

注：“→”表示从深度到可见光流；“←”表示从可见光到深度流；“↔”表示双向；“⊕”表示融合方式是加法；“⊙”表示融合方式是乘法。

型 1在测试集上的 MAE 达到 3.463 kg；相比之下，模型 2 的连接方式使得 MAE 下降 4.445%。这说明体质量回归模型对于空间结构信息更加关注。模型 3 实现了两者信息的相互融合。可以看到，相比于单向流的连接方式，这种采用双向流连接的方式取得了更好的性能，MAE 为 3.019 kg，分别比模型 1 和模型 2 下降 12.821% 和 8.764%。出现这种情况的原因是可见光图像的纹理轮廓信息和深度图像的

空间结构信息对生猪体质量的测定是互补的，单一的连接方式不能很好地学习到两者的互补信息。

模型 4 ~ 6 中 2 个流的连接门控由加法改变为乘法，其余保持不变。试验结果表明，在采用单向连接方式时，模型 4 和模型 5 分别相比于模型 1 和模型 2 的性能有所提升，但是在采用双向连接后，模型 6 的性能反而比模型 3 有所下降。这可能是因为乘法相比加法更能使信号产生较大的变化，双向使用导致传播的信号变化过大，从而干扰网络的表示能力。

最后，采用晚期融合方式的模型 7 在所有的 7 种模型中性能最差。这证明了模型内部早期融合的重要性以及晚期融合不能很好地学习跨模态的特征。在验证了 2 个流双向相加连接方式的优越性后，研究将采用这种融合方式进行后续试验的设计。

2.3　与其他生猪体质量测定方法的比较

现有的基于图像的生猪体质量测定方法主要有常规图像处理模型和深度神经网络模型。为了继续验证本文设计的跨模态特征融合模型 CFF - ResNet 的有效性和优越性，首先采用常规图像处理方法开展试验进行比较。具体而言，用体尺自动检测软件 LSSA_CAU[18] 处理深度图像，从而得到所有样本的体尺参数，包括体长、肩宽和臀高。然后建立多元线性回归模型

$$w_t = 0.254s + 0.245w + 1.608h - 40.795 \quad (6)$$

式中　w_t——生猪体质量，kg

s——体长，cm

w——肩宽，cm

h——臀高，cm

相关测试结果如表 4 所示，其中常规图像处理方法的 MAE、MAPE 和 RMSE 分别为 5.619 kg、6.060% 和 9.853 kg。而本文提出的 CFF - ResNet 模型的 MAE 为 3.019 kg，相比常规图像处理模型减少 46.272%。

表 4　与其他方法对比试验结果

Tab. 4　Comparison of experimental results with other methods

方法	可见光模态	深度模态	平均绝对误差/kg	平均百分比误差/%	均方根误差/kg	参数量	浮点运算量
常规图像处理方法		√	5.619	6.060	9.853		
EfficientNetV2		√	3.527	4.496	6.407	2.03×10^7	2.90×10^9
EfficientNetV2	√		3.530	4.517	6.408	2.03×10^7	2.90×10^9
DenseNet201		√	3.312	3.312	6.106	2.00×10^7	4.34×10^9
DenseNet201	√		3.481	4.420	6.352	2.00×10^7	4.34×10^9
BotNet + DBRB + PFC		√	3.408	4.363	6.285	2.96×10^7	1.76×10^{10}
BotNet + DBRB + PFC	√		3.724	4.705	7.360	2.96×10^7	1.76×10^{10}
CFF - ResNet	√	√	3.019	3.868	5.896	4.70×10^7	8.22×10^9

试验选取其他先进的深度神经网络模型以及生猪体质量测定模型进行对比探究，其中包括改进 DenseNet201[19] 模型、改进 EfficientNetV2[20] 模型、包含 ResNet 和 BotNet 块的双分支以及并行全连接层的 BotNet（BotNet + DBRB + PFC）[6]。试验时，本文分别验证这些方法在单一的可见光图像和深度图像数据集上的性能，并与所提出的 CFF - ResNet 模型进行性能比较。试验结果如表 4 所示。在使用单一模态数据时，基于深度图像的模型准确度普遍优于基于可见光图像的模型，这进一步验证了空间结构信息更有利于生猪体质量的测定。同时，本文所提出方法的生猪体质量准确度优于其他所有方法。相比其他方法中准确度最佳的改进 DenseNet201，MAE 比基于可见光和基于深度模态的模型分别减少 13.272% 和 8.847%。此外，在对比试验中训练的 BotNet + DBRB + PFC 模型的浮点运算量（Floating point operations，FLOPs）最大，然而最佳准确度不及本文提出的 CFF - ResNet 模型和改进 DenseNet201 模型。其中，CFF - ResNet 模型的 MAE 与其基于深度模态的 MAE 相比减少 11.414%。而改进 EfficientNetV2 作为一个轻量化的模型，其浮点运算量在众多模型中较低，然而测试结果中的最佳 MAE 不及其他几种模型。其中，CFF -ResNet 模型的 MAE 与该模型基于深度模态的 MAE 相比减少 14.403%。与以上模型相比，本文提出的改进方法所采用的 ResNet50 主干网络被广泛应用于工程领域，在模型的参数量和性能上达到了平衡。在经过双流跨模态特征学习后，生猪体质量测定性能得到了显著提升。同时，本研究还基于上述跨模态特征学习思路构建了双流 EfficientNetV2 模型，其精度较单模态基准模型有所提升，MAE 为 3.376 kg，但逊色于本研究提出的 CFF - ResNet 模型的 3.019 kg。

综上所述，本文方法具有简单直接、准确率高，并且未产生过多计算量的特点。除此之外，该方法具有通用性，并不局限于特定的骨干网络模

型,便于实际生产环境中的迁移部署以及后续开发,从而实现生猪体质量的智能化测定。

2.4 评估结果分析

表 5 展示了部分生猪的编号、真实体质量、平均预测体质量以及预测准确率,其中大部分生猪个体的体质量预测准确率可以达到 96% 以上。

表 5 生猪真实体质量与预测结果

Tab. 5 Real weight and predicted results of pigs

生猪编号	真实值/kg	预测值/kg	准确率/%
1943489	84.00	84.89	98.94
1954422	105.00	100.80	96.00
1954428	87.00	84.84	97.52
1954896	84.50	83.18	98.43
1954898	74.50	77.26	96.44
1954910	93.00	95.90	96.88
1954916	91.00	95.20	95.38
1954936	87.00	85.55	98.33
1955400	99.00	101.62	97.35
1955412	91.50	88.91	97.17
1955416	86.00	84.43	98.17
1991668	106.50	109.31	97.36

综合以上结果表明,本研究所提出的跨模态特征融合模型可以通过学习可见光和深度两种模态的互补信息,有效地提高生猪体质量测定的精度。

3 结论

(1) 针对现有生猪体质量测定研究仅聚焦于单一模态数据的问题,首先采集并处理得到 9842 组配准的生猪 RGB - D 图像数据,然后提出了跨模态特征融合模型 CFF - ResNet,有效地融合了可见光和深度图像中的外表纹理轮廓和三维空间结构信息,实现了群养环境下无接触的生猪体质量准确测定。消融试验表明,模型能够充分利用跨模态图像信息的互补性,实现跨模态特征的有效融合。对比试验结果表明,本研究所提出的模型取得了最优性能,MAE 为 3.019 kg。

(2) CFF - ResNet 模型在计算量和准确率之间得到了平衡,同时具有一定的可迁移性。未来的研究将聚焦于设计更加高效的跨模态学习模型,同时将考虑扩展数据集,使其包含尽可能多的不同体质量范围的生猪图像数据,从而进一步提高本研究的适用性。

参 考 文 献

[1] 杨侗瑀, 王祖力, 刘小红, 等. 2021 年世界生猪产业发展情况及 2022 年的趋势[J]. 猪业科学, 2022,39(2):34 - 38.
YANG Dongyu, WANG Zuli, LIU Xiaohong, et al. Development of the world pig industry in 2021 and trends in 2022[J]. Swine Industry Science, 2022,39(2):34 - 38. (in Chinese)

[2] BRANDL N, JORGENSEN E. Determination of live weight of pigs from dimensions measured using image analysis[J]. Computers and Electronics in Agriculture, 1996,15(1):57 - 72.

[3] AUGSPURGER N R, ELLIS M. Weighing affects short-term feeding patterns of growing-finishing pigs[J]. Canadian Journal of Animal Science, 2002,82(3):445 - 448.

[4] 杨艳, 滕光辉, 李保明, 等. 基于计算机视觉技术估算种猪体重的应用研究[J]. 农业工程学报, 2006, 22(2):127 - 131.
YANG Yan, TENG Guanghui, LI Baoming, et al. Measurement of pig weight based on computer vision[J]. Transactions of the CSAE,2006,22(2):127 - 131. (in Chinese)

[5] 张建龙, 冀横溢, 滕光辉. 基于深度卷积网络的育肥猪体重估测[J]. 中国农业大学学报, 2021,26(8):111 - 119.
ZHANG Jianlong, JI Hengyi, TENG Guanghui. Weight estimation of fattening pigs based on deep convolutional network[J]. Journal of China Agricultural University, 2021,26(8):111 - 119. (in Chinese)

[6] HE H, QIAO Y, LI X, et al. Automatic weight measurement of pigs based on 3D images and regression network[J]. Computers and Electronics in Agriculture, 2021,187:106299.

[7] HAO Y, LIU Y, WU Z, et al. Edgeflow: a chieving practical interactive segmentation with edge-guided flow[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 1551 - 1560.

[8] WANG J, SUN K, CHENG T, et al. Deep high-resolution representation learning for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(10): 3349 - 3364.

[9] YUAN Y, CHEN X, WANG J. Object-contextual representations for semantic segmentation[C]// Computer Vision-ECCV 2020: 16th European Conference, 2020: 173 - 190.

[10] HE K, ZHANG X, REN S, et al. Identity mappings in deep residual networks[C]// Computer Vision-ECCV 2016: 14th European Conference, 2016: 630 - 645.

[11] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016: 770 - 778.

[12] MI Y, ZHANG X, LI Z, et al. Dual-branch network with a subtle motion detector for microaction recognition in videos[J]. IEEE Transactions on Image Processing, 2020, 29: 6194 - 6208.

[13] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[C]// Advances in Neural Information Processing Systems, 2014.

(下转第 329 页)

(16):146 – 149. (in Chinese)

[89] 付长波. 重庆市两翼地区特种经济动物养殖现状调查及市场分析[D]. 重庆:重庆师范大学,2011.
FU Changbo. A survey and market analysis on special economic animals' breeding in Chongqing[D]. Chongqing: Chongqing Normal University,2011. (in Chinese)

[90] 韩锋,陈绍志,赵荣. 梅花鹿驯养繁殖经济效益评价[J]. 野生动物学报,2017,38(1):22 – 27.
HAN Feng,CHEN Shaozhi,ZHAO Rong. The economic benefit evaluation of captive breeding of sika deer (*Cervus nippon*)[J]. Chinese Journal of Wildlife, 2017,38(1):22 – 27. (in Chinese)

[91] 陈盈盈,鲍连艳,赛道建,等. 大丰自然保护区麋鹿驯养保护的生态对策[J]. 山东师范大学学报(自然科学版),2004(3):74 – 76.
CHEN Yingying,BAO Lianyan, SAI Daojian, et al. The improvement of habitat in Dafeng's deer reserve[J]. Journal of Shandong Normal University (Natural Science), 2004(3): 74 – 76. (in Chinese)

[92] 侯立冰,孙大明,俞晓鹏,等. 江苏大丰麋鹿人工驯养试验研究[J]. 畜牧兽医科学(电子版),2019(1):13 – 14.

[93] STARČEVIĆ V, SIMIĆ M, RISOJEVIĆ V, et al. Integrated video-based bee counting and multi-sensors platform for remote bee yard monitoring[C]//2022 21st International Symposium Infoteh-Jahorina (INFOTEH). IEEE, 2022: 1 – 6.

[94] BJERGE K, FRIGAARD C E, MIKKELSEN P H, et al. A computer vision system to monitor the infestation level of Varroa destructor in a honeybee colony[J]. Computers and Electronics in Agriculture, 2019, 164: 104898.

[95] PRATHAN S, AUEPHANWIRIYAKUL S, THEERA-UMPON N, et al. Image-based silkworm egg classification and counting using counting neural network[C]//Proceedings of the 3rd International Conference on Machine Learning and Soft Computing, 2019: 21 – 26.

[96] 科大讯飞. 猪只盘点挑战赛[EB/OL]. 2021 – 10 – 24[2023 – 05 – 07]. https://challenge.xfyun.cn/topic/info?type=pig-check.

[97] KAY J, KULITS P, STATHATOS S, et al. The caltech fish counting dataset: a benchmark for multiple-object tracking and counting[C]//Computer Vision-ECCV 2022: 17th European Conference, 2022: 290 – 311.

(上接第 282 页)

[14] WANG L, XIONG Y, WANG Z, et al. Temporal segment networks: towards good practices for deep action recognition[C]// European Conference on Computer Vision. Springer, Cham, 2016: 20 – 36.

[15] PASZKE A, GROSS S, MASSA F, et al. Pytorch: an imperative style, high-performance deep learning library[C]// Advances in Neural Information Processing Systems, 2019.

[16] KINGMA D, BA J. Adam: a method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

[17] LOSHCHILOV I, HUTTER F. Sgdr: stochastic gradient descent with warm restarts[J]. arXiv preprint arXiv:1608.03983, 2016.

[18] GUO H, MA X, MA Q, et al. LSSA_CAU: an interactive 3D point clouds analysis software for body measurement of livestock with similar forms of cows or pigs[J]. Computers and Electronics in Agriculture, 2017,138:60 – 68.

[19] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 2261 – 2269.

[20] TAN M, LE Q. Efficientnetv2: smaller models and faster training[C]//International Conference on Machine Learning, 2021: 10096 – 10106.