



東北大學秦皇岛分校

Northeastern University at Qinhuangdao

毕业论文

通过视觉测量鸭子体重

院别	控制工程学院
专业名称	测控技术与仪器
班级学号	20178210
班级序号	170120
学生姓名	陈若愚
指导教师	赵玉良

2021年5月31日

郑重声明

本人呈交的学位论文，是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料真实可靠。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确的方式标明。本学位论文的知识产权归属于培养单位。

本人签名：

日期：2021年5月31日

通过视觉测量鸭子体重

摘要

在畜牧业和食品工业中，在食品生产过程中对动物进行自动称重一直是一个挑战。为了解决这个问题，基于图像的非接触方法已经被开发出来用于给猪和羊称重。这些方法显著提高了肉食品行业的盈利能力和效率。然而，基于图像的方法来称重不规则形状的小动物还没有发展。由于被处理的动物图像往往存在不完整性以及噪声，对于手工提取特征存在困难性，不完整的特征对于图像称重的预测是有精度损失的。

因此，本文提出了一种卷积神经网络(CNN)的深度学习回归模型，使用 2D 图像来准确估计悬挂在生产线上的鸭子的重量，该 CNN 模型自动提取特征，无需人工干预。使用从一个畜牧工厂收集的数据集，训练该 CNN 模型。在两种情况下进行了测试，一种是在简单的场景下，采集的鸭子图像不存在遮挡，完整度较高；另一种是在更复杂的场景下采集的数据训练模型，鸭子图像存在相互之间的遮挡。

对两种场景下视觉测重的方法进行了评估，采集鸭子图像对每只鸭子都会采集到不同角度的图像。针对简单场景本文使用 50 只鸭子作为训练与测试，采用 10 折法与平均绝对精度评估模型，误差为 58.8 克，准确率高达 97.85%，远远超过了传统的图像测重的方法；在复杂场景下采用了 282 只鸭子作为训练与测试，提出了一种随机融合的预处理方法，将精度提升到了 55.6g。无论在何种场景下都使用了小规模数据集进行训练，减少了大规模数据训练的负担性。值得一提的是，据我们所知，本文的工作是第一个通过图像端到端预测动物体重的。

针对本文实验所得到的结论，验证了端到端视觉测重的可行性，总结了端到端视觉称重的挑战与困难，并提出了视觉称重方法潜在的几种方法，希望我们的工作能为推动智慧畜牧业发展贡献一份力量。

关键词：图像称重，深度学习，端到端，随机融合，小规模数据集

Weighing Ducks by Computer Vision

Abstract

In the animal breeding and food industries, automatically weighing animals during the food production process remains a challenge. To address this, image-based non-contact methods have been developed for weighing pigs and sheep. These methods have significantly improved the profitability and efficiency of the meat food industry. However, image-based methods for weighing small animals with irregular shapes have yet to be developed. The incomplete and noisy images of animals being processed are difficult for manually extract features, incomplete features have a loss of accuracy for the prediction of image weighing.

Here, we propose a convolutional neural networks (CNN) regression model that uses 2D images to accurately estimate the weight of ducks hanging on the production line. This CNN model extracts features automatically without human intervention. Using datasets collected from an animal husbandry factory, we trained this CNN model. We attempt to train the model in two environments, one is in a cleaned environment that the ducks don't exist cover in images, which have high completeness; another experience is using the datasets in a complex environment to training a model, that exists occlusion in the images by ducks.

We evaluate the image weighing method in two environments, the datasets collected from each environment will include images from all kinds of angles. In the cleaned environment we using 50 ducks for training and testing, evaluated it using 10-fold cross-validation and mean square deviation (*MAD*), the model achieves a mean abstract deviation (*MAD*) of 58.8 g and an accuracy of 97.85%, which better than two widely used traditional image weighing methods; In the complex environment we collected 282 ducks for training and test, we proposed a stochastic image fusion preprocessing method for image preprocessing, that achieve a mean abstract deviation (*MAD*) of nearly 55.6 g. We either using small scale datasets to train model, that reduces the complicated calculation of large datasets. It is worth mentioning that, to our knowledge, our work is the first work to predict animal weight end-to-end through images.

We proved the feasibility of end-to-end image weighing, summarized the challenges and difficulties of end-to-end image weighing, and proposed several potential methods for image weighing. We hope that our work can contribute to the development of smart animal husbandry.

Key Words: Image Weighing, Deep Learning, End-to-end, Small Scale Datasets

目 录

1 绪论	1
1.1 研究背景	1
1.2 研究现状	1
1.2.1 接触式测重	1
1.2.2 通过 3D 图像视觉称重	2
1.2.3 通过 2D 图像视觉称重	2
1.3 本文难点	3
1.4 主要创新点	4
1.5 章节安排	4
2 图像处理与机器学习基础理论	7
2.1 图像处理	7
2.1.1 数字图像	7
2.1.2 感兴趣区域 ROI	8
2.1.3 图像通道拆分与合并	8
2.1.4 图像运算	9
2.1.5 阈值分割	9
2.1.6 图像形态学操作	10
2.2 机器学习理论	11
2.2.1 机器学习	11
2.2.2 学习算法	11
2.2.3 模型评估	12
2.2.4 过拟合与欠拟合	12
2.2.5 正则化与交叉验证	13
2.2.6 神经网络	14
2.3 计算机视觉深度学习应用	16
2.3.1 卷积神经网络	16
2.3.2 经典卷积神经网络结构	20
2.3.3 基于梯度的优化方法	22

2.4 本章小结.....	24
3 端到端图像称重.....	25
3.1 数据集与预处理.....	25
3.1.1 数据集采集.....	25
3.1.2 数据的预处理.....	26
3.1.3 设备与编程环境.....	27
3.2 卷积神经网络回归模型.....	27
3.3 结果与验证.....	28
3.3.1 超参数优化.....	28
3.3.2 评估鸭子的体重.....	31
3.3.3 与传统特征工程方法的比较.....	32
3.4 讨论.....	34
3.5 本章小结.....	34
4 利用残缺图像进行体重预测.....	37
4.1 数据集与预处理.....	37
4.1.1 数据集采集.....	37
4.1.2 数据预处理.....	38
4.1.3 设备与编程环境.....	39
4.2 卷积神经网络结构.....	39
4.3 体重估算方法.....	40
4.4 结果.....	41
4.4.1 模型精度评估.....	41
4.4.2 融合图像对模型精度的影响.....	41
4.5 不同视角对模型的影响.....	42
4.6 本章小结.....	44
5 总结、讨论与展望.....	45
5.1 研究工作总结.....	45
5.2 不足与展望.....	45
5.3 潜在可挖掘方向讨论.....	46

5.3.1 基于解耦合方法加速预测	46
5.3.2 基于目标检测的通用体重预测算法	47
5.4 本章小结	48
结 论	49
攻读学位期间的学术成果	51
致 谢	53
参考文献	55
附 录	61
附录 A 附加数据图与表.....	61
附录 B 证明.....	63
附录 C 英文原文.....	64
附录 D 中文译文	76

1 绪论

1.1 研究背景

中国每年养鸭超过 30 亿只，是世界上养鸭最多的国家^[1]。鸭子屠宰场的一项关键任务是称量鸭子的体重，以进行日常管理^{[2][3]}、健康监测^{[4][5]}和各种其他管理活动^[6]。与此同时，养鸭农户和屠宰场根据体重来交易鸭子。然而，传统的人工称量鸭子的方法劳动强度大，效率低。因此，养鸭业需要更准确、更高效的鸭子称量方法来提高盈利能力和社会效益。

屠宰加工过程主要以人机结合的分割方式为主，不同于传统的加工制造行业，该过程是活体宰杀并分解分割的过程，是一个由一而多的过程，故此整个生产工艺对过程数据的采集有很迫切的需求。数据对于原料结算，生产结构分析，成本分析等都起到了十分重要的作用。

针对订单进行分级生产（不同的客户对于产成品的规格有独特的约定，比如肯德基需要的鸡肉必须规格统一，也就要求了单体胴体必须在一定的重量范围内）。原料的结算，肉鸡、肉鸭作为屠宰企业的主要生产原料，均为外采的方式获取，故此需以重量为单位与供应方进行结算，称重的输出结果可以作为结算的依据。数据的累积对于分析和定义生产计划，原料采购计划，市场规划三方面均有十分重要的管理意义。

禽肉的重量直接决定了产品的经济效益。禽肉生产分割的第一个环节就是计数和称重，也就是需要通过外延设备，得知生产流水线上单体只数和重量规格。传统的方式为红外点只器和电子称重配合而成。随着生产企业对于生产过程的精益化要求不断提升，传统的方式对于数据的采集和分析已经达不到管理要求，故此亟待引入新的技术方式解决这一问题。

1.2 研究现状

1.2.1 接触式测重

接触式称重系统用于监督管理动物的生长状况，主要借助接触式的电子器件与传感器的方式测量。González-García 等人^[7]开发了一个 Walk-Over-Weighing 的系统，用于给活羊称重。他们塑造了一种门式的空间，当活羊经过此通道空间时，置于侧部的传感器将收集压力信号并进行拟合训练分析，以达到自动称重的目的；Wang 等人^[8]开发了一个

杆状称台系统，用于测量活鸡的体重。该系统通过摄像头图像处理用于定位与计数活鸡的情况，即可自动化地采集活鸡的体重信息。

然而接触式测量存在很大的缺点，如果对象是活物，在非人为干预的情况下，很难短时间将大批量动物测量体重，因为需要动物活动到检测台；如果对象是非活物，则可能需要借助外力将非活物移至测量平台进行称重，从而造成了过大的成本。因此接触式测重的方法可能在精度方面最为准确，但是从时间和成本的方面考虑，接触式测重不是一个很好的方案。

1.2.2 通过 3D 图像视觉称重

非接触称重的最典型方案便是通过图片进行测重。3D 视觉图像的获取方式主要分为两种，一种为通过双目摄像机构建立体视觉，通过相机标定，双目校正，立体匹配等步骤，即可计算深度得到空间信息；另一种方式为通过深度摄像机或者激光雷达直接获取点云的信息，构建空间信息。很多的视觉称重方法应用了 3D 立体视觉作为特征提取的来源：Pezzuolo 等人^[9]在猪的顶部区域部署了深度相机，可以获得猪的深度信息，用于计算猪的体长、前高、后高和胸围。然后他们回归了体重和猪的这些多变量特征之间的关系；Shi 等人^[10]通过双目视觉技术，测量了猪的背部面积，并基于此拟合了猪的背部面积与体重之间关系；Menesatti 等人^[11]使用双摄像头系统从提取羊的肩隆高度，体长和胸部深度，利用这些多变量估计羊的体重；Taşdemir 等人^[12]以荷斯坦奶牛为研究对象，采用图像分析的方法确定奶牛的体尺、臀高、体长和臀宽，构建多元线性回归方法估计奶牛的活重；Buranakarl 等人^[13]提出了一种估计水牛体重的方法，通过扫描仪器对水牛进行了 3 维重建，然后通过 3D 模型提取了身体高度、心脏周长、肩宽、髂宽、坐骨结节宽度、肩膀和回肠的翼之间的长度、回肠翼之间的长度坐骨结节和肩膀之间的长度坐骨结节，再采用多元线性回归分别对公牛与母牛进行体重预测。

然而计算空间信息本就存在较大的复杂性，从空间信息获取特征已存在一定的误差，再用存在误差的特征进一步估算体重会进一步减小模型的鲁棒性。相机的标定等也存在一定的工作量，便捷性不高。

1.2.3 通过 2D 图像视觉称重

通过 2D 图像测量动物的体重信息，即只通过一张照片，而无立体空间信息，预测动物的体重。目前通过 2D 图像来估算动物体重的工作非常多，目前 2D 图像称重的方

法基本都是通过手工的特征工程提取特征，再进行拟合等操作。2D 图像的处理更加简单快捷，通过图像中目标的尺寸与形态可用于与体重建立关联。

常见的 2D 图像估计动物体重的方法包括 2 种，一种为通过计算图像中目标的面积（即目标的像素数目）并与体重进行拟合。该方案已应用于很多目标，例如 Kashiha 等人^[14]通过椭圆拟合法提取图像中猪的背部面积，用于回归体重；Mollah 等人^[15]提取鸡的背部面积用于估测体重；Balaban 等人^[16]通过视野面积预测阿拉斯加鳕鱼的质量，他们的另一项工作^[17]也通过类似的方法预测了 4 种三文鱼的质量。

仅靠像素面积与体重建立关系，很难达到高精度，故部分工作尝试提取更多的特征，尝试提高图像称重的精度。Lines 等人^[18]提出了一种视觉称重鲑鱼的方法，提取了鲑鱼的轮廓信息，遵从了 Humphries 等人 1981 年提出的特征选取方法^[19]，识别了一系列连接鱼类可识别的解剖特征和覆盖全身的桁架的长度，并构建了一个线性模型；Velardo 等人^[20]通过正面照片和侧面照片提取人体的的 7 个特征值，通过多元线性回归的方法预测人体的体重；Banhazi 等人^[21]通过提取猪背部在竖直方向的几个极值点的数值宽度，构建多元线性回归来预测活猪的体重；Tasdemir^[22]等人通过数字图像提取了荷斯坦奶牛的高度、臀高、体长与臀宽，通过回归方程对奶牛体重进行了回归；Wongsriworaphon 等人^[23]通过提取猪的边缘信息，坐标标准化，然后通过人工神经网络进行特征回归；Amraei 等人^[24]利用椭圆图像分割肉鸡的体型，将面积、周长、凸面积、长轴长度和短轴长度等特征输入人工神经网络来估计肉鸡体重；Wang 等人^[25]提取面积、凸面积、周长、偏心度、长轴长度和短轴长度作为特征，并利用这些特征和人工神经网络对行走猪的重量进行估计；Jun 等人^[26]提取了猪背的二值化图像和测量面积、曲率、偏差作为特征，并基于提取的特征和人工神经网络估计了猪的体重。

然而，动物的体重绝不仅仅与部分几个特征有关，例如在 Zong 等人^[27]的工作中就明确指出了仅靠部分特征是无法精准地估计猪的体重的。然而通过人手工提取特征存在困难性，提取何种特征？该特征是否有利于识别？特征之间是否存在耦合？

1.3 本文难点

绝大部分系统对于目标动物的体重预测均存在较大的偏差，例如牛，羊等的体重预估，大多数系统只能在一个群体中获得较好的效果，而有效地进行单物体体重识别还存在较大的困难^[27]。最主要的原因还是与测重的方法有关，用视觉图像进行体重预测，最重要的是特征的选择。目前的研究基本都只是选取了很少一部分的特征，例如图像的面

积，身体的形态学特征等等。然而部分的特征只取决于整个样本的走势，而与个体之间的关系很难完全靠此决策，例如猪可能有同样的骨架，但是身体的肉量不同，导致体重不同。

同时，特征的选择也是相当困难的一面，选择何种特征？该特征是否有利于动物体重的识别？特征之间是否存在耦合？是否缺少了一些特征从而导致了模型性能的下降？数据分析存在一个上限，这个上限是由数据和特征决定的，而采用何种算法只是尝试逼近这个上限，如果这个上限本就达不到一般要求，而采取何种算法都是徒劳的。此外一些复杂的3D视觉系统提取特征本就会存在较大的偏差，并且更容易受到环境噪声的干扰，将受到偏差的积累，这也是无法精确预测单一动物体重的原因。

此外，很多的系统在需要进行称重时，都需要进行设备的校准和安装，例如很多针对猪的体重预测工作都需要改装猪舍^[27]。如果稍微存在偏差，例如通过2D图像的面积进行体重估算，同一只猪在不同远近的摄像机前2D图像的像素面积是不同的，稍有偏差会干扰到线性函数的预测。针对3D系统虽然可以解决远近的问题，但是3D系统依然忽略了面积等因素，仅靠身体骨架结构预测也是存在较大精度误差的。

1.4 主要创新点

本文的主要工作为针对流水线上悬挂的屠宰后的鸭子，用 2D 视觉图像估算鸭子的体重，并以个体为单位精准地预估。本文的主要创新点如下：

1. 首次提出了端到端的视觉称重方案，省去了人工提取特征的复杂操作，仅将整幅图片作为输入，特征则由卷积层自动地提取。该方案不仅在处理上简便，并在模型精度方面远远地超越了传统的特征提取方法，超出了 2 倍的精度。
2. 提出了一种从多视角评估动物体重的方法，当预测某一动物体重时，综合选取各个角度的图像联合评估，或取平均值，或融合图像作为输入。该方法解决了动物多方位视觉预测体重不稳定的问题，提高了模型预测的鲁棒性。
3. 针对复杂场景下图像残缺的问题，提出了一种随机融合的预处理方式，通过在各个角度下的图像随机融合，不仅可以扩大数据集的样本量，同时也增加样本的多样性，有效地提高了破损图像的体重预测精度。

1.5 章节安排

本文主要尝试解决多种场景下对流水线鸭子进行图像称重，加快预测速度，提高模

型精度，并尝试应用到实际场景中。主要的工作从简单场景验证，从复杂场景改善。具体的章节安排如下：

第二章介绍本文涉及技术的相关工作，主要包括基本的图像处理操作，机器学习技术以及其下的细分方向。

第三章和第四章主要介绍本文在图像称重方法上的工作，其中第三章主要是在简单场景下，图像完整度较高情况下验证端到端的策略。第四章是基于第三章理论的基础上，讨论在复杂场景，即图像可能存在残缺的情况下，如何提高模型的预测能力。

第三章首次提出了端到端的动物图像称重的策略，该策略免去了繁琐的人为手工提取特征的过程。我们从工厂中架设了便携式相机正对流水线，收集了50只鸭子，每只鸭子包括各个视觉图像若干。提出了一种卷积神经网络的结构，用于直接预测端到端体重，并讨论了不同模块对模型的影响。在评估单只鸭子的体重时采取多视角策略，从各个视角计算得到的体重取平均作为最终估算体重。采用十折法对模型进行性能评估，平均绝对误差为58.8克，平均相对误差为2.15%。我们也将我们的方法与传统的通过面积，以及提取形态学特征再用多变量拟合的方法，最终我们的精度超越了传统方法的2倍。验证了端到端自动提取特征方法的可行性。

第四章主要针对在复杂的流水线下，鸭子与鸭子之间存在遮挡的情况下，如何通过残损的图像对鸭子进行高精度地体重预测？我们采集了300只鸭子并进行了数据清洗，每只鸭子存在残缺的图像。基于第三章已有的理论，仍然采取端到端的预测方式。本文针对残损图像问题提出了一种随机融合的预处理方法，不仅扩展了样本的多样性，也扩展了样本的信息量，更有效地进行图像称重。最终我们对数据进行了评估，平均绝对误差为55.6克，大大提高了模型在图像存在残缺情况下的精度。

第五章主要针对本文的全部工作进行总结，并针对已有的问题进行展望，围绕着模型速度与鲁棒性提出了一系列潜在的可深度挖掘但未进行实验验证的方向与方法。

2 图像处理与机器学习基础理论

本章主要从原理与发展方面综合介绍本文所需用到的技术知识。具体包括一些基本的图像处理知识以及机器学习，还有计算机视觉的应用方法。本章的图像处理内容主要参考了 Gonzalez 著的《Digital Image Processing》^[28]，机器学习与深度学习的理论主要参考了周志华著的《机器学习》^[29]、李航著的《统计学习方法》^[30]、Goodfellow 等人著的《Deep Learning》^[31]，Bishop 著的《Pattern Recognition and Machine Learning》^[32]。

2.1 图像处理

2.1.1 数字图像

数字图像，即用数字来表示一幅图像，可以理解为一个矩阵 I ，这个矩阵的高和宽分别为 H 和 W ，则一个数字图像的大小便为 $H \times W$ 。其中，整个矩阵由 $H \times W$ 个点组成，我们称之为像素。每一个像素在数字图像中都有一个位置表示， H 方向由 x 表示， W 方向由 y 表示，则获取一个像素即 $I(x, y)$ ，获取的参数即为数字图像在该点的强度^[28]。不像人的视觉系统，数字图像往往是由有限的像素组成，数字图像方便了人们存储视觉信息。同时数字图像也被用于处理一些任务，例如手机等电子产品的显示，医学图像处理，监控摄像，异常检测等等。可以说，数字图像处理对人们的生产、生活等起到了至关重要的作用。我们大部分的设备是以RGB（红、绿、蓝）图像的格式保存彩色图像的，即每个像素都有3个数值。部分图像可能以灰度图的形式保留，则每个像素只有1个数值，如图2.1所示。数值的取值范围为0到255，最小值0一般代表灰色，最大值255一般代表纯黑色。其他的色彩表示，如CMY（青、粉红、黄），CMYK（青、粉红、黄、黑）等。

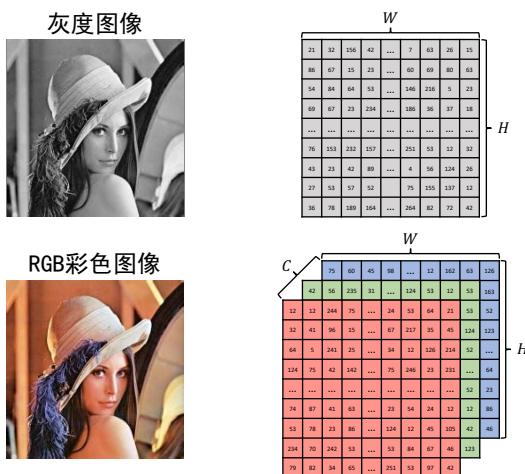


图2.1 数字图像的表示

2.1.2 感兴趣区域 ROI

ROI，即 region of interest，感兴趣的区域。图像处理中往往我们只需要整张图中某一个特定的区域，用框等方法分割出来，用于所需要的处理操作。我们经常需要提取 ROI 区域，例如目标检测中，我们只有提取了一个特定的区域，才能从这个区域中分析物体所属的类别，是图像处理中的常见的操作。

分割特定的兴趣区域，可以用框，圆形，三角形，等各种的多边形，而常用方法是用矩形框勾勒。通常我们只需要知道框的左上坐标 (x_1, y_1) 与右下坐标 (x_2, y_2) 即可勾勒出 ROI 区域，或者知道区域的中心坐标 (x, y) ，以及框的长 w 与宽 h ，如图 2.2 所示。



图 2.2 提取 ROI 区域

2.1.3 图像通道拆分与合并

部分图像，如 RGB 图像，是存在多个通道的，通道的处理也是一个问题。例如酮体吸收红光，则信息量最大区域即在红光区域，则图像的 R 通道包含更显著的特征，而花草等在绿光区域，在 G 通道包含更显著的特征。我们图像处理中常常包括图像通道的分离，从而选择特定的通道，当然通道的顺序也会影响图像的显示效果，如图 2.3 所示。



图 2.3 图像通道顺序排布与分离

2.1.4 图像运算

图像运算，如四则运算加减乘除意义。图像的四则运算，运算数除了是单个的数字外，如二维矩阵和三维矩阵之间的运算，要求有相同的尺寸，即相同的 $H \times W$ 。图像运算有什么作用？例如图像加减法运算中，实例便是图像融合。图像的乘法运算，多应用于类似与的操作，如得到一个二值化掩膜，可以与原图进行乘操作以提取图像中与掩膜有关的内容。图像除操作，经常的便是归一化，将图像的像素数值范围从[0,255]变为[0,1]，从而方便数据运算。在很多的神经网络处理图像中，都需要将图像减去一个平均值，再归一化，以方便数据的训练。

2.1.5 阈值分割

阈值分割，即通过不同的阈值，将图像的像素值进行对应的操作，具体可理解为当图像某位置的像素值高于阈值，如何处理，低于阈值，如何处理？是图像二值化操作中最常用的方法，如我需要对图像中的物体进行检测，背景部分像素强度较低，而物体中像素强度较高，我们便可在他们中设定一个阈值，使高于阈值的像素值设 1 或 255，低于阈值的像素值清零。常见的阈值化分为二进制阈值化，反二进制阈值化，截断阈值化，阈值化为 0，反阈值化为 0。具体的描述与公示见表 2.1，阈值分割的几个例子如图 2.4 所示。

表 2.1 几种阈值分割方法

阈值方法	描述	公式
二进制阈值化	将像素值大于阈值的数设置为自定义数值，否则设置为 0。	$dst(x, y) = \begin{cases} setting & src(x, y) > thresh \\ 0 & otherwise \end{cases}$
反二进制阈值化	将像素值大于阈值的数设置为 0，否则设置为自定义数值。	$dst(x, y) = \begin{cases} setting & src(x, y) \leq thresh \\ 0 & otherwise \end{cases}$
截断阈值化	将像素值大于阈值的数设置为阈值，否则不变。	$dst(x, y) = \begin{cases} threshold & src(x, y) > thresh \\ src(x, y) & otherwise \end{cases}$
阈值化为 0	将像素值大于阈值的数不变，否则设置为 0。	$dst(x, y) = \begin{cases} src(x, y) & src(x, y) > thresh \\ 0 & otherwise \end{cases}$
反阈值化为 0	将像素值大于阈值的数不设置为 0，否则不变。	$dst(x, y) = \begin{cases} src(x, y) & src(x, y) \leq thresh \\ 0 & otherwise \end{cases}$

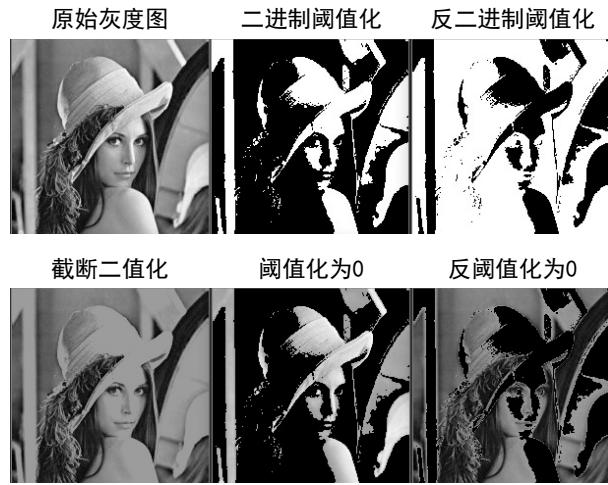


图 2.4 几种图像腐蚀操作的示意图

2.1.6 图像形态学操作

图像形态学操作，主要包括图像的腐蚀与图像的膨胀操作。图像的腐蚀操作可用式 2.1 表示：

$$A \ominus B = \{z | B_z \subseteq A\} \quad (2.1)$$

其中 A 为目标图像， B 为图像腐蚀所用的卷积模板。大致是通过 B 与 A 进行卷积操作，其中要求 B 的尺寸为基数尺寸，因为要遍历每个像素点时，像素点位于核的中心位置。然后计算核 B 覆盖位置卷积运算数值的最小值，并用这个最小值代替该位置像素值，依次遍历图像所有像素。图像腐蚀操作常用于去除图像中比较细小的噪声，一般可控制核函数的形状与大小去除噪声。

图像的膨胀操作可用式 2.2 表示：

$$A \oplus B = \{z | B_z \cap A \neq \emptyset\} \quad (2.2)$$

图像膨胀可看作图像腐蚀的对偶形式，用核 B 的点与原图 A 的点依次迭代，如果核 B 与图 A 覆盖区域非空，则这样的像素点构成的集合可认为是 B 对 A 的膨胀图像。图像膨胀的作用一般是为了去除存在于物体中的一些孔洞，裂纹等。以一个 5×5 的核为例，图像腐蚀与膨胀如图 2.5 所示。



图 2.5 图像腐蚀与膨胀的效果

2.2 机器学习理论

2.2.1 机器学习

从上世纪 50 年代开始，人工智能技术便开始发展并融入到我们的生活中，与我们的生活息息相关。根据张钹院士等人的总结，可将人工智能化为符号主义与连接主义^[33]。符号主义方法试图通过一些通用的知识解决广泛的问题，一系列的方法包括知识工程，专家系统，知识图谱等等方法，以归纳已有的知识产生一定的逻辑推理。连接主义借助了人神经传播的思维方式，希望从刺激等角度产生模式上的感应。无论是符号主义，还是连接主义，都可与机器学习结合，机器学习也是人工智能领域最重要的技术之一。

早期的机器学习方向的人工智能方法，例如 1950 年 Turing 提出的图灵测试，1954 年 Samuel 提出的基于博弈的可学习西洋跳棋程序^[34]，并将机器学习定义为，不需要显示的编程，即可使计算机获取学习能力。1958 年，Rosenblatt 提出了感知机模型^[35]，模拟人脑的神经运行方式。然而 1968 年，Minsky 等人指出感知机无法学习异或操作，且无法处理线性不可分的问题^[36]，使神经网络的研究开始进入瓶颈期。直到 1986 年 Hinton 等人提出了反向传播 BP 算法^[37]，才使得神经网络进一步发展。与此同时 Quinlan 于 1986 年提出了决策树模型^[38]，也是机器学习一种常见的方法。1995 年 Cortes 等人提出了支持向量机模型^[39]，有坚实的可解释性，而且也有非常出色的效果，使 2010 年以前支持向量机一直占据着机器学习的主流算法。2006 年 Hinton 等人提出了深度信念网络^[40]，通过逐层训练的方法解决了网络过深梯度消失等问题，并提出了深度学习的理念。2012 年，Hinton 等人提出的 AlexNet^[41]在 ImageNet 图像识别大赛上获得了冠军，深度学习开始进入到了高潮的阶段。

2.2.2 学习算法

机器学习领域中的学习算法，希望是能从数据中学习经验，并能够根据学习的内容进行预测。通常的人工智能算法分为两种模式，一种即符号主义中根据逻辑与知识库有预先设计的方案，例如区分数字 1 与数字 7，只需要判断 1 和 7 中是否有过长的横线。而另一种即为连接主义，尽管人可以很容易区别猫与狗，但是人很难去描述特征去辨别猫与狗，因此这种情况中我们使用已有的一些样本去学习特征，学习方法，这就是学习算法。

学习算法中，几个重要的部分为，模型的任务是什么？何种评价指标来评估模型？

经验？需要通过模型的任务定义模型，是分类，回归，映射或者图像修复，去噪等等。只有定义了正确的任务，才能让模型有效。其次，如何说明该模型是有效的，需要一个合理的评价标注，例如分类任务上常采用准确率，回归问题上常采用方差或精度，图像修复上常采用相似度进行度量。最后是经验，经验可以理解为在数据集中获取一定的信息，通过该信息改善模型。总之，可以这样理解：通过某经验，可以使模型在任务上取得了更好的评价指标。

2.2.3 模型评估

对于模型的性能评估，有利于证明模型的泛化能力，是模型学习中很重要的一部分。本节将列举出几个常见的模型评估函数，我们定义模型用 f 表示。

在回归任务中，最重要的便是预测结果与真实结果之间的差异，可以通过方差来表示，最常见的性能度量便是用均方差损失函数，一般希望均方差越小越好，也常常被用于模型的损失函数，如式 2.3：

$$E = \frac{1}{m} \sum_{i=1}^m (f(x_i) - y_i)^2 \quad (2.3)$$

而在分类任务上，常用精准度来度量模型，一般希望精准度越高越好，如式 2.4：

$$acc = \frac{1}{m} \sum_{i=1}^m \mathbb{I}(f(x_i) \neq y_i) \quad (2.4)$$

2.2.4 过拟合与欠拟合

通常我们在模型的学习中都期望得到一个良好的学习模型，使得其有更强的泛化能力。所谓泛化能力即为模型在一个领域中表现良好，而在任务相同的其他几个领域依然表现良好，例如我们平时努力学习表现良好，而期末考试一样表现良好。然而模型训练中常常出现两种训练不满意的情况，即欠拟合与过拟合，可用图 2.6 表示。欠拟合情况下，模型对数据的拟合度还存有一定的空间，一般可以采用更换模型或者增加迭代次数的方式解决欠拟合问题。过拟合即模型对经验数据的拟合度过高，而在其他数据上验证时可能效果会变差。例如，一个同学平时学习刻苦，将往年的考试试卷细细地做了一遍，但是到了期末考试阶段，由于思维还在往年卷中，而期末试卷的题型发生了较大的变化，从而使考试成绩不理想，这就是典型的过拟合。过拟合对模型泛化能力的影响非常大，下一节列举一些常见正则化方法。

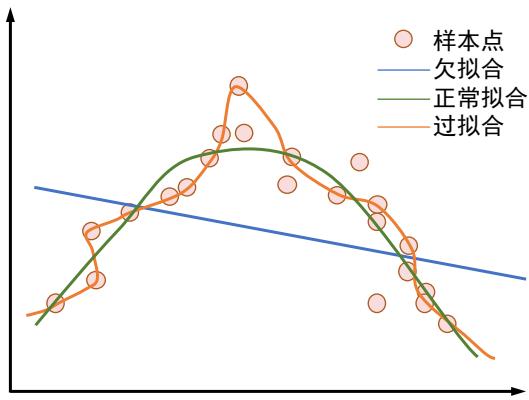


图 2.6 欠拟合与过拟合示意图

2.2.5 正则化与交叉验证

我们在机器学习领域常用的预防过拟合的方法包括正则化与交叉验证。正则化是一种使风险最小化的策略，常常与模型复杂度呈正相关。正则化的基本形式如式 2.5:

$$\min \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i)) + \lambda J(f) \quad (2.5)$$

其中， f 代表机器学习模型， L 代表经验风险，也可理解为损失函数，而 J 代表了正则化项， λ 为一超参数。一般常用的正则化项包括 L_1 范数和 L_2 范数，如式 2.6 和 2.7:

$$L(\omega) = \frac{1}{N} \sum_{i=1}^N (f(x_i; \omega) - y_i)^2 + \lambda \|\omega\|_1 \quad (2.6)$$

$$L(\omega) = \frac{1}{N} \sum_{i=1}^N (f(x_i; \omega) - y_i)^2 + \frac{\lambda}{2} \|\omega\|^2 \quad (2.7)$$

其中 ω 为模型的参数，通过限制模型参数的大小以达到正则化的效果。

交叉验证也是验证模型防止过拟合的一种手段，这里我们介绍两种交叉验证方式，一种为简单的交叉验证，另一种为 S 折交叉验证。简单的交叉验证，即在训练模型时，需要经验数据，我们将经验数据可分为 70% 训练集，15% 验证集，15% 测试集。在模型学习阶段，我们使用训练集用以更新模型的参数，每迭代更新一轮参数，我们将模型在训练集与验证集上各进行一次模型评估，并记录下数值，等迭代结束，该训练数值的集合即称之为学习曲线，我们可以根据学习曲线上训练集与验证集的走向，选择一个合适的点保存参数，然后在测试集上验证，3 者效果都好则为合格的模型，学习曲线如图 2.7 所示。

另一种 S 折验证法，主要是针对数据集量过小的情况，可能存在数据分布不完善，

这时便可以使用 S 折验证法，将数据随机分为 S 份互不相交的数据集，每次选择其中的一个作为测试集，其他的作为训练集。训练与测试 S 次，最终将 S 次模型评估结果作为最后的结果，即为模型的大致情况。

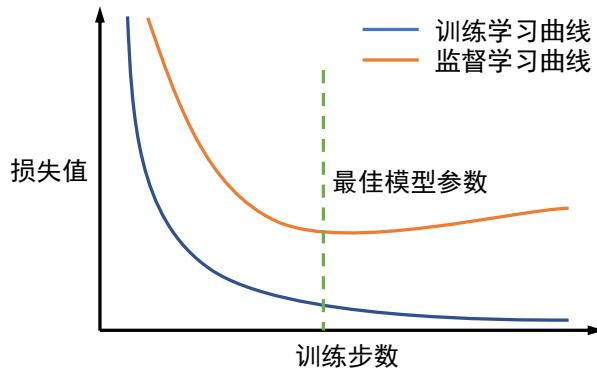


图 2.7 交叉验证学习曲线及模型参数选择

2.2.6 神经网络

神经网络的发展历史已在 2.2.1 节中大致描述，现代的神经网络种类很多，例如卷积神经网络，循环神经网络，Hopfield 网络，深度信念网络等等。这里我们主要讲述多层次感知机与反向传播算法的基础理论推导。感知机均模仿神经元的结构，如图 2.8 所示为神经网络常用的神经单元，其中 x 为输入。

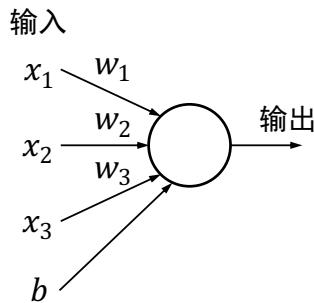


图 2.8 神经元的一般结构

最早感知机的输出取决于设定的阈值，大于阈值为 1，小于阈值为 0，如式 2.8 所示：

$$out = \begin{cases} 0 & \text{if } output \leq thresh \\ 1 & \text{if } output > thresh \end{cases} \quad (2.8)$$

而这样的缺点，即线性不可分问题，因而后面提出了一种改善神经网络非线性的方法。最常见的便是在输出层上用 Sigmoid 函数改善，如式 2.9 所示：

$$out = \frac{1}{1 + \exp(-\sum_j \omega_j x_j - b)} \quad (2.9)$$

Hornik 等人的研究表明^[42], 用多层与足够数量的神经元可以逼近任意的非线性连续的函数。一般我们将神经网络中数据输入的层定义为输入层, 中间层定义为隐藏层, 输出结果的层定义为输出层。图 2.9 展示了多层感知机(神经网络)的一般结构。

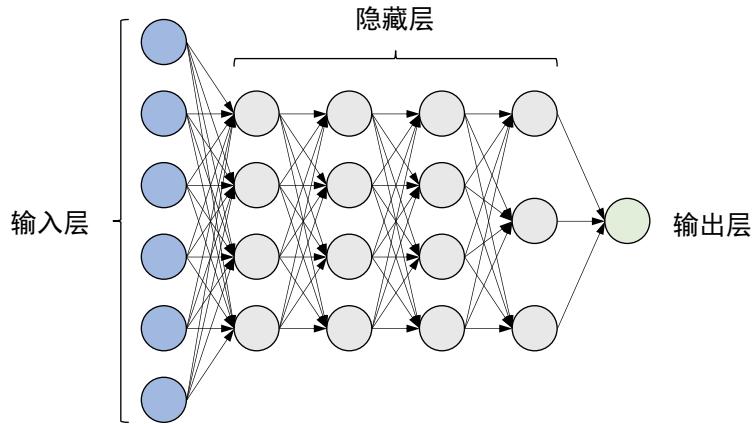


图 2.9 多层感知机的一般结构

假设我们的输入是向量 x , 某一层的输出是向量 y , 则多层感知机的前向传播我们可以写为式 2.10:

$$y = \sigma \left(\sum_j \omega_j x_j + b \right) \quad (2.10)$$

其中 σ 代表改善神经网络非线性的函数, 也称为激活函数, 例如 Sigmoid 函数。当我们需要计算从输入层至输出层的过程, 也称为前向传播过程, 按式 2.10, 每次计算得到的输出值, 将作为下一层的输入值, 直至输出层。总之我们可以定义为, 从输入层到输出层的计算过程为神经网络的前向传播过程。

由于初始化时网络的参数往往不能达到预期的成果, 我们需要用一些经验, 也就是数据改善模型的参数。改善的方法有很多种, 例如基于概率的贝叶斯估计, 本文中主要以基于梯度的方法, 梯度下降法。

首先, 我们要依据评价方式定义一个代价函数, 或称为损失函数, 这里我们采用二次代价函数, 也就是均方差损失函数, 如式 2.11 所示:

$$J(\omega, b) = \frac{1}{2n} \sum_x \|y(x) - a\|^2 \quad (2.11)$$

其中 a 为 x 向量所对应的真实数据, 这样我们每次运算前向传播, 将结果与真实数据做差时, 即可得到损失值, 我们就可以以损失值的最小化为方法来优化模型。在针对更新某参数 ω 时, 多层的数据如何更新? 假设只是一层的隐藏层, 则 ω 关于 J 的导数为 $\frac{\partial J}{\partial \omega}$, 而遇

到多层时，由于根据导数的性质 $\frac{\partial J}{\partial \omega} = \frac{\partial J}{\partial h} \cdot \frac{\partial h}{\partial \omega}$ ，即我们可以根据链式法则，每一层的数据的输出与输入间均求一下导数，然后将这些导数全部乘起来，即为多层次下的导数计算，根据梯度更新这些权重，我们就可以快速让损失函数降下来，以此达到使模型有更好的评估效果。

现在我们已经可以根据参数计算损失函数的导数，如何更新？这里引入一个新的参数，学习率。假设我们已知损失函数对参数的导数，即梯度，我们便知道了下降的方法。但是我们仅知道在此时参数下的损失情况，并不知道周围的参数变化时的损失情况，因此我们需要根据梯度慢慢推进，这便是学习率的要求。学习率不能设置过大，如果过大可能导致损失不减反升，也不能设置过小，否则会导致模型更新速度很慢。我们用 η 来表示学习率，则梯度下降更新规则可如式 2.12 所示：

$$\omega = \omega - \eta \frac{\partial J}{\partial \omega} \quad (2.12)$$

我们可以将整个梯度下降算法总结为表 2.2：

表 2.2 梯度下降算法

梯度下降算法
输入：训练集 $D = \{(x_i, y_i)\}_i^m$
输入：学习率 η
输入：网络模型 f
输入：训练轮数 epoch
1: 随机初始化网络模型 f
2: for i=1-epoch
3: for all (x_i, y_i) in D
4: 根据式 2.11 计算损失值
5: 计算各个变量关于损失值的梯度
6: 根据式 2.12 更新参数
7: end
8: end
9: 将模型参数保存下来
输出：模型的参数

2.3 计算机视觉深度学习应用

2.3.1 卷积神经网络

卷积神经网络是神经网络的一种，在计算机视觉与语音识别领域取得了巨大的成就。最早卷积神经网络相关的卷积与池化等理念在 1982 年 Fukushima 等人^[43]提出，而 1989 年 LeCun 等人^[44]实现了一个完整的卷积神经网络雏形，将卷积等操作与反向传播结合起来。而 1998 年 LeCun 等人^[45]提出了 LeNet 网络，充分定义了卷积神经网络的各个部

分，并成功用 ReLU 激活函数代替了 Tanh 激活函数，加快了模型训练时的收敛速度。而 2012 年 Krizhevsky 与 Hinton 等人^[41]提出的 AlexNet 成功在大规模分类数据集 ImageNet^[46]上取得了巨大的突破，也使得深度学习开始进入高潮。此外 2014 年 Simonyan 等人提出了 VggNet^[47]，说明了网络深度的重要性，同年谷歌提出了 Inception 的结构^[48]，也是倾向于加深网络的深度以提高模型的表达。但是由于链式发展的存在，网络的深度越深，误差难以从输出层传播到输入层，存在梯度弥散的问题。2015 年何凯明等人提出了深度残差学习的结构^[49]，通过残差学习有效的解决了网络过深梯度消失的问题。此后深度学习又有各个方面的发展，例如注意力模块，轻量级等等。本节主要讲述经典的卷积神经网络常见的模块内容。

卷积神经网络的组成常常是前面跟着一堆各式各样的卷积层，最后跟上全连接层用于具体细节的划分。

(1) 卷积

图像处理中常常使用一个 $w \times h$ 的算子，用于对图像进行滤波，提取边缘等操作，这些算子通常认为固定的一个值。而卷积操作也有类似的前向传播形式，只是此处的算子是一个可学习的滤波器。说到滤波器我们必须提及一下二维卷积与图像卷积两个概念，二者本质是不同的，二维卷积本质主要是信号的采样，而图像卷积的本质是滤波。通常我们称卷积操作中的算子变量为卷积核、滤波器等，卷积操作中每一层卷积层的卷积核不一定只有 1 个，通常卷积操作中有很多的卷积核，每个卷积核除了 $w \times h$ 的大小外，还有通道数 c ，取决于输入块的维度，以及核数量 n ，这个通常是人为设定的。

而卷积核的尺寸我们通常称为卷积核大小，也称为感受野的大小，是一个很重要的参量，通常人们会使用 3×3 的卷积核。与图像滤波操作基本相同，除了卷积核数量和感受野外，卷积仍存在两个可选择的参数，即步长和填充。步长即卷积核在 w 和 h 方向依一次性移动的像素格，通过控制步长可以控制输出尺寸，而填充即在图像周围补零，目的是为了使卷积前后输出的尺寸大小不变（非通道长度）。卷积的过程如图 2.10 所示。

假设我们输入层的尺寸为 $W \times H$ ，有 C 个通道。卷积核的尺寸为 $w \times h$ ，通道数为 c ，核的数量为 n 个。假设卷积操作的步长在 w 和 h 方向上我们均设置为 s ，填充均设置为 p ，则输出层的大小 W' 、 H' 如式 2.13 所示：

$$\begin{aligned} W' &= \frac{W - w + 2 \times p}{s} + 1 \\ H' &= \frac{H - h + 2 \times p}{s} + 1 \end{aligned} \tag{2.13}$$

输出层的通道数取决于卷积核的数量，每一个卷积核卷积操作后将产生 1 个通道的输出，将这些输出从通道方向全部叠加在一起即为卷积操作。

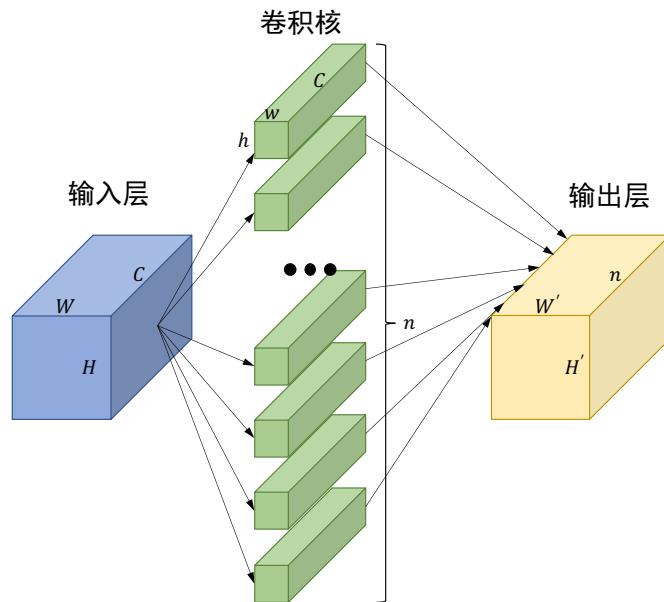


图 2.10 卷积的过程

(2) 池化

卷积操作将有大量的参数，需要一种降维的方式改善需要的大幅度参数，而保证其特征的不变性。由于图像有局部空间性，因此我们引入一种操作为池化操作的方式。池化一般设置一个通道为 1 的窗口，与卷积相同，均存在步长和填充。输入层每个通道都与池化窗口操作。具体分为平均池化与最大池化，平均池化通过计算窗口覆盖面上数值的平均值，而最大池化计算覆盖面上数值最大的值，以此来减小参数量的大小。池化操作如图 2.11 所示。

假设输入层尺寸为 $W \times H$ ，池化窗口的尺寸为 K ，步长为 S ，则池化后的尺寸大小如式 2.14 所示：

$$W' = \frac{W - K}{S} + 1$$

$$H' = \frac{H - K}{S} + 1$$
(2.14)

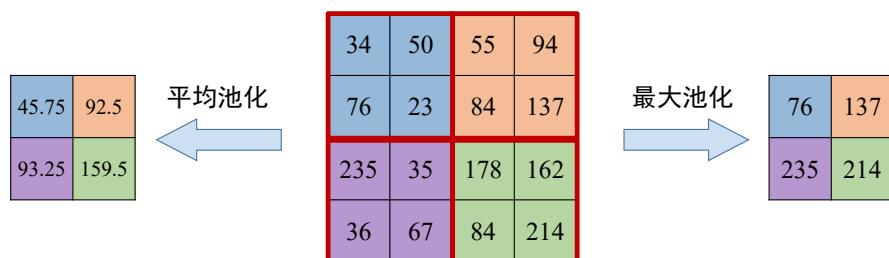


图 2.11 池化过程

(3) 激活函数

激活函数在 2.2.6 中有提及 Sigmoid 函数，我们将讨论各种新式常用的激活函数。激活函数的种类很多，原始的有 Sigmoid 和 Tanh，而后 AlexNet 通过修正线性单元改善了网络训练的收敛速度，同时也产生了很多种激活函数。这里我们介绍 ReLU^[50]、Elu^[51]、Selu^[52]、Softplus^[53]和 Softmax^[54]，具体如表 2.3 所示。其中 Softmax 激活函数常常用于分类任务时使用。

表 2.3 各种激活函数的表达式

激活函数	公式
Sigmoid	$\frac{1}{1 + e^{-x}}$
Tanh	$\frac{e^x - e^{-x}}{e^x + e^{-x}}$
ReLU	$\begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$
Elu	$\begin{cases} x & \text{if } x > 0 \\ \alpha(e^x - 1) & \text{if } x \leq 0 \end{cases}$
Selu	$\lambda \begin{cases} x & \text{if } x > 0 \\ \alpha(e^x - 1) & \text{if } x \leq 0 \end{cases}$
Softplus	$\log(1 + e^x)$
Softmax	$\frac{\exp(x_i)}{\sum_j \exp(x_j)}$

(4) 损失函数

损失函数对于模型的梯度获取至关重要，常见情况下我们常按任务分配损失函数。例如在回归任务中我们常用均方差损失函数如式 2.3 所示。而针对分类任务，我们常常使用交叉熵损失函数，如式 2.15 所示：

$$L = -\frac{1}{m} \sum_{i=1}^m [y_i \ln f(x_i) + (1 - y_i) \ln(1 - f(x_i))] \quad (2.15)$$

其中 m 为批量， y 为真实的标签， f 为模型， x 为模型的输入。

此外，在一些判断图像或向量的相似度上，通常用 L_1 距离和 L_2 距离，如式 2.16 和 2.17 所示：

$$L = \|f(x) - y\|_1 \quad (2.16)$$

$$L = \|f(x) - y\|_2^2 \quad (2.17)$$

(5) Dropout

为了解决模型的训练中存在的过拟合问题, Google 提出了一套训练使用的预防过拟合的方法, 被称为 Dropout^[55]。通常将 Dropout 用于全连接层, 即在前向传播与反向传播时移出部分的节点。我们假设一个模型的输入为 n 的向量, 神经网络的权重均相同, 则输入层每一个节点的梯度均相等, 更新时将一直保持不变。而引入 Dropout 后, 部分节点消失, 可以使各个输入节点的梯度发生变化导致不同, 从而使参数变化开来。

(6) Batch Normalization

一般在模型训练时, 通常会计算数据集中图像各个通道的像素平均值, 在训练模型时先将各个通道的数值减去平均值, 这是一种老方法。而现在常用一种强正则化的方法, 可以替代 Dropout、 L_1 与 L_2 的正则化方法, 为 Batch Normalization^[51]。不仅可以加快训练, 也可以省去负责的预处理。Batch Normalization 通常放在卷积层中, 例如卷积层之前与卷积层之后。该模型主要取决于输入数据的通道, 因此 Batch Normalization 的性能通常对输入批量有要求, 具体的公式表达如式 2.18 所示:

$$BN(x) = \gamma \left(\frac{x - \mu(x)}{\sigma(x)} \right) + \beta \quad (2.18)$$

其中 $x \in \mathbb{R}^{N \times C \times H \times W}$, 即也可以理解为在单独的某像素的通道上做归一化。 $\gamma, \beta \in \mathbb{R}^C$, μ 代表均值, σ 代表方差。如式 2.19 和 2.20 所示:

$$\mu_c(x) = \frac{1}{NHW} \sum_{n=1}^N \sum_{h=1}^H \sum_{w=1}^W x_{nchw} \quad (2.19)$$

$$\sigma_c(x) = \sqrt{\frac{1}{NHW} \sum_{n=1}^N \sum_{h=1}^H \sum_{w=1}^W (x_{nchw} - \mu_c(x))^2 + \epsilon} \quad (2.20)$$

2.3.2 经典卷积神经网络结构

卷积神经网络结构对模型的性能影响至关重要, 例如 Woo 等人曾提出过卷积神经网络影响模型性能的几种结构包括深度、宽度与基度^[56]。本节将描述几种最为经典的卷积神经网络结构。

首先是最早的 VGGNet, 其采用了增加模型深度的方法提高了模型的识别率。VGGNet 的网络层数增加到了 16 层与 19 层, 每经过一次池化操作的下采样, 下一层的卷积核个数将是上一层的 2 倍, 且只采用了 3×3 感受野大小的卷积核。VGGNet 的基本

网络结构图如图 2.12 所示，这里要说明一下，VGGNet 的结构很多，基本满足卷积层均为链式的卷积-池化即可认为是 VGGNet 类型，也是最常见类型，这里我们展示的是原始论文中的 VGG16 网络模型。

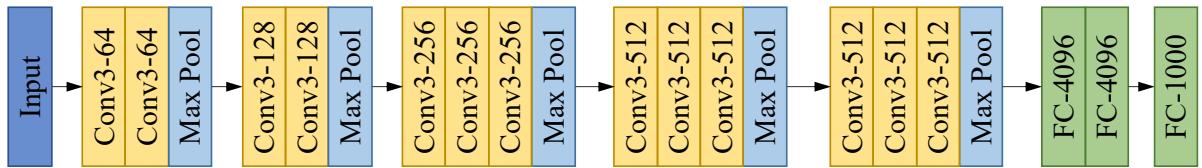


图 2.12 VGG16 网络的基本结构

GoogLeNet 是一种提升网络宽度，增加稀疏参数的网络，其已经有 4 个版本的网络结构了，本节将介绍最简单的 Inception v1 结构。Inception 结构在每一个模块输入后，将分出四路，每一路用不同的卷积方案进行计算，最后将计算图由通道方向连接在一起，Inception v1 结构如图 2.13 所示。最终将多个 Inception v1 模块叠加在一起，组合成了最早的 GoogLeNet，其在 3 个不同的深度输出分类的模型，有利于模型的深度选择。

我们还有必要提一下 Inception v4^[57]，他提出了一种新的卷积组合操作，使模型的感受方式得到了改善。例如我们经常使用 3×3 感受野大小的卷积核，Inception v4 提出了一种新的组合方式，即取代 3×3 ，使用连续的 1×7 和 7×1 卷积核。这样的好处是可以提取横向与纵向两个方向的特征，而没有横向与纵向的特征耦合情况。这样我们通常可以对输入层采用 1×1 的卷积进行降维，然后从一个分支其提取主要的横向与纵向的特征，在用 3×3 的核进一步融合，增强特征的表达，这是 Inception v4 最大的特色。其余的部分结构参考了其他的网络结构，这里不再进行叙述，Inception v4 由很多的模块组成，其中最具特色一种模块如图 2.14 所示。

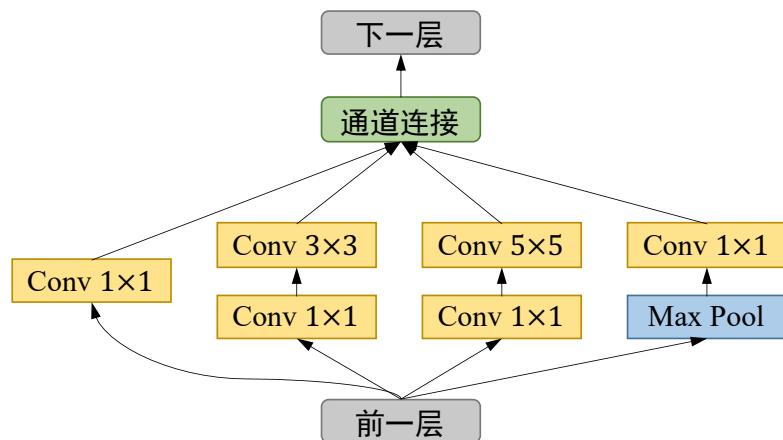


图 2.13 Inception v1 结构

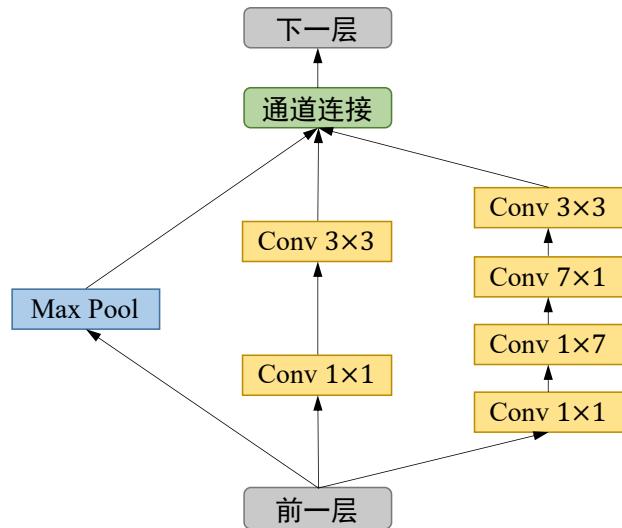


图 2.14 Inception v4 一种模块的结构

还有一个里程碑式的网络结构，即 ResNet。由于网络的反向传播的链式法则，梯度的传播会越来越小，即使增大前面的梯度，也会导致后面梯度爆炸的情况。这种情况使卷积神经网络的深度无法一直扩展，直到残差结构被提出，常见的残差结构如图 2.15 所示。

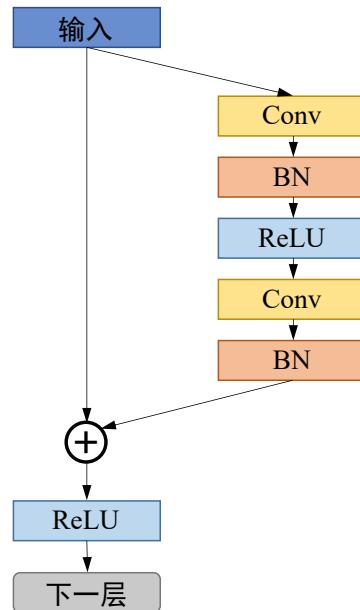


图 2.15 ResNet 的残差模块

以上便是几种最经典的网络结构，其余很多提出的网络结构都有借鉴于以上几种网络，也是最基础的基线。

2.3.3 基于梯度的优化方法

基于梯度的优化方法，有很多种，本节将主要介绍几种。式 2.12 已经提及了梯度下

降算法，而其他的基于梯度的优化方法，本节的内容主要参考了 Runder 等人整理的综述文献^[58]。

首先是批量梯度下降算法，通常我们在运算模型时，不会依次地输入参数，而是批量的输入参数，有利于加快模型训练的速度，也有利于类似 Batch Normalization 的正则化方法等。假设模型的参数用 θ 表示， η 表示学习率， J 表示损失函数，则批量梯度下降可以用式 2.21 表示：

$$\theta = \theta - \eta \nabla_{\theta} J(\theta) \quad (2.21)$$

然而批量的梯度计算可能在大型数据集上存在一个问题，批量数据中部分不利于模型的学习。这里引入了随机梯度下降，也称为 SGD，一次性只输入一个样本数据，可以加快模型的训练，随机梯度下降可以用式 2.22 表示：

$$\theta = \theta - \eta \nabla_{\theta} J(\theta; x^i; y^i) \quad (2.22)$$

小批量梯度下降，我们可以在随机梯度下降的基础上，引入小批量梯度下降算法，扩展后 n 个数据集，这么做可以防止数据集中过多的噪声，同时使模型的训练过程更加地稳定。小批量梯度下降算法可用式 2.23 表示：

$$\theta = \theta - \eta \nabla_{\theta} J(\theta; x^{i:i+n}; y^{i:i+n}) \quad (2.23)$$

然而，模型极有可能陷入狭窄的局部最优，这里模仿物理模型的冲量，冲出模型的沟壑区，称为 Momentum，如式 2.24 所示：

$$\begin{aligned} v_t &= \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta) \\ \theta &= \theta - v_t \end{aligned} \quad (2.24)$$

固定的学习率往往不利于模型的长期训练，因此有的工作会引入模型的学习率衰减函数，这里我们介绍一种自适应学习率的优化方法，Adam。主要通过计算过往的梯度与指数的衰减信息而更新参数，具体算法实现如下：

$$g_t = \nabla_{\theta} J(\theta_{t-1}) \quad (2.25)$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (2.26)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (2.27)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (2.28)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (2.29)$$

$$\theta_t = \theta_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (2.30)$$

β_1 和 β_2 为两个超参数，一般被设置为 0.9 和 0.999。SGD 与 Adam 是两种我们最常用的优化方法。在绝大多数的模型中，一般默认使用 Adam 优化算法，Adam 的缺点就是没有 SGD 的快速收敛性，当然现在也有很多的优化方法提出，将 SGD 与 Adam 的优点结合，这里我们不再叙述。

2.4 本章小结

本章我们总结了图像处理与机器学习的基础理论知识，也就是我们在全篇的技术理论中的基础理论知识都将是基于本章节中提出的一些要点。图像处理部分往往与本文的一些预处理操作有关，机器学习的知识与我们本文中很多的模型相关。因此与本章节技术相关的知识在后面几章中将不再细写，后面的章节将主要叙述基于本章理论的具体的创新性实现。

3 端到端图像称重

正如第一章节中所提出的问题，传统的方法中，仅靠提取部分的特征难以对物体进行单个体精准测量，因为部分特征只代表了整个样本的趋势，而并不代表个体的特性。如要选择更加精准的单个物体重量识别，必须增加有效的特征数量。

在模式识别领域中，特征的提取往往适用于形态有一定规则的图形，例如阿拉伯数字，有明显的特征。但是对于大规模无明显规则数据，例如 ImageNet^[46]中存在大量的人难以用特征描述的类。针对于鸭子与体重之间的关系，为连续的值，人也难以辨分与之强相关的特征，因此手工提取有效特征的方法很困难。而卷积神经网络(CNN)的卷积层则可以有效地自动提取特征，免去了手工提取特征的方式，因此本工作将围绕卷积神经网络展开研究。

本章中，我们针对完整的鸭子图像，尝试将 CNN 引入其中自动提取特征并预测体重。由于需要精准地预测体重，如果我们依靠分类的方式，即设置很多的区间，则不仅会增大数据集规模要求，也会增大算法的负担，同时难以达到相当的精度。因此本章从连续的角度出发，采用回归的方式用于预测鸭子的体重。

3.1 数据集与预处理

3.1.1 数据集采集

本文的数据集采集自一家养鸭的屠宰场，中国山东省红叶食品有限公司。在距离流水线约2米前架设了一架单目摄像机，正对着流水线，现场的环境以及摄像机的安装示意图如图3.1所示。安装摄像机时，保证鸭子出现在摄像机的正中央。初步选择的场景采集的图像完整性较高，鸭子的两只脚被悬挂于铁架上，除了翅膀部分有些交叉，身体主体部分完全的独立出来。为了保证图像的质量尽可能减少噪声干扰，我们在生产线背面设置蓝色背景的底色，生产线以每秒0.5米的速度运行着，相机以每秒30帧的速度拍摄视频。这些图片包括50只鸭子整齐地挂在生产线上。对每只鸭子，使用150张不同角度的图像来估计重量。我们用电子天平给这些鸭子称重作为标签用于评估我们的模型。

数据集包括50只鸭子，每只鸭子都有150张来自不同角度的图像。数据集分为训练集、验证集和测试集。45随机选择鸭子的图片设置为训练集，其余五个鸭子，20%的图像被随机选中作为验证集，其余80%图像作为测试集。训练集用于优化网络的参数，验证集是用于在模型训练过程中验证网络的精度和泛化能力，测试集用于测试训练后的网络

的准确性，最终评估网络的准确性和泛化能力。

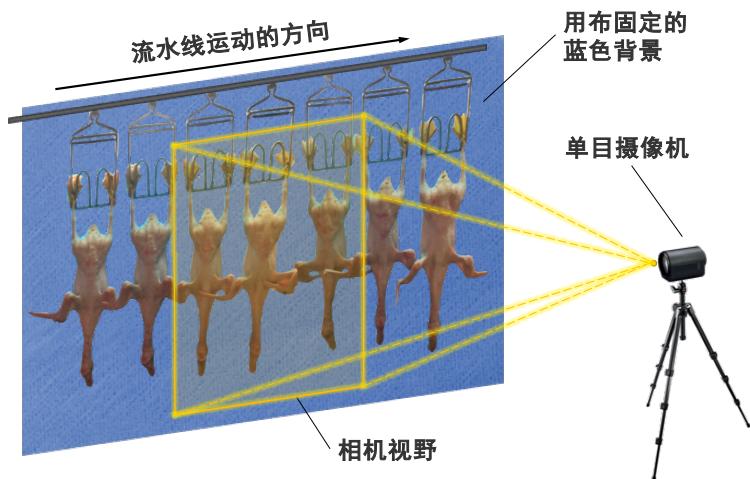


图3.1 图像采集设备的部署与现场环境，环境为简单场景

3.1.2 数据的预处理

为了能单只高精准预测鸭子的体重，我们将鸭子从视频中分离出来。分割的准则即以铁架为参照物，铁架下300像素距离为整只鸭子，这里不再叙述切割的过程。我们主要讨论针对切割下来的鸭子图像，如何做预处理。首先，将低于阈值的像素设置为零，以去除背景噪声。如图3.2所示，图像的R(红色)通道包含的形态学信息最多，鸭子与背景的对比度最好。为了降低模型的计算水平，我们提取了每幅图像的R通道进行进一步分析。小于40的像素值设置为0，大于40的像素值设置为255。除了duck区域中的信息外，还可以通过使用值为1的 5×5 内核进行开运算来删除。考虑到鸭子的脚基本上都是一样的重量，它们的脚被排除在图像之外。铁框下方300像素的部分被用作鸭子图像。提取的单个鸭子的图像从 1015×321 缩小到 253×80 像素，并标记它们的重量。将像素值进一步归一化为[0, 1]。图像切割的结果如图3.3所示，分有腿与无腿两种情况。

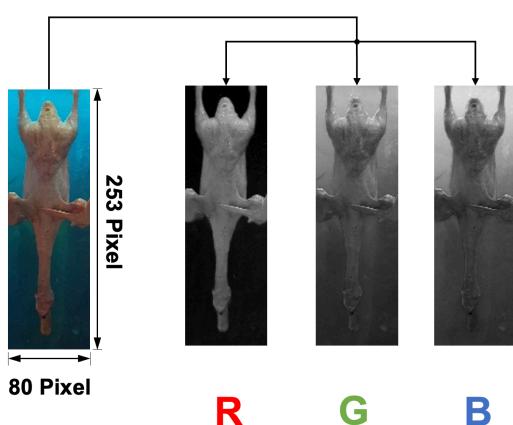


图3.2 鸭子图像的RGB三通道示意图

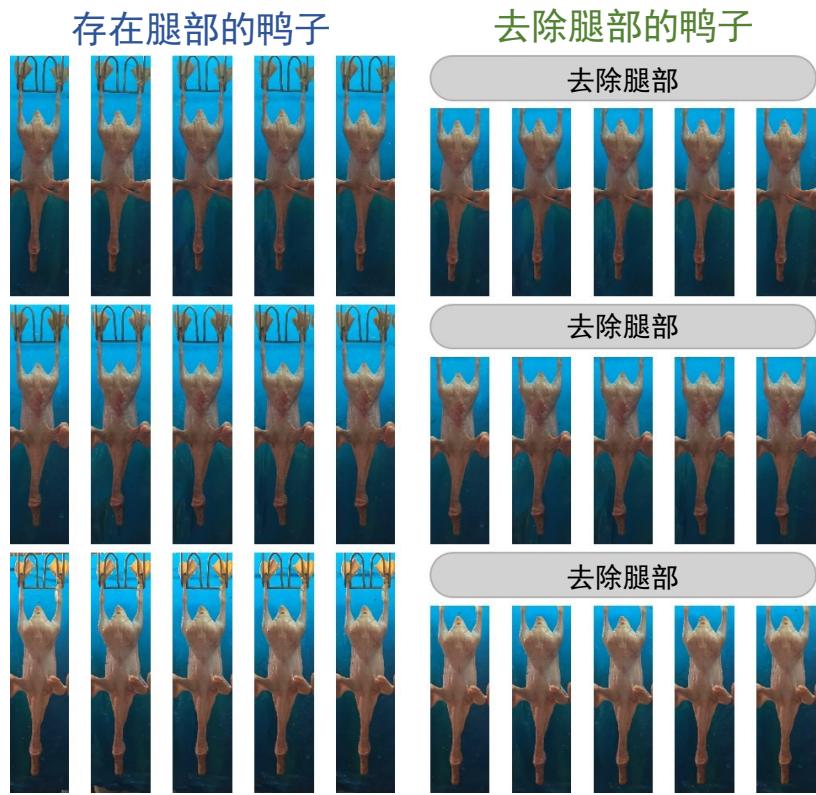


图3.3 将鸭子从视频中分离的结果。每一行都是相同的鸭子，左侧为存在腿部的鸭子，右侧为去掉腿部的鸭子

3.1.3 设备与编程环境

本章的主要工作中，提出的模型使用Python语言（3.7版本）和TensorFlow框架（1.15.0版本）构建，模型的计算与优化在一张英伟达GTX 1080Ti显卡上操作。对比实验中，复现传统方法的程序，包括基于面积的拟合和基于ANN的拟合模型使用MATLAB（2019A版本）。

3.2 卷积神经网络回归模型

我们的模型基于卷积神经网络(CNN)构建回归任务模型。该网络的模型的结构如图3.4所示，一个训练好的模型同时可视化了某一只鸭子作为输入时，预测体重过程的中间结果。模型的输入是一个 253×80 的单通道矩阵，经过归一化，预处理方法如3.1.2节中的描述。我们设计的模型从轻量级考虑，采用最简单的VGG网络^[47]的模块，包括3层卷积层和2个全连接隐藏层，1个输出层。每一层卷积层使用的卷积核的感受野大小为 3×3 ，步长为1，且卷积操作无填充。每一层卷积层由softplus^[53]激活函数激活，并附跟1层最大池化层，池化层由 2×2 的池化窗口，步长为2，无填充。两层全连接层分别含有32与16个节点，并由softplus激活函数激活。输出层仅含有1个节点，无激活函数。

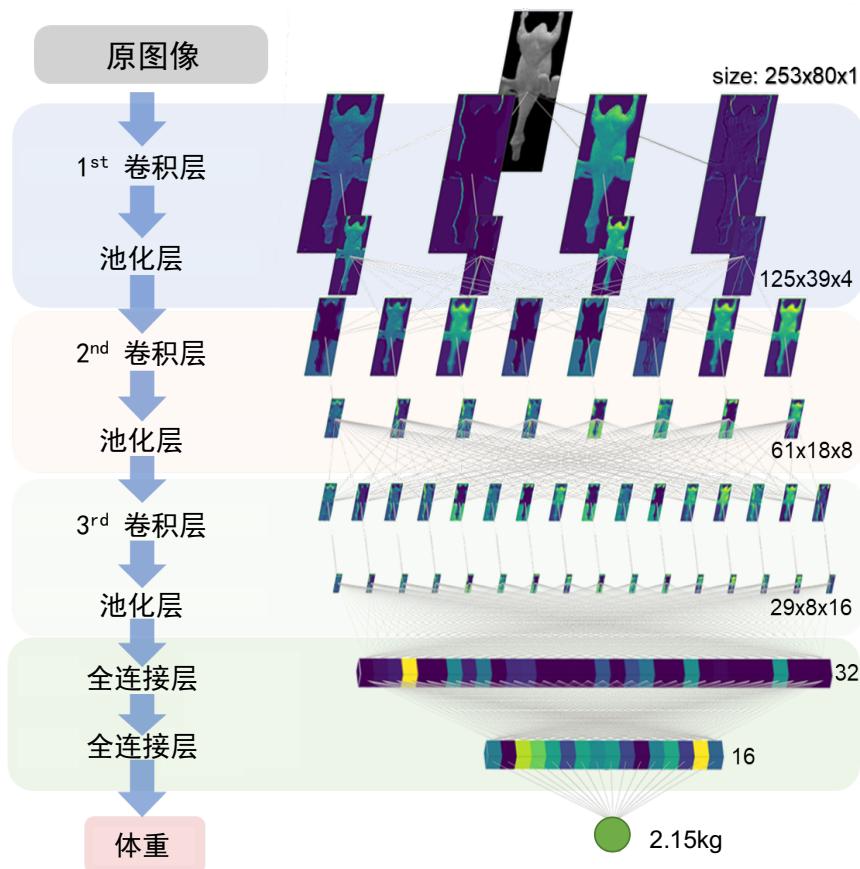


图3.4 卷积神经网络回归模型结构及可视化

模型训练时使用小批量输入计算的梯度，并用Adam^[59]优化方法进行优化，损失函数采用均方差损失函数(Mean square error, MSE)，如式3.1所示：

$$l = \frac{\sum_{i=1}^{N_B} (y_{pred_i} - y_i)^2}{N_B} \quad (3.1)$$

其中 y_{pred_i} 表示模型预测的第*i*只鸭子的体重， y_i 表示真实的标签， N_B 表示批量输入的大小。为了加快模型的训练过程，我们在训练模型时引入了衰减学习率，如式3.2所示：

$$lr = lr_{initial} \times dr^{\frac{gs}{ds}} \quad (3.2)$$

其中 $lr_{initial}$ 表示初始学习率，设为 10^{-5} ； dr 表示衰减率，设为0.99； ds 和 gs 分别表示衰减步数与总步数，其中衰减步数设为1000，即每迭代训练1000步衰减一次学习率。

3.3 结果与验证

3.3.1 超参数优化

在本节中，我们讨论了不同超参数对模型精度的影响。我们使用的超参数包括卷积层的深度、卷积核大小、激活函数、dropout^[55]和模型优化方法。激活函数和优化方法是

最主要的影响因素。使用测试数据集上的误差分布来展示激活函数对模型预测的影响。Adam优化方法是目前唯一能很好地拟合训练数据集的方法，通过学习曲线验证了其对模型精度的影响。我们发现softplus^[53]激活函数和Adam^[59]优化方法可以有效地拟合数据集。利用学习曲线分析了卷积层深度、卷积核大小和dropout^[55]对模型精度的影响。本文采用smooth spline方法对学习曲线进行了平滑处理。

(1) 卷积层的深度：

卷积层的深度是CNN模型性能的关键影响因素^[47]。为了优化回归结果，我们测试了8种不同的CNN结构，如表3.1所示。第一卷积层我们均设置4个卷积核，对于下面的卷积层，每经过一次下采样，下一层的卷积核个数将是上一层的2倍。初始化卷积核采用正态分布初始化方法。使用softplus^[53]激活函数激活所有的卷积层，并使用最大池化。所有这8个网络结构模型都采用相同的全连接网络结构。训练数据集的学习曲线如图A1 a所示，说明所有的网络都很好地拟合了训练数据集。验证数据集的学习曲线如图3.5 a所示，3层CNN模型和5层CNN模型在验证集上的MSE最低，考虑轻量级因素，我们最终选择了3层的CNN模型。

表3.1 测试的8种卷积神经网络的结构，卷积层参数表示为“Conv-<卷积核数量>”，全连接层参数表示为“Dense-<节点数量>”

2层	3层	4层	5层	10层	10层 (2)	13层	16层
Conv-4	Conv-4	Conv-4	Conv-4	Conv-4 Conv-4	Conv-4 Conv-4	Conv-4 Conv-4	Conv-4 Conv-4 Conv-4 Conv-4
Conv-8	Conv-8	Conv-8	Conv-8	Conv-8 Conv-8	Conv-8 Conv-8	Conv-8 Conv-8	Conv-8 Conv-8 Conv-8 Conv-8
Max Pool							
Conv-16	Conv-16	Conv-16		Conv-16 Conv-16 Conv-16	Conv-16 Conv-16 Conv-16	Conv-16 Conv-16 Conv-16	Conv-16 Conv-16 Conv-16 Conv-6
---	---	Conv-32 Max Pool	Conv-32 Max Pool	Conv-32 Conv-32 Conv-32	Conv-32 Conv-32 Conv-32	Conv-32 Conv-32 Conv-32	Conv-32 Conv-32 Conv-32 Conv-32 Max Pool
Dense-32 (dropout rate 0.1)							
Dense-16 (dropout rate 0.1)							
Dense-1							

(2) 卷积核大小:

CNN的卷积核大小对模型的性能至关重要。因此，我们根据权重估计的精度来评估5个核大小。在实验过程中，卷积核的数量是固定的，在一个模型中，所有卷积层的卷积核大小都是相同的。训练数据集的学习曲线如图A1 b所示，说明所有的网络都很好地拟合了训练数据集。图3.5 b为不同核大小模型在验证数据集上的学习曲线。我们选择了 3×3 的卷积核，因为它在验证数据集中MSE最低。

(3) 激活函数:

激活函数影响CNN模型的反向传播。我们测试了五个激活函数：Tanh、Relu、Elu、Selu和softplus^[53]。训练数据集的学习曲线如图A1 c所示，说明所有的网络都很好地拟合了训练数据集。如图3.5 c所示，使用softplus的模型在验证数据集中MSE最低，预测误差最小，如图3.5 d所示，因此我们选择softplus作为激活函数。

(4) Dropout:

为了避免过拟合，使用具有相同参数和结构的模型测试了四种Dropout率。图3.5 e-f比较了这些Dropout率的学习曲线。在图3.5 e中，只有0和0.1的退出率对训练数据集拟合较好，所以我们只考虑这些Dropout率。在图3.5 f中，在验证数据集上，Dropout率为0.1时的MSE最低。因此，在训练期间，Dropout率被设置为0.1。

(5) 优化方法:

Adam^[59]优化方法被认为是最好的一种方法^[58]。在此，我们采用相同的模型，比较了三种优化方法：小批量梯度下降^[58]、Momentum^[60]优化方法和Adam^[59]优化方法。在动量优化方法中，动量项 γ 设为0.9。图3.5 g-h为三种优化方法在训练数据集和验证数据集上的学习曲线。对于两个数据集，Adam优化方法的均方误差最小。因此我们选择了Adam作为优化方法。

因此，根据本节的超参数讨论，我们找到了适用于数据集的几个超参数：卷积层深度，卷积核大小，Dropout率。同时我们发现了2个影响因素较大的超参数，一个是激活函数，使用Softplus激活函数可以大幅度提高模型的鲁棒性。同时，该模型也仅在使用Adam优化方法时能够训练出拟合性较强的模型。因此我们认为Softplus激活函数和Adam优化方法能决定性影响模型的性能，而卷积层深度，卷积核大小和Dropout率在小范围内影响模型精度，需适度调整参数。

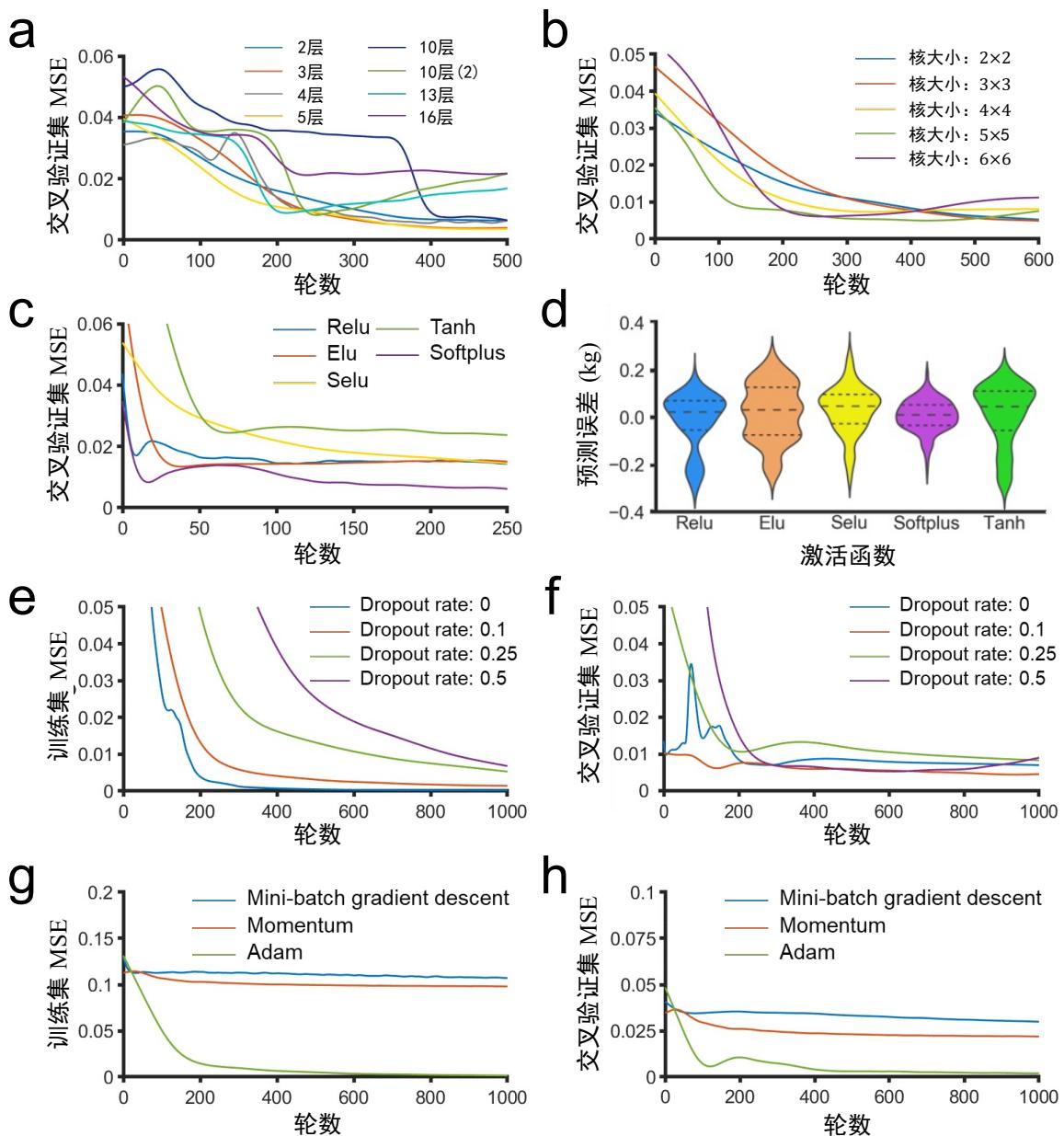


图3.5 模型在不同超参数下的学习曲线（经smooth spline平滑）与预测误差分布。a. 在不同卷积层深度条件下交叉验证集的学习曲线；b. 在不同卷积核大小条件下交叉验证集的学习曲线；c. 在不同激活函数下交叉验证集的学习曲线；d. 在不同激活函数下预测误差分布图；e. 不同Dropout率条件下训练集的学习曲线；f. 在不同Dropout率条件下交叉验证集的学习曲线；g. 在不同优化方法下训练集的学习曲线；h. 在不同优化方法下交叉验证集的学习曲线

3.3.2 评估鸭子的体重

为了评估模型的鲁棒性，我们设计了一个实验。由于我们的样本数比较少，因此我们采用了十折法交叉验证模型。我们进行了10次实验，每次随机从数据集中抽取45只鸭子作为训练集，剩余的5只鸭子中，从每只中随机抽取20%的图片作为验证集，其余的作为测试集，每一只鸭子我们选取150张各个角度的图片。按照此种划分方法可以防止同

一只鸭子同时出现在训练集与测试集中。针对对于同一只鸭子不同角度体重的预测，每幅图像的估计体重如图3.6 a所示，可看出模型对不同角度下鸭子的体重预估波动性较小。为了对体重进行稳定和准确的估计，我们使用这150个评估结果的平均值作为最终的估计体重(EW)。预测误差(PE)定义为鸭子的预测体重和真实标签之间的绝对差。相对误差(RE)被定义为预测误差与真实标签的比值。平均绝对误差(MAD)和平均相对误差被用于评估模型的精度与准确率，如式3.3和3.4所示：

$$MAD = \frac{1}{N} \sum_{i=1}^N \sqrt{(EW_i - y_i)^2} \quad (3.3)$$

$$MRE = \frac{1}{N} \sum_{i=1}^N \frac{|EW_i - y_i|}{y_i} \times 100\% \quad (3.4)$$

其中 N 代表被称重的鸭子的数量， EW_i 为第 i 只鸭子被估计的体重， y_i 是第 i 只鸭子的真实体重。经过十折法，提出的模型的平均绝对误差为58.8克，平均相对误差为2.15%。50只来源于测试集的预测结果与真实标签分布如图3.6 b所示。所有测试图像的预测误差的分布图如图3.6 c所示。

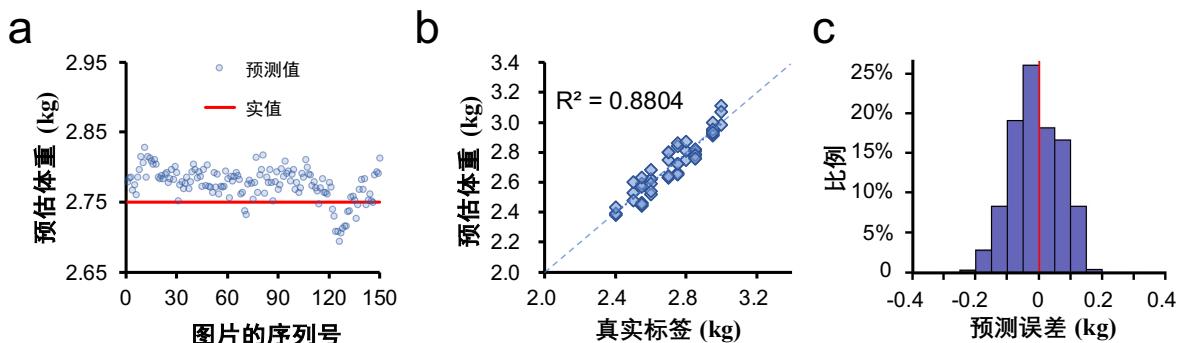


图3.6 模型的测试结果。a. 同一只鸭子不同角度下的体重预测结果。b. 50个预测样本的真实标签与预估体重图。c. 全部测试图像的体重预测误差分布图

3.3.3 与传统特征工程方法的比较

我们将提出的方法与传统的两种视觉称重的方法进行了比较。第一种方法是通过图像的像素面积构建线性函数预测体重，提取鸭子像素面积并构建线性模型的方法如图3.7 a所示。为了减轻成像角度不同导致的计算量与复杂度，本节所提取的特征是从鸭子正中央角度获取的图像。在提取了50只鸭子的图像面积后，我们用面积与体重构建了线形回归方程进行回归预测，回归的结果如图3.7 c所示， MAD 为123.5克， MRE 为4.6%， R^2 为0.7714。这些结果表明，该方法对生产线鸭子图像称重的准确率并不理想。

第二种方法为训练一个人工神经网络ANN模型，通过回归从鸭子中提取的部分特征，如图3.7 b所示。本方法遵从了Amraei等人用于预测鸡体重的工作^[24]。在他们的工作中去除了鸡的颈部、肢节和头，而用剩下的形态学特征提取了面积、周长、凸区域、长轴长度和短轴长度作为特征。为了方便比较，我们采用了类似的方法进行评估。我们使用了一个 39×39 的圆形核进行形态学开运算操作分割出鸭子的身体部分，去除翅膀、四肢和头部。从分割的图像中提取了面积、周长、凸区域、长轴长度和短轴长度作为特征。利用MATLAB的nntool工具箱对含有一层隐层的神经网络模型进行训练。训练采用了默认的参数：tansig作为隐藏层传递函数，MSE作为损失函数，learngdm作为自适应学习方法。表3.2展示了肉鸡和鸭子的体重预估结果，其中采用贝叶斯规则训练的模型MAD最低，为107.0克，MRE为3.79%。图3.7 d为在贝叶斯规则下训练集与测试集估计结果与真实标签情况。这与ANN估算肉鸡体重性能相似，证实本实验的复现正确性。ANN的方法略高于回归面积的方法，因为多了很多特征，但是效果依然比我们提出的方法要差。

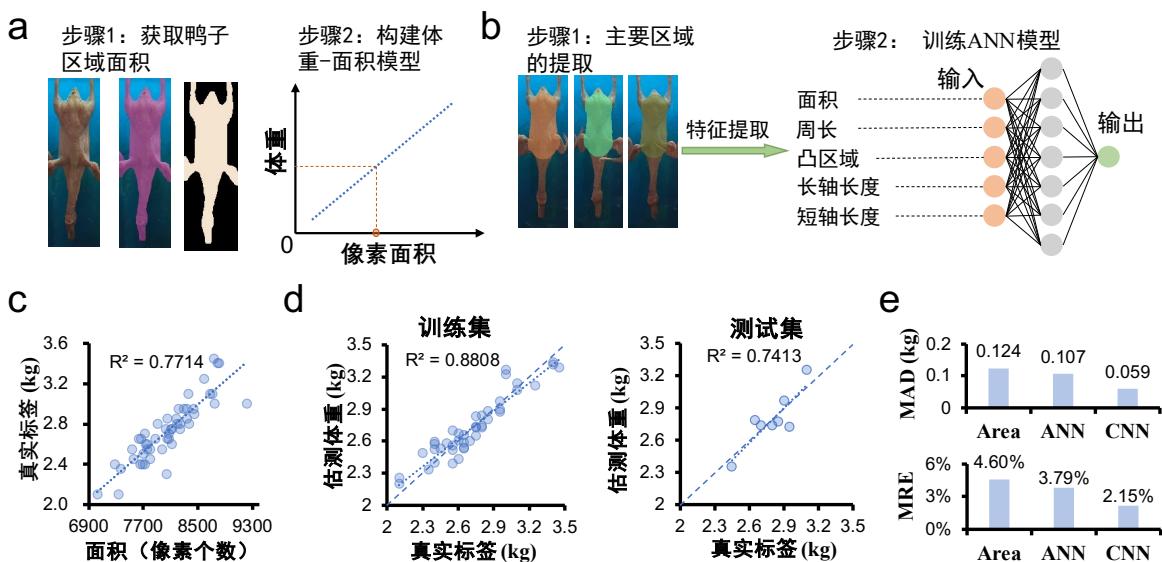


图3.7 a. 像素面积线性回归模型。b. ANN模型。从鸭背区域提取5个特征，利用神经网络对权重进行回归。c. 线性回归模型的线性回归结果。d. 将估计结果与贝叶斯规则训练算法中训练集和测试集的原始权值进行对比。e. 三种方法的比较

比较三种方法的结果可以看出，在二维图像中，CNN模型对鸭子重量的估计效果最好，如图3.7 e所示：ANN方法的估计误差大约是CNN方法的两倍。线性回归模型法在鸭重估算中效果最差。结果表明，我们提出的CNN模型在从二维图像中估计鸭子重量方面优于其他方法。此外，CNN模型可以自动提取特征，避免了人工提取特征所带来的不足，同时也提供了更多的便利。

表3.2 不同训练算法的结果

训练算法	神经元数量	训练集 MAD (克)	训练集 RMSE	鸡训练集 RMSE	测试集 MAD (克)	测试集 RMSE	鸡测试集 RMSE
Gradient descent	7	129.4	161.76	150.66	183.7	217.87	157.12
Levenberg–Marquardt	10	70.9	89.65	89.63	149.7	209.87	91.59
Scaled conjugate gradient	11	114.4	137.59	115.10	113.6	131.55	121.63
Bayesian regulation	9	88.5	109.48	80.68	107.0	121.59	82.37

3.4 讨论

关于3.1.2节中为什么要舍弃掉鸭子的腿，最主要的原因是因为腿部的光线在采集时没有严格控制，导致会有光线差异，影响模型训练，如图3.8 a所示。我们做了一个简单的实验，用上述网络，所有的网络配置与预处理方法均相同，除了输入是否有腿。图3.8 b为单只鸭子在有腿和无腿时的不同角度预测结果分布，可看出去除腿部有利于模型的提升。图3.8 c和d展示了有腿与无腿情况下测试集模型预测误差的分布，有腿情况下MAD为76.5克，无腿情况下MAD为53.6克。

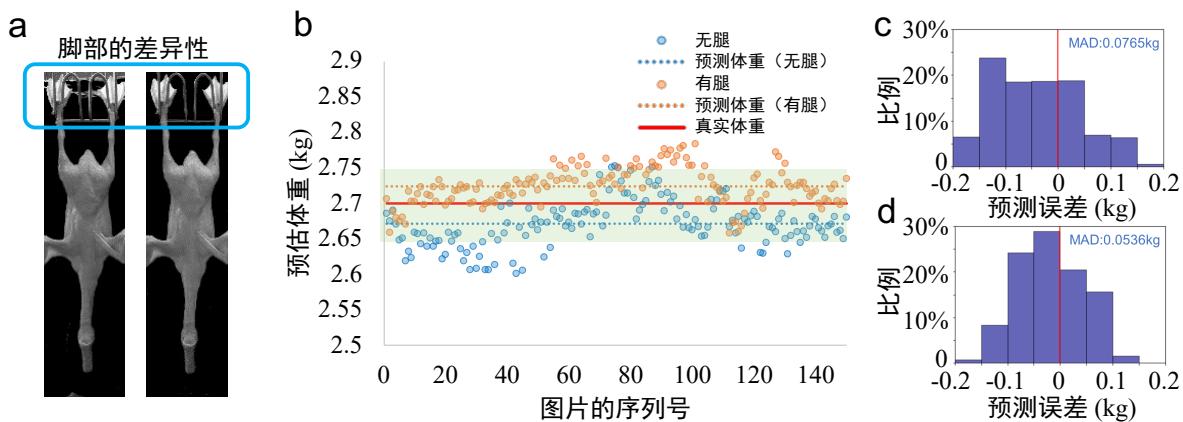


图3.8 a. 数据脚部可能存在较大的光线变化；b. 同一只鸭子在不同角度下有无腿的预测结果；c. 有腿的测试集的误差分布图；d. 无腿的测试集的误差分布图

3.5 本章小结

本章提出了一种基于端到端深度学习的低成本 2D 图像称重的方法来估计鸭子的体重。与人工测量方法相比，这种非接触式测量方法可以在不中断生产线运行的情况下获得鸭子的准确重量，从而大大提高了生产线的效率。我们在生产线前设置了单目可见光

摄像机，拍摄了 50 只鸭子的视频。每只鸭子的重量都是人工测量的，并为同一只鸭子的每一幅图像标记。每只鸭子从不同角度拍摄了至少 150 幅图像，并将这些图像收集为实验数据集。这些图像经过预处理后，我们设计并评估了 8 种类型的 CNN 网络来估计这些鸭子的重量。结果表明，采用 Softplus 激活函数的核大小为 3×3 的三层 CNN 模型验证效果最好。并以 Adam 优化方法和 0.1 的 dropout rate 作为 CNN 网络的最佳训练参数。采用 10 倍交叉验证方法对模型进行估计，得到的测试 MAD 为 58.8 克， MRE 为 2.15%。我们还将 CNN 模型与基于同一数据集的像素面积线性回归模型和 ANN 模型进行了比较。线性回归模型的 MAD 为 123.5 g，ANN 模型的 MAD 为 107.0 g。结果表明，CNN 模型对鸭子重量的估计精度远远高于其他二维图像称重方法。在非接触和准确的方式估计鸭子的重量的能力方面，使我们提出的 CNN 方法在屠宰工厂的生产线上称量鸭子是一个合适的解决方案。

本章证实了端到端地回归训练策略用于动物体重预测的可实施性，免去了人为的手工提取特征，并极大地改善了图像称重的精度。

4 利用残缺图像进行体重预测

在第3章中，我们证实了将端到端的卷积神经网络模型用于自动提取特征并估算体重的可行性，并实现了高精度的评估。然而，真实的现场环境并不一定是完完全全理想的状况。

图像之间存在遮挡影响模型任务的情况很多，例如在COCO数据集^[61]中，作为图像实力分割任务，常常遇到两物体目标距离很近从而出现遮挡的情况。一些工作尝试从损失函数入手^[62]，一些工作尝试，例如医学影像处理，从数据融合^[63]入手。但多是以任务为导向设计解决方法，本文的任务主要是回归模型，并且遮挡情况存在一定的规律性，因此可以尝试从预处理和多尺度角度入手提高模型的精度。

本章以复杂的流水线为研究背景，此场景下鸭子与鸭子之间存在遮挡情况，遮挡物对鸭子体重的预测干扰将相当的大。基于第3章的结论，本章节尝试从预处理及多尺度角度入手，尝试解决在复杂流水线下，残缺图像的端到端体重预测，并期望性能能够达到简单场景下的预测精度。

4.1 数据集与预处理

4.1.1 数据集采集

本章工作以某大型工厂(中国山东红叶食品有限公司)的生产线环境为研究背景。经过一系列的机器加工，工厂需要称量鸭子的重量。最后，处理过的鸭子用它们的右翅挂在生产线上，它们靠得很近，从而出现上一鸭子的屁股被下一只鸭子的头部所遮挡。数据集采集的方式与3.1.1节类似，我们依然通过一个便携式的三脚架单目摄像机作为采集视频的设备。由于鸭子之间存在遮挡，我们并没有将相机正对流水线，而是稍微倾斜了些，目的是为了尽可能减少遮挡，并部署在流水线前近2米左右。然后，我们在生产线的背后固定了一块黑布，以保证背景的简单性，与鸭子图像的可分割性。生产线环境及实验设备摆放如图4.1所示。放置好单目摄像机和黑布后，我们开始拍摄鸭子在生产线上移动的过程，并使用电子秤称重作为标签。

我们总共采集了300只鸭子与对应的标签样本，每个样本中大约包括27张来自不同角度的图像。整个数据集的鸭子体重分布请见附录A的图A2，大于1.8千克的数据分布基本满足正态分布，而小于1.8千克的样本量较小，一般可被认定为不合格产品，且容易通过面积等方式直接去除，因此模型训练与测试仅仅选用了大于1.8千克的数据。

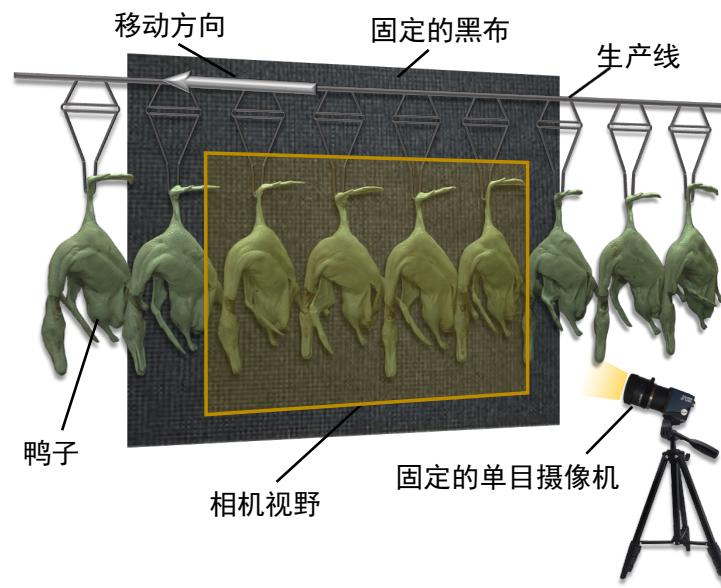


图4.1 生产线的环境与摄像机设备的部署示意图

4.1.2 数据预处理

从视频中提取鸭子图像，估计其权重。图4.2展示了图像的处理过程，目标是在视频的每一帧中分离鸭子并保存为图像格式。图4.2 a所示为视频的原始图像帧，包含来自背景的噪声，将很难检测到每只鸭子的位置。我们采取以下步骤从图像中移除背景噪声。首先将R通道从RGB图像中分离出来，如图4.2 b所示，因为鸭子身体对红光的反射性更强。设定一个阈值40将图像转换为二值化图像，如图4.2 c所示，有很多的噪音存在的背景，引入了1个 5×5 的核函数通过形态学开运算去除噪音，如图4.2 d所示。再消除噪音后，通过二值化图像与RGB图像进行与操作以去除原图像的噪声，如图4.2 e所示。

在去除背景噪声后，我们开始对图像中的鸭子进行检测。我们设定一个高度后开始从二值化图像中水平扫描鸭子的翅膀，并以鸭子翅膀的水平中心点为基准周围固定尺寸提取出鸭子的图像，大小是高700像素，宽500像素，如图4.2 f所示。由于存在旁边鸭子的遮挡情况，这会影响基于章节3方法的预测，证明我们放在了附录B详细叙述。主要的干扰来源于鸭子左侧的上一只鸭子的尾部和右侧下一只鸭子的头部遮挡，我们尝试将这两个部位去掉。我们通过设置一个阈值为40的二值化图像扫描鸭子的头部与颈部的两个轮廓线，由于扫描的轮廓线可能存在抖动，我们使用3次方拟合离散的点形成一条平滑的曲线，如图4.2 f所示，将轮廓线两侧的像素全部清零，即可得到一个较为干净的图像，如图4.2 g所示。在程序运行对鸭子进行分割时，程序会自动记录每一只被切割下来的鸭子的编号，具体的算法流程图见附录A的图A3。

图像融合常常被用于图像处理中，这样做的好处是可以从不同的图像中获取更多的信息。提出了随机图像融合的方法用于增强数据集，可以使破损的图像从其他图像中弥补信息，同时增加更多的图像，过程如图4.2 h所示。我们假设某一只鸭子在不同的角度下总共有 N 张图像，将从中随机抽取 n 张作为随机融合的输入图像，则理论上可以生成 $N!/n!$ 张不同的图像，在 N 与 n 存在较大差距时能生成相当数量的不同的图像。图像融合的方法我们采用加权平均，即图像各个像素点位置的平均值。本章中，我们对每一只鸭子选择各个角度图像 N 为24张用于训练与测试模型，我们设置 n 为6张。每一只鸭子将生成30张随机融合的图像作为该方法的训练集与监督集，部分生成图见附录A的图A4所示。由于本章使用的摄像机无明显成色影响，因此本章我们将RGB三通道均作为输入，输入网络前仅归一化图像的像素值为[0, 1]。

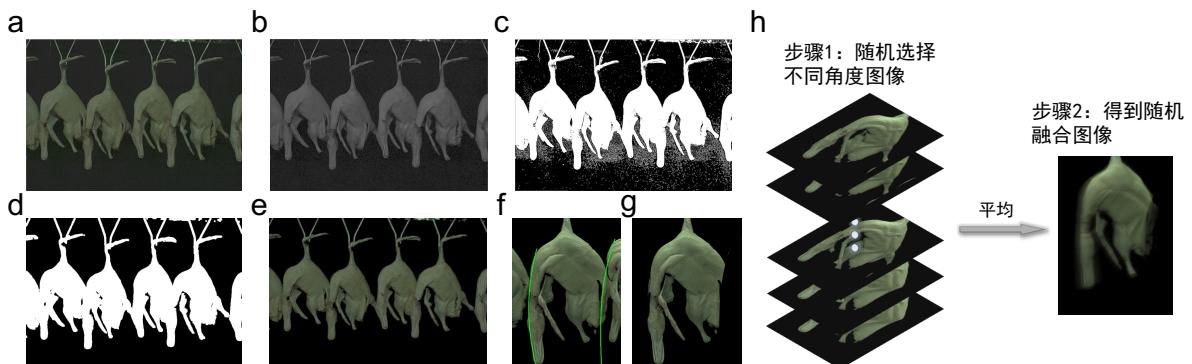


图4.2 图像预处理过程。a-g 切割图像过程展示；h 随机融合方法

4.1.3 设备与编程环境

本章的主要工作中，提出的模型使用Python语言（3.7版本）和TensorFlow框架（1.15.0版本）构建，模型的计算与优化在2张英伟达GTX 1080Ti显卡上操作。

4.2 卷积神经网络结构

不同的CNN网络结构会影响训练过程，尤其是拟合和泛化效果。在之前的第3章研究中，我们证明了 3×3 的卷积核大小和Softplus激活函数对鸭子体重估计有很大的影响，所以在本文中我们继续采用这些参数。Woo等人^[56]总结了网络的三个重要因素：深度、宽度和基数。深度，如VGGnet^[47]和Resnet^[49]；Googlenet^[48]证明了网络宽度可以提高网络性能；Xception^[64]和ResNeXt^[65]的经验表明基数性具有较强的网络性能，但这些方法主要用于分类任务。多尺度网络也广泛应用于深度学习任务中，Godinez等人^[65]提出了一种用于细胞图像分类的多尺度卷积神经网络。在我们之前的工作和Luo等人^[63]的工作

中，我们发现VGGnet类型的网络在回归任务中有很好的表现，所以我们主要讨论了网络的深度、宽度和多尺度因子，不同的网络结构如表4.1所示。其中，所有的结构都采用 3×3 的卷积核，步长为1，无填充，用Softplus激活函数激活。最大池化操作的窗口为 2×2 大小，步长为2，无填充。全连接层均使用Softplus激活函数激活，输出层无激活函数。

表4.1 测试的8种卷积神经网络的结构，卷积层参数表示为“Conv-<卷积核数量>”，全连接层参数表示为“Dense-<节点数量>”

Net-5	Net-5-BN	Net-6-BN	Net-7-BN	Net-6-BN (Behind)	Net-6-BN (Large)	Net-Width	Net-multi-scale		
输入 ($700 \times 500 \times 3$)									
Conv-8	BN Conv-8	BN Conv-8	BN Conv-8	Conv-8 BN	BN Conv-64	BN Conv-4	Average-Pool (1 × 1) BN Conv-4	Average-Pool (2 × 2) BN Conv-4	Average-Pool (4 × 4) BN Conv-4
Max Pool									
Conv-16	BN Conv-16	BN Conv-16	BN Conv-16	Conv-16 BN	BN Conv-128	BN Conv-16	BN Conv-16	BN Conv-16	BN Conv-16
Max Pool									
Conv-32	BN Conv-32	BN Conv-32	BN Conv-32	Conv-32 BN	BN Conv-256	BN Conv-32 Conv-32	BN Conv-32	BN Conv-32	BN Conv-32
Max Pool									
Conv-64	BN Conv-64	BN Conv-64	BN Conv-64	Conv-64 BN	BN Conv-512	Flatten	BN Conv-64 Max Pool	BN Conv-64 Max Pool	BN Conv-64 Max Pool
Max Pool									
Flatten									
Concatenation									
Dense-256 Dropout rate 0.2									
Dense-128 Dropout rate 0.2									
Dense-1									

在训练过程中，训练使用的损失函数遵从式3.1，即仍为均方差损失函数。依然采用衰减学习率如式3.2，参数的配置均相同。本节中的模型在2张GTX 1080Ti显卡上运算，每张显卡的小批量图片数目为200张，优化方法依旧选择Adam方法。此外训练过程中我们选择的Dropout率为20%。

4.3 体重估算方法

本文中的样本中总共包括282只鸭子，体重的分布情况见图A1。我们选择了258只鸭子作为训练集，24只鸭子作为测试集。按照4.1.2节的预处理方法，对每只鸭子都生成30张随机融合的图像作为训练集与监督集。在进行模型的测试时，我们将同一只鸭子各个角度全部的图像进行融合，用融合的图像作为该鸭子的最终评估值。

在评估模型的精度时，我们采取与3.3.2节中式3.3相同的评估方法，即平均绝对误差

MAD。由于本次的样本数量较为充足，因此本节只将训练集作为模型的参数优化，测试集生成的随机融合图像作为监督数据，当模型训练完成后，用测试集的每一只鸭子全部图像的融合图像作为输入，并计算平均绝对误差*MAD*作为模型的最终评估结果。

4.4 结果

4.4.1 模型精度评估

我们针对表4.1的8种网络结构的模型进行了测试，训练的参数均采用了与4.2节描述相同的算法，我们计算了每个模型的*MAD*值，具体见表4.2。尽管从表中认为是Net-Width这个网络结构取得了最好的结果，但是其结果并没有相差太多，很难说是否是因为网络结构的易学习性而提高了模型的精度。因此我们可初步认为，该方法可在精度上达到约55.6克的误差。我们在图4.3中画出了Net-6-BN、Net-Width与Net-multi-scale的学习曲线，其模型的可视化结构图如图A5所示。

表4.2 8种不同网络结构评估的精度

网络名称	Net-5	Net-5-BN	Net-6-BN	Net-7-BN	Net-6-BN (Behind)	Net-6-BN (Large)	Net-Width	Net-multi-scale
<i>MAD</i> (克)	75.4	58.4	57.2	57.8	57.3	64.8	55.6	64.2

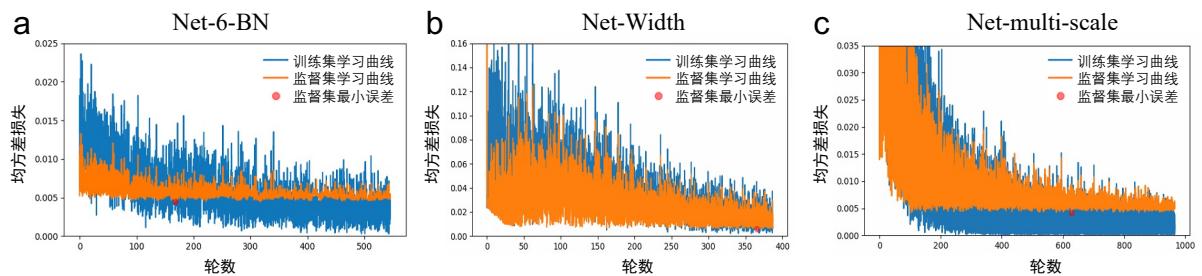


图4.3 三种网络模型的学习曲线

4.4.2 融合图像对模型精度的影响

前面的实验都是在提出的融合图像的数据集方法下训练的，并且各种网络模型都取得了不错的效果。在图像残损的情况下是否是因为图像的随机融合策略而导致的，我们这里简单进行一个实验来证明随机融合图像有助于模型的学习与评估。

为了加快训练，我们选择了表4.1中Net-6-BN的结构作为训练模型。其中一部分数据集为融合图像数据，另一部分数据集为非融合图像数据，训练与测试方法与章节3相同。采用了4.2节的模型训练参数，并且相同划分数据集为训练集与测试集，进行模型训练。

在评估模型时。使用非融合图像数据集的模型的测试精度 MAD 为 69.4 克，而使用了融合图像作为训练集的模型的测试精度 MAD 为 55.7 克（与表 4.2 中不同的原因是由于模型每次训练只有局部最优解，不是定值）。二者之间的差距较大。我们将测试集中每个样本的预测值的绝对误差在融合数据和非融合数据条件下计算出来，并定义了某测试个体中，用融合图像的绝对误差减去非融合图像的绝对误差，即当某个体在非融合图像模型的精度高于融合图像时，数值应为正，反之为负，结果如图 4.4 所示，每一个点代表了测试集中的一个样本。可看出绝大部分个体在通过融合图像后，精度得到了提升，仅部分样本可能因为数据特例高于融合图像精度，因此融合图像可以较大幅度提高模型的精度。

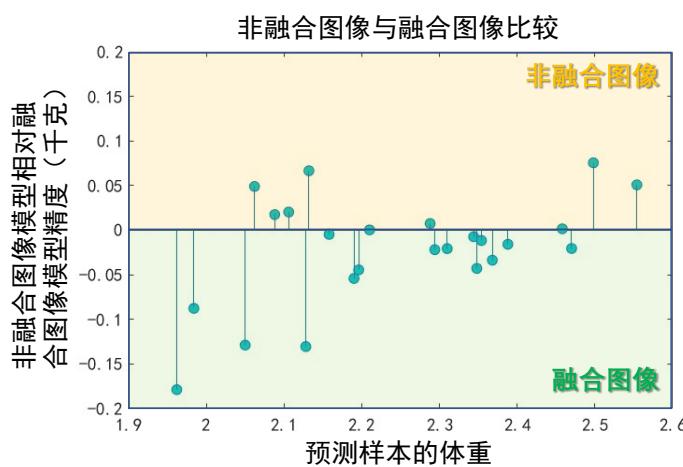


图 4.4 非融合图像与融合图像对模型精度的影响

4.5 不同视角对模型的影响

我们可能会考虑，因为鸭子在不同视角下存在旋转情况，是否某一个特定的角度能促进模型，而某一个特定的角度确能降低模型的训练精度？我们进行了两种实验。第一种实验，类比于异常分布检测：假如我在训练一个模型，训练中将 1 个数字 9 的标签归类于 1，而参与训练。可能模型在此就收到了一定的精度影响，但测试时由于大部分数据是无异议的，从而最后模型评估时，这个数字 9 虽然可能会被判为 1 类，但是相比正常的数字 1，它的置信度会小一些，这就是异常分布现象。

考虑到这种情况，我们可能做出一些假设，如果存在某个角度对模型的精度是存在干扰的，那么这个角度下的预测偏差必然比其他角度的预测偏差要大。因此我们采用表 4.1 中 Net-6-BN 的结构的模型，不使用图像融合而单独训练各个角度的图像的结果，并将测试集上的图像进行预测，只是我们将鸭子的位置分了 27 个位置，相同位置的图像的预测结果聚在一起，我们绘制了他们绝对误差分布的箱型图，如图 4.5 所示。

但是发现各个位置的绝对误差均值基本相同稳定不变，因此从异常分布检测的角度难以说明存在某个视角，该视角下模型会存在较大的干扰。

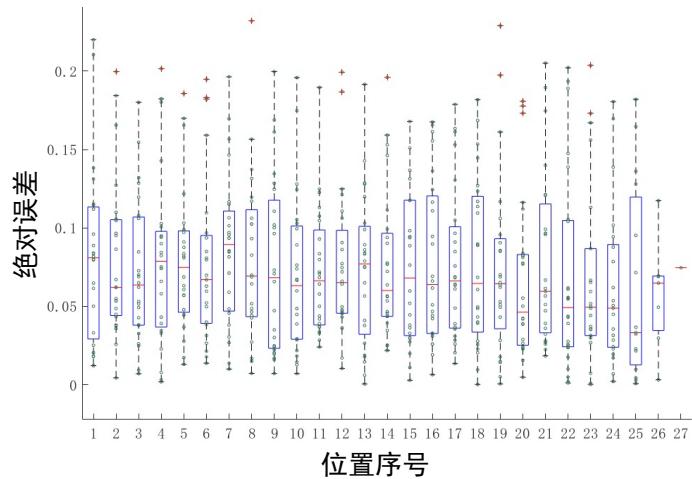


图 4.5 模型对各个不同角度图像预测绝对误差的箱型分布图

紧接着，我们用一种更简便的方法判断不同角度对模型的影响。我们将整个视角分为 4 个区域，每个区域只取靠近中央的 4 张图片作为训练与评估模型。4 个不同的视野如图 4.6 所示，可以看出明显的差异。然后我们以非融合图像的训练和评估策略，评估模型的精确度，模型训练以及数据集配置与前文相同，最终结果如表 4.3 所示。

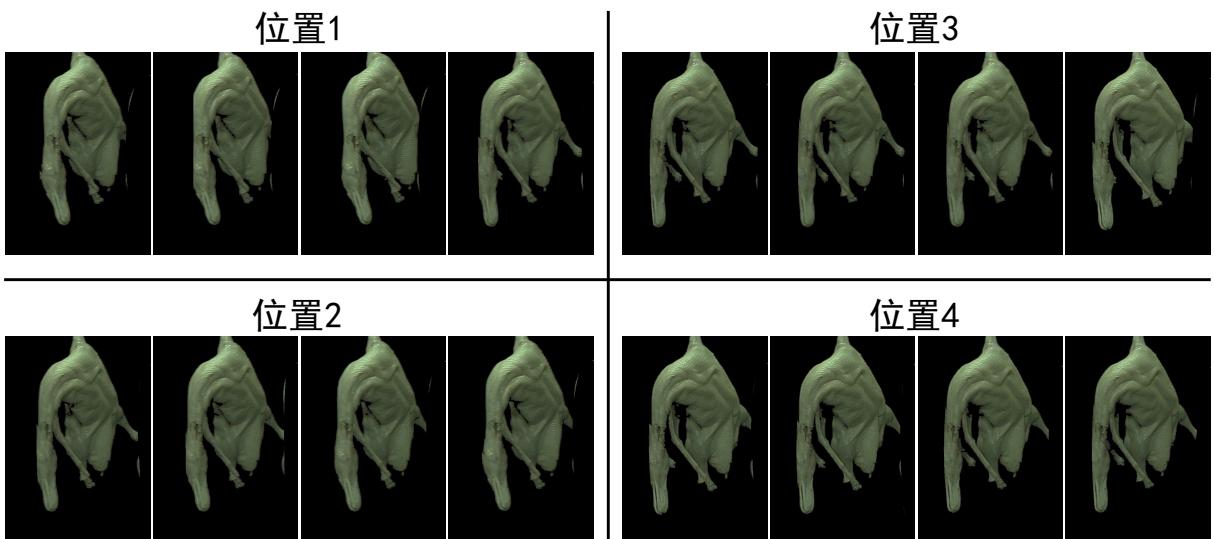


图 4.6 选择的四个不同的小区间

表 4.3 不同区间数据集的模型评估

位置	全部	位置 1	位置 2	位置 3	位置 4
平均绝对误差 (克)	69.4	69.8	68.0	66.0	72.1

各个区间最终结果与整个区间全部训练的结果并没有特别大的差异，因此我们可以得出一个结论：在端到端的模型训练时，即使会存在不同的角度导致不同的视角，只

要满足数据量足够多且有一定的模型依据，可认定模型可以克服不同视角的影响，此结论从第3章的图3.6a也可以得出。

4.6 本章小结

本章的主要工作是基于端到端的图像预测体重方法，在复杂的流水线下，通过残缺的图像进行精确的鸭子体重预测。提出了一种随机融合图像的预处理方法，不仅能够有效地扩大数据集，同时也能有效增强数据集的多样性。我们从工程中采集了282只鸭子，清除了过轻的不合格的鸭子。将数据集分为了训练集与测试集，训练集包括258只，测试集包括24只，针对每只鸭子，我们使用提出的随机融合方法增强，测试集增强的图像用作模型训练时的监督集，在进行测试时我们使用测试集每一只鸭子的全部融合图像作为输入以评估模型精度。我们讨论了部分的网络结构，最终的精度为55.6克，并且在精度上远远超过了不进行图像融合的方法。我们同时也讨论了视角对模型精度的影响，在端到端的模型训练时，即使会存在不同的角度导致不同的视角，只要满足数据量足够多且有一定的模型依据，可认定模型可以克服不同视角的影响。同时我们的方法也打破了很多2D图像识别对于环境严格要求的限制。

本章节成功将端到端的图像预测方法应用到了复杂场景中出现图像残缺的情况下，更进一步证明了端到端图像称重的优越性。

5 总结、讨论与展望

5.1 研究工作总结

本文是作者在东北大学秦皇岛分校读本科期间有关动物的图像称重的工作总结与讨论。本文的主要工作是将自动提取特征的端到端的方式应用于鸭子体重的预测，具体包括一个简单的场景和一个复杂的场景。本文首次提出了端到端的方式应用于图像称重，并超越了传统的称重方法，同时研究了复杂流水线场景下端到端的称重，提高了模型适应恶劣环境的鲁棒性。本文主要的贡献如下：

1. 本文首次提出了一种新的图像称重的方法，基于卷积神经网络的端到端的回归预测，免去了人手工提取特征的过程，而获取到了更多潜在的特征。并提出了多视角图像的体重评估方法，提高了模型的鲁棒性。且我们的精度与传统的回归面积以及提取部分形态学特征进行多元线性回归的方式相比，高出了近2倍。
2. 本文的方法只需要一个便携式的摄像机，部署时只需要对相机视野进行矫正，选择合适的视角，而免去了繁杂的安装操作。为便捷性和快速性提供了新的方法，同时也更加方便前期数据的采集获取，并不干扰工厂的正常工作。
3. 针对于复杂的流水线环境下，图像存在残缺的情况，我们提出了一种随机融合的预处理方法，该方法不仅可以扩大数据集，同时也可以增强数据集的多样性。我们的方法大大提高了模型对残缺图像识别体重的精度，并且在数据充足的情况下，证明了模型的精度与具体的位置无关，打破了2D图像识别对于环境严格要求的限制。

5.2 不足与展望

本文极大地提高了模型视觉称重的精度，具有快速性，便捷性等，但是受作者课业压力与精力以及水平等限制，本文的工作仍有进一步提升的空间，总结如下：

1. 本文主要针对的目标均为死物而非活物，因此没有将算法测试于活物，无法评估算法在活物上的性能。因为活物的变化更大，针对活物的视觉称重与端到端的研究可能还会遇到很多的问题。
2. 本文采用的是小样本，仅在小样本范围内取动物的体重，没有完全考虑到整个工业的情况与隐患，未来可能可以从大样本范围验证模型。
3. 本文在第四章中的数据预处理切割鸭子部分，消耗的时间复杂度过高，是影响速度和精度的一个潜在隐患问题。因此可能需要更新式的方法来解决这个问题。

4. 本文仅仅针对的是固定流水线场景下的鸭子，而非通用的方法，所谓通用的方法即不仅能应用于活物，而且能应用于死物，可以用于预测各种动物，尽管可能不同动物之间所需要的参数不同。

5.3 潜在可挖掘方向讨论

本节主要讨论一些可以继续挖掘的方向，并尽可能从通用测重的方法解决问题。由于作者自身精力的限制，没有时间再去验证这些想法，但是希望将他们提出来为后面的研究者提供帮助与思路。

5.3.1 基于解耦合方法加速预测

解耦合技术，即将数据中有用的信息与无用的信息分离开来，以章节4的鸭子图像为例，附近遮挡的鸭子即无用数据，需要去除使得模型有更高的鲁棒性。第四章采用了人工扫描的方式，增大了模型的复杂度，使得图像的预处理速度变慢了。我们提出一个基于解耦合的模型方法，用于加快模型的预测速度。将方法过程可视为图5.1。

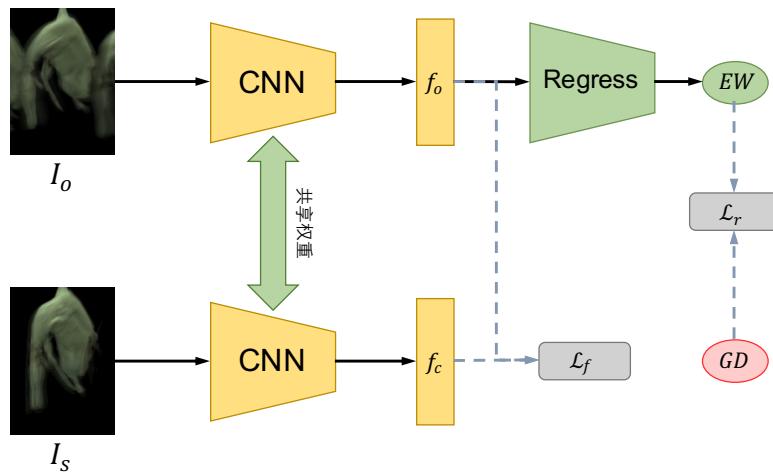


图5.1 提出的解耦合方法假设

假设原始的图像为 I_o ，对应的无其他信息只包含目标图像信息的图像为 I_s ，用于提取特征的卷积层为 $F(I)$ ，提取的特征向量用 f 表示，回归模型为 $R(f)$ ，预测体重结果为 EW ，真实标签为 GD 。此时我们将此作为一个优化问题，输入2张图像，通过相同的卷积层提取特征，分别为 f_o 和 f_c ，此时第一个需要优化的即式5.1：

$$\mathcal{L}_f = \|f_o - f_c\|_2^2 \quad (5.1)$$

这个是为了使卷积层尽可能提取争取的特征，第二个需要优化的便是预测体重的精度，如式5.2：

$$\mathcal{L}_r = |EW - GD|^2 \quad (5.2)$$

因此联立式5.1和式5.2，我们所需要的优化问题为：

$$\arg \min (\mathcal{L}_r + \lambda \mathcal{L}_f) \quad (5.3)$$

其中 λ 为一个超参，流程图如图5.1所示。

如果该方法可行，则我们就可以省去预处理的步骤，大幅度提高模型预测时的速度，更加适用于工程部署。

5.3.2 基于目标检测的通用体重预测算法

大部分情况下我们需要提取图像中的目标，这个是一个针对不同场景存在不同时间的过程，在模型迁移时非常麻烦，因此我们提出一个结合目标检测的方法，用于提高模型的通用性。目标检测的方法非常多，较为传统的有R-CNN^[66]、Fast R-CNN^[67]、Faster R-CNN^[68]、Yolo系列^[69-71]等等。一般大家会想到的方法为，先用目标检测算法将目标检测出来，再将检测到的物体用卷积神经网络预测体重。这是一个可行的方法，但是却不是一个有效的方法，因为模型存在速度问题，目标检测算法已经将占用大量的计算量，再补上一个CNN回归模型只会增大计算量。针对这种方法，我们提出一个联合位置信息的体重预测方法，方法基于Faster-RCNN，如图5.2所示。

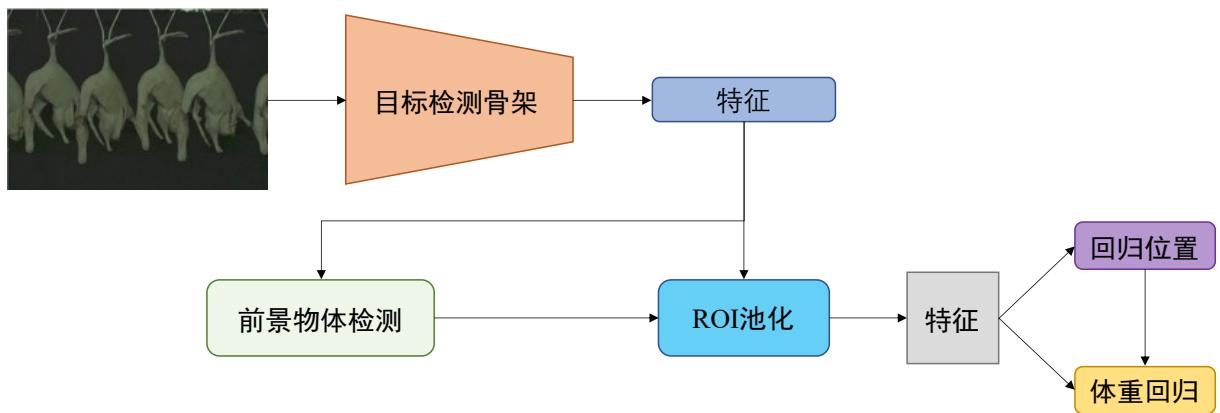


图5.2 基于目标检测方法的通用图像体重预测方法

采用Faster RCNN系列是因为目标的特点，因为我们预测体重往往需要整幅图像，因此不适合Yolo系列的目标检测网络作为骨架。Faster RCNN系列第一步为前景的检测，即物体的检测，然后再用检测到的物体进行回归与分类。我们所提出的方案即将分类的网络更换为体重回归的网络，考虑到经过ROI池化的网络中，相同的目标可能因为前景选择的问题导致不同的尺寸，这样经过体重回归网络必然会得到不同的结果，因此我们在提出的模型中，使模型位置回归网络中提取出一个特征加入到体重回归的网络，这样就会改善网络的单一性，经过训练即可。

5.4 本章小结

本章总结了本文的主要贡献以及存在的一些潜在问题，并根据这些问题讨论了进一步解决的方案。我们提出了一个基于解耦合的模型部署加速的方案，通过该方案我们可以让模型在部署到现场时省去麻烦的预处理操作，而尽可能大幅度提取有用的信息。我们又提出了一个基于目标检测的通用视觉称重模型，这样可以适用于任何场合的训练。希望只需要图像的位置与体重信息的标签，即可训练出一个有效的模型。本章没有进行实验验证，无法保证提出的猜想的可靠性，但希望为后面的研究提供新的思路。

结 论

本文的主要工作是将自动提取特征的端到端的方式应用于鸭子体重的预测，具体包括一个简单的场景和一个复杂的场景。

在简单场景下提出了一种基于端到端深度学习的低成本 2D 图像称重的方法来估计鸭子的体重。与人工测量方法相比，这种非接触式测量方法可以在不中断生产线运行的情况下获得鸭子的准确重量，从而大大提高了生产线的效率。采用 10 倍交叉验证方法对模型进行估计，得到的测试 MAD 为 58.8 克， MRE 为 2.15%。我们还将 CNN 模型与基于同一数据集的像素面积线性回归模型和 ANN 模型进行了比较。线性回归模型的 MAD 为 123.5 g，ANN 模型的 MAD 为 107.0 g。结果表明，CNN 模型对鸭子重量的估计精度远远高于其他二维图像称重方法。在非接触和准确的方式估计鸭子的重量的能力方面，使我们提出的 CNN 方法在屠宰工厂的生产线上称量鸭子是一个合适的解决方案。证实了端到端地回归训练策略用于动物体重预测的可实施性，免去了人为的手工提取特征，并极大地改善了图像称重的精度。

在复杂场景下，解决了通过残缺的图像进行精确的鸭子体重预测。提出了一种随机融合图像的预处理方法，不仅能够有效地扩大数据集，同时也能有效增强数据集的多样性。最终的精度 MAD 为 55.6 克，并且在精度上远远超过了不进行图像融合的方法（69.4 克）。我们同时也讨论了视角对模型精度的影响，在端到端的模型训练时，即使会存在不同的角度导致不同的视角，只要满足数据量足够多且有一定的模型依据，可认定模型可以克服不同视角的影响，同时我们的方法也打破了很多 2D 图像识别对于环境严格要求的限制。

同时我们也对本文的工作进行了总结与问题的总结，并根据已有的问题进行了展望，讨论了可继续挖掘的几个方向。提出了一个基于解耦合的模型部署加速的方案，通过该方案我们可以让模型在部署到现场时省去麻烦的预处理操作，而尽可能大幅度提取有用的信息。提出了一个基于目标检测的通用视觉称重模型，这样可以适用于任何场合的训练。希望只需要图像的位置与体重信息的标签，即可训练出一个有效的模型。以上两种方案只提出了假设但没有实验验证，希望为后面的视觉称重方案提供一个新的思路。

攻读学位期间的学术成果

发表论文

1. Yuliang Zhao, **Ruoyu Chen**, Shuyu Wang, Lianjiang Li, Xiaopeng Sha, Yongliang Yang, Guanglie Zhang and Wen Jung Li. A deep learning model using 2D images that weighs ducks more accurately than conventional non-contact weighing methods. Computers and Electronics in Agriculture, 2021. (Submitted)
2. **Ruoyu Chen**, Liqiang Xiao, Jia Li, Hua Zhang and Xiaochun Cao, Sim2Word: Explaining Similarity with Representative Attribute Words via Counterfactual Explanations. Annual Conference on Neural Information Processing System (NeurIPS), 2021. (Submitted)
3. Liqiang Xiao, Hua Zhang, **Ruoyu Chen** and Xiaochun Cao, Prototypical Guided Contrastive Proposal Encoding for Few-Shot Object Detection. Annual Conference on Neural Information Processing System (NeurIPS), 2021. (Submitted)
4. Jingzhi Li, Lutong Han, **Ruoyu Chen**, Hua Zhang, Bing Han, Lili Wang and Xiaochun Cao, Identity-Preserving Face Anonymization via Adaptively Facial Attributes Obfuscation. ACM International Conference on Multimedia (ACM MM), 2021. (Submitted)

发明专利

1. 赵玉良, 陈若愚, 沙晓鹏, 崔逸丰, 李文超, 一种单目摄像机视觉测重方法, 中国, 公布号: CN110956198A

致 谢

到这里，也大概是我四年本科生涯快要接近尾声了，心里有非常多的感慨。这里我必须要感谢我的母校，东北大学秦皇岛分校，他给我带来了大学四年难忘的经历，锻炼了我。曾经在上大学以前，没有目标性，高考后渐渐也开始失去了挑战性。但是我的母校让我看到了另一面，这个世界很大，我想出去看看。在这里我接触到了以往从未有接触过的东西，提高了我的眼界，也就是从这一刻开始，我燃起了心中挑战欲望的火苗，开始树立起远大的志向，不懈努力。我的心智上巨大的改变，都是我的母校带给我的，我希望未来，我的母校能因为我而骄傲。在这里我获得了优异的成果，锻炼了能力，获得了保研名额，浴火重生了。在本科期间有很多启蒙与帮助我的人，在这里请容我对他们表示感谢。

首先我要感谢我的父母，对我这么多年来的照顾，在我读大学的时候给我最大的支持，让我没有后顾之忧而专心于提高自身的能力。在我需要一些提高自身效率的设备时，父母总是给我支持，让我大学以来用最好的条件锻炼自己。我很感激父母努力为我创造的条件，没有他们的努力我的成功可能还需要付出更大的代价，同时也很感激这 20 年来父母的养育之恩。

感谢我的毕设导师赵玉良老师，本项课题的研究主要基于其课题与基金项目的支持，研究期间为我提供了资源与设备，在学校提供了一个良好的科研环境，并且指导修改我的论文。在实验室期间我锻炼了很多的技能，包括代码工程能力、文献检索与写作能力。感谢本科期间赵老师为我提供的成长资源，奠定了以后做科研的基础。

感谢沙晓鹏老师，是一个很细心的老师，经常与我进行沟通，进行心理疏导，也时常在我的科研上交流指导我，在我的保研上热心地帮助过我，积极在我的未来规划上帮助我，给予我支持与鼓励，与学生非常亲近，也给我提供了非常多宝贵的生活经验。

感谢我的辅导员王晓蕾老师，她在我在校期间提供了很多帮助，经常因为一些急事打电话沟通，保研期间王老师热心地帮我准备着材料，以及在申报个人荣誉时王老师也非常耐心地帮助我，感谢在本科期间老师对我的关心与帮助。

感谢魏永涛老师，他是让我最早接触科创竞赛的老师，在我入学时为了提供了很多的资源，迅速让我的技术能力锻炼了起来，并开阔了我的眼界。

感谢中科院沈阳自动化所的杨永良老师，对我的论文进行了细心的指导，与我讨论并提出自己的见解，以及感谢每一位对我进行过指导，教授过我课程，帮助过我规划未来的老师。

再之，感谢我的学长学姐们，感谢 15 级过控李嘉鹏学长，与他的交流中提高了我的视野，并且在我科研无人帮助时，他与我沟通交流深度学习，提高了我的见解；感谢 16 级测控高君彬学长，在制定自己大学规划时，经常交流并经常给我提供宝贵的意见，让我避开了很多的坑，同时也给我提供了很多的帮助；感谢 16 级通信李一帆学长，本科给了我带来了很多的正能量，推免期间给了我很的建议，促进了我的发展，并一直在我迷茫时鼓励着我前进；感谢 16 级测控杨俊峰学长，在我平时科研压力非常大的时候，经常给我学业上的指导，使我在繁杂的科研任务同时，也保持住了优异的学习成绩；感谢 16 级机械王政、王冬凡学长和 16 级会计李悦莹学姐，在开始时候一起打了大创的比赛，锻炼了我的能力，与你们合作是我的荣幸；感谢 16 级自动化杨丰辰学长，在我保研期间一直在给我提供资料与建议，以及在我平时的学习中积极给我提供资源。还有许许多多帮助过我的校友们，感谢在大学以来对我的帮助。

也要感谢我在学校期间经历过的每一个角落，智能硬件实验室，3D 打印实验室，ROBOCON 实验室，4012 实验室，iCAN，科技创新部，大学生创新创业协会，测控拔尖创新人才实验班。感谢陪伴过我成长的每一位同学与干部，黄锟，皮峻银，赵占宇，韦立凯，张钧杰，诸葛华扬，楼锦程，叶何等等，因为你们，让我的大学更加的精彩。

感谢大四在北京实习期间，操晓春老师和张华老师的照顾与指导，让我提早进入到了博士生的状态，在 AI 领域进一步提高自己的视野和能力。也要感谢北京实习期间吴永林、汤子贤、加小俊、郑卓然、肖立强、袁智博、赵博文等师兄的照顾，以及小伙伴们李葭、张宇锟、王雨露、李嫣等的陪伴。

新的征程就要开始了，因为自己的选择，我不会放弃，我会勇敢地乘风破浪，迎接属于我的明天，乘着年轻，去做想做的事，见想见的人，努力实现自己的理想。

陈若愚

2021 年 5 月 31 日

参考文献

- [1] Li P, Zhang R, Chen J, Sun D, Lan J, Lin S, Song S, Xie Z, Jiang S. Development of a duplex semi-nested PCR assay for detection of classical goose parvovirus and novel goose parvovirus-related virus in sick or dead ducks with short beak and dwarfism syndrome[J]. Journal of Virological Methods, 2017, 249: 165–169.
- [2] Jones T A, Dawkins M S. Environment and management factors affecting Pekin duck production and welfare on commercial farms in the UK[J]. British Poultry Science, 2010, 51(1): 12–21.
- [3] Brown D J, Savage D B, Hinch G N. Repeatability and frequency of in-paddock sheep walk-over weights: implications for flock-based management[J]. Animal Production Science, 2014, 54(5): 582.
- [4] Corrand L, Gelfi J, Albaric O, Etievant M, Pingret J-L, Guerin J-L. Pathological and epidemiological significance of Goose haemorrhagic polyomavirus infection in ducks[J]. Avian Pathology, 2011, 40(4): 355–360.
- [5] Phiri I K, Phiri A M, Ziela M, Chota A, Masuku M, Monrad J. Prevalence and distribution of gastrointestinal helminths and their effects on weight gain in free-range chickens in Central Zambia[J]. Tropical Animal Health and Production, 2007, 39(4): 309–315.
- [6] Gouda G, Ali W. Useful Repercussions on 10 and 12-Week Marketing Body Weights of Laying Ducks with Early Index Selection on Body Weights and Linear Measurements[J]. Journal of Animal and Poultry Production, 2017, 8(8): 233–236.
- [7] González-García E, Alhamada M, Pradel J, Douls S, Parisot S, Bocquier F, Menassol J B, Llach I, González L A. A mobile and automated walk-over-weighing system for a close and remote monitoring of liveweight in sheep[J]. Computers and Electronics in Agriculture, 2018, 153(April): 226–238.
- [8] Wang K, Pan J, Rao X, Yang Y, Wang F, Zheng R, Ying Y. An Image-Assisted Rod-Platform Weighing System for Weight Information Sampling of Broilers[J]. Transactions of the ASABE, 2018, 61(2): 631–640.
- [9] Pezzuolo A, Guarino M, Sartori L, González L A, Marinello F. On-barn pig weight estimation based on body measurements by a Kinect v1 depth camera[J]. Computers and Electronics in Agriculture, 2018, 148(February): 29–36.
- [10] Shi C, Teng G, Li Z. An approach of pig weight estimation using binocular stereo system

- based on LabVIEW[J]. Computers and Electronics in Agriculture, 2016, 129: 37–43.
- [11] Menesatti P, Costa C, Antonucci F, Steri R, Pallottino F, Catillo G. A low-cost stereovision system to estimate size and weight of live sheep[J]. Computers and Electronics in Agriculture, 2014, 103: 33–38.
- [12] Taşdemir Ş, Ürkmez A, Inal Ş. A fuzzy rule-based system for predicting the live weight of holstein cows whose body dimensions were determined by image analysis[J]. Turkish Journal of Electrical Engineering and Computer Sciences, 2011, 19(4): 689–703.
- [13] Buranakarl C, Indramangala J, Koobkaew K, Sanghuayphrai N, Sanpote J, Tanprasert C, Phatrapornnant T, Sukhumavasi W, Nampimoon P. Estimation of Body Weight and Body Surface Area in Swamp Buffaloes using Visual Image Analysis[J]. Journal of Buffalo Science, 2012, 1(1): 13–20.
- [14] Kashiha M, Bahr C, Ott S, Moons C P H, Niewold T A, Ödberg F O, Berckmans D. Automatic weight estimation of individual pigs using image analysis[J]. Computers and Electronics in Agriculture, 2014, 107: 38–44.
- [15] Mollah M B R, Hasan M A, Salam M A, Ali M A. Digital image analysis to estimate the live weight of broiler[J]. Computers and Electronics in Agriculture, 2010, 72(1): 48–52.
- [16] Balaban M O, Chombeau M, Cırban D, Gümüş B. Prediction of the Weight of Alaskan Pollock Using Image Analysis[J]. Journal of Food Science, 2010, 75(8): E552–E556.
- [17] Balaban M O, Ünal Şengör G F, Soriano M G, Ruiz E G. Using Image Analysis to Predict the Weight of Alaskan Salmon of Different Species[J]. Journal of Food Science, 2010, 75(3): E157–E162.
- [18] Lines J A, Tillett R D, Ross L G, Chan D, Hockaday S, McFarlane N J B. An automatic image-based system for estimating the mass of free-swimming fish[J]. Computers and Electronics in Agriculture, 2001, 31(2): 151–168.
- [19] Humphries J M, Bookstein F L, Chernoff B, Smith G R, Elder R L, Poss S G. Multivariate Discrimination by Shape in Relation to Size[J]. Systematic Zoology, 1981, 30(3): 291.
- [20] Velardo C, Dugelay J-L. Weight estimation from visual body appearance[C/OL]. 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS). IEEE, 2010: 1–6.
- [21] Banhazi T M, Tscharke M, Ferdous W M, Saunders C, Lee S H. Improved Image Analysis Based System to Reliably Predict the Live Weight of Pigs on Farm: Preliminary Results[J]. Australian Journal of Multi-Disciplinary Engineering, 2011, 8(2): 107–119.

- [22] Tasdemir S, Urkmez A, Inal S. Determination of body measurements on the Holstein cows using digital image analysis and estimation of live weight with regression analysis[J]. *Computers and Electronics in Agriculture*, 2011, 76(2): 189–197.
- [23] Wongsriwraphon A, Pathumnakul S, Arnonkijpanich B. Image analysis for pig recognition based on size and weight[J]. *IEEE International Conference on Industrial Engineering and Engineering Management*, 2012: 856–860.
- [24] Amraei S, Abdanan Mehdizadeh S, Salari S. Broiler weight estimation based on machine vision and artificial neural network[J]. *British Poultry Science*, 2017, 58(2): 200–205.
- [25] Wang Y, Yang W, Winter P, Walker L. Walk-through weighing of pigs using machine vision and an artificial neural network[J]. *Biosystems Engineering*, 2008, 100(1): 117–125.
- [26] Jun K, Kim S J, Ji H W. Estimating pig weights from images without constraint on posture and illumination[J]. *Computers and Electronics in Agriculture*, 2018, 153(July): 169–176.
- [27] Zong C, Li Z, Zhuang Y, Yu J, Teng G. Comparisons of non-contact methods for pig weight estimation[C/OL]. 2018 ASABE Annual International Meeting. St. Joseph, MI: American Society of Agricultural and Biological Engineers, 2018: 1.
- [28] Gonzalez R C, Woods R E. *Digital Image Processing*[M/OL]. Prentice hall Upper Saddle River, NJ..
- [29] 周志华. 机器学习[M]. 清华大学出版社, 2016.
- [30] 李航. 统计学习方法[M]. 清华大学出版社, 2012.
- [31] Goodfellow I, Bengio Y, Courville A. *Deep Learning*[M/OL]. MIT Press. MIT press Cambridge, 2016.
- [32] Bishop C M. *Pattern recognition and machine learning*[M]. springer. springer, 2006.
- [33] ZHU J, SU H, ZHANG B. Toward the third generation of artificial intelligence[J]. *SCIENTIA SINICA Informationis*, 2020, 50(9): 1281.
- [34] Samuel A L. Some Studies in Machine Learning Using the Game of Checkers[J]. *IBM Journal of Research and Development*, 1959, 3(3): 210–229.
- [35] Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain.[J]. *Psychological Review*, 1958, 65(6): 386–408.
- [36] Minsky M, Seymour A P. Perceptrons[EB](1969).
- [37] Rumelhart D E, Hinton G E, Williams R J. *Learning internal representations by error propagation*[R/OL]. The MIT Press, 1986.

- [38] Quinlan J R. Induction of decision trees[J]. Machine Learning, 1986, 1(1): 81–106.
- [39] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995, 20: 273–297.
- [40] Hinton G E. Reducing the Dimensionality of Data with Neural Networks[J]. Science, 2006, 313(5786): 504–507.
- [41] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84–90.
- [42] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators[J]. Neural Networks, 1989, 2(5): 359–366.
- [43] Fukushima K, Miyake S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition[M]. Competition and cooperation in neural nets. Springer, 1982: 267–285.
- [44] LeCun Y, Boser B, Denker J S, Henderson D, Howard R E, Hubbard W, Jackel L D. Backpropagation Applied to Handwritten Zip Code Recognition[J]. Neural Computation, 1989, 1(4): 541–551.
- [45] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278–2324.
- [46] Deng J, Dong W, Socher R, Li L-J, Kai Li, Li Fei-Fei. ImageNet: A large-scale hierarchical image database[C/OL]. 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 248–255.
- [47] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C/OL]. 3rd International Conference on Learning Representations, ICLR 2015..
- [48] Szegedy C, Wei Liu, Yangqing Jia, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions[C/OL]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015: 1–9.
- [49] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, 2016-Decem: 770–778.
- [50] Li H, Wang J, Tang M, Li X. Deep Sparse Rectifier Neural Networks[C/OL]. Proceedings of the fourteenth international conference on artificial intelligence and statistics. JMLR.org, 2011: 315–323.
- [51] Clevert D-A, Unterthiner T, Hochreiter S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)[J]. arXiv:1511.07289 2015(1997): 1–14..

- [52] Klambauer G, Unterthiner T, Mayr A, Hochreiter S. Self-normalizing neural networks[C/OL]. Advances in Neural Information Processing Systems, 2015.
- [53] Dugas C, Bengio Y, Bélisle F, Nadeau C, Garcia R. Incorporating second-order functional knowledge for better option pricing[J]. Advances in Neural Information Processing Systems, 2001.
- [54] Mikolov T, Kombrink S, Burget L, Černocký J, Khudanpur S. Extensions of recurrent neural network language model[J]. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2011: 5528–5531.
- [55] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research, 2014, 15: 1929–1958.
- [56] Woo S, Park J, Lee J-Y, Kweon I S. CBAM: Convolutional Block Attention Module[C/OL]. Computer Vision - ECCV 2018 - 15th European Conference. .
- [57] Szegedy C, Ioffe S, Vanhoucke V, Alemi A A. Inception-v4, inception-ResNet and the impact of residual connections on learning[J]. 31st AAAI Conference on Artificial Intelligence, AAAI 2017, 2017: 4278–4284.
- [58] Ruder S. An overview of gradient descent optimization algorithms[J]. 2016: 1–14. .
- [59] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization[J]. 2014: 1–15. .
- [60] Qian N. On the momentum term in gradient descent learning algorithms[J]. Neural Networks, 1999, 12(1): 145–151.
- [61] Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick C L. Microsoft COCO: Common Objects in Context[C/OL]. Computer Vision - ECCV 2014 - 13th European Conference, 2014.
- [62] Wang X, Xiao T, Jiang Y, Shao S, Sun J, Shen C. Repulsion Loss: Detecting Pedestrians in a Crowd[C/OL]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 7774–7783.
- [63] Luo G, Dong S, Wang K, Zuo W, Cao S, Zhang H. Multi-Views Fusion CNN for Left Ventricular Volumes Estimation on Cardiac MR Images[J]. IEEE Transactions on Biomedical Engineering, 2018, 65(9): 1924–1934.
- [64] Chollet F. Xception: Deep learning with depthwise separable convolutions[J]. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017, 2017-Janua: 1800–1807.
- [65] Xie S, Girshick R, Dollar P, Tu Z, He K. Aggregated Residual Transformations for Deep

- Neural Networks[C/OL]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 5987–5995.
- [66] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. .
- [67] Girshick R. Fast R-CNN[J]. Proceedings of the IEEE International Conference on Computer Vision, 2015, 2015 Inter: 1440–1448.
- [68] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149.
- [69] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, 2016-Decem: 779–788.
- [70] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[J]. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017, 2017-Janua: 6517–6525.
- [71] Redmon J, Farhadi A. YOLO v.3[J]. Tech report, 2018: 1–6.

附录

附录 A 附加数据图与表

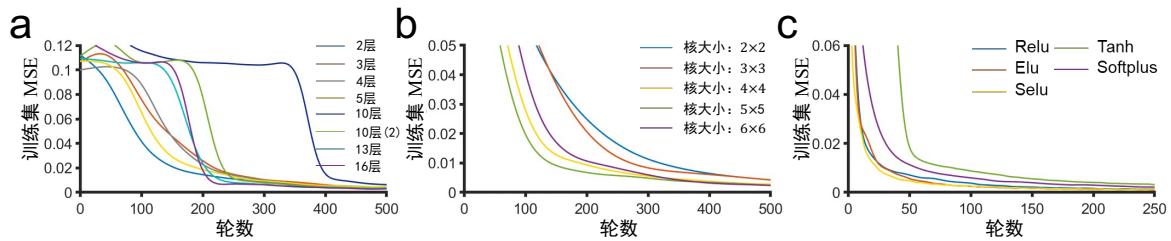


图 A1 CNN 模型在不同超参数下，训练集上的学习曲线（用 Smooth spline 方法平滑过）。a. 不同卷积层深度条件下训练集的学习曲线；b. 不同卷积核大小下训练集的学习曲线；c. 不同激活函数下训练集的学习曲线

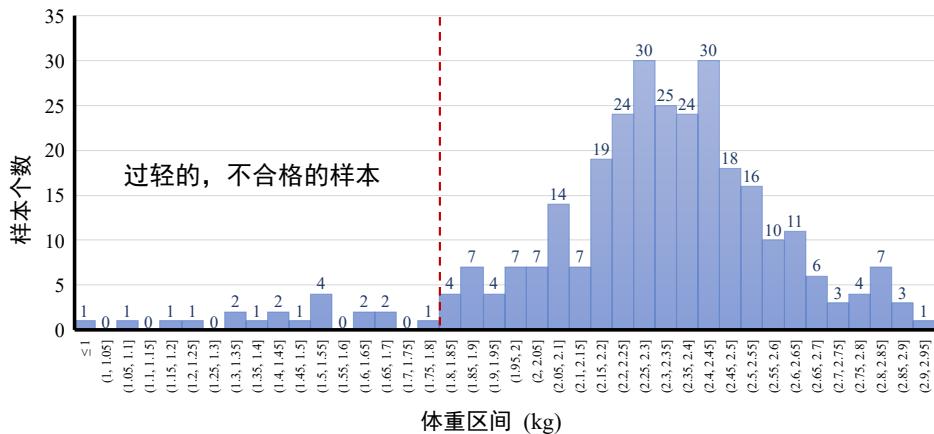


图 A2 采集的数据体重的分布图，以 1.8 千克为界限，只采用大于 1.8kg 的数据训练与评估模型

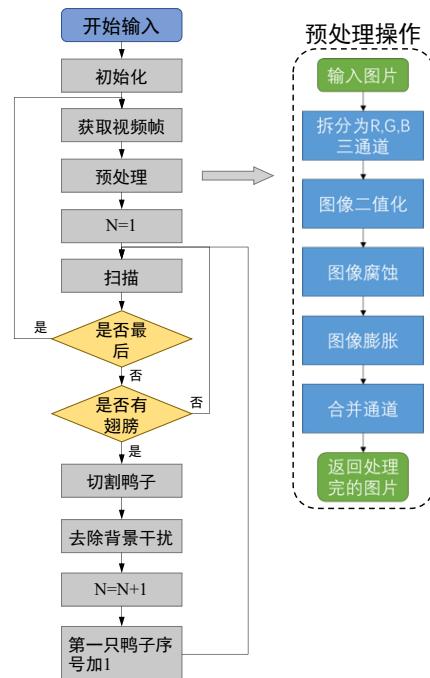


图 A3 切割鸭子图像算法的流程图

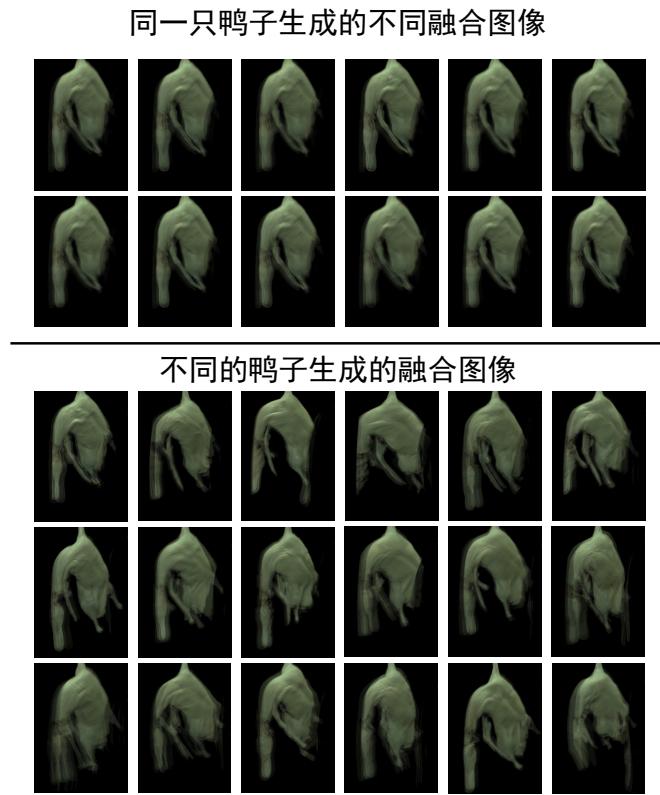


图 A4 随机融合图像结果图

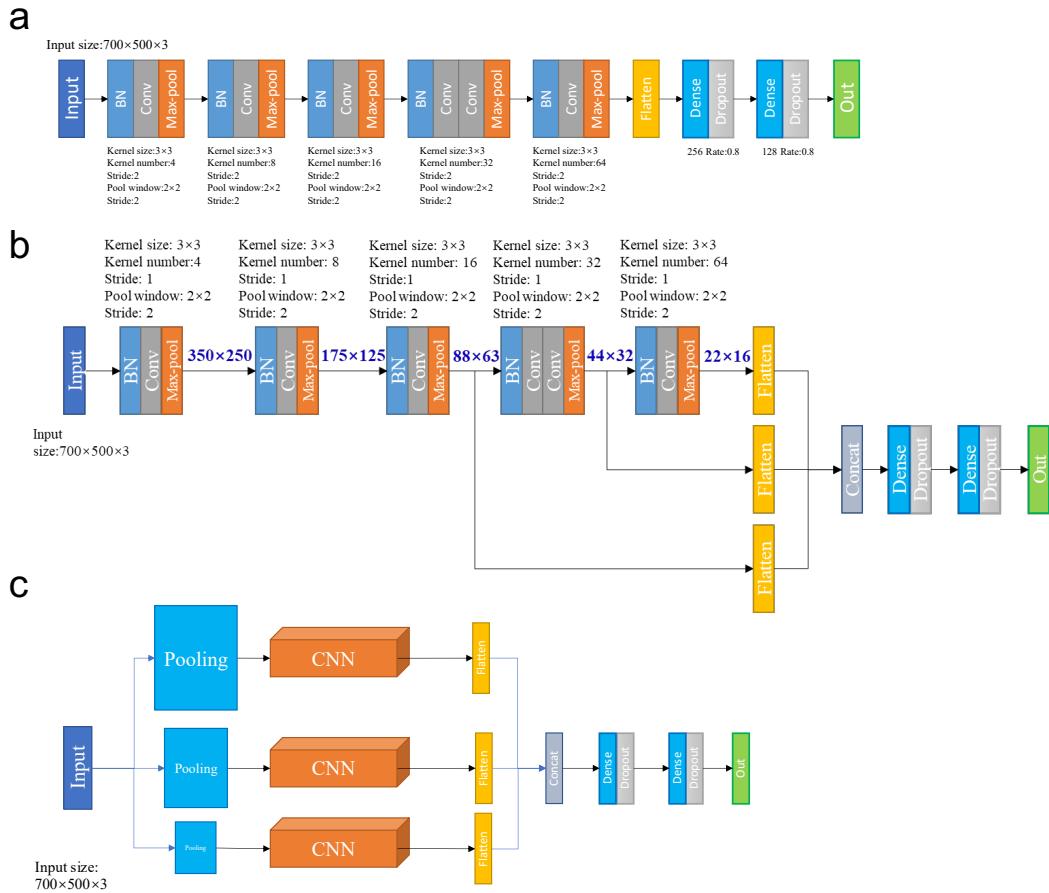


图 A5 a. Net-6-BN 网络结构; b. Net-Width 网络结构; Net-multi-scale 网络结构

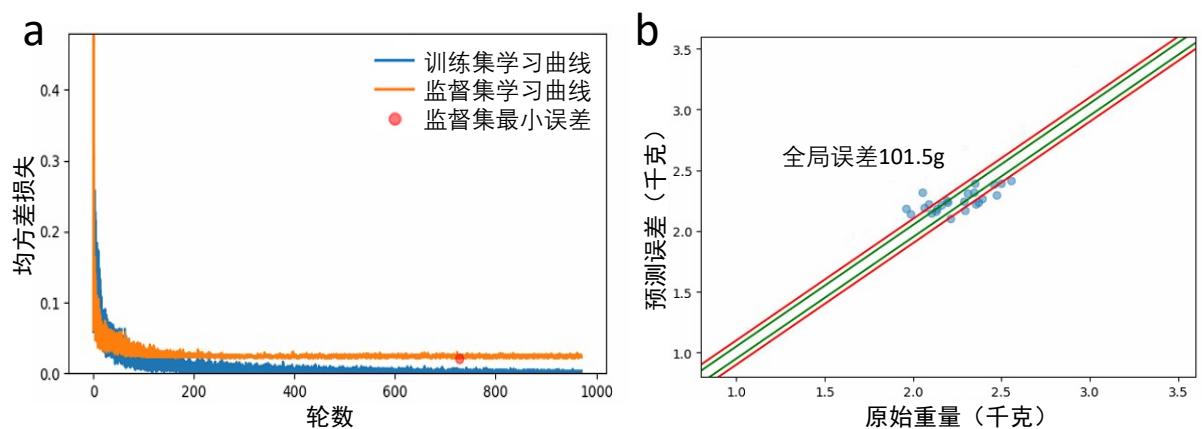
附录 B 证明

本章用于证明图像中多余的部分影响对动物体重的识别率。我们共进行两组实验，一组实验用第3章的方法，一组实验用第4章的图像融合方法，对比模型的精度。数据集我们选用第4章的数据集。本实验使用的卷积神经网络为表4.1中Net-6-BN结构，训练参数配置与4.2节一致。

针对第一种情况，模型的输入数据如图B1所示，两侧的多余物没有被去除，进行单图的端到端的训练与预测，模型的学习曲线如图B2 a所示，模型的预测结果如图B2 b所示，最终的MAD为101.5克。



图B1 未清除多余物的图像



图B2 第一种数据情况下模型结果；a. 学习曲线；b. 预测结果

针对第二种情况，我们将数据集采用4.1.2节的图像随机融合的方式进行预处理。

模型的输入数据如图 B3 所示，进行单图的端到端的训练与预测，模型的学习曲线如图 B4 a 所示，模型的预测结果如图 B4 b 所示，最终的 MAD 为 112.7 克。



图 B3 未清除多余物的融合图像

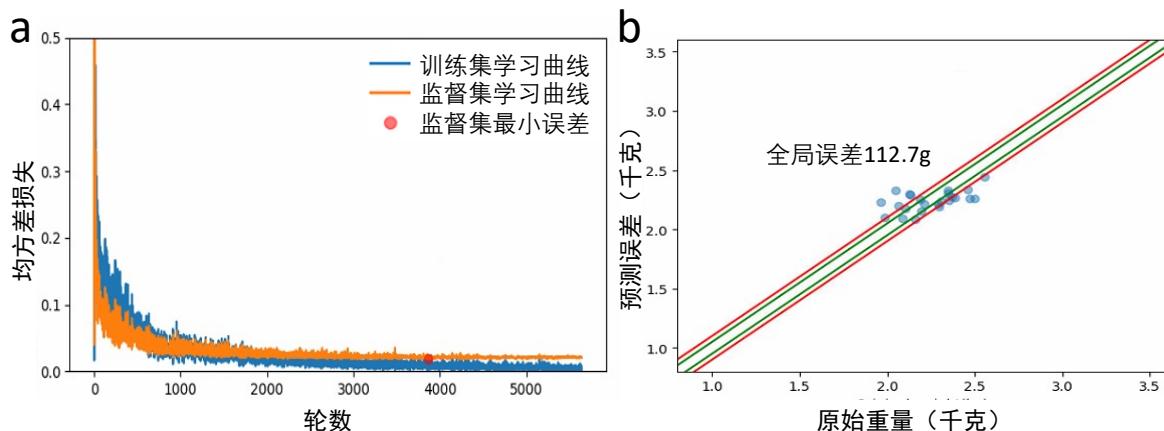


图 B4 第二种数据情况下模型结果；a. 学习曲线；b. 预测结果

最终的结果都比表 4.2 中的结果差非常多，因此我们可以认为目前提出的模型无法有效训练出一个能去除干扰的功能，因此我们在预处理时需要去除这些干扰。

附录 C 英文原文

Comparisons of Non-contact Methods for Pig Weight Estimation

ABSTRACT. Non-contact methods for estimating pig weight is popular in precision livestock farming utilizing computer vision techniques, and has proved success in field applications. The development of pigs body size and body weight estimating based on machine vision can be generally divided into two stages: two-dimensional measurement phase and three-dimensional

measurement phase. In this study, non-contact pig weighing methods from literatures were summarized. The advantages and limitations of each method were assessed by comparisons of weight estimation accuracy, maneuverability, and cost efficiency of the equipment.

Keywords. Non-contact method, Pig weight estimation, precision livestock farming, computer vision techniques

1. Introduction

As the scale of farm increases, even small alterations to production practices can have large impacts on production efficiency and overall profit of a farm. Thus, appropriate automation to improve the efficiency has been increasingly applied in livestock production and becomes a trend. For large scale farms with increasing number of animals to take care of, a number of automation approaches have been accordingly developed. Non-contact estimation of pig's body size and weight using computer vision technique is one of the primary applications for the development of automatic monitoring systems in livestock production. Automatic monitoring animals based on computer vision is a novel method and has been applied to various animal species, especially pigs (Brandl & Jorgensen, 1996). The information of pig body size and weight is crucial for pig production management, and plays an important role in selection of breeding and association with the indices that could reflect pork meat quality, such as back-fat thickness, muscle nutrient compositions, lean meat ratio, etc. (Schofield, 1990; Schofield, Marchant, White, Brandl, & Wilson, 1999). Accurate monitoring daily weight gain of individual pig allows producers to optimize nutritional management practices, predict and control shipping weights, and potentially assist in monitoring herd health (Marchant, Schofield, & White, 1999; Schofield et al., 1999).

Conventional weight measurements are mainly performed manually in pig farms, which suffers a number of drawbacks as manual weight measurement is labor intensive and can cause stress to pigs. Besides, traditional weight measuring equipment is apt to malfunction due to direct exposure to stools, urine, and collision from animals. In contrast, machine vision-based weighing of pigs is a non-intrusive and fast approach, and has been proved accurate in predictions (Wang, Yang, Walker, & Rababah, 2008; Wang, Yang, Winter, & Walker, 2008). Machine vision is used to acquire videos and graphic data for image processing, providing imaging-based automatic inspection of pig weight and subsequent controls. With the development of machine vision technology, pig body size and weight measurements based on machine vision could be divided into two-dimensional measurement and three-dimensional

measurement.

2. Two-dimensional measurement approaches

For approaches using two-dimensional measuring solutions, cameras are mounted either on the top or side of the pig, detecting pig's side and top view pictures. Specific technologies are applied in different studies, and are summarized as follows.

2.1. Visual Image Analysis (VIA)

Schofield (1990) first proposed and used the method of visual image analysis to estimate pigs weight. Images of the pig body's side and back can be acquired from the camera by mounting the camera on the side of the pig body and equipping a 45 degree angled mirror on the top (Fig. 1). This study showed that the correlations between back area and pig weight was the best. Besides, the experimental results also showed that this technology was better for predicting weight robustness and the estimated error was less than 5 %.

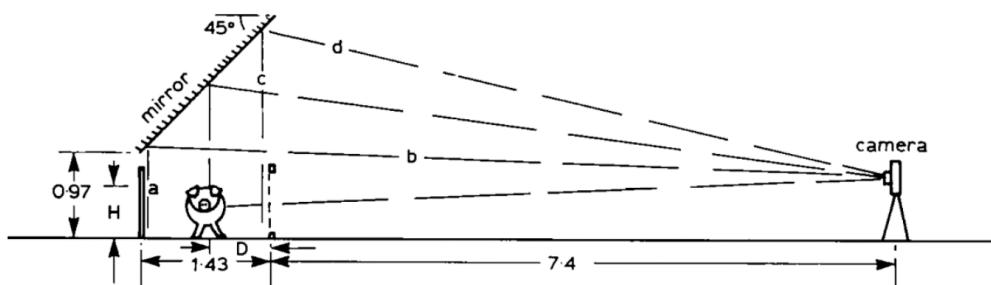


Figure 1. The benchmark image collection system from study of Schofield etc. (1990).

Afterwards, Schofield et al. (1999) built a visual image analysis system which can collect photo automatically in an actual pig farm to perform all image acquisition, processing, and timing operations. They used a low-light (4 lux) monochrome CCD camera equipped with a wide-angle lens (4 mm) for image collection. The camera was installed above the feeding station and two tungsten filament lamps were used as the supplement light. About 2.2 cd/m² illuminance could be achieved during the daytime in the feeding station. The system collected the images of pig back and recorded the growth rate of the three breeds of pigs through a prediction model established with the body weight. The experimental results showed that the estimated average weight error was within 5 %, and in the range of 60-90 kg, the error of some varieties was less than 2 %.

Marchant et al. (1999) tested an image analysis system to estimate pig's weight and calibrated the intercept of the model by using a hand-weighed data. The estimation error of the 80 kg pig was less than 1 %.

Doeschl-Wilson, Whittemore, Knap, and Schofield (2004) used the visual image analysis

to describe different breeds of pigs about body size and body shape, and to assess the changes in body size and shape during growth. Besides, pig growth was monitored by platform scales in the feeding station and visual imaging system mounted above the feeding station. When each pig was in the feeding station, the quality of the living body was measured and a two-dimensional digital image of the pig could be taken by the camera above. And then, the image was processed by an algorithm, the size and shape measurement results could accurately describe the growth of different pigs in different groups. White, Schofield, Green, Parsons, and

Whittemore (2004) used the VIA system and the platform balance to detect three breeds of pigs growth rate. By comparison, the results showed that using VIA system to test the pig was faster and more accurately than using platform balance. However, if the pig stayed in an abnormal position for a long time in front of the camera, the weight estimation of VIA will produced a certain deviation.

Parsons, Green, Schofield, and Whittemore (2007) found that using VIA system was able to estimate the weight of pigs with an average error less than 3.5%. In addition, the system could be added to the Integrated Management System (IMS). The data obtained from VIA system could be used to real time optimize pig s growth models, controlling pig s weight gain and fat thickness.

2.2. Projected area of pig

Minagawa and Ichikawa (1994) placed the pig in a black box which was illuminated with two incandescent lamps and a video camera on the top to take pictures in order to calculate the projected area of the pig (Fig. 2). An index relationship between projected area and body weight was established by using data from 33 pigs with a standard error of 0.9 kg.

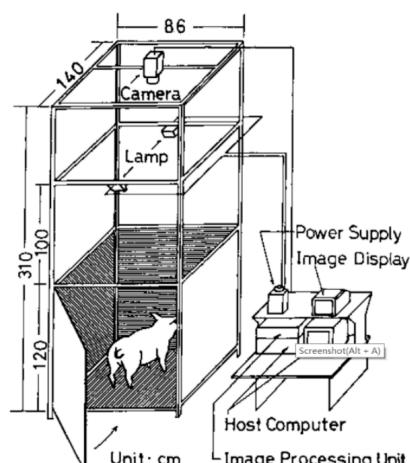


Figure 2. System for taking projected image area of a pig with a camera from above (Minagawa et al., 1994).

2.3. Ruler calibrated camera

Wang, Yang, Winter, and T. Walker (2006) developed contactless pig weight measurement system based on machine vision (Fig. 3). A total number of 187 pigs were used as experimental subjects. They placed a ruler at the height of the pig to calibrate the camera and used Matrox software to extract pigs. Then, a relationship model was established for estimating pig's weight. The results showed that the correlation coefficient between projection area and body weight was the highest ($R^2=0.96$). In addition, there was a good correlation between body width and body weight ($R^2=0.95$).

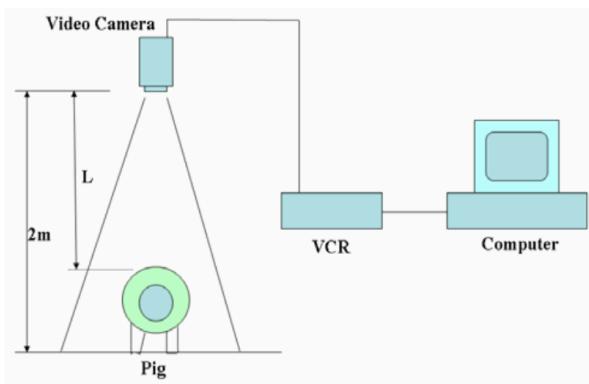


Figure 3. Schematic diagram of the measuring system from the study of Wang et al. (2006).

Two years later, they improved the pig contour extraction algorithm for estimating the weight of pig images (Wang, Yang, Walker, et al., 2008). The machine vision inspection system was used in the experiment included a camera and a computer to realize the process from image acquisition, image processing and feature extraction to analyse data. First, the pig was guided through a passage about 1m width. Second, the camera was mounted on the ceiling and a 40W white fluorescent lamp was placed on each side of the passage, approximately 1 m above the ground, to provide stable illumination. And to obtain a higher contrast of the image, they placed the black rug on the floor of the channel. In the same year, the team improved the machine vision to estimate the pig weight system (Wang, Yang, Winter, et al., 2008). The system did not require the pigs to stay still. As each pig passed through the channel, the camera at the top was constantly shooting. The camera was initially set to "manual" mode in the exposure mode and the aperture was set to the widest position. Because of that, the exposure time per photo at a given light intensity could be minimized to reduce the chance of ambiguity due to pig movement. After that, the captured images were processed to obtain the body size parameters. The relationship between body size and body weight was established by artificial neural network. The experimental results showed that the relative error was 3%.

The measurement accuracy with this method was high, but the method of using a ruler to calibrate the camera was only applicable when the object distance was fixed. The pig height might change because of time and individual changes, this method was difficult to apply in actual production.

2.4. Special marks on pig's body

Yang, Teng, Li, and Shi (2006) planned to collect back and side images of pigs in order to estimate their weight (Fig. 4). They choose the empty pigsty and paint the green paint around. They took the paper pasted on the back of the pig as the calibration object, and draw the rectangular box with the known area on the wall as the reference system. In addition, they placed two high definition digital zoom integrated camera around the pigsty to collect the top view and side view of the pig in real time. The height and area of the object were obtained by the overlook and side view. The correlation between the area of the back, the body height and the weight was studied. After that, the weight of 40 pigs was estimated and the average relative error was 3.2%.

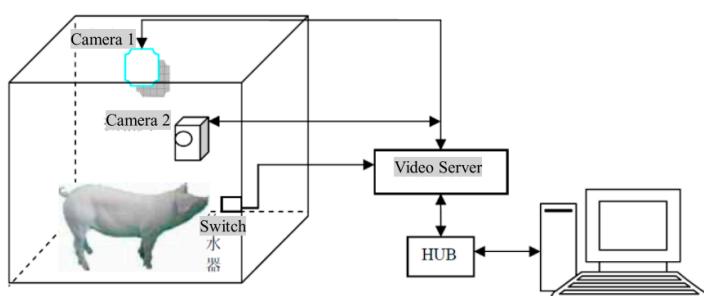


Figure 4. Schematic diagram of the measuring system from the study of Yang et al. (2006).

Since 2013, cameras coupled with special patterns on the pig's back were used for individual pig identification and weight estimation by a Belgian group (Fig. 5) (Kashiha et al., 2013; Kashiha et al., 2014). First, by spraying special patterns on the pig's back and using image processing techniques to locate and identify individual pigs. Second, they used the ellipse fitting to smear the head and tail. And then, the remaining torso area was used to establish the estimation model between the pixel area and body weight. The result showed that the accuracy of pig weight estimation at the population level was 97.5%, the error was 0.82 kg. Besides, it also illustrated that the accuracy at the individual level was 96.2%, and the error was 1.23 kg.

2.5. Calculating the number of pixels in image

Kollis, Phang, Banhazi, and Searle (2007) began to research on pig weight prediction. A rectangular area of 900x1350x364 mm is constructed, and only one pig can be measured at a time. In a Logitech QuickCam Messenger network camera, the pig image is mounted on the

frame above 3M on the measurement area with a resolution of 640x480. The entire system used fluorescent light for lighting and improves the shooting quality. Two 50W downlights were also mounted on the frame, on either side of the camera. After analyzing the collected images, it was found that there was a strong correlation between pig back area, spine length and pig weight.

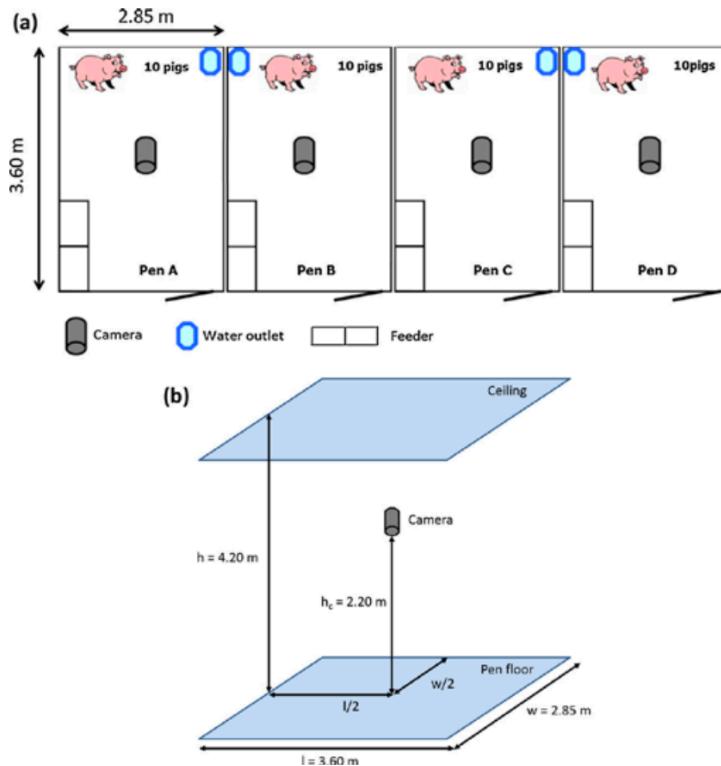


Figure 5. (a) Ground plan of the pens; (b) position of the camera on top of a pen (Kashiha et al., 2014).

Banhazi, Tscharke, Ferdous, Saunders, and Lee (2009) developed a computer-based pig weight estimation system that uses body-scale pixels to estimate body weight. The experiment showed that the average error of estimated pig weight is $\pm 1.18 \text{ kg}$.

Tscharke and Banhazi (2013) improved the original pig weight estimation system to add the function of automatically searching for pictures, identifying the shape of the pig, extracting the body size of the pig and estimating the weight of the pig, and comparing it with the actual weight. The results showed that between 60-120 kg, the group average error was 2 kg, and the average relative error was within 1%. Since the system cannot perform individual identification, only the overall average weight can be estimated.

Estimating the body size based on the number of pixels does not have universality and will no longer apply when the object distance changes.

2.6. User interaction

Wongsriworaphon, Arnonkijpanich, and Pathumnakul (2015) used a non-emergency pig weight estimation system that can be applied in a farm environment. The camera was set up on

the top of a channel to capture the back image of the pig (Fig. 6); afterwards, the contour of the pig body was manually extracted based on user interaction, and the average distance from the center of mass of the pig to the boundary point and the perimeter of the pig were taken as the characteristic values of the image. The characteristic values also as an input to the weight estimation model. The results showed that the average estimation error of the method for body weight is less than 3%, and can be effectively applied in a real farm environment, but the dependence on labor is large and automation is not easy.



Figure 6. The installation of a camera to capture the top-view pictures (Wongsriworaphon et al., 2015)

In the two-dimensional measurement stage, most detection systems only contain one camera. The acquired planar images contain only data of body length, body width or back area, which is not so versatile to establish weight estimation models. When the installation distance or the setting parameters of the camera change, the calibration must be performed again, so it is not practical in the field production.

3. Three-dimensional measurement approaches

With the development of machine vision technology, the cost of the three-dimensional measurement is reduced, and the body weight detection of pigs has entered the stage of three-dimensional measurement. The three-dimensional measurement obtains data including the height information of pig, which makes the estimation of the pig's weight becomes more accurate.

3.1. Stereographic projection technology

Minagawa and Murakami (2001) began to study the prediction of pig body weight by stereographic projection technique (Fig. 7). The slide projector was used as a parallel light source and the grid was divided in the projected slide. The projector was fixed on the top of the drinking water device and the camera was installed beside it. Based on the principle of geometrical optics, the height of the pig was calculated according to the grid spacing between

the ground and the back of the pig body. A mathematical model of body weight prediction was established by using the body height and the back area of the pig with removed ears area. A total of 5 times estimation were conducted for 12 pigs aged from 81 days to 98 days, with an average relative error of 2.1%.

In 2003, they improved the color of the lighting system in order to enhance the contrast between pig image and surrounding environment (Minagawa, Taira, & H. Nissato, 2003). Study showed that clearer pig images were obtained under the red and yellow grids slides. The weight of the 7 Landrace pigs was estimated by the system, with an average error of 1.3%.

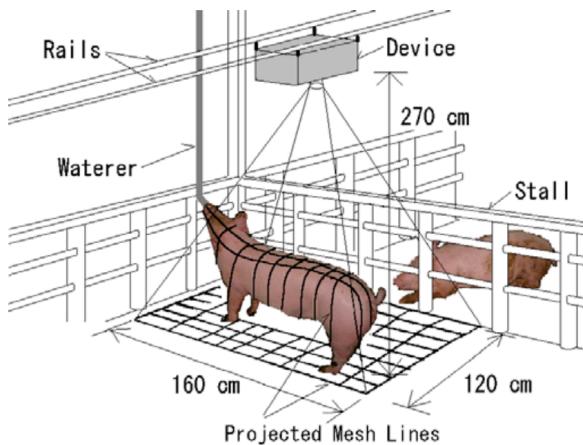


Figure 7. Schematic diagram of the measuring system from study of Minagawa et al. (2003).

Figure 7. Schematic diagram of the measuring system from the study of Minagawa et al. (2003).

With this method, the body height of the pig can be obtained without an additional reference object, but the method of adopting the light projection method is easily affected by the external light, and the application effect in the actual pigsty is poor.

3.2. Binocular stereo technology

Fu and Teng (2009) did some study on pig body size measurement based on the principle of binocular vision, and a pig body size measurement and weight prediction system was designed (Fig. 8). A CCD camera (SSC-E473P, SONY, Japan) combined with customized lens (H051-MP, CBC, Japan) as an image acquisition device installed above the pigs, and the view range of 1.5m – 1.0m was set when camera located at the height of 80.0 cm. Four fluorescent lamps as the lighting source were used as a way to illuminate the pig's body image. The pig body was irradiated on all sides. An extraction algorithm of the pig body measuring point based on the depression structure and the convex hull analysis was developed according to the characteristics of the pig's contour. The results showed that the relative error of the length of pigs was 0.85%, the maximum and minimum detection errors of pig body height and body width were 2.23%, 0.14% and 3.31% and 0.19%, respectively. Through principal component

analysis, a pig weight prediction model based on pig body size parameters was established. The fitting coefficient of the model was 0.955, and the residual error standard deviation was 4.49 kg. The field application showed that the relative error of the average weight prediction was 0.77%, which proved practical value of this system. However, the prediction model of the system used many parameters, such as length, width and height, and key points of the pig body.

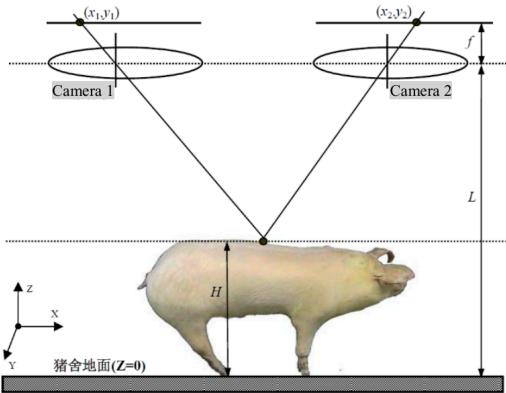


Figure 8. The principle diagram of pig's body dimension detection based on binocular stereovision (Fu et al., 2011).

Liu, Teng, Fu, and Li (2013) combined with the actual situation of pig house, put forward the algorithm for extracting individual information of pig body under complex background, removal algorithm of pig body head and tail based on envelope analysis, and the algorithm of extracting coordinates of measuring points from the complex pig body with a certain bending attitude. A CCD camera (SSC-E473P, SONY, Japan) equipped with customized lens (H051-MP, CBC, Japan) was used as the image acquisition device to collect the RGB colored pig back from above. On the basis of obtaining the contour of the pig body, the cutting points of the head and tail are identified by the envelope analysis with the head and tail removed. Then the coordinates of the 9 measuring points of the body size of the pig body were identified by the extraction algorithm of the body scale. The average relative error of body size was 2.09% and the average relative error of body weight was 3.38%. At the same time, he reconstructed the three-dimensional surface model of the pig specimen [27], used the non-contact 3D scanner (Vivid-910, China), and the measurement distance was 1.5m. The three dimensional point cloud data of the pig body was successfully obtained. The average relative error was 0.17% and the weight measurement error was 2.7%. Therefore, it was feasible to reconstruct the pig body by binocular stereo vision.

Li, Du, Mao, and Teng (2016) and Shi, Teng, and Li (2016) developed a contactless pig weight estimation system based on binocular stereo vision technology. Two black-and-white cameras (acA1300-30gm, Basler, Germany) with customized lens (Computar H0514-MP, CBC,

Japan) were used to capture images. The camera were installed at the top of the feeding area, and the strip fluorescent lamp were used as the light source to make up the scene. A pig body contour extraction algorithm based on the depth image was developed. To solve the problem of gray contour extraction algorithm under the harsh light environment in pig farms, the pig body corner normalization coefficient to screen the measurement point and the distance from the long axis of ellipse to the contour were used to construct a new body ruler detection algorithm. A pig body size and weight noncontact detection software was therefore developed. The detection accuracy of body length, body width, body height, hip width and hip height were 1.44 %, 5.81 %, 4.94 %, 2 % and 1.64 %, respectively. The average relative error of the estimated weight of 268 groups of 4 pigs in the experiment was 2.14%. The full-scale automatic estimation of pig body weight and body weight was realized by this system.

3.3. Study on pig's weight based on depth images

Depth image, also known as distance image or range image, records the distance information between the points and the camera in the scene, and reflects the 3D features of the object surface. The extraction of the object contour based on the depth image can effectively distinguish the image foreground and the background [29]. Kinect is a 3-D somatosensory camera developed by the Microsoft Corp. It has low cost and can quickly obtain the depth information of the object. Compared with the ordinary camera, it has the characteristics of simple calibration and easy to denoise. In recent years, more and more studies have been done on pig body size using Kinect camera.

Kongsro (2014) used Kinect cameras to design a system for estimating the weight of pigs, in which Kinect was mounted on a telescopic aluminum rod (Fig. 9). After the processing the depth image, the body size of the pig was extracted for calculating the volume of the pig, and the weight of the pig was further estimated. The experimental results showed that the system successfully estimated the body weight of two different breeds of pigs, and the detection error was 4%-5%.

Lao et al. (2016) used Kinect camera to design a system for recognizing the behavior of sows. A 3D Kinect camera was installed at 2.20m above the farrowing box and the camera captured images at intervals of 6 seconds. After the initial processing of the image, important features of the sow were extracted, including its center-of-mass coordinates, head coordinates, hip coordinates, shoulder coordinates and waist coordinates. The sows in the depth image were divided into 7 parts (whole body, upper half body, lower half body, head, shoulders, waist and buttocks) and then the average of the depth pixels of each sow was calculated. By setting some

thresholds, the depth values of the pixels in each part are compared with the threshold to determine the behavior of the sow. The experimental results show that the recognition rates of sow lying, sitting standing deafness eating drinking behavior are 99.9%, 96.4%, 99.2%, 78.1%, 97.4%, and 92.7%, respectively, and the recognition rate of different behavioral conversions was 63.9%.



Figure 9. The Kinect prototype for pig weight measurement (Kongsro et al., 2014).

Liu, Teng, Zhang, Li, and Guo (2014) used the Kinect camera as an image acquisition device to collect the hog motion depth image sequence at a speed of 30 frames/s, and performed skeleton extraction, pruning, and other processing on each frame depth image. After the establishment of the pig movement model, a method for calculating the complete step of the pig walking was proposed. Finally, the step frequency characteristics of the pig were extracted by calculating the ratio of the length of the complete step of the sequence to the time of the sequence acquisition. Experimental results show that the accuracy of this method is 82.1%.

The Kinect camera solves the problem of quickly acquiring information about the depth of the pig, but it has flaws that the resolution of the acquired depth image is low and the detection accuracy is worse than 1m. So it is often used in behavior analysis or trait evaluation and the error in obtaining body measurement and body weight information is large.

4. Summary

Machine vision-based body weight measurement technology can be divided into two-dimensional measurement phase and three-dimensional measurement phase. For two-dimensional measurement approaches, most detection systems are equipped with one camera, either mounted on the top or side of the pig. In the phase of three-dimensional measurement, the machine vision equipment obtains the plane parameters as well as the height information of the pigs, which makes the prediction of body side of pig body more accurate, and the accuracy

of the model is also higher comparing with two-dimensional measurement approaches. The current machine vision-based body weight measurement technology for pig body mainly has the following problems

1. Most systems are unable to effectively identify individual pigs and can only detect the average weight of a pig group;
2. Many measuring systems need special facilities and even reconstructions of the pig house to meet the requirement of the machine vision-based weight measuring equipment, which makes the system not practical in field applications.
3. Some systems are lack of automation and require manual assistance;
4. Most of the system's collections of grayscale images are from extracting captured pictures, and the body-scale extraction algorithm is vulnerable to the external environment and pigs' activities.
5. Most two dimensional systems use the plane parameters such as back area, body length, and body width to establish relationship with body weight, neglecting the height of the pig body, which makes the estimation of the pig body shape not accurate;
6. Some systems extract many body-size information and create complex estimating algorithms, but the accuracy of estimation is low;
7. Some systems cost a lot of money, and have poor practicality in actual production;
8. The weight estimation accuracy of a pig group is higher than that of an individual pig. Therefore, the development of a low-cost, high-precision, automated, and practical non-contact body weight measurement system is one of the important research directions in the future.

附录 D 中文译文

猪体重估计的非接触方法比较

摘要：利用计算机视觉技术在精确畜牧养殖中广泛应用于猪体重的非接触估算方法，并在田间应用中取得了成功。基于机器视觉的猪体大小和体重估计的发展一般可分为二维测量阶段和三维测量阶段。本研究总结了文献中的非接触式猪称重方法。通过对设备重量估计精度、可操作性和成本效率的比较，评估了每种方法的优点和局限性。

关键词：非接触法，猪体重估算，精准畜牧养殖，计算机视觉技术

1. 导言

随着农场规模的扩大，即使是对生产实践的微小改变也会对农场的生产效率和整体

利润产生很大的影响。因此，适当的自动化来提高效率已经越来越多地应用于畜牧业生产，并成为一种趋势。对于需要照顾的动物越来越多的大型农场，相应地开发了一些自动化方法。利用计算机视觉技术对猪的体型和体重进行非接触估计是家畜生产中自动监测系统的主要应用之一。基于计算机视觉的自动监测动物是一种新的方法，已应用于各种动物物种，特别是猪(Brandl&Jorgensen, 1996)。猪体大小和体重信息是猪生产管理的关键，在选择育种和与反映猪肉品质的指标（如背膘厚度、肌肉营养成分、瘦肉率等）的关联方面起着重要作用。(Schofield, 1990 年； Schofield, Marchant, White, Brandl, & Wilson, 1999 年)。准确监测个别猪的每日体重增加，使生产者能够优化营养管理做法，预测和控制运输重量，并有可能协助监测牛群健康(Merchant, Schofield, & White, 1999 年； Schofield 等人, 1999 年)。

传统的体重测量主要是在猪场手工进行的，这存在一些缺点，因为手工体重测量是劳动密集型的，会给猪带来压力。此外，传统的体重测量设备由于直接暴露于粪便、尿液和动物碰撞而容易发生故障。相反，基于机器视觉的猪称重是一种非侵入性和快速的方法，并在预测中被证明是准确的(Wang, Yang, Walker, & Rababah, 2008; Wang, Yang, Winter, & Walker, 2008)。机器视觉用于获取视频和图形用于图像处理的数据，提供基于成像的猪体重自动检查和后续控制。随着机器视觉技术的发展，基于机器视觉的猪体尺寸和体重测量可分为二维测量和三维测量。

2. 二维测量方法

对于使用二维测量解决方案的方法，相机安装在猪的顶部或侧面，检测猪的侧面和顶部视图图片。具体技术应用于不同的研究，总结如下。

2.1. 视觉图像分析(VIA)

斯科菲尔德 (1990) 首先提出并使用视觉图像分析方法来估计猪的体重。猪体的侧面和背面的图像可以从相机中获得，方法是将相机安装在猪体的侧面，并在顶部安装一个 45 度角的镜子(图。1)。研究表明，背面积与猪体重的相关性最好。此外，实验结果还表明，该技术具有较好的预测权重鲁棒性，估计误差小于 5%。

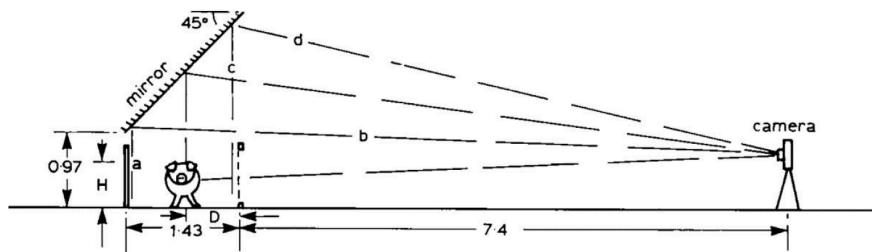


图 1 从斯科菲尔德等人的研究中得出的基准图像采集系统。

随后，Schofield 等人。（1999）建立了一个视觉图像分析系统，可在实际猪场自动采集照片，进行所有图像采集、处理和定时操作。他们使用低光（4 勒克斯）单色 CCD 相机配备广角镜头（4 毫米）的图像收集。在上料站上方安装摄像头，以两盏钨丝灯作为补光灯。约 $2.2\text{cd}/\text{m}^2$ 照度可能是已实现期间的白天在的喂食车站。的系统收集的图像的猪回来还有有记录的增长率的的三个品种的猪通过 a 预测模型成立与的尸体体重。的实验性的结果结果表明，平均重量误差在 5% 以内，在 60-90 公斤范围内，部分品种的误差小于 2 %。

Marchant 等人。（1999 年）测试了一个图像分析系统，以估计猪的体重，并使用手称数据校准模型的截距。对 80 公斤猪的估计误差小于 1%。

迪斯科-威尔森，惠特莫尔，克纳普和斯科菲尔德（2004）使用视觉图像分析来描述不同品种的猪的体型和体型，并评估生长过程中体型和体型的变化。此外，在饲养站用平台秤监测猪的生长，并在饲养站上方安装视觉成像系统。当每头猪在饲养站时，测量活体的质量，并通过上面的相机拍摄猪的二维数字图像。然后，通过算法对图像进行处理，尺寸和形状测量结果可以准确地描述不同猪群的生长情况。

怀特、斯科菲尔德、格林、帕森斯和惠特莫尔（2004）利用 VIA 系统和平台平衡来检测三种猪的生长速度。通过比较，结果表明，使用 VIA 系统对猪进行测试比使用平台平衡更快、更准确。然而，如果猪在摄像机前长期处于异常位置，则 VIA 的重量估计会产生一定的偏差。

帕森斯、格林、斯科菲尔德和惠特莫尔（2007）发现，使用 VIA 系统可以估计猪的体重，平均误差小于 3.5%。此外，该系统还可加入综合管理系统）。从 VIA 系统获得的数据可用于实时优化猪的生长模型，控制猪的增重和脂肪厚度。

2.2. 猪的预计面积

小川和 Ichikawa（1994）把猪放在一个黑匣子里，黑匣子里有两盏白炽灯和一个摄像机在上面照明，以便拍照，以计算猪的投影面积(图 2)。利用 33 头标准误差 $\pm 0.9\text{kg}$ 的猪的数据，建立了投影面积与体重之间的指数关系。

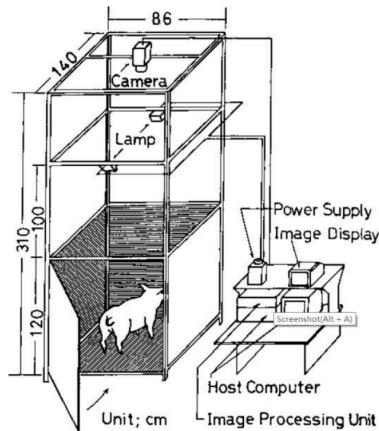


图 2 从上面拿相机拍摄猪的投影图像区域的系统(Minagawa 等人, 1994 年)。

2.3. 尺校准相机

王, 杨, 温特和 T. 沃克(2006)开发了基于机器视觉的非接触式猪体重测量系统(图。3)。共 187 头猪作为实验对象。他们在猪的高度放置了一把尺子来校准相机, 并使用 Matrox 软件提取猪。然后, 建立了估算猪体重的关系模型。结果表明, 投影面积与体重的相关系数最高($R^2=0.96$)。此外, 身体宽度与体重之间存在良好的相关性($R^2=0.95$)。

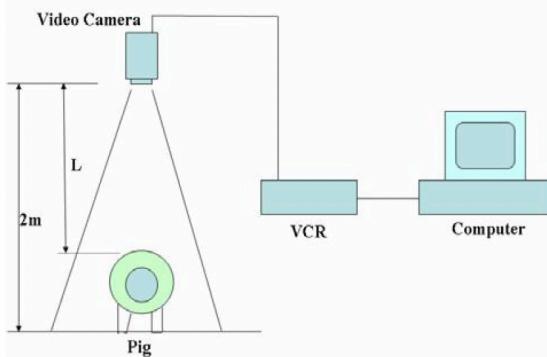


图 3 测量系统原理图来自 Wang 等人的研究 (2006)。

两年后, 他们改进了猪轮廓提取算法来估计猪图像的权重(Wang, Yang, Walker, 等人, 2008 年)。实验采用机器视觉检测系统, 包括摄像机和计算机, 实现从图像采集、图像处理和特征提取到数据分析的过程。首先, 引导猪通过一个大约 1 米宽的通道。第二, 摄像机安装在天花板上, 通道两侧各放置一盏 40W 白色荧光灯, 距地面约 1m, 以提供稳定的照明。为了获得更高的图像对比度, 他们把黑色地毯放在通道的地板上。同年, 该团队改进了机器视觉来估计猪体重系统(Wang, Yang, Winter, 等人, 2008 年)。系统没有要求猪保持静止。当每只猪经过通道时, 顶部的摄像机不断地拍摄。在曝光模式下, 相机最初被设置为“手动”模式, 光圈被设置为最宽的位置。因此, 在给定的光照强度下, 每个照片的曝光时间可以被最小化, 以减少由于猪的运动而产生歧义的机会。

之后，对捕获的图像进行处理，以获得身体大小参数。通过人工神经网络建立体型与体重的关系。实验结果表明，相对误差 $\pm 3\%$ 。

该方法的测量精度较高，但只有在物体距离固定时，才适用于用尺子标定摄像机的方法。由于时间和个体的变化，猪的身高可能会发生变化，这种方法在实际生产中很难应用。

2.4. 猪身上的特殊痕迹

杨、滕、李和施（2006）计划收集猪的背部和侧面图像，以估计它们的体重（图. 4）。他们选择空的猪圈，在周围涂上绿色的油漆。他们以贴在猪背上的纸为标定对象，画出以墙上已知面积为基准的矩形盒。此外，他们在猪圈周围放置了两个高清数字变焦集成相机，实时采集猪的顶部视图和侧面视图。通过俯视和侧视得到物体的高度和面积。研究背部面积，身高与体重的相关性。后，估计 40 头猪的体重，平均相对误差 $\pm 3.2\%$ 。

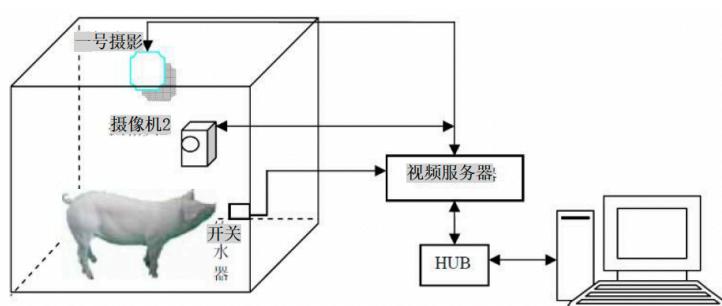


图 4 杨等人研究的测量系统示意图（2006）。

自 2013 年以来，相机加上猪背上的特殊图案被比利时团体用于猪的个体识别和体重估计（图. (Kashiha 等人，2013 年；Kashiha 等人，2014 年)。首先，通过在猪背上喷洒特殊图案，并使用图像处理技术来定位和识别单个猪。第二，他们使用椭圆拟合来涂抹头部和尾部。然后，利用剩余躯干面积建立像素面积与体重之间的估计模型。结果表明，猪体重估计在种群水平上的准确率为 97.5%，误差 $\pm 0.82\text{kg}$ 。此外，还说明了在个体水平上的精度 $\pm 96.2\%$ ，误差较 $\pm 1.23 \text{ 公斤}$ 。

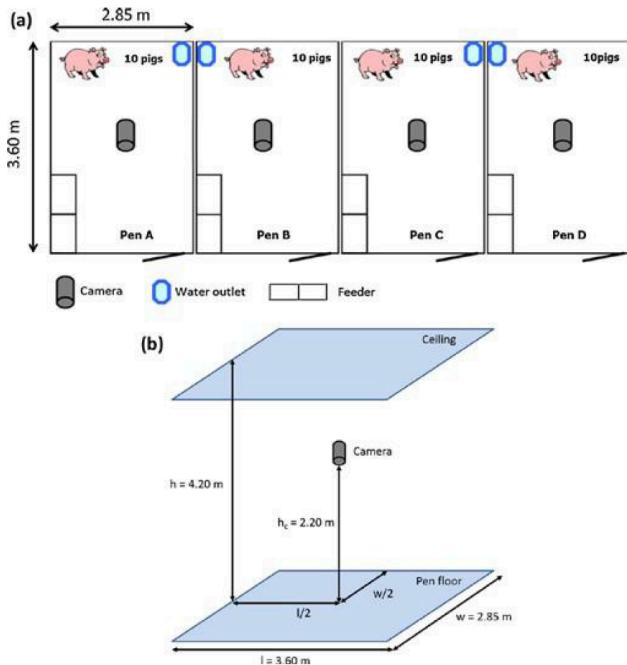


图 5 (a)钢笔的地面计划; (b)照相机在钢笔顶部的位置(Kashiha 等人, 2014 年)。

2.5. 计算图像中的像素数

科利斯、范、班哈齐和塞勒 (2007) 开始研究猪体重预测。构造 900x1350x364mm 的矩形面积，一次只能测量一头猪。在 Logitech 快速 CamMessenger 网络摄像机中，猪图像安装在测量区域 3M 以上的框架上，分辨率为 640x480。整个系统采用荧光灯进行照明，提高了拍摄质量。两个 50W 的筒灯也安装在框架上，在相机的两侧。通过对采集的图像进行分析，发现猪背面积、脊柱长度与猪体重之间存在较强的相关性。

Banhazi、Tscharke、Ferdous、Saunders 和 Lee (2009) 开发了一种基于计算机的猪体重估计系统使用身体尺度像素来估计体重。实验表明，估计猪重的平均误差是 ± 1.18 公斤。

Tscharke 和 Banhazi (2013) 改进了原有的猪体重估计系统，增加了自动搜索图片，识别猪的形状，提取猪的体型大小和估计猪的体重，并与实际体重进行比较的功能。结果表明，在 60-120kg 之间，组平均误差 $\pm 2\text{kg}$ ，平均相对误差在 1% \pm 以内%。由于系统无法进行个体识别，只能估计总体平均权重。根据像素数估计身体大小不具有普遍性，当对象距离发生变化时将不再适用。

2.6. 用户互动

Wongsriworaphon、Arnonkijpanich 和 Pathumnakul (2015) 使用了一种可应用于农场环境的非紧急猪体重估计系统。摄像机设置在通道顶部，以捕获猪的背面图像(图。六);

之后，基于用户交互手动提取猪体轮廓，取猪体质量中心到边界点和猪体周长的平均距离作为图像的特征值。特征值也作为权重估计模型的输入。结果表明，该方法对体重的平均估计误差小于 3%，可以有效地应用于实际的农场环境，但对劳动力的依赖较大，自动化不易。



图 6 安装摄像机以捕获顶部视图图片(Wongsriworaphon 等人, 2015 年)。

在二维测量阶段，大多数检测系统只包含一个摄像机。获得的平面图像只包含身体长度、身体宽度或背部面积的数据，这对于建立体重估计模型来说不是那么通用。当安装距离或摄像机的设置参数发生变化时，必须再次进行校准，因此在现场生产中不实用。

3. 三维测量方法

随着机器视觉技术的发展，降低了三维测量的成本，猪的体重检测进入了三维测量阶段。三维测量得到的数据包括猪的身高信息，使猪体重的估计变得更加准确。

3.1. 立体投影技术

小川和村上 (2001) 开始研究立体投影技术对猪体重的预测(图。7). 幻灯机作为平行光源，网格划分在投影幻灯片中。投影机固定在饮用水装置的顶部，摄像机安装在旁边。根据几何光学原理，根据猪体地面与背部的网格间距计算猪的高度。利用去除耳区的猪体高和背部面积建立体重预测数学模型。对 12 头年龄从 81 天到 98 天的猪进行了 5 次估计，平均相对误差 $\pm 2.1\%$ 。

在 2003 年，他们改进了照明系统的颜色，以增强猪形象与周围环境(Minagawa, Taira, &H)之间的对比。Nissato, 2003 年)。研究表明，在红色和黄色网格幻灯片下获得了更清晰的猪图像。该系统对 7 头长白猪的体重进行了估计，平均误差 $\pm 1.3\%$ 。

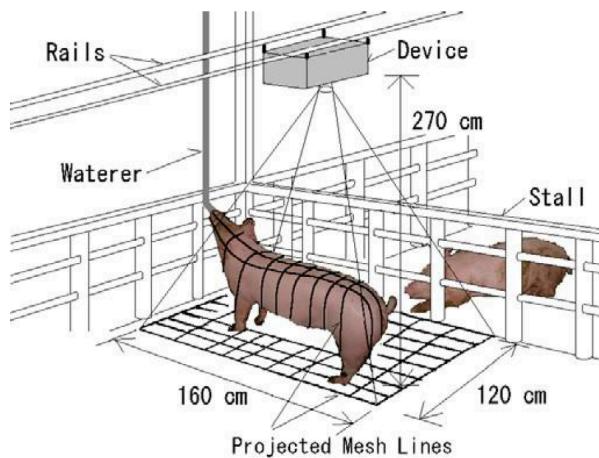


图 7 来自 Minagawa 等人研究的测量系统示意图（2003）。

采用这种方法，可以在不增加参考对象的情况下获得猪的体高，但采用光投影法的方法容易受到外界光线的影响，在实际猪舍中的应用效果较差。

3.2. 双目立体技术

傅和腾（2009）基于双目视觉原理对猪体尺寸测量进行了一些研究，并设计了猪体尺寸测量和体重预测系统(图。8)。一个 CCD 相机(SSC-E473P，索尼，日本)结合定制镜头(H051-MP，CBC，日本)作为一个图像采集装置安装在猪的上方，当相机位于 80.0 厘米的高度时，设置 $1.5 \text{ 米} \times 1.0 \text{ 米}$ 的视角范围。以四盏荧光灯作为照明光源，作为照明猪身体形象的一种方式。猪体四面照射。针对猪轮廓的特点，提出了一种基于凹陷结构和凸壳分析的猪体测点提取算法。结果表明，猪长的相对误差 $\pm 0.85\%$ ，猪体高和体宽的最大和最小检测误差分别 $\pm 2.23\%、\pm 0.14\%$ 和 3.31% 和 ± 0.19 。通过主成分分析，建立了基于猪体大小参数的猪体重预测模型。模型拟合系数为 0.955，残差标准差为 4.49 公斤。现场应用表明，平均重量预测的相对误差 $\pm 0.77\%$ ，证明了该系统的实用价值。然而，该系统的预测模型使用了许多参数，如长度、宽度和高度以及猪体的关键点。

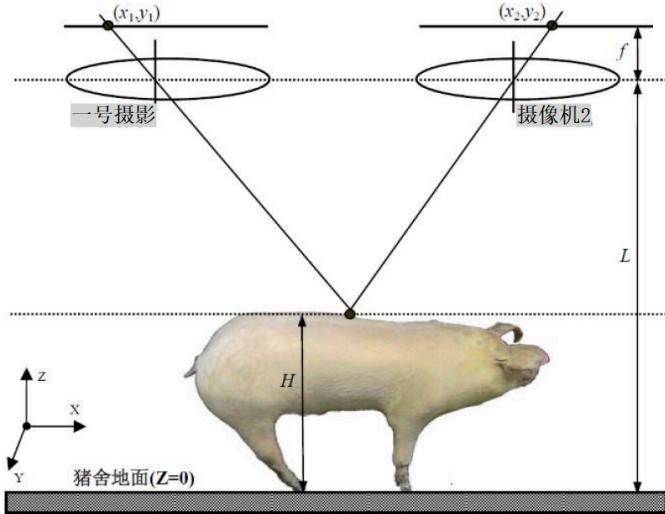


图 8 基于双目立体视觉的猪体尺寸检测原理图(Fu 等人, 2011 年)。

刘, 滕, 付, 李(2013)结合猪舍的实际情况, 提出了复杂背景下猪体个体信息提取算法, 基于包络分析的猪体头部和尾部去除算法, 以及从具有一定弯曲姿态的复杂猪体中提取测点坐标的算法。采用配备定制镜头(H051-MP, CBC, 日本)的 CCD 摄像机(SSC-E473P, 索尼, 日本)作为图像采集装置, 从上面采集 RGB 彩色猪。在获得猪体轮廓的基础上, 通过包络分析识别头尾的切割点, 去除头尾。然后利用体尺提取算法对猪体 9 个测点的坐标进行识别。体型平均相对误差 $\pm 2.09\%$, 体重平均相对误差 $\pm 3.38\%$ 。同时, 他重建了猪标本[27]的三维表面模型, 使用了非接触三维扫描仪(Vivid-910, 中国), 测量距离为 1.5m。成功获取猪体三维点云数据。平均相对误差 $\pm 0.17\%$, 重量测量误差 $\pm 2.7\%$ 。因此, 采用双目立体视觉重建猪体是可行的。

李, 杜, 毛, 腾(2016)和石, 腾, 李(2016)开发了一种基于双目立体视觉技术的非接触式猪体重估计系统。两个黑白相机(acA1300-30gm, Basler, 德国)与定制镜头(ComputarH0514-MP, CBC, 日本)被用来捕捉图像。在投料区顶部安装摄像头, 采用条形荧光灯作为光源进行现场补光。提出了一种基于深度图像的猪体轮廓提取算法。为了解决猪场恶劣光照环境下的灰色轮廓提取算法问题, 采用猪体角归一化系数来筛选测量点和椭圆长轴到轮廓的距离, 构建了一种新的体尺检测算法。研制了猪体大小和体重非接触检测软件。体长, 体宽, 体高, 臀宽和臀高的检测准确率分别为 1.44%, 5.81%, 4.94%, 2% 和 1.64。实验 4 头猪 268 组估计体重的平均相对误差 $\pm 2.14\%$ 。该系统实现了猪体重和体重的全面自动估计。

3.3. 基于深度图像的猪体重研究

深度图像又称距离图像或距离图, 记录场景中各点与摄像机之间的距离信息, 反

映物体表面的三维特征。基于深度图像的物体轮廓提取可以有效地区分图像前景和背景 [29]。Kinect 是微软公司开发的三维体感相机。具有成本低，能快速获取物体的深度信息。与普通相机相比，具有标定简单，易去噪的特点。近年来，利用 Kinect 相机对猪体大小进行了越来越多的研究。

孔斯罗 (2014) 使用 Kinect 相机设计了一个估计猪体重的系统，其中 Kinect 安装在伸缩铝杆上(图。9)。深度图像处理后，提取猪的身体大小用于计算猪的体积，并进一步估计猪的重量。实验结果表明，该系统成功地估计了两种不同品种猪的体重，检测误差为 4%-5%。



图 9 猪体重测量的 Kinect 原型(Kongsro 等人, 2014 年)。

Lao 等人。(2016 年) 使用 Kinect 相机设计了一个系统来识别母猪的行为。在产仔箱上方 2.20m 安装了 3DKinect 摄像机，摄像机每隔 6 秒拍摄图像。在对图像进行初步处理后，提取母猪的重要特征，包括其质量中心坐标、头部坐标、臀部坐标、肩部坐标和腰部坐标。将深度图像中的母猪分为 7 个部分（全身、上半身、下半身、头、肩、腰和臀部），然后计算每个母猪的深度像素的平均值。通过设置一些阈值，将各部分像素的深度值与阈值进行比较，以确定母猪的行为。实验结果表明，母猪躺卧，坐立聋进食饮酒行为的识别率分别为 99.9%，96.4%，99.2%，78.1%，97.4%，92.7%，不同行为转换的识别率分别为 63.9%。

刘，滕，张，李，郭 (2014) 使用 Kinect 相机作为图像采集装置，以 30 帧/秒的速度采集 HOG 运动深度图像序列，并对每个帧深度图像进行骨架提取、剪枝等处理。在建立了猪的运动模型后，提出了一种计算猪行走完全步长的方法。最后，通过计算序列完整步骤的长度与序列采集时间的比值，提取猪的步长频率特性。实验结果表明该方法的准确度为 82.1%。

该 Kinect 相机解决了快速获取猪的深度信息的问题，但存在获取的深度图像分辨率低，检测精度小于 1m 的缺陷。因此，它经常被用于行为分析或特质评价，在获得身体测量和体重信息方面的误差较大。

4. 总结

基于机器视觉的体重测量技术可分为二维测量阶段和三维测量阶段。对于二维测量方法，大多数检测系统都配备了一个摄像头，安装在猪的顶部或侧面。在三维测量阶段，机器视觉设备获得了猪的平面参数以及猪的高度信息，使得猪体侧的预测更加准确，与二次测量方法相比，模型的精度也更高。基于机器视觉的猪体体重测量技术主要存在以下问题：

1. 大多数系统无法有效地识别单个猪，只能检测猪群的平均体重；
2. 许多测量系统需要特殊的设施，甚至需要对猪舍进行改造，以满足基于机器视觉的重量测量设备的要求，这使得该系统在现场应用中没有实际应用。
3. 有些系统缺乏自动化，需要人工协助；
4. 系统的大部分灰度图像集合都是从捕获的图片中提取出来的，而体尺度提取算法容易受到外部环境和猪的活动的影响。
5. 两种体型系统大多利用背部面积、体长、体宽等平面参数与体重建立关系，忽略了猪体的高度，使猪体形状的估计不准确；
6. 一些系统提取了大量的体尺信息，创建了复杂的估计算法，但估计精度较低；
7. 有些系统成本高，实际生产中实用性差；
8. 猪群的体重估计精度高于单个猪。

因此，开发一种低成本、高精度、自动化、实用的非接触式体重测量系统是未来的重要研究方向之一。