基于 DRGB 的运动中肉牛形体部位识别

邓寒冰[1,2]，许童羽[1,2※]，周云成[1,2]，苗 腾[1,2,3]，张聿博[1,2]，
徐 静[1,2]，金 莉[1,2]，陈春玲[1,2]

（1. 沈阳农业大学信息与电气工程学院，沈阳 110866；2. 辽宁省农业信息化工程技术研究中心，沈阳 110866；
3. 北京农业信息技术研究中心，北京 100097）

摘 要：如何解决运动中肉牛关键部位自动识别，是实现肉牛异常行为早期发现的关键。该文通过 Kinect 采集肉牛图像的 2 种模态（Depth 和 RGB）：基于 RGB 模态提出随机最近邻像素比较法，实现肉牛动作样本的自动抓取；基于 Depth 模态提出深度均值法，实现彩色图像背景过滤并保留肉牛形体信息，生成 DRGB 图像样本；基于 Fast R-CNN 设计识别器，参考 AlexNet 设计了 8 种分类网络并比较网络分类精度，选择最优网络作为识别器的基础网络；输入 DRGB 样本对网络的识别部分二次训练，最终得到符合精度要求的识别器。试验证明，RNNPC 的有效数据率为 94%；SelectiveSearch 算法在 DRGB 上产生的候选区域数量减少 90%；识别网络的平均分类精度可以达到 75.88%，处理图像速率为 4.32 帧/s，效果优于原 Fast RCNN，基本可以实现运动中肉牛形体部位识别。

关键词：图像重构；图像识别；算法；肉牛；深度卷积神经网络；目标识别；DRGB
doi：10.11975/j.issn.1002-6819.2018.05.022
中图分类号：S823.9+2；TP391.41 **文献标志码**：A **文章编号**：1002-6819(2018)-05-0166-10

邓寒冰，许童羽，周云成，苗 腾，张聿博，徐 静，金 莉，陈春玲. 基于 DRGB 的运动中肉牛形体部位识别[J]. 农业工程学报，2018，34(5)：166－175. doi：10.11975/j.issn.1002-6819.2018.05.022 http://www.tcsae.org
Deng Hanbing, Xu Tongyu, Zhou Yuncheng, Miao Teng, Zhang Yubo, Xu Jing, Jin Li, Chen Chunling. Body shape parts recognition of moving cattle based on DRGB[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(5): 166－175. (in Chinese with English abstract) doi：10.11975/j.issn.1002-6819.2018.05.022 http://www.tcsae.org

0 引 言

现代肉牛养殖业是中国大力扶植和发展的产业，从目前的牛肉需求来看，中国牛肉需求有望从 2008 年的 608 万 t 上涨到 2020 年的 828 万 t[1]，而与此对应的是国内牛肉供应增长乏力，这就要求养殖户要通过更科学的手段进行肉牛养殖以提高牛肉产量。

在集约饲养的条件下，肉牛异常行为的出现经常是随机的、短暂的，因此如果不能长时间连续观察，很难引起饲养人员的重视，这往往会延长对肉牛疾病的发现时间，给饲养人员造成巨大的经济损失[2]。现代研究发现，肉牛异常行为是由于多种因素综合引起的，包括环境因素、饲料营养、激素、心理和遗传等[3]。所以，引起牛的行为异常原因很复杂，不同性别、不同生长阶段表现也有所不同，因此需要对肉牛进行长时间连续细致观察才能及时发现和预防。

随着大规模图像数据的产生及计算硬件（GPU 等）的飞速发展，基于卷积神经网络的相关方法在各应用领域取得了突破性的成果[4-7]。在深度卷积神经网络（deep convolutional neural network, DCNN）方面，将自动化图像特征提取与分类过程融合，并实现自主学习。国内外研究人员在 DCNN 的基础理论[8]、网络结构设计[9-14]、图像流处理[15]上开展了很多研究。特别是在目标识别等领域已经得到越来越多的认可，例如微软公司设计的 ResNet（大于 1 000 层）在图像分类、目标检测和语义分割等各个方面都取得了很好的成绩[16]。自 2014 年 Ross Girshick 等提出利用 RCNN[17]（regions with CNN feature）方法实现目标识别以后，深度卷积神经网络的已经成为实时目标识别的主要方法，其性能和精度都遥遥领先于当时最优的 DPM（deformable parts model）方法。此后，在实时检测方面，分别出现了基于区域推荐和基于预测边界框的 2 类核心方法：其中区域推荐方法普遍采用滑动窗口来实现，对像素尺寸较小的目标比较敏感，但对图像整体内容没有进行关联分析，如 Fast R-CNN[18]、Faster R-CNN[19]、HyperNet[20]等；而预测边界框方法通常使用预设区域，识别速度快，但会影响图像背景中的小尺寸物体识别精度，如 YOLO[21]、SSD[22]等。

随着各类方法的不断更新和优化，深度神经网络在各研究领域发挥的作用也越来越明显。其中，在农业科研领域深度卷积神经网络已经从理论研究向实际应用转移。在温室环境下已经出现了基于 CNN 的植物花、叶、果实等自动识别原型系统[23-26]；在病虫害识别方面，已经

收稿日期：2017-09-08 修订日期：2018-02-01
基金项目：国家自然科学基金资助项目（31601218，31601217，1601219，31601281）
作者简介：邓寒冰，辽宁沈阳人，讲师，博士，主要从事机器学习与模式识别研究工作。Email：deng_hanbing@126.com
※通信作者：许童羽，辽宁沈阳人，教授，博士，博士生导师，主要从事农业航空和农业遥感研究工作。Email：yatongmu@163.com

出现对害虫分类，病害分类分级的方法[27-31]。目前，针对家禽、水产等大型动物的实时图像处理分析逐渐成为研究热点，文献[32]提出用视频分析方法提取奶牛躯干图像，用卷积神经网络准确识别奶牛个体方法；文献[33]从水产动物视觉检测的图像采集、轮廓提取、特征标定与计算等方面提出了改进措施，对基于计算机视觉测量的动物疾病诊断和分类进行探讨和总结；文献[34]采用改进分水岭分割算法实现运动对群养猪运动轨迹进追踪。随着多类型信息化设备在现代养殖业的使用，数据的多模态特性逐渐成为研究的关注点，利用多模态数据间的内容关联实现算法性能提升和过程优化，已经成为深度学习的一条重要研究方向[35]。特别是在如何利用多模态数据来提高目标识别的精度与速度方面，仍有很多亟待解决的问题。

为此，本文以肉牛为研究对象，拟通过深度卷积神经网络来实现面向多模态数据（深度与 RGB）的肉牛形体部位快速识别。在分类网络的基础上，利用多模态数据对网络部分层中的参数进行精调（fine-tuning），同时利用多模态数据间的映射原理（可用于去除图像背景），降低候选区域的个数，进而加快网络对形体部位的识别速度，以期实现对运动时肉牛的形体部位的定位与识别。

1　样本采集与预处理

由于本文中需要识别的类型较少（头、躯干、腿、尾），因此为了避免过拟合问题，提高样本的多样性，本试验分别于 2016 年 5 月–2017 年 3 月期间在辽宁省法库县牛场进行数据采集。其中训练集和验证集是通过 4 种不同像素的数码相机进行采集约 10 000 幅肉牛完整图像，然后通过人工处理形成约 40 000 幅包括肉牛头部、躯干、尾部、腿部及背景 5 种类型的彩色图像用于网络的训练（80%）和验证（20%）；而对于测试集，本试验利用可采集景深数据的视频设备，采集约 10 组完整视频文件（连续图像序列）。

1.1　测试样本采集的设备选择与场景布置

1）设备选取：本文以微软公司的 Kinect 作为测试集图像采集的设备，该设备能够相同时间维度上采集拍摄范围内的彩色数据（RGB）和深度数据（Depth，即拍摄对象与摄像头的距离值）。其中 RGB 数据是通过高清摄像头获取的，而深度数据是通过红外线收发装置测距来获取的。因此通过 Kinect 可以在同一时间维度上获取 2 种模态的图像数据。

2）场景布置：为了提高采样过程中图像样本的质量，避免由于肉牛之间的相互重叠而造成的局部特征信息丢失，本试验在测试集采样过程中，每次取样限定对 1 头牛进行拍摄。根据官方给出的 Kinect 参数[36]，摄像头的水平拍摄视角为 57°，垂直拍摄视角为 43°，垂直方向的倾斜范围±27°，有效拍摄范围约为 0.5～4.5 m。由于肉牛的平均高度大约为 1.5～1.7 m，为了减少样本图像中的物体形变，将摄像头的垂直高度设置为 1.6 m。

1.2　测试集无效样本过滤方法

利用 Kinect（20～30 帧/s）采集测试样本，平均每小时将会产生 72 000~108 000 幅图像，其中大部分属于“低价值”数据（即未出现肉牛以及肉牛长时间静止）。为了在测试集中减少这类数据，同时保证肉牛动作序列的连续性和完整性，本文提出一种随机最近邻像素比较法（random nearest neighbor pixel comparison, RNNPC），按照时间顺序，在原始样本序列中按序取出相邻 2 幅 RGB 图像，分别在 2 幅图像中抽取具有相同坐标和面积的图像区域，并计算该区域 RGB 三通道的像素差值和，通过比较每组像素差值和与预先设定阈值间的大小关系，来预测图像中的该区域关联的物体是否出现位移，进而筛选保留较为完整连续的动作序列。

为了实现 RNNPC 方法，本文将测试集中原图像序列样本分为 3 种类型（如图 1 所示）：1）静态序列（static sequence, SS）：在连续图像序列中，肉牛处于静止状态或肉牛移出拍摄范围；2）细微动作序列（micro-action sequence, MAS）：在连续图像序列中，肉牛有细微的动作变化，但没有明显的水平或垂直移动，例如出现咀嚼、摇晃尾巴、转头等；3）明显动作序列（obvious-action sequence, OAS）：在连续图像序列中，肉牛有明显的水平或垂直移动，例如行走、卧躺、进食等。

a. 测试样本集中的静态序列
a. Static sequence in test samples set

b. 测试样本集中的细微动作序列
b. Micro-action sequence in test samples set

c. 测试样本集中的明显动作序列
c. Obvious-action sequence in test samples set

注：Kinect 设备可以采集同一时间维度的 RGB 图像和深度图像，但本图中只给出 RGB 图像序列，用于描述不同类型的样本特征。
Note：Kinect device can collect RGB images and deep images at the same time, but in this picture, we only show the RGB image sequences which are used to describe different types of sample features.

图 1　三类测试样本
Fig.1　Test samples of three types

考虑摄像头在采集样本过程中是静止的，因此光照变化和肉牛动作是导致图像像素变化的主要原因。根据这一特点，RNNPC 方法的具体实现如下：

设 t 为样本采集的时间点，$\boldsymbol{M}_t^p$ 表示在 t 时刻获取图像所对应的 p 通道像素矩阵，其中 $p\in\{R,G,B\}$；$\boldsymbol{M}_t$ 与 $\boldsymbol{M}_{t+\Delta t}$ 为 2 个连续时间点获取的图像像素矩阵，Δt 表示相邻帧之间的时间间隔；$\boldsymbol{M}_t^p(x,y)$ 表示 t 时刻所获取的图像的(x, y)位置对应的 p 通道的像素值集合；引入像素距离 $d_{x,y}(M_{t1}, M_{t2})$的概念

$$d_{x,y}\left(\boldsymbol{M}_{t1},\boldsymbol{M}_{t2}\right)=\sum_{p}^{\{R,G,B\}}\left|\boldsymbol{M}_{t1}^{p}(x,y)-\boldsymbol{M}_{t2}^{p}(x,y)\right| \quad (1)$$

由于 $\boldsymbol{M}_{t1}$ 与 $\boldsymbol{M}_{t2}$ 是在不同时间点获得的图像像素矩阵，理论上 $\boldsymbol{M}_{t1}\neq\boldsymbol{M}_{t2}$，因此本文为像素距离 $d_{x,y}(\boldsymbol{M}_{t1},\boldsymbol{M}_{t2})$ 设计了阶跃函数 H_θ

$$H_\theta(d)\begin{cases}1 & d>\theta\\ 0 & d\leqslant\theta\end{cases} \tag{2}$$

式中 θ 表示像素距离阈值，利用函数 H_θ 可以统计相邻像素矩阵间 d 值超过阈值 θ 的像素点总数 N

$$N=\sum_{x=0}^{M^H-1}\sum_{y=0}^{M^W-1}H_\theta\left(d_{x,y}\right) \tag{3}$$

式中 M^H 表示像素矩阵的行数（对应图像高度），M^W 表示像素矩阵的列数（对应图像宽度）；为了使随机位置获取的图像区域能够尽量捕捉到目标移动，这里设随机参数 rand∈(0.5,1)，即该方法可以从相邻图像中选取至少 rand×M^H×M^W 个起始位置随机但空间连续的像素点进行差值计算。此外，本文将像素矩阵中的每个位置都赋予一个随机数 ϑ，且 $\vartheta\in[0,1]$，对于不同位置的 ϑ 不相等，即 $\vartheta(x_1,y_1)\neq\vartheta(x_2,y_2)$。基于 rand 值设置命中函数 T_r

$$T_r(x,y)=\begin{cases}1 & \vartheta(x,y)\leqslant\text{rand}\\ 0 & \vartheta(x,y)>\text{rand}\end{cases} \tag{4}$$

利用式（1）~（4）就可以计算相邻图像之间的相似度

$$s(M_{t1},M_{t2})=\frac{\sum_{x=0}^{M^H-1}\sum_{y=0}^{M^W-1}H_\theta\left(d_{x,y}\right)T_r(x,y)}{M^H\cdot M^W\cdot\text{rand}^2} \tag{5}$$

可以看出 $s(M_{t1},M_{t2})\in(0,1)$，当 $s(M_{t1},M_{t2})$ 趋近于 1，表示相邻图像相似度高，反之表示相似度低。

本文从 Kinect 获取的 RGB 图像样本中选取 3 组序列（分别为静态序列、细微动作序列、明显动作序列）。在给定 $\Delta t=50$ ms 的条件下，通过设置 θ 值来获取每组图像序列的相似度曲线。分别将图 1 中 3 组图像序列作为 RNNPC 方法的输入，通过计算得到的相似度曲线如图 2 所示。可见对于不同的样本类型，相似度曲线呈现出不同的特点。从 3 组序列的曲线分布来看，随着 θ 值的增加，SS 的相似度从 30%左右（图 2a）提高到 97%左右（图 2c），随着 θ 值的增加，由光照造成的像素差异明显减少；在 θ=0 时，3 类曲线的差异不明显（图 2a），而随着 θ 值增加，曲线分布差异逐渐增大，然而当 $\theta\geqslant10$ 时，这种差异又出现减小的趋势（对比图 2b 与图 2c）。可以证明随着 θ 的增大，可以将 3 种不同类型曲线分布差异扩大，但当 θ 超过一定限度时，这差异又出现减弱的趋势，这表明当 θ 增加到一定程度，由目标移动所产生的像素变化将不再明显。因此，考虑减少光照影响，同时扩大相似度曲线分布差异，本文选择 θ=5 作为像素距离阈值。

图 2d 是由 RNNPC 方法获取的一段完整的图像序列样本的相似度曲线。设 $s_{\max}$ 为曲线最大值，$\bar{s}$ 为曲线值的均值，$\boldsymbol{S}_{\max}$ 为曲线局部极大值集合，$\bar{s}_{\max}$ 为局部极大值均值，$\boldsymbol{S}_{\min}$ 为曲线局部极小值集合，$\bar{s}_{\min}$ 为局部极小值均值

$$\bar{s}_{\max}=\sum_{s_j\in\boldsymbol{S}_{\max}}s_j\Big/\left|\boldsymbol{S}_{\max}\right|,\quad \bar{s}_{\min}=\sum_{s_i\in\boldsymbol{S}_{\min}}s_i\Big/\left|\boldsymbol{S}_{\min}\right| \tag{6}$$

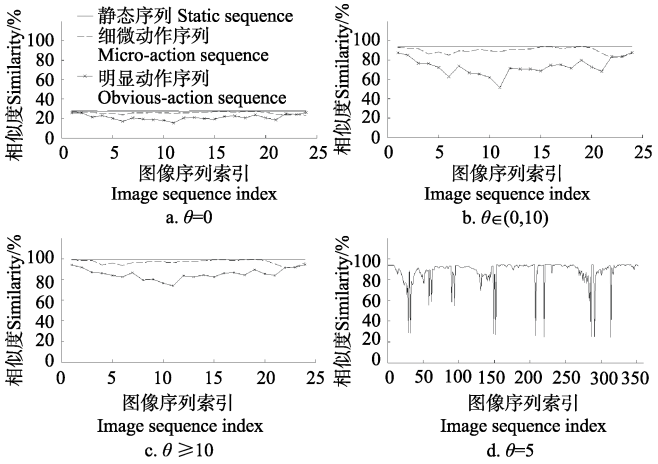


图 2 不同像素距离阈值 θ 下的图像序列相似度曲线

Fig.2 Similarity curve of image sequence of different pixel distance thresholds θ

经过统计，曲线在 y 轴投影落在($s_{\max}$, max{$\bar{s}_{\max}$, $\bar{s}$})区间内表示静态序列；当曲线在 y 轴投影落在(max {$\bar{s}_{\max}$, $\bar{s}$}, min{$\bar{s}_{\max}$, $\bar{s}$})区间内表示细微动作序列；当曲线在 y 轴投影落在(min{$\bar{s}_{\max}$, $\bar{s}$}, $\bar{s}_{\min}$)内表示明显动作序列。

为了检验 RNNPC 方法对于完整视频数据处理的有效性，试验选用 10 段视频进行处理（每段视频 30 min 左右）。根据视频信息的帧率，可以计算出每段视频将产生约 3.6 万帧图像。将自动保留下来的图像序列与人工筛选保留的序列进行比较，结果如表 1 所示。

表 1 随机最近邻像素比较法产生明显动作序列的结果

Table 1 Results of obvious-action sequence by random nearest neighbor pixel comparison(RNPPC)

组号 Group index	总帧数 Total frames	保留的图像序列 Preserved image sequence			节省的存储空间 Saved storage/%
		总数 Totals	错误帧 Errors	正确率 Accuracy/%	
1	36 012	10 260	636	93.8	71.6
2	36 010	9 794	564	94.2	72.9
3	36 002	10 152	688	93.4	71.9
4	36 018	10 090	724	92.8	72.0
5	36 006	10 010	552	94.4	72.2
6	36 012	10 312	644	93.7	71.4
7	36 016	9 986	432	95.6	72.3
8	36 020	10 892	748	93.1	69.8
9	36 008	10 024	722	92.7	72.2
10	36 020	9 864	466	95.2	72.7
平均				93.89	71.9

从试验结果可以看到，利用 RNNPC 方法采集连续图像样本可以节省 72%左右的存储空间，而剩余 38%样本的有效率在 94%左右，样本质量和数量可以满足样本要求。

2 深度信息与 RGB 信息融合

由于本文采用区域推荐原理来生成目标候选框，因此如何利用深度图像来减少连续 RGB 图像序列在测试过程中的产生的候选框数量是本节主要解决的问题。

2.1　深度信息可视化

为了将深度信息进行可视化处理，本文用灰度值来表示深度信息

$$g(x,y)=\frac{i(x,y)-d_{\min}}{d_{\max}-d_{\min}}\times 255 \tag{7}$$

式中 $i(x,y)$表示位于深度值矩阵 $\boldsymbol{I_d}$ 中(x,y)位置的深度值；$g(x,\ y)$表示与 $i(x,\ y)$对应的灰度值；$d_{\max}$ 表示最远拍摄距离；$d_{\min}$ 表示最近拍摄距离。深度值小于 $d_{\min}$ 的像素点灰度值设为 0，而深度值大于 $d_{\max}$ 的像素点灰度值设为 255。图 3 是利用 Kinect 在同一时刻采集的肉牛 RGB 图像以及利式(7)计算得到的深度图像。

a. 肉牛 RGB 图像
a. RGB image of cattle

b. 肉牛深度图像
b. Depth image of cattle

图 3　相同时间维度的 RGB 图像和深度图像
Fig.3　RGB and depth images with same temporal dimension

2.2　RGB 图像与深度图像的映射

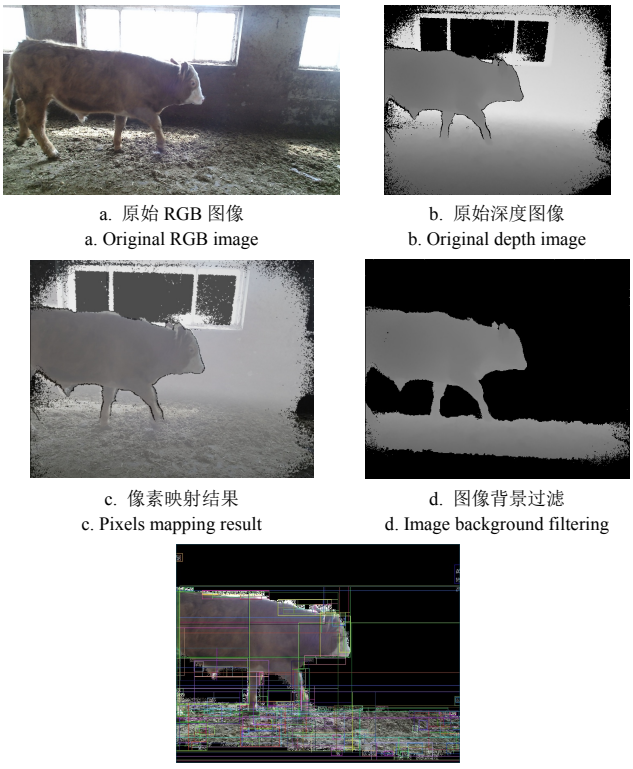
由于 Kinect 的彩色相机和红外相机存在平移距离差，因此在同一时刻采集的原始 RGB 图像与深度图像在内容上无法实现关联。如果能够在目标识别之前尽量去除原图像中的背景信息，就能缩短区域推荐算法的运行时间。所以，需要实现深度图像与 RGB 图像间主要区域的像素点映射。

本文首先利用微软公司提供的开源方法对空间上存在关联的像素点进行标注，然后将深度像素点投影到 RGB 图像上，由于深度图像的大小与分辨率都小于 RGB 图像，因此在处理像素点关联的过程中会损失 RGB 图像部分边缘信息。图 4c 给出了映射效果（只保留映射部分），其中深度图像中的肉牛与 RGB 图像中的肉牛的外沿轮廓几乎完全重合。实现像素点映射就可以建立 RGB 图像与深度图像在内容上的关联，这为下一步去除图像背景信息提供了有效的支持。

2.3　基于深度信息的 RGB 图像背景过滤

利用目标检测算法（Selective Search[37]）来处理原始 RGB 图像，会生成大量的候选区域（2×10^3 以上），其中 90%以上都是无效或重叠候选区域。为了减少无效的候选区域数目，本文利用深度信息将原始 RGB 图像中的背景虑除，并且保证肉牛形体图像的完整。

对于深度图像序列，过滤背景需要在图像序列中找到肉牛移动过程中的灰度区间，同时将区间外的像素信息都过滤掉。然而由于肉牛是移动的，因此其灰度区间也是动态变化的。本文首先要获得被拍摄对象运动时的动态平均灰度值。在 1.2 节中，利用 RNNPC 方法可以用于计算相邻图像的相似度，而相似度是通过像素差值来得到的，因此可以利用 RNNPC 方法间接获得最邻近图像间的像素变化区域，这里设置为 $\boldsymbol{R_C}$，对区域内全部像素点做均值计算，可以得到均值灰度 ρ



a. 原始 RGB 图像
a. Original RGB image

b. 原始深度图像
b. Original depth image

c. 像素映射结果
c. Pixels mapping result

d. 图像背景过滤
d. Image background filtering

e. Selective search 算法产生的候选框
e. Region proposals of selective search method

注：为了展现肉牛 RGB 图像与深度图像的映射效果，图 4c 提高了 RGB 像素的透明度，本示例保留原始灰度图像，并将超过深度范围的（小于 0.5 m 或大于 4.5 m）灰度值设为 0。

Note：In order to exhibit the pixels mapping effect of cattle RGB image and depth image, we improve the transparency of RGB pixels of fig.4c. In this example, we preserve the original gray image and set gray value to 0 when the depth range is less than 0.5 m or greater than 4.5 m.

图 4　深度图像和彩色图像映射结果及在结果对应的候选框
Fig.4　Results of depth and color images mapping and corresponding bounding boxes

$$\rho=\sum_{i=1}^{i\leqslant|\boldsymbol{R_C}|}\frac{g_i}{|\boldsymbol{R_C}|} \tag{8}$$

式中$|\boldsymbol{R_C}|$为 $\boldsymbol{R_C}$ 集合中像素点个数，g_i 为 $\boldsymbol{R_C}$ 集合中第 i 个像素点的灰度值。基于 ρ 值可以设定一个区间系数 δ。对于深度图像 $\boldsymbol{M}$，$g(x,y)$为图像中(x,y)处像素点的灰度值，利用式（9）对全部像素进行处理，则$[\rho-\delta,\rho+\delta]$区间内的像素将被保留下来。

$$g(x,y)=\begin{cases}g(x,y) & g(x,y)\in[\rho-\delta,\rho+\delta]\\ 0 & g(x,y)\notin[\rho-\delta,\rho+\delta]\end{cases} \tag{9}$$

然而经过式（9）处理后，仍会残留很多无效像素点，为了去掉更多的无效信息，本文利用改进后的正态分布函数，对式（9）的结果图像进行二次灰度处理。将 ρ 值作为正态分布函数的期望，通过调整方差 σ 和自定义系数 φ 来改变函数形态

$$g(x,y)=\frac{g(x,y)\varphi}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\left(g(x,y)-\rho\right)^2}{2\sigma^2}\right) \tag{10}$$

其中期望值 $\mu=\rho$，方差 σ 和自定义系数 φ 为人工设定参数。本文这里将对灰度进行两种类型的处理：对于 $\boldsymbol{R_C}$ 集合中的像素点尽量保留原始灰度信息，令式(10)中的 σ=4，φ=15，这样可以保证灰度值在$[\rho-\delta,\rho+\delta]$内的像素点不被

降低像素值；对于不在 $\boldsymbol{R}_C$ 集合内的像素点，要将这些区域的灰度调低至 0 值附近，因此令式(10)中的 σ=1，φ=0.5，这样可以令灰度值在$[\rho-\delta, \rho+\delta]$区间之外的像素点的像素值趋近于 0。从图 4d 中可以看到，式(10)可以将深度图像中的背景信息过滤掉，同时最大程度保留了肉牛整体形体信息。

基于上述方法，可以将过滤后的深度图像中的黑色像素位置标识出来，并将 RGB 图像中相同坐标位置的像素值设为 0，本文将这种过滤背景信息的图像称为 DRGB 图像。图 4e 是利用 Selective Search 算法处理 DRGB 图像而产生的结果。经过统计，候选区域的数量约为 200 个左右，与原始图像的处理结果相比，候选区域数量降低了一个数量级，这会使网络测试过程中减少候选框的生成数量，从输入源头减少了区域推荐和候选边框回归等过程的运行时间。

3 基于 AlexNet 的分类网络训练

3.1 训练样本和验证样本处理

训练集和验证集主要用于训练分类网络模型，是实现目标识别的前提。为了提高样本多样性，在采集图像过程中分别在牛棚内、牛棚外进行拍摄，同时针对肉牛形体大小、形状特点、毛皮颜色以及不同姿态等分别进行拍摄。最后将整体图像进行人工裁剪和标注，形成测试集和验证集，过程如图 5 所示。

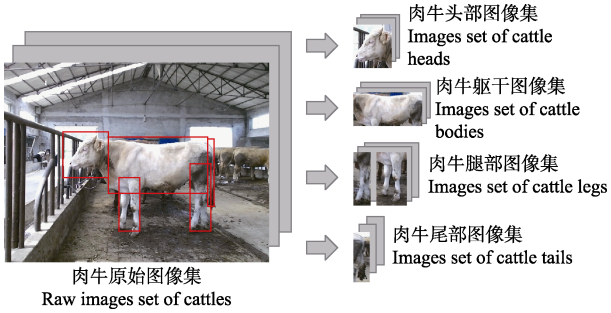


图 5 训练集和验证集样本生成过程

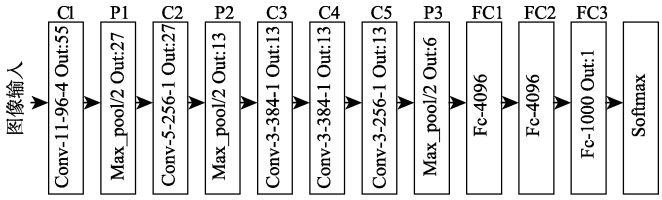
Fig.5 Generation process of training and validation samples set

3.2 AlexNet 网络架构

AlexNet[38]是 Image LSVRC-2102 大赛中的冠军模型，是一种典型的卷积神经网络，如图 6 所示。其中的卷积层主要作用是提取特征，包含一组可以自动更新的卷积核，针对不同的特征提取密集度，卷积核用固定大小的卷积步长（Stride）与来自上一层的图像或特征图作卷积运算，经由激活函数（ReLU）变换后构成卷积特征图，代表对输入图像特征的响应。

AlexNet 设计的结构及训练策略是基于 ImageNet[39]数据集，主要适用于广义的物体识别。若将 AlexNet 直接用于肉牛关键部位的定位和识别，会因数据规模小、数据类别间的纹理差异小而出现损失函数收敛效果差和过拟合等风险[40]。同时，随着网络宽度和深度的增加，其学习能力也会相应的提高，但是训练成本也会呈指数增长。特别是对于固定分类问题，当网络层数过多后，会出现性能下降的问题，因此需要针对具体问题调整网络结构和样本。

首先，肉牛的关键部位的表象通常大小、形状各异，比如躯干的成像面积远大于头、腿和尾部，腿和尾部的成像宽度比头和躯干要窄。为此，本文采用均值像素填充的方式来将不同大小的图像转换为 227×227 大小的 RGB 图像作为网络输入（图 7），避免由于拉抻造成的图像形变。



注：Conv-11-96-4 表示该层为卷积层，卷积核尺寸为 11×11，卷积核数量（输出通道数）为 96，步长为 4，Out:55 表示输出的特征图尺寸为 55×55 dpi；Max_pool/2 该层为池化层，用最大池化操作，步长为 2 像素；FC-4096 表示该层为全连接层，连接数为 4096。

Note: Conv11-96-4 indicates the layer is convolutinal layer with 11×11 kernal size, 96 channels and stride is 4. Out:55 indicates the output feature map scale is 55×55 dpi; Max_pool/2 indicates the layer is pooling layer with max pooling, and the stride is 2 pixels; FC-4096 indicates the layer is full connection layer with 4096 channels.

图 6 AlexNet 网络架构

Fig.6 AlexNet framework

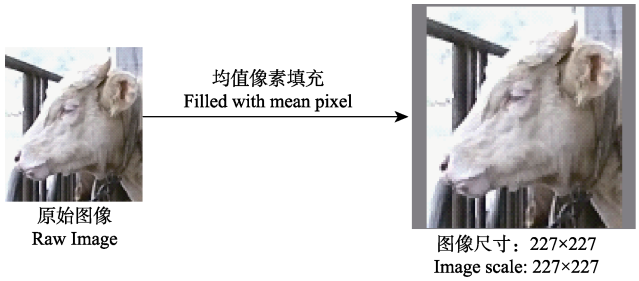


图 7 利用均值像素填充原始图片

Fig.7 Fill original image with mean pixels

针对头、躯干、腿、尾和背景的 5 分类问题，将 AlexNet 的 FC3 层的神经元数量调整为 5 个。未改进的 AlexNet 的参数个数达到 6 000 万个，是为了解决大规模图像分类而设计的，而本试验在类型数量和样本数量上都相对很少。为了提高网络训练效果，在保持 AlexNet 基本结构不改变的前提下，本文配置了 8 种类型分类网络（表 2），每种网络需要训练的参数总数量随着网络层数的递减而递减。其中在全连接层参数不变的前提下，减少卷积层参数对参数总量影响较小（表 2 中网络 I、II、III 比较）；而全连接层对参数总量的影响较大（表 2 中网络 IV 和 V）。

3.3 网络训练方法

本文使用的深度学习框架主要基于 Tensorflow 平台实现（convolutional architecture for fast feature embedding）[41]，计算平台采用单块型号为 NVDIA Tesla K40 的图形处理器（支持 PCI-E 3.0，核心频率为 745 MHz，显存 12 GB，显存频率 6 GHz，带宽 288 GB/s）[42]。由于支持 PCI-E 3.0，这使得 K40 与 CPU 之间的带宽从 8 GB/s 提高到 15.75 GB/s。

采用小批量随机梯度下降法对网络进行训练，在首次训练时只将 batch 数目设置为 32，在每轮训练结束后再将 batch 值提高到原来的 2 倍进行下一次训练，一直增加到 256。采用均值为 0、标准偏差为 0.01 的高斯分布为网络所有层的权重进行随机初始化，偏置（bias）均初始化

为 0，学习速率（lr）设置为 0.01，在训练过程中学习率　　的变化率为 0.1。

表 2　基于 AlexNet 的 8 种分类网络配置

Table 2　Eight kinds of network configuration based on AlexNet

网络编号 Network number	网络配置 Network configurations												层数 Total layers	参数总量 Total parameters/10^6
	C1	P1	C2	P2	C3	C4	C5	P3	FC1	FC2	FC3	输出 Output		
I	11-96-4	max/2	5-256-1	max/2	3-384-1	3-384-1	3-256-1	max/2	4 096	4 096	5	softmax	8	58.29
II	11-96-4	max/2	5-256-1	max/2	3-384-1	3-384-1	–	max/2	4 096	4 096	5	softmax	6	57.40
III	11-96-4	max/2	5-256-1	max/2	3-384-1	–	–	max/2	4 096	4 096	5	softmax	5	56.08
IV	11-48-4	max/2	5-128-1	max/2	3-192-1	3-192-1	3-128-1	max/2	4 096	4 096	5	softmax	8	21.51
V	11-48-4	max/2	5-128-1	max/2	3-192-1	3-192-1	3-128-1	max/2	2 048	2 048	5	softmax	8	14.58
VI	11-24-4	max/2	5-64-1	max/2	3-96-1	3-96-1	3-64-1	max/2	1 024	1 024	5	softmax	8	4.02
VII	11-12-4	max/2	5-32-1	max/2	3-48-1	3-48-1	3-32-1	max/2	512	512	5	softmax	8	1.50
VIII	11-6-4	max/2	5-16-1	max/2	3-24-1	3-24-1	3-16-1	max/2	128	128	5	softmax	8	0.26

在 batch 偏小时（如图 8a 所示），在训练的过程中会遇到非常多的局部极小点，在步长和卷积方向的共同作用下，虽然 loss 值呈现不断减小的趋势，但在整个过程仍然会出现 loss 值跳变的情况。迭代在 60 000 次到 70 000 次之间出现了较大的 loss 值震荡，在 80 000 次迭代之后，loss 值趋于平稳。

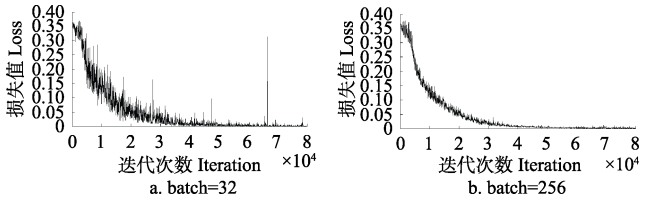


图 8　训练 AlexNet 训练时损失值 loss 收敛情况

Fig.8　Convergence of loss from training AlexNet

为降低 loss 值出现跳变的几率，本文将从以下几个方面对网络进行优化：首先将 lr 调节到 0.02，相当于间接增加了卷积的步长，在一定程度上可以避免训练产生的震荡，越过局部极小点继续向更大的极值点方向进行训练；对于每一层的偏置项从 0 设置为 0.1，限制激活阈值的大小，这样就降低了出现过大误差的概率，避免迭代方向出现较大的变化；继续增大 batch 的值，提高每次训练样本的覆盖率。

通过调整学习率和偏置项，网络训练的收敛性得到了很好的改善，但会带来整体收敛速度过慢的问题，因此需要增加最大迭代的次数。图 8b 是 batch=256 时的 loss 值分布情况，loss 值在 40 000 次迭代是就出现明显的收敛趋势且没有出现 loss 值跳变。因此，本文选择 batch=256 训练分类网络。

根据预先准备的 5 分类 40 000 幅肉牛关键部位图像数据做样本，其中训练集 32 000 幅，测试集 8 000 幅，针对表 2 中 8 种网络结构进行试验。参考 ILSVRC 的评判标准，使用 top-1 错误率（没有被网络正确分类的图像数与样本集图像总数的比例）评价个网络的性能。其中 8 中网络的 top-1 错误率（%）分别为 0.312（网络 I）、0.608（网络 II）、0.763（网络 III）、0.453（网络 IV）、0.598（网络 V）、0.795（网络 VI）、1.276（网络 VII）、6.641（网络 VIII）。

网络 I 和网络 IV 具有较高的分类精度，而网络 VIII 的性能最差。在网络宽度相同的前提下，层数越多分类精度越高（如网络 I 的精度要高于网络 II，网络 II 的精度高于网络 III）；在网络深度相同时，通过增加网络宽度，会使分类精度有所提高（网络 I、IV、V、VI、VII、VIII 的精度递减），这是由于宽度增加使每个卷积层的卷积核数量也会增加，这样可以从输入图像中提取更多的特征，以此来提高网络分类性能。但层数越多（特别是全连接层），网络越宽，参数总量就越大，训练时间就越长，因此根据分类数量和样本数量来调整网络结构，本文为了综合精度和训练时间，选择网络 VI 作为本试验的分类网络。

4　基于 DRGB 的目标识别网络实现

4.1　识别网络设计与精调

目标识别过程，除了要对目标对象进行分类，更重要的是找到目标对象的正确位置。因此在获得高精度分类网络后，需要根据识别对象的特征对分类网络进行参数微调（fine-tuning），同时根据真值区域（ground truth）的位置，对所有候选区域（region proposals）进行合并或删除操作，最终保留概率最大的边框（bounding-box）作为该对象的识别位置。

本文参考了 Fast R-CNN 的实现方法，利用 RoI（Region of Interesting）池化取代分类网络的最后一个池化层，设计出针对肉牛形体部位（头、躯干、腿、尾）的识别网络，如图 9 所示。通过卷积-池化层对输入的整幅图像进行特征提取，并生成特征图；利用 Selective Search 在 DRGB 图像上生成候选区域（如图 9 中的矩形候选区域对应的肉牛头部信息）；RoI 池化层根据候选区域到特征图的坐标投影，从特征图上获取候选区域特征，归一化为大小固定的输出特征，最终由全连接层和 softmax 分类器进行分类和识别，由 bounding box 回归器来进行边框位置定位。由于该识别网络对整幅图像只进行一次连续卷积操作，因此可以做到端到端处理，提高了该模型处理实时目标识别问题的能力。

本文选择网络 VI 作为图 9 的基本网络结构，利用 RoI 池化层替换网络 VI 的最后一个池化层。在 fine-tuning 前，选择 1 000 幅 DRGB 作为参与精调的训练集，通过人工标注肉牛头部、躯干、腿部和尾部等部位的真实区域（ground truth regions，GTRs），利用 Selective search 在每幅 DRGB 上获取 200 个左右的目标候选区域（object

region proposals，ORPs），利用 IoU（intersection over union）来计算 **ORP** 与 **GTR** 的重叠程度，其中 IoU = **ORP** ∩ **GTR**/**ORP** ∪ **GTR**，如果 IoU≥0.5，则该候选区域被标记为对应真实区域的类型（正例），否则被标记为背景（负例）。由于识别网络中负责特征提取部分与网络 IV 的结构一致，可以复用网络 VI 的卷积层进行图像特征提取，因此识别网络可以共享网络 VI 的所有权重参数，包括全部卷积层和 3 个全连接层。将肉牛图像的正、负例区域图像截取出来混入网络 VI 的训练样本，继续对网络进行训练，利用再次训练好的网络 VI 初始化识别网络。

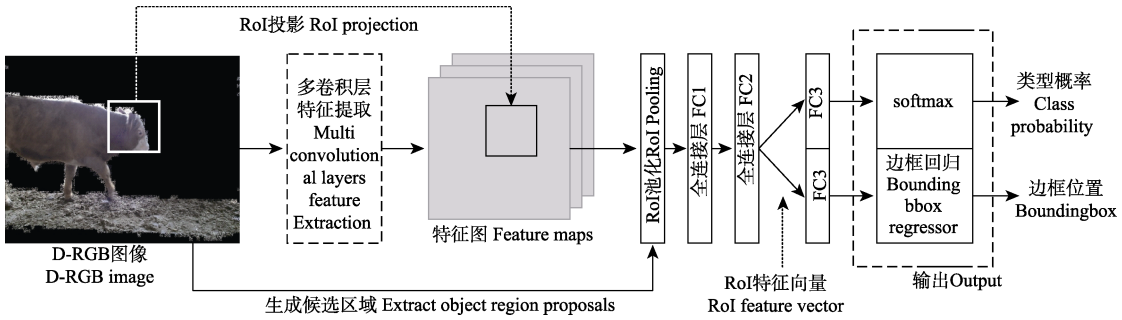


图 9 基于 FR-CNN 的肉牛关键部位识别网络

Fig.9 Recognition network for cattle key parks based on Fast R-CNN(FR-CNN)

4.2 测试与分析

为验证 DRGB 图像序列对网络识别性能的提升，本文同样利用 Fast RCNN 模型对 RGB 图像序列进行识别处理，并比较 2 次测试的平均精度[43]（average precision, AP）、全局平均精度 mAP（mean AP）[43]以及识别速度，结果如表 3 所示。测试结果证明，FR-CNN+DRGB 在检测速度（4.32 帧/s）上远远高于 FR-CNN+RGB 的检测速度（0.5 帧/s），而且前者的 mAP（75.88%）也高于后者的 mAP（68.07%）。其中，FR-CNN+DRGB 网络对肉牛头部的检测效果最好（86.32%），对尾部的检测效果最差（61.25%）。这是由于头部的形状比较单一，而且特征相比于其他部位更加明显；而尾部与腿部存在形状、纹理、颜色的相似性，因此特征相似。利用 FR-CNN+DRGB 网络对一段连续图像序列进行目标识别处理，截取其中一段的识别效果如图 10 所示，从对连续帧处理的结果上看，在肉牛行走过程中牛腿、牛头、牛身都可以很清晰的识别出来，而牛尾本身在行走过程中可能会隐藏在牛腿间，而且形态特征类似于牛腿，因此会在个别图像中没有成功识别，但这并不影响肉牛整体形态的识别。而通过观察可以看出，每个识别的目标基本可以与肉牛形体关键部位对应，实现了对运动中肉牛关键位置的识别。

表 3 肉牛关键部位检测速度和平均精度

Table 3 Detection speed and average precision of cattle key parts

检测网络 Detection network	检测速度 Detection speed /(帧·s^{-1})	平均精度 Average precision AP/%				
		头部 Head	躯干 Body	腿部 Leg	尾部 Tail	平均 Mean(mAP)/%
FR-CNN+DRGB	4.32	86.32	82.61	73.34	61.25	75.88
FR-CNN+RGB	0.5	77.69	75.33	65.86	53.41	68.07

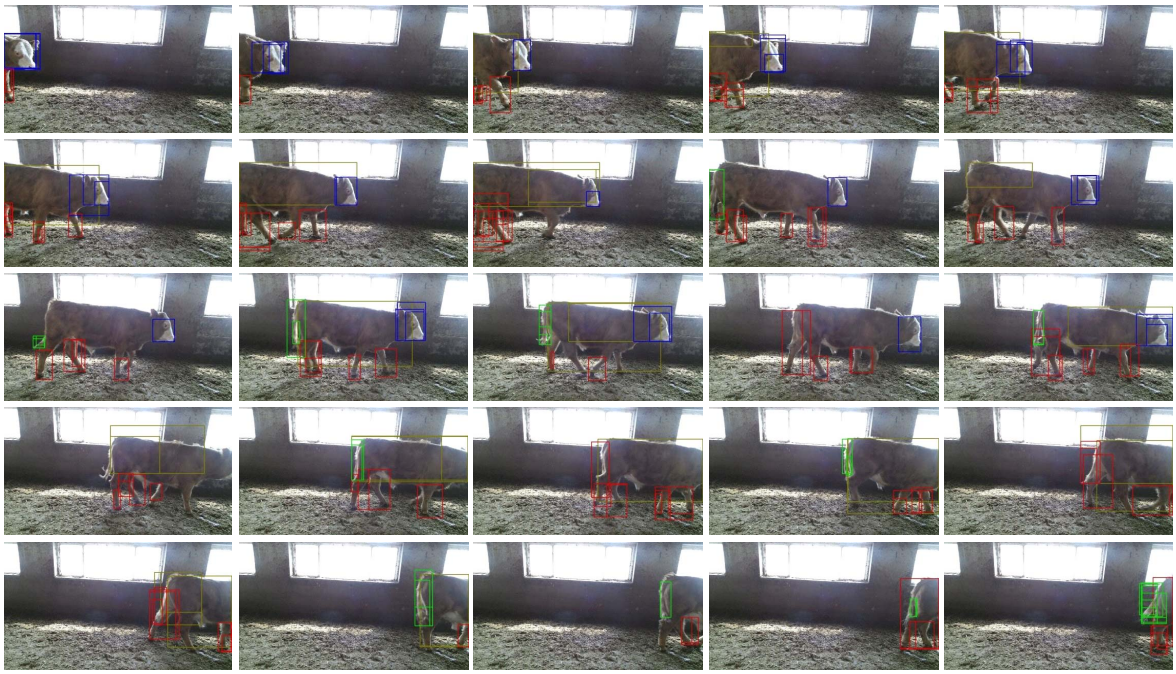


图 10 部分运动中的肉牛形态部位识别结果

Fig.10 Partly body shape parts recognition results of moving cattle

5 结 论

本文利用 Kinect 在相同时间维度下采集肉牛运动过程的 2 种模态信息（Depth and RGB，DRGB），并针对 2 种模态信息进行相应的处理，试验结果表明：利用随机最近邻像素比较法（random nearest neighbor pixel comparison, RNNPC）来自动获取运动中肉牛连续帧图像，可以减少 72%的无效帧数据，且平均有效帧比率约为 94%；将 RGB 图像与 Depth 图像进行像素点映射，并利用 Depth 图像中动态变化区域的均值深度来过滤 RGB 图像背景，生成 DRGB 图像，经 Selective Search 算法测试，目标候选区域可以减少一个约数量级；基于 AlexNet 设计出 8 种分类网络，通过调整深度卷积神经网络结构和参数变化策略，可以提高这 8 类分类网络练时的收敛速度，同时参照 Fast-RCNN 构造了最终目标识别网络。利用 DRGB 样本训练后的识别网络在识别平均分类精度可以达到 75.88%，识别速度可以达到 4.32 帧/s，而利用 RGB 样本训练后的原 Fast RCNN 网络在分类精度上可以达到 68.07%，识别速度可以达到 0.5 帧/s，因此基于 DRGB 的识别网络要优于原生 Fast RCNN。综合上述方法，最终可以实现对运动时肉牛关键部位的识别。

[参 考 文 献]

[1] 国家统计局. 2016 年国民经济和社会发展统计公报[EB/OL].http://www.stats.gov.cn/tjsj/zxfb/201702/t20170228_1467424.html.

[2] 罗锡文，廖娟，胡炼，等. 提高农业机械化水平促进农业可持续发展[J]. 农业工程学报，2016，32(1)：1－11.
Luo Xiwen, Liao Juan, Hu Lian, et al. Improving agricultural mechanization level to promote agricultural sustainable development[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2016, 32(1): 1－11. (in Chinese with English abstract)

[3] 何东健，孟凡昌，赵凯旋，等. 基于视频分析的犊牛基本行为识别[J]. 农业机械学报，2016，47(9)：294－300.
He Dongjian, Meng Fanchang, Zhao Kaixuan, et al. Recognition of calf basic behaviors based on video analysis[J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(9): 294－300. (in Chinese with English abstract)

[4] Yann Lecun, Yoshua Bengio, Geoffrey Hinton. Deep Learning[J]. Nature, 2015, 521: 436－444.

[5] Dahl G E, Yu D, Deng L, et al. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(1): 504－507.

[6] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504－507.

[7] Gawehn E, Hiss J A, Schneider G. Deep learning in drug discovery[J]. Molecular Informatics, 2016, 35(1): 3－14.

[8] Lecun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4): 541－551.

[9] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc, 2012: 1097－1105.

[10] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]// Computer Vision and Pattern Recognition. IEEE, 2015: 1－9.

[11] Srivastava R K, Greff K, Schmidhuber J. Highway networks[EB/OL]. https: //arxiv.org/abs/1505.00387.

[12] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]// International Conference on Learning Representations (ICLR), 2015.

[13] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE Computer Society, Las Vegas, NV, United States, 2016.

[14] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE Computer Society, Honolulu, Hawaii, United States, 2017.

[15] Deng J, Berg A, Satheesh S, et al. ImageNet large scale visual recognition competition 2012(ILSVRC2012) [EB/OL]. http: //www.image-net.org/challeges/ISVRC/2012/.

[16] Farabet C, Couprie C, Najman L, et al. Learning hierarchical features for scene labeling[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(8): 1915－1929.

[17] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2014.

[18] Girshick R. Fast R-CNN[C]// IEEE International Conference on Computer Vision (ICCV), 2015.

[19] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]// Annual Conference on Neural Information Processing Systems (NIPS), 2015.

[20] Tao Kong, Anbang Yao, Yurong Chen, et al. HyperNet: Towards Accurate Region Proposal Generation and Joint Object DetectionTao Kong[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016.

[21] Redmon, J, Divvala, S, Girshick, R, et al. A: You only look once unified, real-time object detection[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016.

[22] Wei Liu, Dragomir Anguelov, Dumitru Erhan, et al. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer Vision (ECCV), 2016.

[23] 田有文，程怡，王小奇，等. 基于高光谱成像的苹果虫伤缺陷与果梗/花萼识别方法[J]. 农业工程学报，2015，31(4)：325－331.
Tian Youwen, Cheng Yi, Wang Xiaoqi, et al. Recognition method of insect damage and stem/calyx on apple based on

hyperspectral imaging[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(4): 325－331. (in Chinese with English abstract)

[24] 周云成，许童羽，郑伟，等. 基于深度卷积神经网络的番茄主要器官分类识别方法[J]. 农业工程学报，2017，33(15)：219－226.
Zhou Yuncheng, Xu Tongyu, Zheng Wei, et al. Classification and recognition approaches of tomato main organs based on DCNN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(15): 219－226. (in Chinese with English abstract)

[25] 贾伟宽，赵德安，刘晓样，等. 机器人采摘苹果果实的 K-means 和 GA-RBF-LMS 神经网络识别[J]. 农业工程学报，2015，31(18)：175－183.
Jia WeiKuan, Zhao Dean, Liu Xiaoyang, et al. Apple recognition based on K-means and GA-RBF-LMS neural network applicated in harvesting robot[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(18): 175－183. (in Chinese with English abstract)

[26] 赵源深，贡亮，周斌，等. 番茄采摘机器人非颜色编码化目标识别算法研究[J]. 农业机械学报，2016, 47(7): 1－7.
Zhao Yuanshen, Gong Liang, Zhou Bin, et al. Object recognition algorithm of tomato harvesting robot using non-color coding approach[J]. Transactions of the Chinese Society for Agricultural Engineering, 2016, 47(7): 1－7. (in Chinese with English abstract)

[27] 杨国国，鲍一丹，刘子毅. 基于图像显著性分析与卷积神经网络的茶园害虫定位与识别[J]. 农业工程学报，2017，33(6)：156－162.
Yang Guoguo, Bao Yidan, Liu Ziyi. Localization and recognition of pests in tea plantation based on image saliency analysis and convolutional neural network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)，2017, 33(6): 156－162. (in Chinese with English abstract)

[28] 谭文学，赵春江，吴华瑞，等. 基于弹性动量深度学习的果体病例图像识别[J]. 农业机械学报，2015，46(1)：20－25.
Tan Wenxue, Zhao Chunjiang, Wu Huarui, et al. A deep learning network for recognizing fruit pathologic images based on flexible momentum[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(1): 20－25. (in Chinese with English abstract)

[29] 王献锋，张善文，王震，等. 基于叶片图像和环境信息的黄瓜病害识别方法[J]. 农业工程学报，2014，30(14)：148－153.
Wang Xianfeng, Zhang Shanwen, Wang Zhen, et al. Recognition of cucumber diseases based on leaf image and environmental information[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2014, 30(14): 148－153. (in Chinese with English abstract)

[30] 王新忠，韩旭，毛罕平. 基于吊蔓绳的温室番茄主茎秆视觉识别[J]. 农业工程学报，2012，28(21)：135－141.
Wang Xinzhong, Han Xu, Mao Hanping. Vision-based detection of tomato main stem in greenhouse with red rope[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2012, 28(21): 135－241. (in Chinese with English abstract)

[31] 郭艾侠，熊俊涛，肖德琴，等. 融合 Harris 与 SIFT 算法的荔枝采摘点计算与立体匹配[J]. 农业机械学报，2015, 46(12): 11－17. (in Chinese with English abstract)
Guo Aixia, Xiong Juntao, Xiao Deqin, et al. Computation of picking point of litchi and its binocular stereo matching based on combined algorithms of Harris and SIFT[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(12): 11－17. (in Chinese with English abstract)

[32] 赵凯旋，何东键. 基于卷积神经网络的奶牛个体身份识别方法[J]. 农业工程学报，2015，31(5)：181－187.
Zhao Kaixuan, He Dongjian. Recognition of individual dairy cattle based on convolutional neural networks[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(5): 181－187. (in Chinese with English abstract)

[33] 段延娥，李道亮，李振波，等. 基于计算机视觉的水产动物视觉特征测量研究综述[J]. 农业工程学报，2015，31(15)：1－11.
Duan Yan’e, Li Daoliang, Li Zhenbo, et al. Review on visual characteristic measurement research of aquatic animals based on computer vision[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(15): 1－11. (in Chinese with English abstract)

[34] 高云，郁厚安，雷明刚，等. 基于头尾定位的群猪运动轨迹追踪[J]. 农业工程学报，2017，33(2): 220－226.
Gao Yun, Yu Hou’an, Lei Minggang, et al. Trajectory tracking for group housed pigs based on locations of head/tail[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(2): 220－226. (in Chinese with English abstract)

[35] Nitish Srivastava, Ruslan Salakhutdinov. Multimodal learning with deep Boltzmann machines[C]// International Conference on Neural Information Processing System (NIPS), 2012: 2222-2230.

[36] Microsoft. Developing with Kinect for Windows[EB/OL]. https://developer.microsoft.com/en-us/windows/kinect/develop.

[37] Uijlings J, Vandesande K, Gevers T, et al. Selective search for object recognition[J]. International Journal of Computer Vision. 2013, 104(2): 154－171.

[38] Alex Krizhevsky, Ilya Sutskever, Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks[C]// Proceedings of the 25th International Conference on Neural Information Processing Systems. 2012-12-03, 1097－1105.

[39] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2014, 115(3): 211－252.

[40] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition[EB/OL]. https://arxiv.org/abs/1512.03385.

[41] Abadi M, Barham P, Chen J, et al. TensorFlow: A system for large-scale machine learning[C]//Usenix Conference on Operating Systems Design & Implementation, 2016.

[42] Nvidia. Nvidia Tesla K40[EB/OL]. www.nvidia.cn/object/tesla_product_literature_cn.html.

[43] Everingham M, Gool L V, Williams C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303－338.

Body shape parts recognition of moving cattle based on DRGB

Deng Hanbing[1,2], Xu Tongyu[1,2※], Zhou Yuncheng[1,2], Miao Teng[1,2,3], Zhang Yubo[1,2], Xu Jing[1,2], Jin Li[1,2], Chen Chunling[1,2]

(1. *College of Information and Electrical Engineering, Shenyang Agricultural University, Shenyang* 110866*, China*; 2. *Liaoning Engineering Research Center for Information Technology in Agriculture, Shenyang* 110866*, China*; 3. *Beijing Research Center for Information Technology in Agriculture, Beijing* 100097*, China*)

Abstract: Body shape parts acquisition and recognition from moving cattle are difficult to realize automatically especially with the similar color and texture frames or images. Usually, the cattle's abnormal behavior is hardly detected by human because abnormal behavior detection needs continuous observation, but we can solve this problem with our method. In this paper, we use the Microsoft Company's smart vision sensor (named Kinect) to collect the information of 2 modals from cattle movement (the depth information and RGB information, DRGB). The depth information can be acquired by the infrared sensor and the depth value means the distance between the object and the sensor. The RGB information can be acquired by the normal camera. Based on the color modal information, we propose a randomized nearest neighbor pixel comparison (RNNPC) method to snatch at continuous action frames without static frames. By comparing the distinction of pixels between 2 adjacent frames, we can judge whether there is obvious or micro actions appearing in these 2 neighboring images. And if the distinction value is more than the threshold, the camera would record the cattle's continuous movements or actions automatically. Based on the depth modal information, we can calculate the mean value of depth which is obtained in the dynamic region by the RNNPC method. And with the mean value of depth, we can filter and transform the invalid pixels into dark pixels in the continuous frames. Meanwhile, the intact shape of the cattle in the original picture is preserved by our method (save the pixels with no change). From the results of filtration, we can see that the original image background area is reduced, and the method (we use SelectiveSearch in this paper) also directly reduces the number of candidate regions compared with the conventional recognition algorithm. According to the training samples scale, the network performance and the characteristic of the key parts of cattle from continuous frames, we adjust the parameters of the deep convolutional neural network, and we set the iterating upper limit and finish the network training when the number of iterating reaches this limit. Finally, with the trained network, we can realize the recognition of key parts from moving cattle in the processed continuous frames. The experimental results show that RNNPC method can save 72% of storage space, and the ratio of valid data for the other 38% of continuous frames can reach 94%. By filtering the invalid pixels, the invalid background information or objects of the original images can be almost removed, and the number of candidate objects generated by the conventional object recognition algorithm can be reduced by an order of magnitude. Compared to the original image, 90% candidate regions are reduced with SelectiveSearch algorithm on DRGB image. By adjusting network parameters, we can improve the convergence speed of deep convolutional neural network in training the samples with similar color and texture, and the single training iteration time does not significantly increase. The average classification accuracy of the net can reach 75.88%, and the image processing rate is 4.32 FPS, and under the same condition, the effect is better than that by the original Fast RCNN. By using the method mentioned above, we can realize body shape parts acquisition and recognition from moving cattle.

Keywords: image reconstruction; image recognition; algorithms; cattle; deep convolutional neural network; object recognition; DRGB