

基于无人机图像的羊群密集计数算法研究

赵建敏, 李雪冬, 李宝山*

内蒙古科技大学信息工程学院, 内蒙古 包头 014010

摘要 羊群羊只计数是畜牧生产过程中较为耗时费力的环节,在草原畜牧业中,统计羊只数量有助于过度放牧监测和草原生态评估。采用无人机(UAV)获取航拍图像,制作了羊群密集计数(USC)数据集,为羊群密集计数研究提供数据支持。在 USC 数据集上,对 MCNN、CSRNet、SFANet、Bayesian Loss 网络模型进行了实验。实验结果表明:由于遮挡,羊群的不规则分布,羊群大小、形状、密度等方面的巨大变化,MCNN、CSRNet、SFANet 模型将假设的高斯核应用于点标注计算真值密度图,很难达到高质量;而 Bayesian Loss 模型提出的 Bayesian 损失函数对每只羊标注点的计数期望进行监督,取得了较好的计数效果,密度图平均绝对误差(MAE)为 3.56,均方误差(MSE)为 5.46,平均相对误差(MRE)为 1.86%,可为草原羊群密集计数提供有益参考。

关键词 图像处理;卷积神经网络;密集计数;羊群计数;密度估计;无人机

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP202158.2210013

Algorithm of Sheep Dense Counting Based on Unmanned Aerial Vehicle Images

Zhao Jianmin, Li Xuedong, Li Baoshan*

School of Information Engineering, Inner Mongolia University of Science & Technology, Baotou, Inner Mongolia 014010, China

Abstract It is a time-consuming and laborious task to manually count the number of sheep in the process of pastoral livestock production. Counting the number of sheep in grassland animal husbandry has been helpful for overgrazing monitoring and grassland ecology assessment. A unmanned aerial vehicle (UAV) is used to obtain aerial images, and a UAV sheep counting (USC) dataset is made to provide data support for the study of flock dense counting. Based on the USC dataset, experiments with network models, included MCNN, CSRNet, SFANet, and Bayesian Loss are carried out. The experimental results show that due to occlusion, irregular distribution of sheep, great changes in sheep size, shape, and density, MCNN, CSRNet, and SFANet models apply the assumed Gaussian kernel to point labeling to calculate the truth density map, which are difficult to achieve high quality. However, the Bayesian loss function proposed by Bayesian Loss model supervises the counting expectation of each sheep's labeling points. The average absolute error (MAE) of the density map obtained by the Bayesian Loss model is 3.56, the mean square error (MSE) is 5.46, and the average relative error (MRE) is 1.86%, which provides a useful reference for the dense counting of grassland sheep.

Key words image processing; convolutional neural network; dense counting; sheep counting; density estimation; unmanned aerial vehicle

OCIS codes 100.4996; 100.2000; 100.3008

收稿日期: 2020-12-15; 修回日期: 2021-01-22; 录用日期: 2021-02-12

基金项目: 内蒙古自然科学基金(2019MS06021, 2019LH06006)、内蒙古自治区科技重大专项(2019ZD025)

通信作者: *libaoshan@imust.edu.cn

1 引言

内蒙古羊养殖业在草原畜牧业中占有重要地位。在实际畜牧生产过程中,羊群羊只计数是必不可少的环节,传统方式主要依靠人工统计,费时费力。自动、准确地预测羊群羊只数量具有重要意义,可以为过度放牧监测和草原生态评估提供数据支持。但草原放牧区域广泛、场景光照变化大、羊群分布不均、羊只之间遮挡严重等,为基于图像的羊群羊只计数提出新的挑战。

目前,国内外学者普遍采用基于目标检测的方法进行计数研究,并且取得了一定的效果^[1-5]。这种方法虽然计数精度高,但在拥挤场景中计算高密度区域的目标数量时,无法很好地克服背景和遮挡干扰等问题,目标检测的难度限制了大规模密集群体计数精度的进一步提高。近年来,人群密集计数方法发展迅猛,传统方法主要包括基于回归的方法和基于密度估计的方法。针对高密度场景中遮挡严重的问题,相关研究人员提出了基于回归的方法,该方法通常通过建立区域图像特征与区域目标数目的回归模型来估计场景中的目标数量^[6]。然而,基于回归的方法会受到尺度和视角急剧变化的影响,导致准确度很难得到进一步提升。2007年,Lempitsky等^[7]开创性地提出了基于密度估计的目标计数算法框架,该算法已经能够利用物体的空间分布信息和语义信息,为后续深度学习的方法奠定了理论基础,但该算法太依赖于密度图的质量。上述方法由于自身的局限性,不适宜处理较为复杂的密集群体分布情况。

随着卷积神经网络(CNN)和计算机视觉计数的迅速发展,同时密集人群计数领域中也出现了更具有挑战性的数据集^[8-12],越来越多的研究者们利用CNN来改进密度估计方法,使得基于CNN的方法在目标计数和密度估计方面取得了巨大成

功^[13-16]。Zhang等^[17]首先采用CNN来解决人群计数问题。Zhang等^[10]提出了一种包含三列卷积核尺寸各异的多列卷积神经网络(MCNN)模型用于人群计数,这是一项明确关注多尺度问题的开创性工作。Li等^[18]提出了CSRNet网络,该网络去除了多列CNN,使用空洞卷积扩大感受野,利用扩张的卷积层来聚集拥挤场景中的多尺度上下文信息。Zhu等^[19]提出了一种基于双路径多尺度融合解码器的解编码网络SFANet来进行人群计数。Ma等^[20]构建了一个概率密度模型,并提出新的Bayesian损失函数来计算估计密度图和真实密度图之间的差异。

不断涌现的人群计数模型对高密度羊群计数有较大的借鉴作用,作为深度学习模型训练的基础,数据集的作用至关重要。本文提出了一种基于无人机图像的草原羊群计数方法,制作了羊群密集计数(USC)数据集,采用Bayesian Loss模型计算羊群羊只数量,有助于实时监测草原放牧强度,为有关部门进行草原管理、草原生态保护等提供决策支持。在USC数据集上,对MCNN、CSRNet、SFANet、Bayesian Loss网络模型进行了对比实验并分析比较了实验结果。实验结果表明,与其他模型利用高斯核将标注点转换为真值密度图的方法不同,Bayesian Loss模型提出的Bayesian损失函数对每个标注点的计数期望进行更可靠的监督,取得了较好的计数效果。

2 羊群密集计数算法

2.1 Bayesian Loss 模型

Bayesian Loss模型由密度图生成网络和计数估计算法两部分组成。密度图生成网络的主干网络是VGG19^[21],输入密集羊群图像后,该网络可以输出估计密度图;计数估计算法通过处理估计密度图实现计数估计。整个模型流程如图1所示。

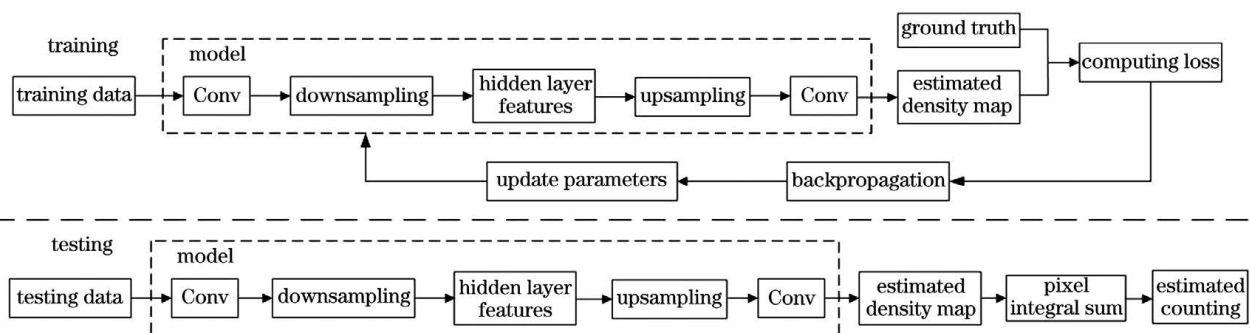


图1 Bayesian Loss模型流程图

Fig. 1 Flow chart of Bayesian Loss model

密度图生成网络首先将 1595 pixel×1432 pixel 的原始图像缩放为 512 pixel×512 pixel 的图像;其次采用 VGG19 网络获取多尺度特征,经过 5 个卷积层、4 个池化层、5 个 ReLU 函数后(卷积层中卷积核大小均为 3×3,池化层采用 2×2 的卷积核进

行最大池化),实现特征图的下采样,得到高维隐层特征;然后对特征图进行上采样,经过 2 个 3×3 的卷积微调后通过 1 个 1×1 的卷积生成 512×512 的估计密度图;最后对密度图像素进行积分得到羊群羊只数量。整个 Bayesian Loss 模型结构如图 2 所示。

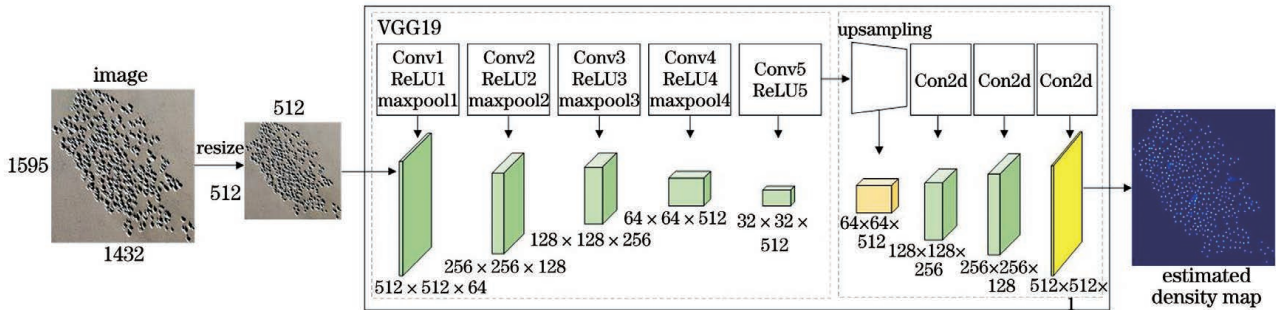


图 2 Bayesian Loss 模型结构图

Fig. 2 Structural diagram of Bayesian Loss model

Bayesian Loss 模型不再依赖点标注生成不完美的伪标签,直接采用数据集所提供的点标注作为弱监督信号。为了使监督信号格式匹配,在估计得到的概

率密度图的基础上又多多进行了一步期望操作,利用离散的概率密度期望值与离散的点标注设计损失函数并进行期望值意义上的回归估计,如图 3 所示。

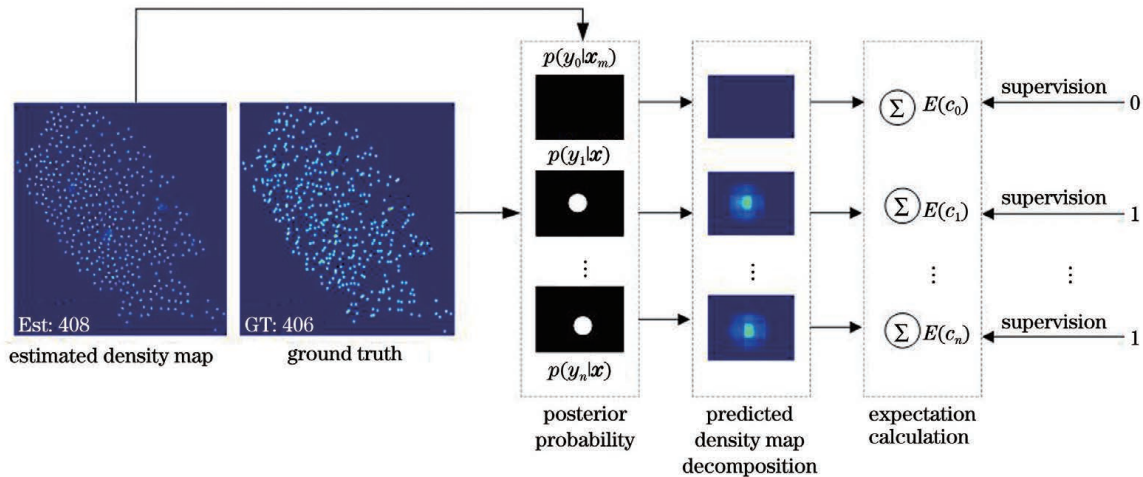


图 3 Bayesian 损失函数示意图

Fig. 3 Schematic diagram of Bayesian loss function

将数据集所提供的点标注作为先验值,使用二维高斯分布来近似目标的似然概率分布,根据给定的标签来构建 x_m 的似然函数(x_m 表示估计密度图中一个二维像素位置),即当出现目标标签为 y_n 的目标时,羊只出现在位置 x_m 的条件概率的表达式为

$$p(x_m | y_n) = N(x_m; z_n, \sigma^2 \mathbf{1}_{2 \times 2}), \quad (1)$$

式中: x_m 表示给定密度图中一个二维像素位置; $y_n = n$ 为对应的标签; $\{(z_n, y_n), n = 1, 2, \dots, N\}$ 表示样本图像的点标注图, N 为羊只总数, z_n 表示羊只标注点位置; $N(x_m; z_n, \sigma^2 \mathbf{1}_{2 \times 2})$ 为 x_m 处的二维高斯分布。根据 Bayesian 定理, x_m 具有 y_n 标签的后验概率为

$$p(y_n | x_m) = \frac{p(x_m | y_n) p(y_n)}{p(x_m)} = \frac{p(x_m | y_n) p(y_n)}{\sum_{n=1}^N p(x_m | y_n) p(y_n)} = \frac{p(x_m | y_n)}{\sum_{n=1}^N p(x_m | y_n)} = \frac{N(x_m; z_n, \sigma^2 \mathbf{1}_{2 \times 2})}{\sum_{n=1}^N N(x_m; z_n, \sigma^2 \mathbf{1}_{2 \times 2})}. \quad (2)$$

后验概率场定义了空间中每一个位置属于某一个目标的概率,基于此,计算每一个目标的期望计数,出现第 n 只羊的期望为

$$E(c_n) = E\left(\sum_{m=1}^M c_n^m\right) = \sum_{m=1}^M p(y_n | \mathbf{x}_m) D^{\text{est}}(\mathbf{x}_m), \quad (3)$$

式中: $D^{\text{est}}(\mathbf{x}_m)$ 是神经网络预测的目标概率密度, $m=1, 2, \dots, M$, M 是估计密度图中的总像素数; c_n^m 表示空间位置 \mathbf{x}_m 对其标签 y_n 的概率,即某一空间位置可能存在羊只的概率大小, c_n 是与 y_n 相关的总数。(3)式利用后验概率 $p(y_n | \mathbf{x}_m)$ 和估计密度图 $D^{\text{est}}(\mathbf{x}_m)$ 进行相乘相加得到每一个目标的期望计数,由于任意给定的目标理想计数期望应该是 1,因此该目标的 Bayesian 损失函数为

$$L^{\text{Bayes}} = \sum_{n=1}^N F[1 - E(c_n)], \quad (4)$$

式中: $F[1 - E(c_n)]$ 为未达到理想期望值 1 的损失函数。

为了更好地对背景像素进行建模, Bayesian Loss 模型引入了一个额外的背景标签 $y_0 = 0$, 令 $p(y_n) = p(y_0) = \frac{1}{N+1}$, 后验概率重写为

$$p(y_0 | \mathbf{x}_m) = \frac{p(\mathbf{x}_m | y_0)}{\sum_{n=1}^N p(\mathbf{x}_m | y_n) + p(\mathbf{x}_m | y_0)}. \quad (5)$$

前景的期望计数和整个背景的期望计数分别为

$$E(c_n) = \sum_{m=1}^M p(y_n | \mathbf{x}_m) D^{\text{est}}(\mathbf{x}_m), \quad (6)$$

$$E(c_0) = \sum_{m=1}^M p(y_0 | \mathbf{x}_m) D^{\text{est}}(\mathbf{x}_m). \quad (7)$$

实际希望得到的结果是背景计数为 0, 每个标注点的前景计数为 1, 因此提出增强损失函数。

$$L^{\text{Bayes}^+} = \sum_{n=1}^N F[1 - E(c_n)] + F[0 - E(c_0)]. \quad (8)$$

为了定义背景可能性, Bayesian Loss 为每个像素构造了一个虚拟背景点 \mathbf{z}_0^m ,

$$\mathbf{z}_0^m = \mathbf{z}_n^m + d \frac{\mathbf{x}_m - \mathbf{z}_n^m}{\|\mathbf{x}_m - \mathbf{z}_n^m\|_2}, \quad (9)$$

式中: \mathbf{z}_n^m 是距离 \mathbf{x}_m 最近的羊只标注点; d 是羊只标注点与虚拟背景点的距离。当某个像素点距离羊只标注点比较远时, 则为该位置分配一个背景标签, 如图 4 所示。

预测过程中, 给定输入预测图片, 对卷积神经网络模型得到的羊群密度估计图 $D^{\text{est}}(\mathbf{x}_m)$ 进行求和, 得到该图片中羊只的预测目标数目,

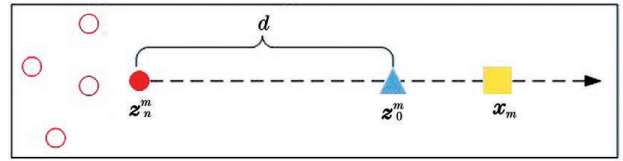


图 4 虚拟背景点几何图

Fig. 4 Geometric diagram of virtual background point

$$C = \sum_{n=1}^N E(c_n) = \sum_{n=1}^N \sum_{m=1}^M p(y_n | \mathbf{x}_m) D^{\text{est}}(\mathbf{x}_m) = \sum_{m=1}^M D^{\text{est}}(\mathbf{x}_m). \quad (10)$$

大量的实验表明了 Bayesian Loss 模型在准确性、鲁棒性和泛化能力方面具有优势, 因此本实验借鉴 Bayesian Loss 提出的训练损失函数对每只羊标注点的计数期望进行监督, 从而进行准确计数。

2.2 密度图

羊群密度图以特殊高亮的形式显示羊群分布位置, 积分值代表图像中的羊只数量, 密集计数图像的标注就是标记图像中羊只的位置, 以羊只身体部位中心点作为羊只位置点进行标注。图 5 为经过自适应高斯核处理后的羊群密度结果图, 每只羊身体部位区域的概率之和为 1, 得到完整的羊群图像密度图后, 对其进行积分求和, 就得到了羊群羊只的数量。为了保存更多的空间信息, 使用归一化高斯核模糊每只羊的标注, 估计图像中每只羊的大小并转换为密度图^[22]。

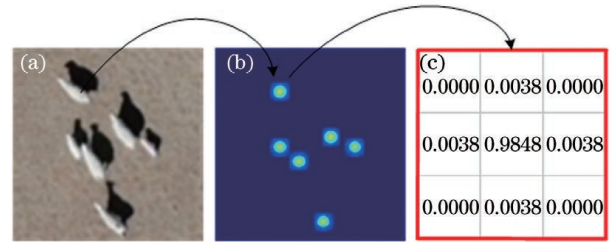


图 5 羊群密度图的生成。(a)原图; (b)羊群密度图;

(c)几何自适应高斯核

Fig. 5 Generation of sheep density map. (a) Original drawing; (b) sheep density map; (c) geometric adaptive Gaussian kernel

图像 \mathbf{x} 中第 i 个坐标 \mathbf{x}_i 为羊只身体部位标记点, 可以用 $\delta(\mathbf{x} - \mathbf{x}_i)$ 函数表示。对于一张具有 N 个羊只标注点的图像, 可以用 $H(\mathbf{x})$ 函数表示,

$$H(\mathbf{x}) = \sum_{i=1}^N \delta(\mathbf{x} - \mathbf{x}_i). \quad (11)$$

使用高斯核滤波器 G_σ ^[7] 与(11)式进行卷积, 即可得到密度图方程 $F(\mathbf{x})$,

$$F(\mathbf{x}) = \sum_{i=1}^N \delta(\mathbf{x} - \mathbf{x}_i) * G_{\sigma} \quad (12)$$

由于透视原理,在用二维彩色图像表示三维实际羊群场景的过程中,每一个像素点所表示的羊身体部位所占面积会存在较大差异,使用上述生成方式会导致生成的密度图不够准确。将图片的视角扭曲考虑到密度图生成中,采用自适应高斯滤波器与(11)式进行卷积,密度图方程重写为

$$\begin{cases} F(\mathbf{x}) = \sum_{i=1}^N \delta(\mathbf{x} - \mathbf{x}_i) * G_{\sigma_i(\mathbf{x})} \\ \sigma_i(\mathbf{x}) = \beta \bar{d}_i \end{cases}, \quad (13)$$

式中: \bar{d}_i 是标记点 \mathbf{x}_i 与最近的 K 只羊之间的平均距离; β 为参数。大量实验表明当 $\beta=0.3$ 时,生成的密度图质量最好^[10]。

3 实验设计及结果分析

3.1 实验环境及参数设置

将数据集按 4:1 分为训练集和测试集,其中训练集含 2808 张图像,测试集含 702 张图像。在羊群数据集 USC 上训练了 MCNN、CSRNet、SFANet 等模型,优化了模型参数。

模型训练和测试所使用的硬件平台为: Intel i7-7700k、32 GB 内存的 CPU, NVIDIA GeForce GTX 1080Ti × 2 的显卡。软件平台为: Ubuntu16.04、CUDA 9.0、cudnn7.5.1。

选用 Adam 优化器训练 MCNN 模型,参数配置为: epoch 为 2000,学习率为 1×10^{-5} ,动量为 0.9,训练迭代 22500 次。将在 USC 羊群数据集中随机挑选的 2528 张图像用于训练,280 张用于验证,702 张用于测试。

选用 stochastic gradient descent (SGD) 优化器训练 CSRNet 模型,参数配置为: epoch 为 400,学习率为 1×10^{-6} ,动量为 0.95,衰退率为 0.0005。将在 USC 羊群数据集中随机挑选的 2808 张图像用于训练,702 张用于测试。由于硬件平台的内存限制,将批处理规模设置为 1。

选用 Adam 优化器训练 SFANet 模型,参数配置为:学习率为 1×10^{-4} ,epoch 为 500。将在 USC 羊群数据集中随机挑选的 2808 张图像用于训练,702 张用于测试。值得一提的是, SFANet 模型是受到特征金字塔^[23]和注意力机制^[24]的启发而被提出的,既能准确计数,又能为拥挤场景的密集计数提供高质量的密度图。

选用 Adam 优化器训练 Bayesian Loss 模型,参数配置为:学习率为 1×10^{-5} ,衰退率为 0.0001,epoch 为 1000,高斯分布方差 σ 为 8。为了提高模型的计数准确率,将输入图像尺寸调整为 512 pixel × 512 pixel,将在 USC 羊群数据集中随机挑选的 2528 张图像用于训练,280 张用于验证,702 张用于测试。利用训练后的网络模型对测试图像进行了一系列实验,并验证了所提算法的性能。

3.2 数据集训练

3.2.1 评价指标

计数算法性能的评价标准主要采用量化评价准则来度量估计值与真实值之间的误差。MCNN、CSRNet、Bayesian Loss 等模型都采用平均绝对误差 (MAE)、均方误差 (MSE) 和平均相对误差 (MRE) 作为评价指标。为能够较好进行对比分析,所提算法也采用上述指标进行羊群计数模型验证。各指标的表达式为

$$E_{\text{MAE}} = \frac{1}{N'} \sum_{j=1}^{N'} |C_j - C_j^{\text{GT}}|, \quad (14)$$

$$E_{\text{MSE}} = \sqrt{\frac{1}{N'} \sum_{j=1}^{N'} (C_j - C_j^{\text{GT}})^2}, \quad (15)$$

$$E_{\text{MRE}} = \frac{1}{N'} \sum_{j=1}^{N'} \frac{|C_j - C_j^{\text{GT}}|}{C_j^{\text{GT}}} \times 100\%, \quad (16)$$

$$C_j = \sum_{l=1}^L \sum_{w=1}^W z_{l,w}, \quad (17)$$

式中: N' 为测试图像数量; C_j 为第 j 幅测试图像的预测羊群羊只数量; C_j^{GT} 为第 j 幅测试图像的真实羊群羊只数量; L 和 W 是估计密度图的宽和高; $z_{l,w}$ 是估计密度图在 (l,w) 处的像素值,对模型输出的估计密度图上的每个像素进行求和得到估计总数。

3.2.2 数据集

由于缺乏公开羊群密集计数数据集,本实验组提出了羊群密集计数数据集 USC。该数据集由搭载 1200 万像素高清镜头的大疆 DJI 御 Mavic Air 固定翼型无人机,在距内蒙古苏尼特左旗某草原牧场地面 10~50 m 高度采集的图像组成。最终整理得到 3510 张草原羊群图像,这些图像包括了光照、背景、羊群数量和羊群分布等不同的情况,数据集中部分样本如图 6 所示。

利用改进的人群密集计数标注工具按照人群标准数据集^[11-12]格式标注 USC 数据集,如表 1 所示,其中 Total 指所有图片总的羊只标注点数;Min 指羊只标注点最少的图像的羊只标注点数;Ave 指所有图片平均羊只标注点数;Max 指羊只标注点最多

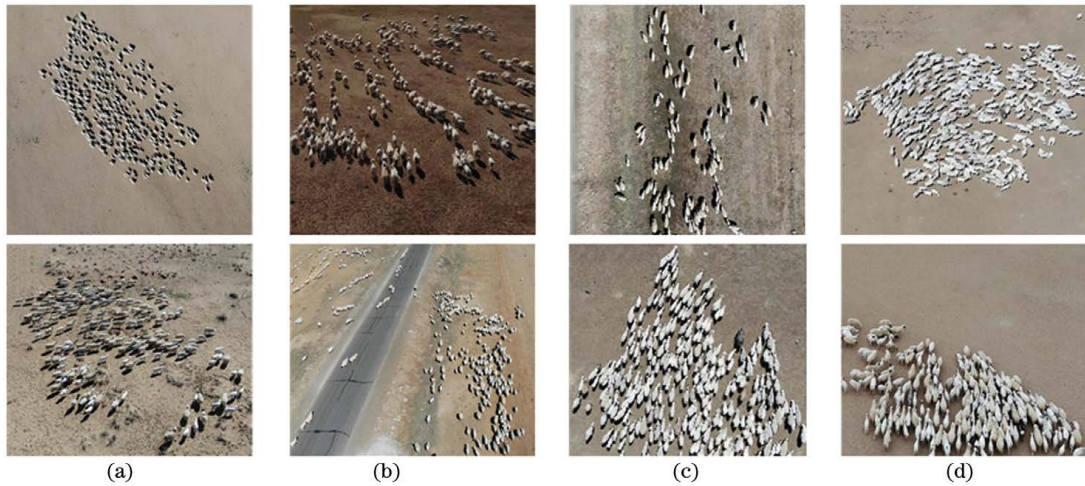


图 6 数据集图片示例。(a)光照不同;(b)背景不同;(c)羊群数量不同;(d)羊群分布不同

Fig. 6 Examples of dataset images. (a) Different illumination; (b) different backgrounds; (c) different number of sheep; (d) different distributions of sheep

表 1 USC 羊群数据集相关信息

Table 1 Related information of USC sheep dataset

Dataset	Number of images	Resolution	Total	Min	Ave	Max
USC	3510	Different	707257	45	201	487

的图像的羊只标注点数。

将羊群密集计数数据集按照 4:1 的比例随机划分为训练集和测试集,训练集和测试集在图像样本空间上没有交集。在此基础上采用数据增强的方法对图像样本进行随机裁剪、旋转、缩放等操作,扩增数据集样本数量,提高训练效果,提升模型性能,增强 CNN 模型的鲁棒性。

3.3 模型评估

3.3.1 Bayesian Loss 模型实验结果

图 7 是测试集中每帧图片预测值和真实值之间的分布,从图中可以看出,部分数据误差较大(第 162,236,332 帧图像等)。原因是这些图像中羊只目标较小且遮挡严重,后续会跟进此问题并进行相关改进工作。

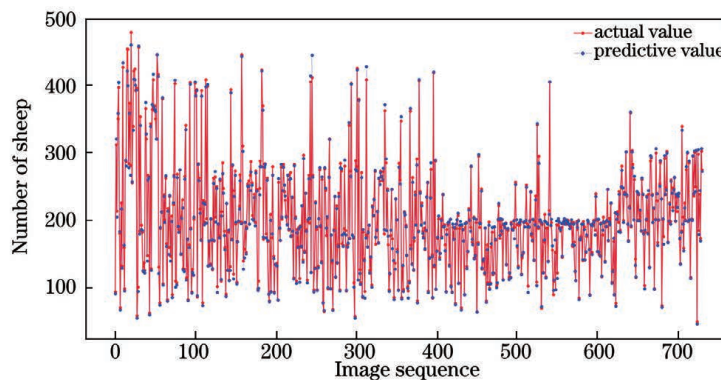


图 7 计数结果

Fig. 7 Counting result

Bayesian Loss 模型中,当高斯分布方差 $\sigma = 16$ 时,生成的密度图质量很模糊,如图 8 所示。预测密度图越暗,越能分清羊只和背景,越亮越不容易分清。从图 8 中可以看到,基本上羊只稀疏区域都是暗的,羊只密集区域比较亮,符合后验概率的计算原

理。 σ 越小图越清晰,因为小的 σ 不容易导致羊只区域重叠。

3.3.2 各个模型实验结果

为全面评估所提模型的性能,对各个模型进行了综合比较,结果如表 2 所示。

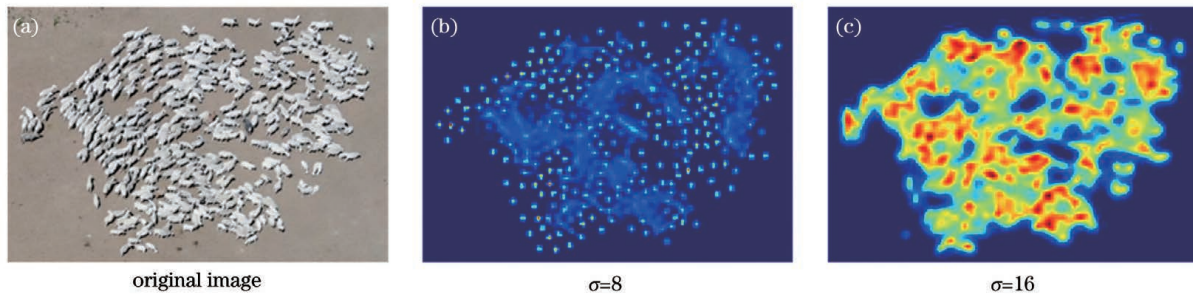


图 8 高斯核参数对模型的影响。(a)原始图像;(b) $\sigma=8$;(c) $\sigma=16$

Fig. 8 Effect of Gaussian kernel parameter on model. (a) Original image; (b) $\sigma=8$; (c) $\sigma=16$

表 2 羊群数据集上不同模型的综合对比

Table 2 Comprehensive comparison of different models on sheep dataset

Method	Backbone	Optimizer	Epoch	Batch_size	Lr_policy	Train time /h	MAE	MSE	MRE /%
MCNN	FS	Adam	2000	1	StepLR	180	43.34	51.05	21.58
CSRNet	VGG16	SGD	400	1	StepLR	52	21.58	29.80	10.90
SFANet	VGG16	Adam	500	8	StepLR	152	5.93	8.14	3.17
Bayesian Loss	VGG19	Adam	1000	16	StepLR	38	3.56	5.46	1.86

由表 2 可知, Bayesian Loss 模型的计数性能最好, 密度图 MAE 为 3.56, MSE 为 5.46, MRE 为 1.86%。

图 9 为不同模型的网络收敛效果图。从图中能

够看出, 所提模型随着 epoch 的增加, 模型的误差整体有同一趋势, 表明 epoch 的设计能够保证模型收敛。

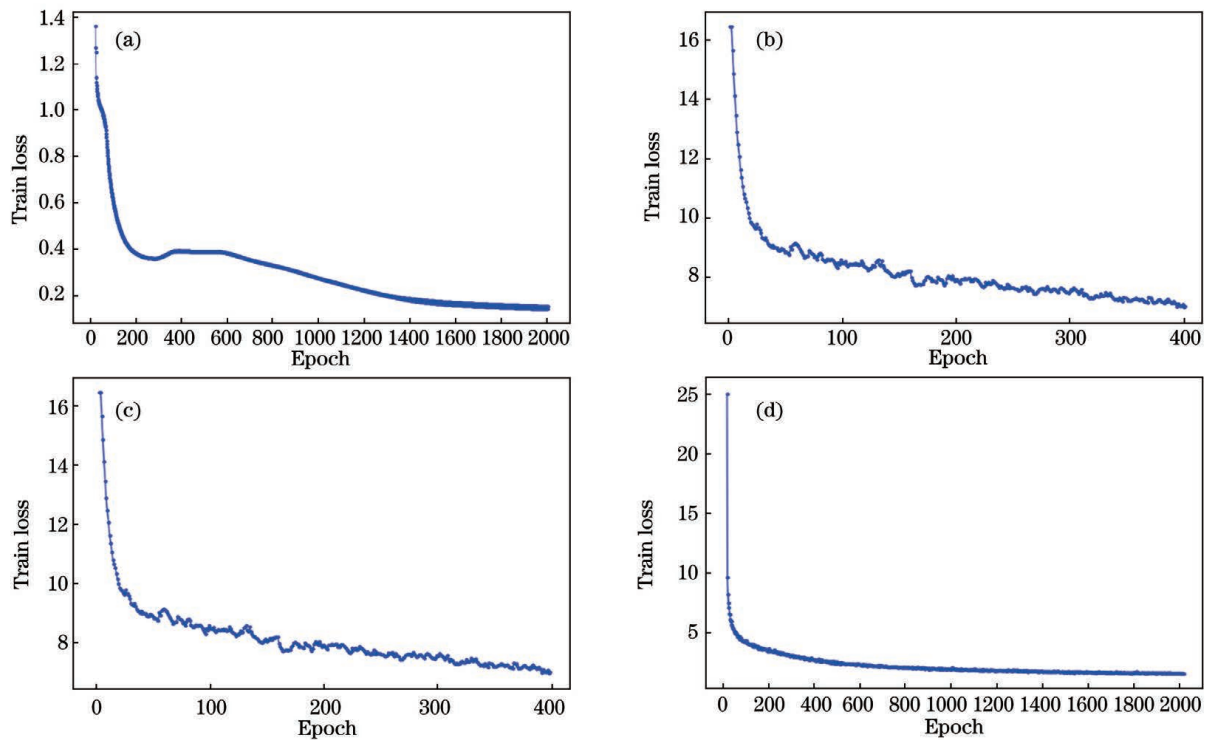


图 9 不同模型训练损失曲线。(a)MCNN;(b)CSRNet;(c)SFANet;(d) Bayesian Loss

Fig. 9 Training loss curves of different models. (a) MCNN; (b) CSRNet; (c) SFANet; (d) Bayesian Loss

3.3.3 各个模型对比实验分析

为全面评估所提模型, 对各个模型在羊群数据

集上的单张图像进行了对比实验, 图 10 展示了 4 个模型的可视化结果, 其中第 1 行是原始图像, 第 2 行

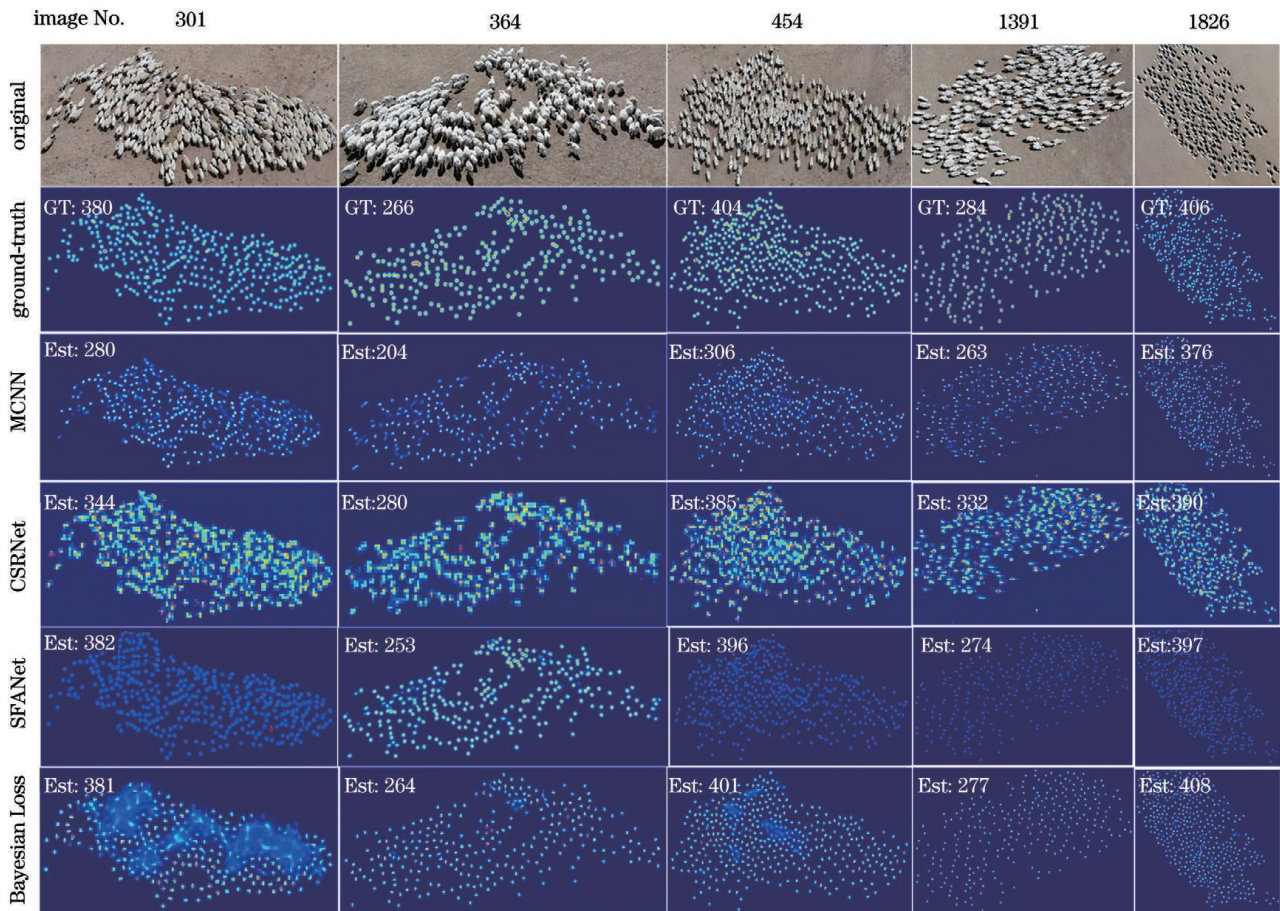


图 10 不同模型的可视化结果

Fig. 10 Visualization results of different models

是原始图像真实密度图,第 3 行是 MCNN 模型的估计密度图,第 4 行是 CSRNet 模型的估计密度图,第 5 行是 SFANet 模型的估计密度图,第 6 行是 Bayesian Loss 模型的估计密度图。从图中可以看出, Bayesian Loss 模型是测试集上计数精度最高的算法。

4 结 论

提出了一个基于无人机图像的背景复杂、密集尺度变化大的大规模羊群密集计数数据集 USC,为深度学习在该领域的研究和应用提供基础的数据支持。借鉴人群密集计数方法,对比实验了主流密集计数模型。定量和定性实验结果表明, Bayesian Loss 模型计数准确率最高,在羊群密集计数研究和应用领域给出有益的尝试,具有一定的应用价值。未来还可针对以下两点开展进一步研究:1)所提出的数据集场景单一,目标尺度变化小,在未来的工作中将继续扩充多场景同时涵盖多尺度羊只目标数据图像;2)由于拥挤场景下

羊群中的羊只互相遮挡,且身体较小,现有的计数模型在处理高遮挡场景时,效果依然不理想,这也是后续改进的重点。

参 考 文 献

- [1] Sarwar F, Griffin A, Periasamy P, et al. Detecting and counting sheep with a convolutional neural network[C]//2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), November 27-30, 2018, Auckland, New Zealand. New York: IEEE Press, 2018.
- [2] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [3] Shao W, Kawakami R, Yoshihashi R, et al. Cattle detection and counting in UAV images based on convolutional neural networks [J]. International Journal of Remote Sensing, 2020, 41(1): 31-52.
- [4] Tian L. Design of sheep number detection system

- [D]. Hohhot: Inner Mongolia University, 2019.
- 田磊. 羊群个数检测系统设计[D]. 呼和浩特: 内蒙古大学, 2019.
- [5] Redmon J, Farhadi A. Yolov3: an incremental improvement[EB/OL]. (2018-04-08)[2018-09-07]. <https://arxiv.org/abs/1804.02767>.
- [6] Chan A B, Vasconcelos N. Counting people with low-level features and Bayesian regression[J]. *IEEE Transactions on Image Processing*, 2012, 21(4): 2160-2177.
- [7] Lempitsky V, Zisserman A. Learning to parse images of articulated bodies[M]//Schölkopf B, Platt J, Hofmann T. *Advances in neural information processing systems 19*. Cambridge: The MIT Press, 2007.
- [8] Chan A B, Liang Z S J, Vasconcelos N. Privacy preserving crowd monitoring: counting people without people models or tracking[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK, USA. New York: IEEE Press, 2008: 10139874.
- [9] Idrees H, Saleemi I, Seibert C, et al. Multi-source multi-scale counting in extremely dense crowd images[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. New York: IEEE Press, 2013: 2547-2554.
- [10] Zhang Y Y, Zhou D S, Chen S Q, et al. Single-image crowd counting via multi-column convolutional neural network[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 589-597.
- [11] Idrees H, Tayyab M, Athrey K, et al. Composition loss for counting, density map estimation and localization in dense crowds[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11206: 544-559.
- [12] Wang Q, Gao J Y, Lin W, et al. NWPU-crowd: a large-scale benchmark for crowd counting and localization[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(6): 2141-2149.
- [13] Zuo J, Ba Y L. Population-depth counting algorithm based on multiscale fusion[J]. *Laser & Optoelectronics Progress*, 2020, 57(24): 241502.
- 左静, 巴玉林. 基于多尺度融合的深度人群计数算法[J]. *激光与光电子学进展*, 2020, 57(24): 241502.
- [14] Gao G S, Gao J Y, Liu Q J, et al. CNN-based density estimation and crowd counting: a survey[EB/OL]. (2020-03-28)[2020-12-15]. <https://arxiv.org/abs/2003.12783>.
- [15] Yu C Y, Xu Y, Gou L S, et al. Crowd counting based on single-column deep spatiotemporal convolutional neural network[J]. *Laser & Optoelectronics Progress*, 2021, 58(8): 0810011.
- 鱼春燕, 徐岩, 缙丽莎, 等. 基于单列深度时空卷积神经网络的人群计数[J]. *激光与光电子学进展*, 2021, 58(8): 0810011.
- [16] Shi Z L, Ye Y D, Wu Y P, et al. Crowd counting using rank-based spatial pyramid pooling network[J]. *Acta Automatica Sinica*, 2016, 42(6): 866-874.
- 时增林, 叶阳东, 吴云鹏, 等. 基于序的空间金字塔池化网络的人群计数方法[J]. *自动化学报*, 2016, 42(6): 866-874.
- [17] Zhang C, Li H S, Wang X G, et al. Cross-scene crowd counting via deep convolutional neural networks[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 833-841.
- [18] Li Y H, Zhang X F, Chen D M. CSRNet: dilated convolutional neural networks for understanding the highly congested scenes[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 1091-1100.
- [19] Zhu L, Zhao Z J, Lu C, et al. Dual path multi-scale fusion networks with attention for crowd counting[EB/OL]. (2019-02-04)[2020-12-10]. <https://arxiv.org/abs/1902.01115>.
- [20] Ma Z H, Wei X, Hong X P, et al. Bayesian loss for crowd count estimation with point supervision[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 6141-6150.
- [21] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2020-12-10]. <https://arxiv.org/abs/1409.1556>.
- [22] Meng Y B, Ji T, Liu G H, et al. Encoding-decoding multi-scale convolutional neural network for crowd counting[J]. *Journal of Xi'an Jiaotong University*, 2020, 54(5): 149-157.
- 孟月波, 纪拓, 刘光辉, 等. 编码-解码多尺度卷积神经网络人群计数方法[J]. *西安交通大学学报*, 2020, 54(5): 149-157.
- [23] Sindagi V A, Patel V M. Generating high-quality

crowd density maps using contextual pyramid CNNs[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 1879-1888.

[24] Liu J, Gao C Q, Meng D Y, et al. DecideNet:

counting varying density crowds through attention guided detection and density estimation [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 5197-5206.