

引用格式:

陈海燕, 甄霞军, 赵涛涛. 基于数据增强的高原鼠兔目标检测[J]. 湖南农业大学学报(自然科学版), 2022, 48(4): 496–500.

CHEN H Y, ZHEN X J, ZHAO T T. Target detection of *Ochotona curzoniae* based on the data augmentation[J]. Journal of Hunan Agricultural University(Natural Sciences), 2022, 48(4): 496–500.

投稿网址: <http://xb.hunau.edu.cn>



基于数据增强的高原鼠兔目标检测

陈海燕, 甄霞军, 赵涛涛

(兰州理工大学计算机与通信学院, 甘肃 兰州 730050)

摘 要: 针对基于卷积神经网络的高原鼠兔目标检测模型在实际应用中缺乏训练数据的问题, 提出一种前景与背景融合的数据增强方法: 首先对训练集数据进行前景和背景的分离, 对分离的前景作图像随机变换, 对分离的背景用背景像素随机覆盖, 得到前景集合和背景集合; 从前景集合和背景集合中随机选取前景和背景, 进行像素融合; 再从训练集中随机选取样本, 将标注边界框区域采用剪切粘贴方法融合到训练图像的随机位置, 得到增强数据集。采用两阶段的弱监督迁移学习训练模型, 第一阶段在增强数据集上对模型预训练; 第二阶段在原始训练集上微调预训练模型, 得到检测模型。对自然场景下高原鼠兔目标检测的结果表明: 在相同的试验条件下, 基于前景与背景融合数据增强的目标检测模型的平均精度优于未数据增强、Mosaic 和 CutOut 数据增强的目标检测模型; 基于前景、背景融合数据增强的目标检测模型的最优平均精度为 78.4%, 高于 Mosaic 的 72.60%、Cutout 的 75.86%和 Random Erasing 的 77.4%。

关 键 词: 高原鼠兔; 样本缺乏; 数据增强; 迁移学习; 样本平衡

中图分类号: TP319

文献标志码: A

文章编号: 1007-1032(2022)04-0496-05

Target detection of *Ochotona curzoniae* based on the data augmentation

CHEN Haiyan, ZHEN Xiajun, ZHAO Taotao

(School of Computer and Communication, Lanzhou University of Technology, Lanzhou Gansu 730050, China)

Abstract: Aiming at the problem that the *Ochotona curzoniae* target detection model based on convolutional neural network lacks training data in practical application, a data augmentation method is proposed by the fusion of foreground and background. Firstly, separate the foreground and the background of the training data, with image transforming the separated foreground randomly and covering the separated background by background pixels, to obtain the foreground set and the background set, respectively. The foreground and background are randomly selected from the foreground set and the background set, respectively and are fused based on pixel addition. Then randomly select a sample from the training set, and use the cut-and-paste method to fuse the labeled bounding box area of the selected sample to the training images' random positions to obtain an augmented data set. A two-stage weakly supervised transfer learning was used as the train the model. The first stage pre-trains the model dependent on the augmented data set. The second stage fine-tunes the pre-training model to obtain the detection model. Under the same experimental conditions, the experimental results of the target detection of *Ochotona curzoniae* in natural scenes show that the average accuracy of the target detection model based on this method is better than that of the target detection model without data augmentation, Mosaic, and Cutout data augmentation. The optimal AP of the target detection model based on data augmentation method by the fusion of foreground and background is 78.4%, which is higher than 72.6% of Mosaic method, 75.86% of Cutout method, and 77.4% of Random Erasing method.

收稿日期: 2021-05-07

修回日期: 2022-03-25

基金项目: 国家自然科学基金项目(62161019、62061024)

作者简介: 陈海燕(1978—), 女, 甘肃陇西人, 博士, 副教授, 主要从事图像处理研究, chenhaiyan@sina.com

Keywords: *Ochotona curzoniae*; lack of samples; data augmentation; transfer learning; sample balance

高原鼠兔目标检测是对其进行种群数量统计及研究种群动态变化的基础^[1-2]。基于卷积神经网络(CNN)的目标检测模型需要大量的训练数据^[3-10],而自然场景下的高原鼠兔多分布于高原山地,采集高原鼠兔图像困难,导致基于 CNN 的高原鼠兔目标检测模型缺乏训练数据^[11]。

在图像分类和目标检测领域,普遍认为模型的性能与训练样本数量的对数成正比^[12]。而由于隐私和安全性等因素的影响,使样本的获取难度大,导致模型缺乏训练数据,引起过拟合^[13]。针对缺乏训练数据导致的模型过拟合问题,通常采用早停止(early stopping)、正则化(regularization)、选择丢弃(dropout)和数据增强(data augmentation)等方法解决^[14]。数据增强的目的是对训练数据集扩充,从而降低模型的过拟合^[15-16]。图像变换是一种常用的数据增强方法,通常采用图像的仿射、扭曲和随机剪切等变换来扩大数据集^[15]。常见的基于图像变换的数据增强方法主要有 Cutout^[17]、Random Erasing^[18]、MixUp^[19]、Mosaic^[3]等。这些基于图像变换的方法生成的图像与原数据集图像有相同的语义信息,对提高训练数据的多样性和泛化能力有限^[20]。SHIN 等^[21]在研究少样本的船舶检测时,使用前景提取和粘贴的方法生成新样本,提高前景目标的位置分布模式,在扩大船舶数据集时提高数据集的多样性,虽然在特定数据集中表现出较好的性能,但并未考虑数据集中前景和背景类别不平衡的问题。当数据集中前景和背景类别不平衡时会导致基于 anchor 机制的目标检测模型在训练中存在正负 anchor 不平衡的问题。文献[20]指出,通过改变前景目标在图像中的位置,可以丰富前景目标的位置分布模式,提高数据集的多样性。文献[22]指出,增加图像中小目标的数量,可以增加与前景目标相交的先验框数量,平衡训练中的正负样本。

受到文献[21]和文献[22]中前景提取和粘贴增强方法的启发,针对模型缺乏训练数据的问题,笔者提出一种前景和背景融合的数据增强方法,记为 FBFAP。在此方法的基础上,结合两阶段的弱监督迁移学习训练模型,第 1 阶段在增强数据集上预训练模型,第 2 阶段在原始数据集上微调预训练模型,

得到检测模型,然后用最终的目标检测模型对高原鼠兔进行检测。

1 FBFAP 数据增强方法

令 I 为原始图像, F 为前景, B 为背景, M 为原始图像分割掩码。如式(1)所示,为了获得原始图像的前景和背景,将原始图像 I 的前景和背景分离,得到前景 F 和背景 B 。对前景 F 及其对应的分割掩码 M 分别进行式(2)和式(3)所示的 f 变换,得到含新前景 \bar{F} 的前景集合和新前景分割掩码 \bar{M} 的集合。对背景图像的前景区域用背景像素随机覆盖,得到含新背景 \bar{B} 的背景集合。从前景集合中随机选择 1 个前景,从背景集合中随机选择 1 个背景,使用基于像素加的方法融合,得到新图像 \bar{I}_1 。随机从标记训练集中选择样本,将标记边界框区域采用剪切粘贴的方法融合到 \bar{I}_1 的随机位置,得到新图像 \bar{I} 。如式(5)所示,其中 \otimes 表示基于像素的相乘, \oplus 表示基于像素的相加。

$$I \xrightarrow{\text{前景、背景分离}} \{F\} \{B\} \quad (1)$$

$$f: F \rightarrow \bar{F} \quad (2)$$

$$f: M \rightarrow \bar{M} \quad (3)$$

$$B \xrightarrow{\text{随机采样覆盖}} \{\bar{B}\} \quad (4)$$

$$\bar{B} \otimes \bar{M} \oplus \bar{F} = \bar{I} \quad (5)$$

式(2)中, f 变换是图像的平移变换。对于分离的背景图像,基于图像中距离相近区域像素模式通常相似的事实,在前景边界框的周围近邻采样一块同样大小的背景区域替换前景区域,获得连续变化的背景;由图像与其分割掩码乘获得前景,并对其进行 f 变换,得到新前景和新分割掩码。将前景与背景基于像素加的方法融合,从训练集中随机选取一个样本,将目标边界框区域粘贴到 \bar{I}_1 的随机位置,并记录位置坐标,获得新图像和对应的位置信息,得到增强数据集。前景、背景融合可以丰富目标位置分布模式,提高训练数据集样本的多样性。区域剪切和粘贴增加数据集中前景的数量,增加与前景相交的先验框数量,平衡训练中的正负样本。

研究^[22]表明,在前景和背景融合中,随着单张图像上融合的前景数量增多,模型的精度反而会下降,融合前景数量为 2 时模型达到最优。以此为依

据，将融合的前景目标数设置为 2。

2 高原鼠兔检测模型的建立及训练

以文献[9]中的 Faster R-CNN 为基础模型，结构如图 1 所示，采用两阶段的弱监督迁移学习方法进行训练。第 1 阶段，在 FBFAP 增强的数据集上对模型预训练，优化器为 Adam，初始学习率为 0.001，每 5 个 epoch，学习率衰减为原来的 0.1；beta1 和 beta2 分别是 0.9 和 0.999；训练 30 个 epoch，

BatchSize 为 4。第 2 阶段，在原始数据集上微调预训练的模型，参数设置与第 1 阶段一致。试验平台为搭载 NVIDIA Titan V 显卡的图形工作站，CUDA 和 CUDNN 版本分别是 10.1.168 和 7.6.1，操作系统为 Ubuntu LTS 16.04。模型实现框架为 Pytorch 1.1 和 Torchvision 0.3，使用 Python 3.5.2 编程。

为了评价 FBFAP 方法的性能，采用查准率(P)、召回率(R)、平均精度(AP)作为评价指标。

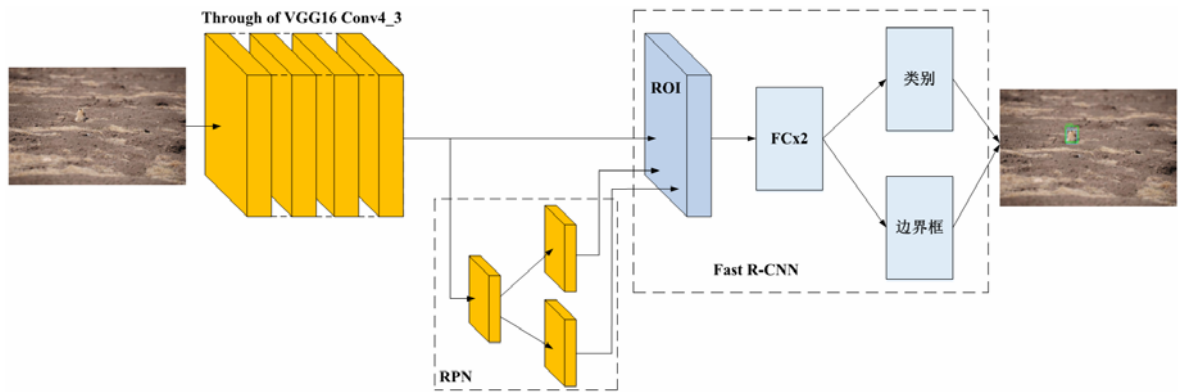


图 1 检测模型结构

Fig. 1 Structure of the detection model

3 高原鼠兔目标检测

3.1 数据采集

高原鼠兔数据为在青藏高原东北部(101°35'36"~102°58'15"E、33°58'21"~34°48'48"N)甘南草原采集的高原鼠兔图像，总共 1100 张，其中 900 张作为训练集，200 张作为测试集，分别记为 D-train 和 D-test。数据集格式与 Pascal VOC^[23]数据集格式一致。为了便于研究，将训练集随机分成大小为 100、200 和 600 的训练集，分别记为 DI-train、DII-train、DIII-train。使用数据增强方法得到的数据集记为增强数据集。

3.2 原始数据集上的检测结果

表 1 所示是在 DI-train、DII-train、DIII-train 和 D-train 训练集上训练的基于 Faster R-CNN 的目标检测模型的 P 、 R 和 AP 值。可以发现，训练数据集越大，模型的性能越好，进一步表明大数据集对深度卷积神经网络训练的必要性。由于 DI-train 和 DII-train 上的性能较差，因此在 DI-train 和 DII-train 上进行研究更具代表性，最后实现对 D-train 增强。

表 1 原始数据集上的查准率与召回率和平均精度

Table 1 The P , R , and AP on the original training sets			%
训练集	P	R	AP
DI-train	49.75	48.10	13.01
DII-train	49.48	68.10	37.80
DIII-train	49.60	90.00	76.71
D-train	49.61	91.43	81.98

3.3 对 DI-train、DII-train 和 D-train 数据增强后的检测结果

使用 FBFAP 方法分别对 DI-train 和 DII-train 增强，并使用两阶段的弱监督迁移学习训练基于 Faster R-CNN 的目标检测模型。表 2 所示为 DI-train 增强数据集大小对模型检测性能的影响。结果表明：FBFAP 方法能够有效提高模型的 R 和 AP；DII-train 增强数据集大小对模型检测性能的影响类似，且在 DI-train 和 DII-train 的增强数据集是 600 时取得最优 AP，分别为 66.74%和 78.40%。

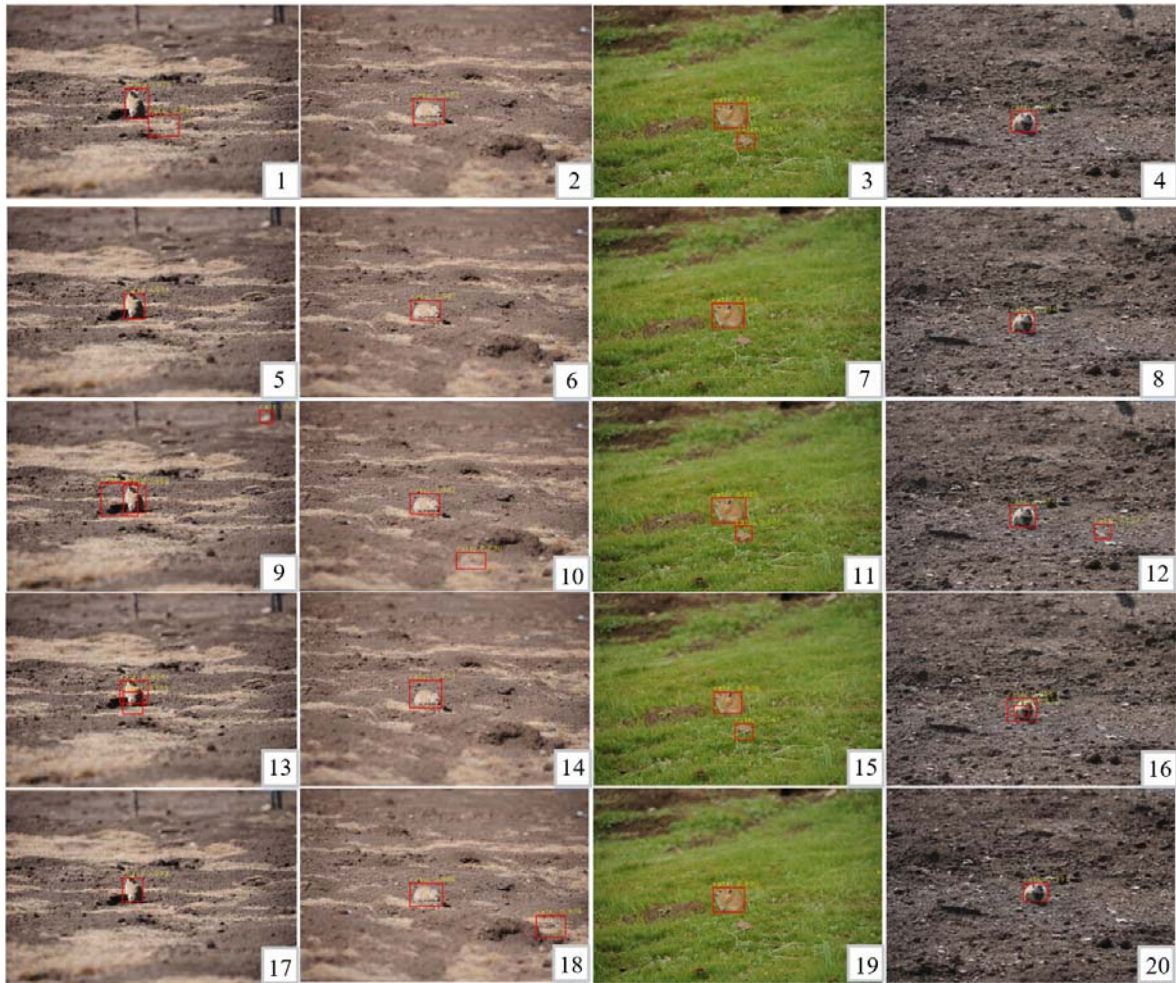
对 D-train 使用 FBFAP 数据增强，增强数据集大小为 600，此时的 P 值为 49.62， R 值为 93.81，AP 值为 83.34，可以发现 FBFAP 提高了模型的召回率和平均精度。

表 2 DI-train 增强数据的查准率与召回率和平均精度

Table 2 The P , R , and AP on the DI-train training sets with data augmentation %			
数据集大小	P	R	AP
100	49.58	56.67	14.50
200	49.75	47.62	19.73
300	49.70	77.62	58.60
400	49.55	78.57	61.86
500	49.57	82.86	64.25
600	49.57	81.43	66.74
700	49.58	83.81	66.35
800	49.57	81.43	66.32
900	49.72	83.33	66.15

为了进一步说明 FBFAP 方法的有效性，使用 Mosaic、Cutout 和 Random Erasing 方法分别对 DI-train 和 DII-train 增强，在相同的增强样本数量和相同的试验条件下训练检测模型，并将他们的检测结果与本检测结果相比较。

图 2 为不同数据增强方法对高原鼠兔目标检测的示例，红色矩形框表示目标检测的结果。可以看出，相对于未数据增强、Mosaic 数据增强、Cutout 数据增强、Random Erasing 数据增强的目标检测方法，基于 FBFAP 的高原鼠兔目标检测方法更准确。



1~4 为未数据增强方法；5~8 为 FBFAP 数据增强方法；9~12 为 Mosaic 数据增强方法；13~16 为 Cutout 数据增强方法；17~20 为 Random Erasing 数据增强方法。

图 2 不同方法的高原鼠兔检测效果

Fig.2 Target detection results test by different methods

参考文献：

[1] 陈海燕，陈刚琦. 基于语义分割的高原鼠兔目标检测[J]. 华中科技大学学报(自然科学版)，2020，48(7)：7~12.

[2] 陈海燕，陈刚琦，张华清. 基于 SegNet 模型的高原鼠兔的图像分割[J]. 湖南农业大学学报(自然科学版)，2020，46(6)：749~752.

[3] BOCHKOVSKIY A, WANG C Y, LIAO H Y M . YOLOv4: optimal speed and accuracy of object detection

- [EB/OL].[2021-03-29].<https://arxiv.org/pdf/2004.10934v1.pdf>.
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Columbus. New York: IEEE, 2014: 580-587.
- [5] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision. New York: IEEE, 2017: 2980-2988.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]//Proceedings of the 14th European Conference on Computer Vision. Berlin: Springer, 2016: 21-37.
- [7] SERMANET P, EIGEN D, ZHANG X et al. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks [EB/OL]. [2021-03-29]. <https://arxiv.org/pdf/1312.6229.pdf>.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 779-788.
- [9] REN S Q, HE K M, GIRSHICK R, et al. 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.
- [10] TIAN Z, SHEN C H, CHEN H, et al. FCOS: fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE, 2017: 9626-9635.
- [11] 张爱华, 王帆, 陈海燕. 基于改进 CV 模型的目标多色彩图像分割[J]. 华中科技大学学报(自然科学版), 2018, 46(1): 63-66.
- [12] SUN C, SHRIVASTAVA A, SINGH S, et al. Revisiting unreasonable effectiveness of data in deep learning era [C]//2017 IEEE International Conference on Computer Vision (ICCV). New York: IEEE, 2017: 843-852.
- [13] LU J, GONG P, YE J, et al. Learning from very few samples: a survey[EB/OL]. [2021-03-29]. <https://arxiv.org/pdf/2009.02653>.
- [14] WU Q F, CHEN Y P, MENG J. DCGAN-based data augmentation for tomato leaf disease identification[J]. IEEE Access, 2020, 8: 98716-98728.
- [15] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data 2019, 6: 60.
- [16] TAKAHASHI R, MATSUBARA T, UEHARA K. Data augmentation using random image cropping and patching for deep CNNs[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(9): 2917-2931.
- [17] DEVRES T, TAYLOR G W. Improved regularization of convolutional neural networks with cutout[EB/OL]. [2021-03-29]. <https://arxiv.org/pdf/1708.04552>.
- [18] ZHONG Z, ZHENG L, KANG G, et al. Random Erasing Data Augmentation [EB/OL]. [2021-03-29]. <https://arxiv.org/pdf/1708.04896>.
- [19] ZHANG H Y, MOUSTAPHA C, YANN N D, et al. MixUp: beyond empirical risk minimization[EB/OL]. [2021-03-29]. <https://arxiv.org/pdf/1710.09412>.
- [20] BANG S, BAEK F, PARK S, et al. Image augmentation to improve construction resource detection using generative adversarial networks, cut-and-paste, and image transformation techniques[J]. Automation in Construction, 2020, 115: 103198.
- [21] SHIN H C, LEE K I, LEE C E. Data augmentation method of object detection for deep learning in maritime image[C]//2020 IEEE International Conference on Big Data and Smart Computing(BigComp). New York: IEEE, 2020: 463-466.
- [22] KISANTAL M, WOJNA Z, MURAWSKI J, et al. Augmentation for small object detection[EB/OL]. [2021-03-29]. <https://arxiv.org/pdf/1902.07296>.
- [23] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (voc) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

责任编辑: 罗慧敏
英文编辑: 吴志立