

基于多尺度类注意力元学习的丝绸图案检测

葛亮, 王直杰

东华大学信息科学与技术学院, 上海

收稿日期: 2021年10月24日; 录用日期: 2021年11月22日; 发布日期: 2021年11月29日

摘要

中国是丝绸的发源地, 丝绸图案作为丝绸的主要元素承载了深远的历史文化。计算机视觉技术在丝绸图案的检测并不多见, 主要因为丝绸图案类别繁多, 总体数量不多, 一般的目标检测技术无法在丝绸这样的小样本数据集上发挥良好的效果。针对该问题, 本文提出了一种适用于小样本丝绸数据集的多尺度注意力目标检测算法, 通过多尺度类向量为目标检测网络提供更多的分类信息, 实现丝绸图案的检测任务。本文针对古代丝绸中常见图案进行数据收集和标定, 并且利用提出的网络实现对不同年代的狮子和花卉进行实验。实验结果表明该方法可以实现小样本丝绸图案目标检测任务, 而且效果优于其他小样本目标检测算法。

关键词

小样本学习, 多尺度, 类注意力, 目标检测

Silk Pattern Detection Based on Meta Learning with Multi-Scale Class Attention

Liang Ge, Zhijie Wang

College of Information Science and Technology, Donghua University, Shanghai

Received: Oct. 24th, 2021; accepted: Nov. 22nd, 2021; published: Nov. 29th, 2021

Abstract

China is the birthplace of silk. Silk patterns, which are the main element of silk, carry far-reaching history and culture. Computer vision technology is rare in the detection of silk patterns, mainly because of the wide variety and small number of silk patterns. The general object detection methods can not play a good effect on a small sample data set such as silk. To solve this problem, this paper proposes a multi-scale attention target detection algorithm suitable for small sample silk

data sets, which provides more classification information for the target detection network through multi-scale class vectors to realize the detection task of silk patterns. In this paper, data collection and data calibration are carried out for common patterns in ancient silk, and the proposed network is used to experiment lions and flowers in different ages. The experimental results show that this method can realize the target detection task of small sample silk pattern, and the effect is better than other small sample target detection algorithms.

Keywords

Few-Shot Learning, Multiple Scale, Class Attention, Object Detection

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

丝绸贸易开辟了古代中国与西域国家的贸易之路, 使得中国文化得以广泛传播。丝绸是丝绸图案的主要载体, 丝绸图案的风格与题材随年代发展不尽相同, 深刻的反映了当时的人文风俗。把目标检测技术用于丝绸研究具有十分深远的历史意义。

目前, 丝绸研究人员对丝绸图案的年代和类别的研究主要依赖人眼识别, 分类结果往往取决于个人经验。使得丝绸图案研究工作效率偏低, 研究结果往往富有主观性而且缺乏一致性。计算机视觉技术日益先进, 目标检测技术在文物识别方面有了广泛的应用, 但是目标检测在丝绸图案上的应用并不多见, 因此对于丝绸目标检测技术的研究十分必要。

主流的目标检测算法可以分为两种类型: 1) 基于候选框的两阶段目标检测算法[1] [2] [3]; 2) 基于 anchor 的单阶段目标检测算法[4] [5] [6]。并且经过各种改进这两种算法在 PASCAL VOC、MS COCO 等公共数据集上的效果不分伯仲。然而简单得将这两类深度学习算法应用到古代丝绸图案检测上并不能达到理想的效果, 其中最主要的原因在于, 丝绸数据种类多数量少的特点, 经过多轮迭代训练不能得出具有泛化性能模型。

针对小样本问题, 人们希望模型能具有人类的学习能力, 只需通过简单的学习就能分辨出不同的事物。学者们纷纷从各个角度提出不同的解决方案。基于微调的迁移学习方法[7], 研究认为直接在小样本空间上的模型缺少泛化性, 提出迁移学习策略。首先让模型在分布相似的大数据集上进行粗略的学习, 初始化目标域的相关参数, 然后固定住底层网络参数, 让具有初始化参数的模型在目标域上继续学习目标域独有的特征, 但是这种方法会忽略源域和目标域的重要目标知识, 降低模型的可传递性[8]。基于度量学习[9] [10] [11]的方法在孪生网络的基础上进行改进, 首先基于网络共享提取查询图像和目标图像的特征, 然后将该特征作为可学习度量网络的输入, 网络输出查询图像和目标图像相似度作为分类结果。但是基于度量学习的模型只适用于单个目标的输入, 缺乏普适性。基于元学习的目标检测算法[12]提出将元学习方法应用在传统的目标检测上, 通过元学习的策略学习一个具有普适性的目标检测模型, 该模型能快速从少量数据中学习数据的分布。Meta R-CNN [13] [14]等提出基于 Faster R-CNN 增加了类注意力机制, 类注意模块与 Faster R-CNN 共享特征提取层, 从 Support 集中提取类别信息作为 Faster R-CNN 分类器的指导信息, 提高模型的准确率。

中国古代丝绸上的图案以有序为美, 同样的图案会重复出现, 并且有大小变换。当 Support 集中的图

案尺寸与 Query 集中的尺寸有较大差别时, 基于元学习的目标检测算法效果有明显下降的趋势。针对该问题本文在 Meta R-CNN 基础上提出多尺度类注意力改进算法。在注意力网络中采用金字塔结构输出不同尺度的类注意力特征图, 再由原型网络针对每一个类别提取类向量, 分别与 Faster R-CNN 中不同尺度的特征图进行匹配, 尽可能找出各种尺度目标物候选框, 提高检测效果。

2. 相关方法

2.1. Faster R-CNN

Faster R-CNN [2]是典型的两阶段目标检测算法在准确率方面保持着一定的优势, 因此, 本文将 Faster R-CNN 框架作为小样本目标检测的底层框架。Faster R-CNN 的网络结构如图 1 所示。Faster R-CNN 主要包括: 1) 特征提取网络: 提取输入图像的特征的高维特征表示最为后续网络的特征输入。目前主流的用于特征提取的网络有 ResNet [15], MobileNet [16], VGG [17]等。2) RPN (Region Proposal Network)网络, 该网络将(1)中的特征图作为输入, 生成一系列不同尺寸的候选框 $\{z_{i,j}\}_{j=1}^{n_i}$, 其中每一个候选框都含有一个目标物或者背景。3) ROI (Region of Interest, ROI)层, RPN 网络生成的候选框因目标本身尺寸有所不同, 这不利于后续网络对候选框进行分类运算。ROI 层计算出每一个候选框的特征图, 然后经过池化操作统一特征图尺度, 这些特征图作为检测和分类网络的输入。4) 分类与回归网络, 该网络计算出每一个候选框的类别概率, 并利用边界框回归对候选框偏移量进行估计, 修正边界框的位置。

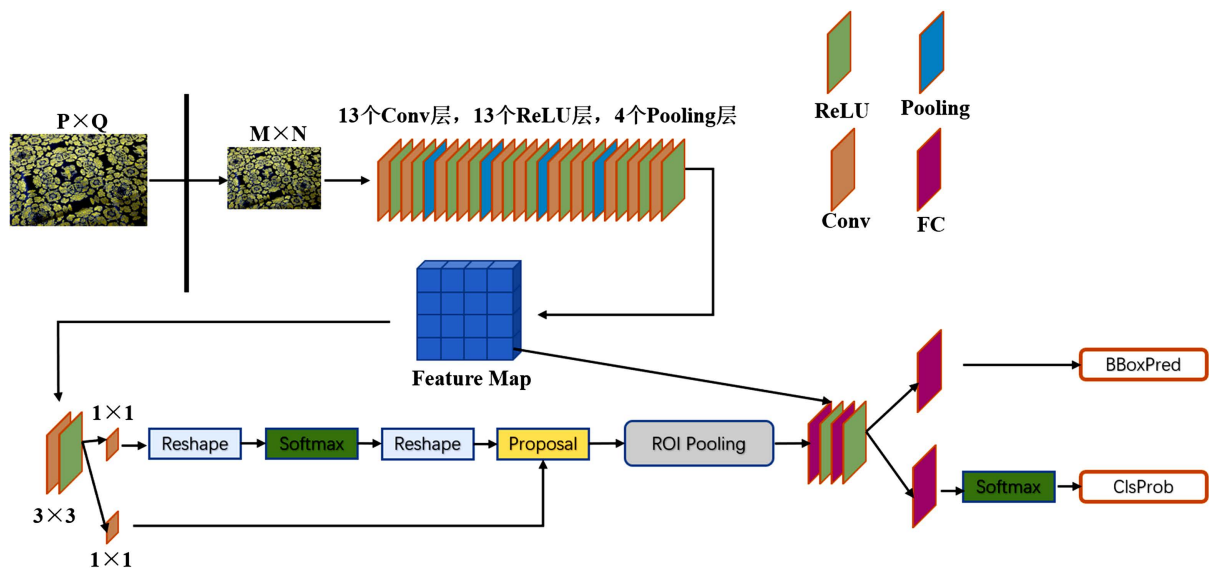


Figure 1. Framework of faster R-CNN model

图 1. Faster R-CNN 网络结构

尽管 Faster R-CNN 在目标检测方面有良好的表现能力, 但是 Faster R-CNN 模型训练非常依赖大量标定好的数据, Faster R-CNN 模型不能很好的解决小样本中的检测问题。面对训练数据不足的情况, 大多数基于深度学习的模型都会产生严重的过拟合现象。

2.2. 元学习

元学习也叫“学会学习”, 元学习以任务为学习单位, 通过学习不同任务的共性, 增强模型的泛化能力, 使模型在缺少训练数据的情况下依然具有较好的表现力[18]。元学习不是简单地学习一个从输入空

间到输出空间的映射模型, 而是希望模型通过对每一项历史任务的学习和经验积累, 从而面对新任务时可以更加容易, 所需要的训练样本更少, 同时还能保证一定的算法精度。以任务为学习单位的元学习模型能够从不同的任务中获得有用的经验, 模型更多的关注自身的学习能力, 然后将这种学习能力抽象成函数 $F(x)$ 若出现新的任务, 模型 $F(x)$ 利用极少量样本继续学习新任务中的信息, 从而快速适应和掌握新任务, 也就是抽象出对应任务 D_i 的函数模型 $f_i(x)$ 。Meta Learning 的结构如图 2 所示。现阶段, 元学习方法在模式识别中使用较为广泛, 但是将元学习用于目标检测的研究并不常见, 主要因为目标检测任务相对于模式识别, 图像分类等任务更为复杂, 目标检测背景更为复杂, 图像中多个目标于背景相互交叉重叠, 对模型的要求更高。

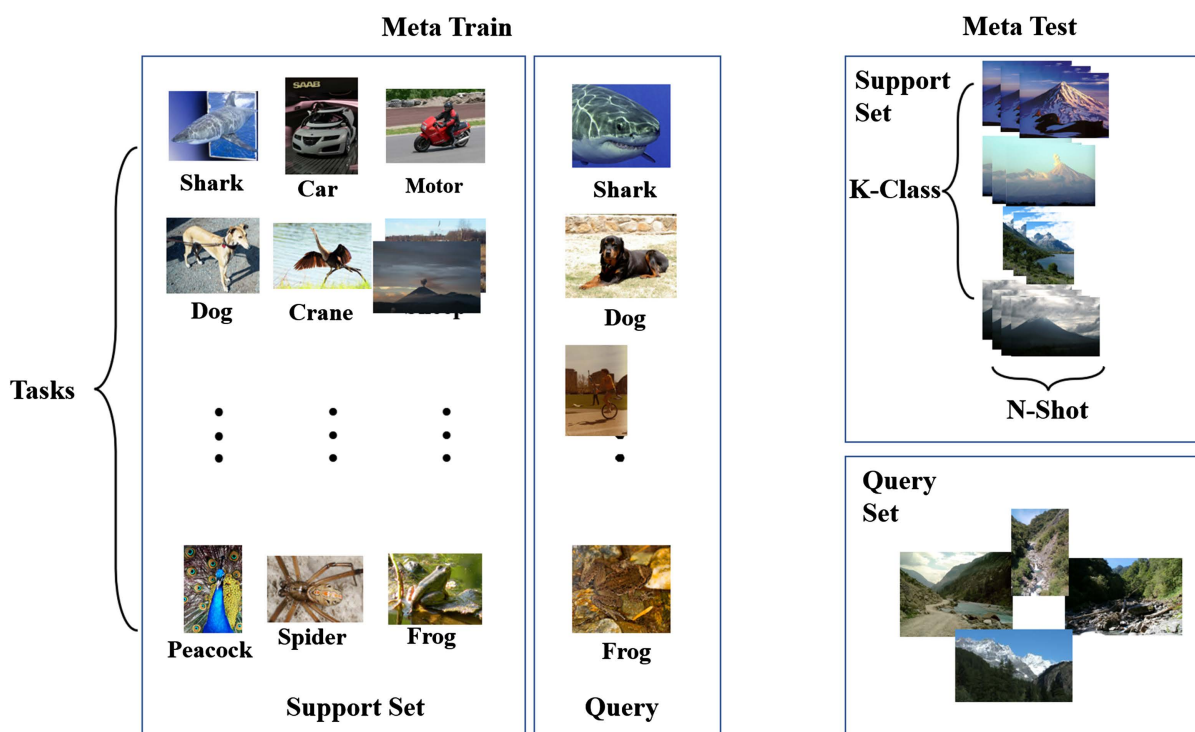


Figure 2. Framework of meta learning
图 2. 元学习结构示意图

2.3. Meta R-CNN

Meta R-CNN 将元学习策略应用在目标检测框架上, 结合 Faster R-CNN 框架, 实现小样本目标检测任务。Meta R-CNN 的目的是学习一个能快速适应小样本数据集的模型。Meta R-CNN 采用标准元学习的学习策略, 以多个小样本目标检测任务作为训练样本, 把快速对仅有少量新类样本的数据集实现检测任务作为训练目标。

对于 K -way N -shot 问题, 在元学习范式中, 每一个任务共有 K 个类别, 每个类别有 N 个样本, 元学习把这 $K \times N$ 样本作为支撑集 S^D , 其中支撑集中的样本都具有实例级标签, 然后针对其中部分类别随机抽取一个样本作为查询集 Q^D , 其中 $S^D \cap Q^D = \emptyset$ 。Meta R-CNN 利用原型网络从 S^D 学习每一个类别的先验知识 P_k 。在 Meta R-CNN 模型中原型网络和 Faster R-CNN 的特征提取网络实现参数共享。

$$P_k = \frac{1}{N_k} \sum_{i=1}^{N_k} f(x_i) \quad (1)$$

其中, P_k 表示 S^D 中第 k 个类别的原型特征, $f(x_i)$ 表示第 k 个类别中第 i 个样本的区域特征。查询集样本通过共享的特征网络生成查询样本的特征 $f(x^Q)$, 然后与支撑集中的每一个原型特征向量 P_k 进行卷积, 生成新的特征图 F_k 。

$$F_k = f(x^Q) \odot P_k \tag{2}$$

将 F_k 特征图作为 Faster R-CNN 网络的输入, F_k 特征图融合了类别信息, 所以 Faster R-CNN 网络在生成区域候选框的过程中更加关注具有类别特征的区域, 于原始 Faster R-CNN 网络相比, Meta R-CNN 中的候选框更有针对性, 大大降低区域候选框的数量。

但是, 在本文中的丝绸图案检测中, Meta R-CNN 网络的效果并不理想, 主要因为古代丝绸图案的表现形式, 同一种图案会以不同的尺寸重复出现, 当 Support 中的图案与 Query 中的图案尺寸相差较大时, 模型的识别能力明显下降。本文针对该问题提出了模型改进。

3. 方法改进

针对 Meta R-CNN 网络在小样本目标检测中对不同尺度目标检测效果不佳问题, 本文提出多尺度类注意力模块, 使用原型网络融合多种尺度的类别特征, 从而 Faster R-CNN 能获得更多的类别信息, 提高目标检测准确率。改进后的小样本目标检测框架如图 3 所示。小样本目标检测主要分为两个部分: 1) 原型网络, 本文利用原型网络完成多尺度类注意力提取。2) Faster R-CNN 网络, Faster R-CNN 网络对融合类别信息特征图实现目标检测。

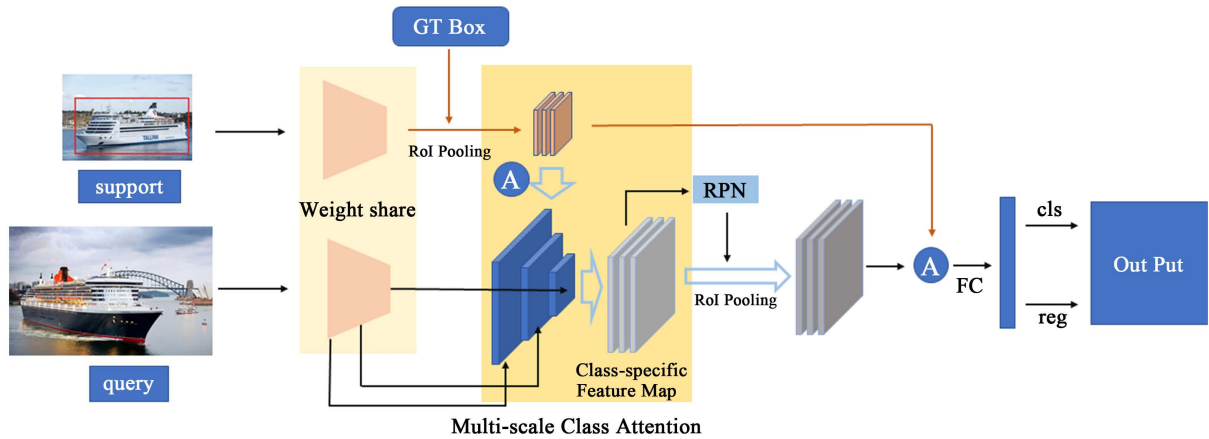


Figure 3. Framework of multi-scale class attention object detection

图 3. 多尺度类注意力目标检测结构

多尺度类注意力向量

本文对支持集样本的多个尺度进行实例级特征提取, 使原型网络提取的类注意力信息融合多个尺度的类别特征提高 Faster R-CNN 识别精度。考虑到 GPU 的算力问题, 本文使用三种尺度的特征提取。提取实例级注意力特征的网络结构如图 4 所示, 特征提取网络与 Faster R-CNN 的特征提取网络参数共享, 原型网络针对每个类别的多尺度特征生成对应的类注意力向量。特征提取的过程先自上而下特征图尺寸依次缩小 1/4, 尺寸越小, 特征图获取的高维信息就越丰富, 较大尺寸的特征图局部特征更丰富。然后做自低向上依次上采样, 考虑到位置信息损失, 上采样的过程中融合相同尺寸的左侧特征。最后将融合高维语义信息特征和浅层局部信息的类别特征作为原型网络的输入, 生成类别注意力特征。

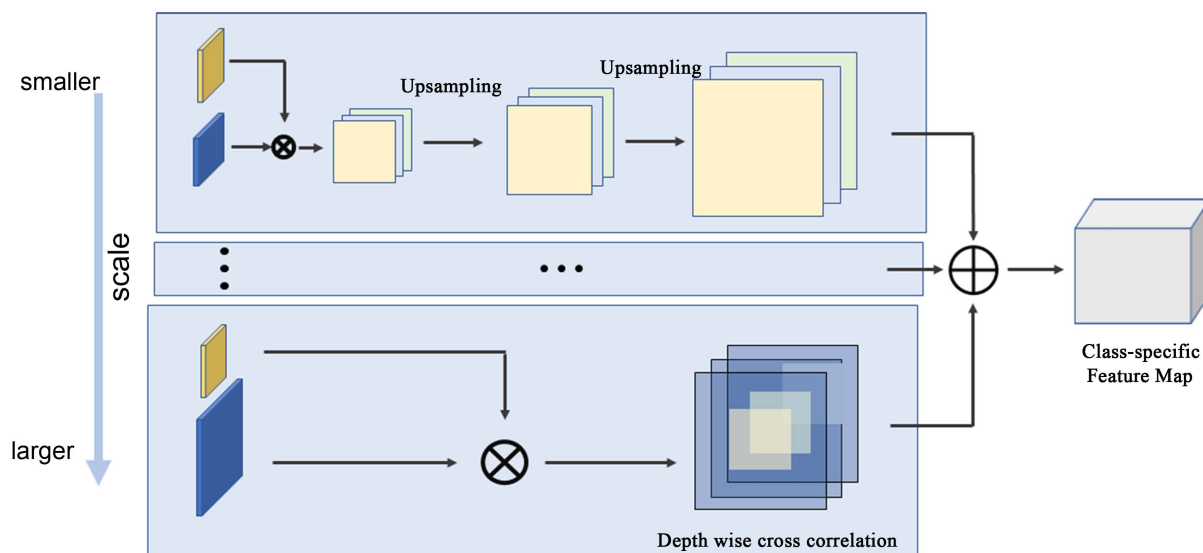


Figure 4. Module of multi-scale class attention

图 4. 多尺度类注意力生成模块

4. 实验及结果分析

4.1. 数据集

为验证本文提出算法的正确性和可靠性, 本文在融合东华大学丝绸数据集的 Pascal VOC2007 上进行 4-way 5-shot 和 4-way 20-shot 的实验。东华大学丝绸数据集由东华大学服装学院提供古代丝绸仿制织物, 东华大学信息科学技术学院进行图像级样本采集, 并且进行实例级样本标注, 该数据集包含 4 个类别, 共有 4000 多张彩色图片。示例图如图 5。本文中实验数据集共包含 24 个类别, 其中 20 类数据属于 Pascal VOC2007 数据集, 6 个类别数据东华大学丝绸数据集。实验数据划分情况见表 1。

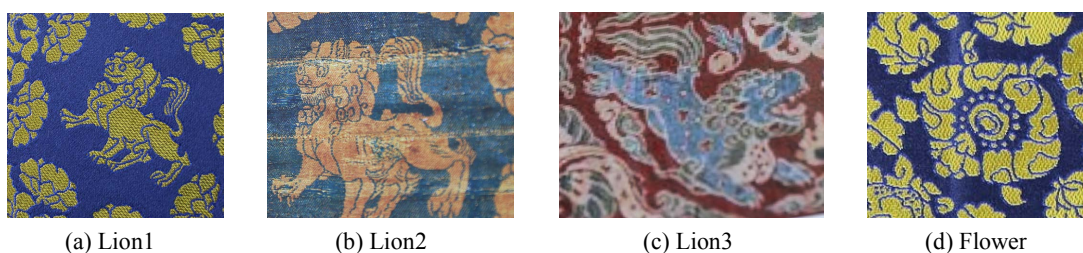


Figure 5. Examples of lions and flowers in silk dataset

图 5. 丝绸狮子和花卉图案示例

Table 1. Dataset summary

表 1. 试验数据集结构

	Meta Train	Meta Test
No. Class	22	4
No. Image	13200	3210
No. Box	25,242	7383
Avg No. Box/Img	1.91	2.3

4.2. 实验环境

实验环境配置如下: Intel Xeon W-2150B 处理器, 内存 32GB, Nvidia GeForce RTX 2080ti 显卡, Ubuntu18.04 操作系统。所有程序均基于开源 Pytorch 框架设计, 并使用 GPU 加速计算。在所有实验中, 使用初始学习率为的 Adam 算法进行训练, 每 3000 次迭代学习率减半。

4.3. 实验结果与分析

基于 4.1 节中的数据集, 本文所提方法与 YOLO-Few-shot [11]、Meta R-CNN [13] [14]等算法进行对比试验。实验采用 4-way 5-shot 和 4-way 20-shot 两种类型的对比。在 4-way 5-shot 实验过程中, 本文将采集到的丝绸数据集的三种不同年代的狮子和一种花卉作为元测试数据集, 最终希望模型能够在对比实验结果如表 2 所示。试验结果表明, 与其他小样本目标检测算法相比, 本文提出的方法在丝绸数据上优于其他算法, 在 PASCAL 中原有的类别的检测效果也有一定的准确率提升, 说明本文方法优化了 Meta R-CNN 模型的效果。

Table 2. Resulting data of experiment

表 2. 试验结果数据

Shot	Baselines	Novel classes					Base classes										mAP
		lion1	lion2	lion3	flower	mean	aero	bike	boat	car	chair	table	cat	sheep	plant	mean	
10	Few-shot [11]	19.60	21.10	17.60	26.70	21.25	73.60	73.10	56.70	76.10	66.80	77.70	42.11	57.40	74.70	66.47	52.55
	Meta R-CNN [13]	21.90	20.80	18.50	22.80	21.00	78.20	68.29	70.10	77.43	45.76	60.19	67.00	68.30	71.29	67.40	53.12
	Meta R-CNN [14]	19.80	19.80	16.20	26.10	20.48	54.10	78.54	56.98	73.12	56.34	76.32	56.87	49.43	69.00	63.41	50.20
	Ours	22.10	20.10	22.99	21.89	21.77	78.90	69.64	71.00	76.50	67.00	77.00	56.89	76.68	70.01	71.51	56.21
20	Few-shot [11]	30.12	35.67	27.87	22.69	29.09	75.87	76.42	67.32	80.12	56.00	69.88	65.00	79.09	76.37	71.79	58.65
	Meta R-CNN [13]	31.22	38.11	30.89	34.00	33.555	76.00	68.00	80.00	74.00	67.00	71.00	68.65	78.32	69.31	72.47	60.50
	Meta R-CNN [14]	32.77	29.99	37.00	33.33	33.2725	67.21	74.52	77.38	71.30	69.80	72.43	71.30	74.00	75.76	72.63	60.52
	Ours	35.00	38.67	31.67	38.00	35.835	77.00	75.67	76.21	72.65	71.87	73.20	72.65	69.33	73.23	73.53	61.93

5. 结论

东华大学服装学院的丝绸图案数据集检测问题, 目前为止, 没有人员提出关于丝绸图案的数据集, 所以基于一般 CNN 网络的目标检测算法在东华大学服装学院的丝绸图案检测中都会存在过拟合问题, 本文提出多尺度类注意力目标检测方法, 利用元学习的训练策略, 训练一个能快速学习小样本数据集的 Faster R-CNN 模型。并且基于原型网络提出多尺度类注意力模块, 多尺度类注意力融合多个尺度的特征信息, 为 Faster R-CNN 的候选框生成提供了更丰富的类别信息, 指导 RPN 网络生成更多与类别相关的候选框, 减少冗余信息。Faster R-CNN 在后续分类过程中也融合了原型类注意力特征, 使模型更关注同类别的信息, 提升识别率。

我们利用 Pascal VOC2007 作为训练集, 检验模型在丝绸图案数据集上的检测能力, 根据表 2 中的数

据集可知, 当模型在 Baseclass 中具有相当水平时, 本文提出的方法在新类的识别上效果最为明显, 说明本文提出的方法具有一定的效果。

参考文献

- [1] Girshick, R. (2015) Fast R-CNN. *Proceedings of the 2015 IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
- [2] Ren, S., He, K., Girshick, R. and Sun, J. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Proceedings of the 28th International Conference on Neural Information Processing Systems*, Vol. 1, Montreal, 7-12 December, 91-99.
- [3] He, K., Gkioxari, G., Dollár, P. and Girshick, R. (2017) Mask R-CNN. *Proceedings of the 2017 IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 2980-2988. <https://doi.org/10.1109/ICCV.2017.322>
- [4] Redmon, J., Divvala, S., Girshick, R., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [5] Shafiee, M.J., Chywl, B., Li, F. and Wong, A. (2017) Fast YOLO: A Fast You Only Look Once System for Real-Time Embedded Object Detection in Video. *Journal of Computational Vision and Imaging Systems*, 3. <https://doi.org/10.15353/vsnl.v3i1.171>
- [6] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016) SSD: Single Shot Multibox Detector. *European Conference on Computer Vision*, Amsterdam, 8-16 October 2016, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- [7] 潘兴甲, 张旭龙, 董未名, 姚寒星, 徐常胜. 小样本目标检测的研究现状[J]. 南京信息工程大学学报(自然科学版), 2019, 11(6): 698-705.
- [8] 李昊. 基于小样本的目标检测算法研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2020.
- [9] Zhang, T., Zhang, Y., Sun, X., Sun, H., Yan, M., Yang, X., et al. (2019) Comparison Network for One-Shot Conditional Object Detection. arXiv preprint arXiv: 1904.02317.
- [10] Karlinsky, L., Shtok, J., Harary, S., Schwartz, E., Aides, A., Feris, R., et al. (2019) Repmet: Representative-Based Metric Learning for Classification and Few-Shot Object Detection. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 5197-5206. <https://doi.org/10.1109/CVPR.2019.00534>
- [11] Kang, B., Liu, Z., Wang, X., Yu, F., Feng, J. and Darrell, T. (2019) Few-Shot Object Detection via Feature Reweighting. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 8419-8428. <https://doi.org/10.1109/ICCV.2019.00851>
- [12] Dong, X., Zheng, L., Ma, F., Yang, Y. and Meng, D. (2018) Few-Example Object Detection with Model Communication. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41, 1641-1654. <https://doi.org/10.1109/TPAMI.2018.2844853>
- [13] Yan, X., Chen, Z., Xu, A., Wang, X., Liang, X. and Lin, L. (2019) Meta R-CNN: Towards General Solver for Instance-Level Low-Shot Learning. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 9577-9586. <https://doi.org/10.1109/ICCV.2019.00967>
- [14] Wu, X., Sahoo, D., Hoi, S. (2020) Meta-RCNN: Meta Learning for Few-Shot Object Detection. *Proceedings of the 28th ACM International Conference on Multimedia*, Seattle, 12-16 October 2020, 1679-1687. <https://doi.org/10.1145/3394171.3413832>
- [15] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep residual Learning for Image Recognition. *Proceedings of the 2016 IEEE conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [16] Howard, A., Zhu, M, Chen B, Kalenichenko, D., Wang, W., Weyand, T., et al. (2017) MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv Preprint arXiv:1704.04861.
- [17] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv: 1409.1556.
- [18] 董博文, 汪荣贵, 杨娟, 薛丽霞. 结合多尺度特征与掩码图网络的小样本学习[J/OL]. 计算机工程与应用, 2021: 1-16. <http://kns.cnki.net/kcms/detail/11.2127.TP.20210416.1554.026.html>, 2021-10-28.