Computer Science and Application 计算机科学与应用, 2021, 11(5), 1579-1587
Published Online May 2021 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2021.115163

Hans汉斯

基于SSD改进的遥感图像目标检测算法

卢启祥

东北大学理学院，辽宁 沈阳
Email: 584669646@qq.com

收稿日期：2021年4月26日；录用日期：2021年5月21日；发布日期：2021年5月28日

摘　要

随着光学遥感技术的发展，遥感图像的目标检测方法也在逐步完善。SSD是一种单级目标检测模型，本文是基于SSD算法应用与遥感图像的目标检测任务并且针对遥感图像与正常图像的区别有针对性的改进。首先，目标检测分类模块和检测模块同时进行优化，由于检测模块内部简单样本和困难样本分布不均匀带来的梯度更新不均衡，导致检测模块收敛慢于分类，针对这个问题提出了一种新的损失函数可以缓解梯度更新不均衡，有效地在训练过程中加快模型的收敛并提高精度。同时，提出了Laplace-NMS方法，对于目标密集情况时后处理效果更好。本文提出的损失函数相对于SSD算法采用的提高了3.47%，同时本文提出的Laplace-NMS算法相对于NMS算法有0.78%的提升。

关键词

深度学习，目标检测，卷积神经网络，图像处理

An Improved Object Detection Algorithm Based on SSD in Remote Sensing Image

Qixiang Lu

College of Science, Northeastern University, Shenyang Liaoning
Email: 584669646@qq.com

Received: Apr. 26th, 2021; accepted: May 21st, 2021; published: May 28th, 2021

Abstract

With the development of optical remote sensing technology, the object detection method of remote sensing image is gradually improved. SSD is a single-stage target detection model. This paper is based on the application of SSD algorithm and the object detection task of remote sensing images, and aimed at the difference between remote sensing images and normal images to make

文章引用: 卢启祥. 基于 SSD 改进的遥感图像目标检测算法[J]. 计算机科学与应用, 2021, 11(5): 1579-1587.
DOI: 10.12677/csa.2021.115163

targeted improvements. Firstly, the classification module and the detection module of object detection are optimized simultaneously. Because of the uneven distribution of simple samples and difficult samples in the detection module, the gradient update is not balanced, which leads to the slower convergence of the detection module than the classification. In order to solve this problem, a new loss function is proposed, which can alleviate the imbalance of gradient updating and improve the accuracy. At the same time, the Laplace-NMS method is proposed, which has better post-processing effect when the target is dense. The loss function proposed in this paper is improved by 3.47% compared with the SSD algorithm, and the Laplace-NMS algorithm proposed in this paper is improved by 0.78% compared with the NMS algorithm.

Keywords

Deep Learning, Object Detection, Convolutional Neural Network, Image Processing

1. 引言

目标检测是光学遥感图像分析中的一个重要问题。同时，深度学习被应用到图像处理中，以海量的数据采集和现代 GPU 矩阵计算的发展取代了传统的方法。基于深度学习的目标检测任务可以应用于实际，在实时检测方面取得了很大进展。自从 R-CNN [1]第一次将深度学习应用于目标检测，目前为止已经发展了很多优秀的框架。其中代表性的二阶段方法 Fast-RCNN [2]对于 R-CNN 模型的最后一层卷积层后加了一个 ROI pooling layer，并且将损失函数改成多任务损失函数(multi-task loss)，将边框回归直接加入到 CNN 网络中训练。Faster-RCNN [3]使用 RPN (Region Proposal Network)代替原来的 Selective Search 方法产生候选窗口，产生候选窗口的 CNN 和目标检测的 CNN 参数共享。FPN [4]提出一种不同分辨率特征融合的方式使得不同层次的特征增强，模型性能显著提升。Cascade-RCNN [5]在二阶段的基础上引入级联几个检测网络达到不断优化预测结果的目的。D2Det [6]通过引入密集的局部回归来预测一个目标建议区域的多个密集盒的偏移量，达到精确定位的目的。

YOLO [7]是经典的非候选框的一阶目标检测方法，对于 Faster R-CNN，其先通过 CNN 得到候选框，然后再进行分类与回归。相比 YOLO，SSD [8]采用 CNN 来直接进行检测，并且提取了不同尺度的特征图来做检测，大尺度特征图可以用来检测小物体，而小尺度特征图用来检测大物体。Few-shot [9]利用少镜头支持集和查询集之间的相似性来识别新对象，减少了误识别。

由于遥感图像是高空拍摄采集，导致了目标的尺度非常小，对模型训练造成了很多问题，同时与传统的目标检测问题相比，遥感图像的目标检测过程中目标的密度非常高，针对以上问题，本文从损失函数和预选框后处理两个角度提出改进。对于 SSD 模型提出了二点改进；1) 为了平衡小目标和正常目标之间的梯度，本文提出了改进的 SmoothL1 损失函数来解决这个问题，并且模型可以收敛到更高的精度。2) 本文提出了 Laplace-NMS 来取代原始的 NMS [10]算法对预选框进行后处理，在不需要重新训练模型的基础上提高检测精度。

2. 相关工作

SSD 是由 Liu 等人 2015 年提出的一种快速高效的检测方法，使用全卷积网络去提取特征并在多个特

征层上预测不同尺度的目标。在生成预测框阶段，SSD 先将所有经过卷积得到的预测偏移信息和与之对应的先验框进行解码，计算出预测框。SSD 结构如图 1 所示。

SSD 是一个基于前馈的卷积网络，基础网络采用了 VGG [11]，它产生一个固定大小的 bounding box 集合，以及这些框中对象类实例存在的分数，然后是一个非最大抑制步骤来产生最终检测。每个特征图连接到最终检测层，这样可以使得网络检测和定位在不同的尺度图像中的对象。此外，这种尺度定位发生在向前传递过程中。不需要重新采样特性映射，从而使 SSD 能够以完整的前馈方式操作，因此 SSD 比较快速。

SSD 框架使用一种改进版的 anchor 算法来实现边界框建议。先验框是固定大小的边界框盒，其尺寸是根据数据集中每个类的地面真值包围盒的尺寸和位置预先计算的。因此称其为“先验”，因为依赖于贝叶斯统计推断，即先验概率分布。SSD 为每个特性映射预测每个类的得分，这些特性映射指定每个框中是否存在一个类实例。

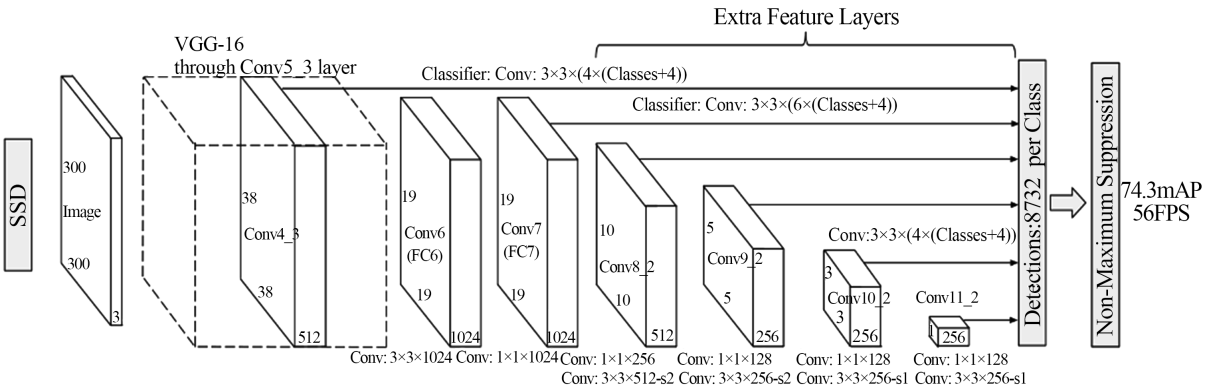

Figure 1. SSD model structure
图 1. 模型结构

损失函数定义为位置误差与置信度误差的加权和：置信度误差部分采用交叉熵损失函数，位置误差部分采用 SmoothL1 损失函数。检测网络预测出多个候选框后，通过非极大抑制算法(NMS)去除重复叠加的候选框，并且通过设置阈值除去置信得分较低的候选框。

SSD 算法损失函数定义为位置误差与置信度误差的加权和：置信度误差部分采用交叉熵损失函数，位置误差部分采用 SmoothL1 损失函数。总损失函数公式如下：

$$L(x,c,l,g)=\frac{1}{N}\left(L_{conf}(x,c)+\alpha L_{loc}(x,l,g)\right)$$

其中回归损失如下：

$$L_{loc}(x,l,g)=\sum_{i\in Pos}^{N}\sum_{m\in\{cs,cy,w,h\}}x_{ij}^{k}\text{smooth}_{\text{L1}}\left(l_i^m-g_j^m\right)$$

其中，l 为先验框的所对应边界框的位置预测值，而 g 是 ground truth 的位置参数。

SmoothL1 损失函数如下：

$$L(x)=\text{smooth}_{\text{L1}}\begin{cases}0.5x^2, & |x|<1\\ |x|, & \text{otherwise}\end{cases}$$

如下公式为分类损失函数：

$$L_{conf}(x,c) = -\sum_{i\in Pos}^{N} x_{ij}^{p}\log\left(c_i^p\right) - \sum_{i\in Neg}\log\left(x_i^0\right) \text{ where } c_i^p = \frac{\exp\left(c_i^p\right)}{\sum_p \exp\left(c_i^p\right)}$$

3. 改进方法

3.1. 改进的 SmoothL1

SSD 中对于候选框的损失函数如果使用的是 L2 损失函数即 MSE 时，随着误差的增大的过程中，有 MSE 损失函数的梯度与误差值成比例增大。当训练过程中误差很大时带来的梯度同样非常的，在卷积神经网络的训练过程中如果梯度过大会导致训练过程不稳定，更有可能产生梯度爆炸，由于人工标注的数据有大量的异常值，MSE 损失函数对于异常值特别敏感。这就导致了在训练过程中模型很难收敛，并且可能加大模型的训练难度。并且在训练初期，预测值与 groud truth 差异过于大时，损失函数对预测值的梯度十分大，使得训练不稳定。L1 对 x 误差的梯度为常数，L1 损失函数的优点在于对于异常和极端值不敏感，在训练初期使模型的梯度更新过程稳定。但是在训练的后期，提高模型精度的过程中，由于 L1 损失函数对于误差的不敏感的特点，预测值与 ground truth 差异很小时，L1 损失对预测值的梯度的绝对值仍然为 1，而 learning rate 如果不变，损失函数将在稳定值附近波动，难以继续收敛以达到更高精度。SmoothL1 在 x 较小时，对 x 的梯度也会变小，而在 x 很大时，对 x 的梯度的绝对值达到上限 1，也不会太大以至于破坏网络参数。因此本文针对回归问题中简单样本和困难样本的梯度和误差是线性关系的问题，提出了新的损失函数，可以通过调节系数 β 来控制简单样本和困难样本关于误差的梯度关系问题。公式如下：

$$L(x) = \begin{cases} \alpha\left(x*\arctan x\right) - \left(\beta/2\right)\ln\left(1+x^2\right), & |x|<1 \\ |x|+C, & \text{otherwise} \end{cases}$$

3.2. Laplace-NMS 算法

NMS 算法是目标检测算法中一个经典的后处理方法，NMS 算法流程如图 2 所示。

Algorithm 1 Non-Max Suppression

```
procedure NMS(B,c)
    B_nms ← ∅
    for b_i ∈ B do
        discard ← False
        for b_j ∈ B do
            if same(b_i, b_j) > λ_nms then
                if score(c, b_j) > score(c, b_i) then
                    discard ← True
        if not discard then
            B_nms ← B_nms ∪ b_i
    return B_nms
```

Figure 2. NMS algorithm steps
图 2. NMS 算法步骤

目标检测算法在训练完成以后，在测试阶段有两处需要计算矩形框的重叠度，第一处是计算先验矩形框和真实矩形框的重叠度，目的是根据重叠度确定先验框所属的类，包括背景类；第二处是计算预测

矩形框和真实矩形框的重叠度，目的是根据重叠度筛选最优的矩形框，使用 Iou 计算重叠度，交并比等于两个矩形框交集的面积与矩形框并集的面积之比。

因模型预测阶段预测大量的重叠的预测框，并且每个框分别带有分数所以我们需要采用算法进行筛选，其中 SSD 算法采用的为 NMS 算法其实现严格按照搜索局部极大值，抑制非极大值元素的思想。

NMS 是为了去除重复的预测框，算法在图片中只有单个物体被检测的情况下具有很好的效果，同时对于多个目标的时候，对于重叠物体无法很好的检测。当图像中存在多个重叠度很高的物体时，NMS 会过滤掉其中置信度较低的一个。当 overlap 阈值越大、proposals boxes 被压制的就越少，结果就是导致大量的 FP (False Positives)，进一步导致检测精度下降与丢失(原因在于对象与背景图像之间不平衡比率，导致 FP 增加数目远高于 TP)当 overlap 阈值很小的时候，导致 proposals boxes 被压制的很厉害，导致 recall 大幅下降。

传统的非极大值抑制算法首先在被检测图片中产生一系列的检测框 B 以及对应的分数 S。当选中最大分数的检测框 M 时，该框从集合 B 中移出并放入最终检测结果集合 D。于此同时，集合 B 中任何与检测框 M 的重叠部分大于重叠阈值的检测框也将随之移除。如果一个物体在另一个物体重叠区域出现，即当两个目标框接近时，分数更低的框就会因为与之重叠面积过大而被删掉，从而导致对该物体的检测失败并降低了算法的平均检测率，例如检测算法本来应该输出两个检测框，但是传统的非极大值抑制算由于绿框的得分较低且绿框和红框的 IOU 大于设定的阈值，因此会被过滤掉，导致只检测出一个物体，显然这样的算法设计是不合理的。NMS 直接粗暴的将和得分最大的 box 的 IOU 大于阈值的 box 的得分置零的方式，没有任何的缓和。

针对 NMS 算法无法对多类别并且重叠的物体无法很好的检测。非极最大抑制算法首先按照得分从高到低对建议框进行排序，然后分数最高的检测框被选中，其他框与被选中建议框有明显重叠的框被抑制。该过程被不断递归的应用于其余检测框。当图像中存在多个重叠度很高的物体时，NMS 算法会过滤掉其中置信度比较低的一个物品，但是因为不同物品的置信度分布不同，直接对比会产生误删的情况。针对以上问题，本文采用了 Laplace-NMS 算法。

Laplace-NMS 是用一个稍微小一点的分数替代原有的分数，而非直接粗暴的置零。Laplace-NMS 与 NMS 算法一样，同样采用贪心算法的策略，并且在原有的模型的基础上可以直接使用不需要重新训练模型，Laplace-NMS 算法吸取了 NMS 算法的整体历程，在置信度运算的时候进行了更改，在算法执行过程中不是简单的对 IoU 大于阈值的检测框删除，而是降低得分。使用的处理置信度的函数为本文提出加权。

本文将 NMS 算法中的加权系数改为：

$$s_i = \frac{1}{\gamma} \mathrm{e}^{-\frac{iou(\mathcal{M},b_i)}{\gamma}}, \forall b_i \notin D$$

通过衰减与检测框 M 有重叠的相邻检测框的检测分数是对 NMS 算法的改进。越是与 M 高度重叠的检测框，越有可能出现假阳性结果，它们的分数衰减应该更严重。与 NMS 算法一样 Soft-NMS 同样无法找到全局最优值，相对于 NMS 算法 Laplace-NMS 在有多种目标或者目标比较密集的情况会有一定的提升，同时在不需要重新训练模型的基础上可以提高检测效果，避免了资源的浪费。图 3 为 Laplace-NMS 算法与 NMS 算法对比的流程图。

因为 NMS 直接将和得分最大的 box 的 IOU 大于某个阈值的 box 的得分置零，Laplace-NMS 用稍低一点的分数来代替原有的分数，而不是直接置零，Laplace-NMS 仅需要对传统的 NMS 算法进行简单的改动且不增额外的参数。Laplace-NMS 具有与传统 NMS 相同的算法复杂度，使用高效。同时 Laplace-NMS 不需要额外的训练，并易于实现，它可以轻松的被集成到任何物体检测流程中。

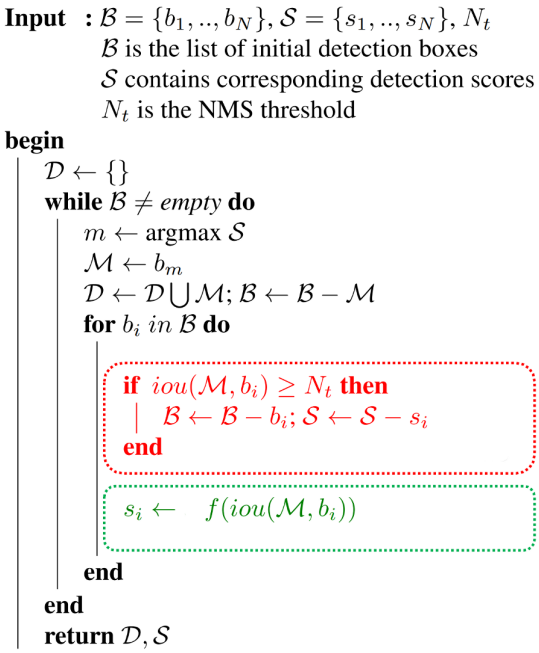

Figure 3. Algorithm flow comparison
图 3. 算法流程对比

4. 实验与分析

在这一节中，我们将评估和比较本文中提出的方法的有效性。具体而言，本实验使用的是 NWPU VHR-10 [12]航空遥感影像数据集，该数据集包含从谷歌和 Vaihingen 数据集剪切并由专家手动标注的高分辨率遥感影像。遥感影像与常规自然影像的主要区别在于，首先，遥感影像的目标尺寸很小，常常聚集在一起。经过一定次数的采样后，目标在预测的特征层中只有大约 1 个像素，因此大小太小无法区分。其次，对于航空遥感影像，背景复杂，复杂的背景会对预测产生很大的影响。

在训练过程中，图像的输入大小是 512 × 512，基础网络采用经过 ImageNet [13]数据集预训练的 VGG 模型，采用 SSD 原本的数据增强方式，并且考虑到航拍图像的特殊性加入了随机角度翻转的数据增强方式。优化方法采用随机梯度下降方法，最小批量为 32，并设置动量为 0.9，权重衰减为 0.0005。对整个训练集迭代 15,000 次。下面实验都采取同样的数据增强和参数确保对照实验的准确性。

表 1 为 SSD 与使用改进后损失函数的 SSD 对比，SSD 原始检测结果为 0.8845，使用改进后的损失函数的 SSD 效果为 0.9192，效果提升了 3.47%。

Table 1. Comparison of loss function results
表 1. 损失函数结果对比

方法	SSD	SSD + AdjustLoss
mAP	0.8845	0.9192

图 4 为 SSD + adsmoothl1 的结果图，图 5 为 SSD 算法的结果图，图中对每个类别的 AP 值进行了展示。通过实验结果可知 ship 的 AP 值从 0.6803 提升到 0.6956，storage 的 AP 值从 0.9074 提升到 0.9282 对于这种小目标本文提出的损失函数对实验结果有明显的提升，特别是 vehicle 从 0.7254 提升到 0.8066 效果提升明显，可以发现越小的目标提升效果越明显。

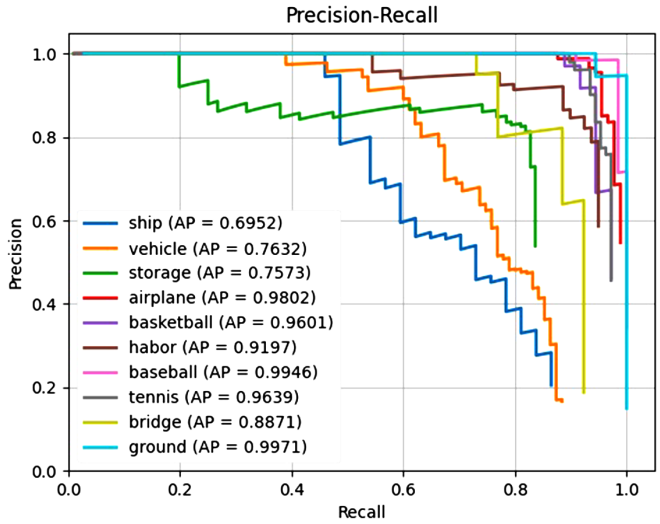

Figure 4. The result of SSD + AdSmoothL1
图 4. SSD + AdSmoothL1 的结果图

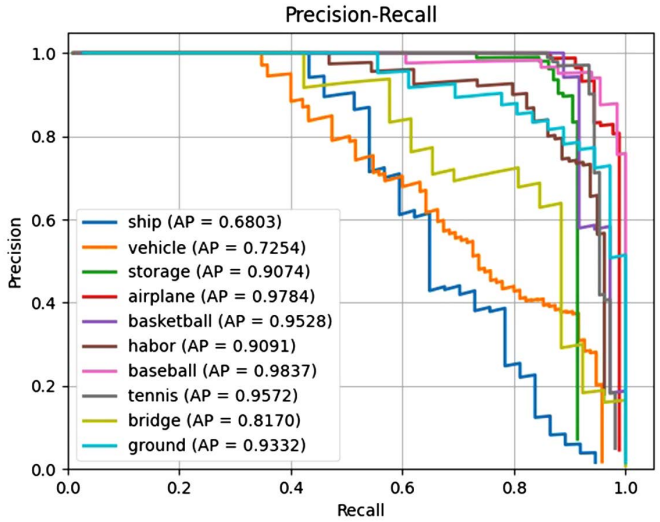

Figure 5. The result of SSD
图 5. SSD 结果图

表 2 为使用 NMS 方法去掉边界框和使用本文提出的 Laplace-NMS 方法的实验对比，效果提升了 0.78%。

Table 2. Comparison of non-maximum suppression methods
表 2. 非极大值抑制方法对比

方法	SSD + AdjustLoss + NMS	SSD + AdjustLoss + Laplace-NMS
mAP	0.9192	0.9270

图 6 为 SSD 使用本文提出的 AdjustLoss 损失函数和 Laplace-NMS 算法结果图，Laplace-NMS 主要是对于重叠度高的目标有一定提升，对于 airplane 这种特别密集且重叠度比较高的目标 AP 值从 0.9784 提升到 0.9978，提升了 1.94%，证明了 Laplace-NMS 方法的有效性。

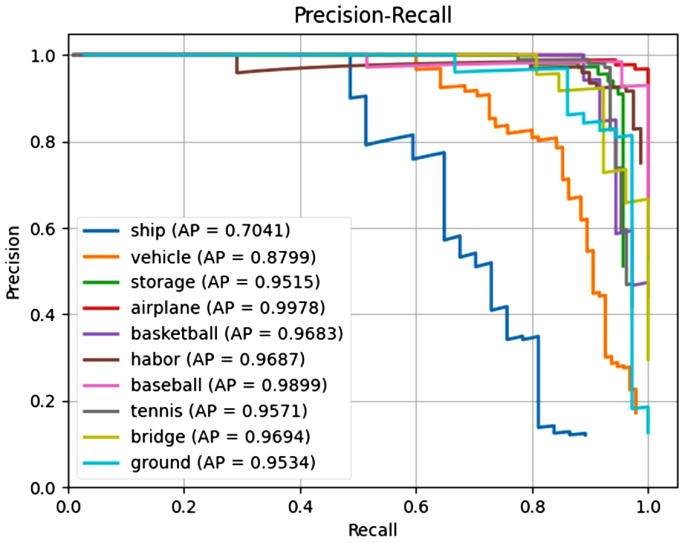

Figure 6. The result of SSD + AdjustLoss + Laplace-NMS
图 6. SSD + AdjustLoss + Laplace-NMS 结果图

5. 结论

本文提出了一种用于遥感图像中的小目标检测的非极大值抑制方法和损失函数。在分析 SmoothL1 损失函数的优缺点的基础上，本文提出了可调节的 SmoothL1 损失函数，本文提出的损失函数可以通过参数调节梯度，能够增加损失函数的灵活性，通过实验证明了本损失函数对于解决小目标难易样本训练过程中有显著作用。通过采用本文提出的 Laplace-NMS 算法相对于 NMS 算法在不改变模型的前提下有一定的提升。

参考文献

[1] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, 24-27 June 2014, 580-587. https://doi.org/10.1109/CVPR.2014.81

[2] Girshick, R. (2015) Fast R-CNN. arXiv:1504.08083 [cs.CV] https://doi.org/10.1109/ICCV.2015.169

[3] Ren, S., He, K., Girshick, R., *et al.* (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

[4] Lin, T.Y., Dollár, P., Girshick, R., *et al.* (2016) Feature Pyramid Networks for Object Detection. arXiv:1612.03144 [cs.CV] https://doi.org/10.1109/CVPR.2017.106

[5] Cai, Z. and Vasconcelos, N. (2017) Cascade R-CNN: Delving into High Quality Object Detection. arXiv:1712.00726 [cs.CV] https://doi.org/10.1109/CVPR.2018.00644

[6] Cao, J., Cholakkal, H., Anwer, R.M., *et al.* (2020) D2det: Towards High Quality Object Detection and Instance Segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16 November 2020, 11485-11494. https://doi.org/10.1109/CVPR42600.2020.01150

[7] Redmon, J., Divvala, S., Girshick, R., *et al.* (2015) You Only Look Once: Unified, Real-Time Object Detection. arXiv:1506.02640 [cs.CV] https://doi.org/10.1109/CVPR.2016.91

[8] Liu, W., Anguelov, D., Erhan, D., *et al.* (2016) SSD: Single Shot MultiBox Detector. arXiv:1512.02325 [cs.CV] https://doi.org/10.1007/978-3-319-46448-0_2

[9] Fan, Q., Zhuo, W., Tang, C.K., *et al.* (2020) Few-Shot Object Detection with Attention-RPN and Multi-Relation De-

tector. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 4013-4022. https://doi.org/10.1109/CVPR42600.2020.00407

[10] Neubeck, A. and Gool, L.J.V. (2006) Efficient Non-Maximum Suppression. 18*th International Conference on Pattern Recognition* (*ICPR*'06), Hong Kong, 20-24 August 2006, 850-855. https://doi.org/10.1109/ICPR.2006.479

[11] Simonyan, K. and Zisserman, A. (2015) Very Deep Convolutional Networks for Large Scale Image Recognition. arXiv:1409.1556 [cs.CV]

[12] Cheng, G., Han, J., Zhou, P. and Guo, L. (2014) Multi-Class Geospatial Object Detection and Geographic Image Classification Based on Collection of Part Detectors. *ISPRS Journal ofPhotogrammetry and Remote Sensing*, **98**, 119-132. https://doi.org/10.1016/j.isprsjprs.2014.10.002

[13] Russakovsky, O., Deng, J., Su, H., *et al.* (2015) ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, **115**, 211-252. https://doi.org/10.1007/s11263-015-0816-y