Journal of Image and Signal Processing 图像与信号处理, 2019, 8(2), 43-50
Published Online April 2019 in Hans. http://www.hanspub.org/journal/jisp
https://doi.org/10.12677/jisp.2019.82007

Hans汉斯

Target Detection and Recognition Based on Improved Faster R-CNN

Jingjing Fang, Jinyong Cheng

School of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Ji'nan Shandong
Email: 412707453@qq.com, cjy@qlu.edu.cn

Received: Mar. 6th, 2019; accepted: Mar. 17th, 2019; published: Mar. 28th, 2019

Abstract

In recent years, with the continuous development of in-depth learning, image research and application based on in-depth learning has achieved excellent results in many fields. RCNN network and full convolution network make the development of target detection technology more and more rapid. Faster R-CNN algorithm has been proposed and widely used in the field of target detection and recognition. In this paper, we mainly study the object detection based on Faster R-CNN algorithm for the image in the data set of self-made office supplies. Compared with RCNN series algorithms, Faster R-CNN proposes a regional recommendation network, and integrates feature extraction, candidate box extraction, boundary box regression and classification into a network, which greatly improves the overall performance. In this paper, an improved Faster R-CNN algorithm based on activation function is proposed. When extracting features, the data set usually has a large number of high-density continuity characteristics, while the activation function is sparse, which solves the problem of target detection of office supplies under small targets and complex background, and improves the detection speed and accuracy.

Keywords

Deep Learning, Object Detections, Region Proposal Network, Feature Extraction

基于改进的Faster R-CNN的目标检测与识别

房靖晶，成金勇

齐鲁工业大学(山东省科学院)，计算机科学与技术学院，山东 济南
Email: 412707453@qq.com, cjy@qlu.edu.cn

收稿日期：2019年3月6日；录用日期：2019年3月17日；发布日期：2019年3月28日

文章引用: 房靖晶, 成金勇. 基于改进的 Faster R-CNN 的目标检测与识别[J]. 图像与信号处理, 2019, 8(2): 43-50.
DOI: 10.12677/jisp.2019.82007

摘　要

近年来，随着深度学习不断的发展，基于深度学习的图像研究与应用已经在很多领域取得了优异的成绩。RCNN网络与全卷积网络等技术框架使得目标检测技术发展越来越迅速。Faster R-CNN算法被提出并广泛应用于目标检测和目标识别领域。在本文中，主要研究了基于Faster R-CNN算法对自制办公用品数据集中的图像进行的目标检测。相较于RCNN系列算法，Faster R-CNN提出了区域建议网络，同时将特征抽取、候选框提取、边界框回归、分类整合到一个网络当中，使得综合性能有很大改进。本文提出基于AlexNet改进的Faster R-CNN算法，在提取特征时，数据集通常具有大量高密度的连续性特征，而激活函数具有稀疏性，解决了目标小且背景复杂情况下的办公用品目标检测问题，提高了检测速度和检测精度。

关键词

深度学习，目标检测，区域建议网络，特征提取

1. 引言

随着深度学习[1]相关技术的快速发展，目标检测[2]与目标识别在生活中的多个领域中有着非常广泛的应用，并且都取得了相当好的效果。目标检测是找出图像或者视频中所有感兴趣的目标，通过判断该区域内是否存在目标来确定目标位置，再进行目标种类识别[3]。是机器视觉领域的核心问题之一。在目标检测中，准确率和时间都是检测方法的衡量标准[4]，所以本文中对办公用品进行识别的好坏与提高人们的工作效率有很大的影响。本文通过减少窗口数量来提高运算效率。由于不同的物体有不同的外观或者形状，再加上光线、背景等因素的干扰，目标检测一直是机器视觉领域最具有挑战性的问题[5]。因此，目标检测的核心问题是：目标有各种形状、不同大小、任何位置。目前主流的目标检测解决思路是通过深度学习算法，进行端到端的训练，即输入图像到输出任务结果一步完成[6]。目标检测的过程是图像－特征提取－分类、回归[7]。

Fast R-CNN 基本实现端到端的检测[8]，但是在选择性搜索(Selective Search，简称 SS)算法[9] [10]提取候选框时需要耗费大量的时间，针对该问题 Faster R-CNN 算法中提出了区域建议网络(Region Proposal Network，简称 RPN) [11]的概念，这个 RPN 网络是利用神经网络自己学习来产生候选区域[12]。在处理办公用品数据集时因为图像背景复杂特征提取不准确，本文在基础的 Faster R-CNN 算法上使用 ReLU 和 Leaky ReLU 激活函数，这个方法很大程度地提高了生成候选区域的可靠程度和目标检测的准确度，并且有效地缩短了预测时间。

2. 基于 Faster R-CNN 算法目标检测与识别

2.1. Faster R-CNN 算法

Faster R-CNN 作为一种 CNN 网络目标检测算法，首先使用卷积层提取输入图像的特征图[13]，该特征图被共享用于 RPN 网络和全连接层[14]。随后用 RPN 网络生成区域建议，通过 softmax 分类器判断候

选区域属于前景还是背景，再利用边界框回归[15]修正候选区域的位置，获得精准的检测框。在 Faster R-CNN 算法中感兴趣区域(Regions of Interest，简称 ROI)池化层收集输入的特征图和区域建议，综合这些信息后对区域建议提取特征图，然后送入全连接层判定目标类别。最后利用区域建议特征图计算区域建议的类别，同时再次使用边界框回归和非极大值抑制[16]获得检测框的精确位置，如图 1 所示。

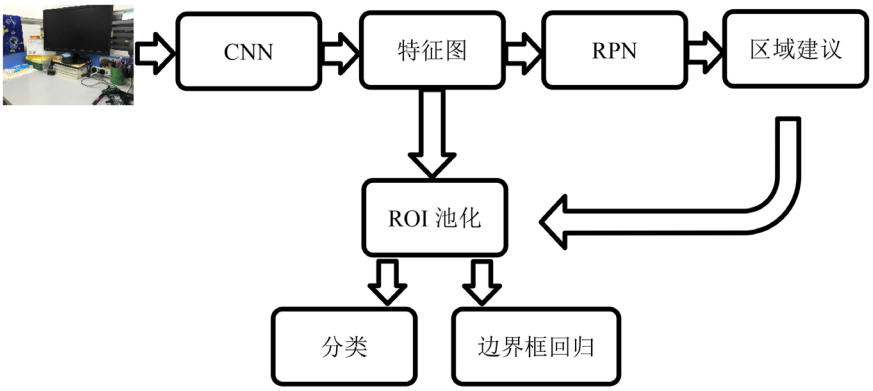

Figure 1. Faster R-CNN algorithm flow chart
图 1. Faster R-CNN 算法流程图

2.2. 区域建议网络

经典的检测方法生成候选框都十分的耗时，如使用滑动窗口法和图像金字塔生成候选框；或者 RCNN 算法使用选择性搜索(Selective Search)方法生成候选框。而 Faster RCNN 算法则改变了使用经典的滑动窗法口和选择性搜索方法，改为使用区域建议网络来生成检测框，这也是 Faster RCNN 的一大优势和特点，候选框的产生速度得到极大地提高。

区域建议网络是一个全卷积网络[17]，它的核心思想是使用 CNN 卷积神经网络直接产生区域建议，使用的方法本质上就是滑动窗口在最后的卷积层上滑动一遍，由于候选区域机制和边框回归可以得到多尺度多长宽比的区域建议。

区域建议网络可以针对生成区域建议的任务进行端到端地训练，同时能够预测出目标的边界框和分数。

区域建议网络的输入可以是任意大小尺寸的图片。RPN 网络将每个特征图的位置编码成一个特征向量；对每一个位置输出一个目标得分和边框回归，换而言之 RPN 网络在每个卷积映射位置上输出该位置的多尺度长宽比候选区域的目标得分和边框回归。

2.3. 非极大值抑制

非极大值抑制算法(Non-Maximum Suppression，简称 NMS)是搜索局部的极大值，同时抑制非极大值元素的过程。目标检测的过程中滑动窗口经过提取特征和分类器分类识别后[18]，每个窗口都会得到一个目标得分并且在同一目标的位置上会产生大量的候选框，这些候选框相互之间可能会有重叠，此时我们需要利用非极大值抑制找到效果最佳的目标边界框，消除冗余的边界框。非极大值抑制的过程是：首先根据置信度的得分对候选框进行排序，选择置信度最高的候选框先输出，将置信度最高的候选框从边界框列表中删除，计算所有边界框的面积，计算置信度得分最高的候选框与其它候选框的重叠度，删除重叠度大于规定阈值的候选框；然后重复之前的步骤，直到边界框列表为空。

IOU 定义了两个候选框的重叠度，设矩形框 $T1$、$T2$，矩形框 $T1$、$T2$ 重叠度的计算公式为：

$\mathrm{IOU}=(T1\cap T2)/(T1\cup T2)$，IOU 就是矩形框 $T1$、$T2$ 的重叠面积占 $T1$、$T2$ 并集面基的比例。

NMS 在广泛应用在计算机视觉领域，例如目标跟踪、目标识别、数据挖掘以及纹理分析等。

3. 相关工作

3.1. AlexNet 模型与激活函数

AlexNet 模型是 2012 年由 Alex Krizhevsky 提出，该模型采用了 8 层神经网络，5 个连接层和 3 个全连接层。AlexNet 使用 ReLU 函数作为 CNN 的激活函数，解决了在深度网络中 Sigmoid 函数造成的梯度弥散问题。

本文使用了 ReLU 函数，它会将一部分神经元的输出为 0，使网络具有稀疏性，并且减少了参数间相互依存的关系，有效地缓解了过拟合问题的发生。同时本文在网络中加入了 Leaky ReLU 函数，解决了 Relu 函数进入负区间后，导致神经元不学习的问题。激活函数的另一个重要特征是：它是可以区分的[19]。有助于在网络中向后推进计算相对于权重的误差梯度时执行反向优化的策略[20]，然后相应地使用梯度下降或者其他优化技术来优化权重以减少误差。

ReLU 激活函数在反向传播求误差梯度时计算量相对较小，节省很多。ReLU 是从底部开始半修正的一种函数，$\mathrm{relu}(x)=\max(0,x)$。当输入 $x<0$ 时，输出为 0，当 $x>0$ 时，输出为 x。ReLU 激活函数能够更加快速的使网络收敛。它不会饱和，即它可以对抗梯度消失问题，至少在正区域 $x>0$ 可以这样，因此神经元至少在一半区域中不会把所有零进行反向传播。Leaky ReLU 函数是经典的 ReLu 激活函数的变体，该函数的输出对于负值输入有很小的坡度。Leaky ReLU 函数的导数总是不为零，$\mathrm{leakyrelu}(x)=\max(0.01x,x)$。因此能够减少静默神经元的出现，允许基于梯度的学习。

3.2. 改进的 Faster R-CNN

在特征提取中，由于大量有效信息的损失，造成特征提取不准确，训练速度慢，本文提出了基于 AlexNet 改进的 Faster R-CNN 目标检测算法，在原始的 CNN 框架中使用 ReLU 和 Leaky ReLU 激活函数，克服梯度消失的问题，加快训练速度。CNN 模型如图 2 所示。

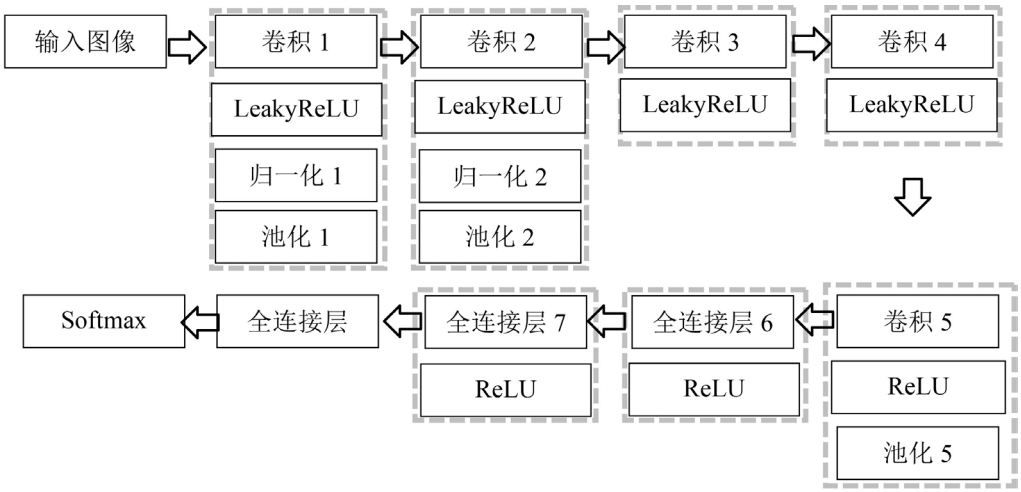

Figure 2. CNN model diagram
图 2. CNN 模型图

在局部响应归一化训练过程中，需要根据正确区域与候选区域的重叠度来训练样本，重叠度是衡量

目标检测准确度一个简单的衡量标准，在训练时需要得出一个预测范围，这个衡量标准跟目标的正确区域与预测的目标候选区域密切相关的，重叠度越高，则相关度越高，反之，相关度越低，计算候选区域与正确区域的重叠度，将重叠度最大或者重叠度大于阈值的候选区域标记为正样本；将与正确区域重叠度小于阈值的候选区域标记为负样本。

4. 实验和分析

主要介绍目标检测使用的数据集以及本文的实验过程和数据分析。

4.1. 实验环境及数据集

本实验中用到了汽车数据集和自制办公用品数据集。汽车数据集包含了 295 张汽车图像；自制办公用品数据集包含 30 类物品，每类 50 张，共 1500 张图片，图片来自网络搜集和日常拍摄。实验中，选取每类数据集的 60%作为训练集，40%作为测试集检测结果最佳。

4.2. 实验过程

CNN 网络结构设置，实验中我们采用 3 * 3 的卷积核进行卷积，采用步幅为 2 的 3 * 3 的空间池区域对数据维度进行下采样操作，在全连接层中添加一个非线性的 ReLU 层。本实验用 25 层的 CNN 网络来进行特征提取。

训练数据，每一个小批量数据包含从一张图像中随机提取的 256 个候选区域，其中前景样本和背景样本各取 128 个，正负比例达到 1:1。如果一个图像中的正样本数小于 128，则多用一些负样本以满足有 256 个候选区域可以用于训练。权重参数设置，前面几层参数是由自制办公用品数据集和汽车数据集分别训练本文提出的 CNN 模型来进行初始化，新增的两层为权重均设置为满足 0 均值，标准差为 0.01 的高斯分布来进行初始化。学习率设置为 0.001，mini-batch 为 1 来进行实验。

在实验中遵循多任务损失定义，将目标函数最小化，FasterR-CNN 中对一个图像的函数定义为：

$$f(m)=\frac{1}{N_c}\sum_m f_c\left(p_m,p_m^*\right)+\lambda\frac{1}{Nr}\sum_m p_m^* f_r\left(t_m,t_m^*\right)，\text{其中 } f(m)=L\left(\{p_m\},\{t_m\}\right) \tag{1}$$

在这里，m 是候选区域的索引，p_m 是候选区域 m 中是目标物体的预测概率。$t_m=\{t_x,t_y,t_w,t_h\}$ 是一个向量，表示预测的候选框的四个坐标参数；t_m^* 是正样本对应正确区域候选框的坐标向量；$L_c\left(p_m,p_m^*\right)$ 是两个类别的对数损失；

$$L_c\left(p_m,p_m^*\right)=-\log\left[p_m p_m^*+(1-p_m)\left(1-p_m^*\right)\right] \tag{2}$$

$L_r\left(t_m,t_m^*\right)$ 是回归损失，用 $L_r\left(t_m,t_m^*\right)=S\left(t_m,t_m^*\right)$ 来计算，S 是 smooth L1 函数。

Fast R-CNN 网络有分类得分层和候选框预测层，这两个同级输出层都是全连接层。分类得分层用于分类，输出 $c+1$ 维数组 p，表示属于 c 类和背景的概率。对每个 ROI 输出离散型概率分布：$p=\left(p_0,p_1,\cdots,p_c\right)$ 其中，p 由 $c+1$ 类的全连接层利用 softmax 分类器计算得出。

候选框预测层用于候选区域的调整，输出边界框回归的位移，输出 $4\times c$ 维数组 t，表示分别属于 c 类时，应该平移缩放的参数。

$$t^c=\left(t_x^c,t_y^c,t_w^c,t_h^c\right) \tag{3}$$

c 表示类别的索引，t_x^c,t_y^c 是指相对于区域建议尺度不变的平移，t_w^c,t_h^c 是指对数空间中相对于区域建议的高与宽。

Faster R-CNN 模型分四步训练，前两步训练区域建议和检测网络，用于 Fast R-CNN 网络，后两步将前两步中的网络结合，创建单个网络进行目标检测。每个步骤有不同的收敛速度，有利于为每个步骤指定独立的训练项。表 1 为 Epoch 为 10 的 RPN 网络训练过程的一部分数据。

Table 1. RPN training process
表 1. RPN 网络训练过程

Epoch	Iteration	Mini-batch loss	Mini-batch Accuracy	Base Learning Rate
1	1	1.5318	21.09%	0.0010
2	50	1.5326	100.00%	0.0010
3	100	1.6181	100.00%	0.0010
4	150	0.8708	96.88%	0.0010
5	200	1.0628	100.00%	0.0010
6	250	1.1254	100.00%	0.0010
7	300	0.8630	100.00%	0.0010
8	350	1.9528	75.10%	0.0010
9	400	0.9821	100.00%	0.0010
10	450	0.8431	100.00%	0.0010

4.3. 实验结果及分析

实验为了验证 Faster R-CNN 算法的可适应性，对自制办公用品数据集和汽车数据集分别进行训练与测试。同时，为了解决 Faster R-CNN 算法在自制办公用品数据集中针对目标较小、背景复杂的办公用品图像，提出了基于激活函数改进的 Faster R-CNN 算法，通过增加卷积层，使用 ReLU 激活函数和 Leaky ReLU 激活函数改进了特征提取的 CNN 模型，在特征提取过程中减少了有效信息的丢失，提高了办公用品数据集检测结果。

从表 2 可以看出，当 epoch 为 2 时 Faster R-CNN 算法在汽车数据集上的检测结果为 77.21%，在自制办公用品数据集上的检测结果为 62.34%；当 epoch 为 5 时，Faster R-CNN 算法在汽车数据集上的检测结果为 78.19%，在自制办公用品数据集上的检测结果为 67.07%；当 epoch 为 8 时，Faster R-CNN 算法在汽车数据集上的检测结果为 88.09%，在自制办公用品数据集上的检测结果为 69.36%；当 epoch 为 10 时，Faster R-CNN 算法在汽车数据集上的检测结果为 92.74%，在自制办公用品数据集上的检测结果为 74.62%。说明当 epoch 为 10 时，检测结果最好。

Table 2. Experimental results of Faster R-CNN in different data sets
表 2. Faster R-CNN 在不同 epoch 下的实验结果

Epoch	2	5	8	10
汽车数据集	77.21%	78.19%	91.60%	92.74%
自制办公用品数据集	62.34%	67.07%	69.36%	74.62%

如表 3 所示，当 epoch = 2 时，本文提出的基于 AlexNet 改进的 Faster R-CNN 在汽车数据集上的检测结果为 78.12%，提高了 0.91 个百分点，在自制办公用品数据集上的检测结果为 62.51%，提高了 0.17 个百分点；当 epoch = 5 时，本文提出算法在汽车数据集上的检测结果为 82.12%，提高了 3.93 个百分点，在自制办公用品数据集上的检测结果为 68.27%，提高了 1.2 个百分点；当 epoch = 8 时，本文提出算法在汽车数据集上的检测结果为 92.74%，提高了 1.14 个百分点，在自制办公用品数据集上的检测结果为

69.62%，提高了 0.26 个百分点；当 epoch = 10 时，本文提出算法在汽车数据集上的检测结果为 94.53%，提高了 1.79 个百分点，在自制办公用品数据集上的检测结果为 76.34%，提高了 1.72 个百分点。由于自制办公用品数据集每类图像数量少于汽车数据集图像数量，因此检测结果较汽车数据集来说有差距，但改进后的算法提高了自制办公用品数据集的检测结果，说明该方法有很好的的迁移性。从本质上来说，卷积网络是一种由输入到输出的映射的过程，它能够学习大量的输入与输出之间的映射关系，用已有的模式训练卷积网络，就可以使该网络具有输入输出之间的映射能力。同时卷积层的中每个神经元连接数据窗的权重是固定的，每个神经元只关注一个特性。由于原始数据具有稠密特性，其包含的信息远多于局部特征点，因此使用激活函数可以完成深层网络的训练，可以更好的提高学习精度，更好更快的提取稀疏特征。本文提出的目标检测算法适用于不同数据集，提高了检测精度，节省了预测时间。

Table 3. Experimental results of Faster R-CNN based on AlexNet in different datasets
表 3. 基于 AlexNet 改进的 Faster R-CNN 在不同 epoch 下的实验结果

Epoch	2	5	8	10
汽车数据集	78.12%	82.12%	92.74%	94.53%
自制办公用品数据集	62.51%	68.27%	69.62%	76.34%

5. 总结

本文提出使用 ReLU 激活函数和 LeakyReLU 激活函数来改进特征提取模型，在卷积层增加 LeakyReLU 激活函数构建 CNN 框架。在实验中通过池化层将特征降维，去处图像处理中的冗余信息，提取重要特征，在一定程度上防止了过拟合。该框架共享卷积核，LeakyReLu 函数对负值数据进行加权优化，既修正数据分布又保留部分负值，解决了当由于负值数据造成权重不再更新引起的梯度死亡问题，有效地处理高维数据，提高了识别精度。在汽车数据集中，基于激活函数改进的 Faster R-CNN 算法检测结果达到 94.53%，在办公用品数据集中达到 76.34%。证明改进的卷积神经网络框架是有效的可行的。但是由于图像中目标较多的问题，我们无法得知正确的输出量，增加了模型的复杂程度，结果提升不是很明显，这是我们下一步要研究的重点。

基金项目

本研究获得山东省自然科学基金(23170807, ZR2017LB024, ZR2018LF004)项目资助。

参考文献

[1] Chavali, N., Agrawal, H., Mahendru, A., *et al*. (2016) Object-Proposal Evaluation Protocol Is “Gameable”. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Sunday, 26 June-1 July 2016, 835-844.

[2] Xie, S., Girshick, R., Dollár, P., *et al*. (2017) Aggregated Residual Transformations for Deep Neural Networks. *Conference on Computer Vision and Pattern Recognition*, 21-26 July 2017, 5987-5995.

[3] Dai, J., He, K. and Sun, J. (2015) Convolutional Feature Masking for Joint Object and Stuff Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 3992-4000.

[4] 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

[5] Hosang, J., Benenson, R., Dollár, P., *et al*. (2016) What Makes for Effective Detection Proposals? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**, 814-830. https://doi.org/10.1109/TPAMI.2015.2465908

[6] Pinheiro, P.O., Collobert, R. and Dollár, P. (2015) Learning to Segment Object Candidates. *Advances in Neural Information Processing Systems*, Montreal, 7-12 December 2015, 1990-1998.

[7] Liu, W., Anguelov, D., Erhan, D., *et al*. (2016) Ssd: Single Shot Multibox Detector. In: *European Conference on*

Computer Vision, Springer, Cham, 21-37.

[8] Girshick, R. (2015) Fast r-cnn. *Proceedings of the IEEE International Conference on Computer Vision*, Araucano Park, 11-18 December 2015, 1440-1448.

[9] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 1.

[10] Uijlings, J.R.R., Van De Sande, K.E.A., Gevers, T., *et al*. (2013) Selective Search for Object Recognition. *International Journal of Computer Vision*, **104**, 154-171. https://doi.org/10.1007/s11263-013-0620-5

[11] Hariharan, B., Arbeláez, P., Girshick, R., *et al*. (2014) Simultaneous Detection and Segmentation. In: *European Conference on Computer Vision*, Springer, Cham, 297-312.

[12] Erhan, D., Szegedy, C., Toshev, A., *et al*. (2014) Scalable Object Detection Using Deep Neural Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 2147-2154.

[13] Szegedy, C., Ioffe, S., Vanhoucke, V., *et al*. (2017) Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning.

[14] Wang, X., Girshick, R., Gupta, A., *et al*. (2018) Non-Local Neural Networks. *The IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-22 June 2018, Vol. 1, 4.

[15] Wei, S.E., Ramakrishna, V., Kanade, T., *et al*. (2016) Convolutional Pose Machines. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 4724-4732.

[16] Neubeck, A. and Van Gool, L. (2006) Efficient Non-Maximum Suppression. 18*th International Conference on Pattern Recognition*, Hong Kong, 20-24 August 2006, Vol. 3, 850-855.

[17] Girshic, R., Donahue, J., Darrell, T., *et al*. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587.

[18] Redmon, J., Divvala, S., Girshick, R., *et al*. (2016) You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 779-788.

[19] Zeiler, M.D. and Fergus, R. (2014) Visualizing and Understanding Convolutional Networks. In: *European Conference on Computer Vision*, Springer, Cham, 818-833.

[20] Gulcehre, C., Moczulski, M., Denil, M., *et al.* (2016) Noisy Activation Functions. *International Conference on Machine Learning*, **48**, 3059-3068.