Computer Science and Application 计算机科学与应用, 2020, 10(12), 2339-2345
Published Online December 2020 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2020.1012247

Hans汉斯

一种基于深度学习的课堂学生学习状态研究

刘秋会，梁明秀，王　林

贵州民族大学数据科学与信息工程学院，贵州　贵阳
Email: 1753879994@qq.com

收稿日期：2020年11月27日；录用日期：2020年12月21日；发布日期：2020年12月28日

摘　要

为了对学生课堂学习情况及教师授课情况进行客观评价，需要掌握学生在课堂上的学习状态，随着计算机视觉技术的发展，对学生课堂学生状态的分析成为可能。本文采用深度学习网络yolov3与dropblock结合对教室监控视频进行分析，检测学生在老师上课时的听课状态，实现对学生在课堂上是否专心听讲的学习状态检测。实验结果表明，通过建议方法得到的学生学习状态与实际人工观察具有很好的吻合度。

关键词

深度学习，课堂学习状态检测，Yolov3，Dropblock

A Study on the Learning State of Classroom Students Based on Deep Learning

Qiuhui Liu, Mingxiu Liang, Lin Wang

School of Data Science and Information Engineering, Guizhou Minzu University, Guiyang Guizhou
Email: 1753879994@qq.com

Received: Nov. 27th, 2020; accepted: Dec. 21st, 2020; published: Dec. 28th, 2020

Abstract

In order to objectively evaluate students' classroom learning and teachers' teaching, it is necessary to master students' learning status in class. With the development of computer vision technology, it is possible to analyze students' classroom learning status. In this paper, the deep learning network yolov3 is combined with dropblock to analyze classroom surveillance video, detect the state of students listening to teachers in class, and realize the learning state detection of whether students are paying attention in class. The experimental results show that the students' learning status obtained by the proposed method is in good agreement with the actual artificial observation.

文章引用: 刘秋会, 梁明秀, 王林. 一种基于深度学习的课堂学生学习状态研究[J]. 计算机科学与应用, 2020, 10(12): 2339-2345. DOI: 10.12677/csa.2020.1012247

Keywords

Deep Learning, Classroom Learning Status Detection, Yolov3, Dropblock

1. 引言

在教室课堂场景下的学生学习状态检测是指在课堂图片中检测出头部状态的过程，状态分为抬头与低头。学生作为受教育的主体，从学生课堂学习状态的研究出发，可作为学生课堂学习效率的评价指标之一。目前教师普遍通过课堂观察以及提问等方式来实时了解学生的课堂学习情况，这会造成课堂信息传递与反馈的滞后性和片面性[1]。并且，随着智能手机等电子设备的普及，目前的课堂教学过程中，出现了大批“低头族”。因此，通过统计“抬头率”可以在一定程度上判断学生的课堂专注度，从而有助于有效地提升课堂教学效率[2]。学生的课堂学习状态与课堂学习效率息息相关。目前课堂场景中对于学生课堂学习状态的研究主要从课堂学习行为[3]、课堂疲劳状态[4]、课堂人脸检测及关注度研究方面[1]展开。

在计算机视觉和模式识别中，近几年深度学习网络得到了广泛的应用，比如，2012 年，以 AlexNet [5]为代表的卷积神经网络(CNN)方法被广泛应用在目标检测领域，精度取得了显著提升。2014 年，GoogLeNet [6]的面世，在保持预算不变的情况下增加网络的深度与宽度，从而实现大规模图片的目标检测。2015 年，ResNet [7]深度残差网络的提出，解决了深度网络过深而浪费现有资源的问题，Fast R-CNN [8]是一种快速基于区域的卷积网络方法，在提高训练和测试速度的同时，提高了检测精度。2016 年，ResNeXt [9]通过重复一个构建块来构建，聚合了一组具有相同拓扑结构的转换，在保持复杂的限制条件下，增加基数也能很好地提高分类精度。而 SSD [10]及 YOLO [11]的提出，让目标检测算法从两个阶段向一个阶段(端到端)迈进。2017 年，YOLOV2 [12]不仅提高了检测速度，而且检测类别高达 9000 种，在数据集 PASCAL VOC 和 COCO 上，是最先进的多尺度训练方法。2018 年，YOLOV3 [13]的提出，不仅简化了网络模型，更是对小目标检测起到了领航的作用。

深度学习虽然应用广泛且效果较好，但是对于不同的研究对象有不同的研究方法，本文针对于教室课堂场景，研究学生学习状态，以头部为目标，把头部状态分为抬头和低头，我们借鉴了 tinyyolov3 与 dropblock 的结合[14]，把 yolov3 与 dropblock 相结合，实验证明它对于网络卷积层是特别有效的正则化方法。

2. 基本原理

2.1. Yolov3

yolov3 [13]提出了一个阶段克服了两种操作缓慢的缺点阶段检测算法。它是一种卷积神经实现端到端的目标检测和识别。它只使用一个 CNN 网络直接预测不同目标的类别和位置节省了大量的时间来检测对象。在这项工作中，我们选择 yolov3 模型提取头部特性。yolo 算法的基本思想是：首先通过特征提取网络对输入特征提取特征，得到特定大小的特征图输出。输入图像为 460 × 460，会分成 13 × 13、26 × 26、52 × 52 的网格，接着如果真实框中某个物体的中心坐标落在某个网格中，那么就由该网格来预测该物体。每个物体有固定数量的边界框，Yolov3 中有三个边界框，使用逻辑回归确定用来预测的回归框。图 2 是 dropblock 与 yolov3 的网络结构。

2.2. Dropblock

过拟合在计算机视觉领域普遍存在，模型在已有的训练集上表现比较好，而在新的未知数据集上表现较差。对于这一现象，在深度神经网络中首次提出了 dropout [15]算法。dropout 一般放在全连接层后。在卷积层中添加 dropout 没有明显的效果。由于卷积层可以通过 drop 掉的神经元附近学习到相似的信息，因此为了在卷积层中防止过拟合现象，出现了 dropblock [16]模块。dropblock [16]是一种用于卷积层的正则化方法。这两种算法的主要区别在于 dropout 随机灭活全连接层的神经元，而 dropblock [16]随机灭活卷积层的单元。在实验中，我们在 yolov3 模型中加入了 dropblock [16]模块，从而获得更好的模型。

dropblock 层以块的形式丢弃特征单元，减少网络对某一特征的依赖。block_size 和 γ 是 dropblock 的两个重要参数。block_size 表示要丢弃的块的大小，而 γ 控制的是要删除活动单元格的数量。block_size 的大小对于所有的特征图都是一样的，不管特征图的分辨率如何。实际上，γ 没有确定的值，但可以按如下方式进行计算

$$\gamma = \frac{1-\text{kepp\_prob}}{\text{block\_size}^2}\frac{\text{feat\_size}^2}{\left(\text{feat\_size}-\text{block\_size}+1\right)^2} \tag{1}$$

其中，kepp_prob 可以理解为灭活中的单元格被保留的概率。有效种子区域的大小为 $\left(\text{feat\_size}-\text{block\_size}+1\right)^2$，feat_size 是特征图的大小。

3. 方法

Yolov3 与 Dropblock 结合

为了提高模型的泛化能力，我们在 yolov3 中加入了 dropblock (下文简称 db)层。在实践中，在 yolov3 中添加了 8 个 dropblock 层。在 yolov3 模型中，在第 1 个卷积层之后加入了第一个 dropblock 层。在 dropblock 层中，根据设定的参数，将这些被激活的神经元块随机灭活。将丢失一些活跃单位的特征图传递给下一层。在第 1、2、3、4、5 个 resnet 模块后添加了第 2、3、4、5、6 个 dropblock 层。在第 1 个上采样层的前面卷积层之间放入第 7 个 dropblock，然后再放最后一层 dropblock 层在第 2 个上采样层的前面卷积层的中间(如图 1 所示)。db1-1 代表的是第一个 dropblock 层，db1-2 代表的是第二个 dropblock 层，依此类推，db1-8 代表的是第八个 dropblock 层；图中 Conv2D 表示卷积层，BN 表示批归一化处理，LeakyReLU 表示激活函数，resunit 表示一个残差单元，resunit*n 表示 n 个残差单元，resn 表示 n 个 r 残差模块，zero padding 表示零填充层；DBL 由卷积层、批归一化处理、激活函数组成，resunit 由两个 DBL 层组成，resn 由一个零填充层与一个 DBL 层和 n 个残差单元组成，block1-1 表示第一个由 1 个 DBL 层和一个卷积层组成的模块，显然，图中有三个这样相同的模块。

4. 实验结果

本实验在 python3.6、框架 tensorflow1.13.1 及 keras2.2.4 环境下进行课堂环境学生抬头低头检测，整个训练过程的学习率及批量尺寸分别为 0.01 及 4，并且迭代 20 次。整个实验在独立显卡 AMD Radeon Pro WX3100 并且有 Intel(R)Core(TM)i7-9700 CPU 和 64GB 储存的台式电脑上进行。

数据集 ClassUD：此数据集是自己创建完成，摄像机的型号为 SNOY HXR-MC2500，其中包含 2820 张教室上课时的学生图片作为训练及测试集，以及 240 张图片作为测试数据，这些测试数据皆是模仿监控视角的位置与高度拍摄所得数据，范围为一个教室的 3~4 排左右，且 240 张图片由 40 分钟视频以 10 秒一张图片的截取方式获得，实验数据分布如表 1 所示。

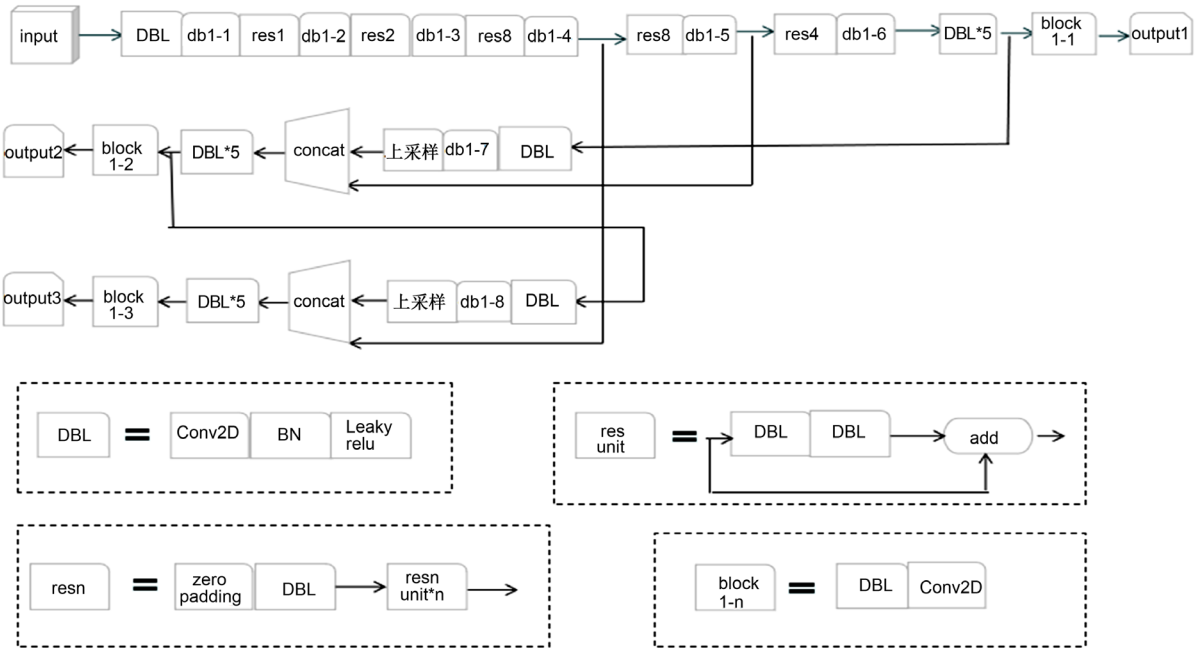

Figure 1. Yolov3 network architecture combined with the db layer
图 1. Yolov3 与 db 层结合的网络体系结构

Table 1. Experimental data distribution
表 1. 实验数据分布

数据集	训练集	验证集	验证集
ClassUD	2538	282	240

实验分为三个部分进行，第一个实验用的是 yolov3 模型在 ClassUD 数据集上进行训练及测试，第二个实验是在 yolov3 的基础模型上加了 dropblock 层(具体网络图如图 2)，同时也是在 ClassUD 数据集上进行训练及测试。第三个实验用的是 yolov3 的精简版 tinyyolov3 在 ClassUD 数据集上进行训练及测试。

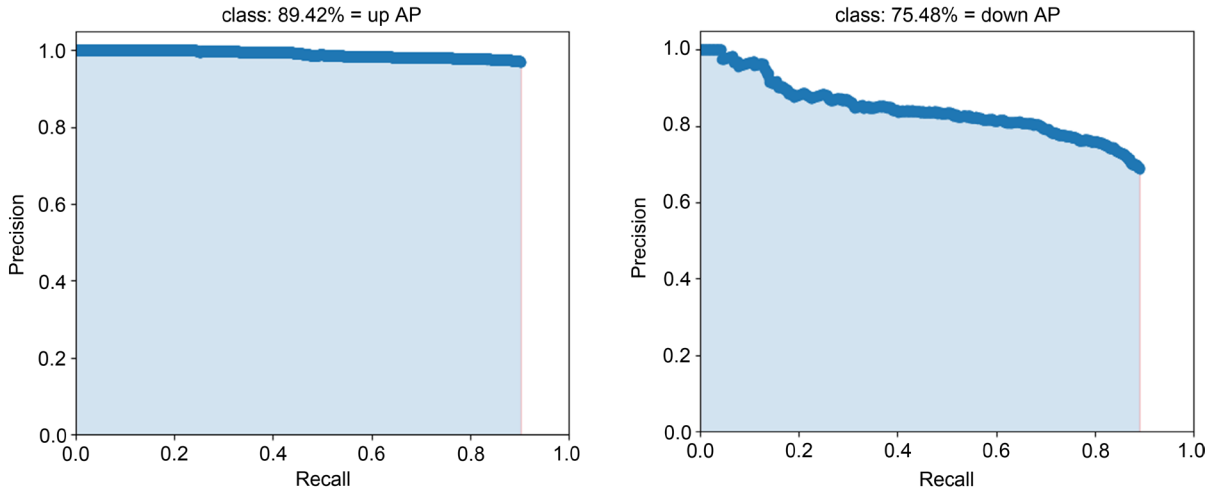

Figure 2. Yolov3 P-R curve
图 2. Yolov3 P-R 曲线图

4.1. 第二部分实验

在这个实验中，我们用 yolov3 加入 dropblock 在数据集 ClassUD 上进行测试，得到的实验结果如图 3。其中 up AP 表示抬头平均精度，down AP 表示低头平均精度；从图中我们可以看出抬头的平均精度为 91.83%，低头的平均精度为 87.01%。

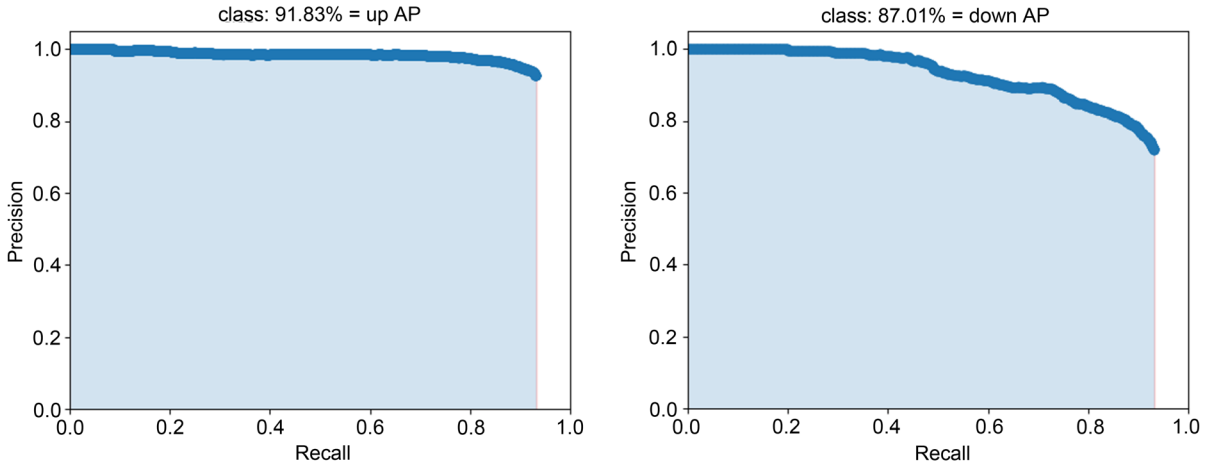

Figure 3. Yolov3 + db P-R curve
图 3. Yolov3 +db P-R 曲线图

4.2. 第三部分实验

在这个实验中，我们用模型 tinyyolov3 在数据集 ClassUD 上进行测试，得到的实验结果如图 4。其中 up AP 表示抬头平均精度，down AP 表示低头平均精度；从图中我们可以看出抬头的平均精度为 78.06%，低头的平均精度为 48.71%。

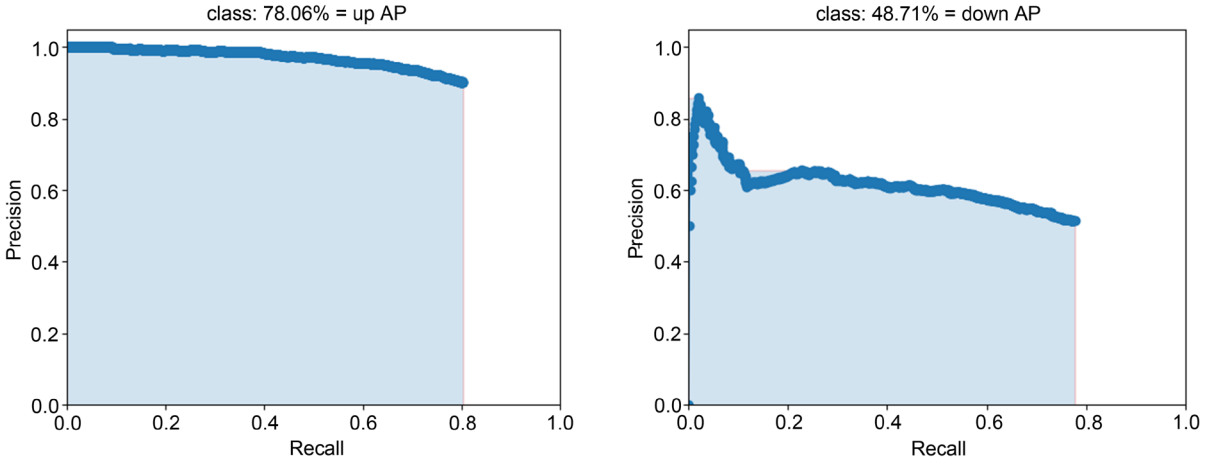

Figure 4. Tinyyolov3 P-R curve
图 4. Tinyyolov3 P-R 曲线图

4.3. 实验结果及分析

从上面的三个实验可以看出，在 yolov3 模型上的抬头检测精度为 89.42%，低头检测精度为 75.48%，经计算均值平均检测精度为 82.45%；在 yolov3+dropblock 模型上的抬头检测精度为 91.83%，低头的检测

精度为 87.01%，计算出均值平均检测精度为 89.42%；在 tinyyolov3 模型上的抬头检测精度为 78.06%，低头检测精度为 48.71%，计算得到均值平均检测精度为 63.39%。yolov3+dropblock 模型相比 yolov3 模型在抬头的检测精度上提高了 2.42%，低头的检测精度提高了 11.53%，均值平均检测精度提高了 6.97%；yolov3+dropblock 模型相比 tinyyolov3 模型在抬头检测精度上提高了 13.77%，低头的检测精度提高了 38.3%，均值平均检测精度提高了 26.03%；yolov3 模型相比 tinyyolov3 模型在抬头的检测精度上提高了 11.36%，低头的检测精度提高了 26.77%，均值平均检测精度提高了 19.06%。

实践中，我们在 yolov3 模型结构中加入了 dropblock 层，并且我们设置 drop_size = 7 和 keep_prob = 0.9。yolov3+dropblock 模型实验结果及 yolov3 (tinyyolov3)模型实验结果如图 5。图 5(a)表示 yolov3 模型下的实验结果；图 5(b)表示 yolov3+db 模型下 drop_size = 7、keep_prob = 0.9 的实验结果；图 5(c)表示 tinyyolov3 模型下的实验结果，可以看出图 5(b)检测的准确率比图 5(a)、图 5(c)情况都要好，这正是我们在 yolov3 中加入 dropblock 层且参数 drop_size = 7、keep_prob = 0.9 的实验结果。其中蓝色的框线表示状态的真实值，绿色的框线表示检测与真实值相符的结果，红色的框线表示错检的结果，粉红色的框线表示漏检的结果。

Figure 5. Partial experimental results
图 5. 部分实验结果

在计算机视觉领域的目标检测中，使用深度学习来做目标检测的实验有很多，每一张图片都包含了不同的目标，我们仅对于我们的研究方向建立了数据集，数据图片中包含一种物体(人)，但我们把她们在课堂学习中的两种状态(抬头低头)看出是两种目标来进行检测，从而有了我们这篇论文的思想。在本论文中，我们通过均值平均精度(mAP)来评估实验模型，评估的结果如表 2，所有的测试均在 ClassUD 数据集上进行。从表可以看出中，当 drop_size = 7、keep_prob = 0.9 时，精度比原有的模型提高了 6.97%，可以看出，在 yolov3+dropblock 模型中，drop_size = 7、keep_prob = 0.9 时好于 yolov3 的情况。而 yolov3 模型又好于 tinyyolov3 的情况。这证明了 dropblock 层对于 yolov3 是有效的，在相同的实验环境下，yolov3 模型的训练时间为 58.79 h，测试每帧图片的时间为 0.57 s，而 yolov3+db 模型的训练时间为 61.51 h，测试每帧图片的时间为 0.59 s，tinyyolo3 模型的训练时间为 14.37 h，测试每帧图片的时间为 0.12 s，虽然在训练过程中 tinyyolov3 的训练速度快于 yolov3，但是它的精度却远远低于 yolov3。在训练时间与测试时间相差不大的 yolov3 模型与 yolov3+db 模型下，显然 yolov3+db 模型对我们的检测任务效果更好。

Table 2. Evaluation results of different models
表 2. 不同模型的评估结果

方法	均值平均精度	训练时间(/小时)	测试时间(帧/秒)
yolov3	82.45%	58.79	0.57
yolov3 + db (7.0.9)	89.42%	61.51	0.59
tinyyolov3	63.38%	14.37	0.12

5. 结论

本文建议的方法利用 yolov3 结合 dropblock 进行教室场景学生课堂抬头低头检测。当数据流入 dropblock 层时，语义信息区域被成块的丢弃，这使得网络不得不集中精力学习剩余语义信息区域中的特征。在 ClassUD 上的抬头低头检测结果证明我们提出的网络结合在性能上比原来的模型要好。该方法有效地提高模型的鲁棒性和泛化能力。但是教室场景的抬头低头检测仍然面临着一系列的问题，如光照、状态不明显(低头幅度较小，可能误检为抬头)、图像质量(摄像头清晰度较低会影响检测效果)和遮挡问题。未来的工作将集中在寻找一种更适合于教室场景抬头低头状态检测的算法，该算法可以实现更好的鲁棒性，提高检测的精度。

参考文献

[1] 唐康, 先强, 李明勇. 基于人脸检测的大学课堂关注度研究[J]. 重庆师范大学学报(自然科学版), 2019, 36(5): 123.

[2] 郭秀兰, 赵佳敏. 本科课堂教学“出勤率、抬头率、满意率”的调查报告[J]. 改革与开放, 2016(19): 108-110.

[3] 左国才, 吴小平, 苏秀芝, 等. 基于 CNN 人脸识别模型的大学生课堂行为分析研究[J]. 智能计算机与应用, 2019, 9(6): 107-110.

[4] 屈梁浩. 基于深度学习的学生课堂疲劳状态的分析与研究[D]: [硕士学位论文]. 重庆: 重庆师范大学, 2019.

[5] Krizhevsky, A., Sutskever, I. and Hinton, G. (2012) ImageNet Classification with Deep Convolutional Neural Networks. 2012 *NIPS*, Lake Tahoe, NV, December 2012, 1097-1105.

[6] Szegedy, C., Liu, W., Jia, Y., *et al*. (2014) Going Deeper with Convolutions. *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, 7-12 June 2015, 1-9. https://doi.org/10.1109/CVPR.2015.7298594

[7] He, K., Zhang, X., Ren, S., *et al*. (2016) Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision & Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90

[8] Girshick, R. (2015) Fast R-CNN. *IEEE International Conference on Computer Vision* (*ICCV*), Santiago, 7-13 December 2015, 1440-1448. https://doi.org/10.1109/ICCV.2015.169

[9] Xie, S., Girshick, R., Dollár, P., *et al*. (2017) Aggregated Residual Transformations for Deep Neural Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 5987-5995. https://doi.org/10.1109/CVPR.2017.634

[10] Liu, W., Anguelov, D., Erhan, D., *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

[11] Redmon, J., Divvala, S., Girshick, R., *et al*. (2016) You Only Look Once: Unified, Real-Time Object Detection. *IEEE Conference on Computer Vision & Pattern Recognition*, Las Vegas, 27-30 June 2016, 779-788. https://doi.org/10.1109/CVPR.2016.91

[12] Redmon, J. and Farhadi, A. (2017) YOLO9000: Better, Faster, Stronger. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 6517-6525. https://doi.org/10.1109/CVPR.2017.690

[13] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement.

[14] Yang, Z., Xu, W., Wang, Z., *et al*. (2019) Combining Yolov3-Tiny Model with Dropblock for Tiny-Face Detection. 2019 *IEEE* 19*th International Conference on Communication Technology* (*ICCT*) *IEEE*, Xi'an, 16-19 October 2019, 1673-1677. https://doi.org/10.1109/ICCT46805.2019.8947158

[15] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *The Journal of Machine Learning Research*, **15**, 1929-1958.

[16] Ghiasi, G., Lin, T.-Y. and Le, Q.V. (2018) DropBlock: A Regularization Method for Convolutional Networks.