

基于改进YOLO v7的血细胞识别计数研究

钟 天

成都信息工程大学, 应用数学学院, 四川 成都

收稿日期: 2023年2月17日; 录用日期: 2023年3月13日; 发布日期: 2023年3月21日

摘 要

血细胞计数具有重要医学意义。而由于传统计数方法存在一定局限性, 本文引入改进YOLO v7, 对血细胞可实行快速且准确的识别计数。改进点为以下三方面: 1) 更改SPPCSPC中Maxpool的尺寸。2) 将第24、37层的卷积层替换为CA注意力。3) 设计了一种上采样整体结构。改进YOLO v7在mAP方面优于原版YOLO v7, 为工程应用带来一定研究意义。

关键词

血细胞, 目标检测, YOLO v7

Research on Blood Cell Recognition and Counting Based on Improved YOLO v7

Tian Zhong

School of Applied Mathematics, Chengdu University of Information Technology, Chengdu Sichuan

Received: Feb. 17th, 2023; accepted: Mar. 13th, 2023; published: Mar. 21st, 2023

Abstract

Blood cell count has important medical significance. However, due to the limitations of traditional counting methods, this paper introduces the improved YOLO v7, which can quickly and accurately identify and count blood cells. The improvement points are as follows: 1) Change the size of Max-pool in SPPCSPC. 2) Replace the convolution layer of layers 24 and 37 with CA attention. 3) An up-sampling architecture is designed. The improved YOLO v7 is better than the original YOLO v7 in terms of mAP, which brings certain research significance for engineering application.

Keywords

Blood Cells, Target Detection, YOLO v7

Copyright © 2023 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

人体血液中的血细胞有三类：白细胞、红细胞、血小板[1]。医学上可根据此三种细胞个数诊断问诊者是否患病。例如：红细胞个数过多，可提示血管阻塞，过少可提示贫血；白细胞数量可提示是否有细菌侵入[2]，血小板个数与高脂血症性急性胰腺炎有关[3]。传统的检测方法需要专业人员使用镜检仪器等检测，耗费大量人力物力。自从 LeNet 问世[4]，卷积神经网络首次用于计算机视觉领域。此后，该领域蓬勃发展，涌现出诸如：AlexNet [5]、VGGNet [6]、GooLeNet [7]、ResNet [8]等架构。而该领域又主要分为三大任务：图像分类、目标检测、图像分割。其中，目标检测技术近年来逐渐应用于各类物体的计数研究。

文献[1]中，探索出一种 YOLO-Dense 的结构，其主要改进为在 YOLO 中引入 Dense 连接，使得 mAP@0.5 达到 86%，获得了比 Faster R-CNN [9]更好的效果。文献[2]中，采用了改进的 YOLO X 对血细胞进行检测，达到了 mAP@0.5 为 92.5%的结果，其主要改进为在 YOLOX 基础上引入了 CBAM 注意力机制、Focal Loss、ASFF。

本文提出改进的 YOLO v7 算法进行血细胞检测，以提升其准确性能。改进的 YOLO v7 的 mAP@0.5 优于 YOLO-Dense 和改进的 YOLO X，mAP@0.5 与 mAP@0.5:0.95 均优于原版的 YOLO v7，这为后续工作的探讨提供一定的思路与启发。

2. YOLO v7 原理

2.1. 目标检测

目标检测是指：在图像或者视频中，找到目标位置并进行分类。其根据阶段数主要分为单阶段与两阶段检测技术。两阶段目标检测的代表模型有 RCNN [10]等。单阶段目标检测的代表模型有：YOLO [11]等。YOLO 系列自 YOLO v3 [12]起，首次采用了 FPN 特征金字塔网络的技术[13]，将浅层的细粒度信息融合进深层的语义信息，以使得模型达到更高的准确性能，而这一点对于小目标检测而言，则更是重要。因为，例如：本文所用数据集中的血小板之类的小目标，其占据像素较少，在多倍下采样的深层，其信息易被污染，浅层信息的融合对于分类和定位任务就显得尤为关键。YOLO v3-SPP [14]中则是首次将 SPP 空间金字塔池化[15]技术运用于 YOLO 系列。其原本是由于非全卷积神经网络。早期，计算机视觉领域的卷积神经网络以末端接全连接结构为主，为了能获得相应尺寸图片的特征转换的向量，SPP 技术被提出。而 YOLO v3-SPP 则从全新角度诠释 SPP 的作用：进一步提取输入特征的多层次感受野下的边缘信息和强信息。YOLO v5 中大量采用了 CSP [16]结构，其本质是一种轻量化思想，它的运用使得模型推理速度提升。YOLO v7 [17]主要采用 ELAN 结构，其融合了 CSP 与 OSA 的思想[18]，并利用 VoVNet [18]作为基础搭建，是一种较优的权衡精度与速度的模型。

2.2. YOLO v7 结构

YOLO v7 结构如图 1 所示:

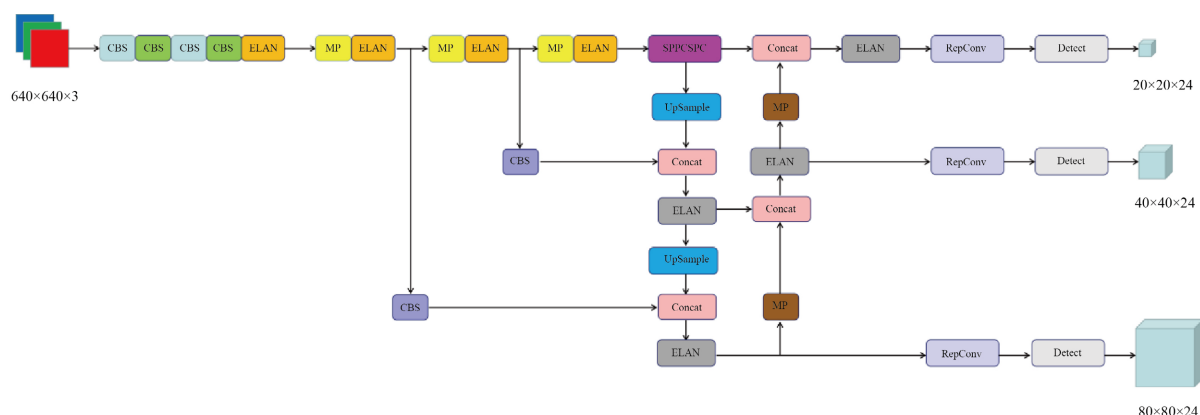


Figure 1. YOLO v7 original structure diagram

图 1. YOLO v7 原版结构图

其中, CBS 包含: Conv、BN、SiLU 激活函数; ELAN 为由卷积层搭建的基础结构; MP 包含: Maxpool 以及 Conv, 并非仅代表 Maxpool; 同样地, UpSample 亦包含: 上采样与 Conv, 并非只代表 Conv。

YOLO v7 结构继承 YOLO v3 开创的 Backbone、Neck、Head 三部分的特点。主干部分首先利用四层单分支卷积结构进行前层信息的处理, 此后在主干中反复堆叠 MP 结构与 ELAN 结构。MP 结构分为两种, 其与 Inception [7] 模块类似, 但其更加精简, 仅采用两个支路, 拥有更少的参数量和更加利于快速推理。ELAN 结构是 YOLO v7 网络中的关键模块, 其也分为两种: 主干中 ELAN 与颈部中 ELAN, 其区别在于聚合到末端的稀疏与稠密。事实上, ELAN 结构受 CSP 思想、VoVNet 思想融合而成, 其要点在于 OSA, 即将前层特征聚合到模块末端。经过 ResNet、DenseNet [19]、SparseNet [20]、VoVNet 等等的演变, 实验上可以表明: 这类 OSA 的结构具有良好的准确性能, 其相对稀疏的连接也带来较好的推理速度, 是一种较优的精度与速度的权衡。ELAN 中的 CSP 相当于是一种修饰, 起到进一步轻量化的作用, 其要点在于将输入特征通道数均分的 $1 * 1$ 卷积, 用法上类似于 BottleNeck, 也可以从梯度流的角度出发, 认为: 这种 CSP 结构能够避免计算过多的重复梯度信息, 从而提高推理速度。SPPCSPC 最初继承自 YOLO v3-SPP。SPPCSPC 与 SPP 的区别在于其运用了 CSP 的思想, 并在 SPP 基础上多使用一个卷积, 以增强特征整合的鲁棒性。RepConv 首次提出于 [21], 其核心在于利用重参数技巧, 将训练时与推理时网络结构区分开。训练时采用多分支结构, 利于准确性能的提升, 而推理时将其算子等价融合为单分支结构, 在不改变结果的情况下能够提高推理速度。

2.3. YOLO v7 检测方法

YOLO v7 检测原理图如图 2 所示。

其中, 粗点代表: 目标的中心点; 虚线框代表锚框; 内侧实线框代表预测框; 各带有箭头的线段表示各长度和指示中心点。

YOLO v7 延续了 YOLO v5 的基于锚框的检测方法。锚框, 是一种先验的预设框, 基于锚框则是指根据预设框作回归以获得预测框。YOLO v7 利用下采样, 形式上将原输入图像分为 $N \times N$ 个网格细胞, 从网络末端的输出结果来看, 其中每一个像素点返回到原图对应一个网格细胞的区域。也有放弃锚框策

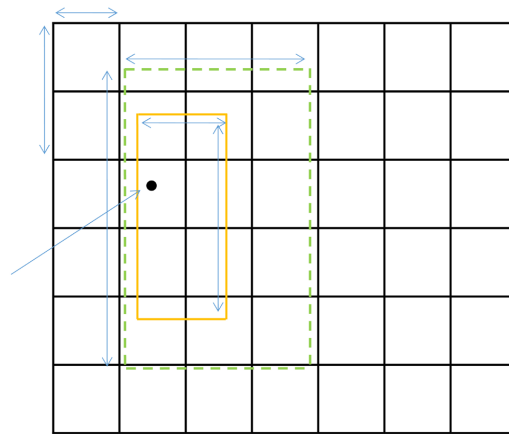


Figure 2. YOLO v7 detection schematic diagram
图 2. YOLO v7 检测原理图

略而使用锚点策略的算法，如：YOLO X [22]、YOLO v6 [23]等等。锚框的获得方法是多样的，可以根据知识的预设、也可以根据已有数据集作 KMeans 聚类获得等方法，但总之：通常，锚框分为三类：对于小尺度、中尺度、大尺度目标检测的锚框。在对于训练集标注时，通常会给出其中心点坐标、真实框的宽高，对于锚框的运用，方法上实际是如公式(1-1)~(1-4)所示，对于中心点的预测采用这样的策略：如果目标落入某个网格细胞，那么就以其左上顶点作为中心点的基准，对与它的偏移量进行预测。关于 YOLO v7 中的分类任务，则继续使用 Sigmoid 函数作激活以预测概率的策略。

$$\hat{P}_w = [2\sigma(t_w)]^2 * P_w \quad (1-1)$$

$$\hat{P}_h = [2\sigma(t_h)]^2 * P_h \quad (1-2)$$

$$\hat{P}_x = 2\sigma(t_x) - 0.5 \quad (1-3)$$

$$\hat{P}_y = 2\sigma(t_y) - 0.5 \quad (1-4)$$

预测各 t 值，以计算左侧各最终的预测值：其中各 Box 的宽高，无论是锚框还是预测框，都已经相对于整图做了归一化；而其中的所谓 x、y 坐标实际上是：当目标落入某 grid cell，以该 grid cell 的左上顶点为(0, 0)，作的偏移量，其已经相对于 grid cell 做了归一化。在 YOLO v2 版本中，中心点的预测仅能限定在落入 grid cell 的开区间内，但这不利于检测的准确性能和模型的收敛速度，因此 YOLO v7 作改进，在公式(1-3)、(1-4)中表明：其中心点横纵坐标范围可以是(-0.5, 1.5)。

3. 改进的 YOLO v7

本文提出三点改进：1) 将原版 SPPCSPC 中的 Maxpool 的尺寸替换为(3, 5, 7)。Maxpool 对输入的特征提取邻域内强信息，在一些任务中可以进行边缘提取。在 SPPCSPC 中可以对主干网络提取的语义信息进行特征整合，抽象出表达更精简的特征。但其感受野尺寸对于模型效果具有影响。在这里，考虑由于数据集像素降低，且其中含有像素较少的血小板目标，为了较多保留信息，而不至于 Maxpool 过多重复地获得某些局部极大值信息，导致信息更加匮乏，本文将原本的(5, 9, 13)尺寸的感受野改为(3, 5, 7)。2) 将第 24 层与 37 层的卷积层替换为 CA [24]注意力。CA 注意力与 CBAM 注意力[25]类似，但其将 CBAM 中的对空间或者通道的全局平均池化改进为：对水平、垂直两方向的平均池化。这样将有利于在信息提取阶段，更完善地实现空间与通道信息的交流。3) 将 Neck 部分原本的上采样与其前的卷积次序替换，并

将上采样从最近邻插值方法替换为 CARAFE [26]。这是受到 Inception 启发,本文中采取替换次序的策略,实质上意在将上采样过程保持在较高的通道维度上。将其原本的 1×1 卷积改至上采样之后,其整体的输出数据结构不变,但由于在较高通道维度上进行上采样,上采样过程中将保留更完善的信息,这对于本文数据集低分辨率且同时具有小目标:血小板的情况是有利的。同时,将 1×1 卷积置于上采样之后,能够进一步整合上采样的特征,对于信息鲁棒性有加强作用。而 CARAFE 对于上下文信息具有一定的感知功能,不同于最近邻插值上采样, CARAFE 可以根据邻域内的信息学习权重以完成上采样,而避免了直接采用原始输入。

4. 实验

数据集: 一共使用 857 个血细胞数据集,其来自于 BCCD 通用数据集。随机抽取 600 个作为训练集,257 个作为测试集。

实验平台: NVIDIA Tesla A40。

模型训练: 本实验采用迁移学习的思想。考虑到数据集中血细胞们像素较低,且数量较少,可能存在较困难的优化过程,因此预先采用了 MS COCO 数据集上预训练权重作为初始化标准。

如图 3、图 4 所示为改进前后的 mAP@0.5 与 mAP@0.5:0.95 的对比曲线图,其中在优化过程稳定后,改进后的曲线位于改进前上方。

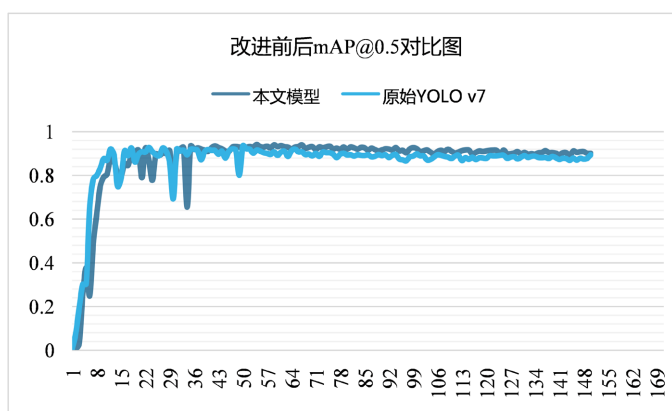


Figure 3. Before and after improvement mAP@0.5 Comparison Chart
图 3. 改进前后 mAP@0.5 对比图

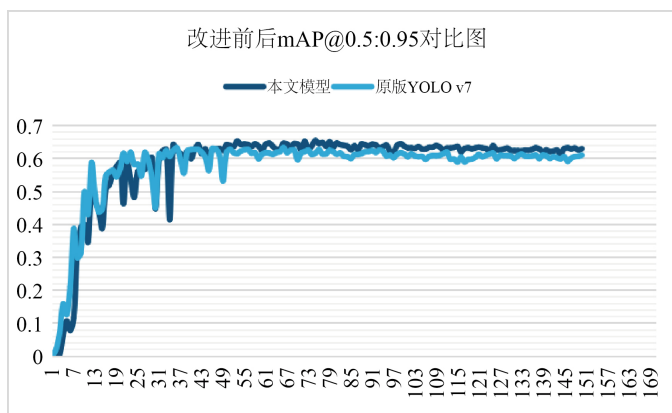


Figure 4. Before and after improvement mAP@0.5:0.95 Comparison Chart
图 4. 改进前后 mAP@0.5:0.95 对比图

5. 模型评估

深度学习目标检测中常用的评估指标有：Recall、Precision、mAP 等等，其中 mAP 的计算方式，借由 MS COCO 数据集的评估标准，常见地又分为两种：mAP@0.5 和 mAP@0.5:0.95。本文采用的就是这两种 mAP 来衡量模型效果，这是因为：以计数为目的任务，常见的是通过记录 NMS 后的预测框个数以完成。但高的 Recall，却可能存在大量误检，NMS 后仍存在大量精确度不高的预测框，导致计数偏多；而高的 Precision，却可能存在大量漏检，NMS 后预测框过少，但在这仅有的少量框内，准确度较高，而导致计数偏少。

在改进前后，评估指标的对比结果如表 1 所示：

Table 1. Comparison results before and after improvement

表 1. 改进前后对比结果

	mAP@0.5	mAP@0.5:0.95
改进前	0.9073	0.6336
调最大池化参数	0.9152	0.6411
本文模型	0.9340	0.6548

6. 结束语

本文对血细胞进行了利用深度学习手段的识别计数研究，并改进了原版的 YOLO v7 算法，使其准确性得到一定的提升。对于后续的研究，若仅以提升准确性为目的，可以考虑先利用 GAN [27] 等技术，将低分辨率的血细胞图像超分辨，这样能够丰富目标信息，也许能够对于其效果起到一定的改善作用。

基金项目

编号：2022-KFJJ-02-001。

参考文献

- [1] 徐晓涛, 孙亚东, 章军. 基于 YOLO 框架的血细胞自动计数研究[J]. 计算机工程与应用, 2020, 56(14): 98-103.
- [2] 易见兵, 黄苏泉, 曹锋, 李俊. 基于改进的 YOLOX 血细胞检测算法研究[J]. 电子测量技术, 2022, 45(22): 177-184.
- [3] 李晓宁, 许克东, 张晶, 邢斌瑜, 申存毅, 王欢. 血清铁蛋白和血小板相关参数与高脂血症性急性胰腺炎患者病情及预后的关系分析[J]. 中国医刊, 2023, 58(2): 160-163.
- [4] LeCun, Y., Bottou, L., Bengio, Y. and Hatan, P. (1998) Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, **86**, 2278-2324. <https://doi.org/10.1109/5.726791>
- [5] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, **60**, 84-90.
- [6] Simonyan, K. and Zisserman, A. (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, 7-9 May 2015, 1-14.
- [7] Szegedy, C., Liu, W., Jia, Y.Q., et al. (2015) Going Deeper with Convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 7-12 June 2015, 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
- [8] He, K.M., Zhang, X.Y., Ren, S.Q. and Sun, J. (2016) Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 770-778.
- [9] Girshick, R., Donahue, J., Darrell, T., et al. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. <https://doi.org/10.1109/CVPR.2014.81>
- [10] Ren, S.Q., He, K.M., Girshick, R. and Sun, J. (2015) Faster R-CNN: Towards Real-Time Object Detection with Re-

- gion Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149.
- [11] Redmon, J., Divvala, S., Girshick, R., *et al.* (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [12] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement. *IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-22 June 2018, 2311-2314.
- [13] Lin, T.-Y., Dollár, P. and Girshick, R. (2016) Feature Pyramid Networks for Object Detection.
- [14] Joseph Redmon. <https://pjreddie.com/darknet/yolo>
- [15] He, K.M., Zhang, X.Y., Ren, S.Q. and Sun, J. (2014) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *ECCV, 13th European Conference*, Zurich, 6-12 September 2014, 346-361.
- [16] Wang, C.-Y., Liao, H.-Y.M., Yeh, I.-H., *et al.* (2019) CSPNET: A New Backbone That Canenhance Learning Capability of CNN.
- [17] Wang, C.-Y., Bochkovskiy, A. and Liao, H.-Y. (2022) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors.
- [18] Lee, Y., Hwang, J.-W., Lee, S. and Bae, Y. (2019) An Energy and GPU-Computation Efficient Backbone Network for Real-Time Object Detection. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, 16-17 June 2019, 752-760.
- [19] Huang, G., Liu, Z. and van der Maaten, L. (2017) Densely Connected Convolutional Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 2261-2269. <https://doi.org/10.1109/CVPR.2017.243>
- [20] Zhu, L.G., Deng, R.Z., Maire, M., *et al.* (2018) Sparsely Aggregated Convolutional Networks. *ECCV, 15th European Conference*, Munich, 8-14 September 2018, 192-208. https://doi.org/10.1007/978-3-030-01258-8_12
- [21] Ding, X.H., Zhang, X.Y., Ma, N.N., *et al.* (2021) RepVGG: Making VGG-style ConvNets Great Again. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 13728-13737. <https://doi.org/10.1109/CVPR46437.2021.01352>
- [22] Ge, Z., Liu, S.T., Wang, F., *et al.* (2021) YOLOX: Exceeding YOLO Series in 2021.
- [23] Li, C.Y., Li, L.L., Jiang, H.L., *et al.* (2022) YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications.
- [24] Hou, Q.B., Zhou, D.Q. and Feng, J.S. (2021) Coordinate Attention for Efficient Mobile Network Design. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 13708-13717. <https://doi.org/10.1109/CVPR46437.2021.01350>
- [25] Woo, S.H., Park, J., Lee, J.-Y. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. *ECCV, 15th European Conference*, Munich, 8-14 September 2018, 3-19.
- [26] Wang, J.Q., Chen, K., Xu, R., *et al.* (2019) CARAFE: Content-Aware Reassembly of Features. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 3007-3016. <https://doi.org/10.1109/ICCV.2019.00310>
- [27] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., *et al.* (2020) Generative Adversarial Networks. *Communications of the ACM*, **63**, 139-144.