

一种新型的肾实质分割网络

张容祥

上海理工大学, 上海

收稿日期: 2022年9月13日; 录用日期: 2022年10月12日; 发布日期: 2022年10月20日

摘要

肾是人体内十分重要的一个组织器官, 因此肾相关的各种病最近几年引起了极大的关注。在这之中, 肾实质就是其中的最常见的一种肾病之一。到目前为止, 关于肾实质病变的诊断主要依赖于临床医生的标注, 从而人工进行判断。这样的方式需要很大的人工与时间成本, 因此, 亟需一种新的方法提升肾实质诊断的效率和精度。本文主要针对小儿肾图, 建立了相关的儿童肾实质数据集。并且, 根据小儿肾图数据集的特征, 我们提出了一种新的分割方法。本方法的核心是细化跳跃连接模块(RSC模块)与transform架构。本文提出的网络不仅提升了分割的精度, 改善了由于分辨率降低从而导致的感受野下降的问题, 而且大大减少了人工标注的时间, 提高了诊断的精度。本文的代码是基于pytorch框架进行的编程, 在小儿肾图数据集进行的实验, 此外, 将本文提出的网络与经典的FCN、SegNet、U-Net和Deeplab-V3+做了对比实验。结果显示本文提出的方法在precision、dice_coeff、recall三种评价指标上, (对比其网络在这三种指标上最优的结果)分别提升了2.547%、4.992%、2.498%, 其效果也得到了专业医生的认可。

关键词

细化跳跃连接模块, 肾实质分割, 深度学习, 语义分割

A New Type of Renal Parenchymal Segmentation Network

Rongxiang Zhang

University of Shanghai for Science and Technology, Shanghai

Received: Sep. 13th, 2022; accepted: Oct. 12th, 2022; published: Oct. 20th, 2022

Abstract

The kidney is a very important tissue and organ in the human body, so various diseases related to

the kidney have attracted great attention in recent years. Among them, the renal parenchyma is one of the most common kidney diseases. Until now, the diagnosis of renal parenchymal lesions has relied mainly on the clinician's annotations to make manual judgments. This method requires a lot of labor and time costs, so there is an urgent need for a new method to improve the efficiency and accuracy of renal parenchymal diagnosis. In this paper, the relevant pediatric renal parenchymal dataset is established for pediatric renal maps. And, based on the characteristics of the pediatric kidney map dataset, we propose a new segmentation method. The core of this method is to refine the jump connection module (RSC module) and transform architecture. The network proposed in this paper not only improves the accuracy of segmentation and improves the problem of declining the sensing field due to the reduction of resolution, but also greatly reduces the time of manual labeling and improves the accuracy of diagnosis. The code in this article is based on the pytorch framework, experiments performed on pediatric kidney map datasets, and in addition, the network proposed in this paper is compared with the classic FCN, SegNet, U-Net, and Deeplab-V3+. The results show that the proposed method is based on three evaluation indicators: precision, dice_coeff and recall. (Comparing the optimal results of its network on these three indicators) it is increased by 2.547%, 4.992%, and 2.498% respectively, and its effect was also recognized by professional doctors.

Keywords

Refine Jump Connection Module, Renal Parenchymal Division, Deep Learning, Semantic Segmentation

Copyright © 2022 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

如下图 1 所示为小儿肾图原始图像, 从图像中可以看出它的边界是十分模糊的。当病人的肾脏发生病变时, 医生会通过注射静脉注射示踪剂, 经过肾图仪监控, 得到左右两肾的肾图, 再加以医生的手工标注, 对肾的健康与否做出诊断。众所周知, 医疗图像的标注是一项很精细的工作, 一旦标注错误, 产生误诊, 会对医生以及患者造成极其严重的损失。一名优秀的影像科医生对于肾图的标注速度大概在二十到三十张每天, 并且医生手工标注的精细程度也会随着工作量的增加而变差, 这样的效率和精度显然是不高的。因此, 对于肾脏病灶区的人工标注, 难度是很大的。

计算机辅助诊断系统[1] [2] (computer aided diagnosis, CAD), 伴随着人工智能领域的发展, 开始逐渐家喻户晓。特别是对于医疗图像领域的研究, 人工智能所附属的图像语义分割, 由于本身效果的优秀, 以及自动分割结果的特性, 受到了很多医生以及医学爱好者们的大力青睐。图像语义分割可以帮助医生快速定位病灶区的位置, 充分地解放了医生的双手和脑力, 这样可以使医生更有精力投入到病情分析和诊断中, 并且大大地提升了诊断结果, 为医生和患者之间搭建了诊断的桥梁, 因此医学图像分割占据了医疗影像研究领域较大的份额。但是肾图[3]分割是一个还未被广泛研究的领域。

本文在 transform 的基础上, 在它本身的解码层中加入了细化跳跃连接模块, 提高了肾实质标注的精度, 同时降低了医生用于肾图标注的成本, 使得诊断的速度大幅度地提升, 帮助医生对患者进行更加准确地处理患者肾脏病灶区。最后我们通过实验与一些比较流行的方法(U-Net [4]、FCN [5]、Deeplab-V3+ [6]、SegNet [7])。

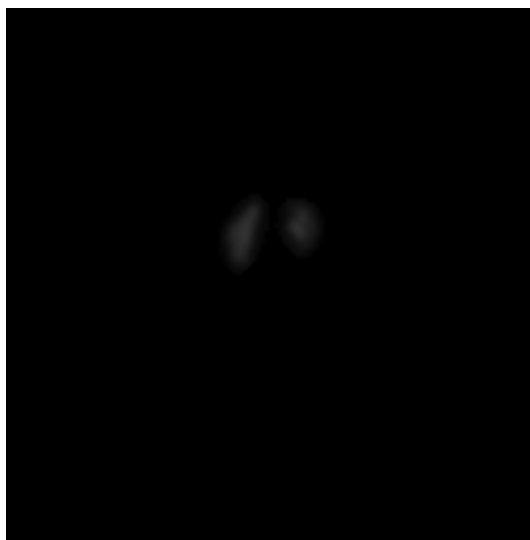


Figure 1.Child kidney diagram
图 1. 小儿肾图

本文的主要贡献如下:

- 1) 在医院提供的小儿肾图原图的基础上,进行了图像预处理,并制作了小儿肾图数据集。
- 2) 提出基于 transform 与细化跳跃连接结合的新型网络该网络,并且将该网络用于肾图的语义分割,本网络改变了传统的卷积网络编写的编解码层,而是用 transform 和细化跳跃连接代替,提升了分割的精度。

2. 图像分割网络

人工智能(Artificial Intelligence)是一个很宽泛的学科,简单来说就是让机器学会人的思维,学会学习预测、视觉识别、语音识别以及智能控制。而神经网络就是实现机器自主学习的一种方式。计算机视觉(CV)是目前人工智能领域的一个热门研究问题,而图像的识别是 CV 领域一个基础的问题。

图像的识别又可以细分为图像分类、目标检测和图像分割这几类任务。目标检测的目的是找出图像中所有感兴趣的目标,同时还需要对他们的位置、大小以及所属类别进行确定。对于图像分割来说,它可以再细分为语义分割、实例分割和全景分割这三类。语义分割是对图像中的每个像素点一一进行类别的分类,实例分割是在语义分割的基础上对同一个类别的不同对象再进行具体的区分,全景分割则是结合了语义分割和实例分割,既可以完成前景和背景的分割,也可以实现同一个类别的不同目标的具体区分。

CNN 的出现开辟了图像分类以及目标检测等任务的新境界,并涌现出了诸如 R-CNN [8]、Fast R-CNN [9]、YOLO [10] [11] [12] [13]等优秀的网络,并在 ImageNet 上取得了极好的成绩,甚至超过了人类。对于图像分类和目标检测来说,更注重的是对图像整体的把握,而图像分割则是需要实现像素级的分割,因此需要更精确的算法。2016 年,何凯明团队在 R-CNN 的基础上,开发了一个用于实例分割的模型 MaskR-CNN [14],当然这个模型最大的优势在于它不是一个一成不变的实例分割框架,它可以通过在网络的输出端增加不同的分支来完成其他的图像处理任务,例如图像分类、目标检测、语义分割和实例分割等。2015 年,一个以全卷积神经网络(Fully Convolutional Networks, FCN)为基础的语义分割模型首次被提出,该模型使用卷积层替换经典的分类网络模型(VGG16)中使用的全连接层,并增加反卷积和跳跃连接(skip connection)的结构,提升了语义分割的准确性。

3. Transform 编码层

下图 2 所示为 transform 编码层的结构:

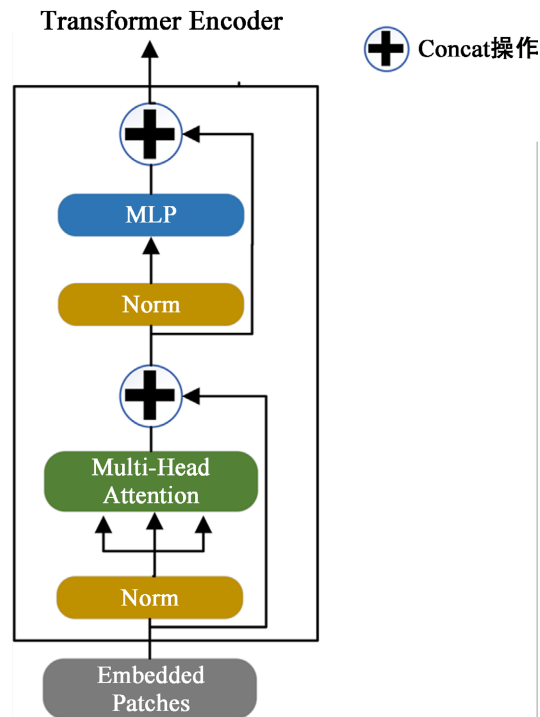


Figure 2. Transformer schematic diagram

图 2. Transformer 示意图

在 transformer 层中, 它每层的输出结果为 $Z^1, Z^2, Z^3, \dots, Z^L$, 其中多个注意力相关的运算输入到多头注意力机制(Multi-Head attention)。它每组 (q, k, v) 所形成的矩阵通过相应的运算生成一个 SA (Self-attention) 即自注意力, 这一过程即自注意力的运算。最后将形成的多个矩阵进行拼接, 在乘以一个参数矩阵 W_o , 便可得到最终的输出。

公式(1)为三元组 (q, k, v) 的计算公式, 其中 $W_Q, W_K, W_V \in R^{L \times C}$ 是随机矩阵, L 代表输入的层数, $Z^L \in R^{L \times C}$ 代表 L 行 C 列的矩阵, L 代表层数, C 代表列数, W_o 代表参数矩阵。

$$q = Z^{L-1}W_Q, \quad k = Z^{L-1}W_K, \quad v = Z^{L-1}W_V \quad (1)$$

SA (Self-attention)计算公式如公式(2)

$$SA(Z^{L-1}) = Z^{L-1} + \text{softmax}\left(\frac{Z^{L-1}W_Q(ZW_K)^T}{\sqrt{d}}\right)(Z^{L-1}W_V) \quad (2)$$

MSA (Multi-Head Self-attention)计算公式如公式(3)

$$MSA(Z^{L-1}) = [SA_1(Z^{L-1}); SA_2(Z^{L-1}); \dots; SA_m(Z^{L-1})]W_o \quad (3)$$

L 层的最终输出结果如公式(4)

$$Z^L = MSA(Z^{L-1}) + MLP(MSA(Z^{L-1})) \in R^{L \times C} \quad (4)$$

细化跳跃连接模块(RSC 模块)

原图经过 transform 层编码后,然后通过多尺度通道的自选择模块,输出流一张分辨率较低的特征图,生成的特征图包含了细节和全局的信息,但是缺乏一些边界和图像的细节信息。

U-Net 最早提出了使用跳跃连接进行特征融合,来补充下采样得到的特征图的细节信息。不同层次的特征图之间会存在语义间隙,传统的跳跃连接结构直接将不同深层信息和浅层信息进行拼接操作,这样显然会产生语义间隙的问题。为了减小语义间隙的影响,我们设计了基于注意力机制细化跳跃连接模块(图 3)。

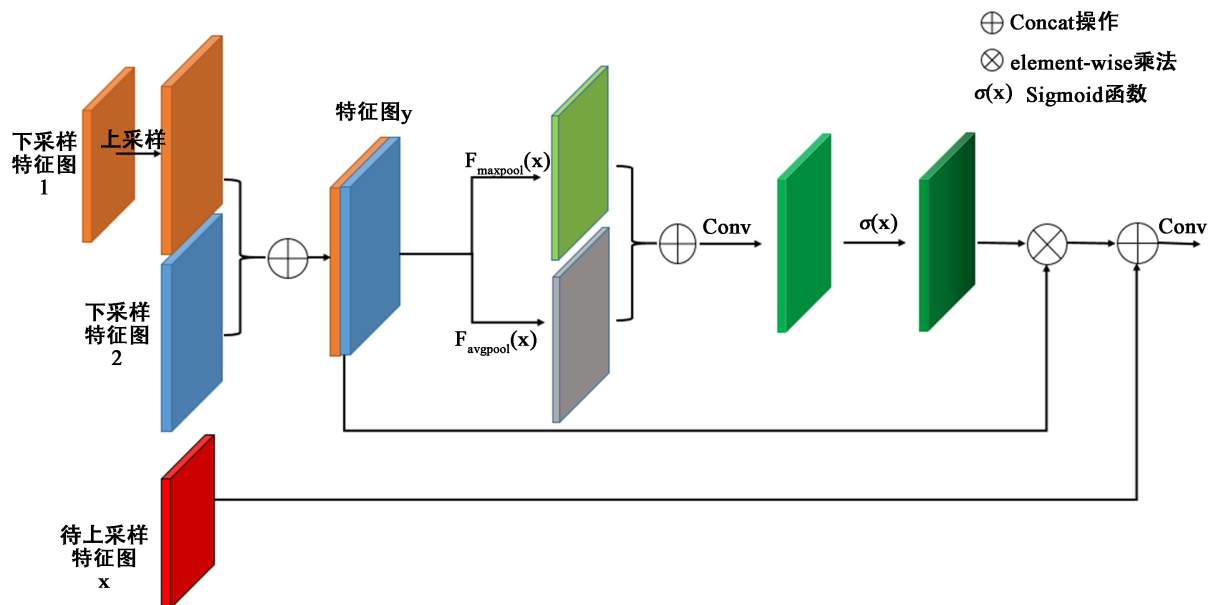


Figure 3. Refine jump connection module (RSC module)

图 3. 细化跳跃连接模块(RSC 模块)

4. 实验结果与分析

4.1. 实验软硬件环境

本文进行试验所使用的实验平台如下表 1:

Table 1. Experiment platform

表 1. 实验平台

实验软硬件环境	环境配置
操作系统	Ubuntu18.04
处理器	Intel(R) Xeon(R) Platinum 6164 CPU @ 1.90 GHz
内存	48 g
显卡	NVIDIA RTX3090
编程语言	Python 3.8.8
深度学习框架	PyTorch 1.10.0 [15]

4.2. 小儿肾图数据集介绍与预处理

本文之中所运用的有关于肾实质的数据是由上海交大医学院所附属的新华医院影像科提供的病人的

原始图(700)套，我们在所提供的数据集上对数据进行了预处理。它的原理是基于肾实质静脉注射的物质进行追踪，最后通过尿液排出所产生的影像，最后通过肾图仪进行收集图像。正向我们下图所看到的。通过肾图仪产生的图像是非常的模糊不堪的灰色图片，因此就需要在我们实验前提前处理下这些数据。每个人通过仪器最终大概会得到 136 张图片。肾实质数据集是在专业的肾科室的影像科医生指导下进行的，在数据处理的过程中，由于前几张的数据，显影剂并没有到达指定的位置，所以我要除去大概一分钟前的灰色图。随后我们对后面的灰度图片，按照时间的顺序，每七张进行融合。一般每个人最终会得到 11 张用于标记的处理图片。最后我们把这 11 张图片利用 OpenCV 进行伪彩色处理，之所以用伪彩色处理，是因为生成的伪彩色的图片会更加有利于随后的标注和训练。我们把每一组图片中的第二张和第三张图片进行标注，并且把原图和我们标注后的标签一同收录到我们的数据集中。经过以上几步操作，我们最终一共得到了肾实质数据集(共 700 张)，最后，通过对得到的数据集图片适当的剪裁，得到 384*384 像素大小的图片。其中打字的流程如下图 4 所示：

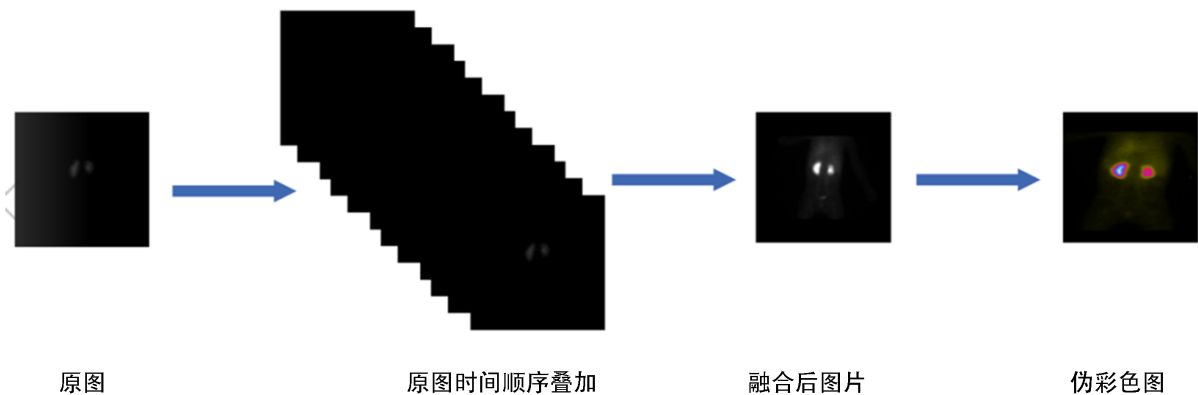


Figure 4. Renal parenchyma data set production process

图 4. 肾实质数据集制作过程

4.3. 损失函数与评价指标

本实验采用了 Recall (回归率)、Precision (精确率)、Dice_coeff (集合相似性)三个人常用于语义分割领域的指标对本文所提出的新型模型进行评价。

本文使用的损失函数为经典的交叉熵损失函数，函数公式如下所示：

$$L_1 = -\sum_{n=1}^N (t_n \ln \rho_n + (1-t_n) \ln (1-\rho_n)) \quad (5)$$

其中， t_n 表示真实标签类别，当 n 为变化类像素时 t_n 取值为 1，否则， t_n 取值为 0； ρ_n 代表预测 n 为变化类像素的概率且 $\rho_n \in [0,1]$ 。 N 是一个样本中总的像素数， n 是样本中一个像素。

$$\text{Precision} = \text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$

$$\text{Dice_coeff} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \quad (8)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (9)$$

$$\text{Jaccard} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (10)$$

实验中我们设置的超参数学习率为 0.002, 设置的 batch-size 为 8, 训练迭代次数设置为小于等于 1000 个 epoch。

4.4. 实验结果

由 4.2 节可知, 由本文制作的肾图数据集包含了 700 张伪彩色图。在实验过程中随机抽取了其中的五分之四作为训练集, 剩下的五分之一作为测试集, 下面具体介绍所做实验。

在本实验中, 为了充分的验证我们所提出的新型网络的效果, 我们与另外四种常用于医学图像分割方向的经典网络分割效果作对比。其中这四个网络包括 FCN, SegNet, U-Net, Deeplab-V3+。实验结果数据如下表 2: FCN, SegNet, U-Net, Deeplab-V3+。

Table 2. Comparison of different network segmentation effects

表 2. 不同网络分割效果对比

	dice_coeff	precision	recall
FCN	0.90910	0.93690	0.92002
U-Net	0.91213	0.93451	0.93301
SegNet	0.89526	0.94037	0.92433
Deeplab-V3+	0.89156	0.92684	0.92139
Our	0.95766	0.96432	0.95632

由实验结果得知, 使用本文提出的网络进行肾图分割, 得到的分割结果在 precision、recall 和 dice_coeff 上均优于运用其余四种经典网络得到的分割结果, 其中比 U-Net 在 dice_coeff 方面高出 4.992%, 在 precision 方面高出 2.547%, 在 recall 上也提升了 2.498%, 总体效果提升明显。

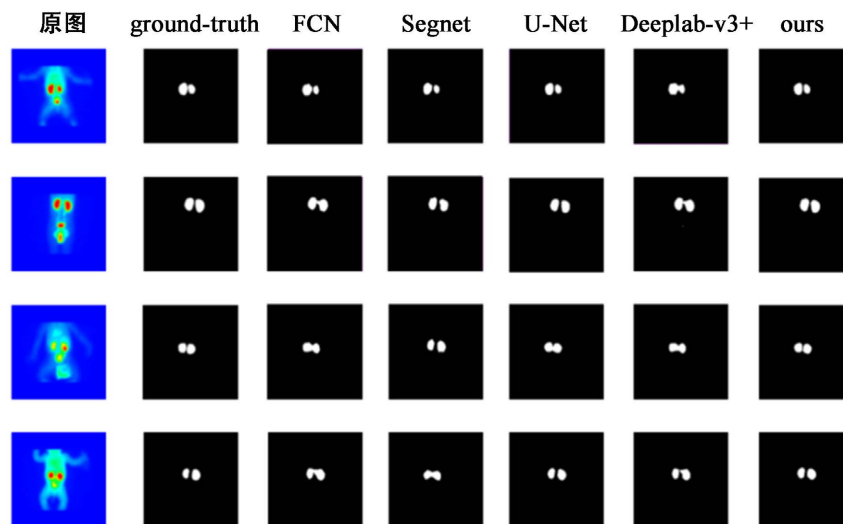


Figure 5. Comparison of different network segmentation effects

图 5. 不同网络分割效果对比图

通过图 5 的效果图我们也可以明显的观察到, 本文提出的网络特别是在细节问题上, 要比其他的四

种经典网络分割的效果要好。

5. 结论

本文提出了一种新型的肾实质分割网络，该网络结合了当下热门的细化跳跃连接模块和 transform 模块，这也是首次把 transform 和细化连接模块应用在肾实质数据集上，正是 transform 和细化连接模块优越的特征提取特点，使得本模型非常适合用于肾实质病灶区的分割诊断。最终的分割效果图也包括了图像的各个尺度的特征，弥补了以前由于卷积导致的个别细节信息丢失的问题。提升了网络的感受野，提高了图像分割的精度，减轻了医生的负担，对于医生的辅助诊断有很高的研究价值。虽然本文提出的网络就，它的精度已经非常高，但是在医疗辅助诊断方面，精度当然是百分之百最好，所以希望未来会有更好的结构替代本文提出的网络。

参考文献

- [1] Chan, H.P., Doi, K., Galhotra, S., Vyborny, C.J., MacMahon, H. and Jokich, P.M. (1987) Image Feature Analysis and Computer-Aided Diagnosis in Digital Radiography. I. Automated Detection of Microcalcifications in Mammography. *Medical Physics*, **14**, 538-548. <https://doi.org/10.1118/1.596065>
- [2] Van Ginneken, B., Romeny, B.M.T.H. and Viergever, M.A. (2001) Computer-Aided Diagnosis in Chest Radiography: A Survey. *IEEE Transactions on Medical Imaging*, **20**, 1228-1241. <https://doi.org/10.1109/42.974918>
- [3] Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., de Lange, T., Johansen, D., et al. (2020) Kvasir-Seg: A Segmented Polyp Dataset. *International Conference on Multimedia Modeling*, Daejeon, 5-8 January 2020, 451-462. https://doi.org/10.1007/978-3-030-37734-2_37
- [4] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-net: Convolutional Networks for Biomedical Image Segmentation. *18th International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, 5-9 October 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [5] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 3431-3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- [6] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H. (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *Proceedings of the European Conference on Computer Vision (ECCV)* 2018, Munich, 8-14 September 2018, 833-851. https://doi.org/10.1007/978-3-030-01234-2_49
- [7] Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 2481-2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- [8] Hu, J., Shen, L., Albanie, S., Sun, G. and Wu, E. (2020) Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**, 2011-2023. <https://doi.org/10.1109/TPAMI.2019.2913372>
- [9] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *Proceedings of the 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] Girshick, R. (2015) Fast R-CNN. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
- [11] Redmon, J., Divvala, S., 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>
- [12] Redmon, J. and Farhadi, A. (2017) YOLO9000: Better, Faster, Stronger. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, 21-26 July 2017, 6517-6525. <https://doi.org/10.1109/CVPR.2017.690>
- [13] Redmon, J. and Farhad, A. (2018) YOLOv3: An Incremental Improvement. arXiv e-prints, arXiv:1804.02767. <https://arxiv.org/abs/1804.02767>
- [14] He, K., Gkioxari, G., Dollár, P. and Girshick, R. (2017) Mask R-CNN. *Proceedings of the IEEE Conference on Computer Vision*, Venice, 22-29 October 2017, 2980-2988. <https://doi.org/10.1109/ICCV.2017.322>
- [15] Paszke, A., Gross, S., Massa, F., et al. (2019) Pytorch: An Imperative Style, High-Performance Deep Learning Library. *36th Annual Conference on Neural Information Processing Systems*, 8-14 December 2019, 8026-8037.