

基于深度学习的宫颈癌智能预测系统实现

刘四旗 高泰恒 李萌奇 王淑娇

(山东协和学院 计算机学院 山东 济南 250000)

【摘要】目前宫颈癌呈现出低龄化的趋势, 早筛早诊尤为重要。细胞检测是早筛的重要方法, 但其也有明显缺点。本文提出了一种基于 U-Net 算法和 YOLO 系列算法的宫颈癌细胞学检测方法, 通过算法改进提高了其分割、检测效果。通过对 U-Net 算法的网络调整, 提高了特征信息的利用率, 提升分割效果, 将分割出的细胞区域更加准确, 提高了其性能。

【关键词】宫颈癌细胞学; U-Net 神经网络; 损失函数

1 前言

宫颈癌是一种对女性健康威胁极大的一种疾病, 近年来宫颈癌发病年龄逐渐减小, 而早期宫颈癌治愈概率高, 提早发现宫颈癌细胞, 可以大大提高患者生存率。

当今宫颈癌检测方法主要通过 HPV 检测、TCT、宫颈活检术、宫颈锥切术进行检测, 而细胞学检测在早筛过程中起到了很大的作用, 传统的宫颈癌细胞学检测早筛自动化程度低, 受医师经验和主观条件制约, 存在着误、漏诊问题。

在利用深度学习进行宫颈癌细胞检测中, 存在着以下难点:

(1) 由于涉及到患者隐私, 导致现有关于宫颈癌细胞数据集较少且质量低, 使用深度学习难以拟合。

(2) 由于宫颈癌病灶部位情况复杂, 即使检测出, 仍需医师进行复核。而使用深度学习为医师节省更多时间成为当前研究的重中之重。

(3) 在宫颈癌检测任务中首先需要图像预处理, 其次进行图像分割与判断。任务量偏大。

2 宫颈癌智能预测系统

细胞分割是保证可以精准快速地进行宫颈癌筛查的重要手段, 随着人工智能技术的发展, 针对图像分割的图像处理算法不断增多, 在执行医学细胞分割任务中, 由于细胞的随机分布导致的细胞在图像中出现粘连重叠, 和细胞之间的差异, 由于物理原因导致的细胞损伤成为细胞分割任务的“痛点”^[1], 解决以上问题成为分割任务的重点。针对图像分割研究中的“痛点”, 目前出现了很多算法解决上述问题, 比如分水岭算法、FCN、U-net 等。语义分割网络 U-Net 算法在 2015 年被提出, 其基于编码与解码思想有效的结合低分辨率与高分辨率信息, 可以更好地执行分割任务。

由于宫颈癌细胞在形态方面会发生改变, 所以采用机器视觉进行判断成为可能, 宫颈细胞检测作为宫颈癌细胞筛查的最后步骤, 起到了决定性作用, 但是由于细胞在图像中占比过小,

检测任务会退化成像素级分类任务, 再加上细胞过小数据标注任务难度大大提升, 目前解决以上问题的方法有采取选择性的数据采样进行 ACM 与 FCNs 之间协同效应的反馈处理^[2]。

本系统采用 TCT/HPV 结合分析, TCT/HPV 为宫颈癌筛查的必要项目, 也是女性日常体检中的必检项目。本系统对 TCT/HPV 医学影像进行预处理后, 使用 YOLOX 目标检测算法对图中的液基薄层细胞进行检测, 并综合 HPV 病毒检测给出辅助诊疗建议, 并将辅助诊疗报告存入系统。较高精确度的辅助诊疗减少了人工阅片的时间, 使医务工作者的工作效率极大的提升, 有助于患者的宫颈癌早筛早诊。

2.1 YOLO 系列在医学图像处理方向的研究与延伸

人工智能在细胞病理学中有着重重要的地位, 可以精准快速地识别疾病是人工智能领域的最重要目标。该目标难点主要来自于前期图像预处理复杂, 原因是因为医学病理特征成因复杂, 且图像复杂度高, 图像随机误差很大(噪音), 细胞之间的重叠。如传统图像的颜色没有识别要求, 为了训练和计算方便采用的降维处理是将其灰化处理, 然后使用分割算法将形态分离, 该过程极为复杂, 最后使用分类和回归来达到识别^[2]。

2.2 U-Net 分割效果

为了获得更好的分割效果, 本文通过将损失函数赋予相互接触的两个细胞之间的背景标签更高的权重, 完成特征提取。对于细胞核的分割对于轻度重叠可以将细胞核分开, 对于重叠面积过大的细胞核采取舍弃裸露面积小的细胞核。但是还是存在着重叠面积过大无法对其分割的问题。

通过对测试集测试得出的实际效果如图 1 所示:

如图所示效果轮廓明显, 但存在了边缘少量残留和重叠难以检测, 为了更加契合目标检测效果, 保证细胞形态特征的完整性, 通过借鉴 ResNet 中的残差模块概念, 设计了一种残差模块, 通过融合两个 3×3 的卷积层进行一次特征变换操作与一次特征相加。网络训练过程中, 采用 mini-batch 小批量输

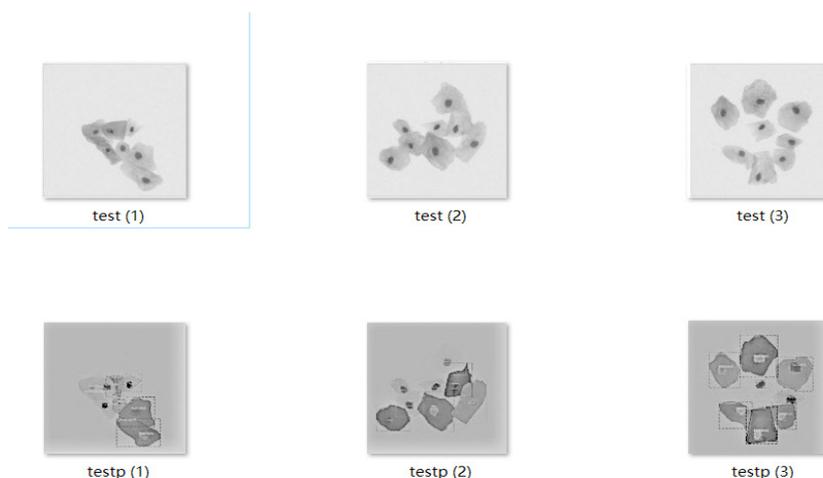


图 1: 实际效果

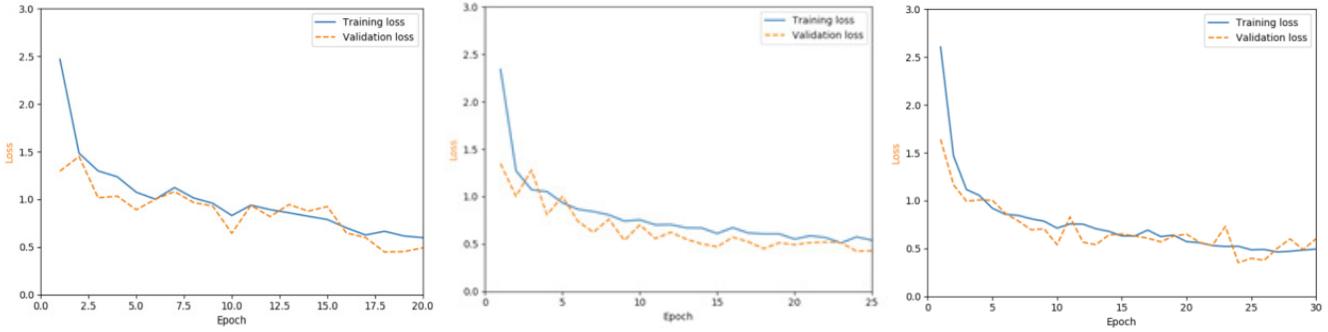


图 2 : Epoch 为 20 loss 曲线 ; 图 3 : Epoch 为 25 loss 曲线 ; 图 4 : Epoch 为 30 loss 曲线

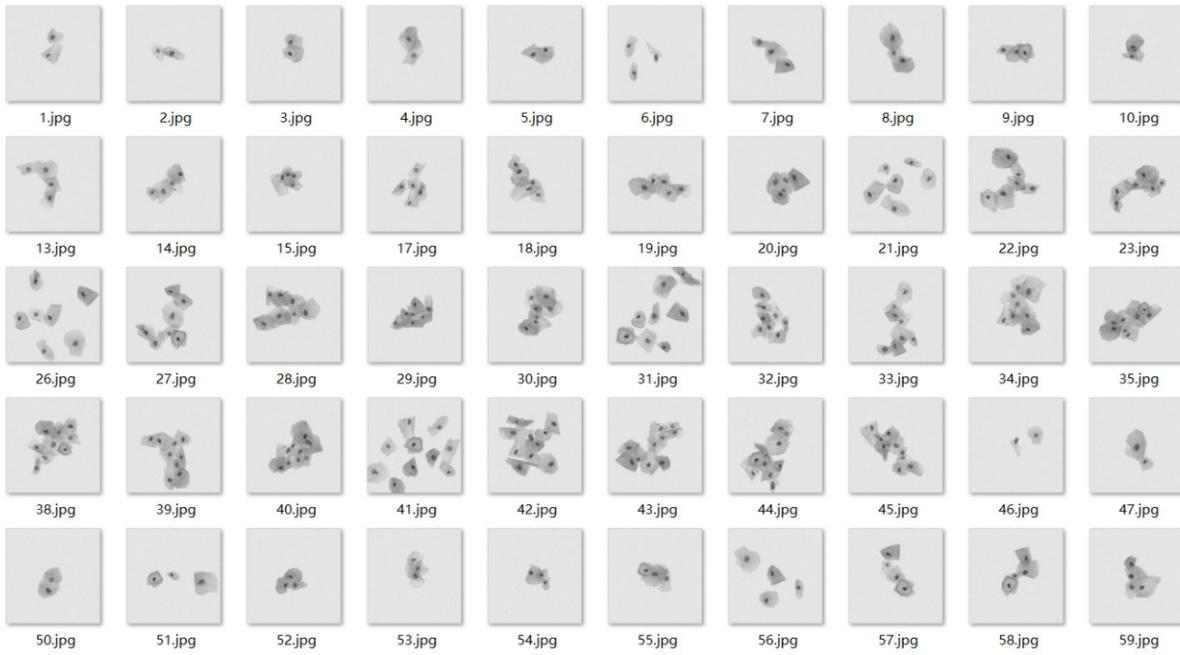


图 5

入和 mini_mask load_image_gt 方法, 在自动迭代调整网络权重时每轮迭代次数为 150, 总的迭代次数为 4000-6000 训练过程中, 将训练轮数设置为 20、25、30 三种不同的方式探讨 epoch 对 loss 值的影响, 在开始时 epoch 的学习率设为 0.00031。(见图 2, 图 3, 图 4)

2.3 数据及数据集标注

本文数据来自 ImageNet 和 IBSI2014 数据集的 1200 张宫颈癌细胞展开对细胞分割任务的研究, 利用了 labelme 进行数据标注工作, 将细胞边缘用线勾出, 进行标注, 在深度学习中, 有更多信息代表着更精确的识别效果, 为了取得更好的效果本文使用了数据集扩增技术, 通过对数据集进行处理, 如使用、缩放、旋转、平移等方式进行扩增, 将数据集充分利用, 针对不平衡数据采取随机欠采样和过采样方法, 进行数据处理, 解决数据不平衡问题。(见图 5)

该数据集包含 1000 余张宫颈癌细胞图像, 结合了宫颈癌细胞的性质, 将细胞分为正常细胞、异常上皮细胞和癌变细胞。为了提高分割的精准度, 同时对干扰细胞进行标注, 虽然会使拟合难度增大但是, 对于细胞的检测与分割却起到了非常重要的作用。

4 总结

本文从宫颈癌的细胞学出发, 通过深度学习, 利用图像分

割网络 U-Net 和图像识别技术 YOLO 系列构建了一套对宫颈癌细胞的检测机制, 做到可以高效且高质地做到对细胞学病理切片的分割与切片, 同时论述了可以用 YOLO 进行宫颈癌细胞检测的可能性, 以此可以提升检测效率, 极大地减少医生对于细胞学检测所消耗的时间, 对于一些基层医院也有较大的意义, 利于提高对宫颈癌的早筛早诊。

参考文献:

[1] 贺淑芳. 宫颈癌筛查精细化管理研究以及 miR-145-5p/FSCN1 致病机制研究 [D]. 南昌大学, 2021.
 [2] Van Grinsven M J J P, van Ginneken B, Hoyng C B, et al. Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images[J]. IEEE transactions on medical imaging, 2016, 35(5): 1273-1284.
 [3] 李成. 基于改进 YOLOv5 的小目标检测算法研究 [J]. 长江信息通信, 2021, 34(09):30-33.

课题项目: 2020 年(第一批)教育部高教司产学合作协同育人项目《基于深度学习的宫颈癌智能预测系统实现》, 项目编号: CKXYRDC003

作者简介:

刘四旗, 2000 年 7 月, 就读于山东协和学院计算机科学与技术专业。