

基于多模型优化的超声图像肿瘤自动识别

古万荣¹ 樊纬江¹ 谢贤芬² 张子烨³ 毛宜军¹ 梁早清¹ 林镇溪¹

1 华南农业大学数学与信息学院 广州 510642

2 暨南大学经济学院 广州 510632

3 华南理工大学数学学院 广州 510641

摘要 随着计算机视觉识别技术的发展,越来越多的研究人员将该技术应用到肿瘤图像的识别上。但由于成本,许多医院仍然采用成本较低的B超等设备,产生了模糊、伪影和多个相似肿瘤噪声区域。目前的方法在清晰图像识别中具有很高的精度,但在超声图像方面却存在低准确度且不稳定的结果,其原因是许多现有算法对模糊、噪声图像误判较高。文中基于R-CNN和PRN的方法快速准确地获取高噪声的超声图像的关键特征,并通过数据增强和形态学滤波的方法确保了识别的稳定性。同时,所提方法还融合了血流信号分类模型增强识别精度。基于实际甲状腺肿瘤图像的数据集可知,提出的方法对比新近算法具有较高的准确性和稳定性。

关键词:深度学习;融合模型;肿瘤识别;神经网络

中图分类号 TN957.52;TP751

Automatic Tumor Recognition in Ultrasound Images Based on Multi-model Optimization

GU Wan-rong¹, FAN Wei-jiang¹, XIE Xian-fen², ZHANG Zi-ye³, MAO Yi-jun¹, LIANG Zao-qing¹ and LIN Zhen-xi¹

1 School of Mathematics and Information, South China Agricultural University, Guangzhou 510642, China

2 School of Economy, Jinan University, Guangzhou 510632, China

3 School of Mathematical, South China University of Technology, Guangzhou 510641, China

Abstract With the development of computer vision recognition technology, more and more researchers apply this technology to the recognition of tumor images. But because of the cost, many hospitals still use low-cost ultrasound and other equipment, resulting in ambiguity, artifacts and many similar tumor noise areas. The present method has high precision in clear image recognition, but it shows low accuracy and unstable result in ultrasonic image. The reason is that many existing algorithms misjudge the modulus and noise image. In this paper, the key features of high-noise ultrasound images are obtained quickly and accurately by R-CNN and PRN methods, and the stability of recognition is ensured by data enhancement and morphological filtering. At the same time, the classification model of blood flow signal is fused to improve the recognition accuracy. Based on the data set of a real Thyroid neoplasm image, the proposed method is more accurate and stable than the new algorithm.

Keywords Deep learning, Fusion model, Tumor recognition, Neural network

多普勒超声是一种常用的医学诊断技术。与其他成像技术相比,其具有实时成像、成本低、无需活检、对患者无损伤等优点。尽管有这些优点,其在超声图像的肿瘤识别方面仍然存在一些问题,如严重斑点、高噪声、低分辨率、对比度不明显以及肿瘤形状无规则。上述缺点导致了肿瘤自动识别过程的准确性和不稳定性的降低。为了克服多噪声图像中的肿瘤识别问题,近年来有许多研究成果。但大多数研究集中在降噪和图像锐化方面,而不是将两者结合起来进行分析,以适应特定的症状和生理特征,以达到适当的水平。此外,目前大多数研究仍然采用传统的特征选择方法,如互信息量和主成分分

析(PCA),有效信息大量丢失。目前人们已经提出了深度学习方法来特征挖掘,并且已经取得了一些成果,但是一些特征提取算法在医学图像处理中并不能很好地应用。特别是它们不适合肿瘤超声图像,准确性和稳定性仍然不足。在现实的医疗中,高误诊率会延误医疗的进行。

在以上背景下,肿瘤图像识别研究至少存在两个主要问题:1)输入数据为低质量的超声图像,换句话说,图像是模糊的,包含很多噪声或模糊的形状边缘;2)主流的特征选择技术,如深度学习和机器学习,对特定的真实世界的甲状腺超声图像适应性较差。为了解决这两个问题,本文提出采用数据

基金项目:广东省自然科学基金(2018A030313437);广东省哲学社会科学项目(GD18CXW01);广东省科技计划项目(2018A070712021);教育部人文社科项目(18YJCZH037);全国统计科学研究重点项目(2019LZ37)

This work was supported by the Guangdong Natural Science Foundation Project (2018A030313437), 13th Five-year Plan Project of Philosophy and Social Science in Guangdong Province (GD18CXW01), Guangdong Science and Technology Program Project (2018A070712021), Ministry of Education Humanities and Social Sciences Research Youth Fund Project (18YJCZH037) and 2019 National Statistical Science Research Key Project(2019LZ37).

通信作者:毛宜军(yijunmao@163.com)

增强、形态滤波等方法保证数据质量,有效地利用数据质量进行特征提取,然后利用 R-CNN 和 RPN 多层深度网络充分挖掘图像海量的潜在特征,这样就可以挖掘出肿瘤区域和噪声区域之间的差异。同时,该算法可以更准确地锁定肿瘤区域,降低误报率。由于该方法的高精度和高效率,肿瘤区域可以被快速锁定,未来可以将其应用于动态视频甲状腺癌的识别中。

1 相关技术背景

医学图像识别一般包括 3 个主要步骤:图像预处理、特征抽取和对象识别。

一般来说,输入的图像在进行图像识别之前需要进行去噪、对比度增强、锐化、区域分割等处理。本文中,超声图像承载着大量的噪声信息,图像的模型和预处理是必不可少的。Yu-Len 等^[2]提出了一种乳腺癌的分水岭分割方法,Gaetano 等^[3]对遥感图像的分割进行了研究。本文主要研究医学图像的去噪问题。一些研究建议对图像分割进行特征缩减。越来越多的研究集中在模糊医学图像的分割上^[6]。然而,现有的预处理方法在清晰彩色图像处理方面有很好的效果,并且在许多应用领域都有很好的性能^[7-8]。甲状腺癌图像只有黑白模具和边缘,有更多的脉冲噪声干扰,因此现有的方法大多不适合本文的应用。

本文基于形态学和纹理特征,从图像中识别良恶性结节外观,从而实现识别算法^[9]。其形态特征包括结节的方向、形状、回声、边缘等,其在许多应用中得到应用^[10-11]。纹理特征包括超声表现和回波模型。传统的特征提取方法主要基于形状^[12]、颜色^[13]和纹理^[12,14],其常用的方法有 SIFT 和 HOG^[15]。

分类器或聚类器用于识别病变区域的分割图像。在进行分类之前,需要筛选出分类性能较强的特征,以提高分类器的泛化能力,并降低计算维数,避免维数灾难。由于良恶性肿瘤的特征具有自相关性和交互性,非线性映射方法可以作为分类器或聚类器。常用的方法有 BP 算法、决策树算法、FLD 算法、SVM 算法、贝叶斯分类器算法、SOM 算法等。然而,传统的机器学习特征提取方法过于依赖经验。在许多新的应用程序场景中,没有太多可以学习的经验。因此,这类场景的特征提取存在较大的误差。

为了解决传统机器学习方法在特征挖掘中的不足,本文提出了一种新的深度学习方法,将深度学习算法应用到医学图像识别和图像分割分析中。在医学图像分割领域,不同的分割方法通常与不同的机器学习或深度学习方法相结合,如 Mocanu 等提出的基于特征学习和集成学习分层视网膜血管分割的方法^[16],该方法也可用于跟踪心室内膜超声图像分割。此外,3D CNN 和全卷积神经网络常用于语义分割和图像分割。例如 Korez 等利用 3D CNN 对 MR 图像进行脊柱分割,Moeskops 等在 MRI 图像中对大脑和胸部肌肉进行分割。在分类和识别领域,有监测视网膜技术的糖尿病患者的转移学习^[19]、皮肤癌的分类转移学习^[20]、DCE-MRI 的自动分类稀疏自动编码器^[21]和检测和评价膝关节关节炎分级使用的卷积神经网络^[22]等方法。随着计算机视觉技术的不断发展,深度学习模型在辅助诊断方面也取得了很大的进展,如利用 Bengtsson 等^[23]的 Deep Belief 网络和 SVM 对 pap-spar

进行分析,对宫颈癌进行识别;Kooi 等^[24]基于 CNN 对乳腺良恶性肿瘤进行识别。这些方法远远优于传统的计算机辅助诊断方法。此外,深度学习算法常用于判断前列腺疾病、乳腺癌、肺癌以及研究细胞微核检测等。

然而,大多数甲状腺肿瘤的超声图像分割方法都不能实现端到端的分割,这就需要人工标注干预。首先,感兴趣区域(ROI)是由具有丰富临床经验的医生在原始超声图像标记。然后利用各种边缘分割算法提取感兴趣区域的轮廓,这就需要大量的人力和精力,需要医生具有较高的临床经验和个人能力。因此,如何能高效且高精度地实现肿瘤感兴趣区域的自动识别具有重要意义。

2 多模型融合的方法

目前深度学习图像识别中的目标检测方法主要有两种:1)基于感兴趣区域提取的两阶段模型,如 R-CNN, SPP-Net, Fast R-CNN, Faster R-CNN, R-FCN 等;2)不提取 ROI 的一阶段模型,如 YOLO, SSD 等。作为计算机视觉领域的重要组成部分,这些图像识别方法主要是利用细胞神经网络进行特征学习,结合一些传统的视觉方法实现目标检测。

超声图像的成像原理是利用超声波穿透人体器官,产生黑白图像。一般来说,噪声区域不仅模糊而且易受干扰,容易形成噪声区域。图 1 给出甲状腺肿瘤超声图像示例。



图 1 甲状腺肿瘤的超声图像

Fig. 1 Ultrasound image of thyroid neoplasm

为了解决甲状腺癌的图像识别问题,本文提出的基本技术框架如图 2 所示。

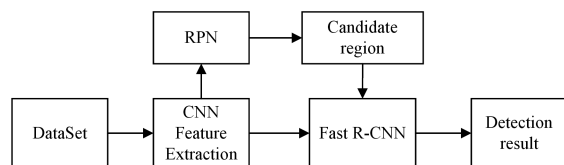


图 2 基于 Fast R-CNN 的甲状腺肿瘤图像识别模型结构

Fig. 2 Structure of thyroid neoplasm image recognition model based on Fast R-CNN

基于快速 R-CNN 的思想,本文提出分别构造 RPN 和快速 R-CNN 来提取和检测图像中的区域。采用不同的卷积网络如 ZF, VGG, ResNet50 等进行特征提取。最有效的网络将用作 RPN 和 Fast R-CNN 的共享网络。将 RPN 生成的候选区域输入到 Fast R-CNN 中,用于软最大分类和边界回归。

图 3 展示了卷积神经网络的基本结构,主要包括输入层(原始超声图像)、卷积层、池层、全连接层和分类层(软最大层)。

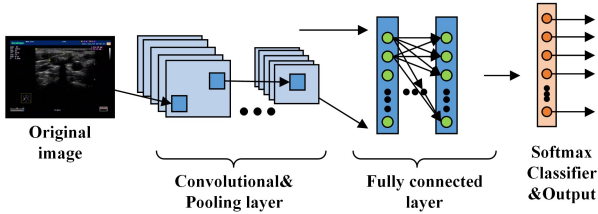


图3 卷积神经网络基本结构

Fig. 3 Structure of convolutional neural network

1) 卷积层: 卷积层是卷积神经网络的重要组成部分, 负责卷积神经网络的许多计算任务。每个卷积层包含多个卷积核(滤波器), 通过平移和滑动扫描整个图像。此外, 这些卷积核可以通过向前传播和向后传播进行自学习。通过卷积核进行图像扫描的结果称为特征图或特征映射(特征映射), 表示每个空间位置对卷积核输入的响应。公式如下:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * w_{ij}^l + b_j^l\right) \quad (1)$$

以二维图像 i 作为输入, 使用二维卷积核 k 进行卷积运算后得到的特征如下:

$$\begin{aligned} \mathbf{S}(i, j) &= (\mathbf{I} * \mathbf{K})(i, j) \\ &= \sum_m^n \mathbf{I}(m, n) \mathbf{K}(i-m, j-n) \end{aligned} \quad (2)$$

一般情况下, 卷积层的结构可由卷积核的大小、卷积核的个数、步长和填充 4 个超参数确定。不同的卷积核可以提取不同的特征。步长决定了卷积核每个运动的长度, 填充可以提取图像的边缘信息, 这基于两个原则: 局部感知和参数共享。

2) ROI 池化层: ROI 池化层使用最大池化操作将 ROI 特征转化为维度较少并且固定的特征图, 大小为 $H \times W$ (如 7×7), 其中 H 和 W 都是超参数, 需要事先设置。每个 ROI 由 4 个参数决定 (r, c, h, w) , 其中 (r, c) 为左上角坐标, (h, w) 为高和宽。ROI 可以利用 $h/H \times w/W$ 大小的尺寸作为滑动窗口, 将大小为 $h \times w$ 的 ROI 转化为 $H \times W$ 的特征图, 从而使得 ROI 池化层能够处理任意尺度的特征图。

3) 损失函数: Fast R-CNN 同样也有两个损失, 分别为分类损失和回归损失, 这种多任务损失通常将所有的损失叠加起来, 从一个损失权重衡量参数来表示。其损失函数定义为:

$$L(p, u, t^u, v) = L_{cls}(p, u) + \lambda [u \geq 1] L_{loc}(t^u, v) \quad (3)$$

其中, $p = (p_0, p_1, \dots, p_K)$ 表示每个类别的预测概率; K 为类别个数, 一共包括 $K+1$ 个类别, 第 0 个类别为背景; u 为真实边框的类别, $t^u = (t_x^u, t_y^u, t_w^u, t_h^u)$ 为第 u 个类别的预测边框的坐标, $v = (v_x, v_y, v_w, v_h)$ 为第 u 类的真实边框的坐标; $[u \geq 1]$ 是指示符函数, 当 $u \geq 1$ 时其值为 1, 否则为 0; λ 为权重衡量参数; L_{cls} 为类别的损失函数, L_{loc} 为边框回归损失函数, 公式如下:

$$L_{cls}(p, u) = -\log p_u \quad (4)$$

$$L_{loc}(t^u, v) = \sum_{i \in \{x, y, w, h\}} smooth_{L1}(t_i^u - v_i) \quad (5)$$

本文采用可以自学习的 RPN 网络生成 ROI, 如图 4 所示, RPN 最终有两个输出, 因此该网络可以看成是一个二类分类器。由于每一个 Anchor 输出一个二类分类值, 因此使用了 k 个 Anchor, 最终输出的是长度为 $2k$ 的向量。而另一个为回归输出, 每个 Anchor 对应一个长度为 4 的向量 (x, y, w, h) , 其中 (x, y) 为 Anchor 的中心坐标, (w, h) 为 Anchor 的长和宽, 则一共包含 $4k$ 个输出。

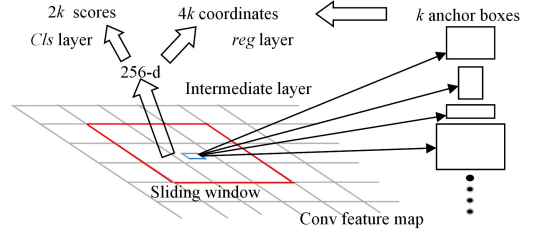


图4 RPN 网络结构图

Fig. 4 RPN network structure

根据上面的定义, RPN 的多任务损失函数 (Loss Function) 见式 (6):

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_t L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_t p_i^* L_{reg}(t_i, t_i^*) \quad (6)$$

其中, i 为每个批量 Anchor 的下标, p_i 为第 i 个 Anchor 是目标类的预测概率, 如果 p_i^* 是目标类别, 则 p_i^* 为 1, 否则为 0; t_i 是一个长度为 4 的向量, 表示预测的边框的参数化坐标, t_i^* 是一个长度为 4 的向量, 表示真实的边框的坐标; N_{cls} 为小批量采样对的大小, N_{reg} 为 Anchor 的数量; λ 为平衡损失函数的权重, 通常设为 10; L_{cls} 为类别的损失函数, L_{reg} 为回归损失函数。

$$L_{cls}(p_i, p_i^*) = -\log[p_i^* p_i + (1-p_i)(1-p_i^*)] \quad (7)$$

$$L_{reg}(t_i, t_i^*) = smooth_{L1}(t_i - t_i^*) \quad (8)$$

$$smooth_{L1}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{if } |x| \geq 1 \end{cases} \quad (9)$$

损失函数中当 $p_i^* = 1$ 时才有损失, 否则不计算损失。 t_i 和 t_i^* 的计算见式 (10)~(17):

$$t_x = (x - x_a) / w_a \quad (10)$$

$$t_y = (y - y_a) / h_a \quad (11)$$

$$t_w = \log(w / w_a) \quad (12)$$

$$t_h = \log(h / h_a) \quad (13)$$

$$t_x^* = (x^* - x_a) / w_a \quad (14)$$

$$t_y^* = (y^* - y_a) / h_a \quad (15)$$

$$t_w^* = \log(w^* / w_a) \quad (16)$$

$$t_h^* = \log(h^* / h_a) \quad (17)$$

其中, x, y, w, h 分别是预测边框的中心坐标, 以及长和宽; x^*, y^*, w^*, h^* 分别是真实边框的中心坐标, 以及长和宽; x_a, y_a, w_a, h_a 分别是 Anchor 边框的中心坐标, 以及长和宽。

4) 训练方式: RPN 的训练可以联合后向传播和 SGD 实现端到端训练, 每次从一张图像中抽取小批量的样本, 每批样本包括正负样本, 由于负样本的数量远远大于正样本的数量, 所以在采样时可以偏向负样本, 例如按照 1:1 的正负比例采样 256 个样本, 如果正样本不满足 128 个, 则可以使用负样本进行填充。

RPN 的一个卷积层可以使用标准差为 0.01 均值为 0 的高斯分布初始化, 而其他的层可以用已经在大规模图像数据库 (Pascal VOC, ImageNet 等) 中训练好的预训练模型 (如 ZF, VGG, ResNet) 初始化, 然后进行微调。

用预训练模型分别初始化 RPN 和 Fast R-CNN 网络, 接着对 RPN 进行训练, 训练完生成 ROI, 然后用这些 ROI 训练 Fast R-CNN 网络, 再用 Fast R-CNN 网络初始化 RPN, 与一开始初始化 RPN 不同的是, 这里固定共享卷积层的参数, 只是微调 RPN 的参数, 接着对 RPN 训练来初始化 Fast R-CNN 网络, 使得 RPN 和 Fast R-CNN 网络能够交替训练。这种训

练方式效果比较好,但是耗时较长。

这种训练方式尝试把 RPN 和 Fast R-CNN 网络融合到一个网络里,在前向传播时 Fast R-CNN 网络使用 RPN 产生的 ROI 进行训练,在反向传播时共享层的梯度来源于 RPN 和 Fast R-CNN 网络。这种情况忽略了边框坐标的梯度,是一种近似的训练,但能够减少 25%~50% 的训练时间。

本文中,RPN 网络损失函数的参数设置为: $N_{cls}=256$, $N_{reg}=2400$, $\lambda=10$ 。ROI 池化层参数设置为: $H \times W=7 \times 7$ 。Fast R-CNN 网络损失函数的参数设置为: $\lambda=1$ 。

3 实验与分析

本文所有的甲状腺结节超声图像均来自中山大学附属肿瘤医院头颈门诊以及头颈二区科。其中包括 16 个病例,368 幅超声图像,包括良恶性图像。但是由于未标注出肿瘤的良恶性,所以本文只考虑肿瘤位置的识别,不包括良恶性的识别。实验将 368 幅超声图像按照 5:2:3 的比例划分为训练集、验证集以及测试集。

在实验中使用近似联合训练方法对模型进行训练,采用基于动量的随机梯度下降优化算法,其梯度更新方式为:

$$v_t = \gamma v_{t-1} + \eta \times g(\theta)$$

$$\theta = \theta - v_t$$

其中, θ 为需要更新梯度的参数; v_t 为 t 时刻的动量; γ 是动量参数且 $\gamma \in [0,1]$; η 为学习率; $g(\theta)$ 为 θ 的梯度。在实验中,将 γ 设置为 0.9,初始学习率 η 设置为 0.001,每迭代 50000 次衰减 0.0005。

3.1 不同模型的准确度对比

根据上述的实验设置,我们分别采用 ZF,VGG16,ResNet50 3 种模型作为基线方法,并以不同的迭代次数对模型进行训练。由于实验数据集数据量不大,所以均采用网络在 ImageNet 上的预训练模型,然后进行微调。实验结果对应的 mAP (Mean Average Precision) 如表 1 和图 5 所示。

表 1 不同迭代次数以及不同网络模型的 mAP

Table 1 mAP for different iterations and different network models (单位:%)

| 网络模型 | 迭代次数 | | |
|----------|-------|-------|--------|
| | 1 000 | 5 000 | 10 000 |
| ZF | 0.691 | 0.762 | 0.697 |
| VGG16 | 0.732 | 0.786 | 0.786 |
| ResNet50 | 0.720 | 0.799 | 0.772 |

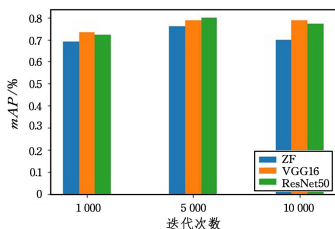


图 5 不同迭代次数以及不同网络模型的 mAP

Fig. 5 mAP histogram with different iterations and different network models

从表 1 和图 5 中可以获取的主要信息有:

(1)VGG16 和 ResNet50 的效果要比 ZF 好,主要原因是 ZF 为小型网络,只有 7 层,而 VGG16 和 ResNet50 均为大型网络,层数达到了 16 层(VGG16)或者 50 层(ResNet50)。通过增加网络层数可以使得网络提取到更深层语义更丰富的信

息,从而提高模型的准确率。

(2)迭代次数为 5000 时效果最好,而迭代次数为 10000 次出现了下降的情况,主要原因是迭代 10000 次时模型出现了过拟合,而迭代 1000 次时欠拟合。由于实验数据量太少,所以整个模型很容易收敛,迭代次数不宜过多。

图 6 显示了 3 种网络模型的 $Loss$ 随着迭代次数的变化折线图。可以发现,3 种网络均在 3000 次迭代时基本收敛,随后基本趋于平稳状态,其中 VGG16 会出现较大的起伏,而 ResNet50 效果最好,不仅 $Loss$ 最小,而且最平稳。所以,后期的系统我们使用 ResNet50 作为 Faster R-CNN 中的特征提取网络。图 6 也可以指导后期网络模型的训练次数,以便减少时间开销和防止过拟合。

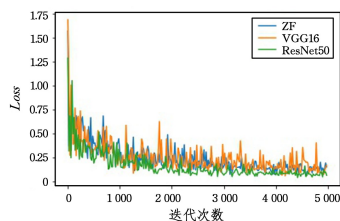


图 6 不同网络模型的 $Loss$ 折线图

Fig. 6 Polyline diagrams of $Loss$ for different network models

3.2 实验分析和调整

根据表 1 可知,3 种模型最好的结果只有 79.9%,还有改进的空间。由于数据集较小,很多因素都会影响实验结果,如数据集划分比例、迭代次数、学习率、数据增强等。下面主要从数据集划分比例、数据增强、学习率因素改进本文模型。

(1)数据集划分比例

由于数据集太小,很容易收敛,不同的划分会使得训练集和测试集比例不同,对模型的结果也会产生影响。如果训练集比测试集大得多,则可能出现过拟合;如果训练集比测试集小或者相差不大,则可能出现欠拟合。因此,我们将迭代次数设置为 3000,为了减少训练时间使用 ZF 网络进行测试,采用 4 种不同的划分比例,分别为 2.5:2.5:5,5:2:3,4:4:2,4:3:3。实验结果如表 2 和图 7 所示。

表 2 不同的数据划分比例下的 mAP

Table 2 mAP of different data partition ratios

| 网络模型 | 划分比例(训练集:验证集:测试集) | | | |
|------|-------------------|-------|-------|-------|
| | 2.5:2.5:5 | 5:2:3 | 4:4:2 | 4:3:3 |
| ZF | 0.775 | 0.762 | 0.848 | 0.778 |

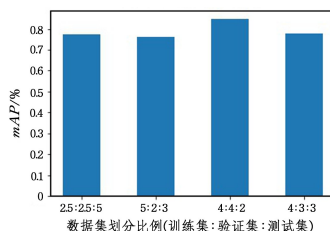


图 7 不同数据集划分比例下的 mAP 柱状图

Fig. 7 mAP histogram with different data set partition ratios

从表 2 和图 8 可以看出,不同的数据集划分比例的确会影响模型的最终效果。可以看出,使用 4:4:2 的划分比例时准确率从 76.2%(5:2:3)提高到了 84.8%。因为数据量小,将更多的数据用于训练更有助于模型学习更多的特征,而且

训练集和验证集的比例比较均匀,能够防止模型过拟合,从而提高模型的泛化能力。这4种划分比例的 $Loss$ 变化折线图如图8所示,可以看出不同的划分比例对收敛速度影响不大,基本都是在迭代3000次时基本收敛,达到稳定状态。

从表2和图8可知,数据集太小,如果用于训练的数据太少使得模型学习到的特征太少,则不容易表达出具体的语义信息;而用于训练的数据太多测试数据太少则使得模型很容易过拟合,对测试数据的表达能力不强,泛化能力弱。为了提高模型的泛化能力,我们使用数据增强技术,例如水平翻转、垂直翻转、随机截取、固定截取、尺度变换、旋转、色彩抖动、边缘增强等。有效的数据增强不仅能扩充训练样本的数量,还能增强样本的多样性,一方面可以避免过拟合,另一方面会带来模型性能的提升,增加图像的平移不变性和旋转不变性,增加卷积神经网络对物体尺度和方向上的鲁棒性。

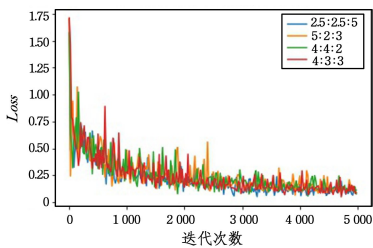


图8 数据集不同划分比例的 $Loss$ 变化折线图

Fig. 8 Polyline diagram of $Loss$ variations for different partition proportions of the dataset

LeNet 在测试时将测试图像归一化,按照固定大小截取4个角以及中心一共5个区域,并将图片水平翻转,增加图像的旋转不变性和平移不变性,提高模型的泛化性能。经过这些操作后,图像数据集可以扩充为原来的10倍,然后用投票机制选择最合适的类别。本文借鉴这种思想,采用随机截取再加上水平翻转的方式增加数据集,其中随机截取的大小为 224×224 。除此之外,还使用图像处理中的边缘增强增加超声图像的对比度。本文使用上述方法对数据集进行数据增强,在4种不同的数据集划分比例中采用ZF网络做了实验,实验结果如表3和图9所示。

表3 数据增强在不同数据集划分下的 mAP 比较

Table 3 Data enhancement mAP comparison under different data set partition (单位: %)

| 是否进行 数据增强 | 划分比例(训练集:验证集:测试集) | | | |
|--------------|-------------------|-------|-------|-------|
| | 2.5:2.5:5 | 5:2:3 | 4:4:2 | 4:3:3 |
| 否 | 0.775 | 0.782 | 0.848 | 0.778 |
| 是 | 0.862 | 0.902 | 0.802 | 0.889 |

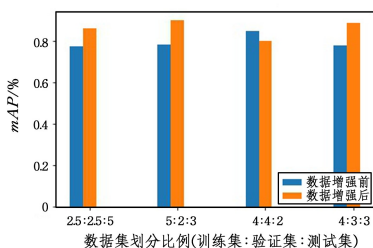


图9 不同数据集划分比例下数据增强前后的 mAP

Fig. 9 mAP histogram before and after scaling of different datasets

从实验结果可知,使用数据增强后准确率普遍上升,4种

划分比例的准确率都在80%以上,其中划分比例为5:2:3的准确率从原先的78%提高到了90%,而未使用数据增强的准确率基本都在70%~80%。使用数据增强后,不仅扩充了样本的数量,同时增加了图像的平移不变性和旋转不变性,提高了模型的鲁棒性和准确度。除此之外,本文对4种划分比例下使用数据增强后的 $Loss$ 收敛变化做了分析,从图10可以看出,数据增强后4种不同划分比例下的 $Loss$ 收敛速度基本保持一致,同样在迭代3000次左右达到稳定状态。图11为不同划分比例下数据增强前后的 $Loss$ 变化,4种划分比例下数据增强前后的 $Loss$ 收敛速度也基本不变。

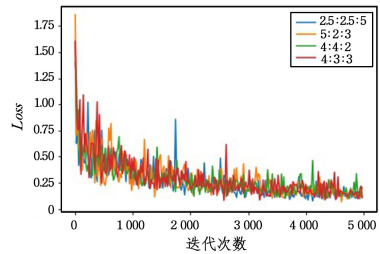


图10 数据集增强前后数据集不同划分比例下的 $Loss$ 变化折线图

Fig. 10 Polyline diagram of $Loss$ changes before and after data set enhancement with different partition ratios

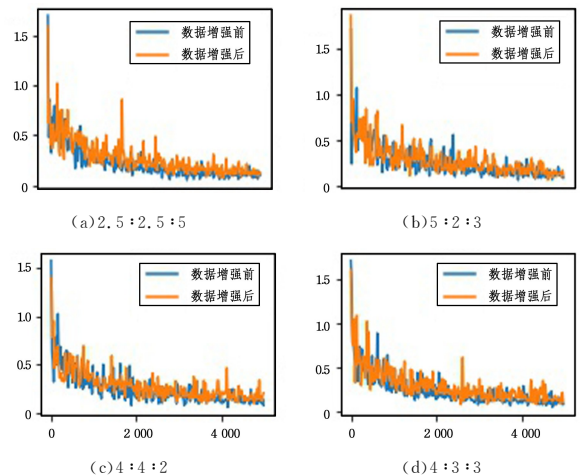


图11 数据增强前后不同划分集的 $Loss$ 变化趋势

Fig. 11 Trends in $Loss$ changes for different partition sets before and after data enhancement

前文采用随机截取的方式做数据增强,但是其缺点是截取出的信息很有可能不完整,如图12所示,所截取的肿瘤只有一半的区域,属于无效信息,这会对模型的识别造成一定的影响。因此我们采用形态学滤波过滤超声图像周边无效的噪声信息,主要提取超声图像中间区域的有效信息。



图12 超声图像随机截取前后的图像

Fig. 12 Ultrasound images before and after random sampling

形态学滤波可用于检测图像中指定的特征,抑制图像中的噪声信息等。膨胀、腐蚀、开闭运算是形态学中基本的变化形式,也有一些高级的转化,例如凸包、连通区域标记、删除小

块区域等。我们采用的基于形态学滤波截取图像流程如图 13 所示,首先对原始图像进行边缘增强,增强图像的对比度,然后利用形态学中的凸包求出图像的最大边界(联通区域),然后对图像进行截取,这是一种具有指导性的截取,而不是随机截取,图 14 为利用形态学滤波对某图像进行截取的过程。

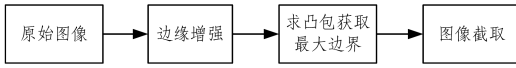


图 13 基于形态学滤波截取图像流程图

Fig. 13 Flow chart of image capture based on morphological filtering

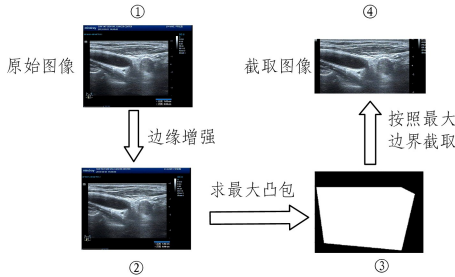


图 14 基于形态学滤波截取图像示例

Fig. 14 Sample image capture based on morphological filtering

基于上面利用形态学滤波截取的方法,我们再次对数据集进行数据增强,实验结果如表 4 和图 15 所示。

表 4 不同数据集划分比例下基于两种不同方式进行数据增强前后的 mAP 比较

Table 4 mAP comparison of different data set partition ratios based on two different ways of data enhancement

| 是否数据增强 | 划分比例(训练集:验证集:测试集) | | | |
|------------|-------------------|-------|-------|-------|
| | 2.5:2.5:5 | 5:2:3 | 4:4:2 | 4:3:3 |
| 否 | 0.775 | 0.782 | 0.848 | 0.778 |
| 是(随机截取) | 0.862 | 0.902 | 0.802 | 0.889 |
| 是(形态学滤波截取) | 0.902 | 0.877 | 0.899 | 0.896 |

(单位:%)

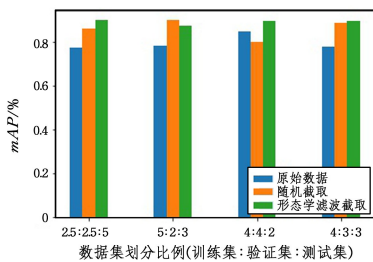


图 15 不同数据集划分比例下基于不同截取方式的 mAP

Fig. 15 mAP histogram with different data set partition ratios based different truncation methods

从表 4 和图 15 可知,采用基于形态学滤波截取的数据增强对模型的准确率有一定的提升,4 种划分比例的准确率基本接近 90%,比基于随机截取的数据增强要好一些,而且更加稳定。基于形态学滤波的数据增强的 4 种划分比例的准确率基本相近,而基于随机截取的数据增强在 4 种划分比例下的准确率起伏较大,例如划分比例为 4:4:2 的准确率波动比较大。除此之外,我们对 3 种不同截取方式的 4 种划分比例的 $Loss$ 收敛变化做了分析,如图 16 所示。可以看出,不同截取方式对于 4 种划分比例的收敛速度基本没影响,基本在 3000 次迭代时达到收敛。

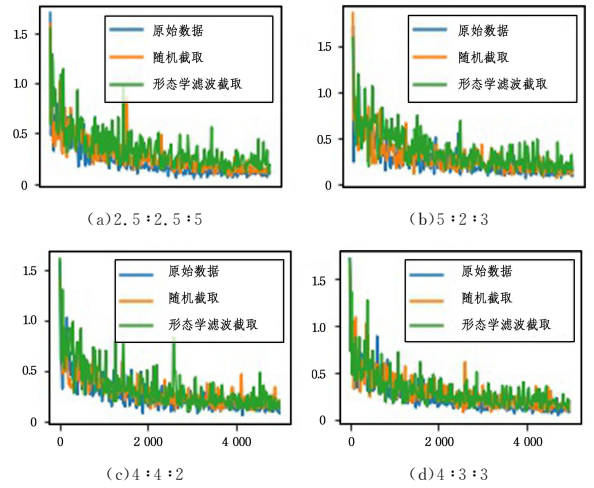


图 16 不同截取方式对于划分比例的 $Loss$ 变化折线图

Fig. 16 Polyline diagram of $Loss$ variations for partition ratios for different truncation methods

(2) 学习率和动量

除了上述的 3 种不同改进方法,一些超参数也会对模型的准确率有影响,例如学习率、动量、滑动窗口大小、平衡权重因子等。下面我们对学习率进行分析,一般来说,学习率越大,网络学习的速度越大。学习率直接影响模型能够以多快的速度收敛到局部最小值,如果学习率太大,有可能取不到极值,会在某一位置反复震荡,而学习率过小,则网络收敛速度太慢,还有可能陷入局部最优。选择合适的学习率不仅能加快收敛速度,减少网络训练时间,还会提高模型的准确率。

下面将数据集按照 4:3:3 的比例划分为训练集、验证集以及测试集,使用基于 ZF 网络的 Faster R-CNN 对基于形态学滤波截取的数据集分别采用 0.1,0.005,0.001,0.0001 的学习率进行 5000 次迭代训练,实验结果如表 5 和图 17 所示。

表 5 不同学习率的 mAP

Table 5 mAP comparison of different learning rates

| 网络模型 | 学习率 | | | |
|------|------|-------|-------|--------|
| | 0.1 | 0.005 | 0.001 | 0.0001 |
| ZF | None | 0.902 | 0.889 | 0.760 |

(单位:%)

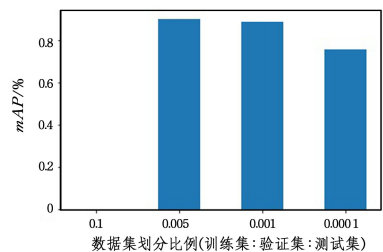


图 17 不同学习率的 mAP

Fig. 17 mAP bars for different learning rates

从表 5 和图 17 发现,当学习率为 0.1 时结果为 None,原因是学习率过大造成浮点数溢出。而学习率为 0.005 时准确率最高,而学习率为 0.0001 时准确率下降。我们对 3 种学习率的 $Loss$ 收敛速度进行分析,如图 18 所示。可以看出,当学习率为 0.0001 时网络收敛速度非常慢,在迭代 5000 次时还未收敛完成,而学习率为 0.001 或者 0.005 时在迭代 5000 次左右时已经基本收敛达到稳定状态。

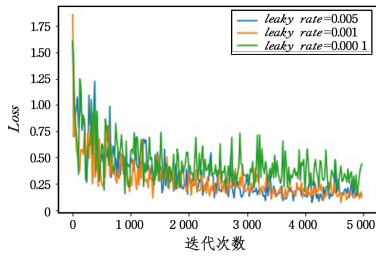


图 18 不同学习率的 Loss 变化折线图

Fig. 18 Polyline graphs of Loss changes for different learning rates

3.3 实验总结

上文对分别使用不同的卷积神经网络作为 Faster R-CNN 的特征提取部分,发现模型的结果不太理想,然后分别采用不同的划分比例,基于随机截取以及形态学滤波截取两种数据增强方法和不同的学习率对模型进行了改进,改进后的模型要比原模型更好,准确率更高,而且模型也相对更稳定。从实验中也可以发现:

(1)更深的网络有助于模型的学习,通过增加网络层数可以使得网络提取到更深层语义更丰富的信息,从而提高模型的准确率。

(2)当数据集较小时,不同的数据集划分比例会对模型的训练造成一定的影响,当划分比例过大时容易造成过拟合,当划分比例过小时容易造成欠拟合,都不利于提高模型的泛化能力。我们可以采取数据增强的方式增大数据集,通常的方式有水平翻转、垂直翻转、随机截取、固定截取、尺度变换、旋转、色彩抖动、边缘增强等。而随机截取的方式有可能使得截取到的信息不完整,可以采用基于形态学滤波的图像处理方式对图像进行截取,减少信息不完整的情况。通过数据增强往往能够提高模型的准确率,增强对图像形变的鲁棒性,提升模型的泛化能力,而基于形态学滤波的数据增强使得模型在提高准确率的同时更加稳定。

(3)一些超参数也会影响模型的效果,例如学习率会对模型的收敛速度产生影响。学习率过大时会造成网络产生振荡,取不到极值,还可能出现溢出,而学习率过小时网络收敛速度过慢,训练时间太长,而且还可能陷入局部最优,所以用合适的学习率对模型进行训练,才能取得更好的效果。

结束语 本文以甲状腺结节超声图像为主要研究对象,基于深度学习图像目标检测技术中的 Faster R-CNN 算法对甲状腺超声图像进行肿瘤定位以及基于 CaffeNet 卷积神经网络对超声图像进行分类,采用深度学习框架 Caffe 以及 Django 框架构建肿瘤图像识别系统,用于识别甲状腺肿瘤超声图像。本文研究的重点主要有:

(1)分析传统的医疗图像识别技术,试图将深度学习目标检测技术应用于医疗图像领域。由于传统技术需要手工提取精确的特征,耗费精力大,而且精确度不高,利用深度学习中的卷积神经网络可以自学习图像特征的特点,通过增加网络层次获取更抽象和丰富的图像语义信息,不仅避免了数据预处理中的特征提取过程,还可以端到端地学习训练,不需要人工干预。

(2)基于 Faster R-CNN 的思想,采用 RPN 网络对超声图像产生用于 Fast R-CNN 判断的候选区域,提高了传统候选区域提取的时间,利用近似联合训练方法减少了网络收敛时

间,并对不同的卷积网络、数据集划分形式、学习率等因素进行对比分析。实现发现,由于 CNN 底层是对图像的边缘信息和轮廓信息的提取,更高层是对图像语义信息和本质信息的抽象,所以利用深层 CNN 更能表征图像语义特征,准确率更好;由于数据集小,不同的数据集划分形式也会对模型产生一定的影响,采用数据增强技术之后可以增加样本的多样性,提高模型的准确率以及泛化能力。

(3)在数据增强中,随机截取的方式可能会造成图像信息不完整,无效的噪声信息也会影响模型的判断。基于传统处理技术,提出基于形态学滤波截取方法,利用形态学滤波中的凸包算法求出图像的最大凸包,根据凸包的最大边界截取图像,避免了图像信息截取不完整的情况,而且能够去除超声图像周边的无效噪声信息。

在以后的工作中,我们将继续研究该算法在 Map-Reduce 框架下的分布式训练和分类。该方法能够快速获得识别结果,提高了该研究的应用前景。此外,我们还可以将该技术框架进一步应用到超声视频识别中。

参考文献

- [1] ALISON N J, DJAMAL B. Ultrasound image segmentation: a survey[J]. IEEE Transactions on Medical Imaging, 2006, 25(8): 987-1010.
- [2] YU-LEN H, DAR-REN C. Watershed segmentation for breast tumorin 2-d sonography[J]. Ultrasound in Medicine & Biology, 2004, 30(5): 625-632.
- [3] GAETANO R, MASI G, POGGI G, et al. Markercontrolled watershed-based segmentation of multiresolution remote sensing images[J]. IEEE Transactions on Geoscience & Remote Sensing, 2015, 53(6): 2987-3004.
- [4] NAIMI H, ADAMOUMITICHE A B H, MITICHE L. Medical imagedenoising using Dual Tree Complex Thresholding Wavelet Transformand Wiener filter[J]. Journal of King Saud University-Computer and Information Sciences, 2015, 27(1): 40-45.
- [5] PHAM V N, LONG T N, NGUYEN T D. Feature-reduction fuzzyco-clustering algorithm for hyperspectral image segmentation[C]// IEEE International Conference on Fuzzy Systems, 2017.
- [6] ZHAO Y, RADA L, CHEN K, et al. Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images[J]. IEEE Transactions on Medical Imaging, 2015, 34(9): 1797-1807.
- [7] FAN X, JU L, WANG X, et al. A fuzzy edge-weighted centroidalvoronoi tessellation model for image segmentation[J]. Computers & Mathematics with Applications, 2016, 71(11): 2272-2284.
- [8] WANG H, HUANG T Z, XU Z, et al. An active contour modeland its algorithms with local and global gaussian distribution fitting energies[J]. Information Sciences An International Journal, 2014, 263(1): 43-59.
- [9] KONUR U, GRGEN F S, VAROL F, et al. Computer aided detection of spina bifida using nearest neighbor classification with curvature scale space features of fetal skulls extracted from ultrasound images[J]. Knowledge-Based Systems, 2015, 85(C): 80-95.
- [10] FAROKHI S, SHEIKH U U, FLUSSER J, et al. Near infrared

- face recognition using zernike moments and hermite kernels[J]. *Information Sciences An International Journal*, 2015, 316(C): 234-245.
- [11] WENG T, YUAN Y, LING S, et al. Clothing image retrieval using color moment[C]// *International Conference on Computer Science & Network Technology*. 2014.
- [12] YAMAGUCHI J, YONEYAMA A, MINAMOTO T. Automatic detection of early esophageal cancer from endoscope image using fractal dimension and discrete wavelet transform[C]// *International Conference on Information Technology-new Generations*. 2015.
- [13] DAS J, ROY H. Human face detection in color images using hsv-color histogram and wld[C]// *International Conference on Computational Intelligence & Communication Networks*. 2015.
- [14] XU X, QUAN C, REN F. Facial expression recognition based on gabor wavelet transform and histogram of oriented gradients [C]// *IEEE International Conference on Mechatronics & Automation*. 2015.
- [15] EMMANUEL A, OLUGBARA O O. Lung cancer prediction using neural network ensemble with histogram of oriented gradient genomic features[J]. *The Scientific World Journal*, 2015, 2015: 786013.
- [16] MOCANU D C, AMMAR H B, LOWET D, et al. Factored four way conditional restricted Boltzmann machines for activity recognition[J]. *Pattern Recognition Letters*, 2015, 66(C): 100-108.
- [17] KOREZ R, LIKAR B, PERNU F, et al. Model-based segmentation of vertebral bodies from mr images with 3d cnns[C]// *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2016: 433-441.
- [18] MOESKOPS P, WOLTERINK J M, VELDEN B H M V D, et al. Deep learning for multi-task medical image segmentation in multiple modalities[C]// *International Conference on Medical Image Computing & Computer Assisted Intervention*. 2016.
- [19] GULSHAN V, PENG L, CORAM M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs [J]. *Jama*, 2016, 316(22): 2402.
- [20] ESTEVA A, KUPREL B, NOVOA R A, et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. *Nature*, 2017, 542(7639): 115-118.
- [21] MANSANET J, ALBIOL A, PAREDES R, et al. Mask selective regularization for restricted boltzmann machines[J]. *Neurocomputing*, 2015, 165(C): 375-383.
- [22] ANTONY J, MCGUINNESS K, CONNOR N E O, et al. Quantifying radio graphic knee osteoarthritis severity using deep convolutional neural networks[J]. 2016.
- [23] BENGTTSSON E, MALM P. Screening for cervical cancer using automated analysis of pap-smears[J]. *Computational and Mathematical Methods in Medicine*, 2014, 2014(2962): 842037.
- [24] KOOI T, LITJENS G, GINNEKEN B V, et al. Large scale deep learning for computer aided detection of mammographic lesions [J]. *Medical Image Analysis*, 2017, 35: 303-312.



GU Wan-rong, born in 1982, Ph.D, assistant professor. His main research interests include machine learning, information retrieval and recommendation.



MAO Yi-jun, born in 1979, Ph.D, assistant professor. His main research interests include machine learning, bioinformatics and algorithm.

(上接第 249 页)

- [6] MICHAILOVICH O, RATHI Y, TANNENBAUM A. Image Segmentation Using Active Contours Driven by The Bhattacharyya Gradient Flow[J]. *IEEE Transactions on Image Processing*, 2007, 16(11): 2787-2801.
- [7] WU H, APPIA V, YEZZI A. Numerical Conditioning Problems and Solutions for Nonparametric i. i. d. Statistical Active Contours[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(6): 1298-1311.
- [8] ZHOU Z, DAI M, WANG T, et al. Prior Distribution-Based Statistical Active Contour Model[J/OL]. *Multimedia Tools and Application*: 1-24. <https://10.1007/S11042-019-08101-2>.
- [9] OSHER S, SETHIAN J. Fronts Propagating with Curvature Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations[J]. *Journal of Computational Physics*, 1998, 79 (1): 12-49.
- [10] PARZEN E. On Estimation of a Probability Density Function and Mode[J]. *The Annals of Mathematical Statistics*, 1962, 33(3): 1065-1076.
- [11] HUANG Y, QIU C, GUO Y, et al. Surface Defect Saliency of Magnetic Tile[J]. *The Visual Computer*, 2018; 1-12.
- [12] KUMAR A, PANG G. Defect Detection in Textured Materials Using Gabor Filters[C]// *Industry Applications Conference*. IEEE, 2000: 1041-1047.
- [13] CHENG M, MITRA N. Global Contrast Based Salient Region Detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 37(3): 409-416.
- [14] GONG M, LI H, ZHANG X, et al. Nonparametric Statistical Active Contour Based on Inclusion Degree of Fuzzy Sets[J]. *IEEE Transactions on Fuzzy Systems*, 2016, 24(5): 1176-1192.
- [15] CHAN T, ZHU W. Level Set Based Shape Prior Segmentation [C]// *Computer Vision and Pattern Recognition*. IEEE, 2005, 2: 1164-1170.
- [16] ZHAO H K, CHAN T, MERRIMAN B, OSHER. S, A Variational Level Set Approach to Multiphase Motion[J]. *Journal of Computational Physics*, 1996, 127: 179-195.



YANG Zhi-wei, born in 1968, bachelor, senior engineer. His main research interests include industrial internet, 5G applications and planning, consulting, design of smart cities, also with the establishment of industrialization and information integration system.