



中国科学院大学
University of Chinese Academy of Sciences

硕士学位论文

基于深度学习的乳腺钼靶分类算法研究

作者姓名: 朱城

指导教师: 赵屹 研究员 中国科学院计算技术研究所

学位类别: 工学硕士

学科专业: 计算机应用技术

培养单位: 中国科学院计算技术研究所

2019年6月

Research on Classification Algorithm of Mammography Based
on Deep Learning

A thesis submitted to the
University of Chinese Academy of Sciences
in partial fulfillment of the requirement
for the degree of
Master of Science in Engineering
in Computer Application Technology
By
Zhu Cheng
Supervisor: Professor Zhao Yi

Institute of Computing Technology, Chinese Academy of Sciences

June, 2019

摘要

乳腺癌是女性最常见的恶性肿瘤之一，传统的基于乳腺钼靶的 CAD 系统在乳腺癌的分类诊断上存在耗时长、较高人为误诊率等问题。目前基于深度学习的乳腺钼靶所使用的整图分类及病灶检测等方法并没有综合考虑医学数据中存在数据部分标注的情况。为此本文主要进行了深度学习技术在乳腺钼靶这种不完全数据下的算法研究。

乳腺钼靶这类医疗影像分类问题有别于传统自然图像，其分辨率大、数据采集困难、数据量小、数据未全标注、类别之间差别微小等问题对研究者们提出了更高的要求和挑战。因此，本文先对医疗影像进行了分析，并在此基础上设计满足医生实际需求和计算机视觉领域要求的评判标准，对数据进行预处理等；为了充分利用医疗影像存在的未标注全的正常样本数据，本文在深度学习物体检测模型 Faster R-CNN 基础之上，引入弱监督学习的方法，在对带有标注信息的恶性乳腺钼靶病例充分学习之上，创新性地利用未带标注信息的正常样本数据，分别设计了两阶段模型 weakFaster R-CNN 和一阶段检测模型 simFaster R-CNN。

weakFaster R-CNN 采用两阶段训练方式，第一阶段专门学习恶性样本信息，第二阶段学习恶性及正常样本之间的区别。但 weakFaster R-CNN 存在耗时长、占用空间大等问题，为此针对数据特点提出 simFaster R-CNN。simFaster R-CNN 通过引入度量学习，创新性地提出了 top likelihood loss 和 similarity loss 等两个损失函数分别优化 RPN 网络的正常样本和学习正负样本之间的微小差异，取得了不逊于 weakFaster R-CNN 的效果。本文并使用该方法在大型公开数据集上 DDSM 上进行了测试，可以媲美目前基于全标注信息的模型训练的结果，充分证明了一阶段模型的可行性和优越性。而且该模型对于医学数据具有普适性，后期可进行更加深入的研究，具有充分的实际应用价值。

关键词：乳腺钼靶，weakFaster R-CNN，simFaster R-CNN，度量学习，similarity loss

Abstract

Breast cancer is one of the most common malignant tumors in women. The traditional CAD system for mammography has lots of problems, including a long time-consuming diagnosis of breast cancer and a high rate of misdiagnosis. And the current methods for mammography based on deep learning which applied for the classification of whole images and the detection of lesions do not comprehensively consider the data without annotations in the medical area. So this paper mainly studies the algorithm of deep learning technology under the circumstances of such incomplete data of mammography.

The image classification problem of mammography is different from that of traditional natural images, which includes the large resolution, the difficulty of data collection, the small datasets, and the not fully labelled of training data. The problem above have put forward higher requirements and challenges to the researchers. So this paper firstly analysed the medical images, formulated the relevant evaluation criteria which took care of doctor's and computer vision's need and preprocessed the data after that. In order to make full use of the unlabelled normal data of medical images, this paper used the deep learning object detection model Faster R-CNN and introduced weak supervised learning. After trained enough target cases of the malignant mammography with labelled information, this paper innovatively made use of normal data without labelled information. And then, this paper proposed a two-stage detection model(weakFaster R-CNN)and a one-stage detection model(simFaster R-CNN)respectively.

weakFaster R-CNN used a two-stage training approach, in which the first stage focused on studying malignant information and the second stage learned the difference between malignant and normal images. Result from the problems of weakFaster R-CNN which took a long time and took up a lot of space, simFaster R-CNN was proposed after then. simFaster R-CNN introduced metric learning, including top likelihood loss and similarity loss, to optimize the loss of normal images in the RPN and the tiny difference between the positive and negative images respectively, finally achieved the effect which

was as good as that of weakFaster R-CNN. This method was also used to be trained at the large public datasets DDSM. Compared to the results of model trained at full annotated information, simFaster R-CNN even can get a better result. Last but not least, the model has universality for medical data, and can be further studied in the future, which makes it have full practical application value.

Keywords: Mammography, weakFaster R-CNN, simFaster R-CNN, Metric Learning, similarity loss

目 录

第1章 引言	1
1.1 背景及意义	1
1.1.1 乳腺分型	1
1.1.2 乳腺癌及乳腺钼靶	2
1.1.3 基于乳腺钼靶诊断的难点	5
1.2 研究现状	7
1.2.1 乳腺影像学计算机辅助诊断方法	7
1.2.2 深度学习技术的发展	9
1.2.3 深度学习在乳腺乳腺上的发展	11
1.2.4 存在的问题及解决方案	12
1.3 主要研究内容	13
1.3.1 需求分析, 确定评判标准	13
1.3.2 实现基于深度学习的乳腺钼靶分类模型	13
1.3.3 设计全新损失函数, 改进并设计全新分类检测模型	13
1.3.4 编程实现, 完成不同模型之间的对比分析	13
1.4 本文结构及内容提要	14
第2章 深度学习相关技术	15
2.1 卷积神经网络	15
2.2 物体检测模型	17
2.2.1 两阶段物体检测模型	19
2.2.2 单阶段物体检测模型	23
2.3 迁移学习	25
2.4 弱监督学习	27
2.5 度量学习	28
2.6 本章小结	29
第3章 数据分析与处理	31
3.1 数据分析	31
3.1.1 医学数据分析	31
3.1.2 乳腺钼靶数据分析	33
3.2 数据处理	37
3.2.1 数据预处理	37

3.2.2 数据集划分	38
3.3 标准制定	38
3.3.1 数据定义	38
3.3.2 基于病例	38
3.3.3 基于图片	41
3.4 本章小结	41
第 4 章 两阶段乳腺钼靶分类检测模型 weakFaster R-CNN	43
4.1 目的	43
4.2 模型图	44
4.3 训练过程与细节	47
4.3.1 训练步骤	47
4.3.2 第一阶段	47
4.3.3 第二阶段	48
4.4 结果讨论与分析	50
4.4.1 基于病例	50
4.4.2 基于图片	51
4.5 本章小结	51
第 5 章 一阶段乳腺钼靶分类检测模型 simFaster R-CNN	53
5.1 目的	53
5.2 模型图	53
5.3 Top likelihood loss	54
5.4 Similarity loss	56
5.5 训练过程与细节	57
5.6 结果讨论与分析	57
5.6.1 基于病例	57
5.6.2 基于图片	59
5.6.3 可视化预测结果	59
5.7 其他工作	61
5.7.1 更换主干网络	61
5.7.2 更换 similarity loss	62
5.7.3 更换权重因子	62
5.7.4 添加模块	63
5.7.5 更换数据集	64
5.8 本章小结	66

第 6 章 总结与展望	69
6.1 总结	69
6.2 展望	69
参考文献	73
作者简历及攻读学位期间发表的学术论文与研究成果	77
.0.1 参加的研究项目	77
.0.2 获奖情况	77

图形列表

1.1 乳腺钼靶分型	1
1.2 乳腺钼靶肿块病灶特点	2
1.3 乳腺钼靶钙化病灶特点	3
1.4 乳腺钼靶结构扭曲病灶特点	3
1.5 乳腺钼靶特征征象病灶特点	4
1.6 良恶性乳腺钼靶数据对比	6
2.1 Lenet 模型图 [1]	16
2.2 计算机视觉主要任务 (a) 图片分类 (b) 物体检测 (c) 物体分割 [2]	18
2.3 R-CNN 模型图 [3]	20
2.4 Fast R-CNN 模型图 [4]	20
2.5 RPN 模型 [5]	22
2.6 YOLO 模型 [6]	23
2.7 SSD 模型 [7]	25
2.8 迁移学习与传统机器学习的不同。 (a) 传统机器学习对不同的学习任务建立不同的模型, (b) 迁移学习利用源域中的数据将知识迁移到目标域, 完成模型建立。	26
2.9 孪生网络图 [8]	29
3.1 基于深度学习的医学数据工作流程	33
3.2 dicom 数据格式	34
3.3 带标注信息的乳腺钼靶图片	35
3.4 本文乳腺钼靶病灶宽高长度直方图 (a) 病灶宽度直方图 (b) 病灶高度直方图	36
3.5 同一女性同一乳房不同角度拍摄图片	36
3.6 带有标注框的正样本及包含疑似恶性病灶的正常样本	37
4.1 weakFaster R-CNN 模型图	45
4.2 残差模块 [9]	46
4.3 weakFaster R-CNN 第一阶段对图片预测结果可视化	49
4.4 weakFaster R-CNN 第二阶段对图片预测结果可视化	51
5.1 simFaster R-CNN 模型图	54
5.2 经过 RPN 训练后得到的恶性样本和正常样本预测得到的 anchors	56

5.3 simFaster R-CNN ROC 曲线	58
5.4 simFaster R-CNN 在天肿数据集上的对正样本预测结果可视化	60
5.5 simFaster R-CNN 在天肿数据集上的对负样本预测结果可视化	60
5.6 SE Block[10]	61
5.7 乳腺钼靶纵横比	64
5.8 DDSM 乳腺钼靶图	64
5.9 simFaster R-CNN 在 DDSM 数据集上的对正样本预测结果可视化	66
5.10 simFaster R-CNN 在 DDSM 数据集上的对负样本预测结果可视化	66

表格列表

1.1 我国乳腺癌基本情况	4
1.2 基于深度学习的 CAD 系统算法 [11]	8
3.1 天肿数据集划分	38
4.1 Resnet101 参数配置 [9]	46
4.2 weakFaster R-CNN 第一阶段训练参数配置	48
4.3 weakFaster R-CNN 第一阶段预测参数配置	48
4.4 weakFaster R-CNN 第二阶段训练参数配置	49
4.5 weakFaster R-CNN 第二阶段预测参数配置	50
4.6 weakFaster R-CNN 第二阶段训练集预测结果	50
4.7 weakFaster R-CNN 基于病例预测结果	51
4.8 weakFaster R-CNN 基于图片预测结果	51
5.1 simFaster R-CNN 基于病例预测结果	57
5.2 simFaster R-CNN 基于图片预测结果	59
5.3 weakFaster R-CNN 预测结果	59
5.4 simFaster R-CNN 使用 Seresnet101 模型预测结果	61
5.5 simFaster R-CNN 使用 contrastive loss 预测结果	62
5.6 simFaster R-CNN 使用不同权重因子预测结果	63
5.7 simFaster R-CNN 添加 DCN 模块预测结果	63
5.8 DDSM 数据集划分	65
5.9 simFaster R-CNN 使用 DDSM 数据集基于图片预测结果	65

第1章 引言

1.1 背景及意义

1.1.1 乳腺分型

乳腺腺体类型依据美国放射学会提出的乳腺影像报告和数据系统(BI-RADS系统)分为四种类型——脂肪型乳腺、散在纤维腺体型乳腺、不均匀致密型乳腺及极度致密型乳腺，而如图1.1可以得知，不均匀致密型乳腺及极度致密型乳腺因腺体遮挡会降低对微小病变的检出率，尤其是非钙化型病变。医学研究表明，其中致密型乳腺的人群罹患乳腺癌的概率比脂肪型增加4-6倍[12]。

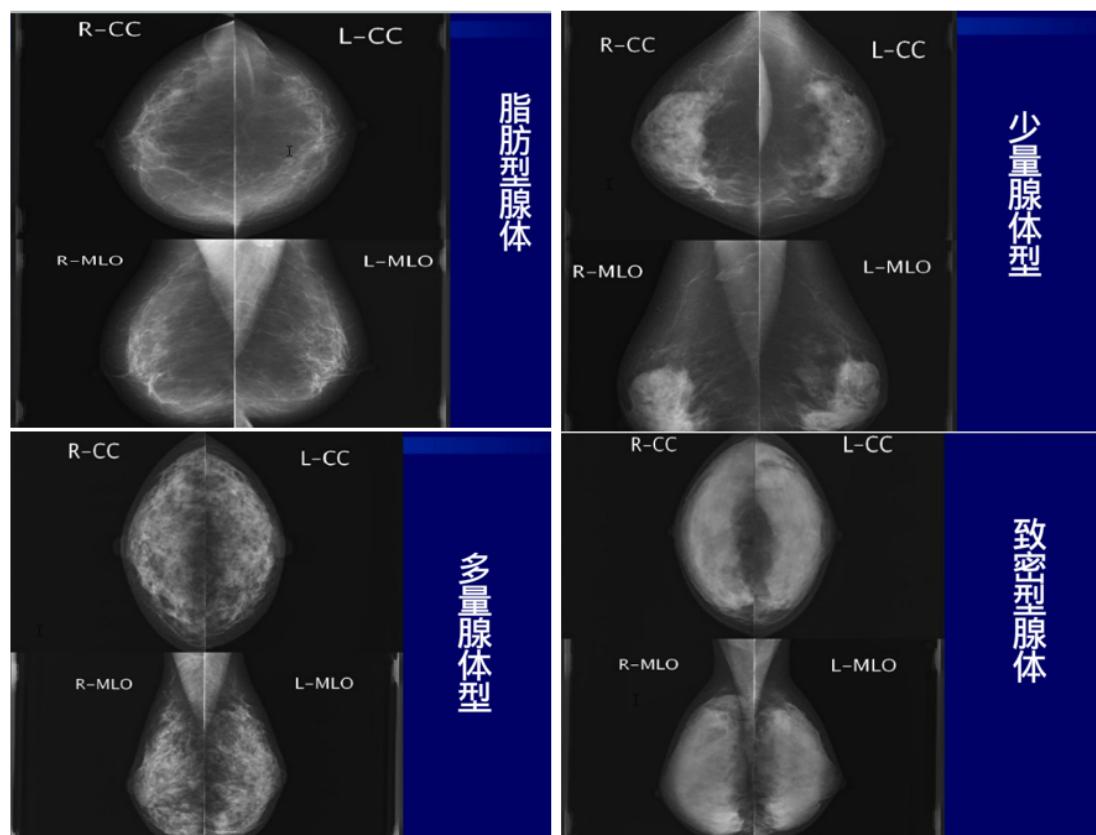


图 1.1 乳腺钼靶分型

Figure 1.1 Mammography target classification

乳腺钼靶问题的病灶分成肿块、钙化(如图1.3)、结构扭曲(如图1.4)、特殊征象(如图1.5)等多种类型。钙化分布多样，形态存在典型良性钙化，高度恶性钙化，钙化和不定性钙化等多种类型，大小一般为0.5mm到1mm之间；结构扭

曲无确定的肿块可见，常见于手术后疤痕，硬化性乳腺病，浸润性导管癌等；特殊征象主要非对称性乳腺组织，局灶性非对称致密类型 [13]。

本文重点研究和考虑肿块类型数据，肿块是对乳腺不同角度拍摄都能显示的病变，其形状包括圆形、软圆型、分叶型、不规则型等，边缘存在清晰、小分叶、模糊、浸润和星芒状等情况，拥有高密度、等密度、低密度、含脂肪密度等不同密度。对于肿块病灶中常见的浸润性导管癌，其大小小于 1cm（如图1.2）[14]。良恶性肿块之间区别并不明显，这需要专门研究病灶所包括的良恶性信息，进而对诊断提供帮助。

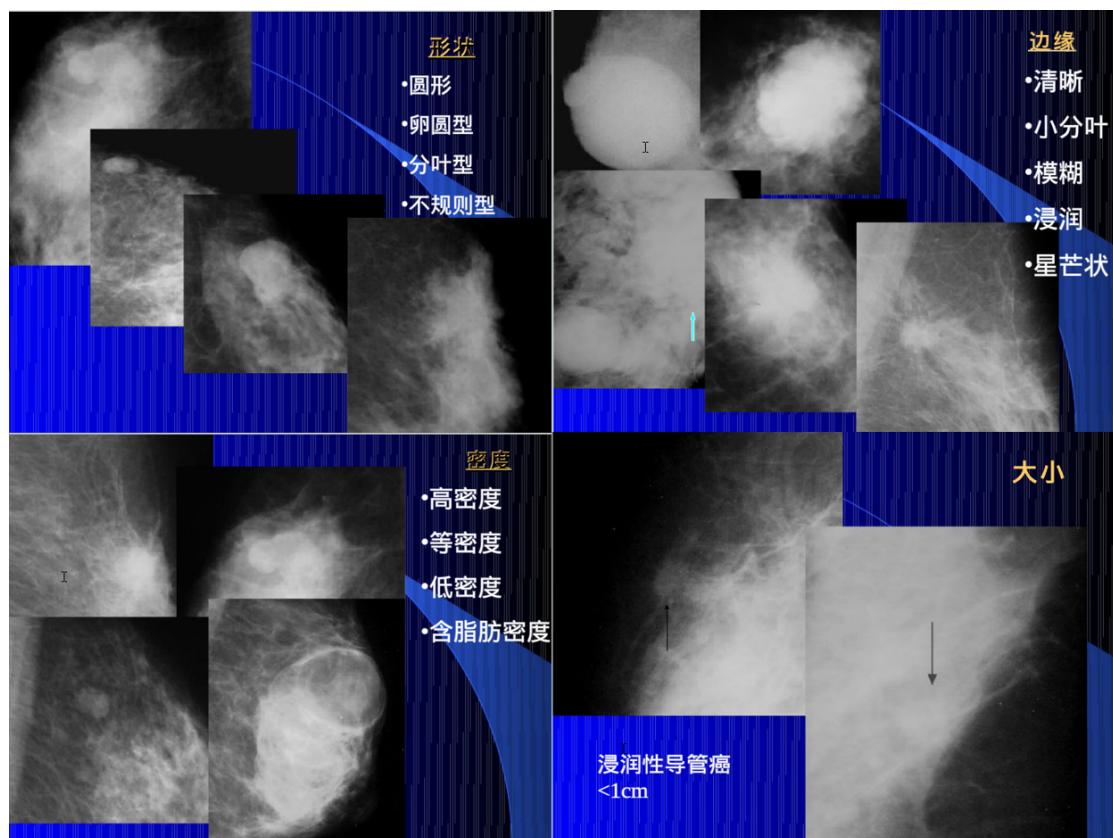


图 1.2 乳腺钼靶肿块病灶特点

Figure 1.2 Mammography's mass characteristics

1.1.2 乳腺癌及乳腺钼靶

乳腺癌是女性最常见恶性肿瘤之一，其发病率逐年上升。在欧美国家，乳腺癌占女性恶性肿瘤的 25%-30%。20 世纪末的统计资料表明全世界每年约有 130 万人诊断为乳腺癌，而有 40 万人死于该病。全球肿瘤流行病统计数据认为乳腺癌是中国女性最常见的癌症，年龄标化率（ASR）为每 10 万人 21.6 例。根据中

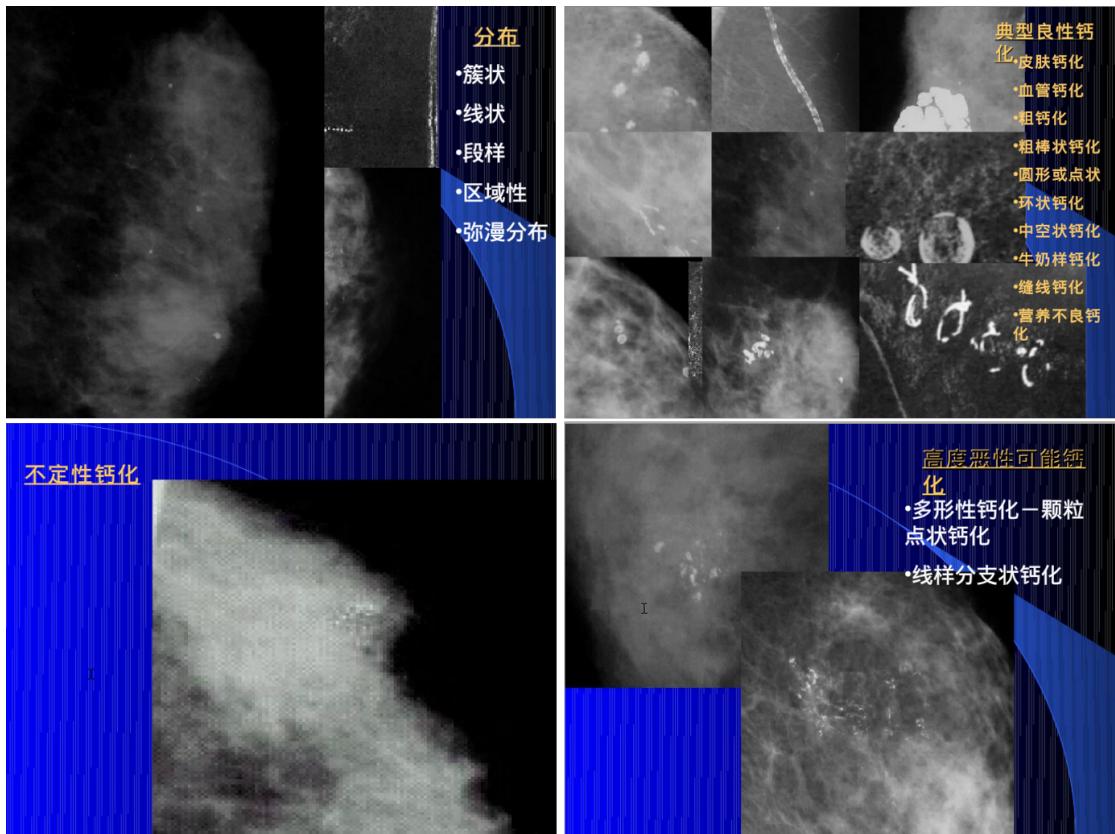


图 1.3 乳腺钼靶钙化病灶特点

Figure 1.3 Mammography's calc characteristics

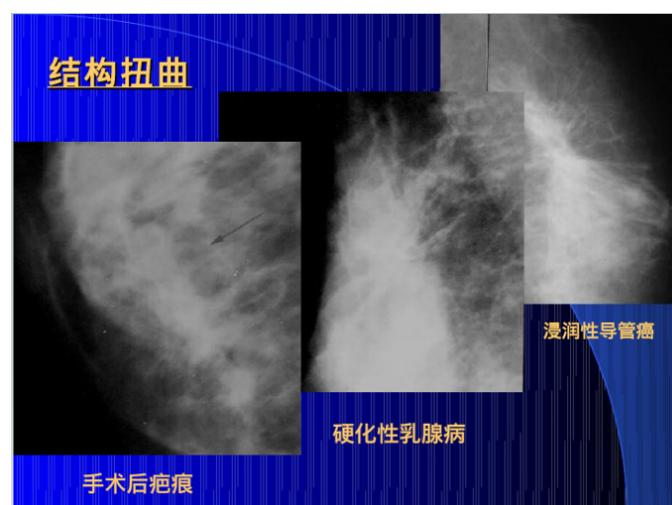


图 1.4 乳腺钼靶结构扭曲病灶特点

Figure 1.4 Mammography's distortion characteristics

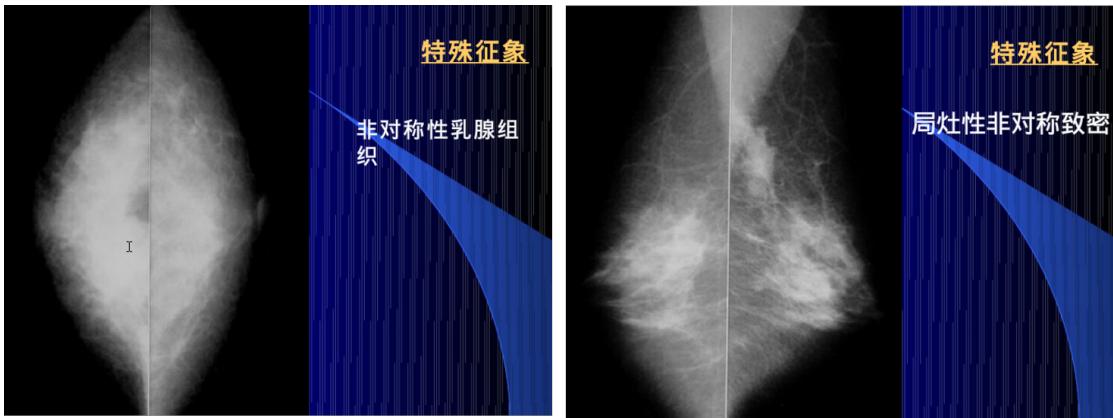


图 1.5 乳腺钼靶特征征象病灶特点

Figure 1.5 Mammography's Characteristic signs characteristics

国国家肿瘤登记中心的数据，乳腺癌是城市女性最常见的癌症，是农村女性第四大常见癌症。城市地区的 ASR（34.3 例/10 万女性）是农村地区的 2 倍（17.0 例/10 万女性）[15]。乳腺癌在城市中的发病率为女性恶性肿瘤的第二位，一些大城市已经上升至第一位，农村中为第五位。乳腺癌已经成为妇女健康的最大威胁，如表1.1所示。

表 1.1 我国乳腺癌基本情况

Table 1.1 Basic situation of breast cancer in China

我国乳腺癌	2012 年人数	2012 年占全球比	至 2030 年增幅	居女性恶性肿瘤排名
发病	18.7 万	11.2%	29.8%	NO.1
死亡	4.8 万	9.2%	47.4%	

随着影像技术的快速发展，X 线、超声等影像技术已逐渐成为肿瘤检出、分期及随访的重要手段，而大量的影像数据在提高疾病诊断准确性的同时也增加了疾病诊断的复杂性和对医师经验的依赖性。乳腺钼靶，全称乳腺钼靶 X 线摄影检查，是目前诊断乳腺疾病简便、可靠的无创性检测手段，具有痛苦小、简便易行和重复性高等优点。钼靶摄片有保存价值，可供前后对比。目前已作为年龄大于 35 岁妇女常规检查（每年可做一次）[16]。由于有少量射线，年龄小于 35 岁的妇女可先做 B 超，当 B 超不能说明问题，怀疑有肿瘤尤其怀疑有乳腺癌的，也可以做钼靶检查。乳腺钼靶可分别针对女性两个乳房进行拍摄，根据角度的不同可以分别得到多张医疗影像资料，目前业内常用的是轴位片（CC）和斜位片（MLO），根据左右乳的不同，一个病人大多数拥有左乳轴位片（L-CC）、左乳斜

位片（L-MLO）、右乳轴位片（R-CC）和右乳斜位片（R-MLO），可以综合上述病症判断病人是否得癌。

随着计算机技术和统计机器学习算法的发展，计算机辅助诊断（CAD）[17]在医学影像领域获得了快速发展。CAD 解决了一些问题，包括对于一些小病灶的发现，包括对于一些病灶形态的分析和病变。其提高医生的诊断准确性的原因在于，在传统诊断方法中，放射科医生的诊断完全是主观判断过程，因而会受到诊断医生经验及知识水平的限制和影响；其次，医生诊断时易于遗漏某些细微改变；再者，不同医师间及同一医师间的阅片差异的影响。而计算机客观的判断对于纠正这些错误和不足具有巨大的优势。

基于钼靶的乳腺癌筛查是目前 CAD 最广泛的应用领域，研究的热点也主要集中在提高肿块和钙化灶的检出准确性，而检出效能主要受乳腺腺体类型和肿瘤组织学类型的影响。钼靶检查归根结底是对钼靶图像的识别问题，当前主要依赖医生进行判读，由于个人经验差异往往导致诊断结论的一致性，即使对于同一位医生，一定的人为失误率也是无法避免的。不仅如此，对于中国女性来说，乳房相对偏小，腺体组织比较多，也进一步提升了目视解读的难度。医学数据可分为恶性，良性和阴性数据，对于医务工作者而言，恶性数据容易标注出具体的恶性病灶位置，而对于良性和阴性数据，难以确切定位病灶位置而且工作量巨大，对于乳腺钼靶数据更是如此，也给基于深度学习的乳腺钼靶的分类检测工作带来全新的挑战与现实的意义。对于我国，医疗现状存在着患者量大图像多的问题；基层医院买得起设备，但没有合适的医生；非疾病高发地区的医生，缺乏实践经验等问题；通过人工智能辅助诊断系统可以方便快捷的帮助医生诊断，比如通过基于钼靶的乳腺癌类型筛查，可以精准地定位出可疑的肿块病灶区域，并且给出肿块良恶性判断，甚至可以通过给出可能的病例报告分析，医生可以根据筛查的结果，得出进一步的诊断分析，可以大幅度提高医生工作效率和诊断准确率。同时能够让低年资医生快速的积累诊断经验，降低学习成本，最终实现多方位帮助医生减轻工作量、辅助医生进行临床诊断的目的。

1.1.3 基于乳腺钼靶诊断的难点

基于深度学习的自然图像识别可以发展的关键原因在于巨大的数据量以及图像数据之间存在边缘，大小，密度等传统计算机视觉的形态学特征。传统的计算机视觉主要采用各种算子来提取图片的特征，之后通过机器学习算法来进行

分类。深度学习通过卷积，下采样等方式可以抽象出图像的高维表达，从而有效地对图片进行分类。但无论如何，图片存在肉眼可以区分的明显特征是设计模型的关键。

医疗影像数据存在区别于传统自然图像，存在着数据采集，数据本身特点带来的特征学习难度，以及区别于传统计算机视觉的评判标准等诸多问题，具体难点表现为：

- 医疗数据采集尤为困难，各个医院标准不一，格式多样。对于单一癌症疾病存在着多种采集方式，并由于采集设备的不同会产生分辨率等众多区别。比如公开数据集 INBREAST[18] 和 DDSM[19–21] 得到的乳腺钼靶数据分辨率就不一样。
- 医疗影像数据一般图片大，良恶性之间区别特别微小，从医生那边得知，基于乳腺钼靶的肿块型乳腺癌判断的唯一标准就是边缘。带有恶性乳腺癌的乳腺钼靶的医学特征是乳腺边缘带有毛刺，而良性的乳腺边缘是光滑的（如图1.6），二者常常无法简单的通过肉眼来人为区分，医生们也需要结合多张图片甚至病理性报告才能得到更加确切的结果。这使得乳腺钼靶的分类判断相比于传统的自然图像分类困难许多。

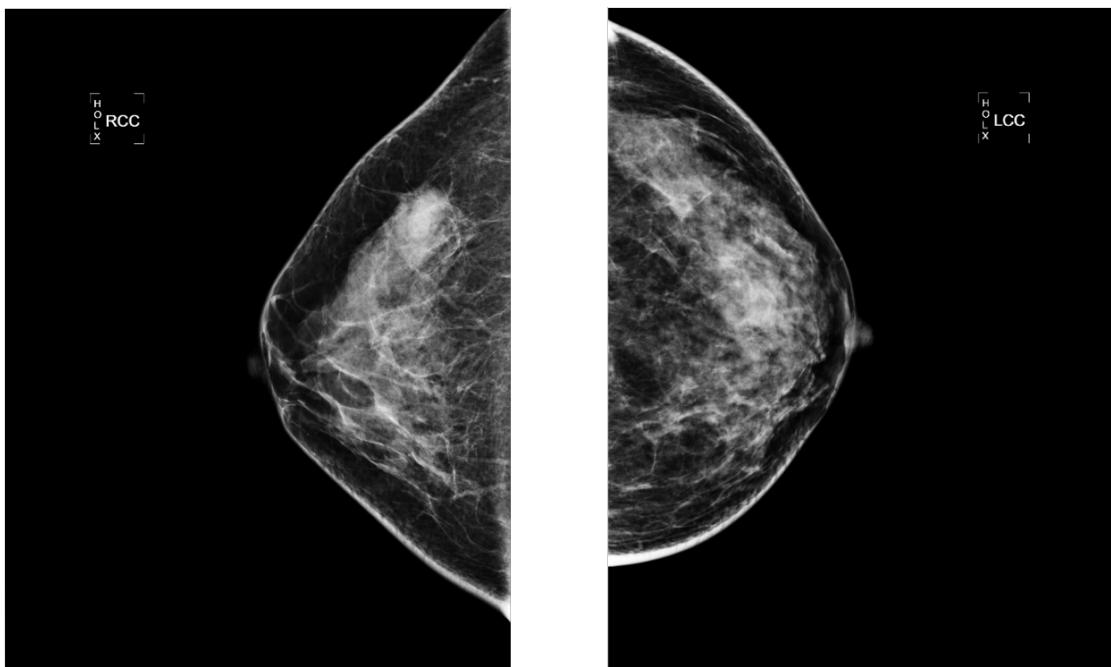


图 1.6 良恶性乳腺钼靶数据对比

Figure 1.6 The malignant and benign mammography

• 医院有自己的标准，比如医生更关注敏感度和特异度，而不怎么关注准确率等传统计算机视觉的评判标准。在医学领域，存在着样本类别极度不均衡等问题，也更需要对此进行设计符合医生实际情况的评判标准。标准制定的有效与否，决定了模型的好坏。

医生的诊断常常会因为人为的误判而导致较高的误诊率，而且乳腺钼靶的分析需要丰富的临床经验，这对于医生要求更高。正因为基于乳腺钼靶诊断同时存在如上诸多难点，如果可以通过训练有足够区分能力的乳腺钼靶分类识别模型，就可以充分降低医生在寻找乳腺钼靶病灶的时间，而且一定程度上降低人为的误诊率，从而达到辅助医生进行诊断的效果，这也充分验证了基于深度学习的乳腺钼靶的分类算法研究的必要性和重大意义。

1.2 研究现状

1.2.1 乳腺影像学计算机辅助诊断方法

随着计算机技术和统计机器学习算法的发展，CAD 在医学影像领域获得了快速发展。CAD 技术理论上可以应用于多种影像技术，通常医学影像学中计算机辅助诊断分为三步。

图像的处理过程（预处理），其目的是把病变从正常结构中提取出来。在这里图像处理的目的是使计算机易于识别可能存在的病变，让计算机能够从复杂的解剖背景中将病变及可疑结构识别出来。通常此过程先将图像数字化（经过一定的 AD 转换），一般用扫描仪将图像扫描，如果原始图像已经为数字化图像，如 DR、CT、MRI 图像则可省去此步。针对不同的病变，需要采用不同的图像处理和计算方法，基本原则是可以较好地实现图像增强和图像过滤，并通过上述设计好的处理过程，计算机得以将可疑病变从正常解剖背景中分离、显示出来。

图像征象的提取（特征提取）或图像特征的量化过程。目的是将第一步提取的病变特征进一步量化，即病变的征象分析量化过程。所分析征象是指对病变诊断具有价值的影像学表现，如病变的大小、密度、形态特征等。数据处理过程。将第二步获得的图像征象的数据资料输入人工神经元网络等各种数学或统计算法中，形成 CAD 诊断系统，运用诊断系统，可以对病变进行分类处理，进而区分各种病变，即实现疾病的诊断。

而基于深度学习的 CAD 系统的算法大体上可分成四类：基于受限玻尔兹曼

机（RBM）的模型，卷积神经网络（CNN），基于自编码（autoencoder）的模型，以及基于稀疏编码的模型（见1.2）[11]

表 1.2 基于深度学习的 CAD 系统算法 [11]

Table 1.2 CAD system algorithm based on deep learning[11]

基于深度学习的算法	算法名称
基于 CNN 的方法	Alexnet,Clarifai,SPP,VGG,GoogleNet 深信度网络（deep belief networks, DBN）,
基于 RBM 的方法	深度玻尔兹曼机（deep Boltzmann machine, DBM）, 深度能量模型（deep energy model, DEM）
基于自编码的方法	稀疏自编码器（sparse autoencoder, SAE）, 去噪自编码器（denoising autoencoder, DAE）, 收缩自编码器（contractive autoencoder, CAE）
基于稀疏编码的方法	稀疏编码空间金字塔配对, 拉普拉斯稀疏编码, 局部坐标编码, 超向量编码

尽管乳腺钼靶影像目前被认为是早期检查乳腺癌最可靠的方式，能够减少 18%-30% 的致死率，但其对比度不高，肿瘤与周围正常组织的分界不明显，这会导致 10%-30% 的癌症漏检。由于这个原因，在最近几十年，相关的和乳腺钼靶结合共同诊断分析的技术不断提出来 [22]，主要包括以下几种：

- 基于乳腺超声图像的 CAD 系统

作为乳腺 X 光影像乳腺病变检查的技术辅助，引入了超声波检查，主要对影像学上致密的乳房提供更精确的评估。当这两种技术结合使用时，鉴别囊性结节和实性结节可减少了 25%-35% 的活检数量，大大提高了有效率。

- 基于 CT 图像的乳腺癌症 CAD 系统

由于乳腺钼靶摄影对病变的显示存在非显著性和误报问题，这会导致不必要的阴性穿刺活检。锥束乳腺 CT 扫描仪 (bCT) 作为一种新的专用影像方法，能够生成高质量的层析数据以提高乳腺组织和结构的可视化，并使病灶显示更显著。

- 基于 MRI 图像的乳腺癌症 CAD 系统

20世纪90年代以来，乳腺MRI就开始被用于表征和检测乳腺病变。MRI的敏感度非常高，可达到78%-98%，但其特异度不足，只有43%-75%。近些年发展起来的计算机辅助诊断软件的目的主要是为了使MRI分析和报告更加便利，或试图进一步突出显示检测到的病变。

综合CAD的发展，现有的乳腺钼靶的CAD系统只能标出可能的区域，但并不能提供诊断和分型，对于中国女性致密性乳腺的特点也很难刻画病灶位置以辅助医生进行诊断，漏诊，误诊的现象时有发生，所以并不能有效地减少工作量。大多数的CAD耗时比较长，常规临床应用存在很大的难度。然而医疗影像的识别分析工作，对于人工智能的需求已然越来越强烈。目前我国医学影像数据的年增长率约为30%，而放射科医师数量的年增长率约为4.1%，放射科医师的数量增长远不及影像数据的增长，这意味着放射科医师在未来处理影像数据的压力会越来越大，甚至远远超过负荷。

1.2.2 深度学习技术的发展

随着谷歌的AlphaGo战胜韩国棋手李世石之后，深度学习从学术界的热点技术，变为大众热捧的热门技术，成为当前人工智能热潮中，最吸引眼球的技术方向。深度学习也从传统的图像、视频、语音识别领域，向文本处理、自然语言理解、人机对话、情感计算等各方面渗透。

深度学习理念，在上世纪六七十年代就有人提出过，但是由于缺乏有效的算法，没有形成规模应用。在上世纪80年代，由于出现了BP算法[23]，多层前馈网络也曾经风靡一时，但是由于误差反向传播(BP)算法，对权值的调整从输出层开始，效果会每层递减，因此训练多层网络速度会变慢，而且多层网络额外增加了很多参数，需要更多的训练样本，由于当时计算能力的限制，最终多层前馈网络也没有流行起来。

随着技术的方法和计算能力的增强，以前制约深度学习的条件逐渐得到了解决，尤其是人们意识到，以前的神经网络，需要研究人员对问题进行深入的研究，提取出问题的关键属性，然后才是设计合适的神经网络，然后训练神经网络来解决这个问题。人们发现，在整个过程中，最困难的地方并非神经网络本身，而是抽取所研究问题的特征，这才是制约神经网络应用的核心问题。基于对这个问题的认识，逐渐出现了利用非监督学习网络来获取所研究问题的特征，然后再采用监督学习方式来训练网络，不仅可以使最为费时费力的特征提取来自动化，

而且还可以对网络进行预训练，降低网络训练工作量以及所需训练样本。正是因为上述原因，深度学习技术才会在十年前逐渐流行起来。

2006 年，神经网络领域的大师 Geoffrey Hinton 教授与其博士生在《Science》和相关期刊上发表了论文，首次提出了“深度置信网络 (DBN)[24]”的概念。与传统的训练方式不同，“深度置信网络”有一个“预训练”(pre-training) 的过程，这可以方便的让神经网络中的权值找到一个接近最优解的值，之后再使用“微调”(fine-tuning) 技术来对整个网络进行优化训练。这两个技术的运用大幅度减少了训练多层神经网络的时间。他给多层神经网络相关的学习方法赋予了一个新名词——“深度学习”。

2012 年，Hinton 教授的研究团队参加了斯坦福大学李飞飞教授等组织的 ImageNetILSVRC 大规模图像识别评测任务 [25]。该任务包括 120 万张高分辨率图片，1000 个类比。Hinton 教授团队使用了全新的多层 CNN (卷积神经网络，Convolutional Neural Networks，简称 CNN) 结构，突破性地将图像识别错误率从 26.2% 降低到了 15.3%。这一革命性的技术，让神经网络深度学习以极快的速度跃入了医疗和工业领域，这才有了后来一系列使用该技术的医学影像公司的出现。深度学习在计算机视觉方面取得的突破性进展是其应用在医疗影像上的主要因素，微软亚洲研究院于 2015 年开发的 ResNet[9] 模型在 ImageNet 自然图片数据集上的识别准确率超过了人类，为丰富多彩的视觉应用提供了很好的基础模块。

深度学习在医疗领域最激动人心的应用，无疑是在在医学诊断方面的应用。谷歌的 DeepMind 和 IBM 的 watson，都在这方面积极布局，尤其是 watson，在某些特定领域，其诊断精度已经超过了人类专家。由于医疗中病例大多数为非结构化文本数据，因此采用多层限制性玻尔兹曼机 (RBM) [26] 堆叠成的深度信念网络 (DBN)，可以自动提取文本病例中的特征，可以有效的学习病历中的知识，同时可以高效地进行诊断。2016 年至 2018 年，《Nature》等顶级学术期刊也将目光投向了基于医学影像的深度学习应用，包括糖尿病视网膜病变诊断 [27]，儿童白内障分类诊断 [28] 和皮肤癌病变诊断 [29] 等等，这些成果表明计算机经过训练甚至可以超过资深专家的诊断准确率，预示了其在计算机辅助诊断方面巨大的应用潜力。2018 年，Cell 杂志上同时报道了基于深度学习的视网膜疾病诊断工具 [30]，证明了 AI 系统在图像诊断方面具有普遍适用性。

目前在乳腺癌问题上，最为成熟的是病理活检方面，如去年发布在《Nature》封面上的一篇关于乳腺癌病理的自动检测模型 [31]，自动检测的准确率首次超过临床医生，达到了 93%-95%。除此外，还包括核磁前列腺癌区域的自动检测、脑部肿瘤的识别和分割，还有应用非常普遍和广泛的乳腺钼靶的钙化点的自动检测，以及胸片的自动检测，所有的方法都依赖于大量的数据以及有效的数据标注，并利用卷积神经网络，递归神经网络等深度神经网络结构自动获取特征表达能力，去掉繁杂的人工特征工程，进行病灶的定性自动检测。

我们有理由相信深度学习同样能在乳腺影像学诊断方面大放光彩，有望加速有关可治疗性疾病的诊断，从而促进疾病的早治疗，最终改善病人的临床结果，取得成功。

1.2.3 深度学习在乳腺乳腺上的发展

采用 CNN，可以极大提高识别率，同时降低对原始图片质量的要求，同时可以降低对训练样本数量的要求，因此 CNN 在医学影像处理方面，应该是目前应用最广泛也是最成功的领域。

目前主流的方法是直接对乳腺钼靶进行整图分类，应用迁移学习 [32] 或者切割出病灶位置进行检测的方式使得结果更加鲁棒。比如 MICCAI2017 收录的 [33] 提出的模型的总体框架是，首先对输入图像进行分割、缩放得到感兴趣的区域，通过 CNN 提取很多区块后，用线性回归分类器得到可能性排序，最后根据机器学习模型得到影像的判定结果输出，其在 INBREAST[18] 乳腺钼靶公开数据集上达到 ACC 为 0.90，AUC 为 0.8586 的结果。LÉVY D 等 [34] 以及 SHEN L. 等 [35] 论文中使用从传统自然数据学习到的 AlexNet[36]，VGG[37] 等模型迁移至公开的乳腺钼靶数据集 DDSM[19–21] 进行验证测试，通过使用热图，移动窗口等方式获取候选框，进而对病灶位置进行切割形成 patch，之后对病灶位置使用数据增广方式，最后 SHEN L. 等 [35] 在测试集上最好的结果 ACC 为 0.90。

除了在使用迁移学习以及切割出病灶位置的思路外，已有的基于深度学习的乳腺钼靶分类算法也考虑到乳腺钼靶的数据特点，进而通过融合不同分类模型、不同分辨率、多个角度乳腺钼靶图片来提高模型的性能。比如 DHUNGEL N 等 [38] 使用 multi-model 的思路，先训练多个分类模型来对单一乳腺钼靶图片进行学习，最后融合每个模型结果达到较好的结果。CARNEIRO N D G 等 [39] 则使用了 multi-scale，由于乳腺钼靶图片较大，可以先将同一张乳腺钼靶图片缩放

到不同尺度，之后融合多个不同分辨率的乳腺钼靶图片来增强对单一乳腺钼靶图片的学习能力。而 GERAS K J 等 [40] 则考虑同一病人下的乳房拥有多个不同角度（MLO 和 CC），通过利用 multi-direction 的思想，结合两个不同角度的乳腺钼靶学习结果，来加强对单一乳房的识别效果。工业界对该问题也是特别关注，比如 Google 的 DeepMind 和腾讯觅影计划在内的工业界目前所采用的方法也是对整图进行分类，其中腾讯觅影计划对肿块检测的良恶性判别的敏感度达到 0.90，特异度达到 0.96。

上述模型均只是考虑到直接使用整图信息进行分类，之后在此基础上使用各种 multi 思路来综合考虑并加强分类的结果。这些模型并没有充分利用到乳腺病灶的信息，而且，所采用的数据来自公开的数据集，常见的是 INBREAST[18] 和 DDSM[19–21]。这些数据集所采用的数据均来自西方女性，并不完全适用于东方女性特有的致密性乳腺。

少数的几种从检测角度对乳腺钼靶进行检测分类的，比如从这一角度考虑的论文 [41–43] 所采用的方法，都是在带有恶性乳腺肿瘤标注信息上的数据集上使用主流的 FasterR-CNN[5] 或 YOLO[6] 模型进行检测。虽然充分利用到了检测信息，但是忽略了大部分的没有带有标注信息的正常数据（包括疑似恶性肿瘤的良性和阴性数据），造成了数据资源的浪费，而且由于没有充分考虑到恶性和正常数据分布的各异性，使得结果容易偏向恶性数据，造成该训练下的检测模型在对正常数据进行检测时，会造成很大的假阳性率。

1.2.4 存在的问题及解决方案

目前的基于深度学习的乳腺钼靶分类算法大多考虑到迁移学习方法。从目前的研究现状可以知道，分类角度主要采用切割 patch 方式提取出病灶信息进而进一步考虑和使用多模型，多角度融合等 multi 思路提升模型效果，这么做对于本文而言，一方面加大了复杂度，而且未充分利用数据自带的检测信息；检测角度则更多的问题在于只考虑到有标注信息的数据，对于实际应用场景，尤其对于医疗领域的影像数据而言，常常存在没法标注的数据，这就使得现有的模型无法解决这问题，并且需要在数据的基础上创新性的设计全新模块。

针对上述问题，本论文期望从天津肿瘤医院获取到的真实数据出发，在基于现有深度学习模型基础上，对数据进行尝试性的研究工作后，提出全新的适用于本数据的模块与方法，从而同时满足性能与效率的要求。

1.3 主要研究内容

由于东方女性偏致密性乳腺，属于致密性乳腺分型，而肿块是众多乳腺癌表现中最为常见的。针对于前面介绍的研究对象乳腺钼靶及其发展现状等，毕设工作将完成基于深度学习的乳腺钼靶分类算法的研究工作，主要研究内容是：针对医院提供的特有的东方女性肿块型致密性乳腺钼靶病例，也就是带有标注的恶性乳腺钼靶和正常乳腺钼靶病例（没有标注的良性和阴性乳腺钼靶病例），基于深度学习等技术，以恶性乳腺钼靶病例图片数据为重点研究对象，并辅以正常乳腺钼靶病例图片数据，使得设计的分类检测模型可以分类出恶性及正常乳腺钼靶病例，且识别恶性乳腺钼靶图片数据中病灶位置。

从本质上而言，本文所要学习的标注信息数据并不全面，所以在具体模型设计时应当充分考虑这一点，并针对模型结果进行相应的损失函数设计以充分利用恶性样本标注信息及未标注的正常样本信息。具体而言包括以下部分：

1.3.1 需求分析，确定评判标准

明确数据采集格式要求，分析医学数据，进行去除噪声数据等操作；根据天津肿瘤医院的实际要求及计算机视觉的知识，设定评判标准。

1.3.2 实现基于深度学习的乳腺钼靶分类模型

调查已有的方法及模型，对比分析优劣；使用原始的 Faster R-CNN，根据数据特点，分阶段训练乳腺钼靶数据并得到分类检测结果。

1.3.3 设计全新损失函数，改进并设计全新分类检测模型

根据数据特点及实现现象，对实验结果进行优劣分析，引入全新损失函数，改进原始 Faster R-CNN 模型。

1.3.4 编程实现，完成不同模型之间的对比分析

基于评判标准及设计的深度学习模型，完成源码的实现，并借助于实验结果对模型模块进行更新与对比分析，从而为基于深度学习的乳腺钼靶及医学影像的算法研究提供新思路。

1.4 本文结构及内容提要

论文的第一章部分主要介绍了乳腺癌及乳腺钼靶的相关医学知识及诊断方法，以及目前所采用的基于深度学习乳腺钼靶的分类检测思路，对于其中的优劣势分别进行了讨论和考虑，并重点介绍了基于深度学习的乳腺钼靶分类算法的研究的必要性与重要性，对于之后的工作进行了整体的思路设计。

论文的第二章详细介绍了本文所使用到的深度学习技术，包括卷积神经网络，基于卷积神经网络技术基础发展的物体检测模型，为解决数据量少而使用的迁移学习，为了学习恶性及正常样本之间区别而引入的度量学习，在设计标准中根据实际情况而使用的弱监督学习等。

论文的第三章具体介绍了医学数据，尤其乳腺钼靶数据的特点，根据数据特点进行了数据预处理，划分以及标准制定，为后期模型的设计与评判提供依据。

论文的第四章主要介绍了从两阶段角度考虑下模型的设计及结果。由于恶性及正常数据之间存在明显的区别，所以在本章中先是考虑对恶性数据进行模型学习，再学习恶性及正常数据之间微弱的区别，通过调参达到一定的效果后，对结果进行分析，从而引出后续的第四章的一阶段模型的设计及研究。

第五章是在第四章已经研究的基础上，试图从降低复杂度的角度考虑提高模型的运行速度，并基于对数据特点的分析上，对 RPN 网络设计 top likelihood loss 损失函数，并对数据之间微弱区别引入 similarity loss 全新损失函数。在保证和两阶段差不多的诊断性能下，提升模型运行的效率。

第六章总结了本文的研究内容，并提出了当前研究方法的不足和未来本领域面临的挑战。

第 2 章 深度学习相关技术

基于深度学习的乳腺钼靶分类问题本质上是传统的计算机视觉分类检测问题。为此，本文应用了卷积神经网络对乳腺钼靶数据进行特征提取，基于目前最成熟的两阶段模型 Faster R-CNN 对乳腺钼靶病灶进行定位分析。考虑到数据量少、训练不易收敛等问题，迁移在传统大数据集上训练好的模型到医疗影像数据进行学习；同时考虑到部分数据未标注，引入弱监督学习的方法，并为了充分挖掘模型区分正常样本和恶性样本的能力，考虑使用度量学习。为此，本文所考虑到的所有的深度学习技术将在本章具体介绍。

2.1 卷积神经网络

深度学习以数据作为输入，经过算法抽象将数据变为最终任务需要的高维特征，完成特征到任务之间的映射。其中，代表算法是神经网络算法，包括深度置信网络、递归神经网络和卷积神经网络等。特别是 CNN，目前在计算机视觉、医疗影像处理等领域发挥着重要的作用，这个也是本文主要涉及到的关键技术。

CNN 是特殊的人工神经网络，区别于神经网络其他模型（如递归神经网络，Boltzmann 机等）。卷积神经网络是一种带有卷积结构的深度神经网络，卷积结构可以减少深层网络占用的内存量，其存在三个关键的操作，第一是局部感受野，第二是权值共享，第三是通过 pooling 层有效地减少网络的参数个数，缓解了模型的过拟合问题。CNN 在众多领域应用尤其是图像相关任务表现优异，诸如图像分类、图像语义分割、物体检测等计算机视觉问题。另外，随着研究的深入，如自然语言处理中的文本分类、软件工程数据挖掘中的软件缺陷预测等问题都在尝试利用卷积神经网络解决，并取得了相比传统方法甚至其他深度网络模型更有的预测结果。

CNN 发展历史中的第一件关键时间是上世纪 60 年代左右的神经科学中，加拿大神经科学家 David H. Hubel 于 1959 年提出猫的初级视皮层中单个神经元的“感受野”概念，在 1962 年发现猫的视觉中枢里存在感受野、双目视觉等其他功能结构，标志着神经网络结构首次在大脑视觉系统中被发现 [44]。

随后，Yann LeCun 等人在 1998 年提出基于梯度学习的 CNN 算法 [1]，并将

其成功用于手写数字字符识别。他们提出的 LeNet 模型^{2.1}是第一个产生实际商业价值的卷积神经网络，同时也为卷积神经网络以后的发展奠定了坚实的基础。

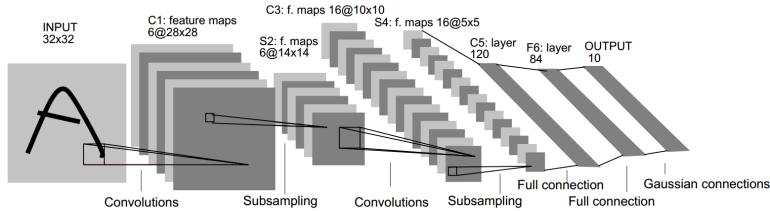


图 2.1 Lenet 模型图 [1]

Figure 2.1 Architecture of Lenet[1]

在 2012 年，在 ImageNet 图像分类竞赛四周年之际，Hinton 等人凭借 AlexNet^[36]以超过第二名近 12% 的准确率获得该竞赛冠军，自此 CNN 在计算机视觉领域开始称霸。在 2015 年，在改进了 CNN 的激活函数后，CNN 在该比赛上性能第一次超过了人类。之后随着网络层数不断加深，模型也越来越复杂，到诸如 MSRA 提出的 152 层 Resnet 模型^[9]甚至上千层网络已司空见惯。

CNN 是一种层次模型，其输入是原始数据，如 RGB 图像、原始音频数据等，经过卷积层的卷积操作，池化层的池化操作以及非线性激活函数映射等一系列操作的层层堆叠，将低层语义不断抽化为高层语义，这一过程为前向传播。最终，CNN 的最后一层将其目标任务形式化为目标函数，通过计算与真实标签之间的损失，经由反向传播算法将损失逐层向前反馈，更新每层参数，并在更新中再次前向传播，如此反复，直到网络模型收敛，达到模型训练的目的。其中几个关键模块分别介绍如下。

- 卷积

在卷积神经网络中，卷积的主要目的是从输入图像中提取特征。通过使用输入数据中的小方块（卷积核）来学习图像特征，卷积保留了像素间的空间关系。每个图像可以被看做像素值矩阵，一个卷积核在输入图像上移动（卷积操作）以生成特征映射。在同一张图像上，不同卷积核的卷积会捕获原始图像中的不同局部依赖关系，进而得到不同的特征信息。

- 池化

空间池化（也称为子采样或下采样）可降低每个特征映射的维度，并保留最重要的信息。空间池化有几种不同的方式：最大值，平均值，求和等。

在最大池化的情况下定义一个空间邻域（例如，一个 2×2 窗口），并取该窗口内最大的元素值。不过也可以取该窗口内所有元素的平均值（平均池化）或所有元素的总和。在实际运用中，最大池化的表现更好。池化的作用是逐步减少输入的空间大小。

- 非线性激活函数映射

每次卷积之后，都进行了另一项称为 ReLU 的操作，ReLU 全称为修正线性单元（Rectified Linear Units），是一种非线性操作。

ReLU 是一个针对元素的操作（应用于每个像素），并将特征映射中的所有负像素值替换为零。ReLU 的目的是在卷积神经网络中引入非线性因素，因为在实际生活中用神经网络学习的数据大多数都是非线性的（卷积是一个线性运算——按元素进行矩阵乘法和加法，所以希望通过引入 ReLU 这样的非线性函数来解决非线性问题）。其他非线性函数诸如 tanh 或 sigmoid 也可以用来代替 ReLU，但是在大多数情况下，ReLU 的表现更好。

- 全连接层

完全连接层是一个传统的多层感知器，它在输出层使用 softmax 激活函数（也可以使用其他分类器，比如 SVM）。“完全连接”这个术语意味着前一层中的每个神经元都连接到下一层的每个神经元。卷积层和池化层的输出代表了输入图像的高级特征。完全连接层的目的是利用这些基于训练数据集得到的特征，将输入图像分为不同的类。除分类之外，添加完全连接层也是一个比较简单的学习这些特征非线性组合的方式，通过在完全连接层的输出层使用了 softmax 激活函数可以使得完全连接层的输出概率之和为 1。

2.2 物体检测模型

从图像中解析出可供计算机分析的特征，是计算机视觉的中心问题。深度学习模型由于其强大的特征学习和表示能力，加之海量的数据量和进步的计算力，成为计算机视觉的热点研究方向。根据后续任务的需要，有三个主要的层次，如图2.2。

一是分类（Classification），即是将图像结构化为某一类别的信息，用事先确定好的类别来对图片进行识别分类。这一任务是最简单、最基础的图像理解任务，也是深度学习模型最先取得突破和实现大规模应用的任务。其中，以 ImageNet

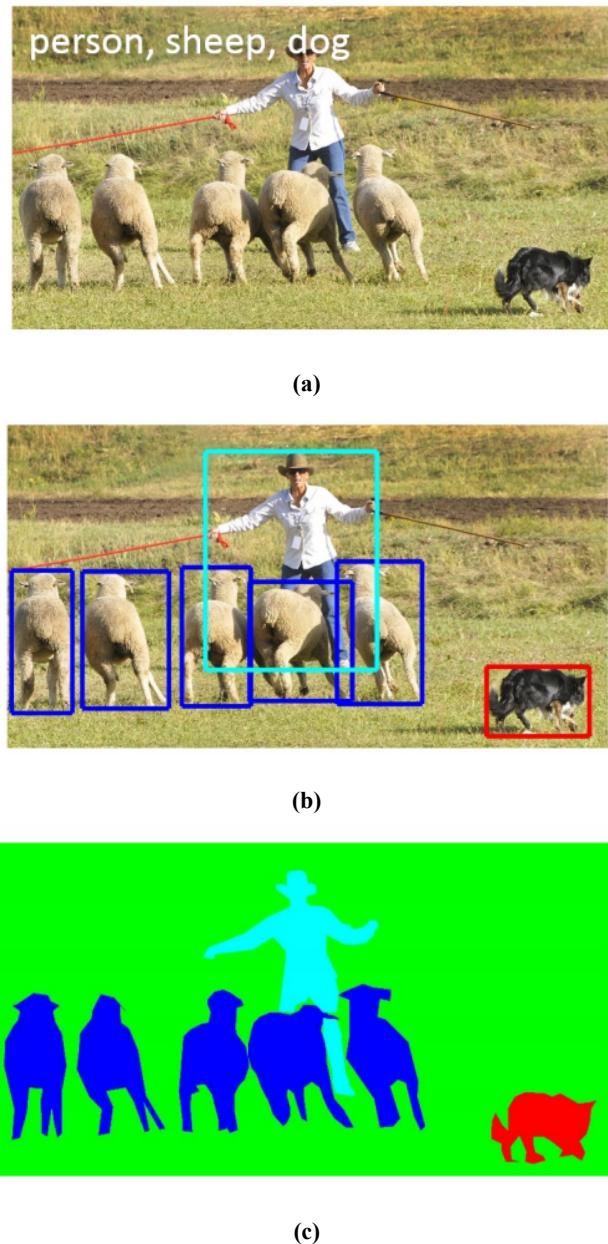


图 2.2 计算机视觉主要任务 (a) 图片分类 (b) 物体检测 (c) 物体分割 [2]

Figure 2.2 Main tasks of computer vision(a)classification(b)detection(c)segmentation [2]

是评测集为代表的每年 ILSVRC 催生了大量的优秀深度网络结构，为其他任务提供了基础。在应用领域，人脸、场景等识别等都可以归为分类任务。

二是检测（**Detection**）。分类任务关心整体，给出的是整张图片的内容描述，而检测则关注特定的物体目标，要求同时获得这一目标的类别信息和位置信息。相比分类，检测给出的是对图片前景和背景的理解，需要从背景中分离出感兴趣的目标，并确定这一目标的描述（类别和位置），因而，检测模型的输出是一个列表，列表的每一项使用一个数据组给出检出目标的类别和位置（常用矩形检测框的坐标表示）。

三是分割（**Segmentation**）。分割包括语义分割（semantic segmentation）和实例分割（instance segmentation），前者是对前背景分离的拓展，要求分离开具有不同语义的图像部分，而后者是检测任务的拓展，要求描述出目标的轮廓（相比检测框更为精细）。分割是对图像的像素级描述，它赋予每个像素类别（实例）意义，适用于理解要求较高的场景，如无人驾驶中对道路和非道路的分割。

物体检测领域是 CNN 应用的相对较为广泛的项目，是本文主要采用的方法。检测模型整体上由基础网络和检测头部构成，前者作为特征提取器，给出图像不同大小、不同抽象层次的表示；后者则依据这些表示和监督信息学习类别和位置关联。检测头部负责的类别预测和位置回归两个任务常常是并行进行的，构成多任务的损失进行联合训练。

2.2.1 两阶段物体检测模型

两阶段物体检测模型因其对图片的两阶段处理得名，也称为基于区域（Region-based）的方法，其中又以 R-CNN 系列工作为这一系列的代表。

- R-CNN[3]

R-CNN（Regions with CNN features）作为 R-CNN 系列的开山之作，模型如图2.3所示。

R-CNN 将检测抽象为两个过程，一是基于图片提出若干可能包含物体的区域（即图片的局部裁剪，被称为 Region Proposal），文中使用的是 Selective Search 算法 [45]；二是在提出的这些区域上运行当时表现最好的分类网络（AlexNet）[36]，得到每个区域内物体的类别。R-CNN 提出 CNN 可用于基于区域的定位和分割物体，而且当监督训练样本数紧缺时，在额外的数据上预训练的模型经过 fine-tuning 可以取得很好的效果。这种基于使用分类任务中训练好的模型作为

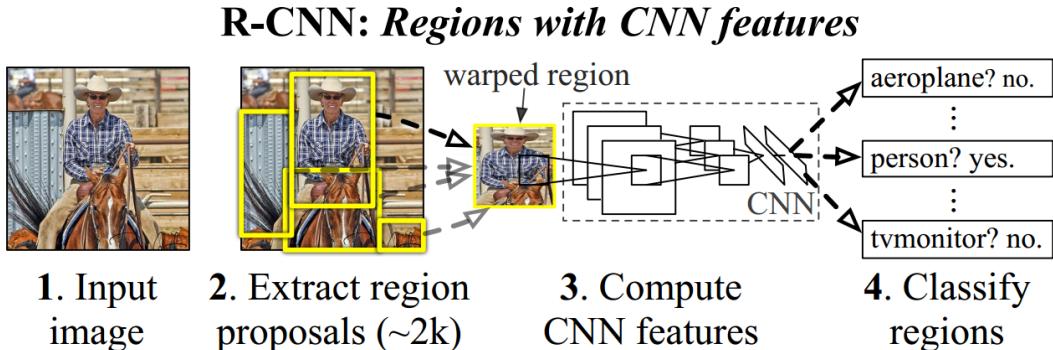


图 2.3 R-CNN 模型图 [3]

Figure 2.3 Architecture of R-CNN[3]

基网络，在检测问题上 fine-tuning 的做法一直被后面的两阶段模型所使用。模型本身存在的问题也很多，如需要训练三个不同的模型（Proposal, Classification, Regression）、重复计算过多导致的性能问题等。尽管如此，这篇论文的很多做法仍然广泛地影响着检测任务上的深度学习方法，后续的很多工作也都是针对改进这一工作而展开。

- Fast R-CNN[4]

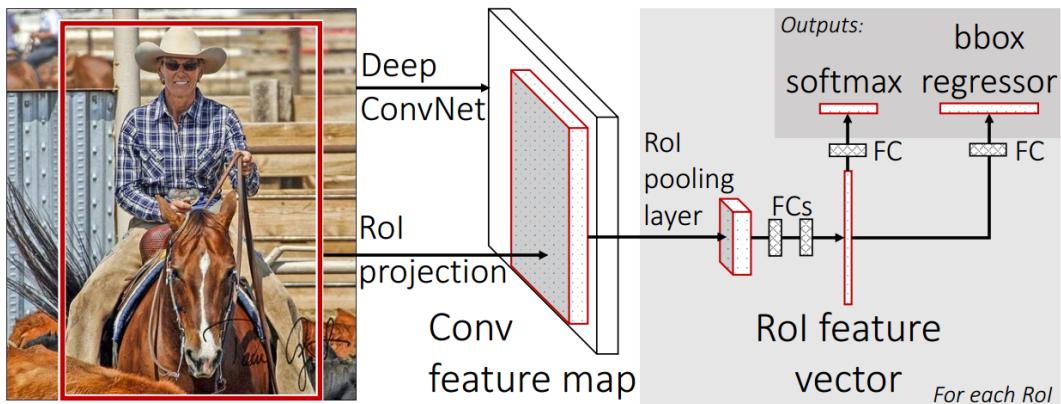


图 2.4 Fast R-CNN 模型图 [4]

Figure 2.4 Architecture of Fast R-CNN[4]

R-CNN 耗时的原因是 CNN 是在每一个 Proposal 上单独进行的，没有共享计算，该模型将基础网络在图片整体上运行完毕后，再传入 R-CNN 子网络，共享了大部分计算，故有 Fast 之名。

如图2.4是 Fast R-CNN 的架构图。图片经过 feature extractor 得到 feature map,

同时在原图上运行 Selective Search 算法并将 RoI (Region of Interest, 实为坐标组, 可与 Region Proposal 混用) 映射到 feature map 上, 再对每个 RoI 进行 RoI Pooling 操作便得到等长的 feature vector, 将这些得到的 feature vector 进行正负样本的整理 (保持一定的正负样本比例), 传入并行的 R-CNN 子网络, 进行分类和回归, 并将两者的损失统一起来。

其中 RoI Pooling 是对输入 R-CNN 子网络的数据进行准备的关键操作。RPN 得到的区域常常有不同的大小, 在映射到 feature map 上之后, 会得到不同大小的特征张量。RoI Pooling 先将 RoI 等分成目标个数的网格, 再在每个网格上进行 max pooling, 就得到等长的 RoI feature vector。文章将 Proposal, Feature Extractor, Object Classification 和 Localization 统一在一个整体的结构中, 并通过共享卷积计算提高特征利用效率, 是最有贡献的地方。

- Faster R-CNN[5]

Faster R-CNN 是 2-stage 方法的奠基性工作, 提出的 RPN 网络取代 Selective Search 算法使得检测任务可以由神经网络端到端地完成。粗略的讲, Faster R-CNN = RPN + Fast R-CNN, 跟 R-CNN 共享卷积计算的特性使得 RPN 引入的计算量很小, 使得 Faster R-CNN 可以在单个 GPU 上以 5fps 的速度运行, 而在精度方面达到 SOTA (State of the Art, 当前最佳)。

Faster R-CNN 的主要贡献是提出 Regional Proposal Networks, 替代之前的 Selective Search 算法。RPN 网络将 Proposal 这一任务建模为二分类 (是否为物体) 的问题。

如图2.5第一步是在一个滑动窗口上生成不同尺寸(scales)和长宽比例(aspect ratio)的 anchor box, 取定 IoU 的阈值, 按 ground-truth 标定这些 anchor box 的正负。于是, 传入 RPN 网络的样本数据被整理为 anchor box (坐标) 和每个 anchor box 是否有物体 (二分类标签)。RPN 网络将每个样本映射为一个概率值和四个坐标值, 概率值反应这个 anchor box 有物体的概率, 四个坐标值用于回归定义物体的位置。最后将二分类和坐标回归的损失统一起来, 作为 RPN 网络的目标训练。

由 RPN 得到 Region Proposal 在根据概率值筛选后经过类似的标记过程, 被传入 R-CNN 子网络, 进行多分类和坐标回归, 同样用多任务损失将二者的损失联合。

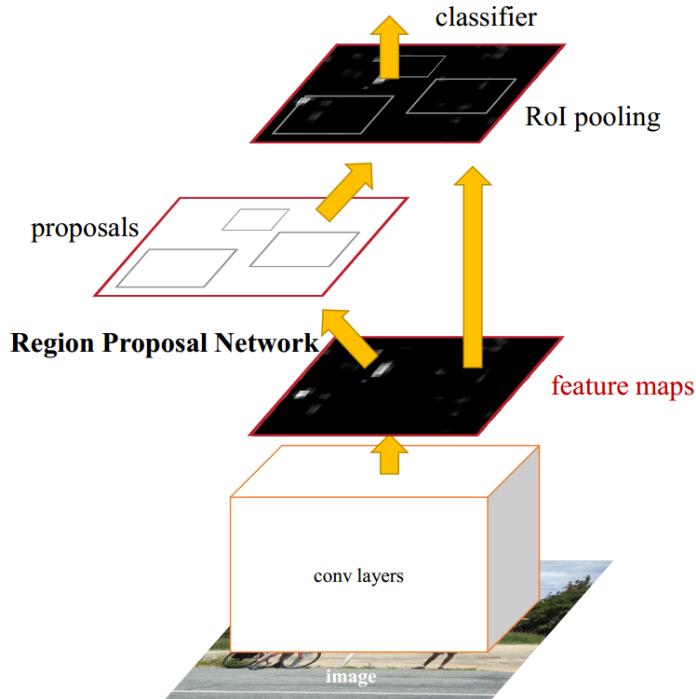


图 2.5 RPN 模型 [5]

Figure 2.5 Architecture of RPN[5]

遵循 multi-task loss 定义，最小化目标函数，Faster R-CNN 中对一个图像的损失函数定义为：

$$L(\{p_i\}, \{u_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (2.1)$$

这个可以考虑为从 anchor box 回归到附近的 ground-truth box。其中 p_i 为 anchor 预测为目标的概率, GT 标签: $p_i^* = \begin{cases} 0 & \text{negative label} \\ 1 & \text{positive label} \end{cases}$, $t_i = \{t_x, t_y, t_w, t_h\}$ 一个向量, 表示预测的 bounding box 包围盒的 4 个参数化坐标, t_i^* 是与 positive anchor 对应的 ground-truth 包围盒的坐标向量, 具体计算如下:

$$\begin{aligned} t_x &= (x - x_a) / w_a; t_y = (y - y_a) / h_a \\ t_w &= \log(w / w_a); t_h = \log(h / h_a) \\ t_x^* &= (x - x_a) / w_a; t_y^* = (y^* - y_a) / h_a \\ t_w^* &= \log(w^* / w_a); t_h^* = \log(h^* / h_a) \end{aligned} \quad (2.2)$$

$L_{cls}(p_i, p_i^*)$ 是两个类别 (目标 vs 非目标) 之间的对数损失, 为交叉熵损失

函数，计算如下：

$$L_{cls}(p_i, p_i^*) = -\log [p_i^* p_i + (1 - p_i^*)(1 - p_i)] \quad (2.3)$$

$L_{reg}(t_i, t_i^*)$ 是回归损失函数，计算如下：

$$L_{reg}(t_i, t_i^*) = R(t_i - t_i^*) \quad (2.4)$$

其中 R 是 smooth L1 损失函数。 $p_i^* L_{reg}$ 表示只有前景 anchor ($p_i^* = 1$) 才有回归损失，其他情况则没有。cls 层和 reg 层的输出分别由 $\{p_i\}, \{u\}$ 控制，这两项分别由 N_{cls} 和 N_{reg} 以及一个平衡权重 λ 归一化。

Faster R-CNN 的成功之处在于用 RPN 网络完成了检测任务的深度化。使用滑动窗口生成 anchor box 的思想也在后来的工作中越来越多地被采用。这项工作奠定了 RPN+RCNN 的两阶段结构，影响了大部分后续工作。

2.2.2 单阶段物体检测模型

- YOLO[6] 如图2.6YOLO 的工作流程如下：

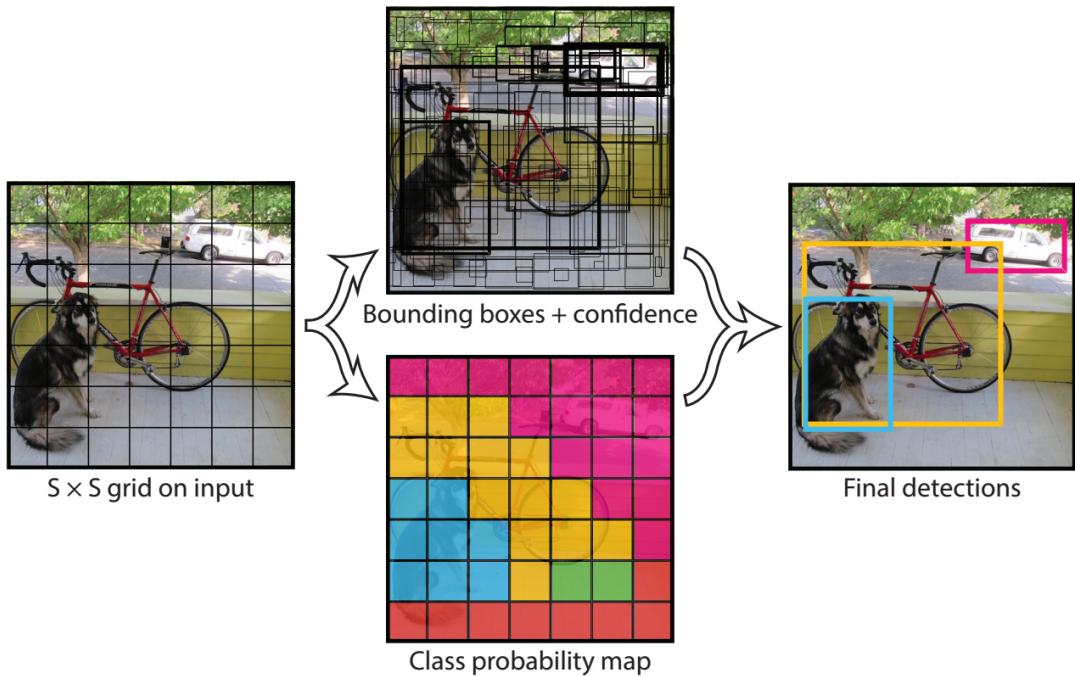


图 2.6 YOLO 模型 [6]

Figure 2.6 Architecture of YOLO[6]

1. 准备数据：将图片缩放，划分为等分的网格，每个网格按跟 ground-truth 的 IoU 分配到所要预测的样本。

2. 卷积网络：由 GoogLeNet[46] 更改而来，每个网格对每个类别预测一个条件概率值，并在网格基础上生成 B 个 box，每个 box 预测五个回归值，四个表征位置，第五个表征这个 box 含有物体（不是某一类物体）的概率和位置的准确程度（由 IoU 表示）。测试时，分数如下计算：

$$\Pr(\text{Class}_i | \text{Object})^* \Pr(\text{Object})^* \text{IOU}_{pred}^{truth} = \Pr(\text{Class}_i)^* \text{IOU}_{pred}^{truth} \quad (2.5)$$

等式左边第一项由网格预测，后两项由每个 box 预测，以条件概率的方式得到每个 box 含有不同类别物体的分数。因而，卷积网络共输出的预测值个数为 $S \times S \times (B \times 5 + C)$ ，其中 S 为网格数，B 为每个网格生成 box 个数，C 为类别数。

3. 后处理：使用 NMS（Non-Maximum Suppression，非极大抑制）[47] 过滤得到最后的预测框

非极大值抑制的方法是：先假设有 6 个矩形框，根据分类器的类别分类概率做排序，假设从小到大属于车辆的概率分别为 A、B、C、D、E、F。

- 从最大概率矩形框 F 开始，分别判断 A、E 与 F 的重叠度 IOU 是否大于某个设定的阈值。
- 假设 B、D 与 F 的重叠度超过阈值，那么就扔掉 B、D；并标记第一个矩形框 F，是保留下来的。
- 从剩下的矩形框 A、C、E 中，选择概率最大的 E，然后判断 E 与 A、C 的重叠度，重叠度大于一定的阈值，那么就扔掉；并标记 E 是保留下来的第二个矩形框。就这样一直重复，找到所有被保留下来的矩形框。

损失函数被分为三部分：坐标误差、物体误差、类别误差。为了平衡类别不均衡和大小物体等带来的影响，损失函数中添加了权重并将长宽取根号。YOLO 提出了单阶段的新思路，相比两阶段方法，其速度优势明显，同时可以实时检测。但 YOLO 本身也存在一些问题，如划分网格较为粗糙，每个网格生成的 box 个数等限制了对小尺度物体和相近物体的检测。

- SSD[7]

如图2.7，SSD（Single Shot Detector）相比 YOLO 有以下突出的特点：

- 多尺度的 feature map：基于 VGG 的不同卷积段，输出 feature map 到回归器中。这一点试图提升小物体的检测精度。
- 更多的 anchor box，每个网格点生成不同大小和长宽比例的 box，并将类别预测概率基于 box 预测（YOLO 是在网格上），得到的输出值个数为 $(C+4) \times k \times m \times n$ ，

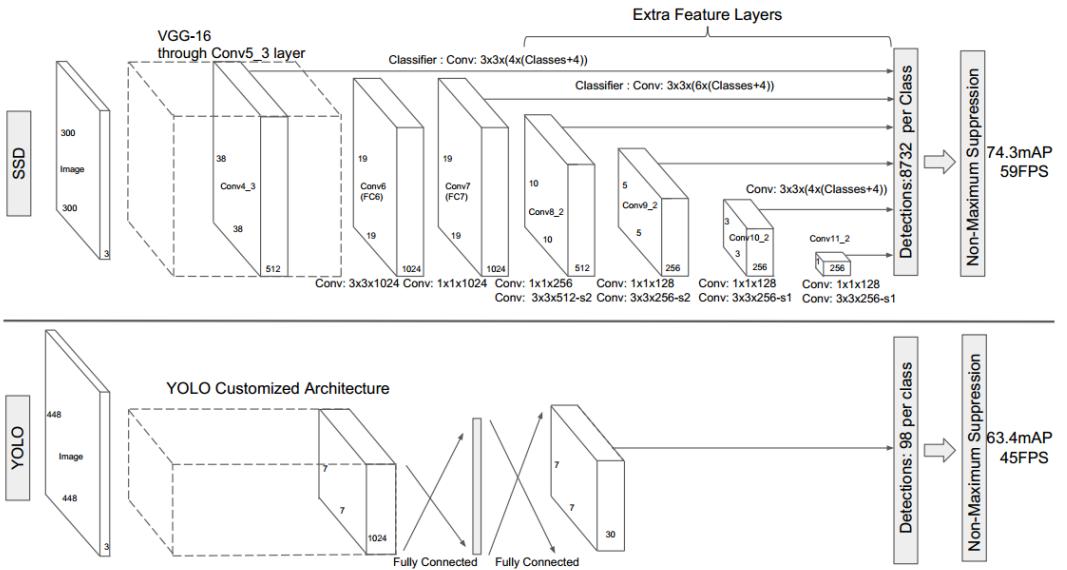


图 2.7 SSD 模型 [7]

Figure 2.7 Architecture of SSD[7]

其中 C 为类别数，k 为 box 个数， $m \times n$ 为 feature map 的大小。

SSD 是单阶段模型早期的集大成者，达到跟接近两阶段模型精度的同时，拥有比两阶段模型快一个数量级的速度。后续的单阶段模型工作大多基于 SSD 改进展开。

2.3 迁移学习

在传统计算机视觉领域，由于图片标注成本小，数据量往往不是一个主要的问题。而对于医学影像，存在需要专业医生标注，数据获取困难等问题，针对现有的少量数据集进行研究就成为核心问题。如图2.8，迁移学习 (Transfer learning)[48] 就是把已学训练好的模型参数迁移到新的模型来帮助新模型训练。考虑到大部分数据或任务是存在相关性的，所以通过迁移学习可以将已经学到的模型参数（也可理解为模型学到的知识）通过某种方式来分享给新模型从而加快并优化模型的学习效率不用像大多数网络那样从零学习。其初衷是节省人工标注样本的时间，让模型可以通过已有的标记数据 (source domain data) 向未标记数据 (target domain data) 迁移，从而训练出适用于 target domain 的模型。

迁移学习按照学习方式可以分为基于样本的迁移，基于特征的迁移，基于模型的迁移，以及基于关系的迁移。基于样本的迁移通过对源域中有标定样本的加

权利用完成知识迁移；基于特征的迁移通过将源域和目标域映射到相同的空间（或者将其中之一映射到另一个的空间中）并最小化源域和目标域的距离来完成知识迁移；基于模型的迁移将源域和目标域的模型与样本结合起来调整模型的参数；基于关系的迁移则通过在源域中学习概念之间的关系，然后将其类比到目标域中，完成知识的迁移。主要的迁移学习方法如下：

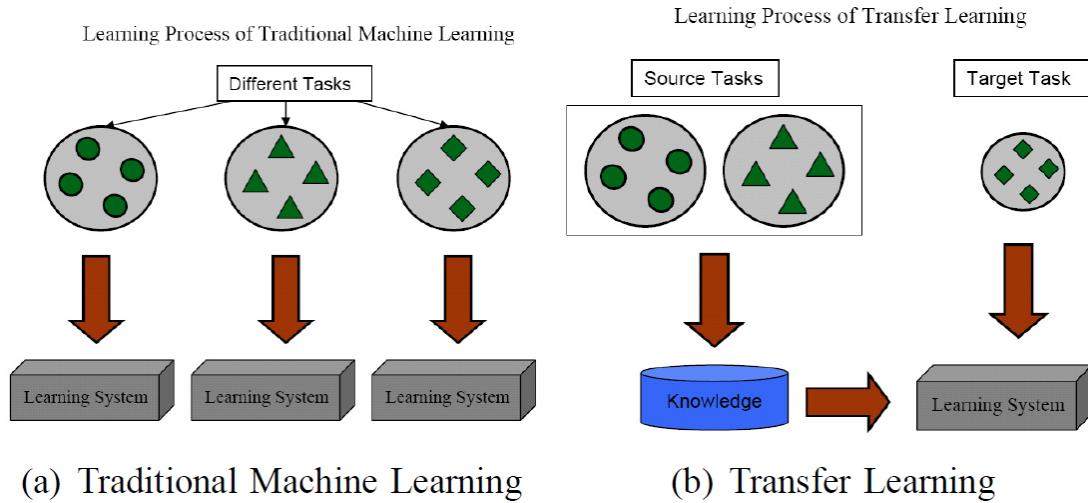


图 2.8 迁移学习与传统机器学习的不同。**(a)** 传统机器学习对不同的学习任务建立不同的模型，**(b)** 迁移学习利用源域中的数据将知识迁移到目标域，完成模型建立。

Figure 2.8 The difference between Migration learning and traditional machine learning. **(a)** Traditional machine learning establishes different models for different learning tasks. **(b)** Migration learning uses data from the source domain to migrate knowledge to the target domain and complete model building.

- ConvNet 固定特征提取器

取出 ImageNet ConvNet 预训练好的模型，移除最后一个全连接层（该层的输出是一个不同的任务，如 ImageNet 1000 级评分），然后把剩下的 ConvNet 作为新数据集的固定特征提取器。模型将计算每个图像包含的激活层之前的隐藏层所提取出来的特征，之后将特征放入一个线性分类器（如线性 SVM 和 softmax 分类器）进行分类从而得到最终的结果。在乳腺钼靶图片的小数据集上，可以采用在传统深度学习模型上固定相关层从而获取浅层特征，进而将获取到的浅层特征与深层特征进行连接重新进行训练，从而得到最终分类结果。

- 微调 ConvNet

在新的数据集上，这种策略不仅是要更换和重训练 ConvNet 的顶层结构，也需要微调网络中训练出来的网络权重。可能微调 ConvNet 的所有层，或把之前

的一些层固定是可能的（由于过度拟合问题）只是微调一些网络的顶层部分。ConvNet 早期特征中包含更多的一般特征（如边缘检测、彩色斑点探测器）应该是许多任务有用，但 ConvNet 的顶层部分由于细节越来越具体，使得在新数据集上并不完全适应。

迁移学习一般要考虑的图片分为几种情况。数据集小并且类似于原有数据集，数据集大而且类似于原有数据集，数据集小但和原有数据集差别大，数据集大但和原有数据集差别大。

对于乳腺钼靶图片问题，数据与传统的计算机视觉数据集差别大，而且数据量小，这属于迁移学习中最难的问题，由于特征之间差别小，所以使用固化顶层模块来训练模型就变得不再有效，需要针对此问题要么提升特征提取器的能力，要么设计全新的模块或损失函数。

2.4 弱监督学习

机器学习在各种任务中取得了巨大成功，特别是在分类和回归等监督学习任务中。预测模型是从包含大量训练样本的训练数据集中学习，每个训练样本对应一个事件或对象。训练样本由两部分组成：一个描述事件/对象的特征向量（或示例），以及一个表示真值输出的标签。在分类任务中，标签表示训练样本所属的类别；在回归任务中，标签是一个与样本对应的实数值。大多数成功的技术，如深度学习，都需要含有真值标签的大规模训练数据集，监督学习技术在具备强监督信息（如大量具备真值标签的训练样本）的情况下取得了很大成功。总的来说，监督学习就是正常用的，有一批高置信的标注数据，通过模型来拟合效果。然而，在许多任务中，由于数据标注过程的成本极高，很难获得强监督信息，因此，使用弱监督学习通常是更好的方式。弱监督学习 [49]，就是很难获取足够量的高置信的标注数据，所以弱监督学习就是来解决这个问题。

通常来说，弱监督可以分为三类。第一类是不完全监督（incomplete supervision），即只有训练集的一个（通常很小的）子集是有标签的，其他数据则没有标签。这种情况发生在各类任务中。例如，在图像分类任务中，真值标签由人类标注者给出的。从互联网上获取巨量图片很容易，然而考虑到标记的人工成本，只有一个子集的图像能够被标注。第二类是不确切监督（inexact supervision），即图像只有粗粒度的标签。第三种是不准确的监督（inaccurate supervision），模

型给出的标签不总是真值。出现这种情况的常见原因有，图片标注者不小心或比较疲倦，或者某些图片就是难以分类。

弱监督方法常见的步骤先是利用监督信息训练，之后去人为预测未监督信息，使其带有人为的标签，然后再次重点去识别分类二者区别。而这也是乳腺钼靶整图分类的常见做法，获取可以病灶区域 patch，进而进行再次细致识别。本文的数据属于给定了监督信息，但信息不够精确的场景，对于物体检测问题而言，同时拥有标注和分类信息才是模型能否学习充分的关键，然而该项目存在只有恶性数据含有标注信息，而正常数据没有标注信息的问题，所以在进行模型具体设计时，需要引入这种弱监督学习的方法。

2.5 度量学习

度量学习 (Metric Learning)[50]，更确切地可以说是距离度量学习 (Distance Metric Learning, DML)、相似度学习。度量学习是人脸识别中常用传统机器学习方法，由 Eric Xing 在 NIPS 2002 提出。分为两种，一种是基于监督学习的，另外一种是基于非监督学习的，根据不同的任务来自主学习出针对某个特定任务的度量距离函数。通过计算两张图片之间的相似度，使得输入图片被归入到相似度大的图片类别中去，其中以孪生网络最为著名。

孪生网络 (Siamese network) [8] 是一种网络结构，通过一个神经网络将样本的维度降低到某个较低的维度。

在低维空间，任意两个样本：

- 如果它们是相同类别，空间距离尽量接近 0；
- 如果它们是不同类别，空间距离大于某个间隔。

Siamese network 就是“连体的神经网络”，神经网络的“连体”是通过共享权值来实现的，如图2.9所示。

孪生网络的目的是为了比较两幅图片是否相似，或者说相似度多少。输入是两张需要对比的图片，输出则是一个相似度数值。孪生网络现在越来越广泛应用于人脸识别，图像块匹配，跟踪等问题，并在此基础上改进和更新了诸多模块和损失函数。Siamese network 的初衷是计算两个输入的相似度，左右两个神经网络分别将输入转换成一个“向量”，在新的空间中，通过判断 cosine 距离就能得到相似度。cosine、exp function、欧式距离均可以作为相似度距离。模型训练的目

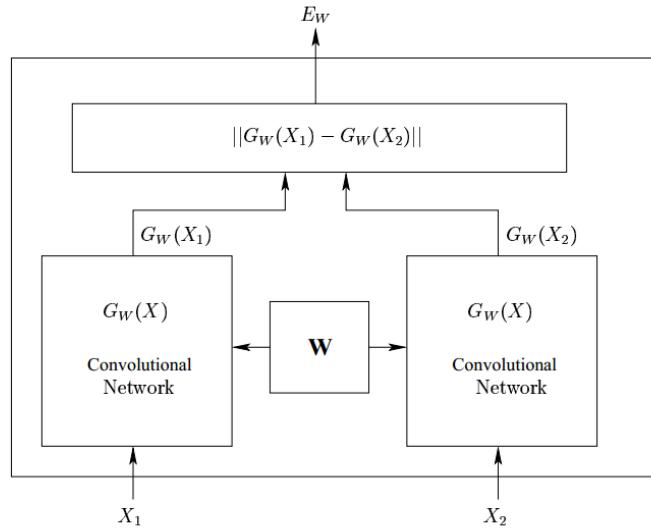


图 2.9 李生网络图 [8]

Figure 2.9 Architecture of Siamese network[8]

目标是让两个相同的输入距离尽可能的小，两个不同类别的输入距离尽可能的大。根据实验分析，cosine 更适用于词汇级别的语义相似度度量，而 exp 更适用于句子级别、段落级别的文本相似性度量。

在本文中，由于需要模型去区分疑似的恶性病灶和恶性病灶之间的区别，所以需要引入李生网络设计思想及其损失函数。

2.6 本章小结

本章在对数据进行充分分析的情况下，具体介绍了本文所使用到的相关深度学习技术，包括卷积神经网络、基于 CNN 的物体检测模型、迁移学习、弱监督学习和度量学习等，为后续项目的开展提供理论支持。

第3章 数据分析与处理

本论文以天津肿瘤医院的 721 位病例图片数据为基础，其中包含 304 例恶性病例，417 例正常病例；图片数有 2908 张，其中包含恶性图片数据有 1254 张，正常图片数据有 1654 张。医学数据存在恶性和正常两大分类，对于恶性数据医生容易标注出具体的癌变病灶位置，对于正常样本，包括拥有疑似恶性病灶的良性数据医生难以精确标出，这使得基于深度学习的模型设计存在一定的难度。每个病例数据大多拥有两个乳房不同角度的图片，这使得实际应用中应当考虑标准方面的设计。

3.1 数据分析

3.1.1 医学数据分析

既往我国大多是医疗机构均采用纸质文件记录医疗数据及医疗活动，这些纸质文件数据属于非结构化数据，利用起来非常困难，虽然早就有巨大的数据量，但无法利用。近年来，我国卫生行政部门大力推进以健康档案、电子病历和公共服务信息平台为基础的区域卫生信息化建设工作。2010 年“十二五”卫生信息化建设工程规划编制工作初步确定了我国卫生信息化建设路线图。随着卫生信息化建设的不断成熟，医疗相关的大数据也在急剧增加。医院信息管理系统（HIS），主要包括电子病历信息系统（医嘱、病程记录、护理记录等）、实验室信息管理系统（检验报告）、医学影像系统（各种医学影像如 MRI、CT、X 光片等），这些系统几乎每分每秒都在产生电子化数据，数据量实在是太大、增长太快，数据量从 MB 到 GB，从 TB 到 PB，对数据处理的实时性、有效性提出更高要求，传统的分析技术无法应付。医学的数据常常是海量的，格式丰富多样，为深度学习的训练提供了有效的数据来源。

另一方面，在传统模式下，数据多数是由不同的应用程序搜集到的，存储格式不一，无法彼此兼容、无法整合，各个数据库就像一个个相互隔离的岛屿，由此产生了“数据孤岛”的概念。比如我国人口死亡登记系统实际上保存着大量的、相对准确的人口死亡信息，包括死亡日期和死亡原因等，这些数据对于评价临床医疗结局无疑是宝贵的原始资料。但是由于各种原因，这部分数据与医院的

医疗方面的数据至今无法整合，令人惋惜。再加上存在着不同的地方存在着器官组织的差异，比如乳腺腺体，东方女性偏致密性，所以基于医学数据的项目在中国目前大部分只能分别考虑，基于东方女性特点的乳腺钼靶分类就变得很有实际意义，而且对模型的可拓展性也提出了较高的要求。

医疗影像数据一般尺度大，以乳腺钼靶为例，尺度可以达到 2000-4000 像素左右，而且格式丰富多样，有包括病人各种信息的 dicom 格式，有包括一个病人多个切片拍摄的 CT 数据等等，所要识别的目标，无论是器官还是肿瘤病灶一般较小。在所有的数据里，影像的数据跟临床病理比较起来，它的标准化、格式化、统一性是最强的。在图像领域，传统的自然图像一般尺度小，所要识别分类的目标一般较大而且目标与目标之间区别较大，即使属于细粒度分类检测问题，细粒度目标之间的差异也是可以通过肉眼进行区分，而在医疗影像数据中类别之间的差异并不明显。在肿瘤识别中，医学数据常分为阴性，良性和恶性三种类型，其中阴性是不包含任何疑似肿瘤病灶的特征，良性包含疑似恶性肿瘤的信息，而恶性包含恶性肿瘤病灶。基于此，医学会提出各种经验性的指标判断一个病人是否得癌，然后基于病理报告进一步确诊。肿瘤的形态，大小，分布往往各异，不具有统计学上的特征，良性肿瘤和恶性肿瘤之间的差别往往只是在于边缘的细微区别，如何有效的识别并区分这细小的不同便成为医学以及深度学习模型得以学习的关键。同时特别不一样的是，影像数据很难像语音数据或者文本数据一样将标注任务外包出去，影像数据的标注大部分只能依靠专业人士，这使得数据的采集在医疗影像问题上变得难度极大，数据量往往较少也成为制约模型能否具有鲁棒性的关键。

对于医生而言，医学数据的指标往往更加关注敏感度，特异度等指标。敏感度 (sensitivity, SEN): 又称真阳性率 (true positive rate, TPR)，即实际患病又被诊断标准正确地诊断出来的百分比。特异度是实际无病按该诊断标准被正确地判为无病的百分比。特异度 (specificity, SPE)，又称真阴性率 (true negative rate, TNR)，即它反映筛检试验确定非病人的能力。对于医生而言，医生希望在保证不过高的误诊率的基础上，能够降低病人的误诊率。医生考虑的层面和计算机视觉也有所不同，如果一个病人拥有多种不同格式以及在不同时间段采集的数据，他们更关心在病人层面判断该病人是否得癌，是否需要进行手术，是否需要做下一步的病理分析等，进而基于实际情况出发，会影响到基于医学数据的深度学习模

型的标准制定。

综上，在医学数据分析的基础上，使用深度学习技术，往往包括以下流程，第一步，对医疗影像数据标注从而获取模型训练所需要的数据资源；第二步，选择合适的模型对数据进行训练，进而对模型进行测试，反馈测试结果给训练模型，对模型进行调参，优化模型；第三步，在模型测试反馈测试结果的同时，也反馈数据标注的问题，进一步优化数据；第四步，基于实际出发，设计相应的合适指标，对模型进行评估，如果最终需要产业化的话，还需要落地模型，对产品评估等。如图3.1。

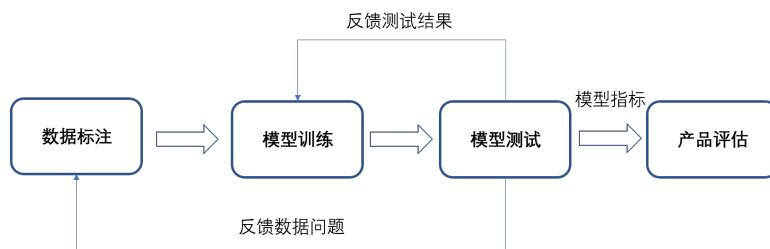


图 3.1 基于深度学习的医学数据工作流程

Figure 3.1 Medical data workflow based on deep learning

在上述流程中，数据标注的制定标准，数据本身的特性以及模型指标的确定都是医学数据进行深度学习技术实践中区别于传统的自然图像的地方。在医学影像数据中，尤其需要对数据进行充分的认识，只有对数据标注提出适应于数据的标注要求及模型评估指标，才可以充分的发挥深度学习的作用，更好地服务于实际情况。

3.1.2 乳腺钼靶数据分析

女性在现在的生活工作中担负起了越来越重要的角色，相应的生活工作压力也随之而来，与此同时疾病也不请自来。乳腺癌是女性最常见的浸润性癌症，也是仅次于肺癌的女性癌症死亡的第二大凶手。女性乳腺是由皮肤、纤维组织、乳腺腺体和脂肪组成的，乳腺癌是发生在乳腺腺上皮组织的恶性肿瘤。乳腺癌中99%发生在女性，男性仅占1%。目前乳腺癌已成为威胁女性身心健康的常见肿瘤。全球乳腺癌发病率自20世纪70年代末开始一直呈上升趋势。美国8名妇女一生中就会有1人患乳腺癌。中国不是乳腺癌的高发国家，但不宜乐观，近年我国乳腺癌发病率的增长速度却高出高发国家1~2个百分点。据国家癌症中心和卫生部疾病预防控制局2012年公布的2009年乳腺癌发病数据显示，全国肿瘤

登记地区乳腺癌发病率位居女性恶性肿瘤的第 1 位，女性乳腺癌发病率（粗率）全国合计为 42.55/10 万，城市为 51.91/10 万，农村为 23.12/10 万。乳腺癌已成为当前社会的重大公共卫生问题。

自 20 世纪 90 年代全球乳腺癌死亡率呈现出下降趋势。究其原因，一是乳腺癌筛查工作的开展，使早期病例的比例增加；二是乳腺癌综合治疗的开展，提高了疗效。乳腺钼靶，全称乳腺钼靶 X 线摄影检查，又称钼钯检查，是目前诊断乳腺疾病的首选和最简便、最可靠的无创性检测手段，痛苦相对较小，简便易行，且分辨率高，重复性好，留取的图像可供前后对比，不受年龄、体形的限制，目前已作为常规的检查。乳腺钼靶以 dicom 格式进行保存，dicom 全称是医学数字图像与通讯，其文件结构如图3.2所示。dicom 整体结构先是 128 字节所谓的导言



图 3.2 dicom 数据格式

Figure 3.2 dicom data format

部分，接着就是四个字节组成的字符串，然后是数据元素 (dataElement) 依次排列到文件结尾。我们要读取 dicom 里面的各种数据就是在各个数据元素中。通俗的讲 dataElement 就是指 tag，就是 dicom 标准里定义的数据字典，每个 dataElement 中的 tag 决定自身或整个文件的某些数据类型或自身 dataElement 内容类别，从前到后按 tag 又可简单分类为：文件元 tag（定义了传输语法），普通 tag（包括图像宽，高，数据传输格式，病人姓名等基本信息），像素 tag（表示 dataElement 存储的是病例的图像数据）。dicom 中存储的信息量丰富，对于本文而言，重点关注的是乳腺钼靶的图像数据，也就是像素 tag。通过对乳腺钼靶 dicom 数据的分析，我们发现乳腺钼靶数据有以下几个特点：

- 医疗影像数据采集困难，使得数据集小

以公开的几个数据集为例，INBREAST[18] 有 116 个病例，410 张乳腺钼靶图片；而另一个比较大的乳腺钼靶公开数据集 DDSM[19–21] 也只有 2620 个病例，而且由于年代久远，缺乏维护，其中很多病例数据缺失；我们所采集的天津

肿瘤医院的数据，耗时医生半年多时间，最终才采集 721 个病例，由此可见医疗影像数据多么宝贵。正因为数据集小，所以对我们的模型训练提出更高的要求，因为深度学习得以发展的一个重要因素正是巨大的数据量。

- 尺寸大，平均有 4000×2000 像素，但病灶占比小，如图3.3所示

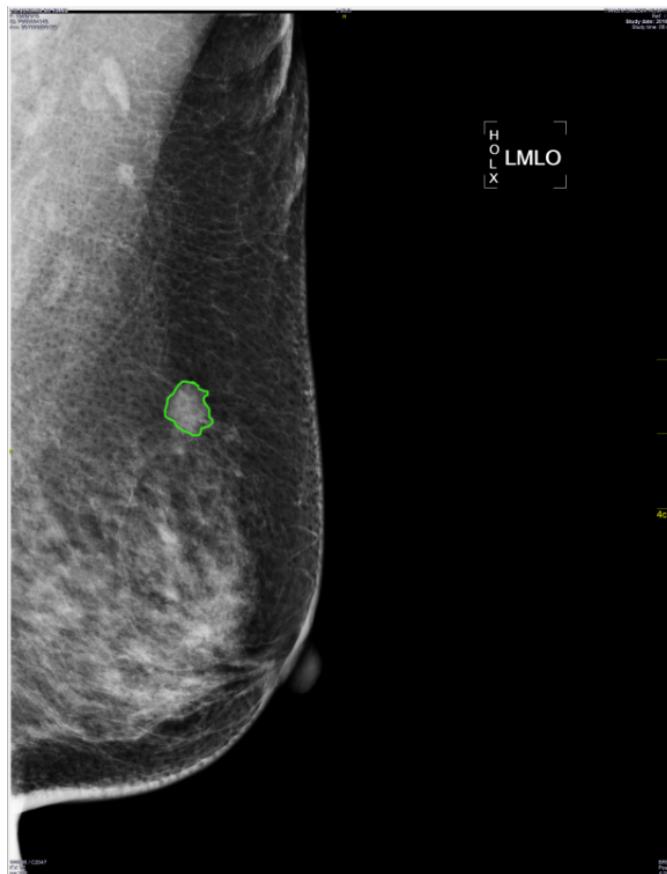


图 3.3 带标注信息的乳腺钼靶图片

Figure 3.3 Mammography with annotation

ImageNet 图像尺寸主要为 299×299 像素，乳腺钼靶尺寸大约是传统自然图像像素的十倍，尺度高，训练的维度高，如果直接载入模型，将导致资源消耗高。ImageNet 需识别物体占据图像的 70%-80%。以我们本文的乳腺钼靶数据为例，其中乳腺钼靶图片尺度为 3328×2560 像素，对乳腺钼靶检测的病灶宽度以高度进行直方图统计，如图3.4，所要识别的病灶占比不到 2%。

- 图片分为良性，恶性及阴性

由于所要考量的问题是肿瘤识别问题，所以图片有良性，恶性和阴性三种类型。按照医学以及实际情况出发，良性和阴性问题常常被判断为正常标签，与恶性标签区分。

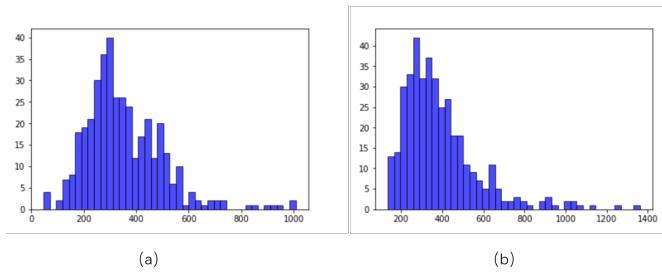


图 3.4 本文乳腺钼靶病灶宽高长度直方图 (a) 病灶宽度直方图 (b) 病灶高度直方图

Figure 3.4 The broad and high length histogram of mammography lesions in this project (a) histogram of lesion width (b) height histogram of lesion

- 正常数据未标注

由于数据工作量大，而且对于良性中疑似恶性病灶的标注困难，所以正常样本并未标注。

- 拥有图片，乳房及病人多层次考虑问题

乳腺钼靶设备会对女性乳房进行左右乳各个角度的拍摄，如图3.5。

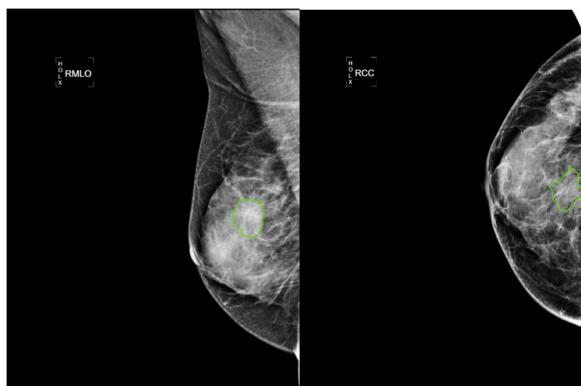


图 3.5 同一女性同一乳房不同角度拍摄图片

Figure 3.5 The same woman takes pictures of the same breast at different angles

在医学上，常用 MLO 及 CC 作为典型图片进行考量。传统的计算机视觉是以单张图片作为考量层次，而医学上更多关心的是基于单个乳房或者单个病人，在我们这个项目中，更多考虑的是病人是否患癌。这两个不同角度拍摄的图片可以决定一个病人的单一乳房是否得癌，只要有一侧乳房得癌，该病人就得癌。一个得癌病人并不是拍摄的所有角度乳房都会有恶性病灶。

- 正常样本中包含恶性样本信息

如图3.6所示，正常样本中，比如右边两张正常样本（良性和阴性）图片中存

在着疑似第一张恶性样本带实线的恶性病灶信息（图中用虚线表示）。

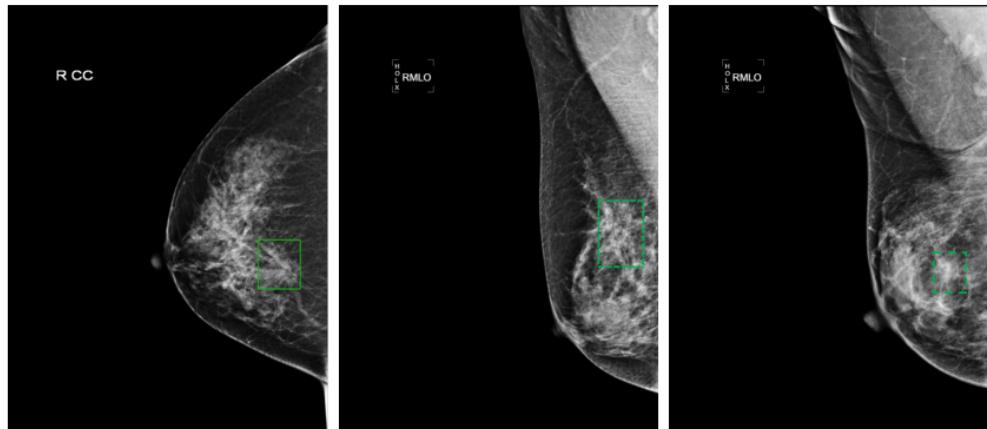


图 3.6 带有标注框的正样本及包含疑似恶性病灶的正常样本

Figure 3.6 The image with labeled malignant lesion(left) and healthy images contain un-annotated benign tumor region (middle) or normal region(right) which are suspect to be malignant in vision

3.2 数据处理

3.2.1 数据预处理

在数据预处理方面，我们依次进行了以下的工作：

- 去除噪声数据，获取图片的标注框文件。
- 转换医学 dicom 文件格式为 jpg 格式，进行灰度值变换，使得像素点最终落在 0-255 之间，拥有 RGB 三通道，以满足深度学习迁移学习的基本要求。
- 原图片尺寸过大，影响训练效率，实际过程中将图片 resize 到 512×512 进行训练。
 - 计算训练集数据的均值，标准差等统计信息；对图片进行可视化查看，确定模型中的具体参数设置，如 anchors 的尺寸 scale，anchors 的长宽比 aspect ratio 等。
 - 对图片进行零值归一化，计算图片的各通道的均值及标准差，得到三通道的均值为 $\text{mean}=(0.494, 0.494, 0.494)$ ，标准差为 $\text{std}=(0.101, 0.101, 0.101)$ 并按照 $\text{channel}=(\text{channel}-\text{mean})/\text{std}$ 对各通道中的各个像素值归一化到 $[-1, 1]$ 。

3.2.2 数据集划分

总病例个数 721 例，其中恶性病例 304 例，良性 + 阴性（正常）病例 417 例，总共图片数有 2908 张，其中有 598 张带有恶性标注框的恶性图片。

以病例为单位，对病例数进行 4: 1 划分为训练集合测试集，划分后的数据分布如表3.1所示。最后所要训练的图片个数有 2336 张，恶性图片 1010 张，良性 + 阴性（正常）图片 1326 张；所要测试的图片个数有 572 张，恶性图片有 244 张，良性 + 阴性（正常）图片 328 张。对于恶性病例下的图片，并不都是包含标注框，也就是病人体现为恶性并不都是所有乳房拍的所有角度都体现为恶性，所以训练集恶性病例下图片有标注框的只有 478 张，测试集下则为 120 张。此时再看训练集数目，带有标注的恶性图片数为 478 张，训练集中共有 2336 张图片，使得恶性及良性图片之间存在着大约 1: 3 的不均衡的数据比例，这需要我们为后期的数据处理进行相应的考量。

表 3.1 天肿数据集划分

Table 3.1 tianzhong data set partition

	恶性	正常	恶性	正常	带有标注的恶性
训练集	244	334	1010	1326	478
测试集	60	83	244	328	120

3.3 标准制定

3.3.1 数据定义

我们将恶性乳腺钼靶定义为正样本（positive），其拥有恶性病灶，所有其他的乳房乳腺钼靶定义为负样本（negative），包括阴性和良性乳腺钼靶，其中良性乳腺钼靶中包含高度疑似恶性病灶的信息。在我们使用的数据集中，所有 positive 都有标注框表示其精确位置，所有 negative 没有任何标注信息。

针对乳腺钼靶数据特点以及医院的实际情况，我们设定了两个标准。

3.3.2 基于病例

本研究针对的是乳腺癌中肿块型病变，浸润性导管癌的检测。乳腺钼靶图像尺寸大，平均有 4000×2000 像素，但病灶占比小；医学上图片经常有良性，恶性

和阴性三种类型，其中良性为含有良性肿块的乳腺，恶性为有恶性肿瘤的乳腺，阴性为不存在肿块的乳腺。综合医学和计算机的特点，positive 为 1，negative 为 0，我们的问题变成两分类问题。

根据实际需求标准，医生希望在病例层面对结果进行分析，也就是该病例下存在至少一张图片为恶性即可即为恶性，该病例下所有图片均为良性该病例才是良性。基于此，我们引入了多示例学习 (MIL) [51] 中关于评判标准的定义。多示例学习可以被描述为：假设训练数据集中的每个数据是一个包 (Bag)，每个包都是一个示例 (instance) 的集合，每个包都有一个训练标记，而包中的示例是没有标记的；如果包中至少存在一个正标记的示例，则包被赋予正标记；而对于一个有负标记的包，其中所有的示例均为负标记。（这里说包中的示例没有标记，而后面又说包中至少存在一个正标记的示例时包为正标记包，是相对训练而言的，也就是说训练的时候是没有给示例标记的，只是给了包的标记，但是示例的标记是确实存在的，存在正负示例来判断正负类别）。通过定义可以看出，与监督学习相比，多示例学习数据集中的样本示例的标记是未知的，而监督学习的训练样本集中，每个示例都有一个已知的标记；与非监督学习相比，多示例学习仅仅只有包的标记是已知的，而非监督学习样本所有示例均没有标记。

针对该医学问题，基于病例情况下，假设我们训练的病例集为 R ，每个病例包含多张乳腺钼靶图片及其标签 $\{(X_i, Y_i)\}_{i=1}^N$ ，其中 N 表示病例个数， $X_i = \{x_{ij}\}_{j=1}^{m_i}$ 表示第 $i - th$ 个病例， x_{ij} 表示第 $i - th$ 病例的第 $j - th$ 张乳腺钼靶图片，对应的预测值为 y_{ij} ，最后对病例预测的值 y_j 为：

$$y_{ij} = \max_k (y_{ijk}) \quad (3.1)$$

其中 k 为预测框的个数，至此，所要求的基于病例预测的值为：

$$y_i = \max_j \left(\left(\max_k (y_{ijk}) \right) \right) \quad (3.2)$$

基于该病例预测的基础上，进行对应的准确率 (Accuracy) 和 AUC (Area under curve) [52]，以及医学上常见指标敏感度 (Sensitivity) 和特异度 (Specificity) 的计算。

- Accuracy (简称 ACC)

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.3)$$

- AUC

考虑到医学数据正负样本不均衡问题，单一的准确率指标无法精确衡量效果，所以引入考虑了召回率和精确率的 AUC 曲线。ROC 曲线（接收者操作特征曲线）是一种显示分类模型在所有分类阈值下的效果的图表。该曲线绘制了以下两个参数：真正例率和假正例率。

真正例率 (TPR) 是召回率的同义词，因此定义如下：

$$TPR = \frac{TP}{TP + FN} \quad (3.4)$$

假正例率 (FPR) 的定义如下：

$$FPR = \frac{FP}{FP + TN} \quad (3.5)$$

ROC 曲线用于绘制采用不同分类阈值时的 TPR 与 FPR。降低分类阈值会导致将更多样本归为正类别，从而增加假正例和真正例的个数。AUC 表示的是曲线下面积，也就是说，曲线下面积测量的是从 (0,0) 到 (1,1) 之间整个 ROC 曲线以下的整个二维面积。

AUC 越大，说明分类效果越好。

- Sensitivity (简称 SEN)

$$SEN = \frac{TP}{TP + FN} \quad (3.6)$$

SEN 也称为真阳性率、召回率 (Recall)，指实际为阳性的样本中，判断为阳性的比例（例如真正有生病的人中，被医院判断为有生病者的比例），计算方式是真阳性除以真阳性 + 假阴性（实际为阳性，但判断为阴性）的比值。灵敏度可以作为避免假阴性的量化指标。

- Specificity (简称 SPE)

$$SPE = \frac{TN}{TN + FP} \quad (3.7)$$

SPE，也称为真阴性率，指实际为阴性的样本中，判断为阴性的比例（例如真正未生病的人中，被医院判断为未生病者的比例），计算方式是真阴性除以真阴性 + 假阳性（实际为阴性，但判断为阳性）的比值。特异度可以作为避免假阳性的量化指标。对于任何测试而言，都需要在灵敏度及特异度之间进行取舍。

3.3.3 基于图片

根据计算机视觉检测领域，采用 Mean Average Precisio(mAP) 对预测结果进行考量。mAP 是目标检测算法中衡量算法的精确度的指标，涉及两个概念：查准率 Precision、查全率 Recall。对于目标检测任务，每一个目标都可以计算出其 Precision 和 Recall，多次计算/试验，每个类都可以得到一条 P-R 曲线，曲线下的面积就是 AP 的值，这个 mean 的意思是对每个类的 AP 再求平均，得到的就是 mAP 的值，mAP 的大小一定在 $[0,1]$ 区间。同时根据对单张图片的 ACC 进行查看，基于一定阈值下，只要该张图片预测出恶性的检测框，那么该张图片就被判断为恶性，只有当该张图片没有预测出恶性检测框，那么该张图片才被判断为正常。据此进行相应的标准计算。

3.4 本章小结

本章充分介绍了乳腺钼靶等医疗影像数据特点，并对数据进行相应的数据处理操作以满足深度学习训练的要求，最后从实际需求和传统计算机视觉角度考虑对本文提出了分别基于病例的 AUC，ACC，SEN，SPE 和基于图片的 mAP 的评判标准。

第 4 章 两阶段乳腺钼靶分类检测模型 weakFaster R-CNN

在目标检测领域，主流方法以 R-CNN 系列为代表的两阶段（生成候选框和预测）以及和 YOLO[53],SSD[7] 为代表的单阶段。

相比单阶段，两阶段检测模型通常含有一个串行的头部结构，即完成前背景分类和回归后，把中间结果作为 Fast RCNN 头部的输入再进行一次多分类和位置回归。这种设计带来了一些优点：

- 对检测任务的解构，先进行前背景的分类，再进行物体的分类，这种解构使得监督信息在不同阶段对网络参数的学习进行指导。
- RPN 网络为 Fast RCNN 网络提供良好的先验，并有机会整理样本的比例，减轻 Fast RCNN 网络的学习负担。

虽然两阶段模型存在空间负担大，时间复杂度高等问题，但对于医生而言，更关注的是准确率以及可解释性，所以本文考虑采用更加成熟稳定的 Faster RCNN 模型。Faster RCNN 分为 Region Proposal Network(RPN) 和 Fast R-CNN 网络，具体可以分为将特征抽取 (Feature Extractor), Proposal 提取, ROI Pooling, bounding box regression, classification 几个部分。RPN 主要负责提取候选框，经过 proposal 提取产生了大量的候选框 (Proposal)，选择出包含物体的 anchors，映射到共享的特征图上，之后将这些 proposal 经过 ROI Pooling 生成统一大小的特征图，再进行后续的 Fast R-CNN 更进一步的精调。所以 Faster R-CNN 能否训练成功的关键在于所提取的特征图是否拥有足够的信息，以及 RPN 能够准确选择出候选框，以及最后的 Fast R-CNN 网络训练的是否能够拥有区分不同类别的能力。

传统的自然图像尺寸一般为 500×500 像素，所要识别的目标占据图片主体，且目标之间在大小，形态等各维度均存在较大的差别。在该项目中，由于肿块占比很小，而且 negative 也常常拥有一些与 positive 中恶性肿块相似的区域，给准确检测带来了很大的困难。

4.1 目的

根据乳腺钼靶上的定义，对于肿块这种病灶类型，positive 和 negative 之间的区别只存在于边缘的区别。对于传统的 Faster R-CNN，如果能够拥有正常样

本中疑似恶性图片的病灶标注，那么这个问题并不太难，相当于让模型进行三分类（positive+negative+背景）检测问题，但问题在于本文并没有标注的 negative 数据。考虑到存在没有标注的信息，而且 positive 及 negative 之间的区别非常微小，所以初步设想是希望能够引入弱监督学习的方式，基于原始的 Faster R-CNN 模型，训练学习 positive 中恶性病灶信息，进而得到 negative 中疑似恶性病灶的标注框，而这也是两阶段乳腺钼靶分类检测算法模型 weakFaster R-CNN 提出的初衷，希望使用两次 Faster R-CNN，第一阶段获取 negative 中疑似恶性病灶的标注框，相当于通过模型人为地给 negative 赋予标签，这可以为之后第二阶段的进一步使用 Faster R-CNN 模型检测学习 positive 及 negative 之间的微弱区别提供帮助。

第一阶段本文先用使用 positive 图片数据训练模型，得到一个可以自动检测恶性肿块的网络，当然由于只训练含有恶性肿块的图片，该网络不可避免也会将许多与之相似的良性肿块判断为恶性肿块。

第二阶段，本文用这个模型训练去预测 negative 图片数据，得到相应的疑似肿块区域，由于已知这些区域非恶性肿块，本文将此归为单独的一类，将此类标注框与原有的恶性标注框共同训练，得到一个具有二分类目标检测的网络，最后将第二阶段得到的模型去预测所有的图片，得到最终的结果。

4.2 模型图

weakFaster R-CNN（模型图如图4.1）由于分别使用 Faster R-CNN 进行两次训练，所以存在两个不同的 Faster R-CNN 模型，这两个 Faster R-CNN 的结构和参数量和原始的 Faster R-CNN 相同。所不同的在于对数据的输入以及数据后处理方面。在数据输入方面，第一阶段只有 positive 图片，第二阶段则包含 postive 和被第一阶段模型预测出包含有高度疑似恶性病灶的 negative 图片数据；数据后处理时，在第一阶段应该将训练好的模型去预测 negative 图片，从而能够与 positive 图片进行融合，继续第二阶段的训练。如图4.1所示，第一阶段预测出的 negative 中病灶位置确实疑似恶性。

- 主干网络

深度残差网络（Deep residual network，ResNet[9]）的提出是 CNN 图像史上的一件里程碑事件，在 ILSVRC 取得了五项第一。从经验来看，网络的深度能够

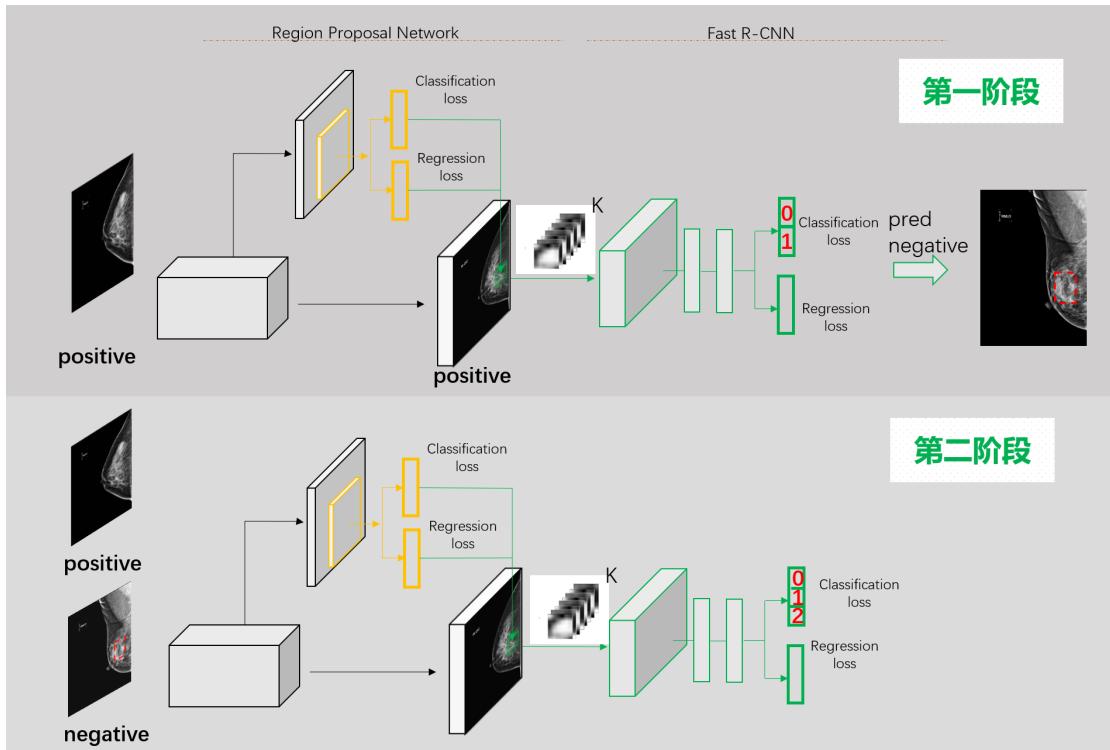


图 4.1 weakFaster R-CNN 模型图

Figure 4.1 Architecture of weakFaster R-CNN

有效提取出模型的高层信息，从而增强模型的学习能力。然而，随着模型层数的加深，以往的模型存在网络深度增加时，网络准确度出现饱和，甚至出现下降的问题。Resnet 通过残差学习（如图4.2）来解决退化问题，有效解决了深度所造成的模型无法收敛的问题，这称得上是深度网络的一个历史大突破。本文采取 Resnet101 作为 Faster R-CNN 的主干网络，借此强大的模型可以充分提取出分类和检测所需要的特征表达，从而提升模型的整体性能。其网络配置如表4.1。

在本文中，无论是两阶段还是一阶段模型，本文均采用该模型作为检测模型的主干网络。具体而言，特征提取模块取 conv1, conv2, conv3 和 conv4，之后再使用 Fast R-CNN 模块取 conv5 作为高层语义的提取器。

- ROI Align

ROI Align 是在 Mask R-CNN[54] 这篇论文里提出的一种区域特征聚集方式，很好地解决了 ROI Pooling 操作中两次量化造成的区域不匹配的问题。原始的 ROI Pooling 过程需要首先根据后面网络要求的输入尺寸求网格大小，如输入特征图尺寸为 5×5 ，而网络输入尺寸为 2×2 ，那么网格按照 $5/2=2$ 个像素点对网格进行切分；其次对于每个小网格，求最大值作为该网格最大池化之后的值。ROI

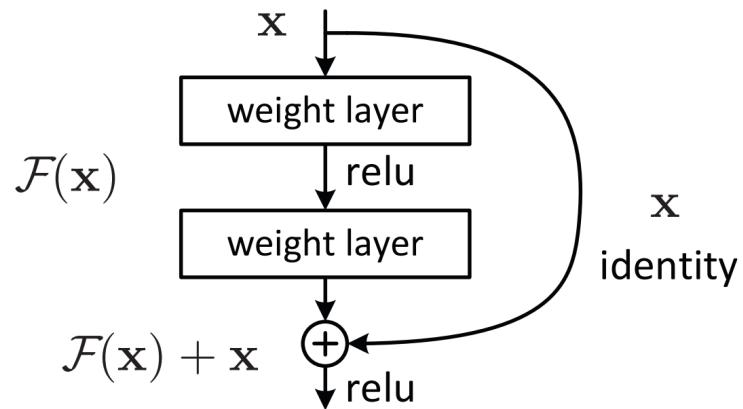


图 4.2 残差模块 [9]

Figure 4.2 residual block[9]

表 4.1 Resnet101 参数配置 [9]

Table 4.1 The config of Resnet101[9]

layer name	output size	101-layer
conv1	112×112	$7 \times 7, 64, \text{stride } 2$
conv2_x	56×56	$3 \times 3 \text{ max pool, stride } 2$
conv2_x	56×56	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
conv4_x	14×14	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$
conv5_x	7×7	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax

Pooling 过程会有两次量化过程，对于一个 Region Proposal，首先从原图经过全卷积网络到特征图，得到的候选框位置可能存在浮点数，进行取整操作从而出现第一次量化；其次，在 ROI Pooling 求取每个小网格的位置时也同样存在浮点数取整的情况。这两次量化的结果都使得候选框的位置会出现偏差。

为了解决上述缺点，本文使用了 ROI Align 的方法。思路很简单：取消量化操作，使用双线性内插的方法获得坐标为浮点数的像素点上的图像数值，从而将整个特征聚集过程转化为一个连续的操作。在检测任务中将 ROI Pooling 替换为 ROI Align 可以提升检测模型的准确性。

4.3 训练过程与细节

4.3.1 训练步骤

- 按照第二张数据预处理和数据划分的方式对数据进行处理；考虑到训练集中恶性和正常样本比达到 1: 3，非常悬殊。实际训练过程中对恶性样本进行上下左右四种方式进行数据扩增，对良性数据只进行左右翻转方式。
- 按照第一阶段思路，只取出训练集中 positive 图片及其标注文件，并按照 Faster R-CNN 输入要求，将训练集数据进行相应操作（包括基本的数据预处理操作）；进行模型调参工作，查看损失曲线下降情况，得到第一阶段训练后的模型。
- 将上述得到的模型对训练集中的所有 negative 数据进行预测，得到良性标注框，取出这部分标注框与原有的医生标注的恶性标注框进行合并，得到第二阶段需要的训练数据。
- 将上述得到的数据给 Faster R-CNN 模型再次训练，再次通过具体的调参，查看损失曲线等方式，最终得到第二阶段的模型。至此，在训练集上已经对所有数据都进行训练过了。
- 将上述得到的模型对所有的测试集图片数据进行训练，之后按照一定的标准得到对应的病例的评判结果。

本文借鉴了迁移学习的思路，对于低层语义采用较小的学习率进行调整，对于高层语义，也就是 conv5 block，本文冻结了除最后两个卷积层外的参数。

4.3.2 第一阶段

第一阶段只考虑训练集中 478 张恶性图片，经过上述数据预处理和扩增方式后，恶性样本总数达到 1912 (478×4)。主要参数设置如表4.2所示。取损失函

表 4.2 weakFaster R-CNN 第一阶段训练参数配置**Table 4.2 The first stage training parameter configuration of weakFaster R-CNN**

参数名称	参数值
Batch_size	2
Learning rate	1e_4
Decay_steps	60
Decay_rate	0.1
Optimizer	adam

数下降足够平稳的 140000iters 对训练集良阴性（正常样本）进行预测，主要参数设置如表4.3所示。

表 4.3 weakFaster R-CNN 第一阶段预测参数配置**Table 4.3 The first stage prediction parameter configuration of weakFaster R-CNN**

参数名称	参数值
CONF_THRESH	0.8
NMS_THRESH	0.3

结果对训练集中的 1858 (2336-478) 张正常样本，预测出 365 张有疑似恶性病灶的图片。对其抽样进行可视化结果如图4.3。其中黄框为 ground truth，粉红框为预测框。可以看出，疑似恶性的图片区分得不错。测试集带有恶性标注的恶性图片共 120 张，预测出 102 张，对其进行 IoU 查看，发现预测出来的 pred_bbox 与真实的 gt 的 IoU 均会大于设置的 0.5 的阈值，也就是说，120 张图片有 18 张没被正确预测，其他预测正确的均预测正确。

4.3.3 第二阶段

开始进行第二阶段的训练。经过第一阶段，训练集中产生了 365 张拥有疑似恶性病灶的正常图片样本，合并原有的训练集恶性图片样本 468 张，是本文第二阶段重点学习的数据。

第二阶段，取原始的恶性病灶图片及第一阶段预测的疑似恶性病灶的正常图片，总共 843 张 (468+365) 图片进行训练，经过特定的预处理和数据扩增操作后，开始进行训练。训练过程中，具体参数设置如表4.4所示。

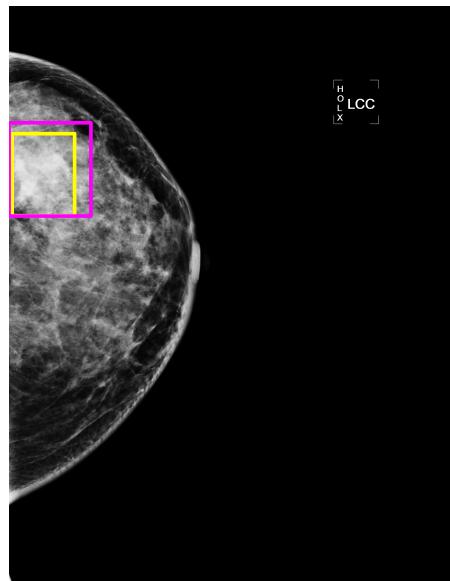


图 4.3 weakFaster R-CNN 第一阶段对图片预测结果可视化

Figure 4.3 Visualization of image prediction results in the first stage of a weakFaster R-CNN

表 4.4 weakFaster R-CNN 第二阶段训练参数配置

Table 4.4 The second stage training parameter configuration of weakFaster R-CNN

参数名称	参数值
Batch_size	2
Learning rate	1e_4
Decay_steps	60
Decay_rate	0.1
Optimizer	adam

使用 260000iters 对训练集进行预测。训练时的主要参数设置如表4.5所示。

表 4.5 weakFaster R-CNN 第二阶段预测参数配置

Table 4.5 The second stage prediction parameter configuration of weakFaster R-CNN

参数名称	参数值
CONF_THRESH	0.7
NMS_THRESH	0.4
RPN_PRE_NMS_TOP_N	1000
RPN_POST_NMS_TOP_N	100
RPN_NMS_THRESH	0.8

4.4 结果讨论与分析

训练集共有 578 个病例，其中恶性病例 244 个，预测结果如表4.6所示。共预

表 4.6 weakFaster R-CNN 第二阶段训练集预测结果

Table 4.6 The prediction result for training set in the second stage of weakFaster R-CNN

	真实为恶性 (244 例)	真实为良阴性 (334 例)
预测为恶性	215	23
预测为良阴性	29	311

测 578 例病人，其中恶性 244 例，正常 334 例，真实为恶性且判断为恶性 (TP) 为 215 例，真实为恶性且判断为阴良性 (FP, 假阳性数目) 为 23 例，真实为良阴性且判断为良阴性 (TN) 为 311 例，真实为良阴性且判断为恶性的 (FN) 为 29 例。

经过在训练集上的查看，可知 weakFaster R-CNN 训练足够充分，开始考虑模型在测试集上的性能。

4.4.1 基于病例

测试集共有 143 例病人，其中恶性病例 60 例，正常病例 83 例。运行结果显示如表4.7所示，从结果上可以看出，weakFaster R-CNN 能够在一定程度上有效地区分恶性样本和正常样本，使得预测结果中的召回率 (SEN) 和特异度 (SPE) 都较高，能够基本满足医院的实际需求。

表 4.7 weakFaster R-CNN 基于病例预测结果**Table 4.7 The prediction results of weakFaster R-CNN based on patient**

	AUC	ACC	SEN	SPE
weakFaster R-CNN	0.8467	0.9123	0.8846	0.9123

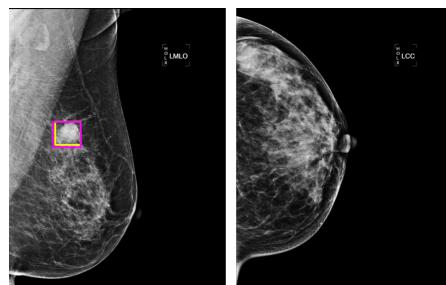
4.4.2 基于图片

运行结果显示如图4.8所示，乳腺钼靶的 mAP 达到 0.40，并经过结果可视化

表 4.8 weakFaster R-CNN 基于图片预测结果**Table 4.8 The prediction results of weakFaster R-CNN based on image**

mAP
weakFaster R-CNN 0.4055

(如图4.4所示) 其中黄框为 ground truth，粉红框为预测框，左边是正样本预测结果，右边是负样本预测结果)，可知模型对于正负样本的区分能力较强。

**图 4.4 weakFaster R-CNN 第二阶段对图片预测结果可视化****Figure 4.4 Visualization of image prediction results in the second stage of a weakFaster R-CNN**

4.5 本章小结

本章对乳腺钼靶数据进行两阶段的考量和设计，第一阶段重点学习恶性样本中恶性病灶的信息，第二阶段着重学习区分正常样本中疑似恶性病灶和恶性病灶的区别，得到了基于 weakFaster R-CNN 训练下的乳腺钼靶数据预测结果。

第5章 一阶段乳腺钼靶分类检测模型 simFaster R-CNN

5.1 目的

考虑到两阶段模型存在训练耗时长、空间占用大、参数量众多难以调节、训练测试反复进行耗时长等问题。通过对乳腺钼靶数据的了解和检测模型的深入理解，本文着手设计更加轻便的一阶段模型，使得模型可以在检测并分类出恶性病灶的同时，能够学习到恶性和正常病灶之间的细微区别。对于医学上的医疗影像问题，这种同时存在恶性及正常样本的问题很多，具体在产业化、标准化中，也需要模型能够更加轻便高效。

两阶段模型的思路清晰，第一阶段重点学习恶性病灶特征，使得 positive 能够及时被检测分类，人为标注出 negative 中疑似恶性病灶区域，将带有疑似恶性病灶的 negative 与 positive 融合后再次训练，第二阶段着重于学习恶性病灶和疑似病灶之间的微弱区别，这种分阶段式的训练方式符合一般逻辑。

通过对损失函数的观察，发现损失函数占比较高的是 Fast R-CNN 的损失，而 RPN 网络的损失并不太高，这说明 RPN 网络能够较为充分地学习到物体（在本文中指的是病灶）的特征。原始的 Faster R-CNN 的 RPN 着重考虑能否区分出物体，其对物体的 ground-truth 在 RPN 训练之前就已给出，而完全忽略了 RPN 会产生大量的高度疑似物体的负 anchors，在乳腺钼靶问题中这个问题尤为突出。由于存在大量高度疑似恶性病灶的 negative（如图3.6所示）在 RPN 网络中没有及时筛选出，使得 Fast R-CNN 的输入数据存在大量带有干扰性的负 anchor 数据，进而影响模型的进一步区分这种差异的能力。

为此，综合损失函数和数据本身的特点，在进行一阶段模型设计时候，需要重点考虑模型的能够识别不同类之间的能力。在此基础上本文引入全新损失函数，对原始的 Faster R-CNN 进行改进，提出 simFaster R-CNN，使得模型能够同时学习恶性病灶特征以及恶性和高度恶性病灶之间的区别，从而加快模型训练与学习。

5.2 模型图

simFaster R-CNN 的网络结构如图5.1所示。

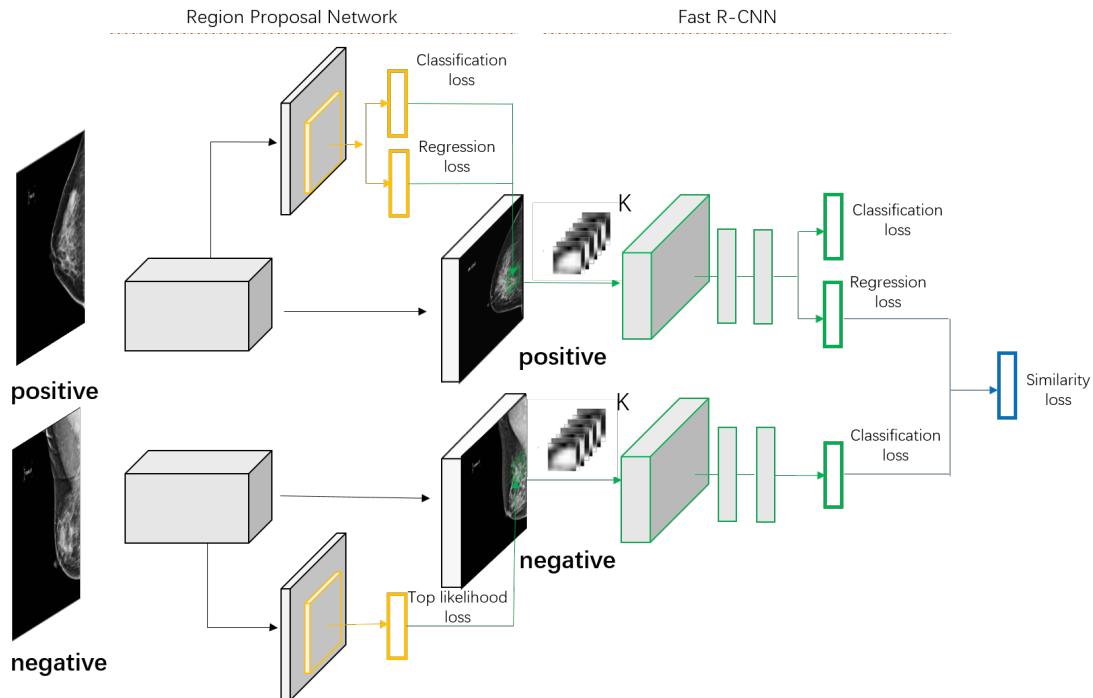


图 5.1 simFaster R-CNN 模型图

Figure 5.1 Architecture of simFaster R-CNN

本模型是基于 Faster R-CNN 模型演变而来。主干部分不变，但输入变成相同数量的恶性和正常样本，二者经过共享的主干网络进行特征提取，获取到对应的特征，共享 RPN 网络。对恶性样本按照原始的 Faster R-CNN 的损失函数，对正常样本则是引入 top likelihood loss。之后将二者生成的候选框经过同样的 ROI Align 操作，共享 Fast R-CNN 的网络，输出时除了进行和 RPN 网络一样的恶性和正常样本的损失计算，在这里，本文同时获取了 K 个（ K 决定于恶性样本生成的候选框中个数）包含高维信息的向量进行相似度损失计算，最后完成整个模型的设计。引出的损失函数目的是为了能够充分降低 negative 被预测为 positive 的概率，同时让模型尽可能地区分开 negative 和 positive 这两个不同类别。为此，本文提出了以下几个 loss。

5.3 Top likelihood loss

在训练 negative 图像时，原始的 RPN 损失将增加 top likelihood loss，使得模型可以降低 negative 预测为 positive 的几率，使得 negative 得以区分开。

在 Faster R-CNN 中，RPN 网络识别的是所要识别的 anchor 是否包含物体。正面标签分配给两种 anchor：

- 与 ground-truth box 具有最高交叉点 (IOU) 重叠的 anchors;
- 与 ground-truth box 的 IOU 大于 0.7。

通常第二个条件足以阻止挖掘 positive，但第一个条件仍然采用是因为在极少数情况下，第二种情况可能没有找到 positive。其中 negative 与 ground-truth box 的 IoU 比率低于 0.3，既不是 positive 也不是 negative 的 anchors 不参与训练。RPN 由以图像为中心的训练抽样策略，每个 RPN 训练的数据量来自单一的图像中的 positive 和 negative 的 anchors，其目标在于优化所有 anchors 的损失函数，每个小批量随机抽样 256 个 anchors 计算损失函数的图像，其中采样的正负 anchors 的比例高达 1: 1。如果其中有少于 128 个阳性样本一个图像，那么负 anchors 将会被补充。

对原始的 Faster R-CNN 的 RPN 研究，可以知道 RPN 的损失只考虑通过上述条件选择出来的 anchors 的 ground-truth 与预测出来的损失，而在真实情况中，存在其他并没有通过上述条件选择出来的，但被 RPN 预测得到的高度疑似恶性的 anchors，表现在分类情况中，就是所判别为物体的概率会特别高。原始的 RPN 训练时会产生大量的负 anchors，如果不考虑这部分 anchors，会使得模型偏向负 anchors。而且当 RPN 考虑引入正常样本情况时，由于负样本不存在标注框，更多的高度疑似恶性病灶的负 anchors 将被产生。如果 RPN 不对这部分负 anchors 进行考虑，将使得模型失去对负 anchors 的学习能力。所以在具体训练时，需要设计损失函数，让模型尽可能降低这部分高度疑似恶性病灶可疑区域的负 anchors，使得模型对于负 anchors 的判断尽可能降低，从而使得 RPN 网络可以偏向正 anchors，提高正 anchors 的分类准确率，从而提高后续给 Fast R-CNN 输入的候选框的准确率。

本文对所有负样本在经过 RPN 训练后得到的负 anchors 按照预测分数进行排名，并取前 256 个负 anchors 进行采样以计算损失函数。这些具有很高预测值的负 anchors，也就是模型会高度判断其为疑似正 anchors 的样本。只要尽可能降低这些具有代表性的易被分错的负 anchors，那么所有正 anchors 被优化为正 anchors 的可能性就会提升。

将负 anchors 的 RPN 损失函数定义为最大似然损失：

$$L_{tloss} = \frac{1}{cls} \sum_{i \in (tops p_i)} L_{cls}(p_i, p_i^* = 0) \quad (5.1)$$

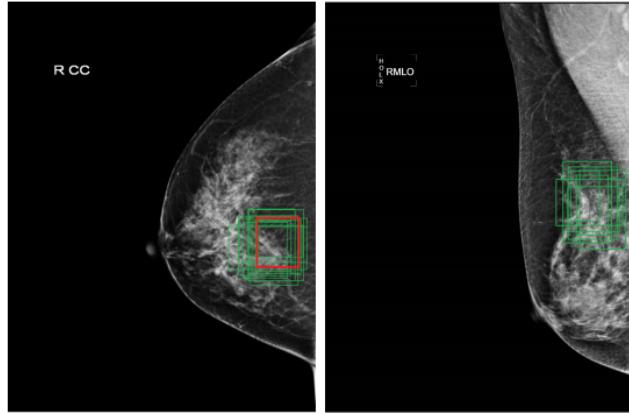


图 5.2 经过 RPN 训练后得到的恶性样本和正常样本预测得到的 anchors

Figure 5.2 The anchors that were got by RPN after training postive and negative

其中， i 是 mini-batch 中 anchors 的索引值， p_i 表示的是每个 anchor 预测为物体的概率值，top likelihood loss 只考虑前 256 个。由于所考虑的是负 anchors，所以 p_i^* 的分类标签总是为 0。分类损失 L_{cls} 和 FasterR-CNN 中的 RPN 网络一样，采用的是交叉熵损失函数。

总而言之，改进后的 RPN 损失函数定义如下：

$$L_{rpnloss} = L_{ploss} + L_{nloss} = L_{pclssloss} + \lambda_1 L_{pregloss} + \lambda_2 L_{tlloss} \quad (5.2)$$

L_{ploss} 是针对正样本所考虑的损失函数，和 Faster R-CNN 中一样，包含分类损失 $L_{pclssloss}$ 和回归损失 $L_{pregloss}$ 。而对于考虑负样本的 L_{nloss} ，只考虑 top likelihood loss L_{tlloss} 。 λ_1, λ_2 是调优参数，在本文中设置为 1。

5.4 Similarity loss

样本经过 RPN 训练后和 NMS[47] 等操作后得到候选框，通过查看 RPN 得到的候选框（如图5.2所示），可以看出正负 anchors 所包含的病灶特征高度类似，需要在 Fast R-CNN 模块尽可能区分开二者。在这部分，本文考虑了人脸识别中常用的孪生网络，也就是学习共享权重网络后的不同类别间的特征向量间的距离。这里引入了 similarity loss，并因此改进 Faster R-CNN 为 simFaster R-CNN。

$$L_{simloss}(X_1, X_2) = \frac{1}{K} \sum_{i=1}^K \|f(X_1^i) - f(X_2^i)\| \quad (5.3)$$

其中 X_1 和 X_2 是从模型 Fast R-CNN 学习到的特征向量选取而来。负候选框和正候选框共享同一个 Fast R-CNN 网络权重， f 是一个距离度量函数，在这里

本文主要选择采用 cosine 函数，降低这个损失函数为了提高识别恶性病灶和疑似恶性病灶的判别能力。但是负候选框是没有病灶信息的。对于所识别出来的疑似恶性病灶的候选框，其分类标签都是为 0 的，没有回归损失。总的 Fast R-CNN 损失函数定义如下：

$$L_{fastloss} = L_{clsloss} + \lambda_3 L_{regloss} + \lambda_4 L_{simloss}, \quad (5.4)$$

式子中包括三项：分类损失 $L_{clsloss}$ ，回归损失 $L_{regloss}$ 和 similarity loss $L_{simloss}$ 。 λ_3, λ_4 是调优参数，在本文中设置为 1 和 0.1。

5.5 训练过程与细节

首先对图片进行如第二章的数据预处理操作，之后使用了水平和竖直两种数据扩增方式来降低过拟合。在训练时，本文使用的是基于 Resnet101 为主干网络的 Faster R-CNN，ROI Align 的 ROI Pooling 方式。每次输入的图片是一对数据，分别是从扩增后的恶性数据集和正常数据集分别选择 batch_size 的图片，在每轮取完时，再对相应的数据集进行 shuffle，以保证每次取出的数据不一样。本文使用了初始学习率为 $1e_4$ ，每经过 9k 迭代次数下降 0.1，总共跑了 45k 迭代次数。调优函数使用的是 Adam，对于每张图片产生的 mini-batch 数值设置为 128。最终测试时设定的得分置信度为 0.5，IoU 的置信度选择的是 Pascal VOV 中的标准值 0.5。最终本文使用了 Nvidia GTX 1080Ti 两张卡跑了两天左右。

5.6 结果讨论与分析

5.6.1 基于病例

对运行各个结果显示如表5.1所示。

表 5.1 simFaster R-CNN 基于病例预测结果

Table 5.1 The prediction results of simFaster R-CNN based on patient

	AUC	ACC	SEN	SPE
Faster R-CNN	0.8467	0.7622	0.9667	0.6145
Faster R-CNN+tlloss	0.8696	0.8531	0.8333	0.8675
simFaster R-CNN	0.9110	0.8881	0.9167	0.8675

从上到下的模型依次是使用原始的 Faster R-CNN 模型直接对正负样本进行训练, 对负样本添加 top likelihood loss 的 Faster R-CNN 模型 (Faster R-CNN+tlloss), 对负样本同时添加 top likelihood loss 和 similarity loss 的 simFaster R-CNN 模型 (Faster R-CNN+tlloss+similoss)。

模型最终预测得到的 ROC 曲线如图5.3。从结果上可以看出, 原始的 Faster

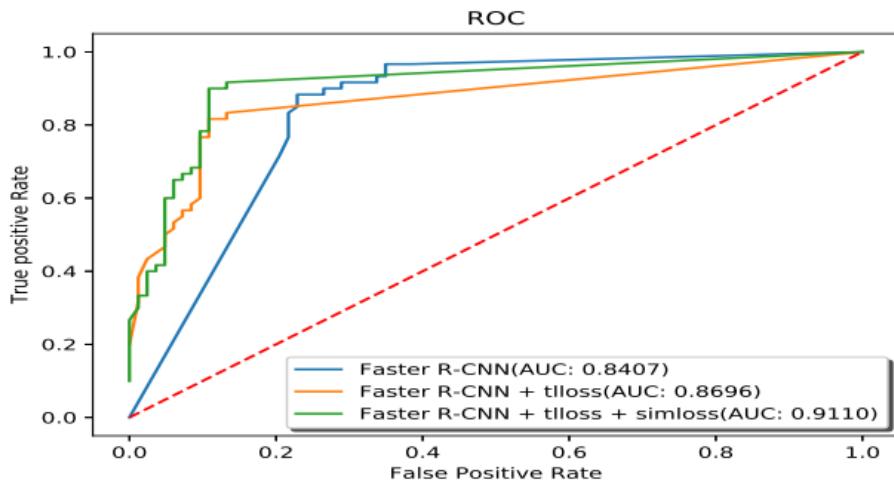


图 5.3 simFaster R-CNN ROC 曲线

Figure 5.3 The ROC curve of simFaster R-CNN

R-CNN 模型由于缺乏对负样本的有效学习, 使得在最终评判标准上存在敏感度上高, 但特异度上很低的情况。在上述原始模型上添加了 top likelihood loss 之后, 模型虽然对于正样本的敏感度下降了, 但是对于负样本的特异度提升, 这表明该 loss 能够有效帮助模型更好地学习负样本的信息。最后, 添加了专门学习正样本和负样本差异性的 similarity loss 后, 模型在保持特异度不变的情况下, 有效提高了敏感度。本文设计的 top likelihood loss 和 similarity loss 在各个指标都取得了远远超过原始的 FasterR-CNN 的结果, 最终 AUC 可以达到 0.91, 甚至超过了临床诊断上医生的识别率 0.72。这结果充分表明了 simFaster R-CNN 通过添加 top likelihood loss 来降低 RPN 中被判错的负 anchors, 以及通过 similarity loss 来学习不同类别的微小差异的有效性, 进而能够在满足精度的情况下, 提升了模型的训练速度。

表 5.2 simFaster R-CNN 基于图片预测结果**Table 5.2 The prediction results of simFaster R-CNN based on image**

mAP	
Faster R-CNN	0.4055
Faster R-CNN+tlloss	0.4219
simFaster R-CNN	0.5353

5.6.2 基于图片

本文还对比了传统的 mAP 指标，该指标综合考虑了单张图片中各个预测框的分类和检测性能。从表5.2中可以看出，该结果也是随着模型的不断完善和优化而不断提升的，并最终在本文提出的 simFaster R-CNN 达到最优的结果。

综合对比 weakFaster R-CNN 的结果（如表5.3），可知 simFaster R-CNN 在 AUC, SEN, mAP 等指标超过了 weakFaster R-CNN，体现出其优越的性能。weakFaster R-CNN 由于第二阶段能够使用初步筛选的疑似恶性乳腺钼靶数据作为训练数据，所以其在良性样本的区分度能力更强，也就是 SPE 指标上会比 simFaster R-CNN 数值高一点。但整体而言，simFaster R-CNN 耗时少，不占用大量空间，取得了不输于 weakFaster R-CNN 的性能。

表 5.3 weakFaster R-CNN 预测结果**Table 5.3 The prediction results of weakFaster R-CNN**

	AUC	ACC	SEN	SPE	mAP
weakFaster R-CNN	0.8467	0.9123	0.8846	0.9123	0.4055
simFaster R-CNN	0.9110	0.8881	0.9167	0.8675	0.5353

5.6.3 可视化预测结果

模型对于正样本预测结果如图5.4，从左到右分别是 Faster R-CNN,Faster R-CNN+tlloss, simFaster R-CNN 对正样本的预测结果，其中黄框为真实框，粉红框为预测框。可以看出，随着模型的不断优化，模型对于 postive 的分类以及回归的精度都得到有效的提升。添加的 top likelihood loss 可以有效降低负 anchors 的召回率，降低 anchors 的假阳性率，从而提高模型的性能。添加 similiarity loss 由于考虑了正负样本的微弱差异，在定位精度上会比只加 top likelihood loss 更高。

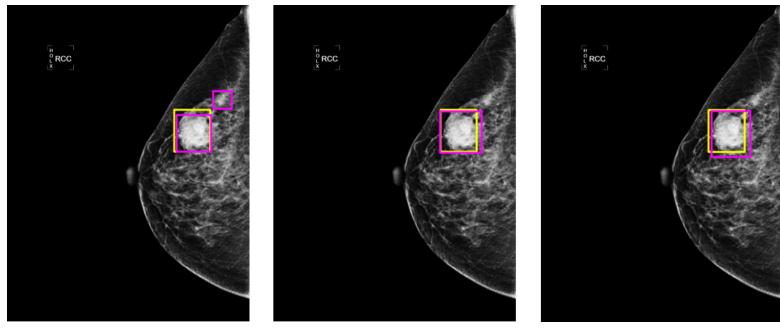


图 5.4 simFaster R-CNN 在天肿数据集上的对正样本预测结果可视化

Figure 5.4 The detection results of positive in vision using simFaster R-CNN at the tianzhong datasets

对于负样本预测结果如图5.5,从左到右分别是 Faster R-CNN,Faster R-CNN+tlloss, simFaster R-CNN 对负样本的预测结果, 其中粉红框为预测框。从对于负样本上的结果上看, 可以清晰地看出模型对于负样本的识别能力随着 loss 的添加而不断地提升。原始的 Faster R-CNN 存在很明显的高假阳性率, 而且其得到的高度疑似恶性病灶的预测框定位并不够精准。添加了 top likelihood loss 会在一定程度上提高模型的定位精度。当使用了 similarity loss 后的 simFaster R-CNN 则同时在分类和定位上远远超过最原始的 Faster R-CNN, 从而极大地降低了模型的假阳性率。

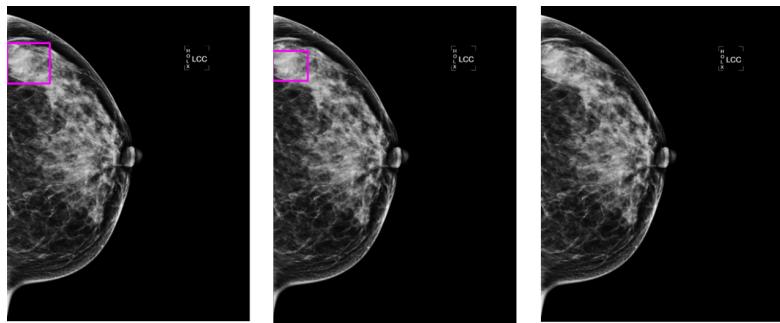


图 5.5 simFaster R-CNN 在天肿数据集上的对负样本预测结果可视化

Figure 5.5 The detection results of negative in vision using simFaster R-CNN at the tianzhong datasets

总而言之, 随着 loss 的不断添加, 模型对于恶性的分类及定位, 对于正常样本的分类都有了更加精准的提升。这也很好地佐证了上述模型的指标大幅提升。

5.7 其他工作

本文还进行了诸多的尝试，比如修改模型的主干网络，用其他的 similarity loss，修改调优函数，主干网络上添加模块等操作，其他公开数据集上的比较。

5.7.1 更换主干网络

本文还对比分析了最后一届 ILSVR 的冠军模型 Senet[10]，top5 的错误率达到了 2.251%，比 2016 年的第一名还要低 25%，其主要考虑了各个特征层之间的关系，通过学习各个特征层不同的权重并应用注意机制，使得关键特征层可以凸显发挥更大的作用。其重要的 SE Block 如图5.6所示。

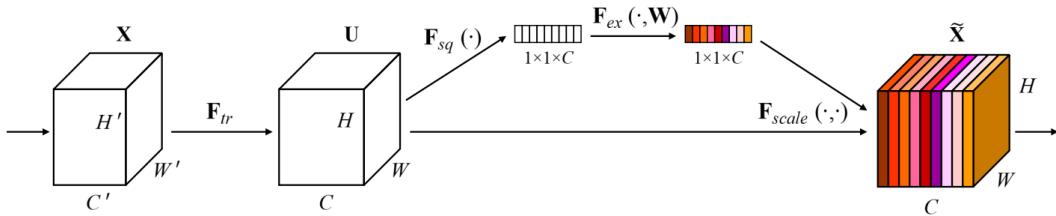


图 5.6 SE Block[10]

Figure 5.6 SE Block[10]

同上，本文取 Seresnet101 作为模型的主干网络，从结果（如表5.4）可以看出，Seresnet101 作为主干网络对于最终结果，并没有达到或超过 Resnet101 的效果。分析原因可能是，传统的自然图像是三通道的，其各个通道之间存在关联信息，使得 SE Block 可以在其中发挥作用，而对于乳腺钼靶数据，其原始图片为单通道灰度图信息，单纯地通过通道叠加变成 RGB 三通道，Seresnet101 并没有学习通道之间的关系，进而不能够对不同通道施加不同权重。Seresnet101 引入额外计算，使得模型在时间复杂度上会比 Resnet101 更高。

表 5.4 simFaster R-CNN 使用 Seresnet101 模型预测结果

Table 5.4 The prediction results of simFaster R-CNN using Seresnet101 as backbone

AUC	ACC	SEN	SPE	mAP
0.8505	0.7972	0.8785	0.7762	0.3865

5.7.2 更换 similarity loss

孪生网络训练中所采用的 loss 是 contrastive loss。这种损失函数可以有效的处理孪生神经网络中的数据对的关系，传统的 Siamese network 使用 contrastive loss，其函数如下：

$$\text{con_loss} = (1 - Y) \frac{1}{2} (D_w)^2 + (Y) \frac{1}{2} \{\max(0, m - D_w)\}^2 \quad (5.5)$$

- D_w 被定义为姐妹孪生网络的输出之间的欧氏距离。 D_w 欧式距离公式如下：

$$\cos_loss = \sqrt{\{G_w(X_1) - G_w(X_2)\}^2} \quad (5.6)$$

- G_w 是其中一个姐妹网络的输出。 X_1 和 X_2 是输入数据对。
- Y 值为 1 或 0。如果模型预测输入是相似的，那么 Y 的值为 0，否则 Y 为 1。
- \max 是表示 0 和 $m - D_w$ 之间较大值的函数。
- m 是大于 0 的边际价值 (margin value)。有一个边际价值表示超出该边际价值的不同对不会造成损失。

相比于 cosine loss，该 loss 考虑了多种数据对情况，包括同正，同负，一正一负等。从结果（如表5.5）可以看出该 loss 对比于 cosine loss，除了 SPE 指标外均取得了超过 cosine loss 的结果。对于本文而言，最终所要考虑的仅是一正一负这种情况，这使得该 loss 对于不同类别之间的区别学习比较充分，而对于同类别，尤其是同负类的相似度学习并不充分，所以更多地推远了恶性样本和正常样本（一正一负）的距离，而并没有有效的拉近恶性样本（同正）或正常样本（同负）之间的距离，在结果上表现为 AUC 和 mAP 会比 cosine 来的大，但对于 SEN 和 SPE 并没有太大的提升。

表 5.5 simFaster R-CNN 使用 contrastive loss 预测结果

Table 5.5 The prediction results of simFaster R-CNN using contrastive loss as similarity loss

AUC	ACC	SEN	SPE	mAP
0.9262	0.8042	0.9500	0.6988	0.5525

5.7.3 更换权重因子

$\lambda_1, \lambda_2, \lambda_3, \lambda_4$ 为调节各个 loss 权重的因子，通过调节不同情况下各个值，可以达到突出或者降低某个 loss 影响的作用。在本文中，默认使用的其值分别是

$\lambda_1=1, \lambda_2=1, \lambda_3=1, \lambda_4=1$ 。为了更好地挖掘这些因子的作用，本文还进行了以下的对比实验。如表5.6所示。

表 5.6 simFaster R-CNN 使用不同权重因子预测结果

Table 5.6 The prediction results of simFaster R-CNN using different lambda

	AUC	ACC	SEN	SPE	mAP
$\lambda_1=1, \lambda_2=0.9, \lambda_3=1, \lambda_4=0.1$	0.9036	0.8811	0.8833	0.8795	0.5266
$\lambda_1=1, \lambda_2=0.1, \lambda_3=1, \lambda_4=0.9$	0.8641	0.8531	0.8333	0.8675	0.5094

可以看出，不同的 $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ 值对于结果影响较大，在本文数据中最好的参数依然是 $\lambda_1=1, \lambda_2=1, \lambda_3=1, \lambda_4=1$ 。而对于后期的 DDSM 数据，最好的权重因子参数分别是 $\lambda_1=1, \lambda_2=0.9, \lambda_3=1, \lambda_4=0.1$ 。不同的参数会对模型最终训练效果产生影响，这需要在实验中进行具体的尝试。

5.7.4 添加模块

对数据进行纵横比查看，如图5.7分析，发现其纵横比落在 1:2, 1:1 和 2:1 三个区间内，且由于乳腺钼靶本身图片像素大，造成的缩放比影响较大。打算使用 Deformable Convolutional Network(DCN)[55]，通过卷积核的学习来对这种影响进行修正。DCN 通过考虑对卷积核中每个采样点的位置添加了一个偏移的变量，通过这些变量，卷积核可以在当前位置附近随意的采样，而不再局限于之前的规则格点。

通过使用 DCN 模块，从最终结果（如表5.7）可以看出，DCN 在结果上并没有太大优势。DCN 的提出是由于传统自然图像上存在超过正常三种纵横比之外的其他纵横比，而对于乳腺钼靶数据而言，纵横比的差异不够突出，DCN 在这方面并没有发挥出他的作用。

表 5.7 simFaster R-CNN 添加 DCN 模块预测结果

Table 5.7 The prediction results of simFaster R-CNN adding DCN

AUC	ACC	SEN	SPE	mAP
0.8149	0.6923	0.8833	0.5542	0.4893

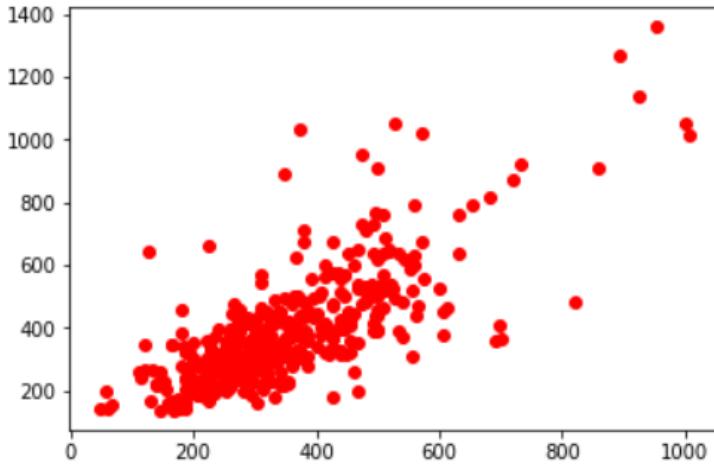


图 5.7 乳腺钼靶纵横比

Figure 5.7 Aspect ratio of mammography

5.7.5 更换数据集

针对这种方法，我还扩展对比了在 DDSM 上的表现。DDSM 数据库是由南佛罗里达大学创建并一直如此广泛用于乳房研究目的，它包含 2620 案件分为 43 卷，每张乳腺钼靶照片包含与之相关的可疑病变的信息。对于恶性图片如图5.8所示。

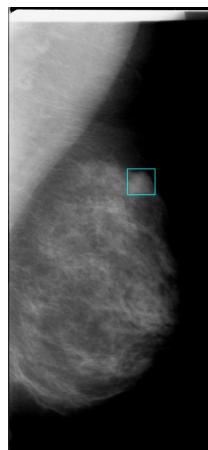


图 5.8 DDSM 乳腺钼靶图

Figure 5.8 DDSM Mammography

相对于本文的东方女性致密性乳腺，DDSM 的乳腺钼靶更多是西方女性的少量腺体型乳腺。使用不同数据集是为了判断模型 simFaster R-CNN 是否具有迁移不同知识的能力。

按照 AL-MASNI M A 等 [42] 中对 DDSM 进行去除存在问题的图片和采集数据，本文最后取出 DDSM 中的包含肿块的部分数据集，训练集中包括 600 张图片，其中 304 张是带有病灶坐标标记信息的恶性样本，296 张不带有标记信息，也就是正常样本；测试集中包括 100 张图片，其中 45 张带有标注信息的恶性样本，55 张为不带有标记信息的正常样本，数据划分如表5.8所示。

表 5.8 DDSM 数据集划分**Table 5.8 Data set partition of DDSM**

数据集	带病灶恶性图片数	正常样本数
训练集	304	296
测试集	45	55

按照和本文数据一样的数据方式，最终对结果进行预测如表5.9所示。

表 5.9 simFaster R-CNN 使用 DDSM 数据集基于图片预测结果**Table 5.9 The prediction results of simFaster R-CNN using DDSM as dataset based on image**

	AUC	ACC	SEN	SPE
Faster R-CNN(no annotation on benign)	0.8945	0.8100	0.8667	0.7636
simFaster R-CNN	0.9446	0.9300	0.9111	0.9455
Faster R-CNN(with annotated benign)	0.9269	0.9100	0.8667	0.9455
AL-MASNI M A 等 [42]	0.8774	0.8552	0.9320	0.7800

考虑到与 AL-MASNI M A 等 [42] 的预测结果进行对比，所以这里计算的所有指标都是基于图片上进行考量。可以看出一阶段模型 simFaster R-CNN 在各个指标上远远超过原始的 Faster R-CNN(no annotation on benign)。在 AUC、ACC、SPE 均超过 AL-MASNI M A 等 [42] 的，SEN 较低原因在于本模型需要考虑到负样本情况，而 AL-MASNI M A 等 [42] 是在基于标注信息训练的最终结果。同时，本文还在弱监督上训练的模型与带有全部标注信息的 Faster R-CNN(with annotated benign) 进行对比，发现结果并不差，这很好的证明了 simFaster R-CNN 想法的有效性。

对 DDSM 的正样本（如图5.9）及负样本（如图5.10）预测结果进行可视化（取置信度为 0.5，从左到右分别是 Faster R-CNN,Faster R-CNN+tlloss, simFaster

R-CNN) 对 DDSM 正样本的预测结果, 其中黄框为真实框, 粉红框为预测框。从结果中可以看出, 随着 loss 的不断增加, 模型对于病灶的预测假阳性率不断下降, 而且定位精度不断提升, 从而也有效证明了 simFaster R-CNN 具备数据迁移学习的强大能力。

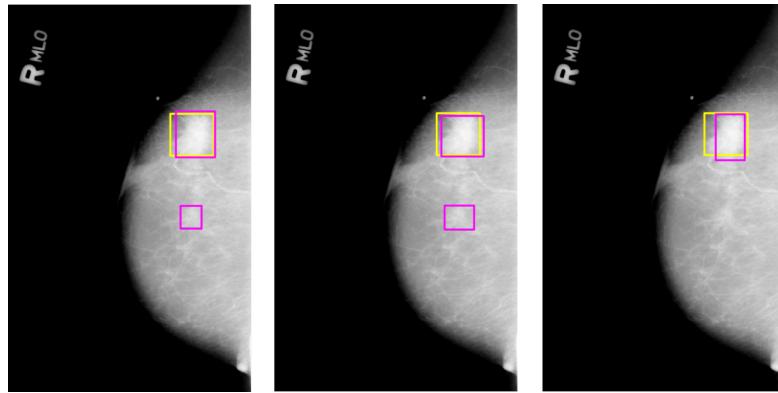


图 5.9 simFaster R-CNN 在 DDSM 数据集上的对正样本预测结果可视化

Figure 5.9 The detection results of positive in vision using simFaster R-CNN based on the DDSM datasets

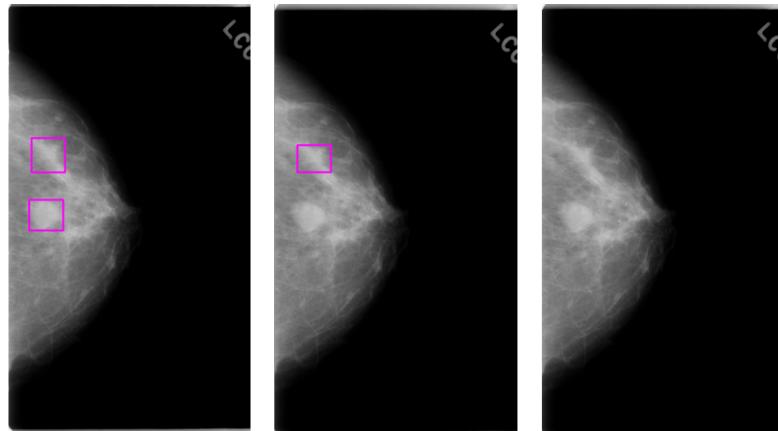


图 5.10 simFaster R-CNN 在 DDSM 数据集上的对负样本预测结果可视化

Figure 5.10 The detection results of negative in vision using simFaster R-CNN based on the DDSM datasets

5.8 本章小结

本章在对两阶段的劣势分析的基础上, 全新设计了一阶段模型 simFaster R-CNN, 其中正样本和负样本共享主干网络, RPN 网络及 Fast R-CNN 网络。在 RPN 阶段, 对大量拥有高度疑似恶性信息的负 anchors 进行损失考虑, 降低其

对真正的正 anchors 的影响；在 Fast R-CNN 阶段，本章引入 similarity loss，着重学习区分恶性样本和正常样本之间的差异，从而提升模型的性能。本章还对 simFaster R-CNN 进行了各种尝试的改良工作，并在公开的大型数据集 DDSM 上与其他人的工作进行了对比分析。

第 6 章 总结与展望

6.1 总结

本文是基于深度学习的基础上，使用从天津肿瘤医院采集的真实数据进行训练，通过对医学数据，尤其是乳腺钼靶的数据的深入分析，结合医院的真实需求，确定评判标准，对数据进行数据预处理以满足传统计算机视觉所要求的数据格式，之后进行数据集划分。考虑到数据集中带有标注的恶性病例数据以及未带标注的正常病例数据，初步设想是设计基于 Faster R-CNN 的先训练恶性病例数据再训练学习恶性病例与正常病例数据之间区别的两阶段模型 weakFaster R-CNN，但由于在训练过程中发现此模型存在耗时长，占用空间等诸多问题，于是创新性地在传统的检测模型 Faster R-CNN 基础上，提出 simFaster R-CNN 模型，针对数据特点分别在 RPN 模块设计针对正常样本的 top likelihood loss 及在 Fast R-CNN 部分针对学习正常与恶性之间区别的 similarity loss，在保证一定精度的前提下，有效地加快了模型训练和测试的速度。

本文是基于广泛存在于医疗影像数据中存在的带有标注的恶性数据及未带标注的正常数据的分析上，在提供一般思路的前提下，提出全新问题，设计全新的损失函数，来为后期的这种数据类型提供开拓性的思路。具体创新点和主要贡献如下：

- 首次考虑医学数据中常见的未全标注信息的弱监督数据；
- 提出适用于医院实际需求的评判标准；
- 使用 weakFaster R-CNN 分阶段学习弱监督数据，完成基于深度学习的乳腺钼靶分类检测工作；
- 提出 simFaster R-CNN，引入度量学习，设计全新损失函数，在学习恶性肿瘤的同时，能够同时区分学习恶性和正常乳腺钼靶之间的区别，满足时间和性能的要求。

6.2 展望

当今世界，以人工智能技术为主要代表的第四次工业革命已经风起云涌，人工智能技术已经逐步走向成熟，各大公司也开始纷纷将技术与实际领域相结合，

以交叉学科的方式将技术应用于产品，从而将产品落地，更好地服务于社会，以及实现公司盈利的前景。在其中，医疗领域无疑以数据量大，数据格式丰富，与国计民生紧紧联系在一起成为各个公司纷纷布局的市场，其中以国家发布的人工智能发展规划中提到的腾讯觅影计划最为著名。通过第一次完完全全从合同签订，到数据采集，标准制定，数据处理，模型设计与分析，模型优化等整个流程走完，我对真实场景的业务数据有了更为深入的了解，并对医疗影像数据有了更为透彻的把握，对于乳腺钼靶图片数据是否患癌也能做个大致的判断。

真实医学场景复杂，数据格式更为多变，每个医院的标准也并不相同，在医学领域的人工智能应用还有漫长的道路需要前进，短短的基于深度学习的乳腺钼靶分类算法只能说是所作的工作的冰山一角。未来肯定会涌现出越来越多的基于深度学习的基于人工智能技术发展起来的医疗影像，乃至医学数据的算法与应用。唯有当技术更好地与领域相结合，技术才能更好地造福于人类，才能更好的服务于社会。

感谢本项目，让我独自负责并全程跟踪一个项目，从而有效锻炼了我对数据的嗅觉以及编码能力。本文还需要进一步研究和改进的是：

- 针对良性和阴性数据之间的更加细微的区别应该单独考虑，所以有必要更加细化 top likelihood loss 和 similarity loss。
- 模型的输入目前是一正一负，在这方面可以考虑引入同正，同负等情况，这样可以增加数据的多样性，一定程度上对数据进行了扩增，从而让模型训练更加充分。
 - 目前针对这种弱监督学习的情况，传统的计算机视觉还有最新研究，比如 Domain Adaptation，异常检测 [56] 等方法，可以适时地引入到乳腺钼靶的分类检测当中。
 - 一个卓越的主干网络可以有效地提高模型的整体预测水平，可以通过引入全新的更加高效的主干网络来提升模型的整体性能。
 - 乳腺钼靶本身的 dicom 格式数据还包含诸多信息，比如病人的年龄，性别等，可以考虑使用结构化数据设计机器学习模型，通过基于深度学习的乳腺钼靶分类检测模型与机器学习模型进行融合，增强模型分类的鲁棒性。
 - 乳腺钼靶还存在钙化，特征征象，结构扭曲等病灶类型。钙化最明显的特征是明亮，数量较多，但难度在于分布各异，大小仅为 0.5mm 到 1mm，这给检

测提出了全新的难度，原有的 simFaster R-CNN 需要着重修改 RPN 模块，提高细小物体的识别能力；对于特征征象，其关键特征在于一点发出放射状，有局灶性收缩或实质的边缘扭曲等现象，其本身并没有良恶之分，可以直接使用传统的 Faster R-CNN 模型训练足矣；对于结构扭曲，其需要通过两个乳房综合判断，所以在模型设计时，需要将单个乳房两个乳房同时输入，判断二者乳房的差异性，进而得出结果。比如钙化病灶较为明亮，

- 由于在医学数据上存在大量的正常样本中存在恶性病灶信息，且正常样本无标注的情况，比如肺结节病灶识别，胃癌病灶识别等等，simFaster R-CNN 模型需要引入更多的数据量以及更多不同的病种数据，对模型进行跨数据的测试，从而让模型能够成为通用的解决这类数据问题的范本。

参考文献

- [1] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
- [2] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]// European conference on computer vision. Springer, 2014: 740-755.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.
- [4] GIRSHICK R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
- [5] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//European Conference on Computer Vision. 2016.
- [8] CHOPRA S, HADSELL R, LECUN Y, et al. Learning a similarity metric discriminatively, with application to face verification[C]//CVPR (1). 2005: 539-546.
- [9] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [10] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
- [11] 陈诗慧, 刘维湘, 秦璟, 等. 基于深度学习和医学图像的癌症计算机辅助诊断研究进展[J]. 生物医学工程学杂志, 2017, 34(2):314r319.
- [12] 胡从依, 柳杰, 刘佩芳. 数字化乳腺 X 线摄影评估乳腺密度, 诊断乳腺癌的研究进展[J]. 中国医学影像技术, 2015, 31(10):1601-1604.
- [13] 杨秋红, 孙一兵. 乳腺癌的影像学诊断及进展[D]. 2005.
- [14] 鲍润贤. 中华影像医学乳腺卷[J]. 北京: 人民卫生出版社, 2002, 66:E2.
- [15] LI N, ZHENG R, ZHANG S, et al. Analysis and prediction of breast cancer incidence trend in china[J]. Zhonghua yu fang yi xue za zhi [Chinese journal of preventive medicine], 2012, 46(8):703-707.

- [16] 冀焕梅,任炜,王丽君,等. 乳腺良恶性肿瘤的钼靶X线鉴别诊断[D]. 2003.
- [17] SATHISH D, KAMATH S, RAJAGOPAL K, et al. Medical imaging techniques and computer aided diagnostic approaches for the detection of breast cancer with an emphasis on thermography-a review[J]. International Journal of Medical Engineering and Informatics, 2016, 8(3):275-299.
- [18] MOREIRA I C, AMARAL I, DOMINGUES I, et al. Inbreast: toward a full-field digital mammographic database[J]. Academic radiology, 2012, 19(2):236-248.
- [19] LEE R S, GIMENEZ F, HOOGI A, et al. Curated breast imaging subset of ddsm. the cancer imaging archive[Z]. 2018.
- [20] LEE R S, GIMENEZ F, HOOGI A, et al. A curated mammography data set for use in computer-aided detection and diagnosis research[J]. Scientific data, 2017, 4:170177.
- [21] CLARK K, VENDT B, SMITH K, et al. The cancer imaging archive (tcia): maintaining and operating a public information repository[J]. Journal of digital imaging, 2013, 26(6):1045-1057.
- [22] 郑光远,刘峡壁,韩光辉. 医学影像计算机辅助检测与诊断系统综述[J]. 软件学报, 2018, 29(5):1471-1514.
- [23] RIEDMILLER M, BRAUN H. A direct adaptive method for faster backpropagation learning: The rprop algorithm[C]//Proceedings of the IEEE international conference on neural networks: volume 1993. San Francisco, 1993: 586-591.
- [24] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7):1527-1554.
- [25] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.
- [26] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]// Proceedings of the 27th international conference on machine learning (ICML-10). 2010: 807-814.
- [27] MORRIS A. Technology watch: Ai can diagnose diabetic retinopathy[J]. Nature Reviews Endocrinology, 2017, 14(2):65.
- [28] LONG E, LIN H, LIU Z, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts[J]. Nature biomedical engineering, 2017, 1(2): 0024.
- [29] 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.

- [30] KERMANY D S, GOLDBAUM M, CAI W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning[J]. *Cell*, 2018, 172(5):1122-1131.
- [31] LIU Y, GADEPALLI K, NOROUZI M, et al. Detecting cancer metastases on gigapixel pathology images[J]. arXiv preprint arXiv:1703.02442, 2017.
- [32] PAN S J, YANG Q. A survey on transfer learning[J]. *IEEE Transactions on knowledge and data engineering*, 2010, 22(10):1345-1359.
- [33] ZHU W, LOU Q, VANG Y S, et al. Deep multi-instance networks with sparse label assignment for whole mammogram classification[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2017: 603-611.
- [34] LÉVY D, JAIN A. Breast mass classification from mammograms using deep convolutional neural networks[J]. arXiv preprint arXiv:1612.00542, 2016.
- [35] SHEN L. End-to-end training for whole image breast cancer diagnosis using an all convolutional design[J]. arXiv preprint arXiv:1708.09427, 2017.
- [36] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.
- [37] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [38] DHUNGEL N, CARNEIRO G, BRADLEY A P. Fully automated classification of mammograms using deep residual neural networks[C]//2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, 2017: 310-314.
- [39] CARNEIRO N D G, BRADLEY A P. Automated mass detection from mammograms using deep learning and random forest[J].
- [40] GERAS K J, WOLFSON S, SHEN Y, et al. High-resolution breast cancer screening with multi-view deep convolutional neural networks[J]. arXiv preprint arXiv:1703.07047, 2017.
- [41] RIBLI D, HORVÁTH A, UNGER Z, et al. Detecting and classifying lesions in mammograms with deep learning[J]. *Scientific reports*, 2018, 8(1):4165.
- [42] AL-MASNI M A, AL-ANTARI M A, PARK J, et al. Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network[C]// 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2017: 1230-1233.
- [43] RIBLI D, HORVÁTH A, UNGER Z, et al. Detecting and classifying lesions in mammograms with deep learning[J]. *Scientific reports*, 2018, 8(1):4165.
- [44] BA L J, CARUANA R. Do deep nets really need to be deep?[J]. *Advances in Neural Information Processing Systems*, 2013:2654-2662.

- [45] UIJLINGS J R, VAN DE SANDE K E, GEVERS T, et al. Selective search for object recognition[J]. International journal of computer vision, 2013, 104(2):154-171.
- [46] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [47] HOSANG J, BENENSON R, SCHIELE B. Learning non-maximum suppression[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4507-4515.
- [48] PAN S J, YANG Q. A survey on transfer learning[J]. IEEE Transactions on knowledge and data engineering, 2010, 22(10):1345-1359.
- [49] ZHOU Z H. A brief introduction to weakly supervised learning[J]. National Science Review, 2017, 5(1):44-53.
- [50] BELLET A, HABRARD A, SEBBAN M. A survey on metric learning for feature vectors and structured data[J]. arXiv preprint arXiv:1306.6709, 2013.
- [51] DIETTERICH T G, LATHROP R H, LOZANO-PéREZ T. Solving the multiple instance problem with axis-parallel rectangles[J]. Artificial Intelligence, 1997, 89(1-2):31-71.
- [52] BRADLEY A P. The use of the area under the roc curve in the evaluation of machine learning algorithms[M]. 1997.
- [53] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[J]. 2015.
- [54] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.
- [55] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 764-773.
- [56] CHANDOLA V, BANERJEE A, KUMAR V. Anomaly detection: A survey[J]. ACM computing surveys (CSUR), 2009, 41(3):15.

作者简历及攻读学位期间发表的学术论文与研究成果

作者简历

2012 年 9 月——2016 年 6 月，在西北工业大学软件与微电子学院获得学士学位。

2016 年 9 月——2019 年 6 月，在中国科学院计算技术研究所攻读硕士学位。

已发表 (或正式接受) 的学术论文:

Learning from Suspected Target: Bootstrapping Performance for Breast Cancer Detection in Mammography(MICCAI 会议在投)

参加的研究项目及获奖情况:

.0.1 参加的研究项目

2017 年 10 月——2018 年 8 月，天津肿瘤医院乳腺钼靶分类项目

2018 年 1 月——2018 年 2 月，阿里天池糖尿病预测比赛

2018 年 8 月——2018 年 9 月，深圳医疗健康大数据创新应用国际大赛

2018 年 9 月——2018 年 11 月，宁波二院胃镜及皮肤镜项目

2018 年 11 月——至今，宁波二院肺结节项目

.0.2 获奖情况

2016 年——2017 年国科大三好学生

阿里天池 fashionAI 服饰属性识别大赛亚军

第四届腾讯广告算法大赛第四名

