



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

博士学位论文

基于深度学习的医学图像内容理解关键技术研究

作者姓名: 陶攀

指导教师: 付忠良 研究员

学位类别: 工学博士

学科专业: 计算机软件与理论

培养单位: 中国科学院成都计算机应用研究所

2018年03月

Research on Key Technologies in Medical Image Processing
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
Doctor of Engineering
in Computer Software and Theory
By
Pan Tao
Supervisor: Professor Zhongliang FU

Chengdu Institute of Computer Applications
Chinese Academy of Sciences

March, 2018

中国科学院大学

研究生学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。

作者签名：

日 期：

中国科学院大学

学位论文授权使用声明

本人完全了解并同意遵守中国科学院有关保存和使用学位论文的规定，即中国科学院有权保留送交学位论文的副本，允许该论文被查阅，可以按照学术研究公开原则和保护知识产权的原则公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密及延迟公开的学位论文在解密或延迟期后适用本声明。

作者签名：

导师签名：

日 期：

日 期：

摘要

对医学图像的内容理解是应用计算机视觉与人工智能进行医学影像分析的最基本问题之一，从二维或三维图像数据中理解图像内容一直是医学图像应用研究的重点领域，涉及到感兴趣目标的去噪、分类、检测、分割和检索等研究内容。由于图像内容理解问题本身的困难性，并且医学图像存在特有的领域先验，如超声特有的斑点噪声，衰减，阴影等因素影响，导致目标受尺度、旋转、形变等而形成不同的成像，使得用计算机对医学图像中的内容进行鲁棒的表达与识别依然是一个严峻的挑战。主要原因之一是不同具体任务分别处于图像抽象的不同层次，如何有效结合低层的图像数据信息和高层的语义信息是解决医学图像的内容理解问题的关键所在。

深度学习为代表的人工智能技术在医学影像分析领域呈现出了非常引人注目的研究进展。但仍有一系列问题难以克服：(1) 不同网络结构对特征学习的影响（如何设计高效结构理论可解释性）；(2) 学习到的特征表示如此高效（能同时应用于分类、检测、分割等）的原因是什么；(3) 深度学习的优异泛化能力从何而来？针对以上问题，论文工作主要涉及深度卷积神经网络自动提取分层递阶图像特征，并在不同层次特征应用到不同抽象层次图像内容理解的研究，论文在深入分析传统计算机视觉算法的基础上，重点研究了深度学习模型的可视化、基于形状先验的统计形状模型，并致力于利用形状先验信息结合深度卷积神经网络解决医学特定目标检测与分割问题。

论文的主要工作和创新之处在于：

针对特征表示的高层语义识别问题，通过构建标准切面数据库，提出了一种基于深度卷积神经网络的超声心动图标准切面自动识别方法，该算法针对网络全连接层占有模型大部分参数的缺点，引入空间金字塔均值池化替代全连接层，获得更多的空间结构信息，利用全局空间金字塔均值池化方法进行微调迁移学习，并大大减少模型参数、降低过拟合风险，通过类别显著性区域将类似注意力机制引入模型可视化过程，详尽分析了数据规模对模型分类精度的影响，并对模型的可解释性和有效性进行了分析。

针对基于深度卷积神经网络的图像分类模型的可解释性问题，通过评估模型特征空间的潜在可表示性，提出一种用于改善理解模型特征空间的可视化方法。给定任何已训练的深度卷积网络模型，引入了通过激活最大化获得的图像可

解释性的正则化方法，结合现有正则化方法提出空间金字塔分解方法，利用构建多层拉普拉斯金字塔主动提升目标图像特征空间的低频分量，结合多层次金字塔调整其特征空间的高频分量得到较优可视化效果。并通过限制可视化区域，提出利用类别显著性激活图技术加以压制上下文无关信息，可进一步改善可视化效果。该模型有效克服了原有可视化方法中由于不能主动调整高低频分量等原因造成的可视化图像语义重复和低效率等问题。

针对自动检测医学图像中指定目标时存在的问题，提出了一种基于深度学习自动检测目标位置和估计对象姿态的算法。该算法基于区域深度卷积神经网络和目标结构的先验知识，采用区域生成候选框网络、感兴趣区域池化策略，引入包括分类损失、边框位置回归定位损失和像平面内朝向损失的多任务损失函数，近似优化一个端到端的有监督定位网络，能快速地对医学图像中目标自动定位，有效地为下一步的分割和参数自动提取提供定位结果。并在超声心动图左心室检测中提出利用检测额外标记点：二尖瓣环、心内膜垫和心尖，能高效地对左心室朝向姿态进行估计。

针对特征表示的底层视觉任务：图像去噪和分割中存在的问题，我们提出了一个有监督多层次残差卷积网络框架，结合不同损失函数学习端到端映射变换；针对医学超声图像的对比度低、存在斑点噪声导致难以分割的问题，提出一种利用沙漏卷积神经网络特征的多尺度形状模型分割方法，自动定位经食道超声心动图中心室并全自动分割心室内外膜。首先，结合梯度方向直方图特征和支持向量机的心室自动检测方法，自动确定分割模型中的初始轮廓；其次，将心室分割任务纳入统计形变模型形状特征点对齐任务框架，通过比较不同外观纹理特征和激活图，包括传统手工设计的特征和利用深度学习自动学习的卷积特征，提出利用堆叠多级沙漏卷积网络建模心室外观的全局和局部信息，统一活动外观模型和局部受限模型的概率形式，采用反向组合梯度下降算法迭代优化分割结果，完成左心室轮廓的自动提取。然后，以医生手动勾勒的轮廓作为“金标准”，通过构造心室分割数据集以评价算法，且提出了扩充数据样本的方法来克服深度卷积网络过拟合问题，进行详尽实验讨论分析了基于不同层级的多级沙漏卷积网络对全局和局部纹理特征建模能力对分割效果的影响。实验结果表明，卷积模块允许网络提取专门用于指定任务的特征，并通过实验显示其优于手工设计的特征。该方法分割效果优于传统形状对齐方法，能够解决自动定位超声心动图中左心室的初始轮廓和弱边界自动分割的问题。

关键词：深度学习，卷积神经网络，医学图像分析，特征表示，深度可视化

Abstract

The understanding of the content of medical images is one of the most basic issues in the application of computer vision and artificial intelligence in medical image analysis. Understanding the content of images from two-dimensional or three-dimensional image data has always been a key area of medical image application research and related to interested targets. Denoising, classification, detection, segmentation and retrieval of research content. Due to the difficulty in understanding the image content itself, and the uniqueness of the medical image, such as the speckle noise, attenuation, shadow and other factors that are unique to ultrasound, the target is subject to different scales, rotations, deformations, etc. to form different images. Robust expression and recognition of the contents of medical images with computers is still a serious challenge. One of the main reasons is that different specific tasks are at different levels of image abstraction. How to effectively combine low-level image data information and high-level semantic information is the key to solving the content understanding problem of medical images.

The artificial intelligence technology represented by deep learning has shown very remarkable research progress in the field of medical image analysis. However, there are still a series of problems that are difficult to overcome: (1) the influence of different network structures on feature learning (how to design efficient structural theory interpretability; (2) why the learned features can work? visualization; (3) excellence in deep learning Where is the generalization ability coming from? Detecting segmentation. To solve the above problems, the dissertation mainly involves the automatic extraction of hierarchical hierarchical image features by deep convolutional neural networks and the application of different levels of features to the understanding of image content at different levels of abstraction. On the basis of in-depth analysis of traditional computer vision algorithms, a deep learning model and a statistical shape model based on shape priors are focused on, and it is devoted to solving medical specific target detection and segmentation problems using shape prior information combined with deep convolutional neural network.

The main work and innovation of the dissertation lies in:

By constructing a standard database, an automatic recognition method based on

deep convolutional neural network for the standard section of echocardiogram is proposed. This algorithm aims at the shortcomings of most of the parameters of the network full-connection layer occupation model, and introduces the spatial pyramid mean-value pool instead of the full connection. At the layer, more spatial structure information is obtained, and the global spatial pyramid mean pooling method is used to perform fine-tuning migration learning, and the model parameters are greatly reduced, the overfitting risk is reduced, and similar attention mechanisms are introduced into the model visualization process through category salient regions. The effect of data size on the classification accuracy of the model was analyzed in detail, and the interpretability and effectiveness of the model were analyzed.

For the interpretability problem of image classification model based on deep convolutional neural network, by evaluating the potential representation of model feature space, a visualization method for improving the understanding of model feature space is proposed. Given any trained deep convolutional network model, a regularization method of image interpretability obtained by maximizing the activation is introduced, a spatial pyramid decomposition method is proposed in combination with the existing regularization method, and a multilayer Laplacian pyramid is constructed. Actively enhance the low-frequency components of the target image feature space, combine multilayer Gaussian pyramid to adjust the high-frequency components of its feature space to obtain a better visualization effect. And by limiting the visualization area, the use of category-significant activation graph techniques to suppress context-free information can further improve visualization. The model effectively overcomes the problems of semantic visual duplication and low efficiency caused by the inability to actively adjust high and low frequency components in the original visualization methods.

Aiming at the problem of automatically detecting the target in the medical image, an algorithm based on deep learning to automatically detect the target position and estimate the pose of the object is proposed. The algorithm is based on prior knowledge of regional deep convolutional neural networks and target structures, and uses region-generating candidate frame networks and pooling strategies for regions of interest to introduce multi-tasks including loss of classification, position regression of borders, orientation loss, and orientation loss in the image plane. The loss function, approximately optimizing an end-to-end supervised positioning network, can quickly locate the

target in the medical image and effectively provide the positioning results for the next segmentation and automatic parameter extraction. In echocardiographic left ventricular detection, it is proposed to use the detection of additional markers: mitral annulus, endocardial cushion, and apex, which can efficiently estimate left ventricle orientation.

For the underlying visual tasks of feature representation: problems in image denoising and segmentation, we propose a supervised multi-level residual convolutional network framework that combines end-to-end mapping transformations with different loss functions. The input is a noisy image and the original image, and the output is a denoised image. In view of the low contrast of medical ultrasound images and the difficulty of segmentation due to speckle noise, a multi-scale shape model segmentation method based on the features of the hourglass convolutional neural network is proposed to automatically locate the center of transesophageal echocardiography and to automatically segment the indoor and outdoor membrane. Firstly, combining the gradient direction histogram feature and the ventricular auto-detection method of the support vector machine, the initial contour in the segmentation model is automatically determined; secondly, the ventricular segmentation task is incorporated into the statistical deformation model shape feature point alignment task framework, and the different appearance texture features are compared. And activation diagrams, including the features of traditional manual design and the convolution features that are learned automatically with deep learning, propose the use of stacked multi-level hourglass convolutional networks to model the global and local information of the ventricular appearance, unifying the active appearance model and the locally constrained model. In the form of probability, the inverse combined gradient descent algorithm is used to iteratively optimize the segmentation results to complete the automatic extraction of left ventricular contours. Then, the contours manually outlined by doctors are used as the "gold standard" to construct the ventricular segmentation data set to evaluate the algorithm, and a method of extending the data samples is proposed to overcome the over-fitting problem of the deep convolutional network. Detailed experiments are discussed and analyzed. The Effects of Different Levels of Multilevel Sandglass Convolution Networks on Global and Local Texture Feature Modeling and Segmentation Effects . The experimental results show that the convolution module allows the network to extract features that are specific to a given task and show that it is superior to manual design features through experiments.

The segmentation effect of this method is better than the traditional shape alignment method. It can solve the problem of automatic segmentation of the initial contour and the weak boundary of the left ventricle in the automatic positioning echocardiogram.

Keywords: Deep learning, convolutional neural network, medical image analysis, feature representation, depth visualization

目 录

目 录	VII
图形列表	XI
表格列表	XIII
第 1 章 绪论	1
1.1 研究背景及现实意义	1
1.1.1 医学影像分析的研究背景	1
1.1.2 课题研究意义	2
1.2 国内外研究现状	3
1.2.1 图像内容理解的研究现状	3
1.2.2 医学图像分析应用于机器学习算法的范例	7
1.2.3 深度学习在医学图像分析应用的研究现状	7
1.3 创新点及全文结构	12
1.4 论文的章节安排	13
第 2 章 基本理论概述	15
2.1 机器学习算法	15
2.1.1 神经网络	15
2.1.2 卷积神经网络	17
2.2 深度卷积神经网络	18
2.2.1 通用分类框架	18
2.2.2 多通路的卷积神经网络结构	20
2.2.3 全卷积网络结构	20
2.3 循环神经网络	21
2.4 无监督深度学习	22
2.4.1 自编码器	22
2.4.2 变分自编码器和深度生成对抗网络	22
2.4.3 受限波兹曼机和深度信念网络	22
2.5 深度学习相关技术	23
2.5.1 激活函数	23
2.5.2 损失函数	26
2.5.3 优化方法	27
2.5.4 正则化方法	29
2.6 小结与讨论	30

第 3 章 特征表示的高层语义应用	31
3.1 超声心动图切面的自动识别方法	31
3.2 Deep-Echo 模型	32
3.2.1 全卷积的网络	33
3.2.2 空间金字塔均值池化层	33
3.3 微调迁移学习	34
3.4 类别显著激活映射图	34
3.5 实验结果和分析	35
3.5.1 实验数据选取和实验方法	35
3.5.2 识别实验结果和分析	36
3.5.3 模型可解释性实验结果分析	36
3.6 小结与讨论	39
第 4 章 特征表示的深度可视化	41
4.1 空间金字塔分解的深度可视化方法	41
4.2 可视化方法的数学模型	42
4.3 梯度更新的可视化方法	43
4.3.1 p 范数正则化方法	43
4.3.2 高斯模糊和 TV 变分	43
4.3.3 基于数据统计先验	44
4.4 空间金字塔分解	44
4.4.1 高斯和拉普拉斯金字塔分解	44
4.4.2 梯度归一化	45
4.4.3 类别激活图限制可视化区域	46
4.4.4 优化方法	46
4.5 实验结果分析和讨论	47
4.5.1 金字塔分解可视化实验结果	48
4.5.2 引入类别显著性的可视化	49
4.6 小结与讨论	49
第 5 章 特征表示的中层结构性语义应用	51
5.1 心室的计算机辅助检测方法	51
5.2 区域卷积神经网络概览	52
5.2.1 物体检测形式化定义	52
5.2.2 区域卷积神经网络的演进	53
5.3 候选区域生成网络及其改进	53
5.3.1 候选区域生成网络模型结构	54

5.3.2 仿射变换候选框	55
5.3.3 朝向回归损失函数	56
5.3.4 带朝向的多任务损失函数	57
5.3.5 困难样例挖掘	57
5.4 实验结果分析和讨论	57
5.4.1 检测 MRI 左心室短轴	58
5.4.2 检测左心室及其朝向	59
5.5 小结与讨论	61
第 6 章 特征表示的底层语义应用	63
6.1 图像的去噪方法	63
6.1.1 相关工作	64
6.1.2 提出的方法	65
6.1.3 实验结果分析和讨论	68
6.1.4 详尽分析模型	68
6.2 形状对齐的心室的分割方法	70
6.2.1 初始位置定位和特征点标注	72
6.2.2 AAM 模型和 CLM 模型	73
6.2.3 结合卷积网络特征的形状对齐模型	75
6.2.4 实验结果分析和讨论	78
6.2.5 小结与讨论	81
第 7 章 总结与展望	83
7.1 研究总结	83
7.2 研究展望	85
参考文献	87
攻读学位期间发表的学术论文与科研成果	101
致 谢	105

图形列表

1.1 David Marr 视觉计算理论的分层结构示意图	4
1.2 局部特征结合机器学习的机器视觉代表性方法举例	4
1.3 传统图像识别方法与深度学习方法对比	5
1.4 统计机器学习方法分类总结	6
1.5 影像组学处理流程	7
1.6 应用深度学习的一些医学影像图集	8
2.1 前馈神经网络（也称为多层感知器）结构示例示意图	16
2.2 经典深度网络结构示意图	17
2.3 深度卷积神经网络结构演进图	18
2.4 AlexNet 网络结构示意图 ^[1]	19
2.5 卷积网络中的各种卷积操作图	21
2.6 各种激活函数比较图	25
2.7 无 Dropout 网络，DropOut 网络和 DropConnect 网络的图示。 ...	30
3.1 Deep-Echo 模型结构示意图	33
3.2 七类标准切面超声心动图及数量分布	35
3.3 Deep-Echo 模型分类的混淆矩阵	37
3.4 不同数据量的平均分类精度	37
3.5 各类切面的原图和显著性热力图	38
3.6 深度模型泛化性能可视化分析	38
4.1 高斯和拉普拉斯金字塔	45
4.2 不同模型类别可视化实验结果	47
4.3 不同正则化方法的可视化效果	48
4.4 金字塔分解正则化可视化效果	49
4.5 引入类别激活图的可视化	49
5.1 区域生成网络模型的各个组件图	54
5.2 考虑物体朝向的区域生成网络模型结构示意图	55
5.3 利用检测分析工具 ^[2] 的检测结果	61
5.4 医学图像目标检测结果示意图	61
6.1 提出网络的整体架构图	65
6.2 RED-NET 的网络结构示意图	66

6.3 不同损失在基准数据集上的定量比较结果表	69
6.4 四种其他噪声类型的去噪比较结果图	70
6.5 初始位置定位结果和特征点标注示意图	72
6.6 定性比较三种不同特征激活图及相应的局部响应映射图	73
6.7 不同特征的形状无关纹理图	76
6.8 具有不同分辨率的特征生成多分辨率注意力图	76
6.9 不同特征 AAM 的左心室内外膜分割性能比较	79
6.10 分割结果对比	80
6.11 用 HG 网络预测心室内外膜成功和失败的案例	81

表格列表

3.1 不同模型分类精度比较	37
5.1 不同模型检测精度的比较表	58
5.2 不同模型的检测精度表	60
5.3 不同旋转角度分类检测性能比较	60
6.1 DeNET-R 和 RED-NET 网络的配置参数表	67
6.2 <i>Set14</i> 上的单一噪声水平的图像去噪结果定量比较表	68
6.3 不同分割方法与专家标注的对比统计	81

第1章 绪论

1.1 研究背景及现实意义

1.1.1 医学影像分析的研究背景

临床医学历经几百年的发展，传统的“视、触、叩、听”已经不能满足现代化医疗的诊断需求，医学影像极大地变革了传统的诊疗体系，成熟的成像模式不断完善，新技术不断涌现，在疾病筛查、早期诊断、治疗方案选择和预后评估等方面发挥着举足轻重的作用。用于医疗诊断的影像使医生能够更早地发现疾病并改善患者预后，介入或术中成像有助于消除和治愈许多检测到的疾病，能更早更有效地诊断身心健康状况，为临床诊疗提供了全面的视角和丰富的信息，迅速地被广泛应用于临床领域。目前临床医学已经无法离开医学影像，并且随着医学影像的发展，临床诊疗将越来越依赖于影像。

在医学成像中，疾病的准确诊断和评估取决于准确的图像采集和图像解释。近年来随着大数据和计算机通讯技术的发展，影像归档和通信系统（Picture Archiving and Communication Systems, PACS）和医学数字成像和通信标准（Digital Imaging and Communications in Medicine, DICOM）等技术的成熟既解决了影像的采集问题，又解决了数据的传输和存储问题。医学影像自 1895 年伦琴发现 X 射线以来，综合利用物理中的各种物质波、光电子技术以及计算机技术，从宏观到微观，由静态到动态，由单模到多模，由 2D 到 3D，形成了各种的成像模式包括 X 射线、超声、计算机断层扫描（Computed Tomography, CT）、磁共振成像（Magnetic Resonance Imaging, MRI）、正电子断层扫描、以及内窥镜和病理切片图像等。

然而，图像解读过程最近才开始受益于不断提升的计算机技术。一旦将医学图像扫描加载到计算机中，研究人员就致力于构建自动化分析系统。早期医学影像分析是通过处理低阶像素（边缘、线）和数学建模（拟合线，圆和椭圆）并应用复合规则系统解决特定任务。而基于训练数据的监督机器学习在医学影像分析中越来越受欢迎，例如活动形状模型（用于分割），图谱（Atlas）方法（分割和配准）以及特征提取和统计分类器的使用（用于计算机辅助检测和诊断），这种模式识别或机器学习方法现在仍然非常流行，并构成了许多成功的商用医学影像分析系统的基础。

医学影像智能分析从简单的计算机辅助检测（Computer Aid Detection,CAD）

发展到如今火热的影像组学 (Radiomics)^[3]，它将影像内包含的所有信息提取出来然后进行综合系统化分析。更确切的说，影像组学是采用自动化算法从影像的兴趣区内提取出大量的特征信息作为研究对象，并进一步采用多样化的统计分析和数据挖掘方法从高通量量信息中提取和剥离出真正起作用的关键信息，最终用于疾病的辅助诊断、分类或分级。

当今世界医疗卫生系统每天都会浪费大量的资源和时间，医学图像的大多数解释都是由医生完成的；然而不同的影像质量和不同的工作流程会导致临床上医生对影像内容的理解具有很大的主观性，不同解译人员之间的巨大差异，容易造成对医学影像内容的错误理解会造成错误诊断，导致很多不必要的额外检查，导致治疗计划的延迟，大大减少了如果早期正确发现的生存率或缓解率。即使对于有经验的临床医生来说，图像内容的精确分析也是具有挑战性和耗时的任务。

1.1.2 课题研究意义

计算机智能化工具，结合机器学习和计算机视觉的人工智能（Artificial Intelligence,AI）算法是促进智能诊断的关键因素，尤其是其代表性方法深度学习，深度学习是有多个处理层的计算模型，能学习具有多层次抽象的数据的表示，不但提高了图像内容理解的准确性，它还在数据分析方面开辟了新的前沿，给许多领域都带来了显著改善，包括语音识别、视觉对象识别、对象检测和许多其它领域，例如药物发现和基因组学等。其能不但改善诊疗以及预后效果，还将作为一个固化已有经验的临床助理，给临床医生的工作方式带来转变，显著提高工作流程效率，而不增加临床医生的负担。

人工智能进入医学影像领域，主要为解决当前医学影像分析面临误诊率高、医师缺口大的问题，其一，如何快速自动准确的分析日益增长的医疗影像设备所产生的具有医学分析和指导价值的结构化和非结构化的海量数据。同时由于标准的数据和规范的标注是医疗人工智能发展的前提，反过来，人工智能可以推动医疗数据的标准化建设。其中结构类影像，比如 X 光、CT，能够非常直观地观察生理结构，判断是否有物理变化的病变，基于人工智能算法实现图像中解剖结构的准确分类和定位是的全自动诊断的基础。而功能类影像能够研究器官对某种物质的代谢能力，从而反映该器官功能，其缺点是不能自主定位异常，不能直接反映真实生理结构，只能通过影像像素和内容综合理解程度来分析代谢的强弱程度，不能实现具有统计学意义的定量分析，诊断结果只能全凭医生的肉眼和经验来判断，导致较高的误诊漏诊率。若结合人工智能算法在定量、定位、精准

量化的基础上，通过与正常数据进行统计比对，大大提高了对病变分析的深度，在实现自动辅助诊断上就具备了现实意义。

其二，有经验的医师缺口大，成长曲线陡峭，医师数量增长远不及影像数据增长，在短时间内理解影像数据给出准确诊断的压力会越来越大，远超负荷。AI 算法未来可以比专业人员更快，更准确地提取大量数据，挖掘模式和预测，加强疾病诊断，提供治疗计划。突破主要来自 AI 领域的深度学习方法，其模拟更高层次抽象并决定了高维特征空间中的最佳决策边界，对某些疾病的影像诊断水平已能达到专家水准，同时通过对深度学习理论的研究人工智能的决策机制，未来或为实现精准诊疗、智慧医疗和保障大众健康带来突破性进展。这对于提升基层医疗服务水平、助推分级诊疗将具有重大意义。

总而言之，结合计算机视觉和深度学习算法，能提升医学图像分析领域模式识别的能力，从解剖结构到疾病，让计算机系统能够自主学习经验数据，不仅能更帮助患者更快速地完成健康检查，同时也可以帮助影像医生减少阅片时间，提升效率，降低误诊的概率，既能为经验不足的医生提供辅助决策建议，也能帮助专家节约时间。

1.2 国内外研究现状

1.2.1 图像内容理解的研究现状

医学影像分析主要涉及图像的内容理解，而实现图像的内容理解是计算机视觉的终极目标^[4]。可根据研究对象和特征表示的目标分成低-中-高三层。底层问题主要是针对图像本身及其内在属性的分析，依据灰度值进一步推断物体的几何结构，常见内容有去噪，滤波和特征增强。高层问题主要是针对图像内容的理解和认知层面的，如识别与分割图像中的特定物体与其行为；场景理解和行为推断。中层介于两者之间，着重于不同层级的特征表示。

计算机视觉的起源可以追溯到 1966 年，以明斯基布置让学生构造一个视觉系统的暑期项目为起点，早期研究主要集中于底层视觉信息处理。Marr^[5] 定义了计算机视觉研究的三个层次，自底向上 (bottom-up) 的分成表达、算法、和实现，将视觉描述为从二维视觉阵列（在视网膜上）到三维描述世界输出的多层表达，从基本简约图 (Primal Sketch)，到 2.5D 深度简约图，再到 3D 模型，如图1.1，希望先把三维结构从图像里面恢复出来，在此基础上再去做理解和判断。

从 20 世纪 80 年代开始，基于逻辑学和知识库推理的专家推理系统在人工智能领域大行其道，计算机视觉的理论方法论随之改变，Biederman 在 Marr 的

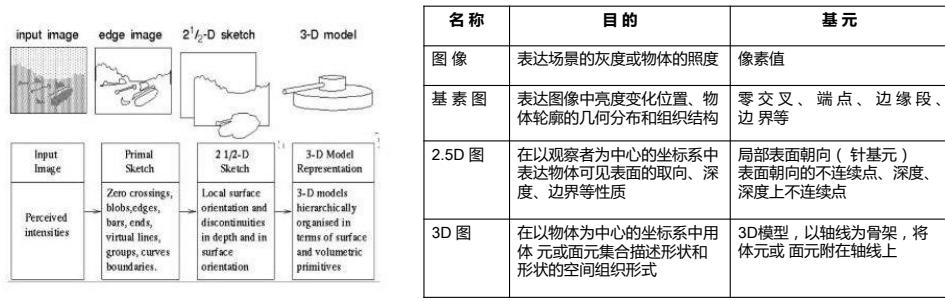


Figure 1.1 David Marr 视觉计算理论的分层结构示意图

基础上提出组成识别理论 (Recognition by Component Theory)。该理论认为通过把复杂对象解译 (Image Parsing) 为简单的部件形状，就可进行视觉统计模式匹配。Treisman 等^[6] 提出特征融合理论 (Feature Intergration)，认为视觉处理是一个以自下而上、局部交互作用的过程。Itti 等^[7] 提出适用于自然图像的高斯金字塔视觉注意机制模型模型。研究发现要让计算机理解图像，不一定先要恢复物体的三维结构。利用包括统计形变模型^[4] 以及能量泛函模型^[8]，对图像进行解译 (Image Parsing)，将物体形状、颜色和纹理等先验特征进行统计模式匹配，寻找一个层次化、结构化的解释是计算视觉的核心问题。

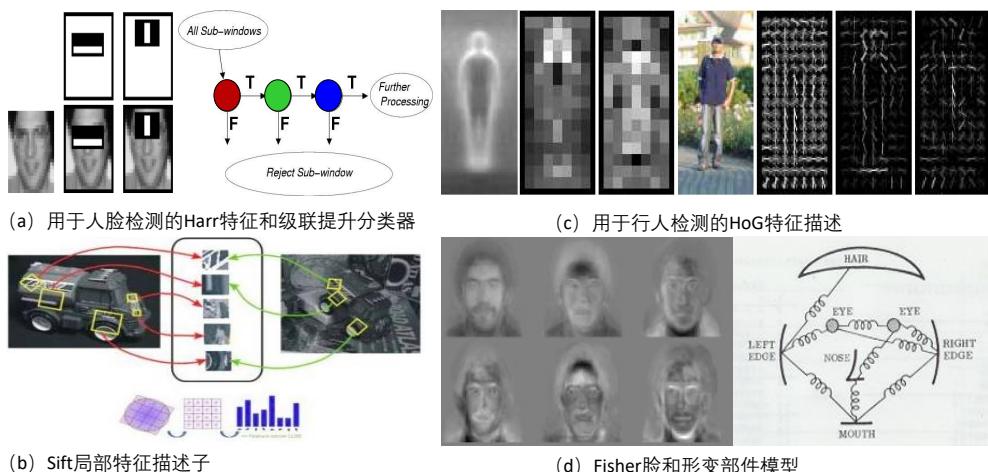


Figure 1.2 局部特征结合机器学习的机器视觉代表性方法举例

进入新世纪后，互联网产生了海量图像数据，大规模数据集也随之出现，导致面向描述符的机器学习方法开始盛行，之前需通过规则、知识或统计模型设计形状、颜色、纹理等先验表征，易受额外影响不具有统计不变性，通过统计手段寻找能够刻画对象最本质的一些局部特征，其进展直接导致诸多应用的出现，如：图像搜索技术进入实用。目标识别领域涌现出空间金字塔 (Spatial Pyramids)^[9]、矢量量化 (Vector Quantization)^[10]，以及在计算机视觉的各个阶段使用的各种机

器学习工具。Lowe^[11] 提出旋转和尺度不变的局部特征描述符 (Scale Invariant Feature Transform,SIFT)，改进后成为模式识别中的经典模型。Viola 和 Jones^[12] 提出基于哈尔特征和级联 AdaBoost 分类器人脸检测框架；Grauman 等^[13] 发展了词袋模型 (Bag of Word, Bow) 用于图像物体识别；Dala 等^[14] 提出方向梯度直方图 (Hog) 特征，利用变形部件模型结合支持向量机进行行人检测。因此，从完全由人类设计的系统转变为由使用计算机提取示例数据进行训练获得特征向量的系统，计算机算法决定了高维特征空间中的最佳决策边界，设计此类系统的关键步骤是从图像中提取判别特征，该过程仍然是由人类手工设计完成的。

故合理的下一步就是让计算机自动从数据学习最佳特征表示，如图1.3揭示了手动设计特征的传统方法与自动提取特征的深度学习方法的区别，模型由多层组成，将输入数据转换为输出（例如疾病存在/不存在），同时学习更高层次的特征。目前最成功的图像分析模型是卷积神经网络 (convolutional neural networks, CNN)，Fukushima 早在 1980 年就提出有关 CNN 的工作^[15]。LeCun 等^[16] 提出的 LeNet 是在手写数字识别领域的第一个成功的实际应用。在医学图像处理中，GPU 首先被引入用于分割，重建和配准，然后被用于机器学习。但在开发各种新技术以有效地训练深度网络之前，CNN 在计算机视觉领域并没太成功。分水岭是 Krizhevsky 等人^[1] 在 2012 年 12 月参加 ImageNet^[17] 识别竞赛挑战做出的贡献，提出 AlexNet 大幅度（领先 10%）赢得该竞赛。随后，使用相关但更深的架构取得了进一步的进展^[18–20]。在计算机视觉各个任务领域，深度卷积神经网络现在已经成为首要选择的技术。

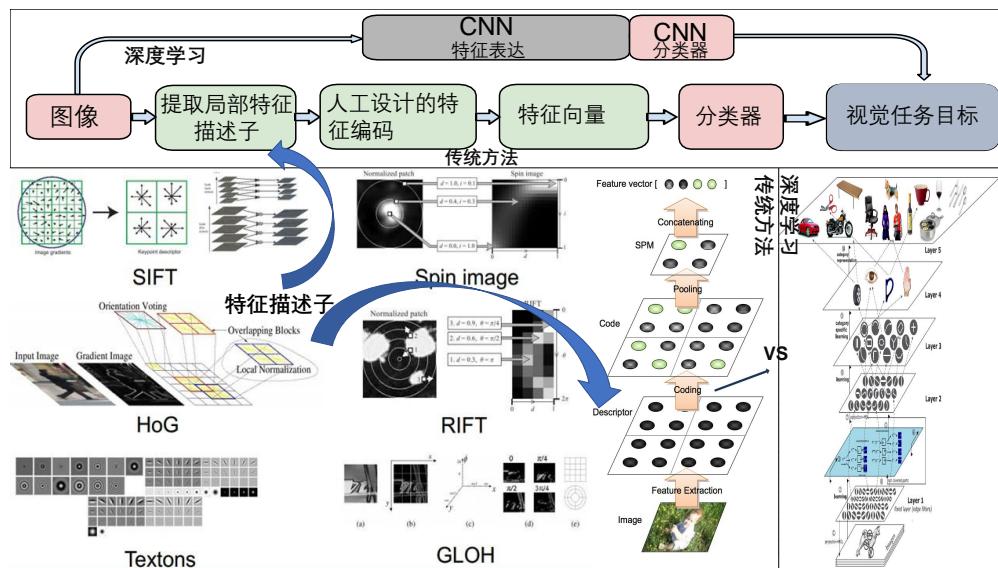


Figure 1.3 传统图像识别方法与深度学习方法对比

另外一些研究者从模拟生物视觉机制出发，通过对感受野、注意机制、颜色特性和光流等方面进行计算机视觉研究。Hubel 和 Wiesel^[21] 对猫视觉皮层细胞的信息处理模式做了深入研究，提出视觉感受野（Receptive Field）概念，进一步发现了视觉皮层通路中对于信息的分层处理机制。Riesenhuber 等^[22] 首次从生物学的角度上模拟建立分层的视觉处理模型（Hierarchical Model and X, HMAX），其模型与文献 [7] 中的模型类似。Berthold 等^[23] 认为人类感知不是对视网膜上降采样图像的解释，而是对光学排列和流动的直接感受体验，提出光流（Optical Flow）用于描述图像灰度的表面活动形式，即获取运动场得到丰富的场景运动和场景结构等信息。

另外一些研究者从模拟生物视觉认知机制出发，McClelland 和 Pitts 等依据神经元连接方式提出神经细胞模型 MP，相关节点单元（Node）是最小的加工单元，节点之间通过通过兴奋和抑制两种连接方式联结成网络，利用“Hebb 规则”的学习机制，指出神经元在不同时刻需具有变化的强度激活值（Activation Value），该理论成为连接主义理论的代表性理论。Rosenblatt 等^[24] 提出了由单层神经元组成的感知机（Perceptron）来完成线性分类任务；Fukushima 等^[15] 提出了神经网络多层的结构，但是并没有提出如何学习这个结构；Rumelhart 和 Hinton 等人^[25] 根据求导的链式法则提出了反向传播（Backpropagation, BP）算法，解决了多层次神经网络所需要的复杂计算量问题。Hopfield 提出利用能量函数的概念来研究一类具有固定权值的循环神经网络的稳定性并付诸电路实现。

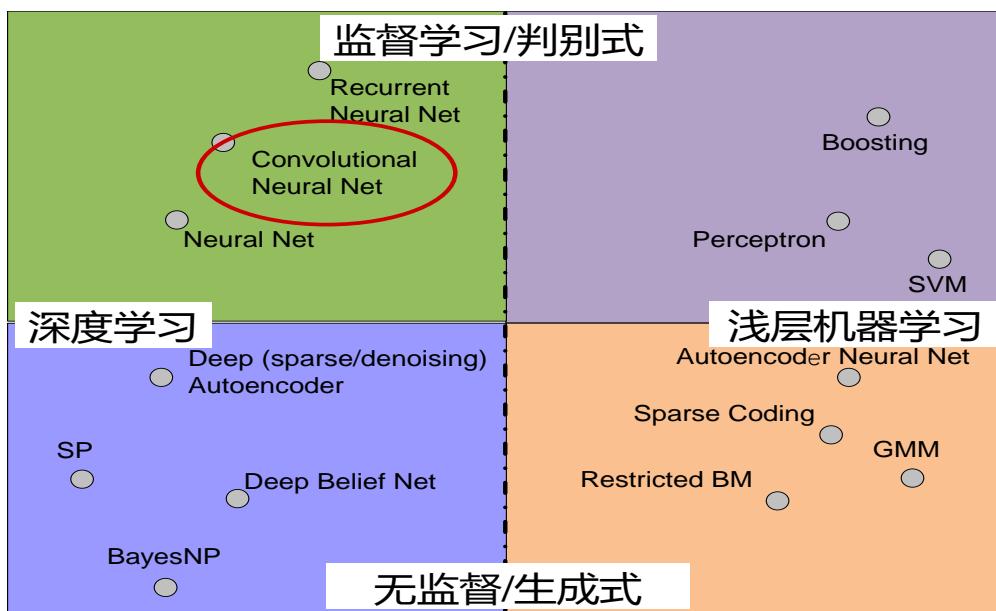


Figure 1.4 统计机器学习方法分类总结

1.2.2 医学图像分析应用于机器学习算法的范例

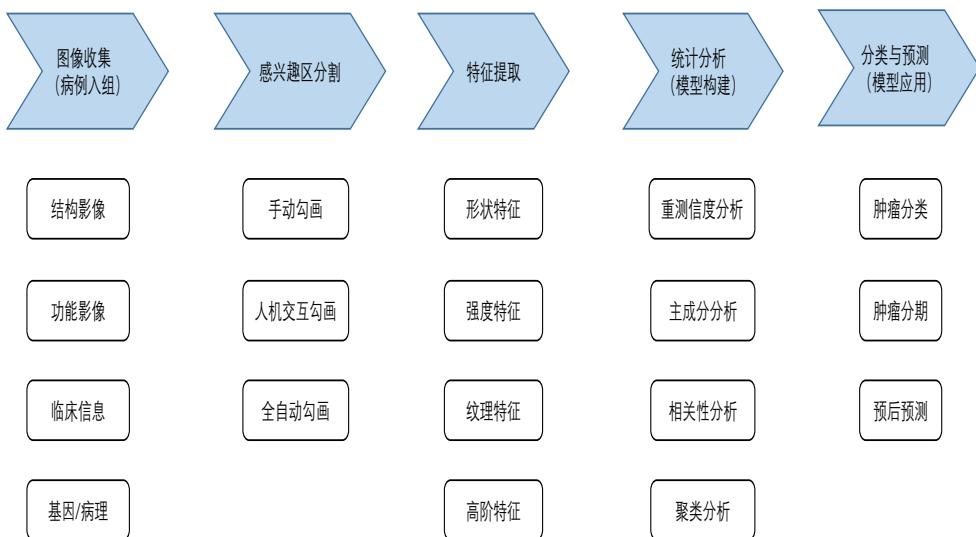


Figure 1.5 影像组学处理流程

以当前传统机器学习算法在医学领域的典型应用影像组学为例介绍相关流程，将影像以高通量方式转换为可挖掘的数据。图1.6给出了一个简单的流程说明，图像收集也就是临床试验中常说的“病例入组”，包含两个要点制定好入组标准是最基础的要求；在结构或功能影像以外包含必要的临床病历信息或基因、病理数据已经成为当前新的研究趋势；感兴趣区分割，通常指的就是获得病灶，比如癌症或肿瘤区域的（勾画），但目前来看，临床用的最多的还是纯手工。这也是影像组学中比较耗费高级人力的一个步骤了；特征提取，图1.6中列出了目前常用的4大类特征。真的是4“大”类，因为每类都可以提取出数十、乃至上百的信息；统计分析，提取出来数百上千的特征进行统计分析，其实核心目的就是减少特征的数量，找到少数的真正关键的特征。找到这些关键特征后，还需要建立起相应的预测或分类模型。分类与预测，千挑万选建立起来的模型肯定要应用才能体现价值。目前影像组学应用最多的领域包括肿瘤分类、肿瘤分期和愈后预测等。

1.2.3 深度学习在医学图像分析应用的研究现状

医学影像分析界业已注意到当前从使用手工设计特征转换到从数据中学习特征的关键进展，在 AlexNet 突破之前，依据模型结构深浅和监督与非监督四个维度度现有机器学习方法进行归类（如图1.4）。通过对有限样本统计理论、和 VC 维泛化能力等机器学习理论的研究，相继提出了各种各样的浅层机器学习模型，

如支持向量机（Support Vector Machines, SVM）、集成学习（Boosting）、稀疏编码（Sparse Coding）等，这些模型基本上可视作只有一层隐层节点（如 SVM、Boosting）的浅层神经网络，神经网络和深度学习背后神经科学的基本理念如上节所述。这些模型无论是在理论分析还是实际应用于医学图像分析中都获得了不小的成功，特别是用于检测异常情况，而且还用于分割等相关领域。尽管有这些发展，但检测的假阳性率相对较高。受限于计算力早期的神经网络通常只有几层，由于理论分析的难度大，训练方法又需要很多经验和技巧，这个时期浅层人工神经网络研究反而相对沉寂。

自 2006 年 hiton 等^[26] 提出深度学习以来，方兴未艾的深度学习真正改变了计算机视觉之前的定义，它正在成为计算机视觉领域机器学习工具最优的选择之一。特别是卷积神经网络已被证明是众多计算机视觉任务的强大工具。世界各地的医学图像分析团队正在迅速进入该领域，并将其和其他深度学习方法应用于各种各样的应用，令人鼓舞的结果正不断涌现，如图1.6，从左上到右下依次为：乳房 X 线图像质量分类^[27]，大脑病灶分割（^[28]），气道树分割中的泄漏检测^[29]，视网膜病变分类（Kaggle 糖尿病视网膜病变图像）^[30]，前列腺分割，结节分型，淋巴结中的乳腺癌转移检测，皮肤病变分类^[31] 和来自 Yang 等人^[32] 的 X 射线骨骼抑制技术。文献^[33–35] 对深度学习应用于医学图像分析做了专门综述。

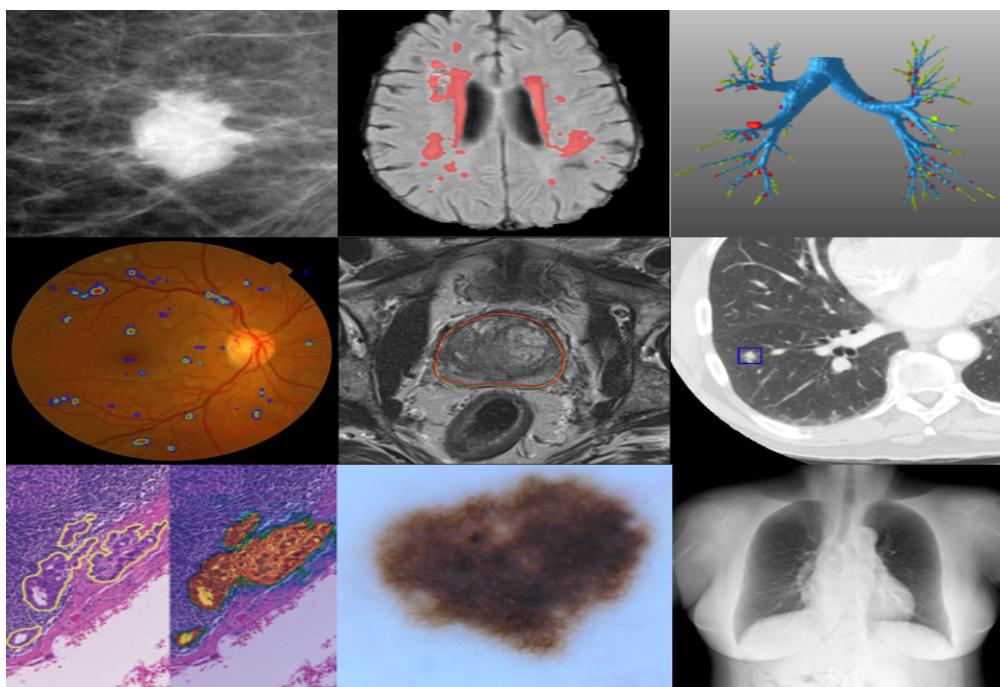


Figure 1.6 应用深度学习的一些医学影像图集

1.2.3.1 分类识别

图像识别分类是深度学习对医学图像分析做出贡献的首要领域之一，尤其是对自然图像进行预训练的 CNN 已经显示出惊人效果，在一些任务中超过了人类。相关研究表明可以调整 CNN 以利用医学图像的内在结构，Chen 等^[36] 利用 CNN 结合领域知识，在胎儿超声心动图标准切面的自动识别问题中取得良好的识别效果。Bar 等^[37] 利用自然图像训练的模型对胸腔 X-射线图像进行特征提取并结合全局特征^[38] 得到最优检测结果。Margita 等^[39] 针对心脏核磁共振图像利用微调迁移从自然图像学习的模型。总之与自然图像分类相比，医学对象分类较多使用预先训练的网络模型，这主要是缺乏高质量标注的医学数据集，且当前卷积网络较难高效结合上下文或三维信息。

1.2.3.2 病变检测

CAD 在医学图像分析领域是一个非常成熟的任务，解剖对象定位（空间或时间），例如定位器官或标记点，一直是分割任务或制定治疗计划等临床工作流程中的重要预处理步骤。传统 CAD^[40] 通过监督方法或经典的图像处理技术（如滤波和数学形态学）检测候选区域，各阶段通常是分离的，并常由手工设计的特征来描述，分类器用于将特征向量映射到实际病变的概率。深度学习采用直接以一个以候选病灶为中心的图像数据区域进行操作，训练一个端到端的 CNN。如利用深度卷积网络进行显微镜图像中细胞检测^[41]、结合深度全卷积网络的 MRI 心室检测与分割^[42,43] 和超声图像解剖结构的检测^[44]。Roth 等人^[45] 使用 CNN 改进三个现有的 CAD 用于 CT 结肠镜检查的结肠息肉，体部 CT 上的硬化性脊柱转移以及体部 CT 上的肿大淋巴结的检测。医学成像中的定位通常需要解析 3D 体积，为了用深度学习算法解决 3D 数据解析，已经提出了几种将 3D 空间视为二维正交平面组合的方法，Setio 等人^[46] 检测在 3D 胸部 CT 扫描中的肺结节，使用不同 CNN 的组合来对每个候选区域进行分类。相关综述请参考文献^[47]。总之与分类识别问题类似，直接解决特定对象检测的问题，需处理如类不平衡、图像的有效像素或体素处理仍是个挑战。

1.2.3.3 分割和形状建模

医学图像中的器官和结构的分割允许定量分析与体积和形状有关的临床参数，它往往是计算机辅助检测的第一步。例如心脏或大脑分中，分割的任务通常被定义为识别组成感兴趣对象的轮廓或内部的体素集。分割是应用医学成像深度学习的论文中最常见的主题，因此也在深度学习方法学中广泛应用，包括

开发独特的基于 CNN 的分割网络结构以及应用循环神经网络（Recurrent Neural Networks, RNNs）进行分割和形状建模。

这些新型 CNN 架构中最代表性的是由 Ronneberger 等人^[48] 提出的 U-Net，将上采样和下采样层的组合，将它们与反卷积和反褶积层之间的所谓跳跃连接相结合，从而直接产生分割图。Milletari 等人^[49] 提出了一种称为 V-Net 的 U-Net 架构的三维变体。RNN 最近在分割任务中变得越来越流行。例如，Xie 等人^[50] 使用空间 RNN 在组织病理学图像中分割肌束膜。Stollenga 等人^[51] 首次在 6 个方向上使用带有卷积层的 3D RNNs。此外，基于深度学习的特征点定位方法^[52] 在形状建模方面也取得令人瞩目的结果。总而言之，医学影像的各个细分领域中深度学习相关方法都大量涌入。但一般需直接针对特定分割任务自定义网络结构，相关综述请参考文献^[53]。

1.2.3.4 新颖的应用和其他医学分析任务

在医学图像的配准、基于内容的医学图像检索、图像重建生成和对比增强、融合图像和文本报告、基因组学和医药领域均能找到深度学习的身影，相关综述文献请参考^[33]。如 Kallenberg 等^[54] 利用无监督的特征学习用于乳房 X 线照相术，解决乳房密度分割和对乳房 X 线照片纹理进行风险评分。Yan 等^[55] 设计了一个多阶段的深度学习回归框架用于图像分类并应用于身体部位识别，它通过多学科的深度学习自动发现了不合理和无信息的局部斑块。苗等人^[56] 提出了一种 CNN 回归方法，用于实时 2-D 和 3-D 配准。Golkov 等人^[57] 提供应用 DL 将扩散 MRI 数据处理减少到单个优化步骤，阐明通过使用 DNN 在逐个体素的基础上的微观结构预测以及健康和损伤组织自动化的无模型分割，展示如何通过深度学习来简化经典数据处理。

1.2.3.5 相关软硬件平台

GPU 和 GPU 计算库（CUDA, OpenCL）的广泛应用促成了深度学习急剧上升的主要原因之一。GPU 是高度并行计算引擎，其执行线程数量比中央处理器（CPU）多一个数量级。使用目前的硬件，深度学习 GPU 的速度通常比 CPU 快 10 到 30 倍。FPGA 和定制的深度学习芯片也获得了相当的关注。

除硬件之外，深度学习方法普及的另一推动力是开源软件库和共享深度模型的广泛可用性。这些库提供了神经网络中重要操作（如卷积、循环网络）的高效 GPU 实现方便用户使用。当前流行的软件包是（按字母顺序）：

- **Caffe**^[58]. 提供 C++ 和 Python 接口，由加州大学伯克利分校的研究生开发。

- **Tensorflow**^[59]. 由 Google 开发提供 C++, Python 和接口.
- **Theano**^[60]. 提供由蒙特利尔 MILA 实验室开发的 Python 库。
- **Mxnet**^[61]. 提供一个 python 和 C++ 接口，是亚马逊的云平台默认深度学习引擎

1.2.3.6 趋势及挑战

深度学习已经渗透到医学影像分析的各个方面，深度学习在医学图像分析中的应用文献首先出现在各种研讨会和会议上 (MICCAI, SPIE, ISBI 和 EMBC 等)，然后在期刊中出现。该主题现已在主要会议和医学影像分析期刊中占据主导地位。得益于深度学习算法的突破以及现代 GPU 高效并行计算能力的提升，引起了巨大的商业兴趣。通用电气，西门子和飞利浦都将深度学习算法整合到成像设备和图像处理系统中以加速自动化定位分割解剖结构，以大大加快工作流程并消除了互操作性差异的问题，或执行自动量化。诸多创业医学影像公司，比如美国的 Enlitic 和国内的 DeepCare，同样使用 AI 快速筛选海量大数据，或者提供即时的临床决策支持。

将深度学习算法应用于医学图像分析面临许多独特的挑战。针对数据的主要挑战不仅是影像数据本身（数据维度和隐私等问题）的可用性，还要获取这些影像的相关注释/标签，也需要考虑通过众包来利用非专家标签减少对专家经验的需求；训练深度学习系统需要仔细考虑如何处理图像中的噪声和不确定性，类别不平衡问题特别严重；能否从一般图像向相关医学领域迁移学习；另一个挑战是平衡深度学习网络中的特征的数量（通常是数千个）与临床特征的数量（通常只有少数）以防止临床特征被淹没；

多数文献都集中在有监督的深度学习方法上以实现分类，检测，分割和标注等。然而应用无监督深度学习的兴趣仍在，并最近又重新开始收到重视，主要是因为世界上可用的大多是未标记数据；并且人类学习在一定程度上以无监督的方式发生，远比监督学习方法高效，如何在不知道具体标签的情况下学会识别物体和结构，或者如何只需要非常有限的监督就可以将这些对象分类。

最后，深度学习方法缺乏理论分析，经常被认为是“黑匣子”。尤其在医学领域可解释性尤其重要，算法系统必须能够以某种方式进行自我评估。这是将深度学习方法应用于医学图像分析，加速临床医生和患者对深度学习应用的接受程度的关键所在。

1.3 创新点及全文结构

所在的研究组多年来与四川大学华西医院合作展开医学影像处理系统中关键技术的研究。博士期间在相关课题资助下，关注于深度学习和医学影像学，着重于机器学习及其在医学影像领域的作用。本文在深度学习理论层面讨论了深度可视化问题，报告了我们在基于深度学习架构的经食道超声心动图标准切面自动分类识别，对一般医学组织对象的定位检测和对超声心动图左心室分割三个主要部分进行研究。

主要研究内容及成果如下：

1) 基于深度特征表示的高层语义解决图像理解的高层任务，应用深度学习进行医学图像的分类识别，通过构建超声心动图标准切面数据库，提出了一种基于深度卷积神经网络的超声心动图标准切面自动识别方法，该算法针对网络全连接层占有模型大部分参数的缺点，引入空间金字塔均值池化替代全连接层，获得更多的空间结构信息，利用全局空间金字塔均值池化方法进行微调迁移学习，并大大减少模型参数、降低过拟合风险，通过类别显著性区域将类似注意力机制引入模型可视化过程，详尽分析了数据规模对模型分类精度的影响，并对模型的可解释性和有效性进行了分析。

2) 基于深度特征表示的可视化解决深度模型的理论分析，针对基于深度卷积神经网络的图像分类模型的可解释性问题，通过评估模型特征空间的潜在可表示性，提出一种用于改善理解模型特征空间的可视化方法。给定任何已训练的深度卷积网络模型，引入了通过激活最大化获得的图像可解释性的正则化方法，结合现有正则化方法提出空间金字塔分解方法，利用构建多层拉普拉斯金字塔主动提升目标图像特征空间的低频分量，结合多层次高斯金字塔调整其特征空间的高频分量得到较优可视化效果。并通过限制可视化区域，提出利用类别显著性激活图技术加以压制上下文无关信息，可进一步改善可视化效果。该模型有效克服了原有可视化方法中由于不能主动调整高低频分量等原因造成的可视化图像语义重复和低效率等问题。

3) 基于深度特征表示的结构语义解决图像内容理解的中层任务，针对自动检测医学图像中指定目标时存在的问题，提出了一种基于深度学习自动检测目标位置和估计对象姿态的算法。该算法基于区域深度卷积神经网络和目标结构的先验知识，采用区域生成候选框网络、感兴趣区域池化策略，引入包括分类损失、边框位置回归定位损失和像平面内朝向损失的多任务损失函数，近似优化一个端到端的有监督定位网络，能快速地对医学图像中目标自动定位，有效地为下

一步的分割和参数自动提取提供定位结果。并在超声心动图左心室检测中提出利用检测额外标记点：二尖瓣环、心内膜垫和心尖，能高效地对左心室朝向姿态进行估计。

4) 基于深度特征表示解决图像内容理解的底层任务，图像去噪和分割分别设计利用了全卷积网络，针对不同的具体任务设计损失函数，提出了一个有监督多层残差卷积网络框架，结合不同损失函数学习端到端映射变换应用于去噪和心室图像的分割；其中分割任务是设计了不同架构的单通道 CNNs 分割模型，与如今流行的分割方法比较，大大提高了分割与识别正确率。

1.4 论文的章节安排

全篇共八章，结构如下：

第一章绪论介绍了应用人工智能进行医学影像分析的研究背景及意义，对当前国内外的研究现状及挑战进行剖析，同时阐述了本论文的研究内容，列出了主要创新点，最后给出了整篇文章的章节安排。

第二章概述了用于医学影像分析的主要深度学习算法和描述了统计形状模型。在这章中简述了各种神经网络，分析网络结构、优化算法和训练中的关键技术；接着介绍了传统的医学图像分割算法中的统计形变模型。

第三章描述了本文的基于 CNNs 的超声心动图标准切面识别方法。该章首先描述了传统 CNNs 算法应用于医学图像中存在的不足；然后提出了一种综全局空间信息的卷积神经网络模型，并可视化分析了模型的有效性；实验结果表明本章方法优于传统 CNNs 模型。

第四章介绍了本文提出的空间金字塔分解的深度可视化算法。首先描述了传统深度可视化方法存在的问题；然后介绍了基于梯度更新的可视化方法；接着提出了利用空间金字塔分解主动调整高低频分量；最后实验验证了本文方法的鲁棒性。

第五章分析了本文提出的基于区域卷积神经网络的心室检测算法。首先，分析了检测算法的形式化定义，概述了物体检测的演进；之后列出了对候选区域生成网络的改进检测物体朝向；并结合带朝向的多任务损失函数进行困难样例挖掘；然后详细从 MRI 及超声二维数据全面验证本文检测方法的有效性和鲁棒性。

第六章分析了本文提出的基于 Encoder-Decoder 网络的去噪算法。首先，分析了本章算法的相关工作；之后列出了提出方法的主要内容；最后实验验证了提出方法的有效性；介绍了基于卷积神经网络的形状对齐算法。首先，指出初始位

置定位和特征点标注方法，即要解决定位检测问题；之后列出了如何结合卷积神经网络特征的 AAM 模型与 CLM 模型；然后详细介绍了基于不同特征的分割结果。

最后总结全文，并展望了未来的研究工作。

第2章 基本理论概述

为了深入了解训练神经网络的工作原理，本章首先介绍相关基本理论概念。从简单前馈网络的结构开始，然后给出使用利用数据的空间或时间属性的高级模型结构，以及无监督深度学习的模型结构，最后给出用于解决特定任务的网络结构和常用的关键技巧。

2.1 机器学习算法

机器学习算法可以分为监督和无监督学习，在监督学习中，输入特征 \mathbf{x} 和标签 y 组成的样例对构成训练集 $\mathcal{D} = \{\mathbf{x}, y\}_{n=1}^N$ ，分类任务中 y 通常表示实例的固定类别；在回归任务中， y 是具有连续值的向量。通过有监督的训练最优拟合训练集，寻找最优模型参数 Θ ，该模型参数依据损失函数 $L(y, \hat{y})$ 预测数据， \hat{y} 表示通过将数据 \mathbf{x} 送入模型函数 $f(\mathbf{x}; \Theta)$ 获得的输出。

无监督学习算法主要处理无标签的数据，需直接对数据进行建模学习结构性质，寻找数据潜在的子空间表示。常见例子是主成分分析和聚类方法，多种损失函数可以指导无监督训练，如重建损失 $L(\mathbf{x}, \hat{\mathbf{x}})$ ，通常对通过降维或加噪声的输入数据进行近似重建。

2.1.1 神经网络

前馈神经网络是一种通用函数近似器^[62]，它构成了大多数深度学习方法的基础，在模型的输出和模型本身之间没有反馈连接；当前馈神经网络被扩展成包含反馈连接时，它们被称为递归神经网络。前馈网络的目标是学习某个函数映射 f^* ，例如对于分类器， $y = f^*(\mathbf{x})$ 将输入 \mathbf{x} 映射到一个类别 y ；前馈网络定义了一个映射 $\mathbf{y} = \mathbf{f}(\mathbf{x}; \theta)$ ，学习优化参数 θ ，使它能够得到训练集最佳的函数近似拟合。图2.1中参数 $\Theta = \{\mathcal{W}, \mathcal{B}\}$ ，其中 \mathcal{W} 为权重， \mathcal{B} 为偏差。输入 \mathbf{x} 与模型参数的线性组合经过非线性变换为激活值 a 。非线性映射称激活函数 $\sigma(\cdot)$ ：

$$a = \sigma(\mathbf{w}^T \mathbf{x} + b). \quad (2.1)$$

传统神经网络中典型激活函数是 Sigmoid 函数和双曲正切函数。复合嵌套多个激活函数的有向无环图为多层神经网络，又称多层感知器（multi-layered perceptrons, MLP）：

$$f(\mathbf{x}; \Theta) = \sigma(\mathbf{W}^T \sigma(\mathbf{W}^T \dots \sigma(\mathbf{W}^T \mathbf{x} + b)) + b). \quad (2.2)$$

其中 \mathbf{W} 是一个包含 \mathbf{w}_k 的列矩阵，与输出中的第 k 个激活值相关联，输入层和输出层之间的层通常被称为“隐藏层”，网络中的每个隐藏层通常都是向量值的，隐藏层的维数决定了模型的宽度（width）。当神经网络包含多个隐藏层时，它通常被认为是一个“深层”神经网络，因此称为“深度学习”。在网络的最后的输出层，激活值通过 $softmax$ 函数映射到 $P(y|\mathbf{x}; \Theta)$ 上得类别分布概率，即

$$P(y|\mathbf{x}; \Theta) = softmax(\mathbf{x}; \Theta) = \frac{e^{\mathbf{w}_i^T \mathbf{x} + b_i}}{\sum_{k=1}^K e^{\mathbf{w}_k^T \mathbf{x} + b_k}}, \quad (2.3)$$

其中 \mathbf{w}_i 表示通向与类 i 相关联的最后节点的输出。图2.1中显示了 MLP 的示意图。

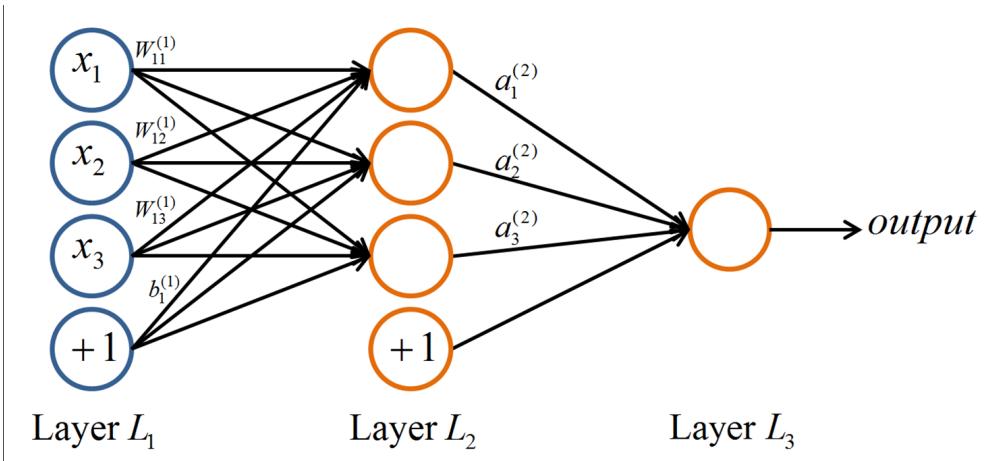


Figure 2.1 前馈神经网络（也称为多层感知器）结构示例示意图

由于神经网络的非线性导致大多数代价函数都变得非凸，需使用迭代的基于梯度的优化，基于极大似然估计的随机梯度下降方法是目前最常用优化的方法，它将参数 Θ 拟合到训练集 \mathcal{D} 。反向传播（back propagation）算法可以用来高效地计算复杂函数的梯度，随机梯度下降中一般使用批量数据的一小部分用于梯度更新，在实践中优化最大可能性等价于最小化负对数似然，它与训练数据和模型分布间的交叉熵等价：

$$\arg \min_{\Theta} - \sum_{n=1}^N \log [P(y_n|\mathbf{x}_n; \Theta)]. \quad (2.4)$$

其通常不直接优化我们感兴趣的目标，例如 ROC 曲线下的面积或用于分割的常用评估度量（例如 Dice 系数），该损失也称为交叉熵代价损失，。

长期以来，深度神经网络（Deep Neural Network, DNN）随着层数加深，优化函数越来越容易陷入局部最优解导致梯度弥散问题难以有效训练。从 2006 年

开始才重新受到欢迎^[26]，指出两种流行的无监督网络结构：堆叠自动编码器和深度置信网络，可以以无监督的方式逐层训练（预训练）DNN。但这些技术相当复杂需要大量手动调参才能产生令人满意的效果。

目前，最流行的模型是以有监督的方式进行端对端训练，通过引入预处理和新的激活函数，极大地简化了训练过程。最流行的网络结构是卷积神经网络和递归神经网络。尽管递归神经网络越来越受欢迎，但 CNN 目前在（医学）图像分析中应用最广泛。以下各节将简要介绍这些方法，从最受欢迎的方法开始，并讨论它们在应用于医疗问题时的差异和潜在的挑战。

2.1.2 卷积神经网络

卷积神经网络（CNN）是多层前馈神经网络的一种特例，是一种专门用来处理具有类似网格结构的数据的神经网络。例如时间序列数据（可以认为是在时间轴上有规律地采样形成的一维网格）和图像数据（可以看作是二维的像素网格）。较普通神经网络引入了感受野、卷积核滤波器组、卷积层、池化层等概念，其隐藏层的神经元设计成跟上一层神经元局部稀疏连接，并利用参数共享来减少模型复杂度。针对图像这种结构化数据，由不同卷积核来探测不同空间位置上的局部统计特征。通过堆叠多层的卷积结构，实现从低层到高层语义空间的抽象映射。MLP 和 CNN 之间有两个关键的区别：首先，网络中的权重以网络对图像

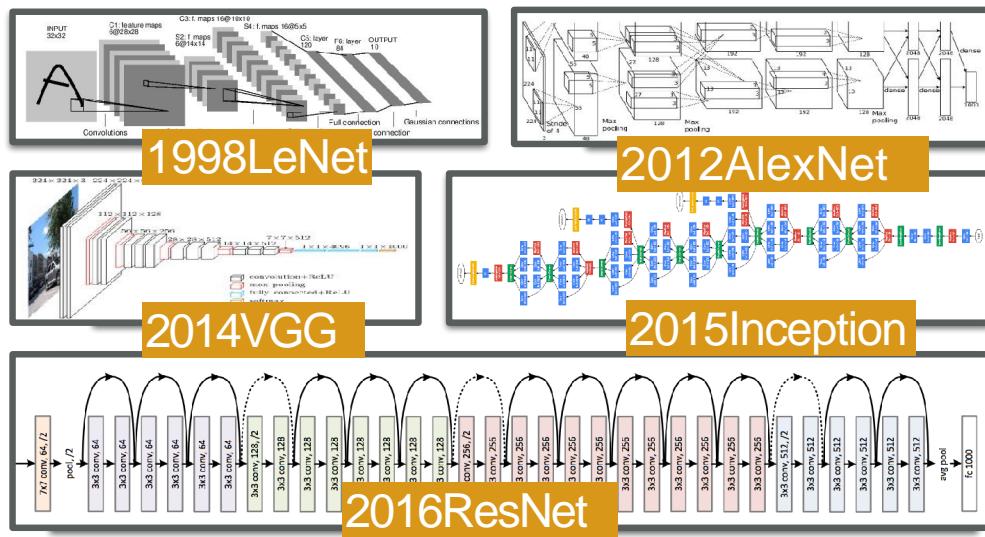


Figure 2.2 经典深度网络结构示意图

执行卷积操作的方式共享。这样，模型不需要为在图像中不同位置出现的同一对象分别检测单独的检测器，从而使网络在输入的平移时保持不变性。它还大大减少了需要学习的参数数量（即权重的数量不再取决于输入图像的大小）。图2.2中

显示了 LeNet-5^[63] 的经典网络结构。

在每一层，输入图像与一组 K 卷积核进行卷积： $\mathcal{W} = \{\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_K\}$ 并添加偏差项 $\mathcal{B} = \{b_1, \dots, b_K\}$ ，每个生成一个新的特征映射（feature map） \mathbf{X}_k 。这些特征受到非线性变换 $\sigma(\cdot)$ 的影响，并且对每个卷积层 l 重复相同的过程：

$$\mathbf{X}_k^l = \sigma(\mathbf{W}_k^{l-1} * \mathbf{X}^{l-1} + b_k^{l-1}). \quad (2.5)$$

其次，在于 CNN 中的池化层（Pooling），其中邻域的像素值使用置換不变函数（通常是最小或平均运算）进行聚合。这会导致一定量的平移不变性，并再次减少网络中的参数数量。在网络的卷积层结束时，通常会添加全连接的层（即常规的神经网络层），其不再共享权重。类似于 MLP，通过 softmax 函数概率归一化提供最后输出层中的激活值，并且使用方向传播优化极大似然损失对网络进行训练，从而产生类别概率分布。

2.2 深度卷积神经网络

鉴于 CNN 在医学图像分析中的流行，将详细介绍广泛使用的模型中最常见的网络结构及其差异，相关的方向的演进请参考图2.3。

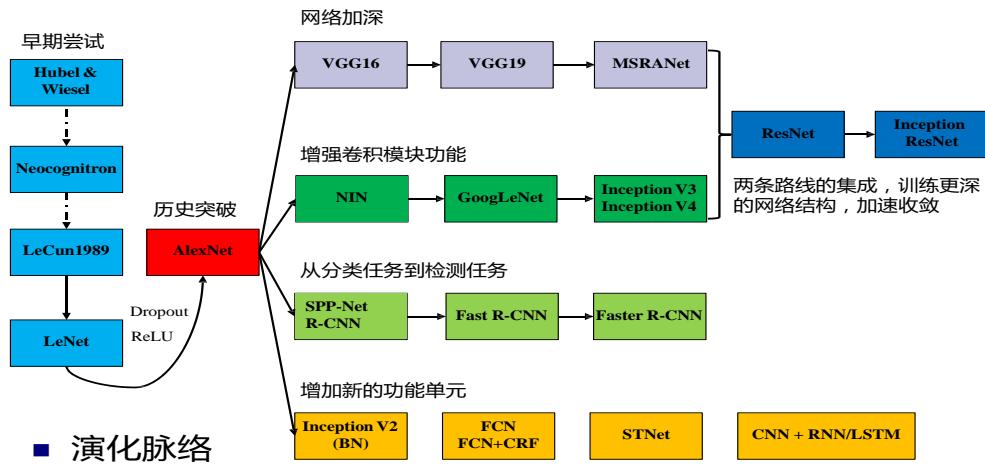


Figure 2.3 深度卷积神经网络结构演进图

2.2.1 通用分类框架

深度 CNN 的典型结构 AlexNet^[1] 是在 LeNet 模型^[63] 的基础上引入修正线性单元（Rectified Linear Units,ReLU）的激活函数和 Dropout 等技术^[1] 进行的改进。模型的激活函数没有采用 Sigmoid 函数或双曲正切函数，而是选择 ReLU 函数，目的是引入更多非线性来加速训练收敛速度，解决多层网络反向传播中梯度

弥散的问题。为了使得每层输入的分布更平稳，一般引入批量归一化层（Batch Normalization, BN)^[64]，采用最大池化层进行下采样，有时也把“卷积-激活-归一化-池化”统称为卷积层。最后需连接全连接层，全连接层就不再保存空间信息，是对低层特征的高层抽象，最终输出指定维度大小的向量，作为该图像的特征向量送入最终的分类器进行分类评估。2012年以来新的网络结构不断涌现，模型

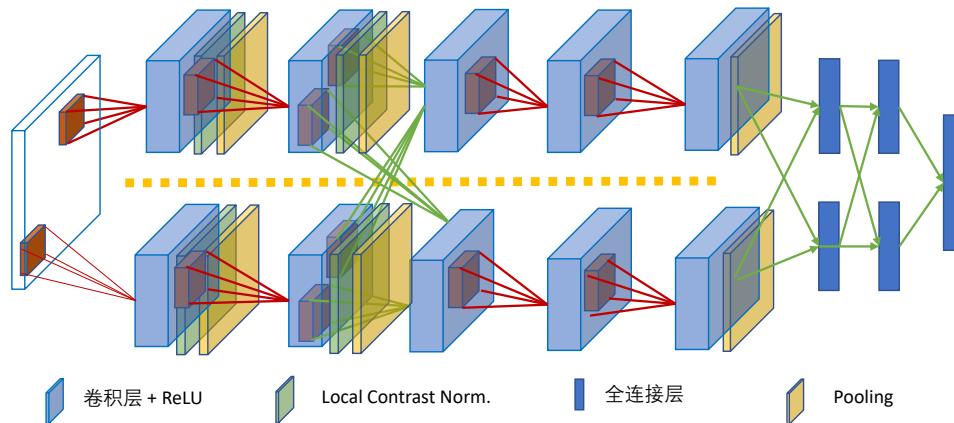


Figure 2.4 AlexNet 网络结构示意图^[1]

更倾向于更深更复杂的结构。堆叠更小尺寸的卷积核，可以用较少的参数来表示类似的函数，这些更深的网络结构在推断时通常具有较低的内存占用量，这使得它们能够部署在诸如智能手机的移动计算设备上。*VGG*^[18]是第一个探索更深层次的网络，网络的结构非常一致，从头到尾全部使用的是 3×3 的卷积和 2×2 的池化层。Szegedy 等^[19]提出了一个名为*GoogLeNet*的22层网络，在深层网络之上，引入了更复杂的构建模块，它使用使用密集连接结构用于估计稀疏 CNN，也称为*Inception*模块^[65]，该模块使用一组不同大小的卷积 $1 \times 1, 3 \times 3, 5 \times 5$ 的卷积取代了公式(2.5)；同时把顶部使用的全连接层替换成平均池化层，可以提高训练过程的效率，并再次减少参数的数量。*ResNet*网络结构^[20]赢得了2015年的ImageNet挑战，其引入跳跃连接，残差块不是直接学习函数，而是仅学习残差，并因此预先调整每一层中学习接近同等函数映射，这样可以有效地训练更深的网络模型，可以认为*ResNet*是不同长度的网络的指数集合。

自2017年以来，ImageNet基准测试的性能已经饱和^[17]，虽然仍有很多分类识别的工作进行网络结构的设计，但很难评估性能的小幅增长是否真的归因于“更好”和更复杂的架构。这些模型所提供的较低内存占用空间的优势通常对于医疗应用来说并不重要。因此，*AlexNet*或其他简单模型（如*VGG*、*ResNet*）仍然很受医学数据分析领域欢迎。

2.2.2 多通路的卷积神经网络结构

将深度学习技术应用于医疗领域的挑战通常在于将现有网络结构适应于例如不同输入格式，例如三维数据。在 CNN 早期应用于这样的体积数据时，通过将感兴趣体积（VOI）划分为切片，将全部 3D 卷积和由此产生的大量参数分开，所述切片作为不同的流馈送到网络。文献 [66] 是第一个将这种方法用于膝关节软骨分割的。类似地，网络可以以多通道方式从 3D 空间中嵌入多个角度的贴片^[67]，这些方法也被称为 2.5D 分类。

根据不同任务需要、不同融合方式可以得出多通路的网络结构，多个输入 CNN 网络的特征图可以在网络的任何处合并融合。如双通道架构^[68] 可应用于多尺度图像分析；在图像检测任务中，为了检测指定对象区域，上下文往往是一个重要的提示，增加上下文最直接的方法是将更大的区域块提供给网络，但这会显著增加网络的参数和内存需求量。因此除了提高分辨率获得局部信息之外，还可引入多通道多尺度网络结构^[69]，一些医学应用也成功地使用该概念^[32,68,70,71]。

2.2.3 全卷积网络结构

分割和去躁是自然和医学图像分析中的一项常见任务，为了解决这个问题，CNN 可分别对图像中的每个像素进行分类，需要在特定像素周围提取区域块。该“滑动窗口”方法的缺点是来自相邻像素的输入块具有巨大的重叠并且多次计算相同的卷积操作，卷积和点积都是线性算子，因此内积可以写成卷积，反之亦然，通过将全连接的层重写为卷积，CNN 可以输入大于其被训练的图像尺寸的任意图像，并生成概率图。但由于池化合并图层，这可能导致输出的分辨率远远低于输入，反卷积^[72] 是为防止这种分辨率下降而提出上采样方法之一，如图2.5。通过将结果拼接在一起，减去由于“有效”卷积而丢失的像素，可以获得最终输出的全分辨率完整输出。将全卷积网络进一步改进提出 U-net 架构^[48]，其包含一个常规的全卷积结构，后面跟着一个上采样部分，其中反卷积用于增加图像大小，通过收缩编码-膨胀解码（Encoder-Decoder），将它与跳跃连接^[20] 相结合，提出 Encoder-Decoder 卷积层。3D 数据也适用类似的方法，如用于体数据的 V-Net 结构，提出包含类似 ResNet 的残余块和随机丢失（Drop）层，损失函数也不是传统的交叉熵，而是直接最小化常用的分割误差测量损失^[49]。

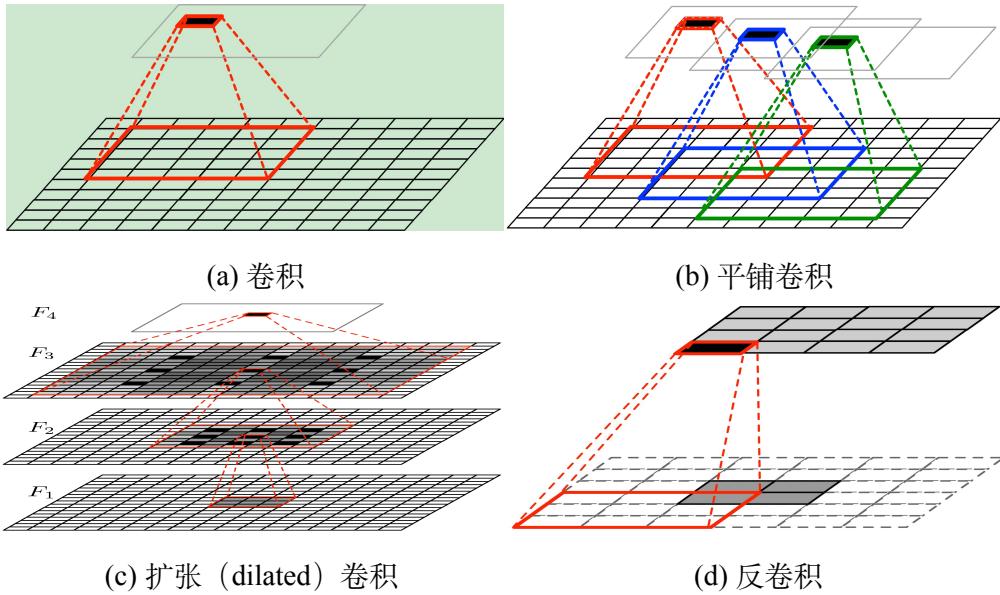


Figure 2.5 卷积网络中的各种卷积操作图

2.3 循环神经网络

传统上 RNN 是为处理序列数据而开发的，可以看作是 MLP 的泛化，输入和输出的长度都不相同，这使得它们适用于诸如机器翻译、语音识别这样的任务，其中源语言和目标语言的句子是输入和输出。在分类设置中，模型学习一个可变长度序列 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T$ ，而不是一个输入向量 \mathbf{x} ，给出的类别概率 $P(y|\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T; \Theta)$ 。

普通的 RNN 在 t 时刻维持一个隐式或隐藏状态 \mathbf{h} ，它是依据其输入 \mathbf{x}_t 和前一状态 \mathbf{h}_{t-1} ：

$$\mathbf{h}_t = \sigma(\mathbf{W}\mathbf{x}_t + \mathbf{R}\mathbf{h}_{t-1} + \mathbf{b}), \quad (2.6)$$

其中加权矩阵 \mathbf{W} 和 \mathbf{R} 是随时间共享的。对于分类，通常会添加一个或多个全连接的层，然后添加 softmax 以将序列映射到类别概率：

$$P(y|\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T; \Theta) = \text{softmax}(\mathbf{h}_T; \mathbf{W}_{out}, \mathbf{b}_{out}). \quad (2.7)$$

由于梯度需要通过时间反向传播 (back-propagation through time, BPTT)，因此 RNN 具有固有的深度 (及时性)，因此与常规深层神经网络一样遭受梯度弥散或爆炸难以训练问题。为此已经设计几种专用单元来解决上下文依赖问题，最早和最流行的是长期短期记忆 (Long Short-Term Memory, LSTM) 单元^[73]，而门控复发单元 (Gated Recurrent Unit, GRU)^[74] 是 LSTM 的最新简化版。

尽管最初提出是针对一维输入，但 RNN 越来越多地应用于图像，在自然图

像中，基于像素的“pixel RNN”被用作自回归生成模型，最终可以生成类似于训练集样本的新图像。对于医疗应用，它们已被用于分割问题，并且在各种竞赛挑战中得到令人鼓舞的结果^[75]。

2.4 无监督深度学习

2.4.1 自编码器

自编码器 (Auto-encoders, AE) 是简单神经网络的一种，通过一个隐藏层 \mathbf{h} 来产生编码 (code) \mathbf{x}' 近似表示输入 \mathbf{x} 。重建其隐藏层 $\mathbf{W}_{h,x'}$ 和相应的偏差 $b_{h,x'}$ ，它们由权重矩阵和输入到隐藏状态和 $\mathbf{W}_{x,h}$ 的偏置 $b_{x,h}$ 来控制。非线性函数用于计算隐藏激活：

$$\mathbf{h} = \sigma(\mathbf{W}_{x,h}\mathbf{x} + \mathbf{b}_{x,h}). \quad (2.8)$$

此外，隐藏层 $|\mathbf{h}|$ 的维度小于 $|\mathbf{x}|$ 。这样，数据被投影到表示输入中的主要潜在结构的较低维度子空间上。如果隐藏层的大小与输入大小相同，并且不再添加非线性激活函数，则模型将简单地学习同等标识函数，正则化或稀疏性约束可以用来改善提升模型训练过程。除了向代价函数增加一个惩罚项，我们也可以通过改变重构误差项来获得一个能学到有用信息的自编码器。去噪自动编码器^[76]是另一种防止模型学习同等标识函数的解决方案。这里模型被训练来重构来自带噪声（通常是椒盐噪声）的输入。相关医疗应用中，自动编码器层经常用（“贪婪算法”）单独训练，通过将自动编码器层放置在彼此之上而形成深度自编码器，然后使用监督训练对整个网络进行微调以进行预测。

2.4.2 变分自编码器和深度生成对抗网络

最近，引入了两种新颖的无监督网络结构：变分自动编码器 (VAE)^[77] 和生成对抗网络 (GAN)^[78]。因很难解决它们配分函数的数值计算而成为众所周知的难题，目前还没发现将这些方法应用于医学图像，但在自然图像中的应用还是值得期待的。

2.4.3 受限玻兹曼机和深度信念网络

受限玻尔兹曼机 (Restricted Boltzmann Machines, RBMs)^[26] 是一种基于能量模型的深度概率模型，由输入层或可见层 $\mathbf{x} = (x_1, x_2, \dots, x_N)$ 和一个带有潜在特征表示的隐藏层 $\mathbf{h} = (h_1, h_2, \dots, h_M)$ 构成。节点之间的连接是双向的，因此给定输入向量 \mathbf{x} 可以获得潜在特征表示 \mathbf{h} ，反之亦然。因此，RBM 是一个生成模

型，我们可以从中进行抽样并生成新的数据点。能量函数被定义为输入和隐藏单位的特定状态 (\mathbf{x}, \mathbf{h}) :

$$E(\mathbf{x}, \mathbf{h}) = \mathbf{h}^T \mathbf{W} \mathbf{x} - \mathbf{c}^T \mathbf{x} - \mathbf{b}^T \mathbf{h}, \quad (2.9)$$

与 \mathbf{c} 和 \mathbf{b} 偏差。系统“状态”的概率通过将能量传递给归一化指数来定义联合概率分布:

$$p(\mathbf{x}, \mathbf{h}) = \frac{1}{Z} \exp\{-E(\mathbf{x}, \mathbf{h})\}. \quad (2.10)$$

计算配分函数 Z 朴素方法（对所有状态进行穷举求和）通常是棘手的。然而，以 \mathbf{h} 为条件计算 \mathbf{v} 形式的条件分布会得出一个简单公式:

$$P(h_j | \mathbf{x}) = \frac{1}{1 + \exp\{-b_j - \mathbf{W}_j \mathbf{x}\}}. \quad (2.11)$$

由于网络是对称的，类似的表达式适用 $P(x_i | \mathbf{h})$.

DBNs^[79] 第一批成功应用深度架构训练的非卷积模型之一，实质上是 AE 层被 RBM 取代的堆叠多层 AE。再次，以无人监督的方式进行单个层次的训练。通过向 DBN 顶层添加线性分类器并执行监督优化来执行最终的微调。

2.5 深度学习相关技术

在本节中，我们将从激活函数，损失函数，正则化和优化四个方面描述 CNN 的主要改进方向。

2.5.1 激活函数

合适选择激活函数能显着提高了 CNN 对于某个任务的性能。在本节中，我们将介绍 CNN 中几种激活函数。

2.5.1.1 ReLU

ReLU 激活函数定义如下:

$$a_{i,j,k} = \max(0, z_{i,j,k}) \quad (2.12)$$

其中 $z_{i,j,k}$ 是位置 (i, j) 第 k 个通道的输出，ReLU 是一个分段饱和线性函数，它将负数部分修剪为零，并保留正数部分（参见图2.6a）。ReLU 的取 max 操作的计算速度比 sigmoid 或 tanh 激活函数快得多，并且它还可以诱导隐藏单元的稀疏性并允许网络获得稀疏表示。文献表明，甚至在没有预训练的情况下，使用 ReLU 可以有效地训练深度网络^[1]。

2.5.1.2 带泄漏的 ReLU

ReLU 单元的一个潜在缺点是只要单元不激活，它就具有零梯度。基于梯度的优化不会调整它们的权重，这可能会导致最初没有激活的单位以后永不会激活。此外，由于恒定的零梯度，它可能会减慢训练过程。为了缓解这个问题，Mass 等^[80]引入了带泄漏（Leaky）的 ReLU（LReLU），其定义如下：

$$a_{i,j,k} = \max(0, z_{i,j,k}) + \lambda \min(0, z_{i,j,k}) \quad (2.13)$$

其中 λ 是 $(0, 1)$ 范围内的预定义参数。与 ReLU 相比，LReLU 使得具有负向值而不是将其映射到常量零点，这使得当单元不活动时允许一个很小的非零梯度。

2.5.1.3 带参数的整流线性单元

不同于在 LReLU 中使用预定义的参数，而是在等式 (2.14) 中给出带参数的整流线性单元 (PReLU)，其自适应学习参数以提高准确性。在数学上，PReLU 函数定义为

$$a_{i,j,k} = \max(0, z_{i,j,k}) + \lambda_k \min(0, z_{i,j,k}) \quad (2.14)$$

其中 λ_k 是第 k 通道的待学习参数。由于 PReLU 仅引入极少量的额外参数，例如，额外的参数数量与整个网络的通道数量相同，因此不会出现过度拟合的额外风险，额外的计算成本可以忽略不计。它也可以通过反向传播与其他参数同时训练。

2.5.1.4 随机整流线性单元

LReLU 的另一个变体是随机整流线性单元 (RReLU)^[81]，在 RReLU 中负值部分的参数从均匀分布中随机抽样，然后在测试中使用固定值（参见图2.6c）。形式上 RReLU 函数定义如下：

$$a_{i,j,k}^{(n)} = \max(0, z_{i,j,k}^{(n)}) + \lambda_k^{(n)} \min(0, z_{i,j,k}^{(n)}) \quad (2.15)$$

其中 $z_{i,j,k}^{(n)}$ 是位置 (i, j) 上第 n 个样例的第 k 个通道的输出， $\lambda_k^{(n)}$ 表示其对应的采样参数， $a_{i,j,k}^{(n)}$ 表示其相应的输出。由于其随机性，它可以减少过度拟合。Xu 等^[81]也对标准图像分类任务中的 ReLU，LReLU，PReLU 和 RReLU 进行评估，并得出结论：在整流激活单元中加入一个非零斜率用于负值部分，可以持续改善性能。

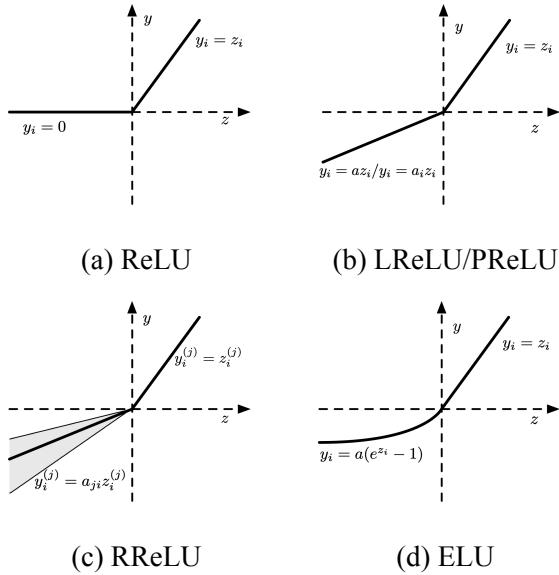


Figure 2.6 ReLU, LReLU, PReLU, RReLU 和 ELU 之间的比较, 参数 λ 对于 Leaky ReLU, 是凭经验预定义的。对于 PReLU 是从训练数据中学习的。对于 RReLU 是一个随机变量, 它是从训练中给定的均匀分布中抽取的, 并在测试中保持固定。对于 ELU 是凭经验预定义的。

2.5.1.5 指数线性修正单元

Clevert 等^[82]引入了指数线性单元 (Exponential Linear Unit, ELU), 它可以更快地学习深度神经网络, 并提高分类精度。像 ReLU、LReLU、PReLU 和 RReLU 一样, ELU 通过将正部分设置为线性变换来避免渐消梯度问题。与同样具有负值部分的 LReLU, PReLU 和 RReLU 相比, ELU 使用饱和度函数作为对噪声更鲁棒的负值部分。ELU 的函数定义如下:

$$a_{i,j,k} = \max(0, z_{i,j,k}) + \min(0, \lambda(e^{z_{i,j,k}} - 1)) \quad (2.16)$$

其中 λ 是用于控制 ELU 为负输入饱和的值的预定义参数。

2.5.1.6 Maxout

Maxout^[78]是一种替代性的非线性函数, 能够近似任意连续函数, 可以在每个空间位置获得跨多个通道的最大响应, 函数定义如下: $a_{i,j,k} = \max_{k \in [1,K]} z_{i,j,k}$, 其中 $z_{i,j,k}$ 是第 i 通道的特征映射。由于 ReLU 实际上是 maxout 的一个特例, Maxout 享有 ReLU 的所有好处, 但参数会相应加倍。

2.5.2 损失函数

深度神经网络设计中的一个重要方面是损失函数的选择，损失函数是模型对数据拟合程度的反映，为特定任务选择合适的损失函数非常重要。我们在本小节中介绍三个有代表性的部分：softmax 损失，铰链损失和对比损失。

2.5.2.1 Softmax 损失

Softmax 损失结合 softmax 函数和交叉熵损失，是多项式逻辑损失和 softmax 的组合。给定训练 $\{(x^{(i)}, y^{(i)}); i \in 1, \dots, N, y^{(i)} \in 0, \dots, K - 1\}$ ，其中 $x^{(i)}$ 是 i 输入图像区域， $y^{(i)}$ 是它的类标签。第 i 输入的第 j 类的预测 $a_j^{(i)}$ 用以下 softmax 函数进行变换：

$$p_j^i = e^{a_j^{(i)}} / \sum_{l=0}^{K-1} e^{a_l^{(i)}} \quad (2.17)$$

Softmax 将预测变成非负值，并将它们归一化以获得类别概率分布。这种概率预测用于计算多元逻辑损失，如下所示：

$$\mathcal{L}_{softmax} = -\frac{1}{N} \left[\sum_{i=1}^N \sum_{j=0}^{K-1} 1\{y^{(i)} = j\} \log p_j^i \right] \quad (2.18)$$

2.5.2.2 铰链损失

铰链 (Hinge) 损失通常用于训练最大间隔的分类器，如支持向量机 (Support Vector Machine, SVM)。多类 SVM 的铰链损失函数由公式 (2.19) 定义，其中 x_n 是给定的特征向量，而 $\ell_n \in [0, 1, 2, \dots, K - 1]$ 指示 K 类中的正确类标签。

$$\begin{aligned} \mathcal{L}_{Hinge} &= \frac{1}{N} \sum_{n=1}^N \sum_{k=0}^{K-1} [\max(0, 1 - \delta(\ell_n, k) w^T x_n)]^p \\ \delta(\ell_n, k) &= \begin{cases} 1, & \text{if } \ell_n = k \\ -1, & \text{if } \ell_n \neq k \end{cases} \end{aligned} \quad (2.19)$$

注意如果 $p = 1$ ，公式 (2.19) 是 L_1 损失，而如果 $p = 2$ ，则是平方铰链损失 L_2 。

2.5.2.3 对比损失

对比 (Contrastive) 损失^[83] 常用于训练孪生多通路网络，属于弱监督方法，用于从标记为匹配或不匹配的数据实例对中学习相似性度量。给定第 i 对数据

$(x_\alpha^{(i)}, x_\beta^{(i)})$, 令 $(z_\alpha^{(i,l)}, z_\beta^{(i,l)})$ 表示其对应的第 l 个 ($l \in [1, \dots, L]$) 层的输出对。图像对通过两个相同的 CNN 分别传递，并将最终图层的特征向量送入到损失函数，对比损失 \mathcal{L}_l 被定义为：

$$\mathcal{L}_{contrastive} = \frac{1}{2N} \sum_{n=1}^N \sum_{l=1}^L (y) d^{(i,L)} + (1 - y) \max(m - d^{(i,L)}, 0) \quad (2.20)$$

其中 $d^{(i,L)} = \|z_\alpha^{(i,l)} - z_\beta^{(i,l)}\|_2^2$, 是 $z_\alpha^{(i,l)}$ 和 $z_\beta^{(i,l)}$ 之间的相似度, m 是影响非匹配对的边界参数。如果 $(x_\alpha^{(i)}, x_\beta^{(i)})$ 是匹配的对, 那么 $y = 1$ 。否则, $y = 0$ 。这种损失函数也被称为单边界参数损失函数。Lin 等^[84] 利用这种单边界损失函数在所有对上对网络进行训练时, 会导致检索结果急剧下降。同时, 仅在非匹配对上进行内部调整时, 性能得到更好的保留。这表明处理丢失函数中的匹配对是造成丢失的原因。虽然非匹配对的召回率是稳定的, 但处理匹配对是召回率下降的主要原因。为了解决这个问题提出了一个双边界损失函数, 它为匹配对添加了另一个边界参数。定义为：

$$\mathcal{L}_{doublecontrastive} = \frac{1}{2N} \sum_{n=1}^N \sum_{l=1}^L (y) \max(d^{(i,L)} - m_1, 0) + (1 - y) \max(m_2 - d^{(i,L)}, 0) \quad (2.21)$$

2.5.3 优化方法

在本小节中, 我们将讨论优化 CNN 训练的一些关键技术, 包括权重的初始化、随机梯度下降、批量归一化等。

2.5.3.1 权重初始化

通常深度 CNN 模型具有大量的参数并且损失函数是非凸的, 导致难以训练。为了实现训练中的快速收敛, 正确的初始化是最重要的先决条件之一。偏置参数可以初始化为零, 而权重参数的设置需打破同一层隐藏单元之间的对称性。例如, 如果我们简单地将所有权重初始化为相同的值, 例如 0 或 1, 那么同一层的每个隐藏单元将得到完全相同的信号。有些启发式方法可用于选择权重的初始大小, 最常用的初始化方法是根据高斯或均匀分布随机设置权重。Glorot 和 Bengio^[85] 提出初始化 m 个输入 n 个输出可根据零均值和特定方差的分布设置权重: $Var(W_{i,j}) \sqrt{6}/\sqrt{m+n}$, 也被称为“Xavier”。但是初始化时强加的性质可能在学习开始进行后不能保持; 可能成功提高了优化速度, 但意外地增大了泛化误

差，这些初始化方法往往不会带来最佳效果。在实践中，我们通常需要将权重范围视为超参数进行调参估计。

2.5.3.2 随机梯度下降

一般使用批梯度下降 (stochastic gradient descent, SGD) 来更新模型参数，采用反向传播算法计算损失函数的梯度。梯度下降算法将目标 $J(\theta)$ 的参数 θ 更新为 $\theta_{t+1} = \theta_t - \alpha \nabla_{\theta} E[J(\theta_t)]$ ，其中 $E[J(\theta_t)]$ 是整个训练集上 $J(\theta)$ 的期望值， α 是学习率。随机梯度下降根据单个随机选取的示例 $(x^{(t)}, y^{(t)})$ 进行梯度更新：

$$\theta_{t+1} = \theta_t - \mu_t \nabla_{\theta} J(\theta_t; x^{(t)}, y^{(t)}) \quad (2.22)$$

实际上，SGD 中的每个参数更新都是针对小批量数据（mini-batch）而不是单个示例，这可以帮助减少参数更新中的变化并且可以导致更稳定的收敛性。收敛速度由学习率 μ_t 控制。在面对小而连续的梯度但是含有很多噪声的时候，带动量 (Momentum) SGD 算法可以很好的加速学习为了加速训练使当前更新梯度取决于历史梯度在相关方向上累积速度向量信息，经典的带动量 SGD 更新由下式给出：

$$\begin{aligned} v_{t+1} &= \theta_t - \mu_t \nabla_{\theta} J(\theta_t; x^{(t)}, y^{(t)}) \\ \theta_{t+1} &= \theta_t + v_{t+1} \end{aligned} \quad (2.23)$$

其中 v_{t+1} 是当前速度矢量，是通常设置动量衰减参数 μ 为 0.9。在梯度下降优化中使用动量的另一种方式 Nesterov^[1]：

$$v_{t+1} = \gamma v_t - \mu_t \nabla_{\theta} J(\theta_t; x^{(t)}, y^{(t)}) \quad (2.24)$$

2.24首先计算当前梯度，然后沿更新的累积梯度方向移动，与经典动量2.23相比，Nesterov 首先沿先前累积梯度的方向移动 v_t ，计算梯度，然后进行梯度更新。这种预期的更新可以防止优化移动太快，并获得更好的性能。一种常用的方法是使用小的恒定学习速率，在初始阶段给出稳定的收敛，然后随着收敛速度的减慢而降低学习速率。除了这种人工手动设置学习率的方法之外，还有很多自适应学习率的方法，如 Adam^[86] 利用梯度的一阶矩估计和二阶矩估计动态调整每个参数的学习率。此外，并行化的 SGD 方法能够提高 SGD 以适合并行的大规模机器学习。

2.5.3.3 批量归一化

批量归一化 (Batch Normalization)^[64] 的提出旨在加速深度神经网络的整个训练过程。对输入减均值的预处理，是加快训练收敛过程的默认操作。他们认为

内部协变量偏移，即深度网络内部节点分布的变化将会减慢网络训练的速度。为此提出了一种称为批量归一化的有效方法来部分缓解这种现象。它通过归一化数据使其服从标准高斯分布来修复了各层输入的方式和差异。批量归一化可以理解为在网络的每一层之前都做预处理，只是这种操作以另一种方式与网络集成在了一起。除了加速训练外，批量归一化还使我们能够使用更高的学习速率而不存在发散风险，并且可以通过防止网络陷入饱和模式来使用饱和非线性。

2.5.4 正则化方法

过拟合在深层 CNN 中是一个不容忽视的问题，可以通过正则化来有效降低过度拟合问题。在下面的小节中，我们介绍一些有效的正则化技术。

2.5.4.1 p 范数正则化

正则化通过增加惩罚模型复杂性的额外项来修改目标函数，若损失函数为 $L(\theta; x; y)$ ，则正则化损失将为：

$$E(\theta; x; y) = L(\theta; x; y) + \lambda R(\theta) \quad (2.25)$$

其中 $R(\theta)$ 是正则化项，并且 λ 是正则化项参数。 p -范数正则化函数通常用作 $R(\theta) = \sum_j \|\theta_j\|_p^p$ ，当 $p = 1$ ， p -范数是凸的，这使得目标函数更容易优化，并使这个函数具有稀疏性，使得参数变小倾向接近零。对于 $p = 2$ ，“2-范数正则化”通常称为“权重衰减”，使得参数更平均；当 $p < 1$ 时， p -范数正则化更多地利用了权重的稀疏效应，但却导致了非凸函数。

2.5.4.2 随机失活

随机失活 (Dropout)^[87] 提供了一大类模型的正则化方法，其可以被认为是以超参数 p 的概率被激活或者被设置为 0。在训练过程中，随机失活可以被认为是对完整的神经网络抽样出一些子集，每次基于输入数据只更新子网络的参数（然而，数量巨大的子网络们并不是相互独立的，因为它们都共享参数）。在测试过程中不使用随机失活，可以理解为是对数量巨大的子网络们做了模型集成，以此来计算出一个平均的预测。

研究指出在使用费舍尔信息矩阵 (fisher information matrix) 的对角逆矩阵的期望对特征进行数值范围调整后，再进行 L2 正则化这一操作，与随机失活正则化

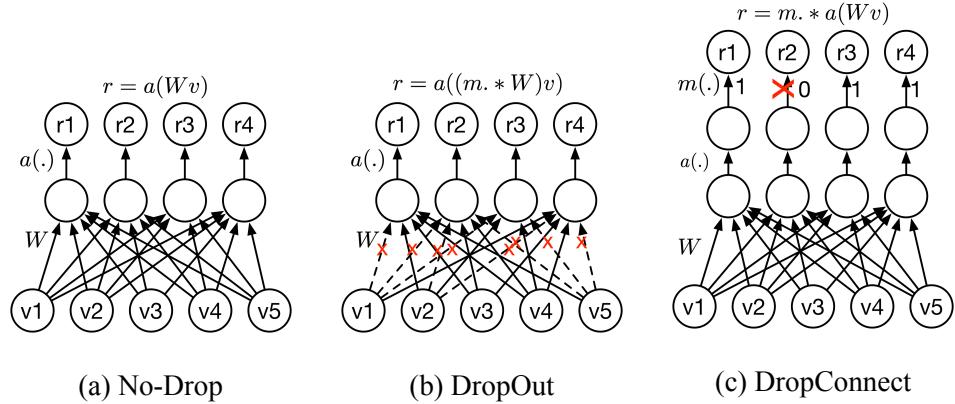


Figure 2.7 无 Dropout 网络，DropOut 网络和 DropConnect 网络的图示。

是一阶等价的。而 DropConnect^[88] 将 Dropout 的思想更进一步改进，DropConnect 不是将神经元的输出设置为零，而是以概率 p 选择完全连接层的权重 W 设置为零。图2.7说明了无 Dropout，Dropout 和 DropConnect 网络之间的差异。

2.6 小结与讨论

本章首先对学习算法和神经网络基本方法进行了梳理，针对卷积神经网络和深度卷积神经网络模型进行了分析，进一步对无监督深度学习进行了概述。同时对深度学习应用于医学图像中涉及的分类、检测和分割应用算法从原理上进行了分析。最后对深度学习中相关核心技术了描述，涉及激活函数、损失函数、优化方法和正则化。

第3章 特征表示的高层语义应用

3.1 超声心动图切面的自动识别方法

在心脏病常规临床检查中，二维实时超声心动图常用于评测心脏的结构和功能。临床超声检查通常主要包括三个步骤：探头扫描不同位置，选取标准切面和对标准切面的测量和诊断^[36]。其中，医师总结出来能更好辅助分析心脏功能结构的特定位置和角度的超声心动图称为标准切面，其正确快速选取不仅对临床诊断具有至关重要的意义，也为病例研究提供比较依据。标准切面的自动识别是超声心动图智能分析和测量的基础。与自然图像相比，医学超声成像质量差，存在斑点噪声和伪影；并且各标准切面类内、类间差异大，使得标准切面的识别成为一个非常具有挑战性的问题。

目前的研究主要集中在利用机器学习和图像处理等方法，进行超声心动图的自动识别、检索及切面内组织结构的定位和分割等。针对超声心动图的自动识别，2004年Shahram等^[89]首次提出采用马尔科夫随机场，设计通用腔室模板检测心脏腔室来辅助三类标准切面识别，但需额外信号来指定处于舒张末期(End-Diastolic, ED)的切面。同样利用处于ED的标准切面，Kevin等^[90]基于多类别提升算法框架，提取哈尔矩形特征训练弱分类器，同样需要检测心脏腔室的空间位置，辅助四类标准切面识别。基于降低特征维度的两层级联方法，把标准切面分类成心尖和胸骨旁两大类，然后进一步区分四类标准切面视频^[91,92]。在文献[89]工作基础上整合局部和全局模板特征，利用多类逻辑提升分类算法，并指出能扩展到任意标准切面^[93]。在对心脏的循环跳动的时空信息进行统计分析的基础上，利用主动外观模型对形状和纹理进行建模，统计追踪一个心动周期并投影到运动空间进行分类^[94]，该方法处理的视频序列。把标准切面视为不同场景，提取低层全局特征来表征不同切面，利用改进核支持向量机进行分类^[95]。这些方法可以归纳为两个阶段：首先根据先验人为设计特征来表征图像；然后利用机器学习中不同分类方法对特征向量进行建模分析得到分类器。然而受限于‘语义鸿沟’问题，根据特定先验人为设计特征，如大多数方法都针对心动周期的某个特定时刻的切面（如ED），会导致模型泛化性能差。

近来，深度卷积神经网络(Convolutional Neural Network,CNN)在大规模自然图像数据集(如ImageNet^[17])上，识别性能远超传统方法^[1]。主要得益于深度学习利用大量标注数据从图像原始像素出发，逐层分级学习中高层的抽象语义特

征^[96]。当前实践中由于深度学习需要大量的标注数据，所以仅在少数医学任务中取得有限的成功应用，且对深度模型的鲁棒性和有效性也缺乏详尽分析。Chen 等^[36] 利用 CNN 结合领域知识，在胎儿超声心动图标准切面的自动识别问题中取得良好的识别效果，但胎儿跟成人超声心动图差异大，具有很大特殊性。Bar 等^[37] 利用自然图像训练的模型对胸腔 X-射线图像进行特征提取并结合全局特征^[38] 得到最优检测结果，并没有对特定医学数据进行迁移训练，仅是作为特征提取器。Margita 等^[39] 针对心脏核磁共振图像利用微调迁移从自然图像学习的模型，但没对模型有效性进行分析。

目前深度 CNN 模型的理论分析工作还不是很完善，能自动学习语义特征的工作机理还是个“黑箱”。对于不同的模型的比较除了准确率之外并没有很好的评价方法，优异的泛化能力从何而来仍是个开放问题。一些工作^[97-100] 通过可视化各层激活值和卷积核来更好理解深度 CNN。对在给定数据集上训练得到的深度 CNN 网络模型，Simonyan 等^[97] 用反卷积来可视化每个神经元的最大激活值。Mahendran 等^[98] 通过对学习到的每层的特征编码进行反编码，建立每层特征编码和原图像的映射关系。Zeiler 等^[99] 试图通过梯度上升方法迭代寻找图像使得最大化激活某个或某些特定的神经元。神经元对图像每个像素的梯度描述了当前像素的怎样改变能影响分类结果。前三个方法均是对已训练的模型进行分析，而类激活映射图（Class Activation Maps,CAM）方法^[100] 用全局平均池化层代替全连接层改进训练过程，分类性能虽略有降低，但能指示出特定类别的显著性判别区域，能很好的解释模型的有效性。

本文提出一种基于深度 CNN 自动识别超声心动图标准切面的方法（Deep Echocardiogram,Deep-Echo）：1) 引入空间金字塔平均池化层代替全连接层，一方面大大减少模型参数，降低过拟合风险；另一方面网络结构变为全卷积网络，使得不用限制输入图像尺寸大小，这对医学超声图像更为重要。2) 为验证该算法的鲁棒性和有效性，针对数据集进行详尽实验，研究分析了深度学习方法的高识别率和优异泛化能力的原因。

3.2 Deep-Echo 模型

将分别从如何构建全卷积网络、全局空间金字塔平均池化层、将类别显著性图纳入可视化过程、如何扩增数据等方面介绍提出的 Deep-Echo 模型。

3.2.1 全卷积的网络

与 GoogLeNet 模型^[19]、ResNet 模型^[20]类似，使用多层卷积层（每层包括 ReLU 层、BN 层和 Pooling 层），用全局平均池化操作替代全连接层。Deep-Echo 模型结构中对最后卷积层输出的特征图，用金字塔平均池化层^[101]代替最大化池化层和全连接层。最后一层输出单元数目为类别的数目，由于实验采用的标准切面有七个类别，因此最后一层输出 7，依次对应相应的类别，采用交叉熵损失函数加 L2 正则化。卷积核数目从 64 开始，每经过一次最大池化层，卷积核数目翻倍，直到 512 为止。学习率初始化为 0.01。具体实验步骤和参数设置见后文实验部分。整个网络结构如图3.2所示。

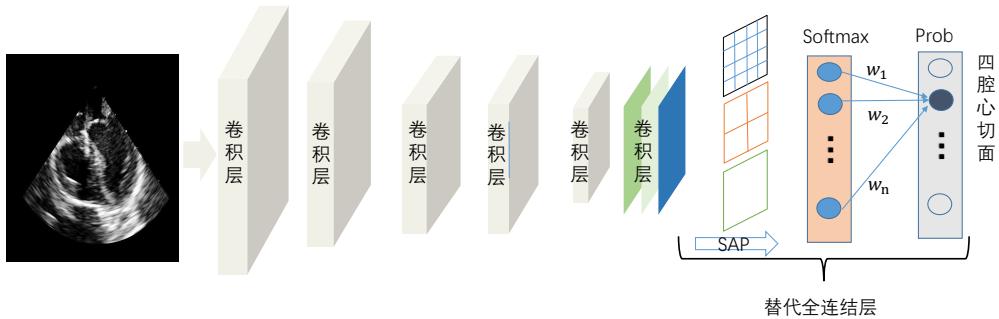


Figure 3.1 Deep-Echo 模型结构示意图

3.2.2 空间金字塔均值池化层

针对深度 CNN 模型中全连接层的两个缺点：全连接层丢失了空间信息，限制了 CNN 只能接受固定尺度的输入，一般只能通过图像尺度归一化的方法来处理不同尺度的输入图像，且使得模型可视化变得不可解释；全连接层参数拥有大约 90% 的模型参数，如 AlexNet 模型^[1] 和 VGG16 模型^[102] 中全连接层参数占全部参数分别为 38M/61M 和 103M/138M，从而导致模型更容易过拟合^[19]。为解决这两个问题，He 等提出空间金字塔池化 (Spatial Pyramid Pooling, SPP) 方法^[101]。SPP 通过使用多个不同大小的池化操作保证固定的特征向量输出，从而允许 CNN 接受任何尺度的输入，增加了模型的尺度不变性，抑制过拟合。与传统的全连接层不同，对每个特征图一整张图片进行多尺度的空间金字塔均值池化，这样每张特征图都可以得到多个尺度的输出。本文方法跟空间金字塔池化网络类似都是三个尺度的空间金字塔池化 (1×1 , 2×2 , 4×4)，其差异在于后不再接多个全连接层，同时用平均池化代替最大化池化，目的在于方便可视化模型的空间位置信息。

3.3 微调迁移学习

利用深度学习进行超声心动图的标准切面识别，仍存在针对小数据量直接训练是否会出现过拟合问题；能否跨领域进行迁移学习，即在自然图像数据集上训练得到的模型能否微调应用到跨领域的超声心动图上。文献 [100] 中指出，用全局平均池化代替全连接层直接随机初始化，从头开始训练模型收敛困难且分类性能下降，故对现有模型进行改造，即针对在自然图像集上预先训练得到的模型如，Alexnet 模型等，变换最后的输出层为所述金字塔平均池化结构，调小学习率后在超声心动图标准切面数据上进行微调迁移学习。

训练时，由于超声心动图的特殊性，人工标注费时费力，对数据集进行扩增能降低人工标注的需求。但扩增数据需注意不能打乱标准切面图像内的局部结构，因此对切面数据只进行水平镜像翻转和旋转。通过引入 BN 归一化层能减轻对 Dropout 的依赖，提高泛化能力，并且本文直接去掉全连接层，故并未采用 Dropout 技术。迁移学习时，由于深度模型中低层的卷积核是跟人类视觉的初级细胞很类似，因此是可以直接迁移复用，高层要针对目标学习判别性信息需进行重新学习^[100]。针对超声心动图的实验支持这样的结论，不同模型的分类准确率都很高，具体实验见后文实验部分。但对于计算机医学辅助诊断而言，模型怎样决策判断比分类准确率更重要。即需解释模型为什么有效和优异的泛化能力从何而来。

3.4 类别显著激活映射图

前文所提模型能高效提取超声心动图标准切面的特征，对超声心动图的单扇形和双扇形标准切面都能很好的识别，甚至对互联网上随意选取的标准切面也能识别。但对模型的有效性和解释性缺乏有力分析，使得对模型决策判断的可信性产生怀疑。针对超声心动图，采用 [100] 提出可视化分析的方法，将其和空间金字塔平均池化结合。对给定图像， $f_j(x, y)$ 表示卷积层 (x,y) 位置上第 j 个神经元的激活值，对第 j 神经元的平均池化操作结果对给定类别 k 的得分函数 S：

$$S_k = \sum_j w_j^k \sum_{x,y} f_j(x, y) \quad (3.1)$$

其中 w_j^k 是第 j 个神经元和第 k 类的连接权重，后接多类多元逻辑损失层，然后由公式3.2可得定义类别激活映射图：

$$M_k = \sum_j w_j^k f_j(x, y) \quad (3.2)$$

其中， M_k 表明在空间 $\square x \square y \square$ 的激活值对该类别分类结果影响的重要性。对类别激活映射图直接双线性插值得到与原图大小相等的显著性图。本文将其和多尺度空间金字塔平均池化结合，得到对多个空间尺度的类别显著激活映射图。值得注意的是，对不同的尺度可设置不同的权重，本文采用同等权重进行融合。该图是对图像空间显著性区域的置信度判别，能辅助可视化分析深度模型的决策过程，在一定程度上解释模型可效性。

3.5 实验结果和分析

3.5.1 实验数据选取和实验方法

本文实验数据来自四川大学华西医院，为临床检查中的经食道超声心动图。所选切面视频包含单扇形和多普勒成像的双扇形两种，其中对双扇形的切面视频，仅取不包含彩色多普勒成像的切面（如图3.3所示）。经专业医师标注的标准切面视频中，至少包含 2-3 个心动周期，并依据医师建议从视频中截取包含一个心动周期的 10 帧图像，并经医师检验筛选后得到最终数据集。

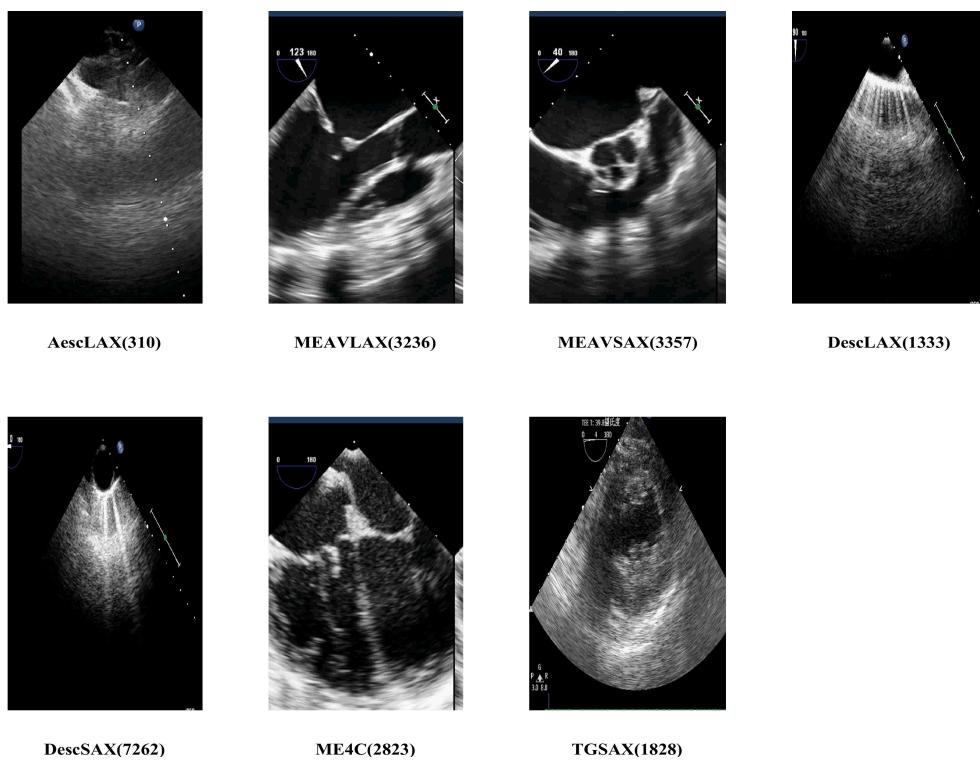


Figure 3.2 七类标准切面超声心动图及数量分布

试验中所用标准切面类别和数量分如布图3.2所示。依据探头在食管中段 (ME) 和经胃底 (TG) 的位置和角度不同，在图3.2中 7 类标准切面分别为：a 为升主动脉长轴 (AescLAX)，b 为主动脉瓣长轴 (MEAVLAX)，c 为主动脉瓣短

轴 (MEAVSAX), d 为降主动脉长轴 (descLAX), e 为降主动脉短轴 (descSAX), f 为食管中段四腔心 (ME4C), g 为经胃底心室短轴 (TGLAX)。其中, d, e, g 为单扇形切面, 其余为双扇形中截取的切面。训练集 (17932 张) 和测试集 (2217 张) 由不同时期采集不同病人对象数据的随机划分。值得注意的是, 所有数据都经过裁剪操作以隐去患者信息。

3.5.2 识别实验结果和分析

本文在构建的超声心动图的数据集上测试分类性能。采用 Caffe 框架^[58] 实现深度卷积网络结构, 预训练模型来自 Caffe model zoo。使用具有 Intel®Core™ i5 3.2GHz 处理器和 12GB 内存的 Tian X GPU 测量所需的时间, 单个切面所需的分类识别时间平均需要 10 毫秒, 基本可满足实时识别。

为验证从自然图像训练的模型能迁移到经食道超声心动图上, 输入图像归一化为 256x256, 网络初始学习率设为 0.001, 迭代一定轮数动态调整学习率大小, 其他参数的设置跟原文献中训练网络结构时一致。三种不同网络结构的深度模型微调前后在同一测试集上的准确率随着迭代次数的增加最后趋于一致, 如表3.1所示, Scratch 表示不经过微调, Finetune 表示经过微调。Deep-echo 模型结构跟 AlexNet 模型类似, 是在其结构基础上去掉全连接层, 用空间金字塔池化层代替, 比 VGG16 和 GoogLeNet 模型的层数更少, 模型结构更简单, 而分类准确率却接近, 表明提出方法的有效性。针对 VGG16 模型和 Google Net 模型也可同样设置, 本文主要关注点不是得到分类精度最优的分类模型, 故并未全部加以实验验证。为验证训练集数据量对深度卷积网络的影响。网络结构采用 AlexNet 模型结合空间金字塔池化层, 在不同数据量上微调, 实验结果如图3.4所示, 数字代表每类至多的数目, 随着数据量的增加, 模型准确率随之提升, 可知针对超声心动图标准切面识别问题, 并不用构建很大的数据集进行识别, 如图3.3中每类至多 500 达到的平均准确率接近使用全部训练集的结果。可推断采用微调技术, 能显著减少深度模型对大数据量的依赖。

为了验证最优模型在不同类别的分类性能, 7 分类的混淆矩阵如图3.4所示, 每行代表实际的类别标签, 每列代表预测的标签。最终的平均分类精度为 97.49%。分类置信度较低的是升主动脉长轴 (AescLAX), 其他各类的准确率都较高。

3.5.3 模型可解释性实验结果分析

深度卷积网络能在标准切面识别问题上得到较高的分类精度, 但仅从分类准确率上评价模型存在局限性。为分析模型的有效性, 采用文中所述可视化方法,

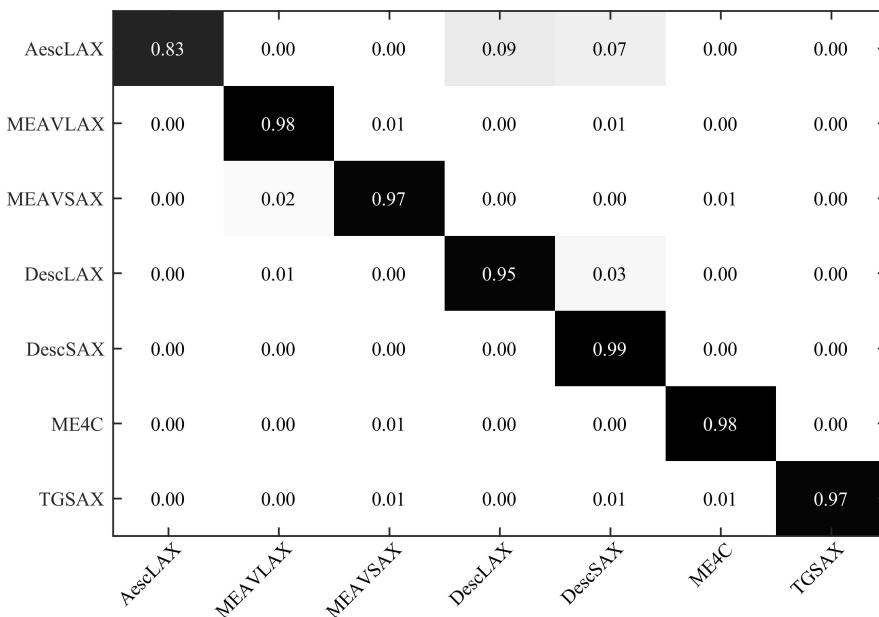


Figure 3.3 Deep-Echo 模型分类的混淆矩阵

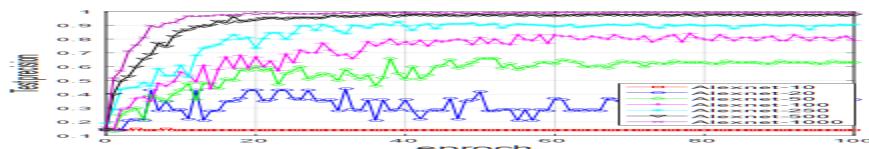


Figure 3.4 不同数据量的平均分类精度

平均分类精度比较		
	Scratch	Finetune
AlexNet	93.35%	93.68%
VGG16	96.66%	96.81%
GoogleNet	97.36%	97.42%
Deep-Echo	97.49%	99.12%

Table 3.1 不同模型分类精度比较

对迁移后的 Deep-echo 模型进行实验。实验结果如图3.5各类切面的原图和显著

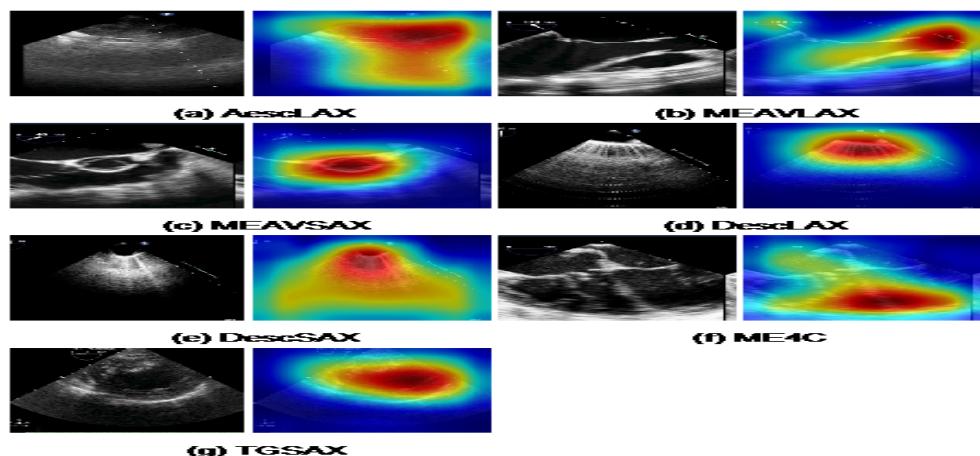


Figure 3.5 各类切面的原图和显著性热力图

性热力图所示，图中为各类切面和对应的类别显著性热力图。类别显著性图中的颜色从蓝到红，表示原图像素中对分类结果影响的重要性是从轻到重。图中结果能很好的解释模型的有效性，并且跟专业医师的判断一致，如图3.5c 中显著性热力图红色区域图定位到图中的圆圈；图3.5d 中定位到的干涉条纹；图3.5f 定位到左心室和右心室的边界等；都跟医师的决策判断依据是一致的。深度模型泛化性

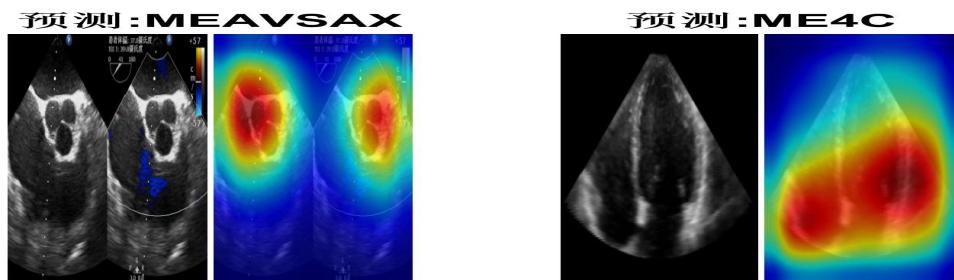


Figure 3.6 深度模型泛化性能可视化分析

能可视化结果如图3.6所示，原图像分别是带彩色多普勒的双扇形切面（图3.6a）和经胸的四腔心切面（图3.6b），这两个图是跟数据集中的经食道超声心动图差异较大，说明深度卷积网络模型确实能对标准切面进行语义分类，表明模型确实能提取到高层语义的特征，深度卷积网络泛化能力优异。可视化结果也能很好的解释模型的有效性，如图3.6中显著性热力图红色区域图定位到图中的圆圈，也是医师认定该切面的关键性结构，图3.6b 定位到左心室和右心室的边界等，都跟医师的决策判断依据是一致的。并且该方法也能作为判断学习模型是否有效的根据，不经过微调的模型虽然能得到较高的分类准确度，并不能得到类似的显著性热力图。

3.6 小结与讨论

本章针对深度特征表示的高层语义在医学图像的具体应用——超声心动图标准切面的自动识别，本文提出了一种基于深度卷积神经网络的超声心动图标准切面自动识别方法，利用所述全局空间金字塔均值池化方法进行微调迁移学习，实验结果表明该方法识别准确率高，并实验分析了数据规模对模型分类精度的影响，结果表明基于深度卷积网络的识别方法应成为超声心动图自动识别的基准方法，接下来会探索更精细类别分类问题，如舒张末期和收缩末期标准切面的识别等。可视化深度模型的实验，对模型的可解释性和有效性进行了分析，推断深度模型的优异的分类性能和泛化能力的原因是可以对类别显著性区域进行判别，采用的可视化方法是对网络模型整体的理解，具体各层特征怎么耦合成语义信息仍需进一步探索。本章研究内容和主要成果以题为《基于深度学习的超声心动图切面识别方法研究》的论文形式已发表在学术期刊上^[103]。

第4章 特征表示的深度可视化

4.1 空间金字塔分解的深度可视化方法

以深度卷积神经网络（Convolutional Neural Network，CNN）为代表的深度学习对计算机视觉和机器学习领域产生了深远影响。但是完全理解深度学习模型的内在工作原理，设计高性能的深度网络结构还是很困难的，一直以来人们普遍将其内部工作原理看成一个“黑箱”，这是由于深度 CNN 存在海量参数，多次迭代更新生成输入输出之间相当不连续和非线性的映射函数；以及对参数的初始状态敏感，存在很多局部最优点。探究 CNN 的运行机制，核心在于它究竟自动提取什么样的特征，经过卷积层、池化层，特征都是分布式表达的，每个特征反映在原图上都会有重叠，故希望建立特征图与原图像之间的联系，即深度可视化。该技术试图寻找深度模型所提取各层特征较好的定性解释，并在设计开发新网络结构方面扮演重要角色。

目前针对 CNN 可视化的研究，主要集中在如何理解 CNN 从海量数据中自动学习到的，能反映图像本质的分层特征表达，即获得网络中隐藏层神经元与人类可解释性概念之间的联系。最直接的方法是展示学习得到的卷积核和相应的特征图，但除了首层卷积核和特征图有直观的解释外，其余各层并没有可解释性。从信号处理的角度看，基于 CNN 高层特征的分类器在输入域，需要较大感知野，才能对以由低频为主的输入图像进行多层非线性响应，并对小的输入改变产生平滑不变输出。同时，由于经过非线性激活函数变换和池化，引入空间不变性获得更好识别性能的同时，也对可视化带来新的挑战。

深度可视化技术可以简单分为三类：基于梯度更新的方法^[97,104–109]；基于特征重建的方法^[98,99,110,111]；基于相关性的方法^[112,113]。基于网络梯度更新的思想是由 Erhan 等^[104]引入，固定模型参数通过梯度更新改变输入值，最大化激活单一神经元或标签类别概率。激活最大化生成的非自然图像还可以是网络模型的对抗样本^[114]。Simonyan 等^[97,105,106]通过梯度上升方法迭代寻找使得最大化激活 CNN 某个或某些特定的神经元的最优图像，其假设神经元对像素的梯度描述了当前像素的改变能影响分类结果的强度。文献 [97] 引入 L2 正则化先验（或称权重衰减），改进可视化效果。Yosinski 等^[107]进一步提出高斯模糊正则化、梯度剪切等技术，其中梯度剪切指的是每次只更新对分类最有利的一部分梯度，改善生成图像质量。文献 [105, 109] 考虑神经元的多面性和利用生成网络作为自然图像

的先验来合成更自然的图像。

Zeiler 等^[99] 提出利用反卷积网络，利用反向传播重构各层特征到像素空间的映射，并用于指导设计调优网络结构，提高分类识别精度。在反卷积过程中利用翻转原卷积核近似作为反卷积核，针对特定特征图在训练集上重新训练。Dosovitskiy 等^[110] 提出通过学习‘上’卷积网络来重建 CNN 各层的特征，指出结合强先验，即使用于分类的高层激活特征也包含颜色和轮廓信息。Mahendran 等^[98,111] 通过对学习到的每层特征表达进行反编码重建，提出利用全变分正则化和自然图像先验，并将 L2 范数正则化推广到 p 范数正则化，得到较优的可视化效果。

本文主要关注前两种方法中的正则化技术，基于相关性分解方法请参考文献 [113]。受文献^[115,116]启发，把用于图像生成的拉普拉斯金字塔，进一步扩展成空间金字塔分解方法，并引入显著性激活图技术进一步改进深度 CNN 的可视化效果。

4.2 可视化方法的数学模型

激活最大化和特征表达反编码重建均是针对已经训练好的模型，对给定输入 $x_i \in R^{C \times H \times W}$ ，其中 C 为颜色通道数，H，W 为图像高和宽。CNN 模型可抽象为函数 $\phi: R^{C \times H \times W} \rightarrow R^d$ ，其第 i 个神经元的激活值为 $\phi_i(x)$ ，对给定图像 x_0 的特征编码 $\phi_0 = \phi(x_0)$ ，定义参数 θ 的正则化项 $R_\theta(x)$ ，寻找使得能量泛函最小化的初始输入 x^* ，其数学模型为

$$x^* = \underset{x}{\operatorname{argmin}}(l(\phi(x), \phi_0)) + \lambda R_\theta(x) \quad (4.1)$$

其中， l 损失比较的是 $\phi(x)$ 和目标 ϕ_0 的差异，选择不同的损失函数定义不同的可视化方法。但该优化通常是一个非凸优化问题，通常采用梯度下降法去寻找局部最优值为

$$x \leftarrow x + \alpha \frac{\partial \phi_i(x)}{\partial x} \quad (4.2)$$

激活最大化方法是文献 [104] 中提出针对深度架构中任意层中的任意神经元所提取的特征，寻找使一个给定的隐含层单元的响应值 $\phi_0 \in R^d$ 最大的输入模式，可由内积形式定义 l 损失为

$$l(\phi(x), \phi_0) = - < \phi(x), \phi_0 > \quad (4.3)$$

式中 ϕ_0 需人工指定，最大化激活的目标可以是全连接层的特征向量，也可以是卷积层某一通道的某一神经元的激活值。特征表达的反编码重建，通过最小化给

定特征向量与重建目标图像特征向量间的损失，一般采用欧式距离来衡量损失误差，定义如下

$$l(\phi(x), \phi_0) = \frac{\|\phi(x) - \phi_0\|^2}{\|\phi_0\|^2} \quad (4.4)$$

但也可利用其它距离度量函数来评价损失。

4.3 梯度更新的可视化方法

用于分类的深度 CNN 提取高层语义信息的同时，丢失了大量低层结构信息。由于首层卷积核大都类似 Gabor 滤波器，导致梯度更新可视化生成图像中包含许多高频信息，虽然能产生大的响应激活值，但对可视化来说导致生成的图像是不自然的。还由于网络模型的线性操作（如卷积）导致对抗样本^[114] 的存在，为得到更类似真实自然图像的可视化结果，需在优化目标函数中引入正则化作为先验。

4.3.1 p 范数正则化方法

对图像来说，像素大小需在一定范围内，直接最大化激活类别概率，生成图像类似随机噪声图像。文献 [97] 通常引入 L2 范数正则化，惩罚过大和过小的极端值，其公式为 $R_\theta(x) = \|x\|_2^2$ 。在文献^[98] 中将其扩展到彩色图像 RGB 通道空间中的 p 范数正则化为

$$R_\theta(x) = \frac{1}{HWC^p} \sum_{h=1}^H \sum_{w=1}^W \left(\sum_{c=1}^C x(h, w, c)^2 \right)^{\frac{p}{2}} \quad (4.5)$$

式中 h, w 表示图像的行和列大小， c 表示颜色通道数，对比发现，文献 [97] 提出的 L2 正则化是忽视各颜色通道的差异的，正则化的力度可通过缩放常量 p 进行控制，即使得图像像素值大小保持在合适的范围内。

4.3.2 高斯模糊和 TV 变分

基于梯度更新可视化方法，引入高斯滤波器主动惩罚高频信息^[107]，高斯模糊核半径大小由高斯函数的标准差控制，可随迭代次数动态调整模糊核大小。

全变分^[98](Total Variance, TV) 跟高斯模糊类似，鼓励可视化生成分片的常量块区域，对离散图像全变分操作可由有限差分来近似求解为

$$R_{TV}(x) = \frac{1}{HWC^\beta} \sum_{hwc} ((x(h, w+1, c) - x(h, w, c))^2 + (x(h+1, w, c) - x(h, w, c))^2)^{\frac{\beta}{2}} \quad (4.6)$$

式中 $\beta = 1$, 但其在可视化过程中, 在图像的平坦区域并不存在边缘, 全变分操作仍沿着边缘方向扩散就会导致出现虚假的边缘, 会引入所谓的“阶梯效应”现象。 $\beta < 1$ 时结合超拉普拉斯先验^[117]能更好匹配自然图像的梯度统计分布, 但对可视化来说反而使得可视化更困难。文献[98]实际实验表明, 跟高斯模糊核一样, 需随迭代次数动态调整 β 大小。

4.3.3 基于数据统计先验

由于常规可视化方法并没有对颜色分布进行建模, 文献^[105]提出通过引入外部自然图像数据, 计算图像色块先验为

$$R_\theta(x) = \sum_p \|x_p - D_p\|_2^2 \quad (4.7)$$

式中 p 为块索引, x_p 表示稠密采样的归一化图像块, D_p 表示自然图像块数据库中距离 x_p 最近图像块。该方法跟文献^[115]中利用参考图像“指导”人脸图像嵌入重建类似。并且基于数据的统计先验可进一步扩展, 引入生成对抗网络, 利用生成网络主动生成自然图像先验^[108]。

4.4 空间金字塔分解

前文介绍的正则化先验主动限制图像空间中高频率和高振幅信息, 生成的可视化图像存在如下问题: 1) 彩色图像的颜色分布仍是不自然的。2) 生成的图像中包含可识别类别对象的多个重复成分, 并且这些部件不能组合成完整的有意义整体。3) 缺乏令人可信的低频细节, 存在棋盘效应, 只是形似。针对这些问题提出利用空间金字塔分解, 主动提升低频信息和调控高频信息以改善生成图像的可视化效果。

4.4.1 高斯和拉普拉斯金字塔分解

拉普拉斯金字塔 (Laplacian Pyramid, LP)^[118] 是由一系列包含带通滤波器在尺度可变的图像上加低频残差组成的。首先通过高斯平滑和亚采样获得多尺度图像, 即第 K 层图像通过高斯模糊、下采样就可获得 K+1 层, 反复迭代多次构建高斯金字塔 (Gaussian Pyramid, GP)。用高斯金字塔的 K 层图像减去其第 K+1 层图像上采样并高斯卷积之后的预测图像, 得到一系列的差值图像即为拉普拉斯金字塔分解图像。拉普拉斯金字塔分解过程(见图 1 所示)包括 4 个步骤: 1) 高斯平滑; 2) 降采样(减小尺寸); 3) 上采样并高斯卷积(图中 expand 操作); 4) 带通滤

波(图像相减)。拉普拉斯金字塔突出图像中的低频分量,拉普拉斯金字塔分解的目的是将源图像分解到不同的空间频带上。

The Laplacian Pyramid

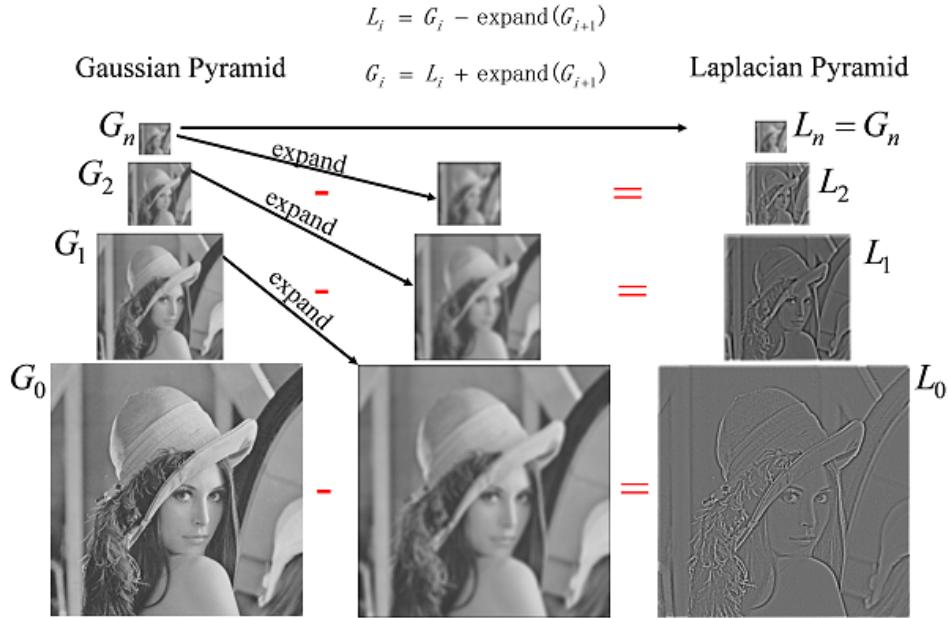


Figure 4.1 高斯和拉普拉斯金字塔

由于自然图像统计特性中的尺度不变性,也称为 $1/f$ 法则^[119],即自然图像集 $I(f_x, f_y)$ 的平均傅里叶功服从 $I(f_x, f_y)^2$ 。在激活最大化可视化深度CNN模型过程中利用提出的高斯和拉普拉斯空间金字塔分解,调整生成梯度图像包含的频谱分量大小。其中空间金字塔分解正则化项为

$$r_\theta(x) = \sum_{k=1}^K [LP_k + GP_k] \quad (4.8)$$

式中 k 代表构建 k 层金字塔分解,本文实验 k 选取为4。 LP_k 为第 k 层的拉普拉斯金字塔分量, GP_k 为第 k 层的高斯金字塔分量。

4.4.2 梯度归一化

基于梯度更新的可视化方法,由于原输入空间中高低频分量混杂在一起,对原输入图像相应的更新梯度进行归一化操作能得到较好可视化效果,即对输入图像每次迭代更新的梯度,则提出梯度归一化操作:

$$g \rightarrow \frac{g}{g.std() + \delta} \quad (4.9)$$

式中 δ 为非负小常量, std 表示梯度矩阵的方差。该梯度中心归一化技术, 可以减少产生重复的对象碎片的倾向, 而倾向于产生一个相对完整对象。梯度归一化的引入同批归一化 (Batch Normalization) 思想类似, 校正 CNN 网络非线性变换引起的“偏移”, 该方法也侧面验证最新提出的分层归一化^[64] 的有效性。

4.4.3 类别激活图限制可视化区域

根据文献 [100] 提出的类别激活图技术, 假设 $f_j(x, y)$ 表示最后的卷积层空间 (x, y) 位置上第 j 个神经元的激活值, 则对 j 神经元的全局平均池化操作结果对给定类别 k 的得分函数 S_k :

$$S_k = \sum_j w_j^k \sum_{x,y} f_j(x, y) \quad (4.10)$$

式中 w_j^k 是第 j 个神经元和第 k 类的连接权重。根据文献 [100], 由式 4.10 可得定义类别激活图 M_k 为

$$M_k = \sum_j w_j^k f_j(x, y) \quad (4.11)$$

式中 M_k 表明在空间 (x, y) 位置的激活值对分类结果影响的重要性。对类别激活映射图直接双线性插值得到与原输入图像大小相等的显著性图。本文利用显著性激活图作为梯度更新的权重因子, 即输入变为原始输入图像与类别激活图的加权乘积。动机是要求网络梯度更新保持在类别显著性区域内, 压制无关背景信息的生成。具体详情请参见第四章实验部分。

4.4.4 优化方法

深度 CNN 模型优化策略的核心是随机梯度下降法, 常用方法是带动量的随机梯度下降法为:

$$V_t = \mu V_{t-1} - \alpha * \nabla f(x_i) \quad (4.12)$$

$$x_{t+1} = x_t + V_t \quad (4.13)$$

式中 μ 为动量因子表示保持原更新方向的大小, 一般选取 0.9, x_t 为在 t 时刻待更新的梯度, α 为学习率; 文献^[98,111] 采用自适应梯度 (Adaptive Gradient, AdaGrad)^[120] 的变种算法, 根据历史梯度信息自适应调整学习率。同时文献 [121] 采用的二阶优化算法针对纹理和艺术风格重建问题, 得到比用基于一阶随机梯度下降算法更优的可视化效果。但本文通过实验对比发现对各种优化方法对生成图像质量影响不大, 从简选择带动量的随机梯度优化方法。

4.5 实验结果分析和讨论

基于梯度更新的可视化方法主要用于激活最大化和特征重建，但文献 [122] 指出用随机未训练的 CNN 模型也能较好重建原图像，表明特征编码重建不能很好解释训练得到 CNN 模型的内在工作机理。故本文实验主要关注在对 ImageNet 公开数据集上预先训练得到的分类模型进行激活最大化可视化实验。

4.5.0.1 不同深度模型的类别可视化

实验选取的深度模型来自于开源社区的 Caffe model zoo，不同的 CNN 模型如：AlexNet 模型^[1]，Vgg-19 模型^[18]，Google-CAM 模型^[100]，GoogleNet 模型^[19]，ResNet 模型^[20]，其分类识别性能依次从低到高，模型的复杂程度依次递增。本文实验默认采用提出的梯度归一化，并引入多分辨率、随机扰动和剪切等小技巧作为通用设置，提高可视化效果。



Figure 4.2 不同模型类别可视化实验结果

为比较不同深度 CNN 模型学习相同样本时特征图的差异，根据式4.1，给定高斯噪声生成随机图像作为输入，指定可视化物体类别向量（见图4.2所示，类别为所有类别中的第 13 类布谷鸟），施加前文提出不同正则化项的组合： p 范数、高斯模糊和金字塔分解正则化。

图4.2结果表示 5 种 CNN 模型在相同正则化方法和相同梯度更新策略下的可视化效果，对比图4.2中 a, b, c 发现随着网络模型深度的增加，可视化难度增大分类性能同可视化效果一致；Vgg-19 模型由于跟 ResNet 模型卷积核大小类似，且比 AlexNet 首层卷积核小（7 和 3），即可视化效果倾向生成比 AlexNet 更大尺寸的物体。而由图4.2中 a, d, e 对比可知，由于 GoogleNet 模型中卷积层的卷积核大小不一，使得可视化结果中引入更多细节。综合可知，基于 GoogleNet 模型的可视化效果最好，后面实验均是在其模型的基础上进行实验比较。

4.5.0.2 不同正则化方法的类别可视化

为验证不同正则化方法对理解深度模型的特征表达的影响，采取前文所述的不同正则化方法，可视化效果结果见图4.3所示，从上到下依次可视化类别为

金甲虫，海星，蝎子，酒壶，卷笔刀。

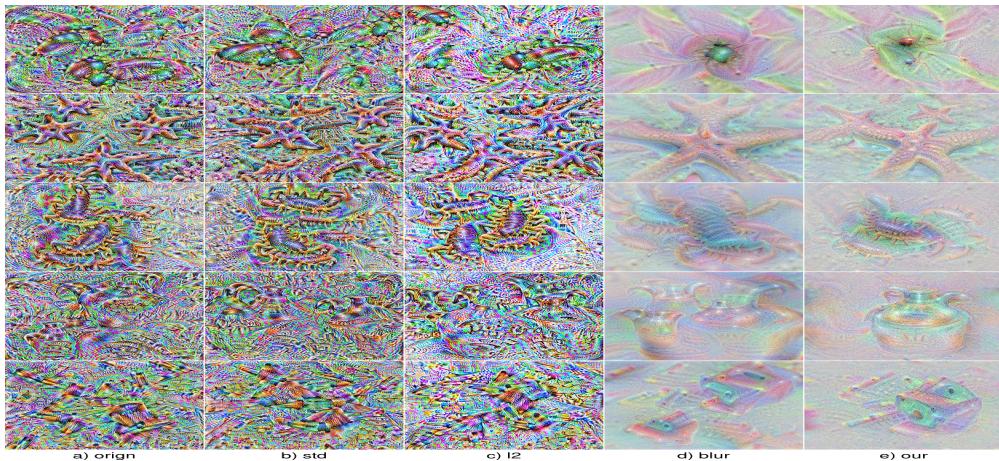


Figure 4.3 不同正则化方法的可视化效果

图4.3(a)列仅施加默认设置和不加梯度归一化的结果,由于输入的随机性,并不能保证每次都生成有意义的可视化结果,但引入本文提出的梯度归一化后,能大概率生成可视化结果见图4.3(b)列所示,图4.3(c)列表示只采用 p 范数正则化,跟文献[97]一致取2,使得图像更平滑,但仍与真实图像相差较大。通过前文理论分析和实验验证,全变分跟高斯模糊作用类似,本文采用根据迭代轮数动态调整高斯模糊核大小,具体是在刚开始采用较大值希望生成物体大概轮廓,随迭代逐渐调小模糊核使得更多细节生成,具体见图4.3(d)。但是这个参数无法自适应设置为最优,对图像高低频分量无法调整控制,而本文提出的利用金字塔分解正则化方法能从粗到细调整,产生较优结果见图4.3(e)列所示。

4.5.1 金字塔分解可视化实验结果

为验证提出金字塔分解正则化方法,对中间层卷积核的可视化,采用前文提出式4.8,指定深度CNN模型中不同卷积层中不同通道,利用前文提出的带动量的梯度更新策略,可视化结果见图4.4,其中从上到下依次为GoogleNet模型低中高层不同通道的可视化结果,与文献[99]一致,低层多尺度分辨率生成的纹理见图4.4首行所示,中层是一些物体部件,见图4中间行所示蜜蜂的局部结构,而高层是更完整的抽象概念见图4.4下层中完整的花瓣。对比图4.4(b)、(c)列,可验证拉普拉斯金字塔主动分解提升图像部分低频成分,而高斯金字塔分解生成的图像中高频细节更突出。

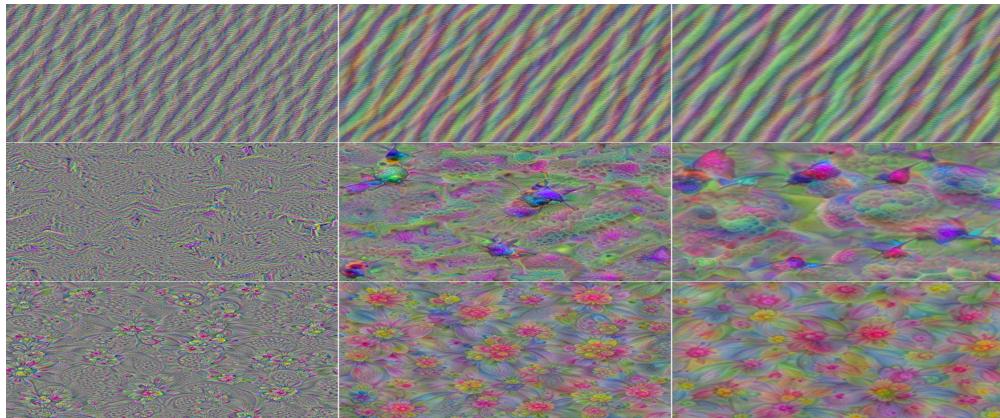


Figure 4.4 金字塔分解正则化可视化效果

4.5.2 引入类别显著性的可视化

通过观察之前可视化结果可知，生成的图像中除了该类别外仍有许多额外的上下文信息（见图4.2中鸟类别的树枝），这些信息与模型的分类能力相关联，可通过引入类别激活图可改善可视化效果。迭代更新过程中依据采用式4.11，使用类别激活图作为加权因子限制迭代更新区域。

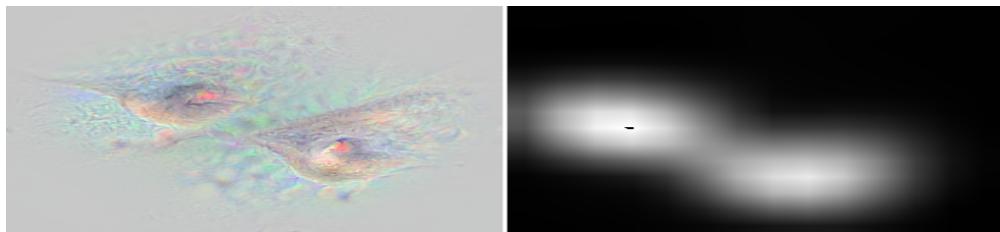


Figure 4.5 引入类别激活图的可视化

实验结果见图4.5(a)所示，具体实验设置和图2采用的参数一致，使用提出的金字塔分解正则化技术，图4.5(b)为图4.5(a)相应的类别激活图，图4.5(a)结果表明与类别无关的上下文信息得到压制，但仍存在两个类别中心。

4.6 小结与讨论

本文针对理解深度 CNN 特征空间存在的问题，提出一种用于改善深度 CNN 分类模型的可视化方法。其中通过改善激活最大化可视化技术来产生更具有全局结构的细节、上下文信息和更自然的颜色分布的高质量图像。该方法首先对反向传播的梯度进行归一化操作，在常用正则化技术的基础上，提出使用空间金字塔分解图像不同频谱信息；为限制可视化区域，提出利用类别显著激活图技术，可以减少优化产生重复对象碎片的倾向，而倾向于产生单个中心对象以改进可视

化效果。激活最大化可显示 CNN 在分类时关注什么。这种改进的深度可视化技术将增加我们对深层神经网络的理解，进一步提高创造更强大的深度学习算法的能力。该方法适用于基于梯度更新的可视化领域，是对网络模型整体的理解，具体各层特征怎么耦合成语义信息仍需进一步探索，深度 CNN 模型如何重建一个完整的类别概念，仍是一个开放性问题。本章研究内容和主要成果以题为《空间金字塔分解的深度可视化方法》的论文形式已发表在学术期刊上^[123]。

第5章 特征表示的中层结构性语义应用

5.1 心室的计算机辅助检测方法

计算机辅助检测 (Computer-Aided Detections, CADs) 是医学影像诊断过程中的一项重要任务，是进行相关结构功能测量的前提条件。其中，二维图像的目标组织结构自动检测是 CADs 技术的核心基础。在临床实践中，医生需整合不同模态、不同位置方向且以不同比例显示的图像信息，目前的研究主要关注如何使检测过程快速自动化。由于医学影像自身的特殊性，比如缺乏大量高质量标注数据；大多数医学目标组织结构存在非刚性形变；图像背景前景的区分不明显等，导致组织结构自动定位比较困难。现有大多数 CADs 系统在临床实际应用中表现不佳的原因是：检测结果的敏感性和特异性都较低，诊断效能低^[124]。

不同模态的医学图像中，如超声、计算机断层扫描和核磁共振等，都存在目标身体器官自动定位的问题。以左心室 (Left Ventricle, LV) 检测为例，大多数 LV 定位方法主要依据位置、时间和形状的假设。基于位置的方法仅假设心室在图像的中心，该方法并不对不同病人心室位置的差异性以及图像的尺寸变化进行考虑，效果较差；基于时间的方法，假设左心室是图像中唯一的运动对象，然而这种方法敏感性高，除心室的运动伪影之外，还存在其它运动的器官，如 Schollhuber^[125] 针对 MRI 短轴使用时空信息并消除运动伪影，由分层模式匹配算法定位包含 LV 的感兴趣区域，其通过使用互信息图像配准使运动伪影最小化，随后估计特征强度—时间曲线进行像素分类和边界的提取，得到最终分割结果；基于形状的方法将 LV 视为圆 (短轴)、椭圆 (长轴)，然而该方法通常针对异常形状的 LV 容错性差，如 Lu 等^[126] 使用大津阈值度量圆形程度，然后进行霍夫变换定位 LV 位置。也可搜索每个切片的质心，并用三维最小二乘拟合去除异常值，得到分割结果^[127]。

不依据具体的强先验假设，机器学习算法可通过区分前景目标对象和背景来解决目标结构自动检测的问题。如 Kellman 等^[128] 提出了一种使用概率集成提升树来估计 LV 姿态和用空间间隔学习 LV 短轴边界的方法。Zhou 等^[129] 在超声心动图中通过规一化集成提升回归学习非线性映射以定位 LV，其团队后来提出针对多个器官的特异性置信最大化分类器，整合更高的自由度以改善回归定位任务的精度。Liu 等^[130] 通过利用基于子模块函数优化理论的多标记搜索策略来进行标记点的检测。Zheng 等^[131] 在实现器官定位的同时，通过组合优化置信度

来估计目标器官的位置、缩放及朝向等参数值。前述机器学习算法都基于弱先验知识，启发式设计相关特征，结合滑动窗口策略，选择分类器进行分类判断窗口中内容以估计相应位置。

近来通用物体检测领域取得巨大进展，主要得益于深度学习能利用大量标注数据，从原始像素出发，逐层分级学习中高层抽象语义特征^[96]。区域卷积神经网络^[132]在大规模自然图像数据集(如ImageNet^[17])上，识别性能远超传统方法^[1,132]。当前实践中由于深度学习需要大量的训练数据，所以仅在少数医学任务中取得有限的成功应用。深度学习方法用在定位检测问题时可分为两个阶段^[133]:候选框位置选取和窗口内容类别分类。如利用深度卷积网络进行显微镜图像中细胞检测^[41]、结合深度全卷积网络的MRI心室检测与分割^[42,43]和超声图像解剖结构的检测^[44]。

上述方法大都关注特定目标结构的检测分割，而本文专门针对目前CADs普遍存在的检测定位问题，基于改进的生成候选框的快速区域深度卷积神经网络(Faster RCNN)^[134]方法，提出一种医学目标结构检测框架:1) 在区域生成网络的基础上引入空间变换损失使得候选框生成网络能捕捉目标的空间变换参数；2) 采用在线困难样例挖掘策略，加快训练收敛过程，提高检测小目标的准确度；3) 并基于目标先验知识，针对左心室提出利用检测二尖瓣环、心内膜垫和心尖位置，高效估计左心室姿态参数。4) 为验证该算法的鲁棒性和有效性，分别针对两个具体CADs应用进行实验分析。

5.2 区域卷积神经网络概览

5.2.1 物体检测形式化定义

若用 r 来表示图像中的矩形窗口区域，令 R 表示由对象检测系统提供的所有候选窗口的集合，将有效定位标记定义为 R 的子集，使得标记位置内内容“不重叠”，令 Y 来表示所有有效标记位置的集合。并合并常用的非最大值抑制(Non-maximum suppression, NMS)过程，给定图像 x 和窗口评分函数 f ，物体检测算法流程可定义为1。

形式化定义物体检测过程见公式5.1，式中参数定义请参考算法1。

$$y^* = \arg \min_{y \in Y} \sum_{r \in Y} f(x, r) \quad (5.1)$$

通常公式5.1可通过贪心搜索的方法来完成，算法将联合最小化在算法1中产生假阳例的数量和最大化检测窗口评分函数，即寻找具有最大得分但同时不重叠的

Algorithm 1 物体检测算法

输入: (x, f)	$\triangleright x$ 为图像, f 为窗口得分函数
$D :=$ 所有候选框 $r \in$ 使得 $f(x, r) > 0$	\triangleright 一般采用滑动窗口
按 f 排序 D 使得 $D_1 \geq D_2 \geq D_3 \geq \dots \geq D_n$	
令 $y^* := \{\}$	
for $i = 1$ to n do	\triangleright 一般采用非极大值抑制
若 D_i 和 y^* 中任意候选框不重叠	
$y^* := y^* \cup D_i$	
End for	
Return: y^* , 物体的目标位置.	

滑动窗口位置集合。

5.2.2 区域卷积神经网络的演进

2014 年 Girshick 等^[132] 提出区域卷积神经网络 (Region-based Convolutional Neural Network, RCNN)，对每一候选框窗口都进行一次前向传播，这将导致冗余计算，时间复杂度高，为解决这一问题，He 和 Ren 等提出 SPP-net^[101] 和 Fast RCNN^[133] 加以改进，不再把每一候选窗口均送入网络，而是仅对图像特征提取一次，把原图中候选区域投影到卷积特征图上，然后对投影后的区域特征图进行空间感兴趣区域池化 (ROI Pooling) 得到固定长度的特征向量。其中 Fast RCNN 中的兴趣区域池化是 SPP-Net 中多尺度空间金字塔池化的特例，仅用单一尺度的金字塔池化操作。RCNN 及其改进的 Fast RCNN 都依赖于人为设计的候选框生成方法，如选择性搜索等。为减少生成候选框的计算时间，Faster RCNN 提出区域生成网络 (Region Proposal Networks, RPN)，区域生成网络和检测网络共享提取特征的卷积层，仅提取几百个或者更少的高质量预选窗口，且召回率较高 (导致更少的假阳例)。但现有的通用物体检测算法均是假设候选框为矩形，不能解决旋转朝向问题。

5.3 候选区域生成网络及其改进

本章将分别从候选区域生成网络模型的结构、仿射变换候选框区域的生成、空间变换损失函数的设计、模型训练方法等方面介绍本文所提出框架，并结合 Faster RCNN 模型提出端到端的目标检测方法。

5.3.1 候选区域生成网络模型结构

候选区域生成网络将一图像（任意大小）作为输入，输出目标候选框的集合和每个候选框内有无目标的概率估计，如图5.1右图所示，RPN 在卷积层后接两个全卷积层完成候选区域生成功能，以实现增加滑动窗口操作。该模型使用全卷积网络 [20] 处理任意大小的图片输入，为了和目标检测网络^[133]共享计算，在特征提取的过程中同时计算目标检测所需的感兴趣区域的初始估计，在最后一个共享卷积层输出的特征映射图上滑动小网络，卷积特征映射图上 $n \times n$ 大小空间窗口作为该网络全连接的输入，本文 n 取 3。每个滑动窗口映射到一个低维向量上（如图5.1左上中 256-d），该向量输出给两个全连接层——候选框位置定位回归层和候选框类别分类层。原文中采用类别无关分类损失，即仅区分该候选框内是否包含物体（前/背景），本文将其扩展为类别相关的分类损失。

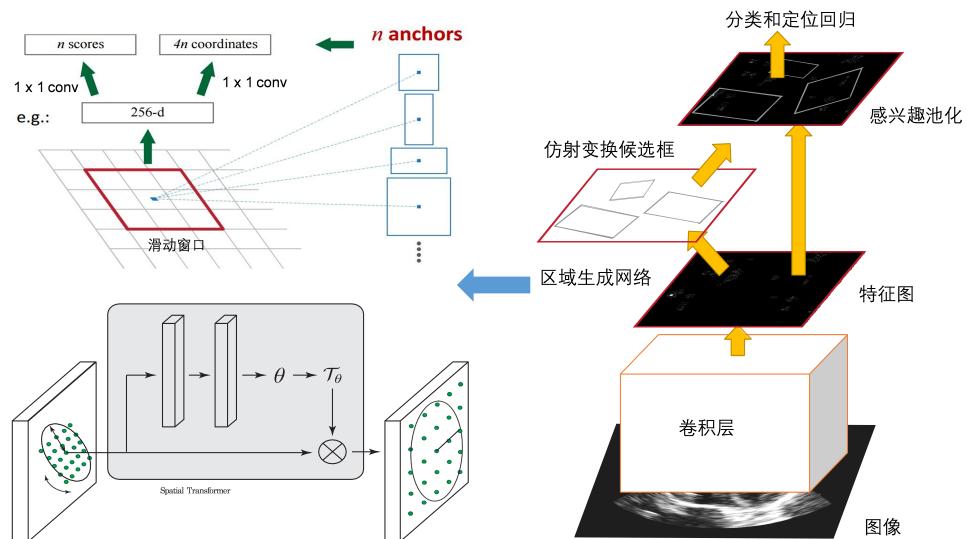


Figure 5.1 左上：引入空间不变性的 anchor 机制左下：空间变换网络右：Faster R-CNN 带仿射变换的检测模型框架

为引入空间尺度不变性，采用多尺度和多纵横比的“参考”框（anchor），如图5.1左上所示。该机制可看作是金字塔型参考框的回归，避免了枚举多尺度、多纵横比的图像或卷积核。在每一个滑动窗口的位置，同时预测 k 个参考区域，回归层有 $4k$ 个输出，即 k 个 box 的坐标编码，多元逻辑回归分类层输出 $(c + 1) \times k$ 个（物体类别数 c 加背景类的）概率估计。候选框由相应的 k 个 anchor 的参数化表示，每个 anchor 以当前滑动窗口中心为中心，并对应一种尺度和长宽比，我们使用 3 种尺度和 3 种长宽比，在每一个滑动位置就有 $k = 9$ 个 anchor。对于大小为 $w \times h$ 的卷积特征映射，总共有 $w \times h \times k$ 个 anchor。

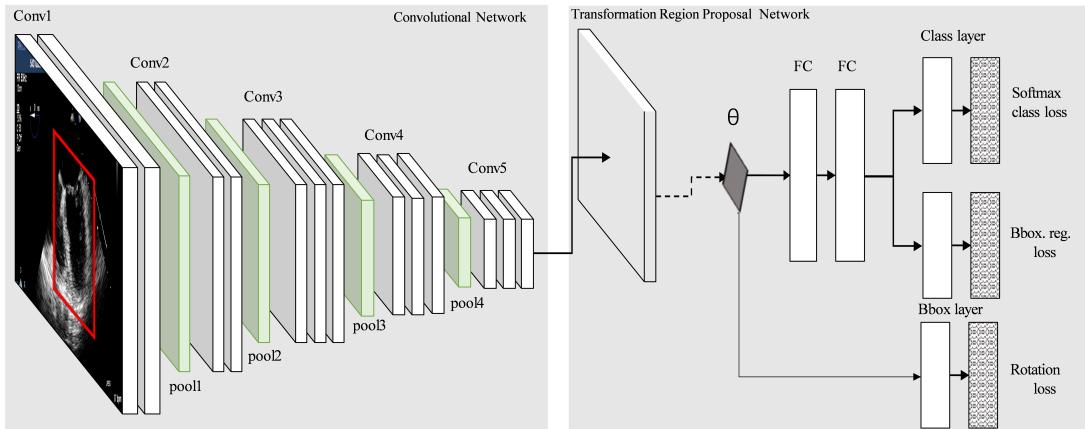


Figure 5.2 考虑物体朝向的区域生成网络模型结构示意图，图中 conv 表示卷积层，pool 表示池化层，FC 表示全连接层，softmax class loss 表示多任务损失中的分类损失，Bbox.reg loss 表示候选框回归定位损失，Rotation loss 表示文中的针对变换参数 θ 的 Von Mise 损失。

5.3.2 仿射变换候选框

为检测物体的姿态，结合空间变换网络^[135]（见图5.1左下），提出带仿射变换的候选框生成算法。之前候选框生成方法仅考虑固定尺度和宽高比的矩形框，并未考虑物体的旋转朝向，二维空间仿射变换可表示为：

$$\begin{pmatrix} x_i^s \\ y_i^s \\ 1 \end{pmatrix} = \tau_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} \quad (5.2)$$

式中 $(x_i^s, y_i^s)^T$ 为输入特征图中目标坐标系下的网格点， τ_θ 为变换矩阵， $(x_i^t, y_i^t)^T$ 为输出特征图中目标坐标系下的采样网格点。其中由于图像的坐标不是中心坐标系，宽高坐标需归一化表示，如 $-1 \leq x_i^s, y_i^s \leq 1$ ，且采用图形学中齐次坐标表示。公式5.2能用六个参数定义对输入特征图的裁剪、平移、旋转和缩放等变换。该公式进一步简化为只考虑旋转变换：

$$\begin{pmatrix} x_i^s \\ y_i^s \\ 1 \end{pmatrix} = \tau_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} \quad (5.3)$$

其中 α 表示绕图像中心顺时针旋转角度，通常变换后的像素并不是在相应网格的整数值，常用双线性插值进行近似，变换后的候选框送入感兴趣区域池化层，后接多任务损失函数。实质是把空间变换层嵌入到 RPN 网络中，并且引入有监督的损失以指导空间变换。

5.3.3 朝向回归损失函数

旋转朝向的周期性会导致两个问题：(1) 要优化的损失函数不能区分对于周期性损失，简单地将模运算符应用于网络的输出会导致不可靠的损失，不能再被鲁棒地优化。(2) 由大多数参数模型中执行的矩阵向量积产生的回归输出是固定的线性运算。为此提出旋转朝向回归损失 $L_{VM}(o, o^*)$ ，第一个问题可以通过采用 Von Mise 分布^[136] 来解决损失函数不连续性，其近似服从于单位圆上的正态分布：

$$p_{VM}(\varphi \mid \mu, k) = \frac{e^{k \cos(\varphi - \mu)}}{2\pi I_0(k)} \quad (5.4)$$

其中 p 指相应的概率密度函数， φ 指角度， μ 是分布的平均角度， k 与近似高斯方差成反比，而 $I_0(k)$ 是阶数为 0 的修正贝塞尔函数，利用余弦函数来避免不连续性，可以得出以下损失函数：

$$C_{VM}(\theta \mid t, k) = 1 - e^{k(\cos(\theta - t) - 1)} \quad (5.5)$$

式中 θ 为预测旋转角度大小， t 为真实旋转角度大小，称 t 为目标值， k 为控制损失函数尾部的简单超参数。由角度 φ 正余弦组成的二维向量 $y = (\cos \varphi, \sin \varphi)$ 替代表示，利用自然语言处理文献中广泛使用的余弦代价函数 [31] 来解决使用线性操作来预测周期值的问题：

$$C_{cos}(y \mid t) = 1 - \frac{y \times t}{\|y\| \|t\|} \quad (5.6)$$

在神经网络框架中的实现是相对简单的，因为所需要的是全连接层和归一化层，前向传播公式：

$$f_{BT}(x \mid W, b) = \frac{Wx + b}{\|Wx + b\|} \quad (5.7)$$

式中 $W \in R^{n \times 2}$ 和 $b \in R^2$ 是来自全连接层的可学习参数，然后反向传播归一化损失的导数为

$$\partial_{x_i} \frac{x}{\|x\|} = \partial_{x_i} \frac{x}{\sqrt{\sum_j x_j^2}} = \frac{\sum_{j \neq i} x_j^2}{(\sum_j x_j^2)^{\frac{3}{2}}} = \frac{\sum_{j \neq i} x_j^2}{\|x\|^3} \quad (5.8)$$

式中归一化确保输出值被联合学习，通过比较 CVM 和 $Ccos$ ，最终朝向回归损失函数为

$$L_{VM}(y \mid t) = 1 - e^{k(y \times t - 1)} \quad (5.9)$$

与式 5.6 相似，主要区别在于存在 e ，它将目标值附近的错误“下推”，实际上是比较小地惩罚小错误。

5.3.4 带朝向的多任务损失函数

多任务损失分别存在于 RPN 及检测网络中，图 2 中显示的是检测网络结构示意图。每一个候选框均送感兴趣池化层，后接两层的全连接层和多元逻辑回归分类损失（图5.2中 Softmax loss），候选区域回归定位损失（图5.2中 Box.reg loss）和旋转朝向回归损失（图5.2中 Rotation loss）：

$$L(p, p^*, t, t^*, o, o^*) = L_{cls}(p, p^*) + \lambda[p^* > 0]L_{box}(t, t^*) + \mu[p^* > 0]L_{VM}(o, o^*) \quad (5.10)$$

式中，分别代表预测类别分类概率，候选框偏移量和感兴趣区域内物体的朝向大小；表示标记类别为背景，表示框内是否有目标的指示函数，分别表示物体的候选框标记和真实朝向。为两个损失的相应平衡权重大小，详细形式如下：

$$L_{(cls)}(p, p^*) = - \sum_c \log p_c^* \quad (5.11)$$

$$L_{box}(t, t^*) = - \sum_{i \in (x, y, w, h)} smooth_{L1}(t^* - t) \quad (5.12)$$

$$smooth_{L1}(x) = \begin{cases} 0.5x^2 & if |x| < 1 \\ |x| - 0.5 & else \end{cases} \quad (5.13)$$

$L_{cls}(p, p^*)$ 和 $L_{box}(t, t^*)$ 是公式5.4中的分类损失和相应的平滑 L1 损失， c 代表类别数。

5.3.5 困难样例挖掘

由于医学数据样本标注困难，数量相对较少，一般假设与目标位置矩形框有重叠的候选框是有较大概率是难以区分的，结果也可能是次优的，因为在其他位置可能存在更难区分的样本，导致模型收敛变慢，误警率高。在每次迭代训练过程中采用在线困难样例挖掘方法（Online Hard Example Mining, OHEM）^[137]，对所有候选框的损失进行排序，由于相似候选框重叠区域的损失很接近，可采用非极大值抑制策略限制候选框的数目，选择前 k 个最大损失作为困难样例，反向传播其相应的梯度，其他候选框的梯度不进行回传，即不更新模型权重。

5.4 实验结果分析和讨论

为验证提出的自动检测算法的有效性和正确性，本节将分别采用一个公开可用的 MRI 数据集，和我们收集的来源于四川大学华西医院麻醉科的经食道超

声心动图数据集（不包含患者信息）上进行实验。相关实验代码请参考¹。

5.4.1 检测 MRI 左心室短轴

纽约大学提供的公用数据集^[138]包含 33 名患者的心脏 MRI 体数据，以及 LV 心内膜和心外膜的手动分割结果。该数据集中的大多数切片包含心脏疾的病例切片。该数据集使用 GE Genesis Signa MRI 扫描仪，采取 FIESTA 方案扫描获得。每个患者的 20 个序列帧包含 8-15 个短轴切片，大小为 256x256，厚度为 6-13 mm，像素分辨率为 0.93-1.64 mm。为了检验所提出方法定位性能，取 14 个体数据形成 1176 个切片作为训练集，其余作为测试集。本实验中不使用旋转朝向损失，评价指标采用文献 [42] 中定量评估计算左心室短轴（SAX）定位的准确度，敏感性和特异性。

Table 5.1 不同模型检测精度的比较表

Method	Accuracy	Sensitivity	Specificity
Baseline[134]	95.06%	73.91%	97.56%
VGG16	96.35%	74.68%	99.26%
ResNet101	98.66%	76.81%	99.16%
VGG16_OHEM	96.56%	79.42%	99.07%
ResNet101_OHEM	99.49%	83.12%	99.40%

为评价不同深度模型对检测效果的影响，实验的检测模型选取 VGG16^[18] 和 ResNet101^[20]，训练方法采取端到端的近似联合优化，OHEM 表明训练过程中采用困难样例挖掘方法，即在训练中只选择损失占前 70% 的样本进行反向传播。训练参数及实现跟文献 [134] 中一致，迭代次数为 1000，以文献 [134] 方法作为基准（表5.1中 Baseline），评价指标采用通用的定位精度、敏感性和特异性，结果见表5.1所示，在测试集上最优检测准确度 99.49%，敏感性 83.12%，特异性为 99.40%，与基准检测模型相比精度提高了超过 3%，同时提高了约 1.5% 的特异性。

另一方面，敏感性是最容易提高的参数，平均超过 8%，模型不能正确定位为大尺寸的心脏，导致较小 LV 切片的高 FP，降低了整体系统性能。而困难样例挖掘的方法没有显著提高特异性，因为 TN 和 FP 都降低。考虑到心脏异常的高变异性导致心脏形状的大变异性，所提出的算法均能成功定位 LV 短轴，当检测

¹<https://github.com/taopanpan/echodetection>

出心室短轴时，可大致确定心室中心点（如图5.4(a)所示），利用二腔心(2CH)和四腔心切面(4CH)均垂直于短轴切面的先验，找到与 SAX 的 2CH 和 4CH 交集在 SAX 平面上投影，然后得到投影线在 2D 图像上相交的位置，即为左心室的 3D 位置（如图5.4(b)所示）。

5.4.2 检测左心室及其朝向

MRI 左心室短轴的检测由于组织结构相对简单，且噪声少。为验证提出算法的通用性，针对超声图像左心室长轴切面检测心室、二尖瓣环、心内膜垫和心尖位置，并估计左心室朝向。主要包含单扇形和多普勒成像的双扇形两种由专业医师标注食管中段四腔心(ME4C)的标准切面视频构成，视频中至少包含 2-3 个心动周期，依据医师建议从视频中截取 5 帧，并经医师检验手工筛选后得到 900 张 ME4C 切面，对切面内左心室(LV)，二尖瓣环、心内膜垫和心尖位置进行人工标注作为“金标准”。其中随机选取 100 张作为测试集，其余作为训练集。训练时采用提出的联合多任务损失，以 VGG16 网络作为检测的预训练的模型为例，在 RPN 中添加空间变换网络实现了各个候选框的空间变换，并施加旋转朝向损失。VGG16 网络特征提取器包括 13 个卷积层，并输出 512 个 conv5 特征图，空间变换网络包括具有两个同样卷积池化层组成的定位网络，其由 20 个卷积核大小为 5、步长为 1 和核大小为 2 的池化层构成，两层全连接层回归得出 6 个仿射变换参数，其中，全连接层的激活函数需选择为双曲正切函数，权重高斯初始化，而变换参数初始化为 $[1, 0, 0, 0, 1, 0]^T$ 。其它跟 Faster RCNN 中设置一致，其中 λ 、 μ ，分别取 0.1 和 0.001；训练方法采取端到端的近似联合优化，迭代轮数为 50000。评价指标采用平均检测精度(mean average precision, mAP)，是多个类别平均检测精度的平均值。表二显示使用提出方法分别在 VGG16 模型和 ResNet101 模型上，结合困难样例挖掘训练方法得出的测试结果，其中 OHEM 表示相应模型结合在线困难样例挖掘方法的检测结果，STN 表示结合提出带朝向损失的空间变换网络的检测结果，在测试集上，针对左心室的 AP 最优可达 99.12%，结果表明提出算法在不同基础模型上均可提高检测精度。

为验证提出算法在检测左心室位置的同时可以回归学习左心室的姿态参数、预测左心室的朝向变换，超参数 k 跟文献 [136] 一致，交叠比大于 0.5 时估计姿态参数，人为标定心室朝向存在较大偏差，但可以根据二尖瓣环、心内膜垫和心尖位置估算出心室朝向角度作为对照。由于 ME4C 切面中心室的大概朝向的分布范围在 $[-45^\circ, 45^\circ]$ 之间，通过手工构建训练集，训练样本旋转以 15° 为间隔的指

Method	MAP	lv	apx	left	right
VGG16_OHEM	80.11%	90.12%	65.46%	81.27%	83.53%
VGG16_OHEM_STN	82.05%	90.92%	66.57%	86.16%	84.46%
ResNet101_OHEM	83.06%	95.72%	66.39%	85.25%	84.83%
ResNet101_OHEM_STN	85.59%	99.12%	67.89%	87.66%	87.48%

Table 5.2 不同模型的检测精度表，其中 LV 表示左心室，Apx 代表心尖，left 代表二尖瓣环，right 代表心内膜垫

定角度。通过分析相关估算结果和预测结果，可以发现二者具有很大的一致性。左心室检测结果和旋转朝向结果见表5.2，检测结果如图5.4(c,d) 所示，更多实验结果请参考给定开源地址。

Method	-45°	-30°	-15°	+15°	+30°	+45°	Avg
Compute	66.65%	78.02%	87.39%	85.53%	75.83%	62.31%	75.94%
Pred	73.09%	81.72%	81.75%	89.56%	80.48%	70.31%	80.76%

Table 5.3 不同旋转角度分类检测性能比较，Compute 表示根据额外标记计算得到的结果，Pred 表示模型预测结果

为了更详细地评估模型性能，使用检测分析工具^[2] 分析了心尖位置的检测结果，如图5.3显示模型可以准确（白色区域）检测到心尖位置，召回率在 84-87% 左右，并且比“弱”标准（小于 0.1 交叠比）高得多。针对心尖位置的定位精确度较低，这是因为医师在标定心尖位置时有很大的随意性，且目标尺寸较小，与类似对象类别有更多的混淆。

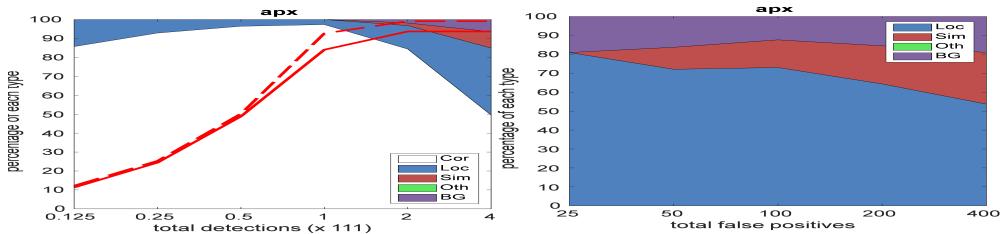


Figure 5.3 利用检测分析工具^[2]的检测结果, 左图显示 apx 检测精度的累积分布: 正确的 (Cor) 或定位不准确 (Loc) 的假阳性, 与之混淆类似类别 (Sim) 与其他类别 (Oth) 或背景 (BG)。固体红色线是以“强”标准 (大于 0.5 交叠比), 反映精确度随检测增加而变化。红色虚线使用“弱”标准 (大于 0.1 交叠比)。右图显示排名靠前的假阳性类型的分布。

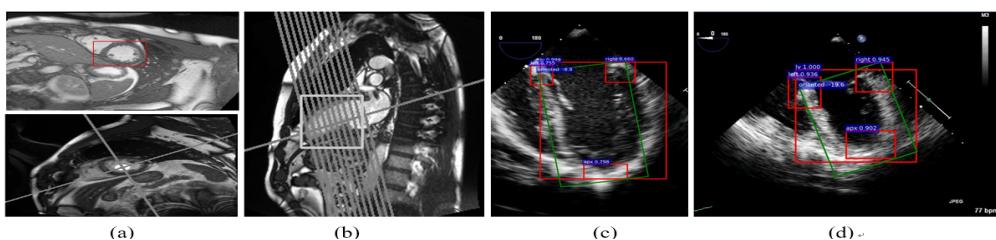


Figure 5.4 (a,b) 表示不同 MRI 图像检测左心室结果, (c,d) 两图表示超声心动图的 ME4C 切面的左心室、二尖瓣环、心内膜垫和心尖位置及旋转角度的检测结果。

5.5 小结与讨论

本文利用深度学习来解决医学图像计算机辅助检测问题, 设计并验证了自动检测 MRI 短轴和超声心动图中 LV 长轴切面的方法, 在通用物体检测 Faster RCNN 框架的基础上, 针对 RPN 引入空间变换, 结合带朝向损失的多任务损失, 探索解决图像平面内物体旋转角度检测的问题, 并利用困难样例挖掘策略加快迭代训练。在公共 MRI 数据集和自主收集的超声心动图数据上进行详尽实验验证, 在多个评估指标方面提供更好的测试结果, 但该方法仍耗费较多的标注数据, 探索需要更少标注数据的检测算法是将来的工作目标。本章研究内容和主要成果以题为《基于深度学习的医学计算机辅助检测方法研究》的论文形式已发表在学术期刊上^[139]。

第 6 章 特征表示的底层语义应用

6.1 图像的去噪方法

图像去噪的目的是从噪声图像中恢复出原图像，这是低级视觉任务中的一个经典问题。由于无噪声的原图像通常是未知的，因此这个问题本质上是不适当的，即它是一个欠定的映射问题，通常映射变换不唯一。一般来说，残差图像 F 可以表示为 $F = y - f(x)$ ，其中 x 是噪声图像， f 是映射函数，它接收输入图像并将其转换为输出图像。 y 是理想的无噪声图像。通过应用不同类型的映射函数，相同的数学模型适用于大多数其他低层级的视觉图像任务，如图像去模糊，去马赛克和超分辨率等。

最近，无论是从高级到低级视觉任务，深度神经网络在计算机视觉领域表现出了卓越的性能。众所周知，神经网络能够将任何可测量的函数逼近到期望的准确度^[62]。在图像去噪场景下，回归框架中的神经网络试图在某些输入噪声分布下逼近潜在条件期望。当以监督方式训练前馈神经网络时，一个关键因素是选择损失函数来测量输出和真实图像之间的差异。最广泛使用的是像素损失，其通过失真图像和参考图像像素的强度差异以及相关的峰值信噪比^[140]来计算。但是，像素损失不能捕捉到感知差异，并且与人类感知图像质量的关联性很差^[141,142]。这是因为当使用像素损失时隐含地做出的许多假设不能被满足-其主要仅依据于图像的局部灰度特征来处理噪声；相反，人类视觉系统对噪声的敏感性取决于局部亮度，对比度和结构等综合因素^[140]。

基于纯粹的学习策略，为图像去噪设计一组深度神经网络已被证明优于其他被广泛采用的传统方法^[143]。但是，所有这些基于学习策略的工作都存在一个问题：如果输入无噪图像，则学到的模型仍会降低并没噪声的原图像的质量。所以他们工作就是必须限定在给定噪声水平下才有效。用于去除噪声的通用标准算法应该能够处理不同级别的噪声水平，针对具体不同噪声水平需要一系列网络，通常这是不切实际的，甚至是不现实的，因为我们不知道真实图像具体的噪声水平和类型。

我们的主要贡献简要概述如下：

1. 我们提出了一个深层的全卷积网络结构，用于学习图像残差，针对图像变换对残差映射函数进行建模，直接学习噪声分布。

2. 通过对像素和感知损失函数的优缺点，训练具有融合底层和高层特征表示信息的变换网络，高效地得到高质量的去噪图像。
3. 为使单个神经网络适用于不同噪声水平，基于网络的统计概率估计，对输入不添加噪声，迫使学习到的网络模型对干净的图像不再进行降级处理；添加不同噪声水平的随机噪声送入网络使得网络模型能针对所有噪声水平进行去噪。
4. 在基准数据集上进行详尽实验，结果表明所采用的新型损失层改进了当前仅基于像素损失的去噪方法。

6.1.1 相关工作

针对图像去噪当前已经提出了相当多的方法，一些方法有选择地平滑噪声图像的部分区域，目的是在保留图像细节的同时“平滑”噪声。一些方法将图像信号变换到可以容易地从信号中分离出噪声的变换域。最近一些方法利用图像的“非局部”统计：基于相同图像中的不同区域在外观上通常相似的假设，提出块匹配的3D滤波(BM3D)算法^[144]，通过协作滤波在变换域中对非局部相似区域块进行分组，BM3D已经成为自然图像去噪的基准测试方法。

虽然手工设计的BM3D是一种高效算法，但基于学习的方法已经广泛用于图像去噪。神经网络方法和其他去噪方法最显著的区别在于，它们通常自动直接从带噪声图像中学习图像变换，而不是依赖人类先验。最近，由于深度神经网络的快速发展，许多新型的神经网络已经应用于图像去噪问题，如堆叠式稀疏自动编码器^[145–149]，多层感知器^[143,150]，深度卷积网络^[141,151–156]。

堆叠式自动编码器^[145]建立了使用去噪标准作为无监督目标的代价函数，以指导学习高层次语义级别的特征表示。去噪性能可以很容易地测量和直接优化。但是这种方法的目标是分类，Xie等人^[146]提出了一种替代的监督训练方案-堆叠式稀疏降噪自动编码器，该方法将稀疏编码和深度网络结合起来，用去噪自动编码器进行预训练，成功地将最初设计用于无监督特征学习的自动编码器，变为适用于图像去噪和缺失自动补全任务。

Burger等^[143]提出了一种基于具有普通多层感知器去噪算法，通过在大型数据集上训练学习，其性能优于BM3D。然而，他们的方法适合于特定水平的噪声，并不能很好地推广到其他噪声水平。Jain等人^[151]提出了结合深度卷积神经网络和无监督学习过程，发现卷积网络比基于小波和马尔科夫随机场方法具有更好的去噪性能。

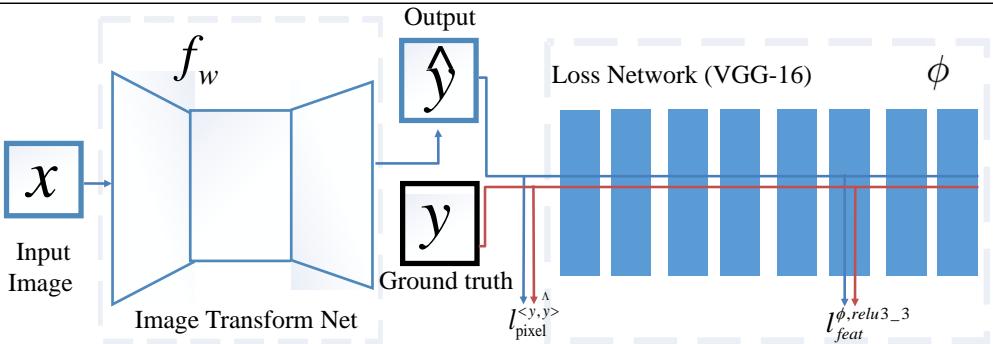


Figure 6.1 提出的网络的整体架构示意图。图像变换网络包含卷积（编码器）和解卷积（解码器）层。我们使用预先训练的图像分类的损失网络的特征表示来定义感知损失函数，这些函数测量输出和真实标签的感知差异，损失网络的特征表示在训练过程中保持不变。

迄今为止，通过在大数据集上训练具有像素损失函数的深度神经网络用于图像去噪任务，但可能会遭受一般性问题，即像素损失与感知图像质量的相关性很低^[141]。最近有论文已经在图像生成领域，使用感知损失来优化感知目标的图像的质量，其取决于从卷积网络提取的高级特征之间的伽马矩阵的相似性^[157]。文献 [153, 158] 的工作与我们的工作特别相关，其训练前馈神经网络以学习图像变换，使用预先训练用于图像分类的损失网络来定义感知损失函数，以测量输出和真实标签的感知差异。不过他们专注于风格转换和图像超分辨率。我们的网络应用具有编码器-解码器结构的卷积和反卷积层，并且使用对称跳跃连接来加速收敛速度，图像的噪声通过卷积来捕获，并通过解卷积来恢复无噪声图像的细节，可以看作是学习具有对称跳跃连接的变换函数。

Zhao 等人^[141] 研究了包括感知损失在内的多种损失的表现，并提出了一种新颖的可微分误差函数，从感知目标设计出一些新的损失层。Wang 等人^[150] 也与我们的研究特别相关，他们研究了自然图像块在线性变换方面的分布不变性，他们展示了如何使一个现有的深度神经网络学习多级别的高斯噪声分布。然而，与上述方法不同，本文通过训练具有感知损失函数的前馈变换网络，基于学习优化的图像变换方法，同时通过显式训练不同级别的噪声并使原始图像同样作为输入，使单个深度神经网络在不同级别的加性高斯白噪声下工作良好。

6.1.2 提出的方法

所提出的框架主要包含一系列卷积层和反卷积层，如图6.1所示。它由两部分组成：一个图像转换网络 f_w 和一个损失网络 ϕ ，用于定义几个代价损失函数 ℓ_1, \dots, ℓ_k 。旨在学习由权重 W 参数化的深度残差卷积神经网络；它通过映射 $\hat{y} = f_w(x)$ 将输入图像 x 转换成输出图像 f_w 。每个损失函数计算标量值 $\ell_i(\hat{y}, y_i)$ ，

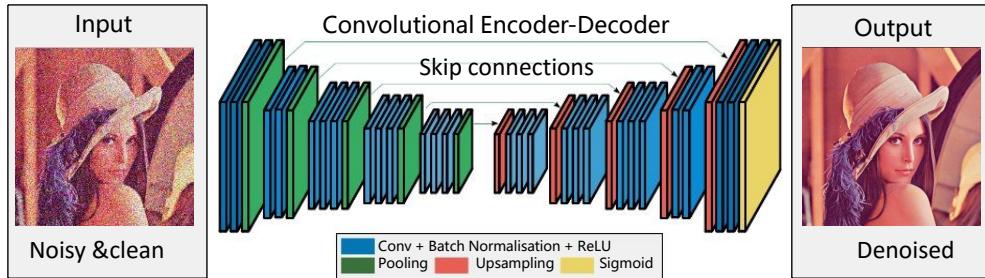


Figure 6.2 RED-NET 的网络结构示意图

测量输出图像 \hat{y} 和目标图像 y_i 之间的差异。学习目标是使用随机梯度下降来训练以最小化损失函数的加权组合：

$$W^* = \arg \min_W E_{x,\{y_i\}} \left[\sum_{i=1} \lambda_i \ell_i(f_W(x), y_i) \right] \quad (6.1)$$

从这个公式中，我们可以看到优化算法的任务是找到最接近图像变换的映射函数 f_W 。同时我们也希望 $f_W(y_i)$ 近似于图像 y_i ，所以我们现在通过在不同情况下选择合适的权重 W 来将图像去噪问题统一在一个统一的框架中。损失网络 ϕ 用于定义一个特征空间损失 ℓ_{feat}^ϕ 和一个像素损失 ℓ_{mse} 。对于图像去噪，输入图像 x 是一个有噪声的输入，真实图像 y 是目标图像。

6.1.2.1 编码器-解码器卷积结构

该框架完全是编码器-解码器卷积模型，NohH 等 [72,153,159–161] 已经提出了用于无监督和有监督深度学习的编码器-解码器的神经网络。

我们的图像转换网络大致遵循 [153] 提出的网络架构 RED-NET 用于图像去噪，网络结构如图6.2所示。基于该网络结构，在卷积之后添加批量归一化 [64] 和 ReLU 非线性层；并在网络中插入一些残差网络的模块 [20]，输出层使用 sigmoid 函数来确保输出图像的像素范围为 $[0, 1]$ 之间，但是我们不使用任何池化层降低分辨率，而是使用不同步长卷积来进行下采样和上采样。除了第一层和最后一层使用 9×9 的卷积核之外，所有卷积层都使用 3×3 的卷积核。由于图像变换网络是完全卷积的，因此在测试时它们可以应用于任何尺寸的图像输入。

RED-NET 和我们的去噪网络（DeNET）的区别在于我们的网络插入了一些残差模块并在不同模块间引入了跳跃连接。在图像去噪的过程中，图像内容的细节可以通过感知损失函数进行补偿，两个网络的具体参数配置在表 6.1 中描述。

将两种学习策略应用于编解码网络的内部模块以使训练更有效，跳跃连接每两个卷积层传递到它们的镜像解卷积层。He 等 [20] 使用残差连接来训练非常深的网络进行图像分类，他们认为残差连接使网络很容易学习具有很多层的网络模

Table 6.1 DeNET-R 和 RED-NET 网络的参数配置。“conv3”和“deconv3”代表大小为 3×3 的卷积和反卷积核。32,128 和 512 是每个卷积和反卷积之后的特征映射的数量。“ c ”是输入和输出图像的通道数量。

DeNET-R	RED-NET
$(\text{conv9-32}) \times 6$	$(\text{conv3-128}) \times 6$
$(\text{conv3-64}) \times 6$	$(\text{conv3-256}) \times 6$
$(\text{conv3-128}) \times 3$	$(\text{conv3-512}) \times 3$
Residual block $\times 5$	
$(\text{deconv3-64}) \times 2$	$(\text{deconv3-512}) \times 2$
$(\text{deconv3-32}) \times 6$	$(\text{deconv3-512}) \times 6$
$(\text{deconv9-3}) \times 6$	$(\text{deconv3-512}) \times 6$
$(\text{deconv3-}c)$	$(\text{deconv3-}c)$

型（如 152 层）；这对于底层任务的图像变换网络来说是一个吸引人的特性，因为在大多数情况下，输出图像应该与输入图像共享大部分内容结构。因此，我们网络的主体由多个残差块组成，每个残块包含两个 3×3 大小卷积核的卷积层。

6.1.2.2 像素损失函数

像素损失是目标 \hat{y} 和输出图像 y 之间的（归一化）欧几里德距离。若二者大小为 $C \times H \times W$ ，那么像素欧几里得损失定义为均方误差（MSE）：

$$\ell_2(\hat{y}, y) = \frac{1}{CHW} \|\hat{y} - y\|_2^2 \quad (6.2)$$

这种损失函数可能会引入网格棋盘效应，因此可引入了 ℓ_1 范数正则化损失加以缓解。这两种损失对像素错误的权衡是不同的： ℓ_1 不会过度惩罚更大的错误值，因此它们可能具有不同的收敛性质，计算 ℓ_1 损失很简单：

$$\ell_1(\hat{y}, y) = \frac{1}{CHW} |\hat{y} - y| \quad (6.3)$$

其应用反向传播的求导也是很简单的，对整个图像的每一像素 p 来说，

$$\partial \ell_1 / \partial p = \text{sign}(\hat{y}(p) - y(p)) \quad (6.4)$$

其中，在整个图像上计算 ℓ_1 的导数将针对图像中的每个像素进行反向传播。

6.1.2.3 感知损失函数

为了测量图像之间感知语义差异提出感知损失函数，跟文献 [141] 中提出的手工设计结构相似性（SSIM）损失不同，可利用用于图像分类的预先训练网

络作为损失网络 ϕ 。在我们所有的实验中， ϕ 都是 16 层 VGG 网络^[18]，其是在 ImageNet 数据集^[17]上预先训练的。并不是鼓励输出图像 $\hat{y} = f_W(x)$ 的像素完全匹配目标图像 y 的像素，而是鼓励它们具有由损失网络 ϕ 定义计算的相似特征表示。让 $\phi_j(x)$ 为图像 x 的第 j 层卷积网络 ϕ 的激活值；然后 $\phi_j(x)$ 将是卷积层形状大小为 $C_j \times H_j \times W_j$ 的特征映射。特征表示损失是特征表示之间的（归一化平方和）欧几里得距离：

$$\ell_{feat}^{\phi,j}(\hat{y}, y) = \frac{1}{C_j H_j W_j} \|\phi_j(\hat{y}) - \phi_j(y)\|_2^2 \quad (6.5)$$

欧几里德距离又称为 ℓ_2 范数距离，如文献 [98] 中指出最小化网络初级特征相似性，可以寻找生成图像 \hat{y} ，往往会产生与 y 一致的图像。使用特征损失来训练我们的图像转换网络输出图像 \hat{y} 与感兴趣的图像 y 相似，但不会强制它们完全匹配。为了鼓励输出图像 \hat{y} 的空间平滑度，遵循之前关于特征重建^[98]的工作，并使用变分正则化损失函数 $\ell_{TV}(\hat{y})$ 。

6.1.3 实验结果分析和讨论

在本节中，使用编码器-解码器的卷积神经网络对我们各种实验设置进行分析，然后评估我们的模型在一些不同的损失函数组合设置下的去噪性能。最后，我们探讨如何让一个神经网络处理不同程度的噪声。

6.1.4 详尽分析模型

我们通过同时最小化多个损失函数： ℓ_2 -均方误差 (MSE) 损失， ℓ_1 -范数损失和特征损失，在具有不同标准偏差 σ 的噪声图像对来训练模型。训练集中图像大小为 256×256 ，并通过添加 σ 大小的高斯噪声生成带噪声和无噪声图像对，使用 Adam^[86] 的自适应优化算法进行批随机梯度下降算法进行训练，学习速度为 1×10^{-3} ，不使用权重衰减或 Dropout 等正则化技术。

去噪实验在 14 个三通道彩色图像的通用基准测试图像 *Set14* 上执行，将零均值和标准差 σ 的加性高斯噪声添加到测试图像中，以测试去噪方法的性能，评价

Table 6.2 *Set14* 上的单一噪声水平的图像去噪结果定量比较表

Sigma	Noisy PSNR / SSIM	RED-NET ^[153] PSNR / SSIM	Ours (ℓ_2) PSNR / SSIM	Ours (ℓ_1) PSNR / SSIM	Ours (ℓ_{feat}) PSNR / SSIM	Ours (ℓ_{mix}) PSNR / SSIM
$\sigma = 10$	28.16 / 0.7041	34.81 / 0.9402	34.35 / 0.8912	33.40 / 0.8930	31.05 / 0.7680	33.16 / 0.7680
$\sigma = 30$	18.88 / 0.3389	29.17 / 0.8423	28.73 / 0.8205	29.76 / 0.991	26.70 / 0.6845	30.15 / 0.8681
$\sigma = 50$	14.79 / 0.2038	26.81 / 0.7733	26.40 / 0.8205	26.79 / 0.8325	25.69 / 0.6411	27.09 / 0.8312
$\sigma = 70$	12.43 / 0.1391	25.31 / 0.7206	25.39 / 0.7105	26.13 / 0.7250	17.89 / 0.6650	26.20 / 0.7180
$\sigma = 100$	10.26 / 0.0901	-	18.40 / 0.4215	20.17 / 0.4680	17.31 / 0.3640	19.16 / 0.4695

	Noisy image,sigma= 50	ℓ_2 loss denoised image	ℓ_1 loss denoised image	feat loss denoised image	best image,sigma= 50
Ground Truth					
PSNR / SSIM		29.11 / 0.8833	29.27 / 0.8841	19.61 / 0.6560	29.31 / 0.8946

Figure 6.3 不同损失在基准数据集上的定量比较结果表，在来自 *Set14* 数据集的图像上以不同的损失类型去噪结果。我们以 F-16 图像的 PSNR / SSIM 为例。

指标选取参考文献^[141,153] 中的峰值信噪比(PSNR)和结构相似性度量(SSIM)^[140]。

作为基准模型，我们使用 RED-NET^[153] 作为比较对象，它是一个全卷积网络，具有卷积和反卷积层，损失函数为像素损失。为了说明 RED-NET 和我们的模型在数据，训练和网络结构方面的差异，我们使用 ℓ_2 对相同的标准偏差 σ 进行图像变换网络训练，在使用像素损失的基础上进行消融实验，添加特征损失函数（见 Section6.1.2），以允许从预训练损失网络传输语义知识作为有监督的信号指导去噪的去噪网络。

首先，与像素损失 ℓ_1 和 ℓ_2 结果相比， ℓ_1 在去噪性能方面做得更好，同时还原了锐利的边缘和细节，但存在网格棋盘效应。如图6.3所示， ℓ_2 图像中的机翼以及 ℓ_2 图像中的主体的红色块元素。这是因为 ℓ_2 惩罚更大的错误值，无论图像中的底层结构如何，都会产生小的错误，其结论与文献 Zhao2015 一致。

此外，若只有特征损失时，在放大数倍分辨率下才能看到轻微的网格棋盘效应，结果为 ℓ_{feat} 如图6.3所示，与基准方法相比，会损害其 PSNR 和 SSIM。我们再次看到，与其他方法相比，我们的 ℓ_{feat} 模型在边缘和细节方面做得很好，比如机翼。 ℓ_{feat} 损失不会不加区分地锐化，与 ℓ_{pixel} 损失相比， ℓ_{feat} 损失锐化了机翼和骑手的边界边缘，但背景仍然是漫反射的，这表明 ℓ_{feat} 损失可能是更了解图像语义。

由于采用的 ℓ_{pixel} 和 ℓ_{feat} 损失共享相同的架构，数据和训练过程，它们之间的所有差异都是由于 ℓ_{pixel} 和 ℓ_{feat} 损失。 ℓ_{pixel} 损失产生较少的视觉伪像和较高的 PSNR 值，但 ℓ_{feat} 损失在重建细节方面做得更好，从而导致令人满意的视觉效果。

最后，我们可以观察到单个损失可以在所有级别的高斯噪声中工作，从而允

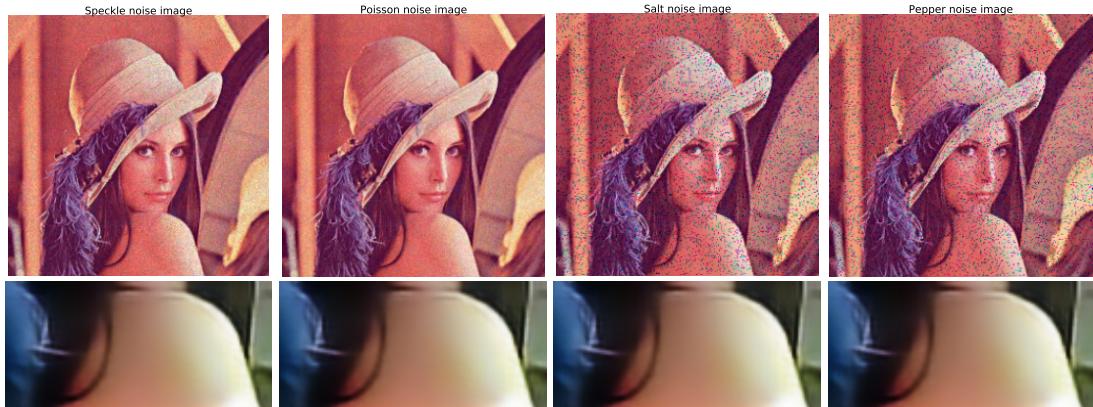


Figure 6.4 四种其他噪声类型的去噪性能比较图，其中变换网络仅用高斯噪声进行训练。

Up: 四种类型的噪声图像：散斑噪声，泊松分布噪声，椒盐噪声，胡椒噪声。 **Down:** 从相应的噪声类型中去噪输出。

许显著减少通用神经网络驱动的去噪算法的训练时间。原因可能是学习图像变换函数可以从任何级别对高斯分布进行建模。如图6.4所示，即使对于其他类型的噪声，如斑点噪声，泊松分布噪声，椒盐噪声或胡椒噪声等，也有较好的去噪能力。

6.2 形状对齐的心室的分割方法

医学图像分割着重提取具有特殊含义的区域，如组织、肿瘤等，并使分割结果尽可能地接近解剖结构。进而辅助医生进行病情分析，诊断及制定治疗方案。如超声心动图可用于评估心室功能的各项参数，如左室容积、射血分数和行程容积等，其定量分析优于定性解释，特别是对于室壁运动和心室体积的估计。然而当前许多方法需指定初始输入，需要专家知识，如需手动勾勒短轴横截面，手动分析很耗时，也取决于观察者主观分析。自动或半自动分割算法是目前进行客观评价所必需的工具。目前虽已研究出各种分割方法，至今还没有一种能够统一适用于各种图像及不同部位的有效方法。由于解剖结构的个体差异较大，分割对象结构性质的千差万别；又由于噪声、伪影和容积效应等影响，使得已有分割算法远未达到理想效果。同时因无法完全用数学模型来简单描述所面临的问题；人们对分割结果预期目标互不相同等原因，只能针对特定问题和具体的需求，在精确度、鲁棒性和效率等关键指标上做出权衡^[162]。

Hansson 等^[163] 提出了贝叶斯概率图模型对心内膜概率进行建模分析，该方法使用左心室和心房相对位置的先验知识。基于能量泛函的活动轮廓及其扩展的水平集方法，如 Marsousi 等^[164] 提出了一种，结合外力和采用多分辨率策略使用

B 样条自适应活动轮廓模型应用于超声心动图左心室心内膜分割。然而这些技术对初始化和参数选择非常敏感。在现有分割方法中，统计形变建模是用于可视化器官变化几何和功能模式的有效工具^[165]，典型建模的方法有可变形模板、点分布模型、图模型等。其分割是在有限的变化范围内进行的，变化范围通常由已知形状来定义。

统计形变模型是医学图像分割任务常用方法，其中表观建模又可分为全局和局部表观建模。基于局部表观的主动形状模型（Active Shape Model, ASM）^[166]和基于全局外观主动外观模型（Active Appearance Model, AAM）^[167]用于超声心动图分割已被证明是非常有效的^[162,168,169]。原始 ASM 在超声图像中存在许多缺陷^[165]，因为它基于边缘灰度特征，也无法解决边缘缺失问题，局部受限模型^[170]（Constrained Local Model, CLM）引入特征点局部区域外观模型加以改进。而 AAM 适合于 2D 和 3D 超声心动图中对左心室的复杂外观建模，因为它具有描述形状和图像强度的典型变化（包括伪影）的能力^[171]。

近来，级联形状回归模型^[172]在特征点定位任务上取得较大突破，该方法使用回归模型，直接学习从表观到形状（或者形状模型的参数）的映射函数，进而建立从表观到形状的对应关系。此类方法不需要复杂的形状和表观建模，简单高效，在可控场景和非可控场景均取得不错的对齐效果。此外，基于深度学习的特征点定位方法^[52]也取得令人瞩目的结果。深度卷积神经网络结合形状回归框架可以进一步提升定位精度。但是基于级联形状回归和深度学习方法一般需要的数据量较大，不能直接适用于医学图像分割场景。

现有心室分割方法很少考虑心室的检测问题^[173]，默认操作是将平均形状手动放置于感兴趣区域，这导致最后的分割结果受初始位置影响较大。针对以上问题，我们提出一种基于沙漏卷积网络特征的多尺度形状对齐方法应用于超声心动图的左心室分割，在几个量化评价标准上的结果表明我们方法的有效性。

本工作提出的主要贡献如下：

- 初始阶段，提出利用物体检测算法准确检测左心室位置，为后续分割自动化放置初始轮廓提供辅助，并构造心室分割数据库以评价算法，且针对训练深度卷积网络提出了扩充数据样本的方法。
- 提出利用全卷积神经网络学习外观和局部特征，构造多级沙漏卷积网络自动提取的特征融合了多种注意力图的上下文信息，实验详细比较了不同特征激活图的分割效果，在超声心动图心室分割任务上验证了基于深度学习

的方法优于传统手工设计的特征。

- 综合分析了多种特征外观纹理和多种特征激活图，并克服 AAM 和 CLM 算法的缺点，利用各自的概率解释去统一全局 AAM 和局部 CLM 算法，得到最优的心室分割效果。

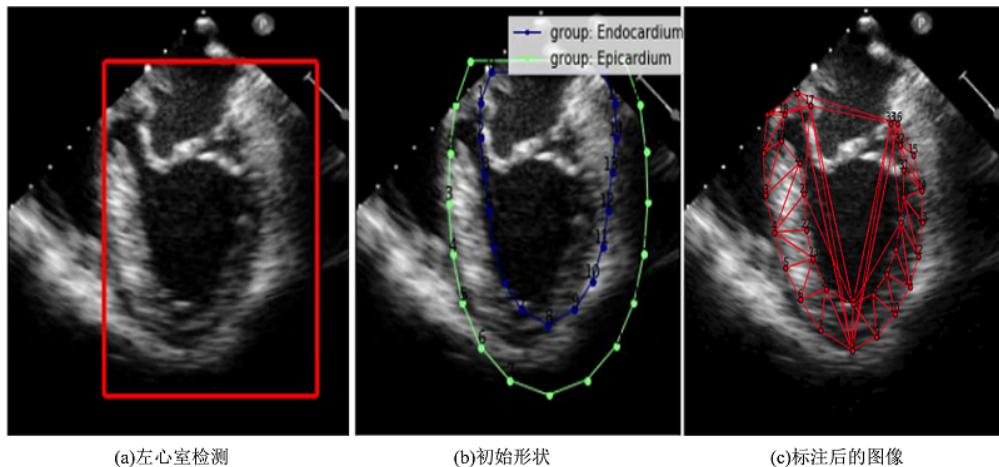


Figure 6.5 初始位置定位结果和特征点标注示意图

6.2.1 初始位置定位和特征点标注

检测左心室为下一步的分割和参数自动提取提供定位结果时，并未采用基于哈尔特征的稀疏积分图，结合提升回归分类器^[174]和标注数据，将扫描窗口中外观映射为位移矢量，学习回归函数的方法。而是针对形变问题，基于图形结构的变形部件模型，使用梯度直方图 (Histograms of Oriented Gradients, HOG) 特征^[14]，结合线性支持向量机分类器和滑动窗口检测思想，对左心室进行检测。在实验数据上能 100% 检测到左心室位置，检测结果如图6.5(a) 所示，其中形变部件模板如图6.6(a) 所示，能清晰看出内外膜轮廓。

斑点噪声和伪影的存在，使得难以定义一组生理上一致的特征点（不能表示相同的区域），从而难以构建有意义的统计表观模型。左心室特征点的标注同文献 [173] 中一致，其中 Centripetal Catmull-Rom 曲线能够在减少特征点数量的同时得到形状一致的特征点，选用了 34 个特征点。如图6.5(b)，外层曲线表示心外膜 (0-16)，内层曲线表示心内膜 (0-16)，图像的标注后的图像和生成纹理时的三角网格如图6.5(c) 所示。

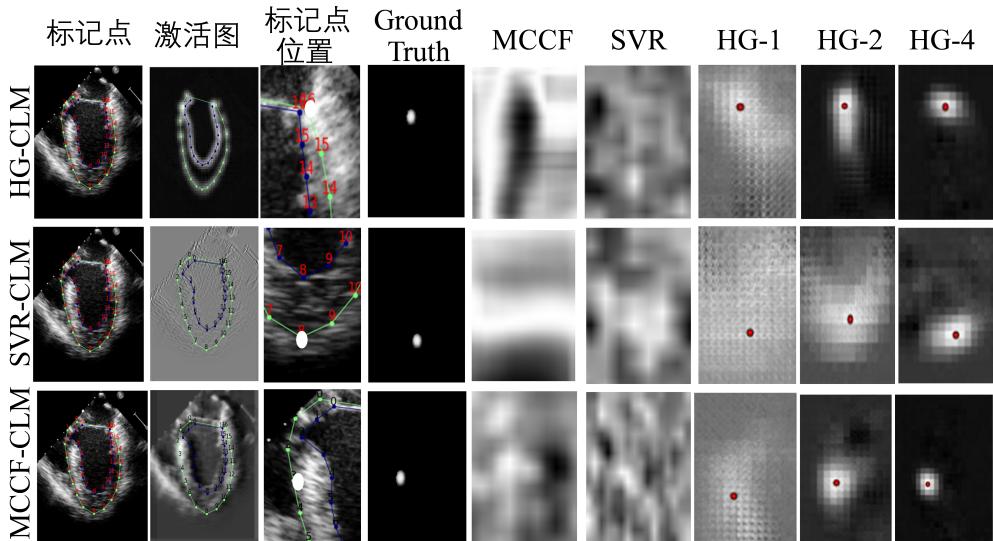


Figure 6.6 定性比较三种不同特征激活图及相应的局部响应映射图, MCCF 通过多通道相关滤波器近似响应图, 且用 RLMS 算法移动到最优位置。SVR 基于支持向量机简单地选择最大响应位置。HG-n 表示所用不同 HG 模块数的局部响应图, n 取 1, 2, 4。

6.2.2 AAM 模型和 CLM 模型

6.2.2.1 基于 AAM 的分割

AAM 是常用医学图像分割方法之一, 是用来解释特定对象形状和外观视觉变化的参数生成模型。设 m 个图像内标记点集合的坐标 $v_i = (x_i, y_i)^T \in R^2$, 则第 k 个形状向量可定义为: $s_k = (x_1, y_1, \dots, x_n, y_n)$ 。AAM 对新图像进行分割时, 拟合策略通常被构造为最佳形状 p 和纹理 c 参数的正则化搜索过程。最小化参数同时依赖于所有标记点位置全局测量偏差:

$$\begin{aligned} p^*, c^* &= \arg \min_{p, c} R(p, c) + D(i[p], c) \\ &= \arg \min_{p, c} \|p\|_{\Lambda^{-1}}^2 + \|c\|_{\Sigma^{-1}}^2 + \frac{1}{\sigma^2} \|i[p] - \bar{a} + Ac\|^2 \end{aligned} \quad (6.6)$$

式中 R 是惩罚形状和纹理变形的正则化项, D 是量化给定全局测量偏差的数据项。 Λ 和 Σ 为对角矩阵包含与形状和纹理特征向量相关联的特征值, σ^2 是图像噪声估计。原始匹配算法使用的是线性回归方法。可以通过假设以下的形状和纹理的概率生成模型来获得式6.6的概率解释 [16,17]:

$$\begin{aligned} s &= \bar{s} + Sp + \epsilon p \sim N(0, \Lambda)\epsilon \sim N(0, \sigma^2 I) \\ i[p] &= \bar{a} + Ac + \epsilon c \sim N(0, \Sigma)\epsilon \sim N(0, \sigma^2 I) \end{aligned} \quad (6.7)$$

式6.7为形状模型和外观模型的概率解释, 其假设 ϵ 服从零均值, σ^2 方差的高斯分布。给定模型参数 $\Theta = (\bar{s}, S, \Lambda, a, A, \Sigma, \sigma^2)$, 可以很容易地定义最大似然过程

来推断最佳形状和纹理参数：

$$p^*, c^* = \arg \min_{p,c} \ln(i[p]|p, c, \Theta) \quad (6.8)$$

通过考虑先验分布的最大后验来估计带正则化项的最优形状 p 和最优纹理参数 c :

$$p^*, c^* = \arg \min_{p,c} \ln p(p|\Lambda) + \ln p(c|\Sigma) + \ln(i[p]|p, c, \Theta) \quad (6.9)$$

公式6.9与公式6.6定义的优化问题等价。

6.2.2.2 基于 CLM 的分割

相对于 AAM, ASM 只使用特征点边缘灰度或轮廓线模型来进行点匹配, 而 CLM 通过其形状标记点邻域内候选块来定义对象的纹理, 同时利用与 AAM 类似的全局形状作为全局约束。针对初始化形状的各个标记点, 用检测器对局部区域进行判别, 作用类似滤波器, 可获得激活得分响应图, 标记点被正确对齐与否的概率可以定义为:

$$p(l_i|x_i, I) = \frac{1}{1 + \exp l_i C_i(I, x_i)} \quad (6.10)$$

式中 $l_i \in (1, -1)$ 指示定位正确与否, C_i 是区分标记点 x_i 对齐与否的分类器, 可使用不同分类器, 例如逻辑回归^[1]、多通道相关滤波 (MCCF) 的平方误差总和最小滤波器 (MOSSE)^[2]和支持向量回归机 (SVR)^[3] 等。拟合 CLM 涉及到解决以下优化问题^[4]:

$$\begin{aligned} p^* &= \arg \min_p R(p) + \sum_{i=1}^v D_i(x_i, I, p) \\ &= \arg \min_p \|p\|_{\Lambda^{-1}}^2 + \sum_{x_i \in S} \sum_{j=1}^k \frac{w_i}{\rho^2} \|x_i - y_i\|^2 \end{aligned} \quad (6.11)$$

式中 $w_i = \frac{1}{1 + \exp l_i C_j(I, x_j)}$, Λ 是计算与形状特征向量相关联的特征对角矩阵和 ρ^2 是形状噪声估计。公式6.10可跟 AAM 一样改写为概率形式^[20]:

$$p^* = \arg \max_p \ln p(p|\Lambda) + \ln p(l_i|p, I, x_i, \Sigma) + \ln(i[p]|p, c, \Phi) \quad (6.12)$$

已经提出了不同的方法仿真模拟真实的响应映射 $p(l_i|I, x_i)$, 最常用的是 [19] 的非参数方法 (RLMS), 它将真实的响应图近似为:

$$\sum_{y_j \in \varphi_{x_i}} p(l_i = 1|I, y_j) N(x_i, \rho^2 I) \quad (6.13)$$

式中当前标记点位置 x_i 是根据先前的概率生成形状模型定义的。将 6.12 代入 6.11，得以下优化问题：

$$p^* = \arg \min_p -\ln p(p|\Lambda) - \sum_{i=1}^v \ln p(l_i = 1|p, I, x_i, \Phi) \quad (6.14)$$

这相当于由 6.11 定义的优化问题，式中响应映射在所有像素位置 $y_{j=1}^k$ ，视真正的标记点位置 y_i 作为潜在变量可评估 φ_{x_i} ，式 6.13 可以使用 EM 算法迭代地求解 [28]。

6.2.3 结合卷积网络特征的形状对齐模型

6.2.3.1 超声组织特征纹理特异性灰度归一化

形状及外观模型利用 PCA 通过计算高维椭球分布的质心和主轴来模拟多维高斯分布。在标准 AAM 灰度归一化后，像素的灰度分布或多或少是高斯分布，使得平均灰度为 0 且方差为 1。而超声心动图灰度直方图具有非高斯分布特征，直方图峰值处于非常低的灰度值，并且倾向于指数下降。这是超声图像的固有属性（尤其是斑点噪声），或多或少地独立于心室的组织类型，大致服从反指数分布或卡方分布^[162]，其宽度范围和偏移量变化很大，进一步的视频信号处理引入更多的偏移和增益变化，导致直方图峰值偏移，灰度范围可能会有很大差异。所以，在应用归一化之前，执行文献 [162] 提出的非线性归一化来处理偏斜和偏移的灰度分布。

6.2.3.2 结合不同外观特征的全局 AAM

全局 AAM 产生精确的拟合结果依赖于形状无关纹理的表示能力，对超声心动图使用图像灰度作为原始纹理来建立活动外观模型导致拟合不准确，影响分割性能。同时标注数据困难，少量数据样本的外观变化较难建模，且心室腔体和腔壁有明显不同的纹理，提出可利用 HOG 特征、稠密 SIFT 特征以及后文提出的卷积网络特征，结合多尺度活动外观模型的左心室分割方法。不同特征的形状无关纹理直接影响 AAM 分割性能，图 6.6 表示采用灰度（图 6.6(b)），hog 特征（图 6.6(c)）构建的 AAM 模型的形状无关纹理可视化结果。AAM 的参数空间的维度很大使得它们难以优化，此外还对不准确的初始化非常敏感。

6.2.3.3 CLM 中的特征激活图

CLM 算法最重要的一步是计算响应图，通过评估各个像素位置的标记点对齐概率，帮助准确地定位标记点。比较常用的多通道相关滤波 (MCCF)^[175] 和支

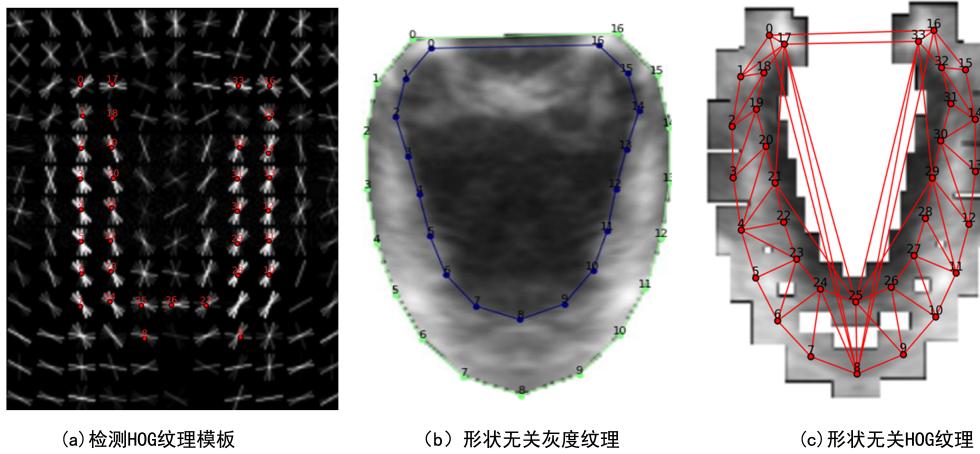


Figure 6.7 不同特征的形状无关纹理图

持向量回归机 (SVR)^[176] 的特征激活映射图可知，在超声心动图分割任务中，这些基于手工设计的特征效果差且不据有可解释性（见图6.6）。在我们的模型中，这是由堆叠多级沙漏全卷积网络 (Hourglass Network, HG)^[177] 完成的，围绕当前估计的所有标记点位置 $n \times n$ 像素区域作为感兴趣区域输入，并且输出在每个像素位置评估标记点概率响应图（见图6.6），网络结构如图6.7所示。

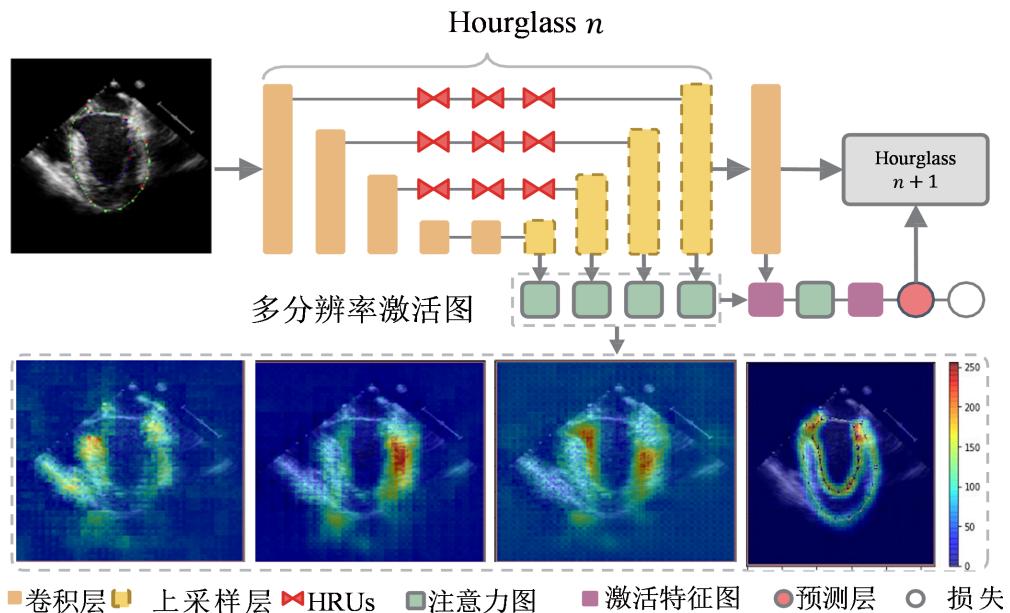


Figure 6.8 在每个沙漏网络中，从具有不同分辨率的特征生成多分辨率注意力图，这些图加和成单一的注意力图，它用于生成激活特征图。

图6.7中的网络基本组件是一种基于残差网络^[20]。“沙漏”型网络结构是拓扑对称的，能够捕获和整合来自不同尺度和分辨率的信息。如图三中卷积层为残差模块，其是 3×3 大小卷积核组成的卷积层、批归一化层和修正线性激活单

元层来提取特征，同时用跳跃连接保留原始信息的统称。所有卷积层不改变数据尺寸，只改变通道数。在最大池化（max-pooling）下采样操作之前，它分离单个通路以将当前信息保留。在上采样（反卷积或最近邻插值）操作之前，添加与原始图像大小相同的特征图。在两次下采样操作的处理之间，为获得不同分辨率的注意力图^[178]，同使用另一个残差模块提取来的特征图进行加权乘积得到激活特征图。对于 $H \times W \times 3$ 的输入图像，每一个 HG 级的激活特征图都会生成一个 $H/4 \times W/4 \times K$ 的预测概率响应图，K 表示标记点个数。对于每个响应图，都比较其与真值标记点附近高斯分布的欧式误差，作为损失函数中继监督（intermediate supervision）训练所有模块。详细的网络参数和训练过程在6.2节中给出。在公式6.13的迭代中，将感兴趣区域图像输入 HG 网络，输出了评估单个标记点对齐的概率响应图。将标记点 i 拟合到位置 x_i 遵循以下等式：

$$\pi_{x_i}^i = p(l_i = 1, I_{x_i}) \quad (6.15)$$

式中 l_i 表示第 i 个标记点，图像的位置 x_i 处的图像 π_i ，响应映射 i 用于最小化等式6.13。消融实验表明，增加 HG 模块数，显著影响分割性能。

6.2.3.4 统一 AAM 和 CLM 的概率解释

整体和局部模型之间的差异在于提取外观向量以及构建变形模型的方式不同。基于整体或局部外观表示的选择高度依赖于建模对象及其内部结构的性质。针对医学图像分割问题，局部图像特征的位置并不总是对应于由专家人类观察者绘制的期望轮廓。因此，轮廓的确切位置不能总是从最强的概率响应图来确定，而是应该由专家观察者提供的示例建模学习得到。为了结合全局和局部框架，采用一种可变形模型拟合问题的概率解释，式6.9和式6.13可以重写为以下优化问题：

$$p^*, c^* = \arg \max_{p, c} \ln p(\Lambda) + \ln p(c|\Sigma) + \ln(i[p]|p, c, \Theta) + \ln p(l_i = 1_{i=1}^v | p, I, x_i, \Phi) \quad (6.16)$$

其中 $R(p, c)$ 对应于复杂形状和纹理变形的正则化项， $D(I, p, c)$ 表示全局未对准度量，并对应于 AAM 拟合中的数据项， $\sum_{i=1}^v D(x_i, I, p)$ 表示对应于 CLM 拟合中数据项的 v 个标记点对齐的局部偏差度量。

6.2.3.5 模型匹配代价函数的优化

等式6.16可以通过反向组合用于拟合 AAM 的梯度下降算法和用于拟合 CLM 的 RLMS 算法来优化，如结合投影反向组合（PIC）算法^[176] 和 RLMS 算法，增量

形状参数 ∂p^* 的最优解由下式给出：

$$\partial p^* = -H^{-1}b \quad (6.17)$$

其中：

$$\begin{aligned} H &= \Lambda^{-1} + \frac{1}{\sigma^2} J_a^T P J_a + \frac{1}{\rho^2} J_s^T J_s \\ b &= \Lambda^{-1} p - \frac{1}{\sigma^2} J_a^T P (i[p] - \bar{a}) - \frac{1}{\rho^2} J_s^T (\mu - s) \end{aligned} \quad (6.18)$$

其中 H 是反向位置的海森矩阵 (Hessian)。 $J = \nabla \bar{A} \frac{\partial W}{\partial o}|_{p=0}$ 和 $P = I - AA^T$ 分别是反向组合雅各比矩阵 (Jacobian) 和投影运算。或通过将交替反向组合 (AIC) 算法^[176] 与 RLMS 组合：

$$\begin{aligned} H &= \Lambda^{-1} + \frac{1}{\sigma^2} J_a^T J_a + \frac{1}{\rho^2} J_s^T J_s \\ b &= \Lambda^{-1} p - \frac{1}{\sigma^2} J_a^T (i[p] - \bar{a}) - \frac{1}{\rho^2} J_s^T (\mu - s) \end{aligned} \quad (6.19)$$

在这种情况下，Jacobian 被定义为 $J = \nabla (\bar{A} + \sum_{i=1}^m c_i A_i) \frac{\partial W}{\partial o}|_{p=0}$ ，有关如何计算 $\frac{\partial W}{\partial p}$ 的更多细节有兴趣的读者请参考^[176]，最佳纹理参数 c^* 由式 5 给出，且两种算法仍然使用式 6.18 定义的完全相同的更新规则得到 ∂p^* 的最优值。

6.2.4 实验结果分析和讨论

6.2.4.1 数据集增强和评价标准

本实验采用 Philips CX50 和 IE33 所采集的带乳头肌和无乳头肌心脏四腔心经食道超声图像，共 45 个视频。专家标注 (ground truth) 由四川华西医院的麻醉科医生完成，其结果作为“金标准”。在训练过程中，我们用大致相同尺度的图像以心室为中心裁剪图像，并将图像缩放到 256x256 的大小作为输入。然后我们随机旋转、镜像翻转和缩放扩增数据集（包括图像和注释），其中需要注意的是要标注标记点有无的模版以应对标记点缺失的情况，最后扩增 10 倍获得 4240 个训练样本作为训练集，而 167 张的测试集不做任何数据扩充。实验所有方法均使用前文提出的左心室检测算法估计轮廓初始位置。评价指标采用人脸对齐任务中常用的评价标准，使用平均点对点误差归一化欧式距离 (NMSE)：

$$E_i = \frac{\frac{1}{n} \sum_{j=1}^n |x_{i,j} - x'_{i,j}|_2}{|lt_i - rb_i|_2} \quad (6.20)$$

式中表示 n 个特征点的两个形状 x_j 和 x'_j ， lt 和 rb 是真实形状边界的左上点和右下点的位置。归一化能够使性能测量与实际心室尺寸或缩放系数无关。本文采用 NMSE 的累积误差分布函数 (Cumulative error distribution, CED) 进行性能评估。

同时计算两个形状之间的距离，然后统计测试集中所有形状与专家标注形状之间的距离的均值和方差。训练 HG 网络模型我们采用 tensorflow 框架，初始学习率为 1×10^{-3} ，网络参数由 Adam 算法^[179]优化，网络中开始是步长为 2，核大小为 7×7 的卷积层，将分辨率由 256 降到 64，以减少 GPU 占用，其后是残差模块和一串下采样层组成的 HG 模块，整个网络中的所有残差模块输出特征数都是 256，相关代码见¹。本文实验采用三种方式：一是将比较不同特征的 AAM 和 CLM，以验证使用单独全局和局部模型的最优分割效果；二是，在统一 AAM 和 CLM 的条件下，比较不同特征激活图对最终分割效果的影响；三是，在同样使用 HG 网络特征的条件下，将使用的 HG 模块数设为 1、2、和 4，比较不同数值下的分割效果。

6.2.4.2 不同特征的 AAM 和 CLM 分割结果

实验中，选取三个尺度的 AAM 模型，变形扭曲函数选择的是薄板样条曲线映射扭曲函数，平均形状作为参考形状获得形状无关纹理，优化算法统一为 PIC，每个尺度最大拟合 30 步。

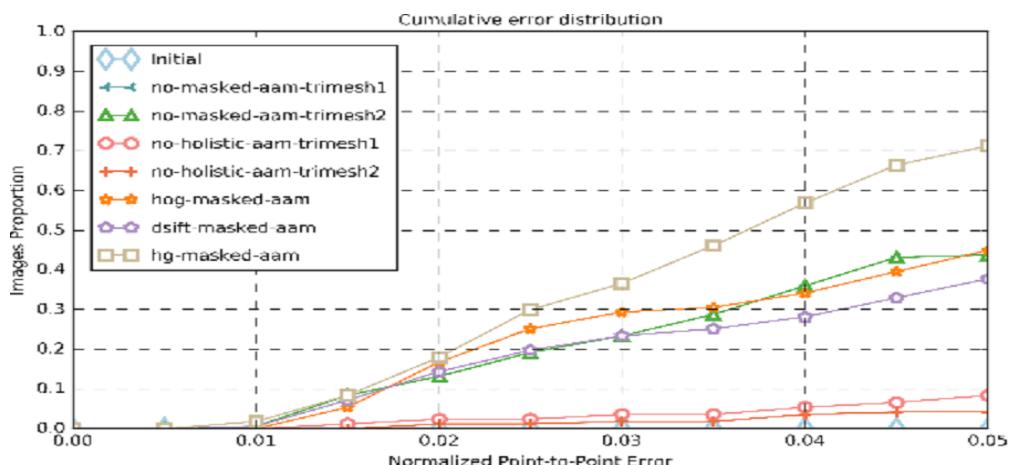


Figure 6.9 不同特征 AAM 的左心室内外膜分割性能比较

公式 1 中外观纹理的对齐较大程度上依赖三角网格的划分，与人脸对齐的差异是，心室分割中的三角网格并不总是都有一定实际意义，本文对比实验了两种的三角网格（图 1c,3b）。同时由于心外膜周围区域较难定义特征点及定位，实验发现基于块的全局 AAM（图 3c 和图 5 中 masked）普遍优于全局 AAM(图 5 中 holistic) 的方法。

分割性能见图 6.9，外观特征比较了原始像素（no）、dsift^[11]、HOG 和 HG 特

¹sst

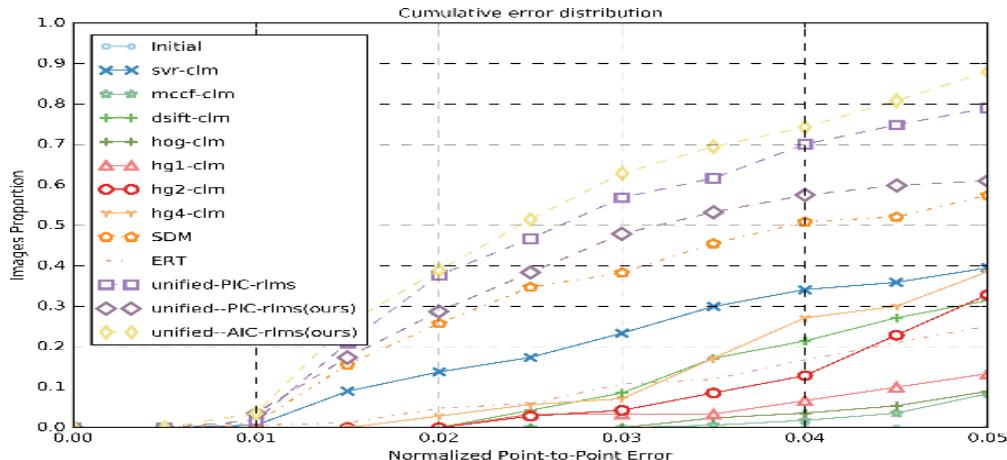


Figure 6.10 分割结果对比

征，结果表明采用 HG 网络自动特征的分割效果远优于手工设计的特征（图6.9中 hg-masked-aam），其中 dsift（8 个通道）和 hog（32 个通道）效果比只使用灰度的结果要好；实验结果表明只使用原始像素，即使用第 3.1 节提出的超声组织特征纹理特异性灰度归一化，得到的形状无关的外观纹理（图6.7b）与真实心动图差异仍较大，导致分割效果较差（图6.9中 no 曲线），主要原因是因为 AAM 方法对初始值敏感，之前文献^[162,163]中实验验证时仅是根据真实形状施加噪声扰动作为初始值^[173]，这不符合实际情况，本文提出心室检测作为初始轮廓的放置依据。

基于不同特征的 CLM 分割效果如图 6 所示，CLM 方法相比 AAM 方法的分割结果较差，主要是由于针对超声图像的分割极易陷入局部极值，无论是基于判别分类 SVR 还是基于概率生成 MCCF 的 CLM 模型分割结果都较差，即使结合 HG 特征改进效果也不明显，这主要是因为 HG 网络是基于特征点周围服从高斯分布的假设训练得到，这对超声心动图明显不十分合适，这也是下一步需要改进的方面。而随着层级的加大得到更多的全局信息，CLM 分割效果逐渐提升（图 6 中 hg1,2,4），但仍劣于基于判别分类回归的 SVR 方法。

6.2.4.3 结合最优的 AAM 和 CLM 分割结果

结合前文提出基于 4 级 HG 网络特征的 AAM 和 CLM 模型，克服两者相应缺点，能得到本文的最优结果（图6.10中 unified-PIC-rlms 表示采用文献^[176]提出的方法），其中 PIC 和 AIC 分别表示前文提到对 AAM 模型两种迭代算法，rlms 表示对 CLM 模型的优化方法。同时跟基于级联形状回归的 ert 算法^[172]和 sdm 算法^[180]进行实验比较，相应实验参数设置同原论文，结果表明提出方法的结果的有效性。

方法	A	B	C1	C2	C3	C4
均值	59.5	72.7	54.8	57.2	55.6	57.8
方差	21.4	25.2	21.8	20.7	21.5	20.3

Table 6.3 不同分割方法与专家标注的对比统计

计算预测形状与专家标注形状之间的距离，然后统计这些距离的均值和方差，得到的统计结果见表6.3。表中 A 代表结合 4 级 HG 特征的 AAM(错误率阈值为 0.03); B 代表结合 4 级 HG 特征的 CLM 方法；用统一 AAM 和 CLM 结合 4 级 HG 特征表示本方法，C1 代表本方法下错误率阈值为 0.05 的内膜分割结果；C2 代表本方法下错误率阈值为 0.05 的外膜分割结果；C3 代表本方法下错误率阈值为 0.02 的内膜分割结果；C4 代表本方法下错误率阈值为 0.02 的外膜分割结果。结果表明从总体形状间的平均距离上能看出提出内膜分割明显优于外膜分割结果，验证方法的有效性（表6.3）。由图6.11a 中可见，本文方法结果与专家标注比较接近。

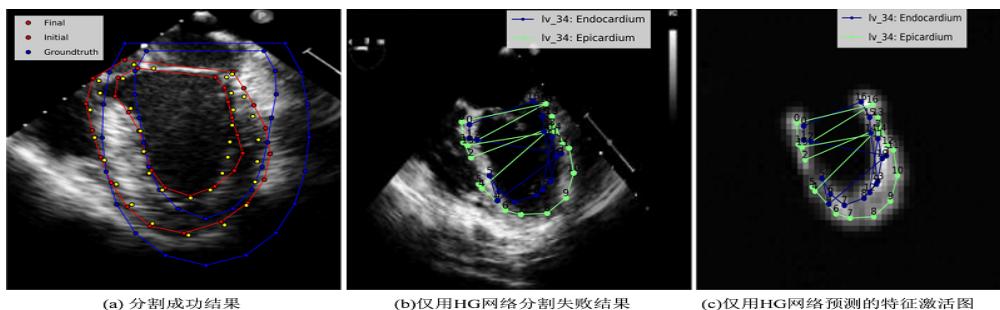


Figure 6.11 用 HG 网络预测心室内外膜成功和失败的案例

从分割失败的案例中能得知，尽管基于 HG 网络能综合建模心室外观的全局和局部特征，且该特征确实是对内外膜的响应（图6.11c），但由于特征点之间并没有形状和顺序信息，有可能导致分割失败。

6.2.5 小结与讨论

本文提出了一种基于沙漏卷积神经网络特征的统计形状模型分割方法，针对医学图像的组织分割任务，在自动检测左心室提供初始化轮廓的基础上，通过统一全局 AAM 和 CLM 模型的概率解释，综合两种方法的优点自动同时分割左心室内膜和外膜。在心室分割数据集上的实验结果表明，本文提出的自动分割方法在准确度和可解释性方面优于许多已有的分割方法。因此，本文的方法是可行的和有效的。本章去噪的研究内容和主要成果以题为《Perceptual Loss with Fully

Convolutional for Image Residual Denoising》的论文形式已发表在学术会议上^[139], 分割的研究内容和主要成果以题为《基于沙漏卷积网络的多尺度形状模型分割方法》的学术论文已投稿在审^[181]。

第7章 总结与展望

现有的医学图像分析方法和相关的研究内容具有内容博杂、应用对象复杂、专业性强的特点。由于医学图像的本身复杂性，不同模态图像成像原理不同，即使同一部位的不同模态的图像内容都不一致，目前没有一种对所有模态医学图像都适用的全自动分析方法以满足临床需求。而深度学习的出现打破了现状，深度 CNN 在图像处理、视频、语音识别和自然语言文本处理中均取得了突破。本文中我们主要从计算机视觉的角度对最近 CNN 在医学图像领域取得的进展进行了研究。首先概述了深度学习在医学图像领域的研究现状，其中分别介绍了不同应用领域的研究现状，还介绍了相关软硬件平台的发展情况，并给出了当前的趋势和面临的挑战；然后讨论了 CNN 在不同方面取得的进步的相关基础理论：比如层的设计结构、激活函数、损失函数、正则化、优化等方面的基础知识；除了从 CNN 的各个方面回顾其进展，我们结合 CNN 在计算机视觉任务上应用到医学图像分析领域的具体任务，其中包括医学超声图像分类、超声图像和磁共振图像中心室的检测以及超声心动图心室分割；并从分类识别引申出对深度模型的可解释性问题的研究，即深度可视化分析。将分别详细总结研究内容并展望相关研究方向。

7.1 研究总结

针对特征表示的高层语义识别问题，构建超声心动图的标准切面数据库，提出了一种基于深度卷积神经网络的自动识别方法，该算法针对网络全连接层占有模型大部分参数的缺点，引入空间金字塔均值池化替代全连接层，获得更多的空间结构信息，利用全局空间金字塔均值池化方法进行微调迁移学习，并大大减少模型参数、降低过拟合风险，同时通过类别显著性区域将类似注意力机制引入模型可视化过程，详尽分析了数据规模对模型分类精度的影响，并对模型的可解释性和有效性进行了分析。

针对基于深度卷积神经网络的图像分类模型的可解释性问题，通过评估模型特征空间的潜在可表示性，提出一种用于改善理解模型特征空间的可视化方法。给定任何已训练的深度卷积网络模型，引入了通过激活最大化获得的图像可解释性的正则化方法，结合现有正则化方法提出空间金字塔分解方法，利用构建多层拉普拉斯金字塔主动提升目标图像特征空间的低频分量，结合多层次高斯金

字塔调整其特征空间的高频分量得到较优可视化效果。并通过限制可视化区域，提出利用类别显著性激活图技术加以压制上下文无关信息，可进一步改善可视化效果。该模型有效克服了原有可视化方法中由于不能主动调整高低频分量等原因造成的可视化图像语义重复和低效率等问题。

针对自动检测医学图像中指定目标时存在的问题，提出了一种基于深度学习自动检测目标位置和估计对象姿态的算法。该算法基于区域深度卷积神经网络和目标结构的先验知识，采用区域生成候选框网络、感兴趣区域池化策略，引入包括分类损失、边框位置回归定位损失和像平面内朝向损失的多任务损失函数，近似优化一个端到端的有监督定位网络，能快速地对医学图像中目标自动定位，有效地为下一步的分割和参数自动提取提供定位结果。并在超声心动图左心室检测中提出利用检测额外标记点：二尖瓣环、心内膜垫和心尖，能高效地对左心室朝向姿态进行估计。

针对特征表示的底层视觉任务：图像去噪和分割中存在的问题，我们提出了一个有监督多层残差卷积网络框架，结合不同损失函数学习端到端映射变换；针对医学超声图像的对比度低、存在斑点噪声导致难以分割的问题，提出一种利用沙漏卷积神经网络特征的多尺度形状模型分割方法，自动定位经食道超声心动图中心室并全自动分割心室内外膜。首先，结合梯度方向直方图特征和支持向量机的心室自动检测方法，自动确定分割模型中的初始轮廓；其次，将心室分割任务纳入统计形变模型形状特征点对齐任务框架，通过比较不同外观纹理特征和激活图，包括传统手工设计的特征和利用深度学习自动学习的卷积特征，提出利用堆叠多级沙漏卷积网络建模心室外观的全局和局部信息，统一活动外观模型和局部受限模型的概率形式，采用反向组合梯度下降算法迭代优化分割结果，完成左心室轮廓的自动提取。然后，以医生手动勾勒的轮廓作为“金标准”，通过构造心室分割数据集以评价算法，且提出了扩充数据样本的方法来克服深度卷积网络过拟合问题，进行详尽实验讨论分析了基于不同层级的多级沙漏卷积网络对全局和局部纹理特征建模能力对分割效果的影响。实验结果表明，卷积模块允许网络提取专门用于指定任务的特征，并通过实验显示其优于手工设计的特征。该方法分割效果优于传统形状对齐方法，能够解决自动定位超声心动图中左心室的初始轮廓和弱边界自动分割的问题。

7.2 研究展望

虽然在实验的测量中，CNN 获得了巨大的成功，但是，仍然还有很多工作值得进一步研究。首先，鉴于最近的 CNN 变得越来越深，它们也需要大规模的数据库和巨大的计算能力，来展开训练。人为搜集标签数据库要求大量的人力劳动。所以，大家都渴望能开发出无监督式的 CNN 学习方式。

同时，为了加速训练进程，虽然已经有一些异步的 SGD 算法，证明了使用 CPU 和 GPU 集群可以在这方面获得成功，但是，开放高效可扩展的训练算法依然是有价值的。在训练的时间中，这些深度模型都是对内存有高的要求，并且消耗时间的，这使得它们无法在手机平台上部署。如何在不减少准确度的情况下，降低复杂性并获得快速执行的模型，这是重要的研究方向。

其次，我们发现，CNN 运用于新任务的一个主要障碍是：如何选择合适的超参数？比如学习率、卷积过滤的核大小、层数等等，这需要大量的技术和经验。这些超参数存在内部依赖，这会让调整变得很昂贵。最近的研究显示，在学习式深度 CNN 架构的选择技巧上，存在巨大的提升空间。

最后，关于 CNN，依然缺乏统一的理论。目前的 CNN 模型运作模式依然是黑箱。我们甚至都不知道它是如何工作的，工作原理是什么。当下，值得把更多的精力投入到研究 CNN 的基本规则上去。同时，正如早期的 CNN 发展是受到了生物视觉感知机制的启发，深度 CNN 和计算机神经科学二者需要进一步的深入研究。有一些开放的问题，比如，生物学上大脑中的学习方式如何帮助人们设计更加高效的深度模型？带权重分享的回归计算方式是否可以计算人类的视觉皮质等等。我们希望这篇文章不仅能让人们更好地理解 CNN，同时也能促进 CNN 领域中未来的研究活动和应用发展。

参考文献

- [1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[C]//Advances In Neural Information Processing Systems. 2012: 1–9.
- [2] HOIEM D, CHODPATHUMWAN Y, DAI Q. Diagnosing error in object detectors[J]. Lect. Notes Comput. Sci., 2012, 7574 LNCS(PART 3): 340–353.
- [3] LAMBIN P, RIOS-VELAZQUEZ E, LEIJENAAR R, et al. Radiomics: Extracting more information from medical images using advanced feature analysis[J]. Proceedings of SPIE—the International Society for Optical Engineering, 2015, 73(4): 389–400.
- [4] COOTES T F, C.J.TAYLOR. Statistical Models of Appearance for Computer Vision[R]// University of Manchester: M. 2004: 1–124.
- [5] MARR, DAVID. Vision: A computational investigation into the human representation and processing of visual information[J]. Quarterly Review of Biology, 1982, 8.
- [6] TREISMAN A, GELADE G. A feature-integration theory of attention[J]. Cognitive Psychology, 1980, 12(1): 97–136.
- [7] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254–1259.
- [8] KASS M, WITKIN A, TERZOPOULOS D. Snakes: Active contour models[J/OL]. International Journal of Computer Vision, 1988, 1(4): 321–331.
- [9] LAZEBNIK S, SCHMID C, PONCE J, et al. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories[J/OL]. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR’06), 2006, 2: 2169–2178.
- [10] YANG J, YU K, GONG Y, et al. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification[J]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2009: 1794–1801.
- [11] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91–110.
- [12] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features [J/OL]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2001, 1: 511–518.

- [13] GRAUMAN K, DARRELL T. The pyramid match kernel: Discriminative classification with sets of image features[C]//Proceedings of the IEEE International Conference on Computer Vision: II. 2005: 1458–1465.
- [14] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection[C/OL]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05): volume 1. IEEE, 2005: 886–893.
- [15] FUKUSHIMA K, MIYAKE S. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position[J]. Pattern Recognition, 1982, 15(6): 455–469.
- [16] LECUN Y. Handwritten digit recognition with a back-propagation network[J]. Advances in Neural Information Processing Systems, 1990, 2: 396–404.
- [17] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211–252.
- [18] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C/OL]//Int. Conf. Learn. Represent. San Diego, USA, 2015: 1–14.
- [19] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[J/OL]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015, 07-12-June-2015(2): 1–9.
- [20] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C/OL]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR): volume 7. 2016: 770–778.
- [21] HUBEL D H, WIESEL T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. Journal of Physiology, 1962, 160(1): 106–154.
- [22] RIESENHUBER M, POGGIO T. Hierarchical models of object recognition in cortex. [J/OL]. Nature neuroscience, 1999, 2(11): 1019–25.
- [23] HORN B K, SCHUNCK B G. Determining optical flow[J/OL]. Artificial Intelligence, 1981, 17(1-3): 185–203.
- [24] ROSENBLATT F. The perceptron: A probabilistic model for information storage and organization in the brain[J]. Psychological Review, 1958, 65(6): 386–408.
- [25] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[M]. [S.l.]: MIT Press, 1988: 533–536.
- [26] HINTON G E, SALAKHUTDINOV R R. Reducing the Dimensionality of Data with Neural Networks[J/OL]. Science, 2006, 313(July): 504–507.
- [27] SICKLES E A. Use of computer-aided diagnosis in mammographic interpretation: De-

- tection versus distraction[C]//Radiological Society of North America 2006 Scientific Assembly and Meeting. [S.l.: s.n.], 2006.
- [28] ABRÀMOFF M D, LOU Y, ERGINAY A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning.[J]. Investigative Ophthalmology Visual Science, 2016, 57(13): 5200.
- [29] CHARBONNIER J P, RIKXOORT E M, SETIO A A, et al. Improving airway segmentation in computed tomography using leak detection with convolutional networks[J]. Medical Image Analysis, 2016, 36: 52.
- [30] GRINSVEN M J J P V, GINNEKEN B V, HOYNG C B, et al. Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1273–1284.
- [31] 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.
- [32] YANG W, CHEN Y, LIU Y, et al. Cascade of multi-scale convolutional neural networks for bone suppression of chest radiographs in gradient domain.[J]. Medical Image Analysis, 2016, 35: 421.
- [33] LITJENS G, KOOI T, BEJNORDI B E, et al. A survey on deep learning in medical image analysis[J]. Medical Image Analysis, 2017, 42(1995): 60–88.
- [34] SHEN D, WU G, SUK H I, et al. Deep Learning in Medical Image Analysis[J]. Annual Review of Biomedical Engineering, 2017, 19(1): 221–248.
- [35] GREENSPAN H, VAN GINNEKEN B, SUMMERS R M. Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1153–1159.
- [36] CHEN H, NI D, QIN J, et al. Standard Plane Localization in Fetal Ultrasound via Domain Transferred Deep Neural Networks[J]. IEEE Journal of Biomedical and Health Informatics, 2015, 19(5): 1627–1636.
- [37] BAR Y, DIAMANT I, WOLF L, et al. Chest pathology detection using deep learning with non-medical training[C]//IEEE International Symposium on Biomedical Imaging. [S.l.: s.n.], 2015: 294–297.
- [38] OLIVA A, TORRALBA A. Modeling the shape of the scene: A holistic representation of the spatial envelope[J]. International Journal of Computer Vision, 2001, 42(3): 145–175.
- [39] MARGETA J, CRIMINISI A, Cabrera Lozoya R, et al. Fine-tuned convolutional neural nets for cardiac MRI acquisition plane recognition[J/OL]. Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization, 2017, 5(5): 339–349.

- [40] VAN G B, SCHAEFER-PROKOP C M, PROKOP M. Computer-aided diagnosis: how to move from the laboratory to the clinic[J]. Radiology, 2011, 261(3): 719–32.
- [41] AKRAM S U, KANNALA J, EKLUND L, et al. Cell Segmentation Proposal Network for Microscopy Image Analysis[M/OL]//Deep Learn. Data Labeling Med. Appl. 2016: 21–29.
- [42] EMAD O, YASSINE I A, FAHMY A S. Automatic localization of the left ventricle in cardiac MRI images using deep learning[C/OL]//2015 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE, 2015: 683–686.
- [43] PARK S R, LEE J. A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis MRI[J/OL]. arXiv Prepr., 2016: 1–21.
- [44] CHEN H, ZHENG Y, PARK J H, et al. Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images[J]. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2016, 9901 LNCS: 487–495.
- [45] ROTH H, LU L, LIU J, et al. Improving Computer-aided Detection using Convolutional Neural Networks and Random View Aggregation.[J/OL]. IEEE transactions on medical imaging, 2016, PP(99): 1.
- [46] SETIO A A A, CIOMPI F, LITJENS G, et al. Pulmonary nodule detection in ct images: False positive reduction using multi-view convolutional networks[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1160–1169.
- [47] SHIN H C, ROTH H R, GAO M, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1285–1298.
- [48] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation[M/OL]//Miccai. 2015: 234–241.
- [49] MILLETARI F, NAVAB N, AHMADI S A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation[C/OL]//2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016: 565–571.
- [50] XIE Y, ZHANG Z, SAPKOTA M, et al. Spatial Clockwork Recurrent Neural Network for Muscle Perimysium Segmentation[C/OL]//OURSELIN S, JOSKOWICZ L, SABUNCU M R, et al. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Cham: Springer International Publishing, 2016: 185–193.
- [51] STOLLENGA M F, BYEON W, LIWICKI M, et al. Parallel Multi-Dimensional LSTM, With Application to Fast Biomedical Volumetric Image Segmentation[M]//CORTES C,

- LAWRENCE N D, LEE D D, et al. Advances in Neural Information Processing Systems 28. [S.I.]: Curran Associates, Inc., 2015: 2998–3006.
- [52] TRIGEORGIS G, SNAPE P, NICOLAOU M A, et al. Mnemonic Descent Method: A Recurrent Process Applied for End-to-End Face Alignment[C/OL]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016: 4177–4187.
- [53] ZHANG Z, XING F, SU H, et al. Recent Advances in the Applications of Convolutional Neural Networks to Medical Image Contour Detection[J/OL]. arXiv preprint, 2017, 1(1): 12–24.
- [54] KALLENBERG M, PETERSEN K, NIELSEN M, et al. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring[J]. IEEE Trans Med Imaging, 2016, 35(5): 1322–1331.
- [55] YAN Z, ZHAN Y, PENG Z, et al. Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1332–1343.
- [56] MIAO S, WANG Z J, LIAO R. A cnn regression approach for real-time 2d/3d registration [J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1352–1363.
- [57] GOLKOV V, DOSOVITSKIY A, SPERL J I, et al. q-space deep learning: Twelve-fold shorter and model-free diffusion mri scans[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1344–1351.
- [58] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding[J/OL]. Proceedings of the ACM International Conference on Multimedia, 2014: 675–678.
- [59] ABADI M, BARHAM P, CHEN J, et al. Tensorflow: A system for large-scale machine learning[C]//OSDI. [S.I.: s.n.], 2016.
- [60] AL-RFOU' R, ALAIN G, ALMAHAIRI A, et al. Theano: A python framework for fast computation of mathematical expressions[J]. CoRR, 2016, abs/1605.02688.
- [61] CHEN T, LI M, LI Y, et al. Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems[J]. CoRR, 2015, abs/1512.01274.
- [62] HORNIK K, STINCHCOMBE M, WHITE H. Multilayer feedforward networks are universal approximators[J]. Neural Networks, 1989, 2(5): 359–366.
- [63] JARRETT K, KAVUKCUOGLU K, RANZATO M, et al. What is the best multi-stage architecture for object recognition[C/OL]//Proceedings of the IEEE International Conference on Computer Vision. 2009: 2146–2153.
- [64] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training

- by Reducing Internal Covariate Shift[C/OL]//Proc. 32nd Int. Conf. Mach. Learn.: volume 37. Lille, France, 2015: 448—456.
- [65] LIN M, CHEN Q, YAN S. Network In Network[J/OL]. arXiv preprint, 2013: 10.
 - [66] PRASOON A, PETERSEN K, IGEL C, et al. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network.[C]//Medical Image Computing & Computer-assisted Intervention: Miccai International Conference on Medical Image Computing & Computer-assisted Intervention. [S.l.: s.n.], 2013: 246–253.
 - [67] ROTH H R, LU L, SEFF A, et al. A new 2.5d representation for lymph node detection using random sets of deep convolutional neural network observations.[J]. Med Image Comput Assist Interv., 2014, 17(1): 520–527.
 - [68] KAMNITSAS K, LEDIG C, NEWCOMBE V F, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation[J]. Medical Image Analysis, 2017, 36: 61–78.
 - [69] FARABET C, COUPRIE C, NAJMAN L, et al. Learning hierarchical features for scene labeling[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(8): 1915–1929.
 - [70] MOESKOPS P, VIERGEVER M A, MENDRIK A M, et al. Automatic segmentation of mr brain images with a convolutional neural network.[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1252–1261.
 - [71] SONG Y, ZHANG L, CHEN S, et al. Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning.[J]. IEEE transactions on bio-medical engineering, 2015, 62(10): 2421.
 - [72] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(4): 640.
 - [73] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[M]. [S.l.]: Springer Berlin Heidelberg, 1997: 1735–1780.
 - [74] CHO K, MERRIENBOER B V, GULCEHRE C, et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation[J]. Computer Science, 2014.
 - [75] STOLLENGA M F, BYEON W, LIWICKI M, et al. Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation[J]. Computer Science, 2015.
 - [76] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion[J]. Journal of Machine Learning Research, 2010, 11(12): 3371–3408.

- [77] KINGMA D P, WELLING M. Auto-encoding variational bayes[J]. Computer Science, 2013.
- [78] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [J]. Advances in Neural Information Processing Systems, 2014: 2672–2680.
- [79] HINTON G E. Deep belief networks[J]. Scholarpedia, 2009, 4(6): 5947.
- [80] MAAS A L, HANNUN A Y, NG A Y. Rectifier Nonlinearities Improve Neural Network Acoustic Models[C/OL]//Proceedings of the 30 th International Conference on Machine Learning: volume 28. 2013: 6.
- [81] XU B, WANG N, CHEN T, et al. Empirical Evaluation of Rectified Activations in Convolutional Network[J/OL]. arXiv preprint, 2015.
- [82] CLEVERT D A, UNTERTHINER T, HOCHREITER S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)[J/OL]. International Conference on Learning Representations, 2015(1997): 1–14.
- [83] SHAHAM U, LEDERMAN R R. Learning by coincidence: Siamese networks and common variable learning[J]. Pattern Recognition, 2017.
- [84] ZHU H, LONG M, WANG J, et al. Deep Hashing Network for Efficient Similarity Retrieval[C]//Proceedings of the 30th Conference on Artificial Intelligence (AAAI 2016): number 1. [S.l.: s.n.], 2016: 2415–2421.
- [85] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[J/OL]. Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS), 2010, 9: 249–256.
- [86] KINGMA D, BA J. Adam: A method for stochastic optimization[J]. Eprint Arxiv, 2014.
- [87] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J/OL]. Journal of Machine Learning Research, 2014, 15(2): 1929–1958.
- [88] WAN L, ZEILER M, ZHANG S, et al. Regularization of neural networks using dropconnect[C]//International Conference on Machine Learning. [S.l.: s.n.], 2013: 1058–1066.
- [89] EBADOLLAHI S, CHANG S F C S F, WU H. Automatic view recognition in echocardiogram videos using parts-based representation[J]. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., 2004, 2.
- [90] Kevin Zhou S, PARK J H, GEORGESCU B, et al. Image-based multiclass boosting and echocardiographic view classification[C/OL]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition: volume 2. 2006: 1559–1565.

- [91] OTEY M, BI J, KRISHNA S, et al. Automatic view recognition for cardiac ultrasound images[C/OL]//MICCAI: International Workshop on Computer Vision for Intravascular and Intracardiac Imaging. 2006: 187–194.
- [92] ROY A, SURAL S, MUKHERJEE J, et al. Modeling of Echocardiogram Video Based on Views and States[M]//Computer Vision, Graphics and Image Processing. [S.l.]: Springer Berlin Heidelberg, 2006: 397–408.
- [93] PARK J H, ZHOU S K, SIMOPOULOS C, et al. Automatic Cardiac View Classification of Echocardiogram[C/OL]//2007 IEEE 11th International Conference on Computer Vision. IEEE, 2007: 1–8.
- [94] BEYMER D, SYEDA-MAHMOOD T, WANG F. Exploiting spatio-temporal information for view recognition in cardiac echo videos[C/OL]//2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2008: 1–8.
- [95] WU H, BOWERS D M, HUYNH T T, et al. Echocardiogram view classification using low-level features[C]//Proceedings - International Symposium on Biomedical Imaging: number 1156822. 2013: 752–755.
- [96] RAZAVIAN A S, AZIZPOUR H, SULLIVAN J, et al. CNN features off-the-shelf: An astounding baseline for recognition[C/OL]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2014: 512–519.
- [97] SIMONYAN K, VEDALDI A, ZISSERMAN A. Deep Inside Convolutional Networks Visualising Image Classification Models and Saliency Maps[C/OL]//Int. Conf. Learn. Represent. 2014: 1–8.
- [98] MAHENDRAN A, VEDALDI A. Understanding deep image representations by inverting them[C/OL]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015: 5188–5196.
- [99] ZEILER M D, FERGUS R. Visualizing and Understanding Convolutional Networks [M/OL]//FLEET D, PAJDLA T, SCHIELE B, et al. Computer Vision – ECCV 2014: 13th European Conference: 8689 LNCS. Zurich: Springer International Publishing, 2014: 818–833.
- [100] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Learning Deep Features for Discriminative Localization[C/OL]//2016 IEEE Conference on Computer Vision and Pattern Recognition Learning: volume 111. 2015: 2921–2929.
- [101] HE K, ZHANG X, REN S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916.
- [102] CHATFIELD K, SIMONYAN K, VEDALDI A, et al. Return of the Devil in the Details:

- Delving Deep into Convolutional Nets[C/OL]//Proceedings of the British Machine Vision Conference 2014. British Machine Vision Association, 2014: 6.1–6.12.
- [103] 陶攀, 付忠良, 朱锴王莉莉. 基于深度学习的超声心动图切面识别方法[J/OL]. 计算机应用, 2017, 37(5): 1434–1438.
- [104] ERHAN D, BENGIO Y, COURVILLE A, et al. Visualizing higher-layer features of a deep network[R/OL]//Univ. Montr.: number 1341. Montréal, Canada, 2009: 1–13.
- [105] LENCI K, VEDALDI A. Understanding image representations by measuring their equivariance and equivalence[C/OL]//2015 IEEE Conf. Comput. Vis. Pattern Recognit. IEEE, 2015: 991–999.
- [106] SZEGEDY C, ZAREMBA W, SUTSKEVER I. Intriguing properties of neural networks [C/OL]//Int. Conf. Learn. Represent. 2014: 1–10.
- [107] YOSINSKI J, CLUNE J, NGUYEN A, et al. Understanding Neural Networks Through Deep Visualization[C/OL]//Deep Learning Workshop, International Conference on Machine Learning (ICML). 2015.
- [108] NGUYEN A, DOSOVITSKIY A, YOSINSKI J, et al. Synthesizing the preferred inputs for neurons in neural networks via deep generator networks[J/OL]. arXiv, 2016: 1–29.
- [109] NGUYEN A, YOSINSKI J, CLUNE J. Multifaceted Feature Visualization: Uncovering the Different Types of Features Learned By Each Neuron in Deep Neural Networks [C/OL]//Proc. Work. Vis. Deep Learn. Int. Conf. Mach. Learn. 2016: 23.
- [110] DOSOVITSKIY A, BROX T. Inverting Visual Representations with Convolutional Networks[C/OL]//2015 IEEE Conf. Comput. Vis. Pattern Recognit. 2015: 184–199.
- [111] MAHENDRAN A, VEDALDI A. Visualizing Deep Convolutional Neural Networks Using Natural Pre-images[J/OL]. Int. J. Comput. Vis., 2016: 1–23.
- [112] CAO C, LIU X, YANG Y, et al. Look and Think Twice: Capturing Top-Down Visual Attention with Feedback Convolutional Neural Networks[C/OL]//2015 IEEE Int. Conf. Comput. Vis. IEEE, 2015: 2956–2964.
- [113] BACH S, BINDER A, MONTAVON G, et al. Analyzing Classifiers: Fisher Vectors and Deep Neural Networks[C/OL]//Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 2016: 17.
- [114] GOODFELLOW I J, SHLENS J, SZEGEDY C. Explaining and Harnessing Adversarial Examples[C/OL]//Int. Conf. Learn. Represent. San Diego, USA, 2014: 484–485.
- [115] HUANG B. FaceNet : A Unified Embedding for Face Recognition and Clustering[C]// 2015 IEEE Conf. Comput. Vis. Pattern Recognit. Boston, USA: [s.n.], 2015: 815–823.
- [116] DENTON E, CHINTALA S, SZLAM A, et al. Deep Generative Image Models using a

- Laplacian Pyramid of Adversarial Networks[C/OL]//Adv. Neural Inf. Process. Syst. 28. 2015: 1486—1494.
- [117] KRISHNAN D, FERGUS R. Fast image deconvolution using hyper-laplacian priors [C/OL]//Y. Bengio, SCHUURMANS D, LAFFERTY J D, et al. Adv. Neural Inf. Process. Syst.: volume 28. Curran Associates, Inc., 2009: 1—9.
- [118] BURT P, ADELSON E. The Laplacian Pyramid as a Compact Image Code[J]. IEEE Trans. Commun., 1983, 31(4): 532—540.
- [119] VAN DER SCHAAF A, VAN HATEREN J H. Modelling the power spectra of natural images: statistics and information.[J/OL]. Vision research, 1996, 36(17): 2759—70.
- [120] DUCHI J, HAZAN E, SINGER Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization[J/OL]. J. Mach. Learn. Res., 2011, 12: 2121—2159.
- [121] GATYS L A, ECKER A S, BETHGE M. Texture Synthesis Using Convolutional Neural Networks[J]. Adv. Neural Inf. Process. Syst., 2015: 262—270.
- [122] HE K, WANG Y, HOPCROFT J. A Powerful Generative Model Using Random Weights for the Deep Image Representation[C/OL]//Adv. Neural Inf. Process. Syst. 28. Barcelona, Spain, 2016: 1—8.
- [123] 陶攀, 付忠良朱锴. 空间金字塔分解的深度可视化方法[J]. 哈尔滨工业大学学报, 2017, 49(11): 68—73.
- [124] CHENG J Z, NI D, CHOU Y H, et al. Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans[J/OL]. Sci. Rep., 2016, 6(1): 24454.
- [125] SCHÖLLHUBER A. Fully Automatic Segmentation of the Myocardium in Cardiac Perfusion MRI[J/OL]. Engineering in Medicine, 2008, 3(22): 12—19.
- [126] LU Y, RADAU P, CONNELLY K, et al. Segmentation of Left Ventricle in Cardiac Cine MRI: An Automatic Image-Driven Method[M/OL]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009: 339—347.
- [127] PETITJEAN C, DACHER J N. A review of segmentation methods in short axis cardiac MR images[J/OL]. Med. Image Anal., 2011, 15(2): 169—184.
- [128] KELLMAN P, LU X, JOLLY M P, et al. Automatic LV localization and view planning for cardiac MRI acquisition[J/OL]. J. Cardiovasc. Magn. Reson., 2011, 13(Suppl 1): P39.
- [129] ZHOU S K, GEORGESCU B, ZHOU X S, et al. Image based regression using boosting method[C]//Proc. IEEE Int. Conf. Comput. Vis.: I. 2005: 541—548.
- [130] SHE Y, LIU D C. An Interactive Editing Tool from Arbitrary Slices in 3D Ultrasound Volume Data[M/OL]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007: 2447—2451.

- [131] ZHENG Y, COMANICIU D. Marginal Space Learning for Medical Image Analysis [M/OL]. New York, NY: Springer New York, 2014: 199–256.
- [132] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C/OL]//2014 IEEE Conf. Comput. Vis. Pattern Recognit. IEEE, 2014: 580–587.
- [133] GIRSHICK R. Fast R-CNN[C/OL]//2015 IEEE Int. Conf. Comput. Vis.: 2015 Inter. IEEE, 2015: 1440–1448.
- [134] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[C/OL]//Adv. Neural Inf. Process. Syst. 2015: 1–10.
- [135] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial Transformer Networks [C/OL]//Proc. 28th Int. Conf. Neural Inf. Process. Syst.: volume 25. Montreal, Canada, 2015: 2017–2025.
- [136] BEYER L, HERMANS A, LEIBE B. Biternion nets: Continuous head pose regression from discrete training labels[J]. Lect. Notes Comput. Sci., 2015, 9358: 157–168.
- [137] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training Region-based Object Detectors with Online Hard Example Mining[J/OL]. Comput. Vis. Pattern Recognit., 2016.
- [138] ANDREOPoulos A, TSOTSOS J K. Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI[J]. Medical Image Analysis, 2008, 12 (3): 335–357.
- [139] 陶攀, 付忠良, 朱锴, 等. 基于深度学习的医学计算机辅助检测方法研究[J]. 生物医学工程学杂志, 2018, 6(2): 12–19.
- [140] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli. Wavelets for Image Image quality assessment: From error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600–612.
- [141] ZHAO H, GALLO O, FROSIO I, et al. Is L2 a Good Loss Function for Neural Networks for Image Processing?[J]. arXiv preprint, 2015.
- [142] ZHANG L, ZHANG L, MOU X, et al. A comprehensive evaluation of full reference image quality assessment algorithms[C]//2012 19th IEEE International Conference on Image Processing. IEEE, 2012: 1477–1480.
- [143] BURGER H C, SCHULER C J, HARMELING S. Image denoising: Can plain neural networks compete with BM3D?[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 2392–2399.
- [144] DABOVK, FOIA, KATKOVNIKV, et al. Image denoising by sparse 3-d transform-domain collaborative filtering.[J]. *IEEE Trans. Image Processing*, 16(8):2080–2095, 2007.

- [145] VINCENTP, LAROCHELLEH, BENGIOY, et al. Extracting and composing robust features with denoising autoencoders.[J]. In *Proc. Int. Conf. Mach. Learn.*, pages 1096–1103, 2008.
- [146] XIEJ, XUL, AND E C. Image denoising and inpainting with deep neural networks.[J]. In *Proc. Advances in Neural Inf. Process. Syst.*, pages 350–358, 2012.
- [147] AGOSTINELLI F, ANDERSON M R, LEE H. Adaptive multi-column deep neural networks with application to robust image denoising[J]. *Advances in Neural Information Processing Systems*, 2013: 1493–1501.
- [148] LI H M. Deep Learning for Image Denoising[J]. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 2014, 7(3): 171–180.
- [149] SKRIBTSOV P V, SURIKOV S O. Regularization Method for Solving Denoising and Inpainting Task Using Stacked Sparse Denoising Autoencoders[J]. *American Journal of Applied Sciences*, 2016, 13(1): 64–72.
- [150] WANG Y Q, MOREL J M. Can a Single Image Denoising Neural Network Handle All Levels of Gaussian Noise?[J]. *IEEE Signal Processing Letters*, 2014, 21(9): 1150–1153.
- [151] JAIN V, SEUNG S. Natural Image Denoising with Convolutional Networks[C]// BOTTOU D K, SCHUURMANS D, BENGIO Y, et al. *Advances in Neural Information Processing Systems*: volume 21. [S.l.]: Curran Associates, Inc., 2009: 769–776.
- [152] WU Y, ZHAO H, ZHANG L. Image Denoising with Rectified Linear Units[M]//Chu Kiong Loo, Keem Siah Yap, Kok Wai Wong, Andrew Teoh Beng Jin K H. *Neural Information Processing*. Springer International Publishing, 2014: 142–149.
- [153] MAO X J, SHEN C, YANG Y B. Image Denoising Using Very Deep Fully Convolutional Encoder-Decoder Networks with Symmetric Skip Connections[J]. *arXiv preprint*, 2016.
- [154] EIGEN D, KRISHNAN D, FERGUS R. Restoring an Image Taken through a Window Covered with Dirt or Rain[C]//2013 IEEE International Conference on Computer Vision. IEEE, 2013: 633–640.
- [155] WU Y, ZHAO H, ZHANG L. Image Denoising with Rectified Linear Units[M]//Chu Kiong Loo, Keem Siah Yap, Kok Wai Wong, Andrew Teoh Beng Jin K H. *Neural Information Processing*. [S.l.]: Springer International Publishing, 2014: 142–149.
- [156] WANG X, TAO Q, WANG L, et al. Deep convolutional architecture for natural image denoising[C]//2015 International Conference on Wireless Communications & Signal Processing (WCSP). IEEE, 2015: 1–4.
- [157] DOSOVITSKIY A, BROX T. Generating Images with Perceptual Similarity Metrics based on Deep Networks[Z]. [S.l.: s.n.], 2016.

- [158] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution[J]. arXiv Preprint, 2016.
- [159] NOHH, HONGS, AND B H. Learning deconvolution network for semantic segmentation. [J]. In *Proc. IEEE Int. Conf. Comp. Vis.*, pages 1520–1528, 2015.
- [160] HONGS, NOHH, AND B H. Decoupled deep neural network for semi-supervised semantic segmentation.[J]. In *Proc. Advances in Neural Inf. Process. Syst.*, 2015.
- [161] DONGC, LOYC C, HEK, et al. Image super-resolution using deep convolutional networks.[J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(2):295–307, 2016.
- [162] BOSCH J G J, MITCHELL S C S, LELIEVELDT B B P, et al. Automatic segmentation of echocardiographic sequences by active appearance motion models[J/OL]. *IEEE Transactions on Medical Imaging*, 2002, 21(11): 1374–1383.
- [163] HANSSON M, BRANDT S S, LINDSTRÖM J, et al. Segmentation of B-mode cardiac ultrasound data by Bayesian Probability Maps[J]. *Medical Image Analysis*, 2014, 18(7): 1184–1199.
- [164] MARSOUSI M, EFTEKHARI A, KOCHARIAN A, et al. Endocardial boundary extraction in left ventricular echocardiographic images using fast and adaptive B-spline snake algorithm[J]. *International Journal of Computer Assisted Radiology and Surgery*, 2010, 5(5): 501–513.
- [165] SANTIAGO C, NASCIMENTO J C, MARQUES J S. A new ASM framework for left ventricle segmentation exploring slice variability in cardiac MRI volumes[J/OL]. *Neural Computing and Applications*, 2017, 28(9): 2489–2500.
- [166] COOTES T, TAYLOR C, COOPER D, et al. Active Shape Models-Their Training and Application[J/OL]. *Computer Vision and Image Understanding*, 1995, 61(1): 38–59.
- [167] COOTES T F T, EDWARDS G J, TAYLOR C J. Active appearance models[J/OL]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 24(6): 681–685.
- [168] MITCHELL S C, BOSCH J G, LELIEVELDT B P F, et al. 3-D active appearance models: Segmentation of cardiac MR and ultrasound images[J]. *IEEE Transactions on Medical Imaging*, 2002, 21(9): 1167–1178.
- [169] VARGAS-QUINTERO L, ESCALANTE-RAMÍREZ B, Camargo Marín L, et al. Left ventricle segmentation in fetal echocardiography using a multi-texture active appearance model based on the steered Hermite transform[J/OL]. *Computer Methods and Programs in Biomedicine*, 2016, 137: 231–245.
- [170] CRISTINACCE D, COOTES T. Automatic feature localisation with constrained local models[J]. *Pattern Recognition*, 2008, 41(10): 3054–3067.

- [171] VAN STRALEN M, HAAK A, LEUNG K E Y E, et al. Full-cycle left ventricular segmentation and tracking in 3D echocardiography using active appearance models[C/OL]// 2015 IEEE International Ultrasonics Symposium (IUS). IEEE, 2015: 1–4.
- [172] KAZEMI V, SULLIVAN J. One Millisecond Face Alignment with an Ensemble of Regression Trees[J/OL]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014(August): 1867–1874.
- [173] 纪祥虎, 高思聪, 黄志标, 等. 基于 Centripetal Catmull-Rom 曲线的经食道超声心动图左心室分割方法[J]. 四川大学学报(工程科学版), 2016, 48(5): 4–10.
- [174] ZHOU S K, ZHOU J, COMANICIU D. A boosting regression approach to medical anatomy detection[C/OL]//2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007: 1–8.
- [175] GALOOGAHI H K, SIM T. Correlation filter cascade for facial landmark localization [C/OL]//2016 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2016: 1–8.
- [176] ALABORT-I MEDINA J, ZAFEIRIOU S. A Unified Framework for Compositional Fitting of Active Appearance Models[J/OL]. International Journal of Computer Vision, 2017, 121(1): 26–64.
- [177] NEWELL A, YANG K, DENG J. Stacked Hourglass Networks for Human Pose Estimation[J/OL]. European Conference on Computer Vision, 2016, 9912(4): 483–499.
- [178] CHU X, YANG W, OUYANG W, et al. Multi-Context Attention for Human Pose Estimation[J/OL]. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- [179] KINGMA D P, BA J. Adam: A Method for Stochastic Optimization[J/OL]. CoRR abs/1412.6980, 2014: 1–15.
- [180] XIONG X, De la Torre F. Supervised Descent Method and Its Applications to Face Alignment[C/OL]//2013 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2013: 532–539.
- [181] 陶攀, 付忠良, 朱锴. 基于沙漏卷积网络的多尺度形状模型分割方法[J]. 工程科学技
术学报, 2018: 1–9.

攻读学位期间发表的学术论文与科研成果

已发表论文

1. Pan Tao,Zhongliang Fu,Lili Wang,Kai Zhu. **Perceptual Loss with Fully Convolutional for Image Residual Denoising.** *Pattern Recognition.CCPR(EI)*. 2016. 122–132.DOI:10.1007/978-981-10-3005-5-11
2. 陶攀, 付忠良, 朱锴, 王莉莉. 金字塔分解的深度可视化方法, 哈尔滨工业大学学报 (EI), 2017,49(11):60-65,DOI:10.11918/j.issn.0367-6234.201612087
3. 陶攀, 付忠良, 朱锴. 基于深度学习的医学计算机辅助检测算法, 生物医学工程学杂志 (EI), 已录用, 2017
4. 陶攀, 付忠良, 朱锴, 王莉莉. 基于深度学习的超声心动图切面识别方法研究, 计算机应用 (中文核心), 2017.DOI:10.11772/j.issn.1001-9081.2017.05.1434
5. Xianghu Ji,Lili Wang,Pan Tao,Zhongliang Fu. Landmark Selecting on 2D Shapes for Constructing Point Distribution Model. *Pattern Recognition.CCPR(EI)*. 2016. 318–331.DOI:10.1007 /978-981-10-3002-4-27
6. Lili Wang, Zhongliang Fu,Pan Tao. **Four-chamber plane detection in cardiac ultrasound images based on improved imbalanced AdaBoost algorithm , IEEE, IICCBDA(EI)** 2016,299-303. DOI:10.1109/IICCBDA.2016.7529574

国家发明专利

1. 纪祥虎, 高思聪, 陶攀, 王莉莉. 用于统计形状模型的特征点辅助标注方法. (申请号:201510672503.8) 专利公开号:CN105205827 A.2015

项目经历

1. 2015–2016 四川省科技创新苗子工程——基于自动分割技术的左心室可视化及功能评价临床教学平台 (编号: 2015060)

项目描述: 本项目主要目标在于使用机器学习方法对左心室进行分割, 得到左心室轮廓及结构和心功能参数; 使用可视化技术对心脏左室进行三维立体结构教学。帮助麻醉医生学员快速学习掌握超声心动图中左心室结构

项目职责：在项目中主要负责超声图像中心脏器官的自动定位和分割，分别利用机器学习的方法对超声图像中的左心室定位，和 AAM 方法对肾脏进行分割。

项目成果：形成论文两篇，专利一项，期间主要研究了基于深度学习的图像预处理方法，基于形状对齐模型进行心室分割，及基于深度级联回归模型进行心室边界分割算法等

2. 2015-2015 阿里巴巴大规模图像搜索赛（38 名共 843 支参赛队伍）

本项目目标是从海量图像中检索最相同或似的 Top20 图像

主要负责使用深度学习模型对图像进行特征抽取，同时配合队友进行图像检索等其他工作，其中用时一个月根据 matconvnet 写了一个 C++ 版本的 CNN 框架的 API，从中获得了处理百万级数据的经验，获得了使用 Open-BLAS 处理大型矩阵运算的经验

项目收获：形成论文一篇，熟悉了深度学习提取语义特征进行实例检索的各项关键技术

3. 2015-2017 四川科技支撑计划-医学图像挖掘与心脏智能诊疗系统关键技术研究

项目描述：本项目主要目标在于使用机器学习方法对超声心动图标准切面进行自动识别。超声图像标准切面分类模块，包括图像预处理、特征提取和分类器模型构建实现标准切面自动识别分类；基于云端的海量切面视频的语义检索模块等

项目职责：项目参与人

任务分工：图像预处理、特征提取、分类器建模、视频语义检索

项目成果：发表论文三篇，分别研究了基于深度学习理论可视化分析其有效性，基于深度特征的超声图像标准切面自动识别算法等

4. 2013-2014 四川省科技支撑项目，华西医院合作项目-医学可视化模拟教学和诊断系统

项目描述：项目旨在为无经验的心脏外科医生和学员提供可视化的教学方案，同时通过机器学习和图形图像处理方法对三维心脏进行开放式建模，以提出一种基于心脏开放模型的智能诊疗综合系统

项目职责：在项目中负责超声图像处理和基于机器学习的病理挖掘工作。

任务分工：图像预处理

项目成果：参与撰写专利两项，对超声仪器，心脏疾病临床基本知识有较全面的了解；设计了针对心脏超声图像的分割识别方法，以及病理挖掘方法；学习了基于偏微分方程的图像去噪和基于水平集的分割方法

在审和 Working 论文

1. 陶攀, 付忠良. 基于 Fast-rcnn 的医学实例检索方法研究, Working, 2015
2. 陶攀, 付忠良. 基于超声心动图的左心室分割综述, Working, 2015
3. 陶攀, 付忠良. 基于形状对齐的超声心动图左心室分割方法, 工程科学学报 (在审), 2017
4. 陶攀, 付忠良. 基于形状对齐的超声心动图左心室分割方法, 工程科学学报 (在审), 2017
5. 陶攀, 付忠良. 基于 CNN-LSTM 的超声心动图左心室分割方法, Working, 2017

参与项目编写和申请

1. 2016 四川科技支撑计划-医学图像挖掘与心脏智能诊疗系统关键技术研究
2. 2016 基于医学图像建模的心功能评价系统研发与应用
3. 2015 国科控股技术创新项目-交互式视觉仿真关键技术研究与产品应用示范
4. 2014 西部之光项目-基于医学图像建模的评价系统
5. 2014 数字化医疗辅助设备关键技术研发—基于机器智能的三维可视化手术诊疗仿真平台

获奖及荣誉

1. 2015 中国科学院研究生院“三好学生”荣誉称号
2. 2016 中国科学院大学优秀学生干部
3. 2017 中国科学院博士国家奖学金

致 谢

转眼博士求学生涯即将结束，我要衷心感谢所有关心爱护我、帮助支持我的老师同学、好友以及家人。

首先，我要感谢我的导师付忠良研究员，在博士四年及硕士两年期间，付老师以其广博的知识、耐心指导学生的人格魅力、严谨的治学态度以及创新的科学精神深深地影响了我，在科学研究以及日常生活各方面都给予我最大的支持，不仅悉心传授我专业知识，更重要的是注重培养我做科研以及创新的能力，在教授理论知识的同时，注重理论联系实际。同时，他以身作则的态度给我树立了良好的榜样，在培养我专业技能的同时注重人格的培养，使我真正成为一个对社会有用的人。这些将使我终生受益。

感谢四川大学华西医院的宋海波医生以及其助手，使我对医学图像处理产生浓厚的兴趣，非常感谢姚宇老师及成都计算机所各位老师给我提供良好的学习环境，感谢他们对我学术研究的帮助和指导。

感谢我父亲陶朝重和母亲陶爱湘的养育之恩，父母一辈子务农，辛苦抚养我们兄妹两个长大，谨以本文给我最敬爱的父亲！

最后向参加论文评审和答辩的专家老师们表示感谢！