



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  
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：

日期：

导师签名：

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，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。性能，但仍有一系列问题难以克服：(1) 不同于传统的基于局部视觉特征的表征方法，深度表征在语义层面对图像进行整体的刻画，因而呈现出对局部细节表征不够突出，且对图像空间位置、几何形变比较敏感的特点；(2) 基于局部表征的方法可以利用局部特征之间的空间关系对图像匹配进行几何校验，以实现更加精确的匹配，而深度表征则难以利用这一性质对检索性能进行增强；(3) 现有的方法多使用具有人工标注的公共基准数据集对检索算法的性能进行验证，无法实现对任意查询实时响应的检索质量评估，不便于搜索引擎根据需要对检索结果进行修正。

针对以上问题，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、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：

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，利用构建多层次拉普

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。

针对图像去噪中存在的问题，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。输入是带噪声的图像和原图像，输出的是去噪后的图像。

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，图像内容理解

## Abstract

Machine learning is used in the medical imaging field, including computer-aided diagnosis, image segmentation, image registration, image fusion, image-guided therapy, image annotation, and image database retrieval. Deep learning methods are a set of algorithms in machine learning, which try to automatically learn multiple levels of representation and abstraction that help make sense of data. This in turn leads to the necessity of understanding and examining the characteristics of deep learning approaches, in order to be able to apply and refine the methods in a proper way.

The aim of this work is to evaluate deep learning methods in the medical domain and to understand if deep learning methods (random recursive support vector machines, stacked sparse auto-encoders, stacked denoising auto-encoders, K-means deep learning algorithm) outperform other state of the art approaches (K-nearest neighbor, support vector machines, extremely randomized trees) on two classification tasks, where the methods are evaluated on a handwritten digit (MNIST) and on a medical (PULMO) dataset. Beside an evaluation in terms of accuracy and runtime, a qualitative analysis of the learned features and practical recommendations for the evaluated methods are provided within this work. This should help improve the application and refinement of the evaluated methods in future.

Results indicate that the stacked sparse auto-encoder, the stacked denoising auto-encoder and the support vector machine achieve the highest accuracy among all evaluated approaches on both datasets. These methods are preferable, if the available computational resources allow to use them. In contrast, the random recursive support vector machines exhibit the shortest training time on both datasets, but achieve a poorer accuracy than the afore mentioned approaches. This implies that if the computational resources are limited and the runtime is an important issue, the random recursive support vector machines should be used.

**Keywords:** University of Chinese Academy of Sciences (UCAS)



## 目 录

摘要 .....	I
Abstract .....	III
符号列表 .....	VII
第 1 章 绪论 .....	1
1.1 研究背景及现实意义 .....	1
1.2 国内外研究现状 .....	3
1.3 创新点及全文结构 .....	5
1.4 论文的章节安排 .....	6
第 2 章 相关知识 .....	9
2.1 深度卷积神经网络的组件 .....	9
2.2 AAM 模型和 CLM 模型 .....	10
2.3 小结与讨论 .....	11
第 3 章 超声心动图切面的自动识别方法 .....	13
3.1 Deep-Echo 模型 .....	14
3.2 实验结果和分析 .....	16
3.3 小结与讨论 .....	20
第 4 章 空间金字塔分解的深度可视化方法 .....	21
4.1 可视化方法的数学模型 .....	22
4.2 梯度更新的可视化方法 .....	22
4.3 空间金字塔分解 .....	24
4.4 实验结果分析和讨论 .....	26
4.5 小结与讨论 .....	29
第 5 章 心室的计算机辅助检测方法 .....	31
5.1 区域卷积神经网络概览 .....	32
5.2 候选区域生成网络及其改进 .....	33
5.3 实验结果分析和讨论 .....	37
5.4 小结与讨论 .....	41

---

<b>第 6 章 图像的去噪方法</b>	43
6.1 相关工作	44
6.2 提出方法	45
6.3 实验结果分析和讨论	48
6.4 小结与讨论	51
<b>第 7 章 形状对齐的心室的分割方法</b>	53
7.1 初始位置定位和特征点标注	54
7.2 结合卷积网络特征的形状对齐模型	56
7.3 实验结果分析和讨论	58
7.4 小结与讨论	62
<b>第 8 章 总结与展望</b>	63
<b>参考文献</b>	65
<b>攻读学位期间发表的学术论文与科研成果</b>	75
<b>致 谢</b>	79

## 符号列表

### Characters

Symbol	Description	Unit
$R$	the gas constant	$\text{m}^2 \cdot \text{s}^{-2} \cdot \text{K}^{-1}$
$C_v$	specific heat capacity at constant volume	$\text{m}^2 \cdot \text{s}^{-2} \cdot \text{K}^{-1}$
$C_p$	specific heat capacity at constant pressure	$\text{m}^2 \cdot \text{s}^{-2} \cdot \text{K}^{-1}$
$E$	specific total energy	$\text{m}^2 \cdot \text{s}^{-2}$
$e$	specific internal energy	$\text{m}^2 \cdot \text{s}^{-2}$
$h_T$	specific total enthalpy	$\text{m}^2 \cdot \text{s}^{-2}$
$h$	specific enthalpy	$\text{m}^2 \cdot \text{s}^{-2}$
$k$	thermal conductivity	$\text{kg} \cdot \text{m} \cdot \text{s}^{-3} \cdot \text{K}^{-1}$
$T$	temperature	K
$t$	time	s
$p$	thermodynamic pressure	$\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-2}$
$\hat{p}$	hydrostatic pressure	$\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-2}$
$\mathbf{f}_b$	body force	$\text{kg} \cdot \text{m}^{-2} \cdot \text{s}^{-2}$
S	boundary surface	$\text{m}^2$
V	volume	$\text{m}^3$
$\mathbf{V}$	velocity vector	$\text{m} \cdot \text{s}^{-1}$
$u$	x component of velocity	$\text{m} \cdot \text{s}^{-1}$
$v$	y component of velocity	$\text{m} \cdot \text{s}^{-1}$
$w$	z component of velocity	$\text{m} \cdot \text{s}^{-1}$
$c$	speed of sound	$\text{m} \cdot \text{s}^{-1}$
$\mathbf{r}$	position vector	m
$\mathbf{n}$	unit normal vector	1
$\hat{\mathbf{t}}$	unit tangent vector	1
$\tilde{\mathbf{t}}$	unit bitangent vector	1
$C_R$	coefficient of restitution	1
$Re$	Reynolds number	1
$Pr$	Prandtl number	1

---

$Ma$	Mach number	1
$\alpha$	thermal diffusivity	$\text{m}^2 \cdot \text{s}^{-1}$
$\mu$	dynamic viscosity	$\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-1}$
$\nu$	kinematic viscosity	$\text{m}^2 \cdot \text{s}^{-1}$
$\gamma$	heat capacity ratio	1
$\rho$	density	$\text{kg} \cdot \text{m}^{-3}$
$\sigma_{ij}$	stress tensor	$\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-2}$
$S_{ij}$	deviatoric stress tensor	$\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-2}$
$\tau_{ij}$	viscous stress tensor	$\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-2}$
$\delta_{ij}$	Kronecker tensor	1
$I_{ij}$	identity tensor	1

### Operators

Symbol	Description
$\Delta$	difference
$\nabla$	gradient operator
$\delta^\pm$	upwind-biased interpolation scheme

### Abbreviations

Acronym	Description
ANFO	Ammonium Nitrate Fuel Oil
CFD	Computational Fluid Dynamics
CFL	Courant-Friedrichs-Lowy
CJ	Chapman-Jouguet
EOS	Equation of State
JWL	Jones-Wilkins-Lee
TVD	Total Variation Diminishing
WENO	Weighted Essentially Non-oscillatory
ZND	Zel'dovich-von Neumann-Doering

## 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。

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 (X-ray)，超声 (Ultrasound, US)，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 (Positron Emission Computed Tomography , PET )，单光子断层扫描 (Single-Photon Emission Computed Tomography, SPECT)，以及内窥镜，病理切片等。

医学影像分析也从传统的计算机辅助检测发展到火热的影像组学，但世界医疗卫生系统每天都会浪费大量的资源和时间，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。同时不同的影像质量和不同的工作流程会导致临幊上医生对影像内容的理解具有很大的主观性，对于通常在临床筛查和风险评估时所获得的医学图像，这些结构可能具有相当模糊的边界和低对比度，即使对于有经验的临幊医生来说，图像级解释也是具有挑战性和耗时的任务。

“当前影像诊断主要依赖人工阅片完成，然而，日益增加的图像数据也为人工阅片带来极大挑战。为了给医生提供有效的辅助诊断信息，智能图像处理技术正变得越来越重要。以机器学习和图像处理技术为基础的计算机辅助诊断 (computer aided diagnosis, CAD) 逐渐成为医学领域的研究热点 [1]。基于机器学习的 CAD 主要包括四方面的内容：图像预处理；感兴趣区 (region of interest, ROI) 的分割；特征提取、选择与分类；肿瘤区域的识别 (分类或者分割) [5]。其中，高效特征的提取尤为关键 [6]。目前，基于传统的浅层机器学习结构的 CAD 系统，高度依赖人工选择的特征，以及分类器对特征的整合。而且，由于传统的浅层学习结

构无法满足实际应用中对复杂函数建模的要求 [7]，所以难以区分高维特征之间的关系，通常需要降维处理。因此，我们需要简化及优化 CAD 技术中的特征选择的过程，以提高 CAD 系统进行辅助诊断的准确度。

“近年来方兴未艾的深度学习技术 [8] 作为一类多层神经网络学习算法，可通过深层非线性网络结构学习特征，并且通过组合低层特征形成更加抽象的深层表示（属性类别或特征），实现复杂函数逼近，表征输入数据分布式表示，从而可以学习到数据集的本质特征 [7]。因此，深度学习算法应用于 CAD 系统具有以下优势：第一，作为一种数据驱动的自动特征学习算法，可以直接从训练数据提取特征，从而大大减少特征提取的工作量以及人工干预的影响；第二，通过神经网络内在的深层结构可以表征特征之间的交互及层次结构，从而揭示高维特征之间的联系；第三，特征提取、特征选择及特征分类。三个核心步骤可以在同一个深层结构的最优化中实现 [6]。由此可见，深度学习有望解决基于传统浅层机器学习的 CAD 问题，从而大大提高辅助诊断能力。

近来，结合机器学习和计算机视觉的人工智能算法被用来帮助临床医生，提高医师对患者影像的理解，从而改善诊断，治疗以及由此产生的预后效果，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。机器学习已被用于医学图像分析，计算机及其运行的算法可以比人类科学家或医学专业人员更快，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。机器学习通常始于机器学习算法系统，该系统计算被认为在进行感兴趣的预测或诊断中是重要的图像特征。然后，机器学习算法系统识别这些图像特征的最佳组合，以对图像进行分类或计算给定图像区域的一些度量。图像中解剖结构的准确分类和定位是的基于图像全自动诊断的基础。

### 1.1.2 课题研究意义

人工智能进入医学成像领域，数据科学革命大约在五年前随着 IBM Watson 和 Google Brain 的出现而开始。他说，2012 年推出的深度学习算法确实推动了人工智能的发展，到 2014 年，机器正确读取放射学研究的比例开始下降，准确度达到了 95% 左右。回顾旧的检查，以帮助医院找到病人可能没有意识到病情的新病人。通过在卫生系统中记录的所有先前的胸部 CT 检查来帮助识别可能患有肺癌的患者。总的来说，人工智能提供了一个重要的机会来增强和增强放射科的阅读，而不是取代放射科医生。医疗领域的人工智能和机器学习将继续得到改善，影响疾病预防和诊断，

西门子医疗集团率先将人工智能（AI）算法引入心脏回波系统，以加速自动化。几年前，飞利浦医疗保健公司也在其 Epiq 超声系统中引入了 AI 的元素。它需要一个三维回波数据集采集和自动分析图像，以确定心脏的解剖，标签，然后切

片的最佳标准视图呈现。这消除了互操作性差异的问题，因为软件将总是选择基于机器学习的最佳视图，该机器学习使用数千个代表患者解剖变异谱的先前检查。这对于操作人员来说要积累相同的知识需要花费数年的时间。其他供应商也引入了深度学习算法的元素来帮助分析超声心动图或执行自动量化。下一代回声系统将结合更多的人工智能功能，通过自动完成耗时的任务和扩大超声检查员的工作量，从而进一步改善工作流程，从而提高工作效率，始终保持准确。

人工智能算法通过识别模式来读取医学图像。人工智能系统使用大量检查进行训练，以确定来自 CT，磁共振成像（MRI），超声或核成像扫描的正常解剖结构。然后使用异常情况训练 AI 系统的眼睛以识别异常，类似于计算机辅助检测软件（CAD）。然而，与 CAD 只是放射科医生可能想要仔细研究的区域不同，AI 软件具有更多的分析认知能力，基于更多的前几代 CAD 软件的临床数据和阅读体验。出于这个原因，正在帮助开发医学人工智能的专家经常将认知能力称为“有效的 CAD”。

通过学术界及企业界的共同努力，我们获得了大量的各种各样的医学影像数据，并形成了多个面向全球公开的医学影像数据库。大量医学影像数据的不断采集与积累，为手术导航精确制导提供了契机，但如何充分有效地利用这些影像，也给临床医生及医学影像信息工作者的研究提出了巨大的挑战。处理此如巨大数量的医学影像数据，亟待解决的问题便是如何综合利用不同模式间影像的互补信息，以及消除不同影像维度及分辨率带来的影响。分析和处理这些大数据，从中准确挖掘有用信息，需要我们提出更快速、鲁棒的计算方法。越来越进步的计算机辅助技术，如模式识别技术、数据挖掘技术等与大规模图像识别和机器学习技术结合，为大规模的数据分析和处理提供了可能。传统手工勾勒分割及配准方法工作量大，耗时多，且受医生自身经验的影响。而传统分割与配准算法缺乏自主学习性，需要算法设计者手动设计特征，且适用于单一种类图像，当学习对象改变时，需要重新进行训练。最近计算机视觉领域流行的深度学习类算法，对不同模式及维度的图像进行训练，使得算法具备自主学习性及普适性，便于更加精确地辅助肿瘤手术，因此为临床提供了新的诊疗契机。

## 1.2 国内外研究现状

### 1.2.1 医学图像分析应用计算机视觉的研究现状

计算机视觉在超声心动图中的应用。心脏回声有一些挑战，医学图像分析可以解决。例如，研究人员建议使用计算机视觉自动分割解剖结构，检测和分类先天性心脏缺陷，实时导管定位等。标准视图采集是心脏超声最基本的任务，也可以通过医学图像分析。标准视图获取为了找到标准的心脏视图，软件应该从超声波扫描期间的多个帧中选择合适的二维平面。在这里，出现了不同的挑战，如分析二维帧，

三维体积，二维时间序列或四维时空图像相关（STIC）体积。因此，随着医学图像数据量的不断增长，我们可以期待医学图像分析软件很快成为超声系统的重要组成部分。

医学图像分析是计算机视觉的实际应用-计算机科学的一个分支，涉及数字图像（包括数字视频帧）中的对象和特征识别。计算机视觉算法通过一系列过程来分析图像，类似于人类视觉系统所执行的过程。在经过初步预处理（包括去噪，滤波和特征增强）之后，软件在图像分割的过程中将图像分解成有意义的区域。然后，算法提取重要的特征，并基于这些特征对图像中的对象进行分类。此外，医学图像分析算法通常执行图像配准 - 映射两个以上相同解剖结构的图像以检测任何差异或变化。基于机器学习，分类是医学图像分析软件最复杂的功能。每个 AI 系统都使用机器学习方法作为其“大脑”。这些算法允许计算机记住大量信息，并在学习完成后使用它来分析类似的信息。这就是为什么这种方法在计算机视觉中得到如此广泛的应用在图像数据集（例如超声图像数据集）上进行训练，然后软件识别真实世界图像中的熟悉特征（例如，在实时超声扫描中）和在此基础上作出相关的结论。这些系统的准确性随着输入数据的数量而增加。从数百个图像开始，它们显示出不错的结果，并且在处理了数以千计的图像和更多图像之后，它们的准确度接近 100%。当然，这也取决于所使用的架构，随着机器学习方法的发展，用于医学图像分析的算法显示出更好的结果。

### 1.2.2 医学图像分析应用人工智能的研究现状

医学影像技术在我国医疗系统中的发展时间比较短，所以在技术方面还不够成熟，但是随着医疗技术以及影像技术的不断发展，首先，医学影像技术呈现出来的信息必然会更加具有敏感性、直观性以及特异性；其次，现在对影像的分析都是定性分析，在未来必然会向着定量的方向发展，不再仅仅给出疾病的诊断结果，而是向着提供手术路径的方向发展；再次，影像信息的采集与显示都还是二维图像，随着数字成像技术的不断发展必然会向着三维全数字化发展；最后，目前，放射科在使用影像技术进行疾病诊断的过程中使用的都还是单一技术，随着影像技术的不断进步，未来会逐渐引进新的影像技术，向着综合方向发展。通用电气，西门子和飞利浦是超声心动图供应商之一，将深度学习算法整合到回声软件中，帮助自动从三维超声数据集提取标准成像视图。这是飞利浦 Epiq 系统的一个例子，该系统使用供应商的解剖智能软件来定义解剖结构，并自动显示解剖标准诊断视图，无需人工干预。这可以大大加快工作流程并减少操作员之间的差异。

包括几家分析公司和创业公司在内的其他公司则展示了使用 AI 快速筛选大量大数据的软件，或者为适当的使用标准提供即时的临床决策支持，最好的测试或成像来进行诊断甚至提供差异诊断。飞利浦将 AI 作为其具有自适应智能的新型 Illumeo 软件的一个组件，该软件可自动获取相关的放射科先前的检查结果。用户

可以在特定的 MPI 视图中点击解剖结构的区域，AI 将查找并打开先前的成像研究以显示相同的解剖结构，切片和方向。对于肿瘤学成像，在图像中点击几次肿瘤，AI 将执行自动量化，然后对先验进行相同的测量，呈现肿瘤评估的并排比较。这可以显着减少与肿瘤跟踪评估和加速工作流程相关的时间。

基于人工智能（AI）的医学图像分析采用卷积神经网络，支持向量机，模糊逻辑系统等机器学习方法从医学图像中提取意义。最先进的计算机视觉软件为诊断人员提供了基于证据的技巧，消除了可能的疑惑并确保了诊断的一致性。标准视图位置是超声心动图中的关键步骤，因为这些帧包含基本的诊断数据。从超声波检查自动捕捉标准飞机可以加快扫描，并使其更加准确。仔细研究这方面的研究将证明这不是一个猜测。标准视图的计算机辅助检测不断支持临床医生。

### 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，通过分析，抓住其中的关键问题，即影响手术精确度的术前诊断及规划中的肿瘤分割与手术进行中的肿瘤配准问题，并对上述问题进行研究及解决算法的改进。

主要研究内容及成果如下：

1) 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。

2) 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，利用构建多层次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) 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) 针对图像去噪中存在的问题，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。输入是带噪声的图像和原图像，输出的是去噪后的图像。基于传统卷积神经网络的心室图像分割方法研究，传统手工分割方法费时、精确度低，易受操作者经验影响，而传统机器学习算法需要手动筛选设计特征，普适性低，因此本文提出能够综合运用多模图像信息且自动提取结构性特征的卷积神经网络方法。本文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

全篇共五章，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，对当前研究现状及难点进行剖析，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结构。

第二章描述了本文基于传统卷积神经网络的肿瘤分割算法。在这章中，首先分析了传统肿瘤分割算法及学习类算法存在的弊端；接着介绍了传统的医学图像分割算法；然后分析了传统分割类方法存在的问题与不足，引出单通道 CNNs 模型并将其应用到肿瘤图像分割中；最后给出基于 CNNs 的肿瘤分割与识别架构，同时测试了不同实验数据下算法的性能并与当前流行算法进行了比较。

第三章描述了本文的基于多通道 CNNs 的多体位肿瘤分割方法。该章首先描述了传统 CNNs 算法中存在的不足；然后提出了一种综合局部及全局信息的多通道卷积神经网络模型用于精确分割；最后给出该算法的整体流程图；实验结果表明本章方法优于传统 CNNs 模型，且配准精度能与目前最流行的肿瘤分割算法相匹配。

第四章介绍了本文提出的深度迭代 Log-demons 的配准算法。首先描述了传统配准框架在存在大的变形时失效的问题；然后介绍了医学图像配准中的基本概念；接着提出了深度迭代配准框架；最后提出了基于 CNNs 的图像预配准技术，并将其揉合到本文提出的双层迭代配准框架下，实验结果验证了本文方法的鲁棒性。

第五章分析了本文提出的基于 PCA 的相似性测度算法。首先，分析了本章算法提出的问题背景，即要解决传统配准中运算速度慢的问题；之后列出了类似的相似性测度原理及方法；然后详细介绍了基于 PCA 的相似性测度优化，即充分利用

图像中的主要特征并结合传统测度进行优化；最后从三维及二维数据全面验证本文方法的有效性和鲁棒性。

最后结论部分总结全文，并展望了今后的研究工作。



## 第 2 章 相关知识

### 2.1 深度卷积神经网络的组件

#### 2.1.1 网络结构

深度 CNN 是多层前馈神经网络的一种特例。隐藏层的神经元设计成跟上一层神经元局部连接，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。

深度 CNN 的典型结构是在 LeNet 模型<sup>[1]</sup>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<sup>[2]</sup>进行了改进。为 CNN 模型的网络结构示意图。定义图像数据为，且其类别标签，其中和， $k$  为类别数，作为网络输入，输入层的，即原始图像作为输入，第层输出个大小为的特征图。第一层为由个特征图作为输入的卷积层，特征图大小为。第层第特征图定义为。计算公式为：(1) 其中为偏置矩阵，为连接第层第个特征图和第层第个特征的卷积核。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。其函数表达式为：

其中表示对第层的激活函数，该层一般嵌入在卷积层后。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)，如图 1 中所示。最大池化层进行下采样，有时把“卷积-激活-归一化-池化”统称为卷积层。最后需连接全连接层 (图中 Fc 层表示)，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  $K$  维的向量，作为该图像的特征向量送入最终的分类器进行分类评估。

#### 2.1.2 反向传播算法

图 1 卷积网络模型结构示意图 Fig.1 The structure of convolutions model 深度 CNN 模型的分类器与传统方法不同的是：把特征提取过程中的卷积核参数和分类器的参数整合到端到端的模型中。对一个有监督的多分类问题，特征提取过程可表示为得分函数， $W$ ,  $b$  是各层可学习的参数包括卷积核  $K$ , 偏置  $B$  和全连接层的权值参数。对第个样本的得分函数分类误差的交叉熵损失函数可定义为：

通过最小化 Softmax 函数的非负对数似然 (公式 5)，能带来归一化的概率解释。一般采用 L2 损失正则化技术提升分类泛化性能。全部  $N$  个样本的损失函数  $L$  为公式 6 所示。其中表示正则化参数。模型最小化方法采用反向传播算法，通过

带动量的批随机梯度下降算法不断调整参数使得模型整体误差函数不断降低。并通过使用权重衰减项和 Dropout 技术控制过拟合。具体实现详情请参考文献 [10]。

## 2.2 AAM 模型和 CLM 模型

### 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，则第  $i$  个形状向量可定义为： $\mathbf{x}_i$ 。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：(1) 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。和为对角矩阵包含与形状和纹理特征向量相关联的特征值，是图像噪声估计。原始匹配算法使用的是线性回归方法 [6]。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 1 的概率解释 [16,17]：(2) (3) 式 2, 3 为形状模型和外观模型的概率解释，其假设服从零均值，方差的高斯分布。给定模型参数，可以很容易地定义最大似然 (ML) 过程来推断最佳形状和纹理参数：(4) 通过考虑先验分布的最大后验 (MAP) 来估计带正则化项的最优形状  $p$  和最优纹理参数  $c$ ：(5) 公式 5 与公式 1 定义的优化问题等价。

### 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：(6) 式中指示定位正确与否， $C_i$  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 [9]、多通道相关滤波 (MCCF) 的平方误差总和最小滤波器 (MOSSE) [18] 和支持向量回归机 (SVR) [16] 等。拟合 CLM 涉及到解决以下优化问题 [19]：(7) 式中， $\Lambda$  是计算与形状特征向量相关联的特征对角矩阵和  $\mathbf{2}$  是估计的形状噪声。公式 7 可跟 AAM 一样改写为概率形式 [20]：(8) 已经提出了不同的方法仿真模拟真实的响应映射，最常用的是 [19] 的非参数方法 (RLMS)，它将真实的响应图近似为：(9) 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 9 代入 8，得以下优化问题：(10) 这相当于由 8 定义的优化问题，式中响应映射在所有像素位置可评估，视真正的标记点位置  $y_i$  作为潜在变量，式 10 可以使用 EM 算法迭代地求解 [28]。

图 2 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。Fig 2 Qualitative comparison between the three local detector strategies. The MCCF approximates the response map by multi-channel correlation filter and uses RLMSalgoritm to move to the nearest mode of the density. The SVR simply chooses the maximum detector response based on SVM. HG-n represents the number of different HG modules used to obtain the local response map, n take 1,2,4.

### 2.3 小结与讨论



## 第3章 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<sup>[3]</sup>。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<sup>[4]</sup>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<sup>[5]</sup>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<sup>[6,7]</sup>。在文献[4]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<sup>[8]</sup>。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<sup>[9]</sup>，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<sup>[10]</sup>。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(Convoluted Neural Network,CNN)在大规模自然图像数据集(如ImageNet<sup>[11]</sup>)上，识别性能远超传统方法<sup>[2]</sup>。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<sup>[12]</sup>。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<sup>[3]</sup>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<sup>[13]</sup>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<sup>[14]</sup>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。Margeta 等<sup>[15]</sup>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<sup>[16-19]</sup>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<sup>[16]</sup>用反卷积来可视化每个神经元的最大激活值。Mahendran 等<sup>[17]</sup>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<sup>[18]</sup>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<sup>[19]</sup>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 识别超声心动图的方法（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。

### 3.1 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。

#### 3.1.1 全卷积的网络

与 GoogLeNet 模型<sup>[20]</sup>、ResNet 模型<sup>[21]</sup>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<sup>[22]</sup>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3.1所示。

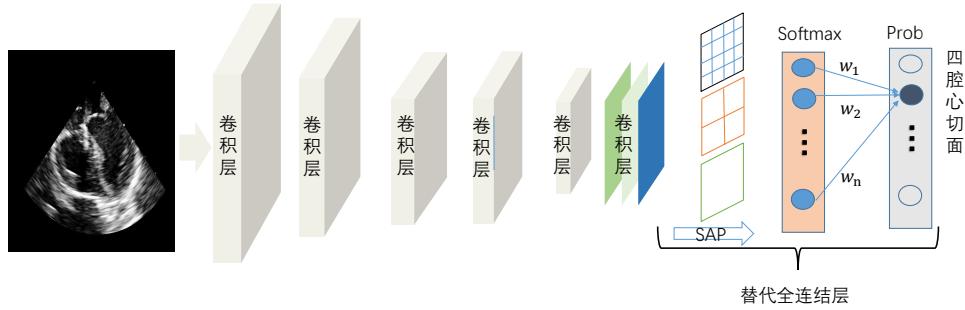


图 3.1: Deep-Echo 模型结构示意图

### 3.1.2 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<sup>[2]</sup> 和 VGG16 模型<sup>[23]</sup> 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<sup>[20]</sup>。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<sup>[22]</sup>。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 ( $1 \times 1$ ,  $2 \times 2$ ,  $4 \times 4$ )，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。

### 3.1.3 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 [19] 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<sup>[19]</sup>。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。

### 3.1.4 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 [19] 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  $(x, y)$  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。

## 3.2 实验结果和分析

### 3.2.1 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3.2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。

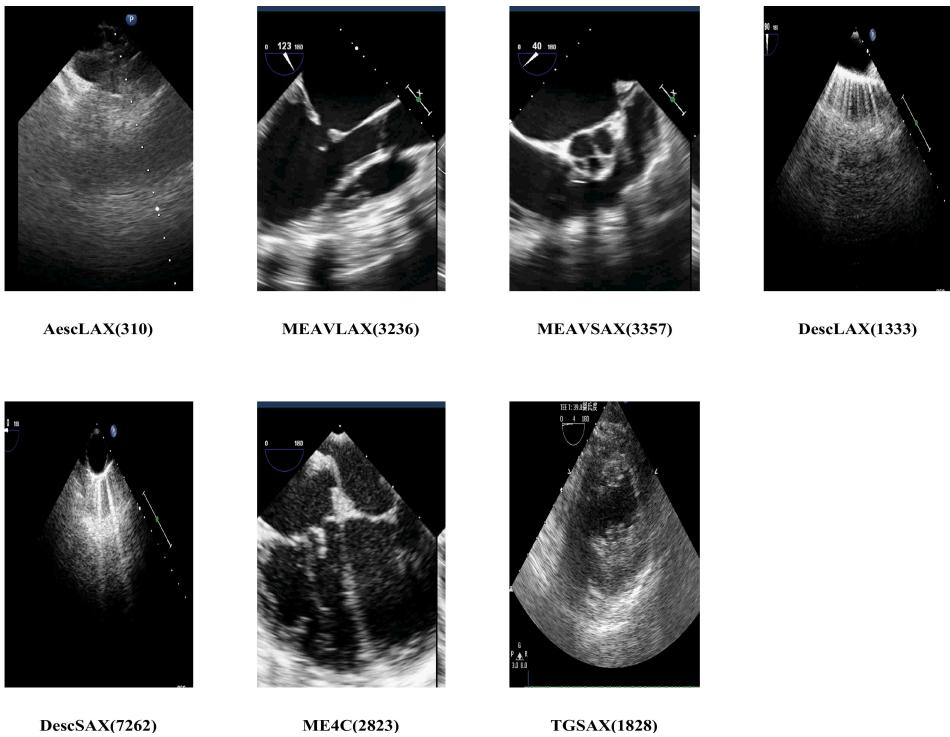


图 3.2: 七类标准切面超声心动图及数量分布

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。

### 3.2.2 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<sup>[24]</sup> 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

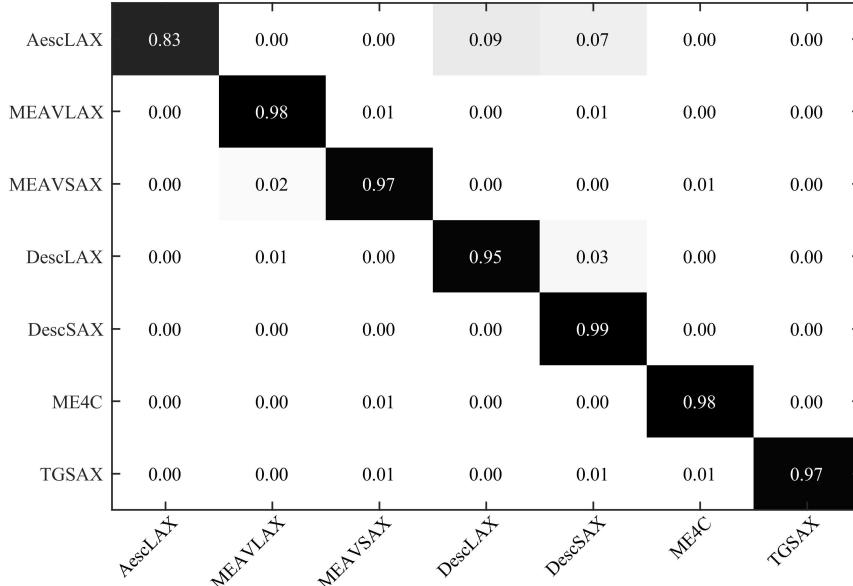


图 3.3: 不同数据量的平均分类精度

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3.3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3.2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。

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	<b>97.49%</b>	<b>99.12%</b>

表 3.1: 不同模型分类精度比较

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3.3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。

### 3.2.3 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，对迁移后的 Deep-echo 模型进行实验。

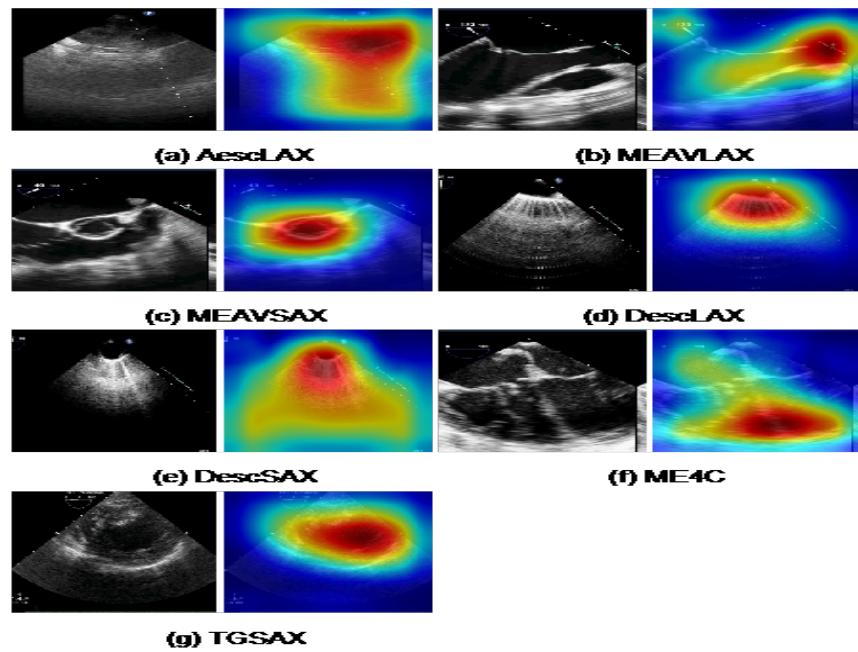


图 3.4: 各类切面的原图和显著性热力图

实验结果如图3.4各类切面的原图和显著性热力图所示，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3.4c 中显著性热力图红色区域图定位到图中的圆圈；图3.4d 中定位到的干涉条纹；图3.4f 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。

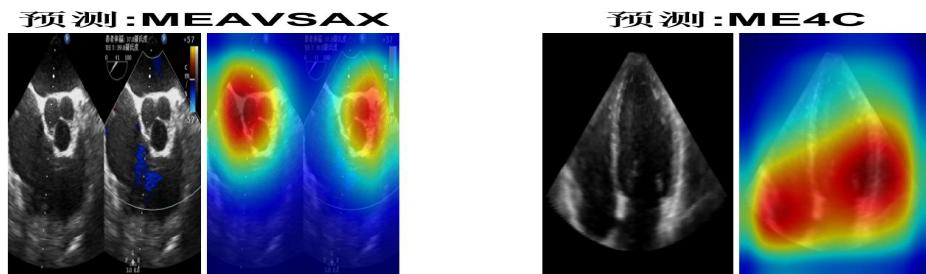


图 3.5: 深度模型泛化性能可视化分析

深度模型泛化性能可视化结果如图3.5所示，原图像分别是带彩色多普勒的双

扇形切面（图3.5a）和经胸的四腔心切面（图3.5b），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。图 7 中可视化结果也能很好的解释模型的有效性，如图3.5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3.5b 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。

### 3.3 小结与讨论

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。

## 第 4 章 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<sup>[16,25–30]</sup>；基于特征重建的方法<sup>[17,18,31,32]</sup>；基于相关性的方法<sup>[33,34]</sup>。基于网络梯度更新的思想是由 Erhan 等<sup>[25]</sup> 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<sup>[35]</sup>。Simonyan 等<sup>[16,26,27]</sup> 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<sup>[16]</sup> 引入 L2 正则化先验（或称权重衰减），改进可视化效果。Yosinski 等<sup>[28]</sup> 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<sup>[26, 30]</sup> 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<sup>[18]</sup> 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<sup>[31]</sup> 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<sup>[17,32]</sup> 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 [34]。受文献<sup>[36,37]</sup>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。

#### 4.1 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  $\theta$  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  $\phi_0$  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 [25] 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  $\phi_0$  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。

#### 4.2 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<sup>[35]</sup>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。

#### 4.2.1 $p$ 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 [16] 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<sup>[17]</sup> 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 [16] 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。

#### 4.2.2 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<sup>[28]</sup>，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<sup>[17]</sup>(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<sup>[38]</sup> 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 [17] 实际实验表明，跟高斯模糊核一样，需随迭代次数动态调整  $\beta$  大小。

#### 4.2.3 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<sup>[26]</sup> 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<sup>[36]</sup> 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<sup>[29]</sup>。

### 4.3 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。

#### 4.3.1 高斯和拉普拉斯金字塔分解

拉普拉斯金字塔 (Laplacian Pyramid, LP)<sup>[39]</sup> 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。

由于自然图像统计特性中的尺度不变性, 也称为  $1/f$  法则<sup>[40]</sup>, 即自然图像集  $I(f_x, f_y)$  的平均傅里叶功谱服从  $I(f_x, f_y)^2$ 。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。

#### 4.3.2 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

## The Laplacian Pyramid

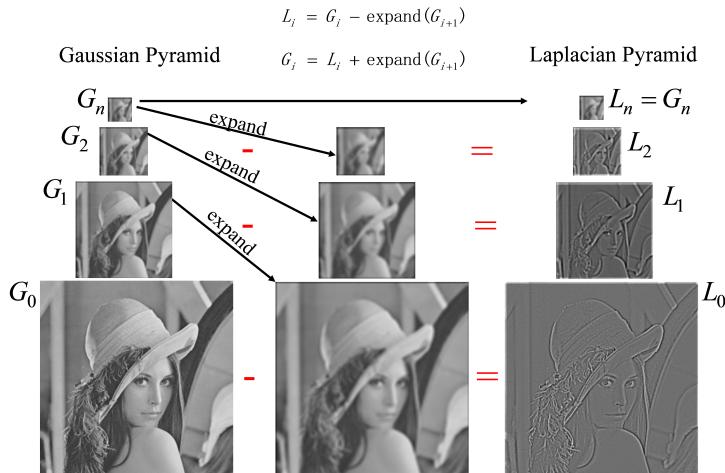


图 4.1: 高斯和拉普拉斯金字塔

每次迭代更新的梯度，则提出梯度归一化操作：

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

式中  $\delta$  为非负小常量， $\text{std}$  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<sup>[41]</sup> 的有效性。

### 4.3.3 类别激活图限制可视化区域

根据文献 [19] 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 [19]，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。

#### 4.3.4 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  $\mu$  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， $\alpha$  为学习率；文献 [17,32] 采用自适应梯度 (Adaptive Gradient, AdaGrad) [42] 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 [43] 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。

### 4.4 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 [44] 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。

#### 4.4.0.1 不同深度模型的类别可视化

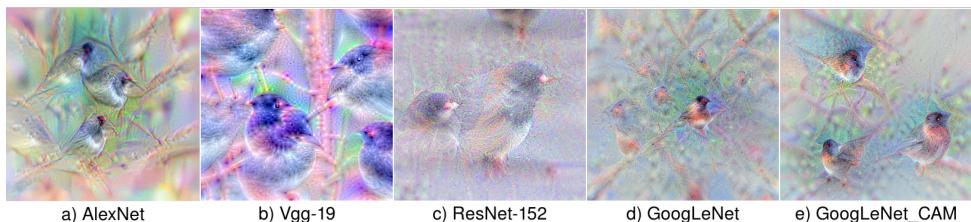


图 4.2：不同模型类别可视化实验结果

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 [2]，Vgg-19 模型 [45]，Google-CAM 模型 [19]，GoogleNet 模型 [20]，ResNet 模型 [21]，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。

为比较不同深度 CNN 模型学习相同类别时特征图的差异，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。

#### 4.4.0.2 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。

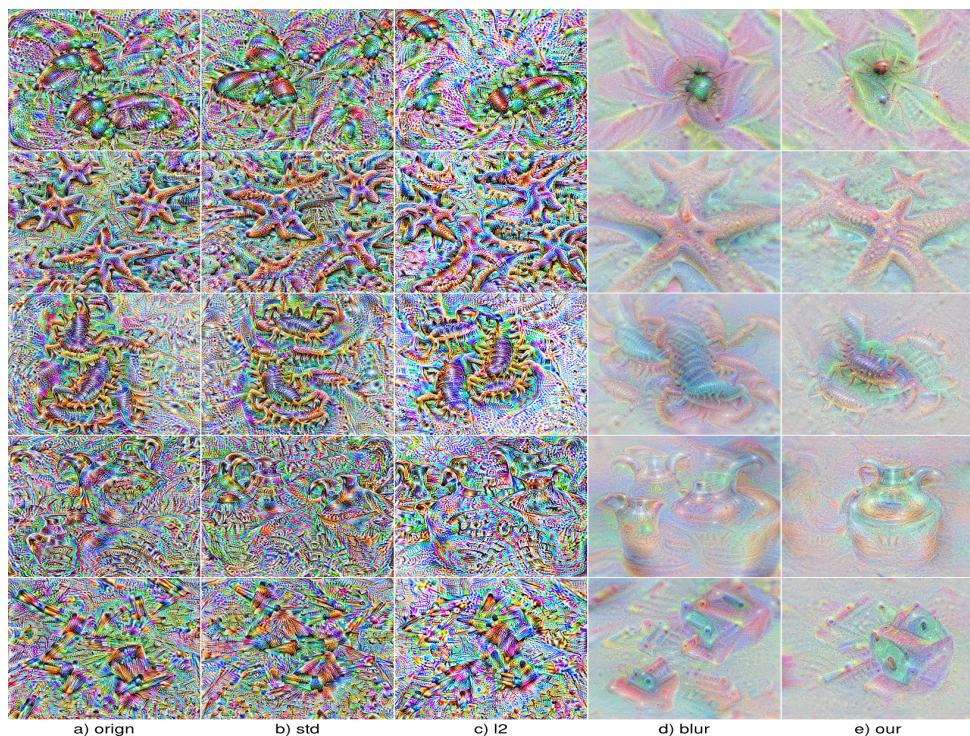


图 4.3: 不同正则化方法的可视化效果

图4.3中 (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<sup>[?]</sup><sup>[2]</sup> 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。

#### 4.4.1 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<sup>[18]</sup>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。

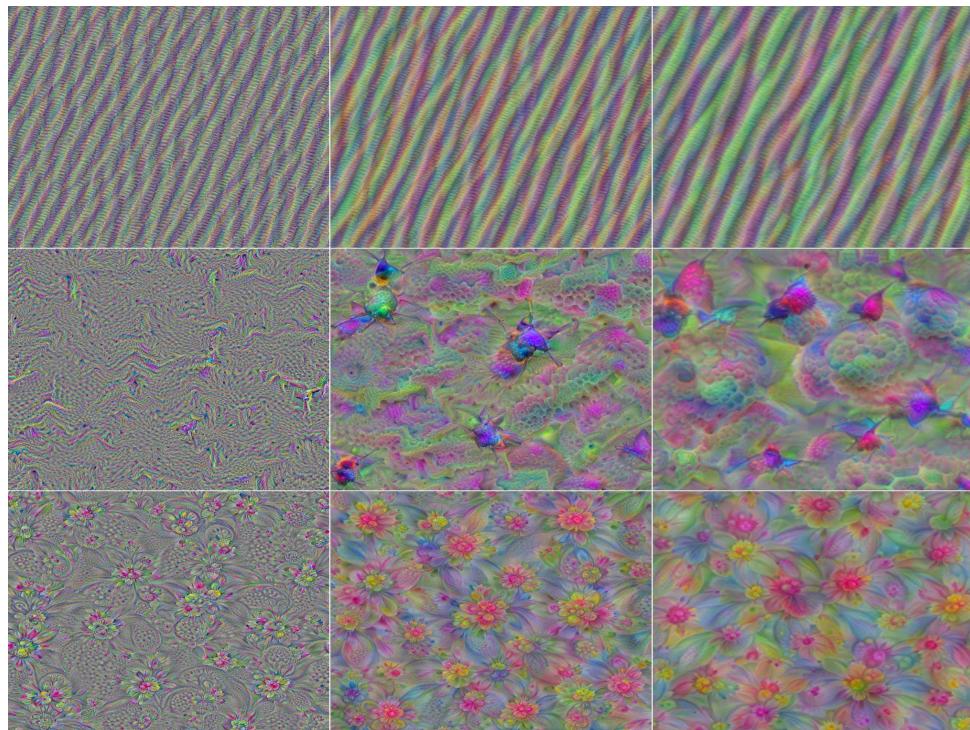


图 4.4: 金字塔分解正则化可视化效果

#### 4.4.2 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。

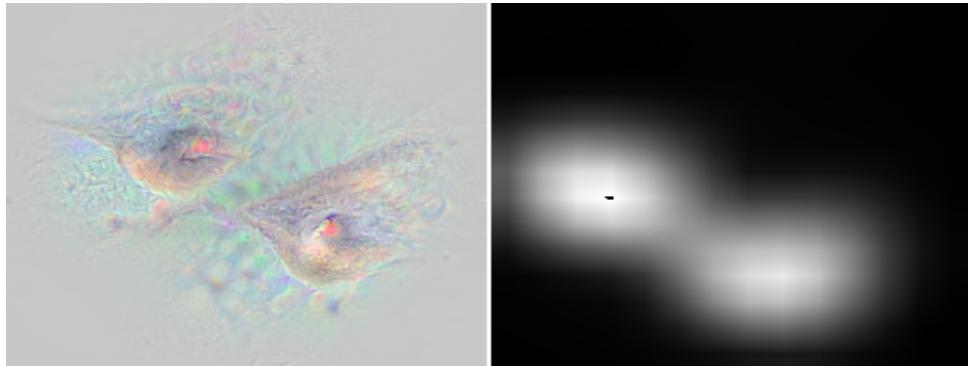


图 4.5: 引入类别激活图的可视化

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。

## 4.5 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。



## 第 5 章 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<sup>[46]</sup>。

不同模态的医学图像中，如超声、计算机断层扫描（Computed Tomography, CT）和核磁共振（Magnetic Resonance Imaging, MRI）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<sup>[47]</sup> 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，随后估计特征强度一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<sup>[48]</sup> 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<sup>[49]</sup>。

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<sup>[50]</sup> 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<sup>[51]</sup> 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<sup>[52]</sup> 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<sup>[53]</sup> 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<sup>[12]</sup>。区域卷积神经网

络<sup>[54]</sup>在大规模自然图像数据集(如ImageNet<sup>[11]</sup>)上,识别性能远超传统方法<sup>[2,54]</sup>。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<sup>[55]</sup>: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<sup>[56]</sup>、结合深度全卷积网络的MRI心室检测与分割<sup>[57,58]</sup>和超声图像解剖结构的检测<sup>[59]</sup>。这些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)<sup>[60]</sup>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。

## 5.1 区域卷积神经网络概览

### 5.1.1 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:

表 5.1: 物体检测算法流程

---

Input: 图像 $x$ , 窗口得分函数 $f$
1: $D :=$ 所有候选框 $r \in R$ 使得 $f(x, r) > 0$
2: 按 $f$ 排序 $D$ 使得 $D_1 \geq D_2 \geq D_3 \geq \dots \geq D_n$
3: 令 $y^* := \{\}$
4: for $i = 1$ to $n$ do
5: 若 $D_i$ 和 $y^*$ 中任意候选框不重叠
6: $y^* := y^* \cup D_i$
7: end for
8: Return: $y^*$ , 物体的目标位置.

---

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5.1。

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

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5.1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滑动窗口位置集合。

### 5.1.2 区域卷积神经网络的演进

2014 年 Girshick 等<sup>[54]</sup> 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<sup>[?]</sup> 和 Fast RCNN<sup>[55]</sup> 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。

## 5.2 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。

### 5.2.1 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<sup>[55]</sup> 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。

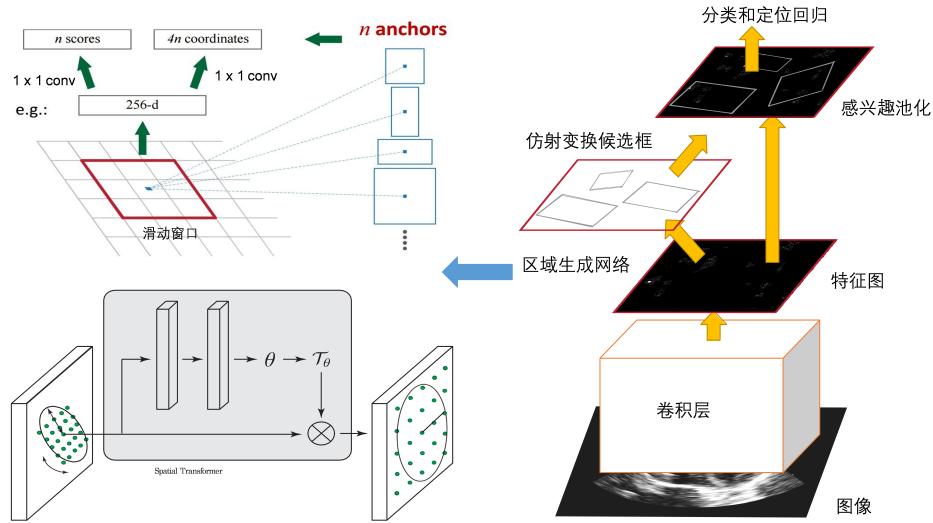


图 5.1: 左上: 引入空间不变性的 anchor 机制 左下: 空间变换网络右: Faster R-CNN 带仿射变换的检测模型框架

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)(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。

### 5.2.2 仿射变换候选框

为检测物体的姿态, 结合空间变换网络<sup>[61]</sup> (见图5.1左下), 提出带仿射变换的候选框生成算法。之前的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:

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

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,  $\tau_\theta$  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,

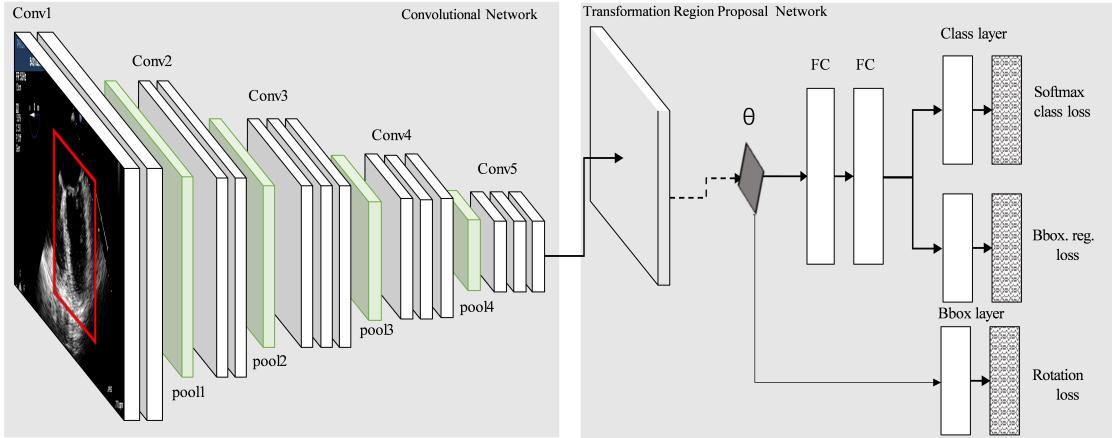


图 5.2: 考虑物体朝向的区域生成网络模型结构示意图, 图中 conv 表示卷积层, pool 表示池化层, FC 表示全连接层, softmax class loss 表示多任务损失中的分类损失, Bbox.reg loss 表示候选框回归定位损失, Rotation loss 表示文中的针对变换参数  $\theta$  的 Von Mise 损失。

宽高坐标需归一化表示, 如  $-1 \leq x_i^s, y_i^s \geq 1$ , 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  $\alpha$  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。

### 5.2.3 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 [62] 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:

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

其中  $p$  指相应的概率密度函数,  $\varphi$  指角度,  $\mu$  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：

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

式中  $\theta$  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  $\varphi$  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：

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

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：

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

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  $C_{VM}$  和  $C_{\cos}$ ，最终朝向回归损失函数为

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

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”，实际上是较小地惩罚小错误。

#### 5.2.4 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。

### 5.2.5 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)<sup>[63]</sup>, 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。

## 5.3 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<sup>1</sup>。

### 5.3.1 检测 MRI 左心室短轴

纽约大学提供的公用数据集<sup>[64]</sup>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病的病例切片。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 [57] 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。

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<sup>[45]</sup> 和 ResNet101<sup>[21]</sup>, 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

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

Method	Accuracy	Sensitivity	Specificity
Baseline[?]	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	<b>99.49%</b>	<b>83.12%</b>	<b>99.40%</b>

表 5.2: 不同模型检测精度的比较

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 [60] 中一致, 迭代次数为 1000, 以文献 [60] 方法作为基准(表5.2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5.2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)。

### 5.3.2 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  $[100010]^T$ 。其它跟 Faster RCNN 中设置一致，其中  $\lambda$ 、 $\mu$ ，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。

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	<b>85.59%</b>	<b>99.12%</b>	<b>67.89%</b>	<b>87.66%</b>	<b>87.48%</b>

表 5.3: 不同模型检测精度，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

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 [62] 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  $15^\circ$  为间隔的指定角度。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 5.3，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	$-45^\circ$	$-30^\circ$	$-15^\circ$	$+15^\circ$	$+30^\circ$	$+45^\circ$	Avg
Compute	66.65%	78.02%	87.39%	85.53%	75.83%	62.31%	75.94%
Pred	<b>73.09%</b>	<b>81.72%</b>	<b>81.75%</b>	<b>89.56%</b>	<b>80.48%</b>	<b>70.31%</b>	<b>80.76%</b>

表 5.4: 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 [65] 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。

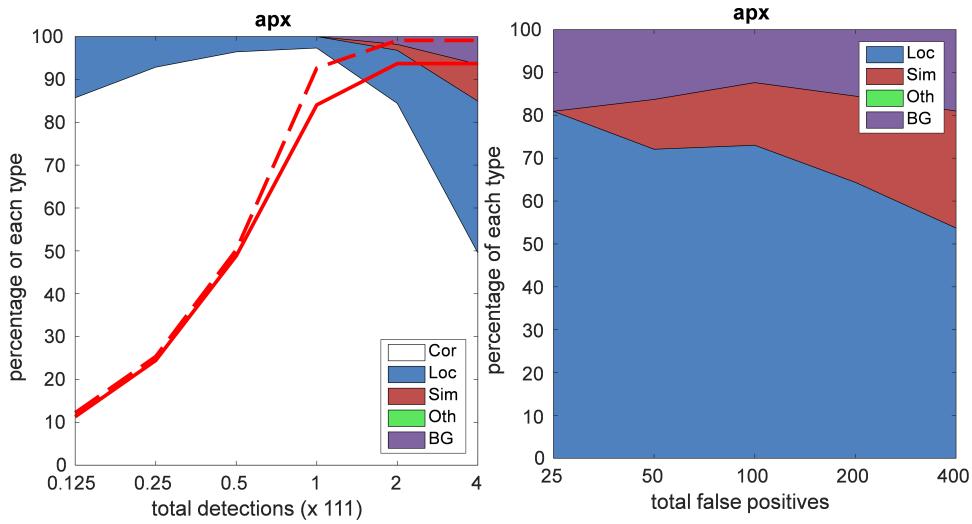


图 5.3: 左图显示 apx 检测精度的累积分布: 正确的 (Cor) 或定位不准确 (Loc) 的假阳性, 与之混淆类似类别 (Sim) 与其他类别 (Oth) 或背景 (BG)。固体红色线是以“强”标准 (大于 0.5 交叠比), 反映精确度随检测增加而变化。红色虚线使用“弱”标准 (大于 0.1 交叠比)。右图显示排名靠前的假阳性类型的分布。

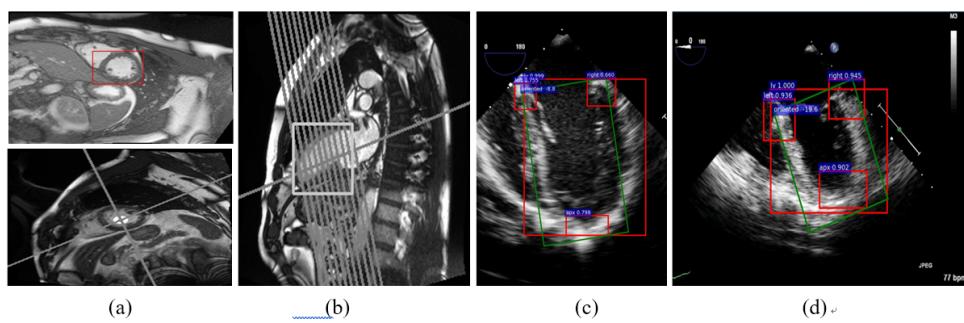


图 5.4: (a,b) 表示不同 MRI 图像检测左心室结果, (c,d) 两图表示超声心动图的 ME4C 切面的左心室、二尖瓣环、心内膜垫和心尖位置及旋转角度的检测结果。

#### 5.4 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。



## 第 6 章 图像的去噪方法

图像去噪的目的是从噪声图像中恢复出干净的图像，这是低级视觉任务中的一个经典问题。由于干净的图像通常是未知的，因此这个问题本质上是不适当的。换句话说，它是一个欠定的映射问题，其中图像变换不是唯一的。一般来说，残差图像  $F$  可以表示为  $F = yf x$ ，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。 $y$  是理想的清洁图像。通过适应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级别成像问题，如图像去模糊，去马赛克和超分辨率。

最近，深度神经网络在计算机视觉领域表现出了卓越的性能，从高级到低级任务。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<sup>[66]</sup>。在图像去噪设置中，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值。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地面真实图像之间的差异。最广泛使用的是每像素损失，其通过失真和参考图像像素的强度差异以及相关的峰值信噪比<sup>[67]</sup> 来计算。但是，每像素损失不能捕捉到感知差异，并且众所周知与感知图像质量的关联性很差<sup>[78?]</sup>。这是因为当使用每像素丢失时隐含地做出的许多假设不被满足。可以说，最重要的是独立于图像的局部特征来处理噪声；相反，人类视觉系统（HVS）对噪声的敏感性取决于局部亮度，对比度和结构<sup>[67]</sup>。

基于纯粹的学习策略，为图像去噪设计的一组深度神经网络已被证明优于其他被广泛接受的方法作为最先进的<sup>[68]</sup>。但是，所有这些工作都存在一个问题：如果输入无噪音，则学习模型也会降低干净的图像质量。所以他们唯一的工作就是在他们接受训练的给定噪音水平下工作。用于噪声消除的标准通用算法应该能够处理不同级别的噪声，这种限制似乎需要一系列这样的网络，每个噪声级别一个。这是不切实际的，甚至是不现实的。因为我们不知道噪音水平和真实图像的类型。

我们的主要贡献简要概述如下：

1. 我们提出了一个非常深的全卷积体系结构，用于图像残差去噪。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. 结合每像素和感知丢失函数的优点，训练具有低和高级别信息的转换网络，生成高质量的去噪图像。
3. 为使单个神经网络适用于所有噪音级别，我们研究网络的统计规律：使输入从不同噪音级别添加随机样本，输入也可以是干净的图像，因为学习的图像

函数必须对干净的图像进行身份验证，训练有素的网络可以自动处理不同的级别。

4. 我们试验了一些常见的基准图像。结果表明我们的网络的优点和所提出的新模型损失层克服了其他最近的图像去噪方面的最新技术方法。

## 6.1 相关工作

已经提出了许多用于图像去噪的方法。一些有选择地平滑噪声图像的部分，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传送到可以容易地从信号中分离噪声的替代域。最近的方法利用图像的“非本地”统计：相同图像中的不同补丁在外观上通常相似。块匹配和 3D 过滤（BM3D）算法<sup>[69]</sup>通过协作过滤在变换域中对非局部相似补丁进行分组。BM3D 已经成为图像去噪的基准。

虽然 BM3D 是一种设计良好的算法，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方法最显着的区别在于，它们通常自动直接从干净而嘈杂的图像中自动学习图像转换，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<sup>[70–74]</sup>，多层感知器<sup>[68,75]</sup>，卷积网络<sup>[76–82]</sup>。

叠加去噪自动编码器<sup>[70]</sup>建立了使用去噪标准作为无监督目标的价值，以指导有用的不同级别的表示的学习。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 等人提出了一种替代的监督训练方案 - 叠加稀疏降噪自动编码器（SSDA），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（DA）进行预训练，成功地将 DA 最初设计用于无监督特征学习，适用于图像去噪和盲目修补的任务。

Burger 等<sup>[68]</sup>提出了一种基于斑块的算法，该算法在具有普通多层感知器的大型数据集上学习，其性能优于 BM3D。然而，他们的方法适合于单一级别的噪音，并且不能很好地推广到其他噪音级别。Jain 等人<sup>[76]</sup>提出了深度卷积神经网络和无监督学习过程，该过程综合了特定噪音模型的训练样本。他们发现卷积网络在小波和马尔科夫随机场（MRF）方法中提供了可比较的并且在某些情况下更好的性能。

到目前为止，通过训练具有每像素损失函数的深度神经网络，已经解决了所有用于图像去噪任务的神经网络。但是在图像处理的情况下，这可能会受到特别的限制，因为每像素丢失与感知图像质量的相关性很低<sup>[78]</sup>。一些最近的论文已经使用优化来生成目标是感知的图像，这取决于从卷积网络提取的高级特征<sup>[83]</sup>。<sup>[79,84]</sup>的工作与我们的工作特别相关，因为他们训练前馈神经网络以进行图像转换，他们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，以测量输出的感知差异和地面实况。不过，他们专注于风格转换和图像超解像。后来的工作提出了编码器和解码器图像细节的卷积和解卷积层，并且使用对称跳跃连接来加速训练。

但是图像的噪声只能通过卷积来捕获，并通过解卷积来恢复图像细节。我们的网络可以看作是具有对称跳转连接的整个转换函数。

Zhao 等人<sup>[78]</sup>研究了包括感知驱动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。从感知动机指标设计出一些新的损失层，它仍然专注于低级像素结构。Wang 等人的另一项工作<sup>[75]</sup>也与我们的研究特别相关，他们研究了自然图像斑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在整个各级高斯噪声。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，将图像变换任务和基于优化的图像变换方法的优点相结合。同时通过显式训练不同级别的噪声并使原始图像作为输入，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。

## 6.2 提出方法

The proposed framework mainly contains a chain of convolution layers and deconvolution layers, as shown in Figure 6.1. There consists of two components: an *image transformation network*  $f_W$  and a *loss network*  $\phi$  that is used to define several *loss functions*  $\ell_1, \dots, \ell_k$ . Aim at learning the deep residual convolutional neural network parameterized by weights  $W$ ; it transforms input images  $x$  into output images  $\hat{y} = f_W(x)$ . Each loss function computes a scalar value  $\ell_i(\hat{y}, y_i)$  measuring the difference between the output image  $\hat{y}$  and a *target image*  $y_i$ . Learning objective is trained using stochastic gradient descent to minimize a weighted combination of loss functions:

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

From this formulation, we can see that the task here is to find a mapping function  $f_W$  that best approximates the image transformation. Meanwhile we also want the  $f_W(y_i)$  approximates the image  $y_i$ , so we now treat the image denoising

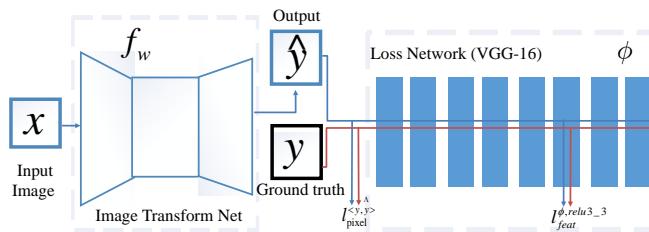


图 6.1: The overall architecture of our proposed network. The image transformation network contains layers of convolution (encoder) and deconvolution (decoder). We use a loss network pretrained for image classification to define perceptual loss functions that measure perceptual differences in output and ground truth label. The loss network remains fixed during the training process.

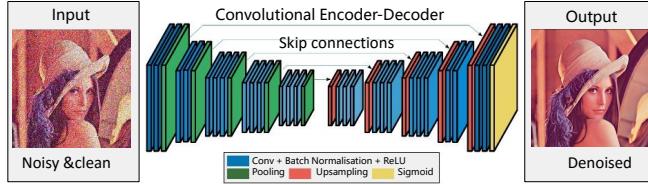


图 6.2: The architecture of RED-NET and our image transformation network DeNET can be viewed as the RED-NET inserted some residual block.

problems in a unified framework by choosing appropriate weights  $W$  with different situations. The loss network  $\phi$  is used to define a *feature space loss*  $\ell_{feat}^\phi$  and a *per-pixel loss*  $\ell_{mse}$  that measure differences in feature and image space. For image denoising, the input image  $x$  is a noisy input, the ground-truth image  $y$  is also input image and target image.

### 6.2.1 Encoder-decoder Architectures

The framework is fully Encoder-decoder convolution model. Combining layers of encoder and decoder [79,85–88] have been proposed for unsupervised and supervised deep learning.

Our image transformation networks roughly follow the architectural guidelines set forth by [79]. They propose RED-NET for image denoising, the network as shown in Figure ???. Base on the architecture , The batch normalization [89]and ReLU nonlinearities layers are added after convolution. And insert some residual blocks [90] into the network. The output layer, which uses a sigmoid function to ensure that the output image has pixels in the range  $[0, 1]$ . But we do not use any pooling layers, instead using strided and fractionally strided convolutions for downsampling and upsampling. Other than the first and last layers which use  $9 \times 9$  kernels, all convolution layers use  $3 \times 3$  kernels. Since the image transformation networks are fully-convolutional, at test-time they can be applied to images of any size.

The difference between RED-NET and ours DeNET is that our network insert some residual block and introduce residual connection. The noise is eliminated step by step after each layer. During this process, the details of the image content can be compensate by the perceptual loss function. The specific configurations of the two networks are described in Table 6.1.

Two learning strategy is applied to inner blocks of the encoding-decoding network to make training more effective. Skip connections are passed every two convolutional layers to their mirrored deconvolutional layers. He et.al[90] use residual connections to train very deep networks for image classification. They argue that residual connections make the network to learn the identify function easily; this is

表 6.1: Configurations of the DeNET-R and RED-NET networks. "conv3" and "deconv3" stand for convolution and deconvolution kernels of size  $3 \times 3$ . 32,128 and 512 is the number of feature maps after each convolution and deconvolution. " $c$ " is the number of channels of input and output image. i.e.,  $c = 3$ .

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 $\times 5$	
$(\text{deconv3-64}) \times 2$	$(\text{deconv3-512}) \times 2$
$(\text{deconv3-32}) \times 6$	$(\text{deconv3-512}) \times 6$
$(\text{deconv9-3}) \times 6$	$(\text{deconv3-512}) \times 6$
$(\text{deconv3-}c)$	$(\text{deconv3-}c)$

an appealing property for image transformation networks, since in most cases the output image should share structure with the input image. The body of our network thus consists of several residual blocks, each of which contains two  $3 \times 3$  convolution layers.

### 6.2.2 Per-pixel Loss Functions

The *pixel loss* is the (normalized) Euclidean distance between the output image  $\hat{y}$  and the target  $y$ . If both have shape  $C \times H \times W$ , then the pixel Euclidean loss is defined as Mean Squared Error(MSE):

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

This loss function can introduce splotchy artifacts, So we also examine the  $\ell_1$ -norm loss. The two losses weigh errors differently: $\ell_1$  does not over-penalize larger errors and consequently, they may have different convergence properties.Computing the  $\ell_1$  loss is straightforward:

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

The derivatives for the back-propagation are also simple, for each pixel  $p$  in the whole image,

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

Note that, although  $L_{\ell_1}$  is computed on the whole image, the derivatives are back-propagated for each pixel in the image. The network trained with  $\ell_1$  provides a significant improvement for several of the issues discussed above.

### 6.2.3 Perceptual Loss Functions

We define *perceptual loss functions* that measure perceptual and semantic differences between images, other than the hand design SSIM loss in<sup>[78]</sup>. They make use of a *loss network*  $\phi$  pretrained for image classification. In all our experiments  $\phi$  is the 16-layer VGG network<sup>[91]</sup> pretrained on the ImageNet dataset<sup>[92]</sup>. Rather than encouraging the pixels of the output image  $\hat{y} = f_w(x)$  to exactly match the pixels of the target image  $y$ , we instead encourage them to have similar feature representations as computed by the loss network  $\phi$ . Let  $\phi_j(x)$  be the activations of the  $j$ th layer of the network  $\phi$  when processing the image  $x$ ; if  $j$  is a convolutional layer then  $\phi_j(x)$  will be a feature map of shape  $C_j \times H_j \times W_j$ . The *feature feat loss* is the (squared, normalized) Euclidean distance between feature representations:

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

The Euclidean distance also can be alternated by  $\ell_1$ -norm distance. As demonstrated in<sup>[17]</sup>, finding an image  $\hat{y}$  that minimizes the feature feat loss for early layers tends to produce images that are visually indistinguishable from  $y$ . Using a feature feat loss for training our image transformation networks encourages the output image  $\hat{y}$  to be perceptually similar to the target image  $y$ , but does not force them to match exactly. To encourage spatial smoothness in the output image  $\hat{y}$ , we follow prior work on feature inversion<sup>[17]</sup> and make use of *total variation regularizer*  $\ell_{TV}(\hat{y})$ .

## 6.3 实验结果分析和讨论

In this section, We provide an analysis on our experiments setting with fully encoder-decoder convolutional neural network. Then evaluate denoising performance of our models under some different loss function setting. At the end we explore how to make one neural network handle different levels of noise.

### 6.3.1 Analysis On Model Details

We train models to perform single and multi level of standard deviation  $\sigma$  by minimizing some loss function:  $\ell_2$ -Mean Squared Error(MSE) loss,  $\ell_1$ -norm loss and feature feat loss at layer `relu1_1` from the VGG-16 network  $\phi$ . We train image size with  $256 \times 256$  from the training set, and prepare noisy inputs by add a Gaussian kernel of width  $\sigma$ . We train with a batch size of 10 using Adam<sup>[93]</sup> with a learning rate of  $1 \times 10^{-3}$  without weight decay or dropout.

Denoising experiments are performed on the standard 14 common benchmark images Set14. As a common experimental setting in the literature, additive Gaussian noises with zero mean and standard deviation  $\sigma$  are added to the test image to test the performance of denoising methods. We report PSNR and SSIM<sup>[67]</sup>, computing both on the three channel color image, following<sup>[78,79]</sup>.

As a baseline model we use RED-NET<sup>[79]</sup> for its state-of-the-art performance. It is a fully convolutional network with convolutional and deconvolutional layers trained to minimize per-pixel loss. To account for differences between RED-NET and our model in data, training, and architecture, we train image transformation networks for the same standard deviation  $\sigma$  using  $\ell_2$ ; these networks use identical data, architecture, and training as the networks trained to minimize other loss function. We train denoising networks with the per-pixel loss typically used<sup>[78,79]</sup>, also with a feature feat loss (see Section 6.2) to allow transfer of semantic knowledge from the pretrained loss network to the denoising network as supervised signal guided denoising.

First of all, Compared to the per-pixel loss  $\ell_1$  and  $\ell_2$  result,  $\ell_1$  does a better good job at denoising performance and meanwhile restore sharp edges and fine details. As show in Figure 6.3 the wing in the  $\ell_2$  image and the red color block elements of the body in the  $\ell_2$  image. This is because  $\ell_2$  penalizes larger errors, but is more tolerant to small errors, regardless of the underlying structure in the image; The conclusion is consistent with the literature<sup>[78]</sup>.

Moreover, Results for  $\ell_{feat}$  are Show in Figure 6.3 when only with the feature feat loss gives rise to a slight cross-hatch pattern visible under magnification, which harms its PSNR and SSIM compared to baseline methods. Again we see that our  $\ell_{feat}$  model does a good job at edges and fine details compared to other models, such as the wing. The  $\ell_{feat}$  model does not sharpen edges indiscriminately; compared to

表 6.2: Quantitative single-level image denoising results on the Set14; we report average PSNR and SSIM on each dataset. each  $\sigma$  value we train identical networks, one with a per-pixel loss  $\ell_1, \ell_2$  and another with a feature feat loss  $\ell_{feat}, \ell_{mix}$  is combine with  $\ell_1$  and  $\ell_{feat}$ . Best results are shown in bold.

Sigma	Noisy PSNR / SSIM	RED-NET <sup>[79]</sup> PSNR / SSIM	Ours ( $\ell_2$ ) PSNR / SSIM	Ours ( $\ell_1$ ) PSNR / SSIM	Ours ( $\ell_{feat}$ ) PSNR / SSIM	Ours ( $\ell_{mix}$ ) PSNR / SSIM
$\sigma = 10$	28.16 / 0.7041	<b>34.81 / 0.9402</b>	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 / 0.8591	26.70 / 0.6845	<b>30.15 / 0.8681</b>
$\sigma = 50$	14.79 / 0.2038	26.81 / 0.7733	26.40 / 0.8205	26.79 / <b>0.8325</b>	25.69 / 0.6411	<b>27.09 / 0.8312</b>
$\sigma = 70$	12.43 / 0.1391	25.31 / 0.7206	25.39 / 0.7105	26.13 / <b>0.7250</b>	17.89 / 0.6650	<b>26.20 / 0.7180</b>
$\sigma = 100$	10.26 / 0.0901	-	18.40 / 0.4215	20.19 / 0.4680	17.31 / 0.3640	<b>19.16 / 0.4695</b>

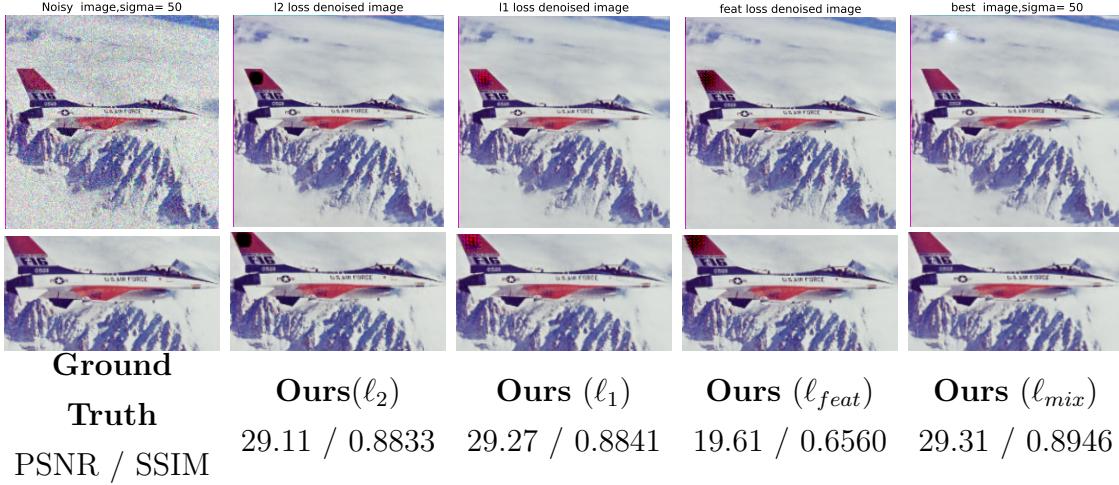


图 6.3: Denoising results with different loss type on an image from the Set14 dataset. We report PSNR / SSIM for the F16-plane image as a example.

the  $\ell_{pixel}$  model, the  $\ell_{feat}$  model sharpens the boundary edges of the wing and rider but the background mountain remain diffuse, suggesting that the  $\ell_{feat}$  model may be more aware of image semantics.

Since our  $\ell_{pixel}$  and our  $\ell_{feat}$  models share the same architecture, data, and training procedure, all differences between them are due to the difference between the  $\ell_{pixel}$  and  $\ell_{feat}$  losses. The  $\ell_{pixel}$  loss gives fewer visual artifacts and higher PSNR values but the  $\ell_{feat}$  loss does a better job at reconstructing fine details, leading to pleasing visual results.

Last, we can observe that a single model can work across all levels of Gaussian noise, thereby allowing to reduce significantly the training time for a general-purpose neural network powered denoising algorithm. The reason for this may be that the learned image transformation function can model the Gaussian-like distribute from any levels and other types. Interesting is that shown in Figure 6.4, even for other type of noise ,such as speckle noise,poission-distributed noise,salt noise or pepper noise, which the model trained for Gaussian noise have ability for image denoise.

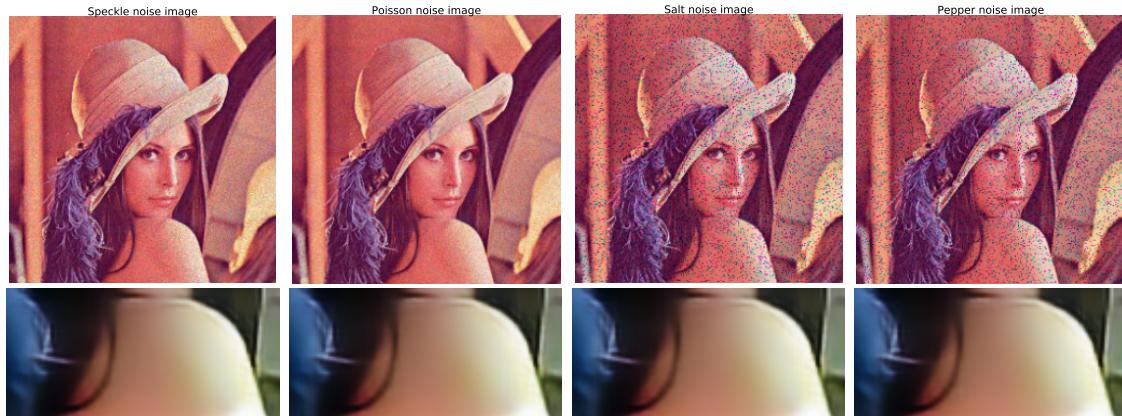


图 6.4: Comparison denoising performance of four other noise types ,which transformation networks trained only with Gaussian noise. **Up:** The four types noise image:Speckle noise,Poisson distributed noise,Salt noise,Pepper noise. **Down:** Denoising sub-outputs from corresponding noise type.

#### 6.4 小结与讨论



## 第 7 章 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<sup>[94]</sup>。

Hansson 等<sup>[95]</sup> 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<sup>[96]</sup> 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<sup>[97]</sup>，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）<sup>[98]</sup> 和基于全局外观主动外观模型（Active Appearance Model, AAM）<sup>[99]</sup> 用于超声心动图分割已被证明是非常有效的<sup>[94,100,101]</sup>。原始 ASM 在超声图像中存在许多缺陷<sup>[97]</sup>，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<sup>[102]</sup>（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表观建模，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<sup>[103]</sup>。

近来，级联形状回归模型<sup>[104]</sup> 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<sup>[105]</sup> 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<sup>[106]</sup>，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：

- 1) 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。
- 2) 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。
- 3) 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。

## 7.1 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<sup>[107]</sup>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<sup>[108]</sup>，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7.1(a) 所示，其中形变部件模板如图7.2(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 [106] 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7.1(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7.1(c) 所示。

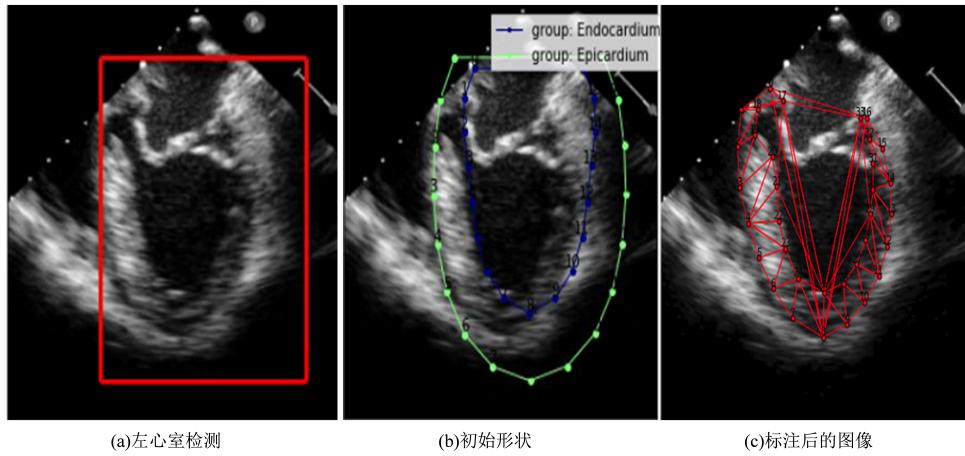


图 7.1: 初始位置定位结果和特征点标注示意图

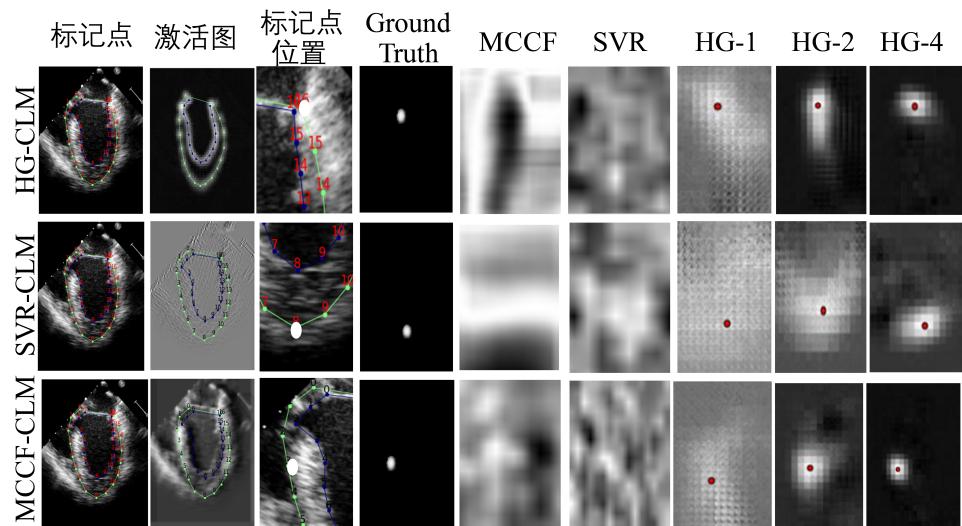


图 7.2: 定性比较三种不同特征激活图及相应的局部响应映射图, MCCF 通过多通道相关滤波器近似响应图, 且用 RLMS 算法移动到最优位置。SVR 基于支持向量机简单地选择最大响应位置。HG-n 表示所用不同 HG 模块数的局部响应图, n 取 1, 2, 4。

## 7.2 结合卷积网络特征的形状对齐模型

### 7.2.1 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<sup>[94]</sup>，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 [94] 提出的非线性归一化来处理偏斜和偏移的灰度分布。

### 7.2.2 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7.2表示采用灰度（图7.2(b)），hog 特征（图7.2(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。

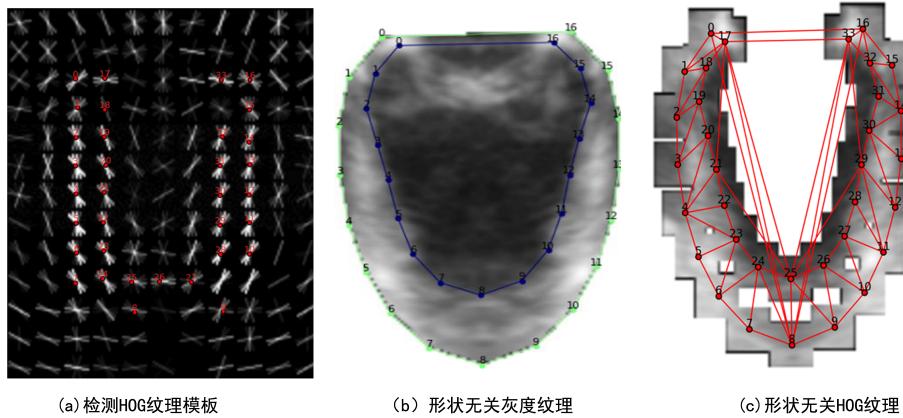


图 7.3: 不同特征的形状无关纹理图

### 7.2.3 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)<sup>[109]</sup> 和支持向量回归机 (SVR)<sup>[110]</sup> 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7.2）。在我们的模型中，这是由堆叠多级沙漏全卷积网络 (Hourglass Network, HG)<sup>[111]</sup> 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7.2），网络结构如图7.3所示。

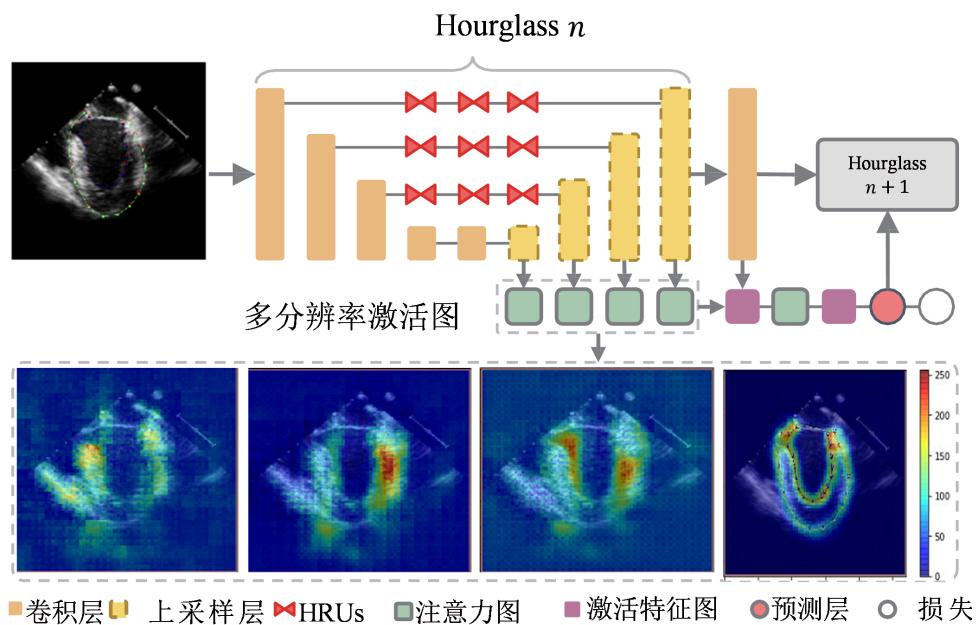


图 7.4: 在每个沙漏网络中，从具有不同分辨率的特征生成多分辨率注意力图，这些图加和成单一的注意力图，它用于生成激活特征图。

图7.3中的网络基本组件是一种基于残差网络<sup>[21]</sup>。“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  $3 \times 3$  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<sup>[112]</sup>，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74.1 节中给出。在公式 10 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：(11) 式中  $l_i$  表示第  $i$  个标记点，图像的位置  $x_i$  处的图像  $I_{x_i}$ ，响应映射  $i$  用于最小化等式 10。我们的消融实验实验表明，增加 HG 模块数，显著影响分割性能。

#### 7.2.4 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 5 和式 10 可以重写为以下优化问题：(12) 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，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。

#### 7.2.5 模型匹配代价函数的优化

等式 12 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<sup>[110]</sup> 和 RLMS 算法，增量形状参数  $p^*$  的最优解由下式给出：(13) 其中：(14) 其中是反向位置的海森矩阵 (Hessian)。和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<sup>[110]</sup> 与 RLMS 组合：(15) 在这种情况下，Hessian 和 Jacobian 被定义为和，有关如何计算  $W/p$  的更多细节有兴趣的读者请参考<sup>[110]</sup>，最佳纹理参数  $c^*$  由式 5 给出，且两种算法仍然使用式 13 定义的完全相同的更新规则得到  $p^*$  的最优值。

### 7.3 实验结果分析和讨论

#### 7.3.1 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）：(16) 式中表示  $n$  个特征点的两个形状和， $l_t$  和  $r_b$  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  $1 \times 10^{-3}$ ，网络参数由 Adam 算法<sup>[113]</sup> 优化，网络中开始是步长为 2，核大小为  $7 \times 7$  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。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。

### 7.3.2 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。

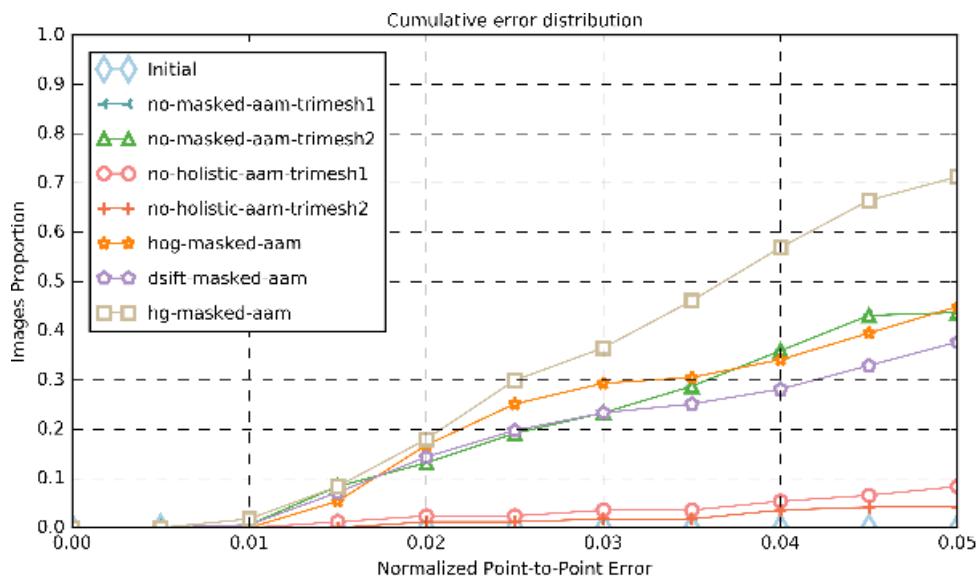


图 7.5: 不同特征 AAM 的左心室内外膜分割性能比较

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。

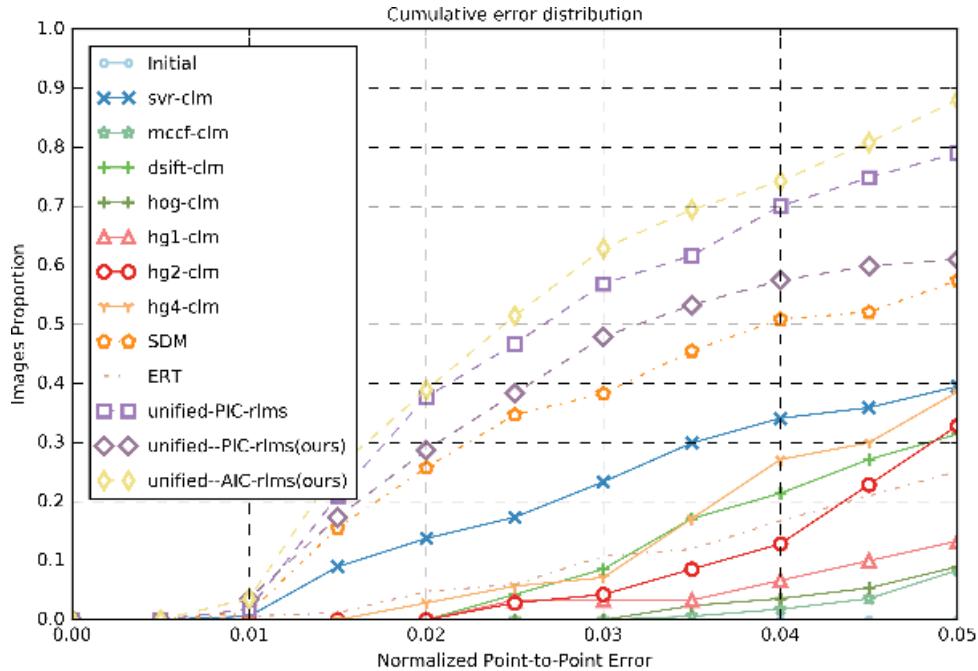


图 7.6: 分割结果对比

分割性能见图7.5，外观特征比较了原始像素（no）、dsift<sup>[114]</sup>、HOG 和 HG 特征，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7.5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7.3b）与真实心动图差异仍较大，导致分割效果较差（图7.5中 no 曲线），主要原因是因为 AAM 方法对初始值敏感，之前文献<sup>[94,95]</sup> 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<sup>[106]</sup>，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。

### 7.3.3 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7.6中 unified-PIC-rlms 表示采用文献<sup>[110]</sup>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<sup>[104]</sup>和 sdm 算法<sup>[115]</sup>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

表 7.1: 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7.1。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7.1）。由图7.7a 中可见，本文方法结果与专家标注比较接近。

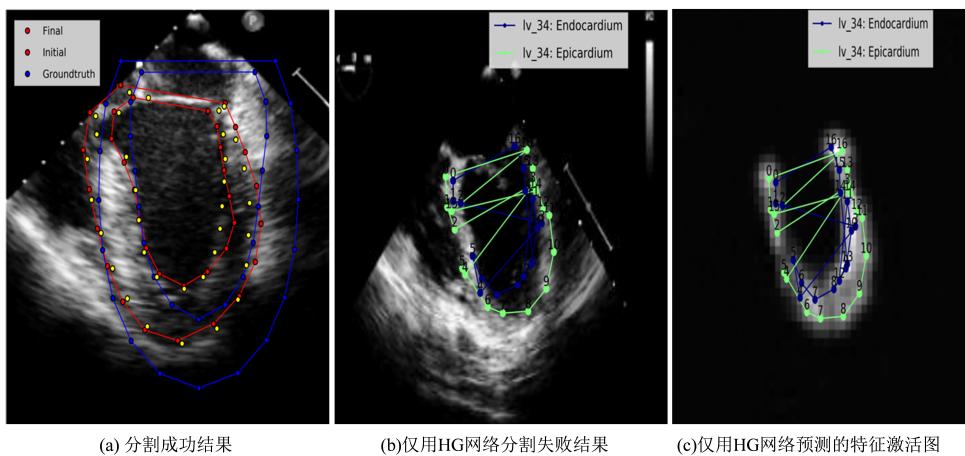


图 7.7: 用 HG 网络预测心室内外膜成功和失败的案例

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7.7c），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。

#### 7.4 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。

## 第 8 章 总结与展望



## 参考文献

- [1] 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.
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[C]//Advances In Neural Information Processing Systems. 2012: 1–9.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.

- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.

- [24] 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.
- [25] 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.
- [26] LENC 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.
- [27] SZEGEDY C, ZAREMBA W, SUTSKEVER I. Intriguing properties of neural networks[C/OL]//Int. Conf. Learn. Represent. 2014: 1–10.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] DOSOVITSKIY A, BROX T. Inverting Visual Representations with Convolutional Networks[C/OL]//2015 IEEE Conf. Comput. Vis. Pattern Recognit. 2015: 184–199.
- [32] MAHENDRAN A, VEDALDI A. Visualizing Deep Convolutional Neural Networks Using Natural Pre-images[J/OL]. Int. J. Comput. Vis., 2016: 1–23.
- [33] 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.
- [34] 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.
- [35] GOODFELLOW I J, SHLENS J, SZEGEDY C. Explaining and Harnessing Adversarial Examples[C/OL]//Int. Conf. Learn. Represent. San Diego, USA, 2014: 484–485.
- [36] 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.
- [37] 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.

- [38] 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.
- [39] BURT P, ADELSON E. The Laplacian Pyramid as a Compact Image Code[J]. IEEE Trans. Commun., 1983, 31(4): 532–540.
- [40] VAN DER SCHAAF A, VAN HATEREN J H. [Z].
- [41] 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.
- [42] 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.
- [43] GATYS L A, ECKER A S, BETHGE M. Texture Synthesis Using Convolutional Neural Networks[J]. Adv. Neural Inf. Process. Syst., 2015: 262–270.
- [44] 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.
- [45] 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.
- [46] 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.
- [47] SCHÖLLHUBER A. Fully Automatic Segmentation of the Myocardium in Cardiac Perfusion MRI[J/OL]. Engineering in Medicine, 2008, 3(22): 12–19.
- [48] 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.
- [49] 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.
- [50] 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.
- [51] ZHOU S K, GEORGESCU B, ZHOU X S, et al. Image based regression using boosting method[C/OL]//Proc. IEEE Int. Conf. Comput. Vis.: I. 2005: 541–548.
- [52] 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.
- [53] ZHENG Y, COMANICIU D. Marginal Space Learning for Medical Image Analysis [M/OL]. New York, NY: Springer New York, 2014: 199–256.

- [54] 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.
- [55] GIRSHICK R. Fast R-CNN[C/OL]//2015 IEEE Int. Conf. Comput. Vis.: 2015 Inter. IEEE, 2015: 1440–1448.
- [56] 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.
- [57] 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.
- [58] PARK S R, LEE J. A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis MRI[J/OL]. arXiv Prepr., 2016: 1–21.
- [59] 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.
- [60] 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.
- [61] 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.
- [62] 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.
- [63] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training Region-based Object Detectors with Online Hard Example Mining[J/OL]. Comput. Vis. Pattern Recognit., 2016.
- [64] 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.
- [65] HOIEM D, CHODPATHUMWAN Y, DAI Q. Diagnosing error in object detectors [J]. Lect. Notes Comput. Sci., 2012, 7574 LNCS(PART 3): 340–353.
- [66] HORNIK K, STINCHCOMBE M, WHITE H. Multilayer feedforward networks are universal approximators[J]. Neural Networks, 1989, 2(5): 359–366.
- [67] 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.

- [68] 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.
- [69] 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.
- [70] VINCENTP, LAROCHELLEH, BENGIOY, et al. Extracting and composing robust features with denoising autoencoders.[J]. In *Proc. Int. Conf. Mach. Learn.*, pages 1096–1103, 2008.
- [71] 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.
- [72] 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.
- [73] LI H M. Deep Learning for Image Denoising[J]. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 2014, 7(3): 171–180.
- [74] 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.
- [75] 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.
- [76] 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.
- [77] 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.
- [78] ZHAO H, GALLO O, FROSIO I, et al. Is L2 a Good Loss Function for Neural Networks for Image Processing?[J]. arXiv preprint, 2015.
- [79] 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.
- [80] 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.
- [81] 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.
- [82] 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.
- [83] DOSOVITSKIY A, BROX T. Generating Images with Perceptual Similarity Metrics based on Deep Networks[Z]. [S.l.: s.n.], 2016.
- [84] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution[J]. arXiv Preprint, 2016.
- [85] NOHH, HONGS, AND B H. Learning deconvolution network for semantic segmentation.[J]. In *Proc. IEEE Int. Conf. Comp. Vis.*, pages 1520–1528, 2015.
- [86] HONGS, NOHH, AND B H. Decoupled deep neural network for semi-supervised semantic segmentation.[J]. In *Proc. Advances in Neural Inf. Process. Syst.*, 2015.
- [87] LONGJ, SHELHAMERE, AND T D. Fully convolutional networks for semantic segmentation.[J]. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, pages 3431–3440, 2015.
- [88] 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.
- [89] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. Computer Science, 2015.
- [90] HEK, ZHANGX, RENS, et al. Deep residual learning for image recognition.[J]. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, volume abs/1512.03385, 2016.
- [91] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. CoRR, 2014, abs/1409.1556.
- [92] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database[C]//[S.l.: s.n.], 2009: 248–255.
- [93] KINGMA D, BA J. Adam: A method for stochastic optimization[J]. Eprint Arxiv, 2014.
- [94] 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.
- [95] 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.
- [96] 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.

- [97] 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.
- [98] 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.
- [99] 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.
- [100] 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.
- [101] 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.
- [102] CRISTINACCE D, COOTES T. Automatic feature localisation with constrained local models[J]. *Pattern Recognition*, 2008, 41(10): 3054–3067.
- [103] 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.
- [104] 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.
- [105] 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.
- [106] 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.
- [107] 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.
- [108] 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.
- [109] 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.

- 
- [110] 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.
  - [111] 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.
  - [112] 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.
  - [113] KINGMA D P, BA J. Adam: A Method for Stochastic Optimization[J/OL]. CoRR abs/1412.6980, 2014: 1–15.
  - [114] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91–110.
  - [115] 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.



## 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,ICCCBDA(EI)* 2016,299-303.DOI:10.1109/ICCCBDA.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 OpenBLAS 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 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！