

《Light-weight Spatio-Temporal Graphs for Segmentation and Ejection Fraction Prediction in Cardiac Ultrasound》^[1]复现和优化

邓小龙

摘要

超声心动图参数的准确和一致预测对心血管的诊断和治疗非常重要。左心室分割可以用来得出心室容积、射血分数（EF）和其他相关的测量值。在论文中，作者提出了一种新的自动方法，称为 EchoGraphs，用于预测射血分数，并通过检测解剖学关键点对左心室进行分割。其基于图卷积网络（GCNs）的直接坐标回归模型被用来检测关键点。GCNs 可以根据每个关键点的局部外观，以及所有关键点组合的全局空间和时间结构来学习表示心脏形状。作者在 EchoNet 基准数据集上评估了 EchoGraphs 模型。与语义分割相比，GCNs 显示出准确的分割，并在鲁棒性和推理运行时间方面有所改进。EF 是与分割同时计算的，EchoGraphs 也获得了最先进的射血分数估计。

关键词：超声心动图；左心室分割；射血分数估计

1 引言

心力衰竭仍然是全世界住院和死亡的主要原因，迫切需要个性化和前瞻性的护理解决方案^[2]。超声心动图（Echocardiography），也称心脏超声成像（Cardiac Ultrasound Imaging），是评估心血管功能不可或缺的临床工具。由于其便携性、低成本和实时性，它通常被临床用于作为第一次成像检查^[3]。虽然许多传统方法，例如 Active Contours、Level-sets 和 Active Shape 模型等^[4]。已经被用于自动分割超声图像中的解剖结构，但所实现的精度受到超声成像的各种问题的限制，例如低信号噪声比、变化的斑点噪声、低图像对比度、边缘脱落和由诸如致密肌肉和肋骨等结构投射的阴影。

左心室射血分数是一项评价心脏功能的关键指标。它由心跳的舒张末期（ED）和收缩末期（ES）的容积计算得到。心跳的舒张末期（ED）和收缩末期（ES），分别对应于左心室的最大容积和最小容积的时期。在论文中，作者提出了 EchoGraphs 框架，以预测射血分数，并通过检测解剖学关键点对左心室进行分割。

2 相关工作

2.1 关键点回归

以前的一些工作已经显示了使用深度学习进行左心室分割^[5-10]和直接 EF 估计^[9,11-12]的高度精确自动化结果。对于分割，以前的工作大多依赖于语义分割，它输出输入图像的像素分类。在某些情况下，额外的模块被包含在隐含的形状约束中^[8,13]，但一般来说，这些方法没有明确地优化感兴趣的形状。这对于像左心室分割这样的问题来说是次优的，因为病人之间的形状是一致的，而确实发生的形状变化对于许多诊断来说是至关重要的^[14-15]。此外，注释通常以关键点的形式提供，使关键点回归成为更自然的学习形式。

2.2 图卷积网络

医学成像中最先进的语义分割技术是 nnU-Net^[16]，一个以 U-Net 为骨干的学习管道。nnU-Net 的成功源于其基于数据集特征的许多网络超参数的增强和自动优化。图卷积网络（GCNs）作为一种整合多模态的学习方法已经越来越受欢迎。作为一种整合可能不是网格结构的多模态数据的学习方法，已经越来越受欢迎^[17]。然而，GCN 也非常适合于根据图像或视频特征预测跨时空的关键点位置。对于这些问题，图的节点代表关键点，图的边缘代表点之间的学习关系。GCNs 在诸如视频中的姿势预测^[18]和一些医学成像问题中的分割等问题上取得了最先进的成果^[19-20]。在这项工作中，我们将图卷积方法用于超声心动图的分割。

3 本文方法

3.1 本文方法概述

EchoGraghs 框架如图 1 所示。本文的方法分为两部分：单帧 GCN 和多帧 GCN。单帧 GCN 中，CNN 编码器的输出被转发给具有四个螺旋卷积层的 GCN。为每个关键点回归一个坐标向量。多帧 GCN 中，由 F 帧组成的图像序列被送入一个 CNN 视频编码器，输出一个特征表示。这个特征表示被传递给一个预测 EF 值的回归器模块，以及输出两个关键帧的关键点坐标的时空 GCN。当这些帧的位置未知时，可以选择添加 ED/ES 分类器。

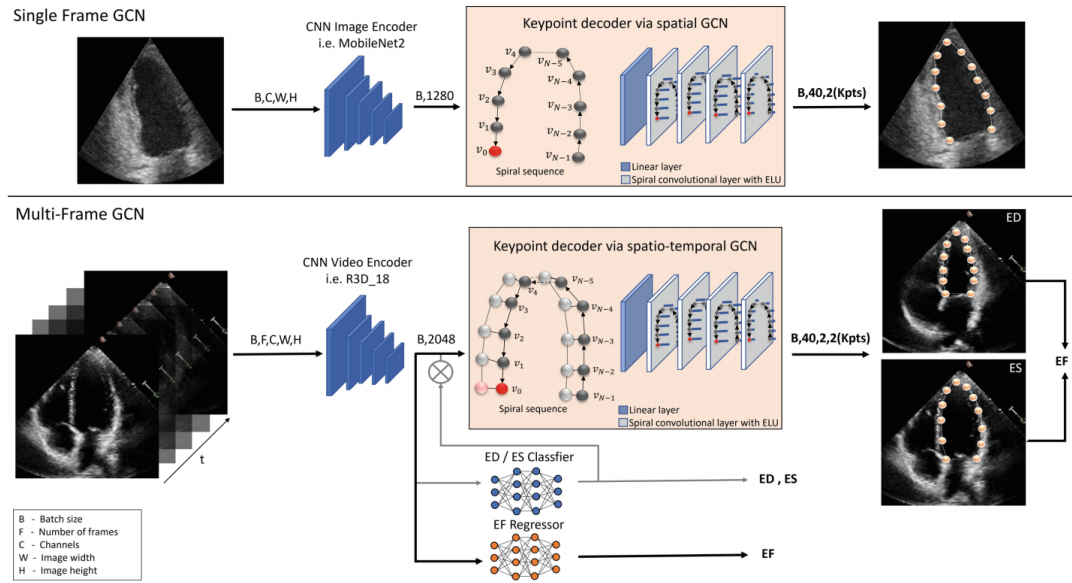


图 1: EchoGraghs 整体框架图

3.2 单帧 GCN

左心室是一个解剖形状，可以用一个封闭的轮廓来描述，并通过从轮廓上取样的有限数量的点来近似。给定一个超声帧，我们可以把左心室的轮廓解释为一个无向图 $G = (V, E)$ 。节点 $V = \{v_i | i = 1, \dots, N\}$ 代表轮廓点， E 代表各点之间的连接。一个轮廓点的空间变化最有可能影响到邻近的点，但也可能影响到不直接靠近的其他点。图卷积神经网络（GCNs）可以被看作是 CNNs 的概括，它不需要固定的网格，但可以通过使用边聚集邻居的信息并应用加权项来操作任何非欧氏结构的数据^[21]。虽然可以使用几种图形网络模型，但所提出的方法依赖于螺旋卷积。^[22] 这个算子在信息传递过程中强制执行相邻节点的固定排序来计算节点更新， $x_i^k = \gamma^k(\|_{j \in S(i,l)} x_j^{k-1})$ ，其中 $S(i,l)$ 是相邻节点 x_j 的固定螺旋式串联， γ 是一个多层感知器。选择螺旋连接是因为它们可以明确地模拟相邻节点的感应偏差。选

择螺旋连接是因为它们可以明确地模拟相邻节点的感应偏差，并且在计算上很有效率。尽管该方法的灵感来自于封闭的网状结构，并且螺旋线被定义在网状顶点上。尽管该方法的灵感来自于封闭的网格结构，并且螺旋线是在网格顶点上定义的，但它也可以应用于圆形结构，如左心室，它也可以应用于左心室这样的圆形结构，通过定义起始指数，然后以时钟的方式生成序列。循序渐进的方式生成序列。提出的图解码器结构包括一个初始密集层，进一步压缩输入特征向量，然后是四个螺旋卷积层，并由一个指数线性激活单元补充激活单元。在最后一层，为每个节点预测两个值，这两个值代表图像中的坐标像素位置。

单帧 GCN 允许使用 MobileNetV2、ResNet_18 或 ResNet_50 等常见的 CNN 编码器网络。输入的单帧图像通过编码器网络，提取其关键特征，送入关键点解码器（Keypoint Decoder）。关键点解码器由 1 个线性层和 4 个螺旋卷积层组成，在最后一层中，为每个关键点预测出两个值，表示图像中的关键点的坐标位置。

3.3 多帧 GCN

多帧 GCN 与单帧 GCN 结构类似。由于输入是整个视频序列，编码器使用了可以编码时间维度的 R3D_18 网络。射血分数定义为 $EF = (EDV - ESV) / EDV$ 。EDV 为左心室舒张末期容积，ESV 为收缩末期容积。虽然可以使用 Simpson 双面方法从提取的关键点计算心室容积，从而估计 EF，但根据早期的工作，发现直接从特征向量估计可以获得更准确的 EF 预测。因此，在多帧 GCN 中添加了一个 EF 回归器，使用了 4 层 MLP 网络（每个线性层后面都有一个指数线性激活单元），它可以从编码器中获取特征向量并输出一个单一的 EF 预测值。同时，在关键点中产生一个 EF 预测，可以获得置信度检查和可解释的结果。

除此之外，当输入一个未标记的视频时，网络需要知道在 ED、ES 帧在哪来应用分割。为了预测 ED、ES 帧，作者引入了一个 ED/ES 分类器，一个基于 4 个线性层的 MLP 网络。其中每一层都经过归一化和 ReLU 激活。该网络输出两个数组，代表每个帧是 ES 或 ED 的可能性。然后使用加权交叉熵损失，将每个数组与 Ground Truth 索引位置相匹配。

4 复现细节

4.1 与已有开源代码对比

本次实验的 EchoGraphs 框架代码是开源的。但在 GCN 网络中，缺乏低级特征和高级特征的权衡。为了解决这个不平衡的问题，我们引入了 PRM 模块^[23]。PRM 模块结构如图 2。我们在 Encoder 后添加了 PRM 模块。PRM 模块从 Encoder 尾部获取特征作为输入，通过 PRM 模块后尺寸不变。在 PRM 模块中，顶层路径有利于保留局部特征，中间路径的设计是为了在通道上对特征进行重新加权，而底部路径则是为空间注意力而提出的。低级别的精确空间信息和高层次的判别语义信息。空间信息有利于关键点的定位，而语义信息有利于关键点的分类。而这些特征对最终预测的贡献是不同的。

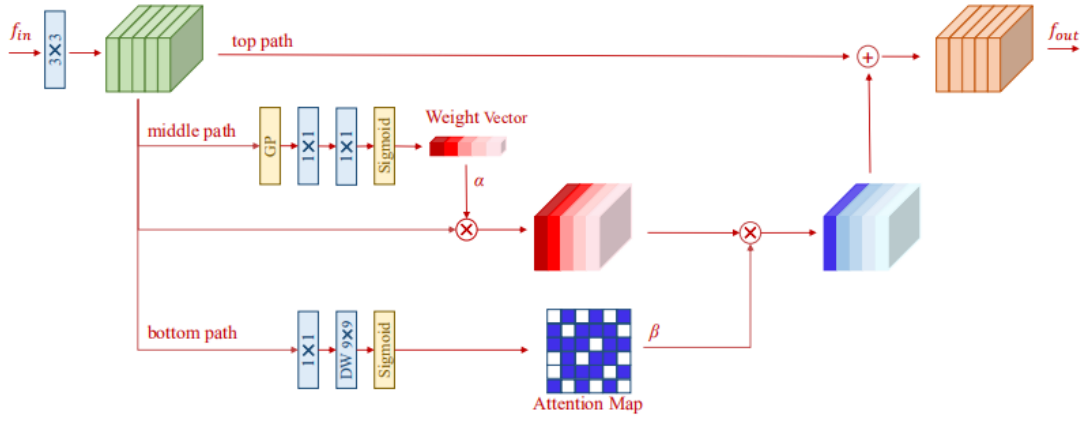


图 2: PRM 模块。GP 表示全局池化层，DW 表示深度可分离卷积。 α 表示权重向量。 β 表示注意力图。

4.2 实验环境搭建

本次实验环境在 Linux 系统中，实验所需的软件包和库为：Python3.8、PyTorch1.9、CUDA10.2、OpenCV、torchvision、scipy、tensorboard 等。

数据集来源于 Echonet-Dynamic 数据集^[9]，是一个大型超声心动图视频数据集，包括 10030 个标记的超声心动图视频和医学专家注释。

单帧 GCN 的训练、预测是使用一张 NVIDIA GeForce RTX 2080Ti 上。

5 实验结果分析

5.1 左心室分割

评价左心室分割性能，使用了 Dice 分数、平均关键点误差（Mean Keypoint Error, MKE）作为评价指标，Runtime 是以毫秒为单位，每一帧的运行时间是指在没有预处理或增强的情况下，对模型进行一次正向传递。EchoGraphs 的分割性能如表 1 所示。

Model	Backbone	Dice (%)	MKE (%)	Runtime [cpu/gpu]	Parameters
EchoNet	DeepLabV3	91.7 ± 4.2	2.5 ± 1.2	33.65/4.94	39.6 M
nnU-Net	U-Net	92.8 ± 3.6	2.3 ± 1.2	14.86/1.05	7.3 M
EchoGraphs (ours)	MobileNetv2	91.6 ± 4.0	2.3 ± 1.0	2.45/0.68	4.92 M
EchoGraphs (ours)	ResNet18	91.8 ± 4.0	2.3 ± 1.0	2.68/0.46	12.1 M
EchoGraphs (ours)	ResNet50	92.1 ± 3.8	2.2 ± 0.9	6.73/1.05	27.1 M

表 1: 左心室分割结果

对单帧 GCN 进行复现，经过实验调参并优化得到的复现结果如表 2 所示。在加入 PRM 模块，分割性能获得了一点提升。可视化结果图如 3 所示。

Model	Backbone	MKE (%)
EchoGraphs	MobileNetV2	2.13
EchoGraphs	ResNet18	2.02
EchoGraphs	ResNet50	2.00
EchoGraphs with PRM	ResNet50	1.94

表 2: 复现结果

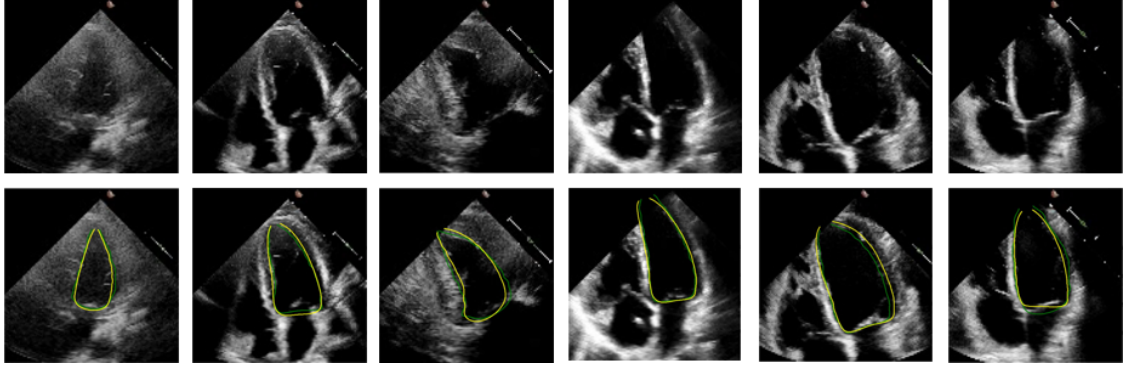


图 3: 分割可视化结果图。上: 为原图像, 下: GT 和分割结果, 黄线为 GT, 绿线为分割结果。

5.2 射血分数预测

评价射血分数估计, 使用了 MAE (Mean Absolute Error)、RMSE (Root Mean Squared Error) 和 R^2 (Correlation) 三个评价指标。EchoGraphs 的预测性能如 3 所示。

Method	Input	Frames	MAE	RMSE	R^2
EchoNet (MC3)	single heartbeat	32	4.22	5.56	0.79
Transformer (M.)	single heartbeat	128	5.32	7.23	0.64
Regression only (ours)	single heartbeat	16	4.28	5.75	0.72
EchoGraphs (ours) using keypoints	single heartbeat	16	4.66	6.30	0.73
EchoGraphs (ours) with regression head	single heartbeat	16	4.01	5.36	0.81
EchoNet	whole video	-	4.05	5.32	0.81
Transformer	whole video	-	5.95	8.38	0.52
EchoGraphs (ours) - Peak computation	whole video	-	4.30	5.86	0.65
EchoGraphs (ours) - ED/ES classifier	whole video	-	4.23	5.67	0.79

表 3: 射血分数预测结果

6 总结与展望

这次实验主要的工作是复现 EchoGraphs 网络框架, 以及加入 PRM 模块的一些优化。论文中的结果分割精度低于 nnU-Net, 但 nnU-Net 包含测试时间的增强和大量的超参数优化。这些同样的优化可以应用于 EchoGraphs, 并可能进一步提高性能。在本次实验中发现, EchoGraphs 框架的单帧 GCN 对单帧图像的分割效果较好, 但多帧 GCN 对视频与其他 SOTA 方法相比还是稍显逊色。以及使用 ES/ED 分类器来识别 ES、ED 帧, 效果较差。未来可以从为关键点形状的约束设计损失函数、识别关键帧等方面考虑进行优化。

参考文献

- [1] THOMAS S, GILBERT A, BEN-YOSEF G. Light-weight spatio-temporal graphs for segmentation and ejection fraction prediction in cardiac ultrasound[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part IV. 2022: 380-390.
- [2] SAVARESE G, STOLFO D, SINAGRA G, et al. Heart failure with mid-range or mildly reduced ejection fraction[J]. Nature Reviews Cardiology, 2022, 19(2): 100-116.
- [3] AKKUS Z, ALY Y H, ATTIA I Z, et al. Artificial intelligence (AI)-empowered echocardiography interpretation: a state-of-the-art review[J]. Journal of clinical medicine, 2021, 10(7): 1391.
- [4] CHEN C, QIN C, QIU H, et al. Deep learning for cardiac image segmentation: a review[J]. Frontiers in Cardiovascular Medicine, 2020, 7: 25.
- [5] GILBERT A, MARCINIAK M, RODERO C, et al. Generating synthetic labeled data from existing anatomical models: an example with echocardiography segmentation[J]. IEEE Transactions on Medical Imaging, 2021, 40(10): 2783-2794.
- [6] JAFARI M H, GIRGIS H, LIAO Z, et al. A unified framework integrating recurrent fully-convolutional networks and optical flow for segmentation of the left ventricle in echocardiography data[C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. 2018: 29-37.
- [7] LECLERC S, SMISTAD E, ØSTVIK A, et al. Deep learning segmentation in 2D echocardiography using the camus dataset: Automatic assessment of the anatomical shape validity[J]. arXiv preprint arXiv:1908.02994, 2019.
- [8] OKTAY O, FERRANTE E, KAMNITSAS K, et al. Anatomically constrained neural networks (AC-NNs): application to cardiac image enhancement and segmentation[J]. IEEE transactions on medical imaging, 2017, 37(2): 384-395.
- [9] OUYANG D, HE B, GHORBANI A, et al. Video-based AI for beat-to-beat assessment of cardiac function[J]. Nature, 2020, 580(7802): 252-256.
- [10] SMISTAD E, ØSTVIK A, SALTE I M, et al. Real-time automatic ejection fraction and foreshortening detection using deep learning[J]. IEEE transactions on ultrasonics, ferroelectrics, and frequency control, 2020, 67(12): 2595-2604.
- [11] KAZEMI ESFEH M M, LUONG C, BEHNAMI D, et al. A deep Bayesian video analysis framework: towards a more robust estimation of ejection fraction[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23. 2020: 582-590.

- [12] REYNAUD H, VLONTZOS A, HOU B, et al. Ultrasound video transformers for cardiac ejection fraction estimation[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24. 2021: 495-505.
- [13] PAYER C, ŠTERN D, BISCHOF H, et al. Integrating spatial configuration into heatmap regression based CNNs for landmark localization[J]. Medical image analysis, 2019, 54: 207-219.
- [14] BALTABAEVA A, MARCINIAK M, BIJNENS B, et al. Regional left ventricular deformation and geometry analysis provides insights in myocardial remodelling in mild to moderate hypertension[J]. European Journal of Echocardiography, 2008, 9(4): 501-508.
- [15] MARCINIAK M, GILBERT A, LONCARIC F, et al. Septal curvature as a robust and reproducible marker for basal septal hypertrophy[J]. Journal of hypertension, 2021, 39(7): 1421.
- [16] ISENSEE F, JAEGER P F, KOHL S A, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation[J]. Nature methods, 2021, 18(2): 203-211.
- [17] AHMEDT-ARISTIZABAL D, ARMIN M A, DENMAN S, et al. Graph-based deep learning for medical diagnosis and analysis: past, present and future[J]. Sensors, 2021, 21(14): 4758.
- [18] WANG J, YAN S, XIONG Y, et al. Motion guided 3d pose estimation from videos[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16. 2020: 764-780.
- [19] GOPINATH K, DESROSIERS C, LOMBAERT H. Graph domain adaptation for alignment-invariant brain surface segmentation[C]//Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Graphs in Biomedical Image Analysis: Second International Workshop, UNSURE 2020, and Third International Workshop, GRAIL 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 2. 2020: 152-163.
- [20] TIAN Z, LI X, ZHENG Y, et al. Graph-convolutional-network-based interactive prostate segmentation in MR images[J]. Medical physics, 2020, 47(9): 4164-4176.
- [21] GONG S, CHEN L, BRONSTEIN M, et al. Spiralnet++: A fast and highly efficient mesh convolution operator[C]//Proceedings of the IEEE/CVF international conference on computer vision workshops. 2019: 0–0.
- [22] BOURITSAS G, BOKHNYAK S, PLOUMPIS S, et al. Neural 3d morphable models: Spiral convolutional networks for 3d shape representation learning and generation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 7213-7222.
- [23] CAI Y, WANG Z, LUO Z, et al. Learning delicate local representations for multi-person pose estimation [C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16. 2020: 455-472.