Cardiac Computed Tomography Angiography Plane Prediction and Comprehensive LV Segmentation

Davis Marc Vigneault[1] DVIGNE01@STANFORD.EDU
[1] *Department of Radiology, Stanford University, Stanford, CA, USA*

Ashish Manohar [1,2,3] ASHMAN@STANFORD.EDU
[2] *Division of Cardiovascular Medicine, Department of Medicine, Stanford University, Stanford, CA, USA*
[3] *Cardiovascular Institute, Stanford University, Stanford, CA, USA*

Abraham Hernandez [2] XBRAHAM@STANFORD.EDU
Krista Tin Chi Wong [2] KRISTAW@STANFORD.EDU
Fanwei Kong [4] KONGF@WUSTL.EDU
[4] *Department of Mechanical Engineering and Materials Science, Washington University in St. Louis, St. Louis, MO, USA*

Tea Gegenava [2] GEGENAVAT@YAHOO.COM
Koen Nieman [1,2,3] KNIEMAN@STANFORD.EDU
Dominik Fleischmann [1,3] D.FLEISCHMANN@STANFORD.EDU

Editors: Under Review for MIDL 2025

Abstract

The use of cardiac computed tomography angiography (CCTA) has dramatically increased over the past decade, with an increasingly recognized role for functional assessment; however, reformatting these datasets into standard cardiac planes and performing quantitative analysis remains time consuming and disruptive to clinical workflows. Here, we propose a fully automated, volumetric, end-to-end trained network for simultaneous detection of standard cardiac planes and comprehensive left ventricular (LV) segmentation in the predicted short axis coordinate system. The architecture consists of a coarse segmentation module, a transformation module, and a fine segmentation module. The coarse segmentation module provides an initial segmentation of the full field of view (FOV) axial images at low resolution. The transformation module predicts the rotations corresponding to the standard cardiac planes (short axis, SAX; two chamber, 2CH; three chamber, 3CH; and four chamber, 4CH) and reformats the source volume into the predicted SAX coordinate system at high resolution. Finally, the fine segmentation module segments the narrow FOV, high resolution SAX volume. The dataset consisted of 313 CCTA studies partitioned into training, validation, and testing in an 80:10:10 split. Architectural decisions are justified using ablation experiments. On the test set, the proposed architecture achieved high quality segmentations (Dice scores of 0.955, 0.928, and 0.808 for the bloodpool, myocardium, and trabeculations, respectively) and accurate plane predictions (mean angle errors of 9.126°, 9.480°, 9.030°, and 8.840° for the SAX, 2CH, 3CH, and 4CH planes, respectively). This fully automated pipeline has the potential to replace current manual workflows, expediting the availability of standard cardiac planes and quantitative analysis for interpretation.

Keywords: Cardiac computed tomography angiography (CCTA), segmentation, spatial transformer network (STN)

1. Introduction

The use of cardiac computed tomography angiography (CCTA) in the United States increased 85% over the previous decade (Reeves et al., 2021), with outpatient, inpatient, and emergency department exams all more than doubling in frequency. This trend is likely to continue or accelerate owing to the increasing availability of scanners capable of performing high quality cardiac exams, incorporation of coronary CT angiography as a Class I recommendation in the AHA/ACC clinical practice guidelines on the evaluation of chest pain (Gulati et al., 2021), and doubling of reimbursement by Medicare in the United States starting in 2025 (Maxwell, 2024). Moreover, there is an increasingly recognized role of retrospectively ECG-gated cine acquisitions for functional assessment (Peper et al., 2020), with incremental value over coronary CTA alone (Seneviratne et al., 2010). Reformatting these images into standard cardiac planes is critical for standardized comparison between exams, wall thickness measurements, and myocardial segment classification; however, this processing is time consuming and usually requires a third party software package outside the standard clinical PACS system.

The literature on medical image segmentation is extensive, with deep neural networks yielding excellent performance over the past decade. Ronneberger et al. (2015) first introduced the U-Net, a highly successful 2D encoder-decoder architecture with skip connections. Since then, a plethora of modifications to the U-Net have been proposed. Residual, recurrent, and residual-recurrent versions have been described (He et al., 2016; Milletari et al., 2016; Alom et al., 2019). Oktay et al. (2018) added attention gates, using saliency maps to preserve only relevant activations. Additional connections within (Huang et al., 2017; Jegou et al., 2017) or between (Zhou et al., 2020) the network layers have been added to enhance information flow. Multiple U-Nets have been combined into “cascaded” networks, which in their simplest form provide the predictions of one U-Net module as an input to a second U-Net module (Liu et al., 2021) and in more sophisticated implementations densely connect the network layers of successive U-Net modules (Wu et al., 2023). Most recently, more sophisticated approaches using transformers (Chen et al., 2021a; Cao et al., 2021) and graph neural networks (Kong et al., 2021) have also been proposed. Many of these concepts have been applied to CCTA segmentation (Bruns et al., 2020; Li et al., 2021; Jun Guo et al., 2020; Wang et al., 2022; Kong et al., 2021).

Combined segmentation and detection of standard cardiac planes from CCTA has received much less attention. The most closely related work (Chen et al., 2021b) describes a method for prediction of SAX, 2CH, 3CH, and 4CH planes by branching a single fully connected layer from the bottleneck of a volumetric U-Net. This approach has several shortcomings in that (a) separate models are required for each plane, (b) only a single fully connected layer is explored in predicting the plane parameters, (c) the images are not reoriented into the cardiac coordinate system to improve upon the segmentations, (d) end-to-end training is not employed, and (e) the sample size is relatively small (100 patients).

Therefore, the purpose of this study was to develop a fully automated, volumetric, end-to-end trained network for simultaneous detection of the standard cardiac planes (SAX, 2CH, 3CH, and 4CH) and comprehensive left ventricular (LV) segmentation (bloodpool, myocardium, and trabeculations) in the predicted SAX coordinate system.

2. Methods

2.1. Dataset

The dataset consisted of 313 CCTA studies randomly partitioned into training, validation, and testing in an approximately 80:10:10 split (250 training, 30 validation, and 33 testing). Cases were obtained as part of routine clinical practice and were retrospectively collected with IRB approval. Final clinical diagnoses were normal ($N = 89$), hypertrophic cardiomyopathy ($N = 106$), LV non-compaction ($N = 46$), and dilated cardiomyopathy ($N = 72$). Acquisitions were retrospectively ECG-gated and reconstructed at mid-diastole. Studies were obtained from one of four scanners: SOMATOM Force (Siemens Healthineers; $N = 275$), SOMATOM Definition Flash (Siemens Healthineers; $N = 36$), Lightspeed VCT (General Electric Healthcare; N=1), or Sensation 64 (Siemens Healthineers; $N = 1$). Slice thicknesses were 0.75 mm for Siemens and 0.625 mm for GE scans. The median reconstructed field-of-view diameter was 190.0 mm (interquartile range: 173.0–209.0 mm), with a median in-plane pixel spacing of 0.37 mm (interquartile range: 0.34–0.41 mm).

2.2. Proposed Architecture

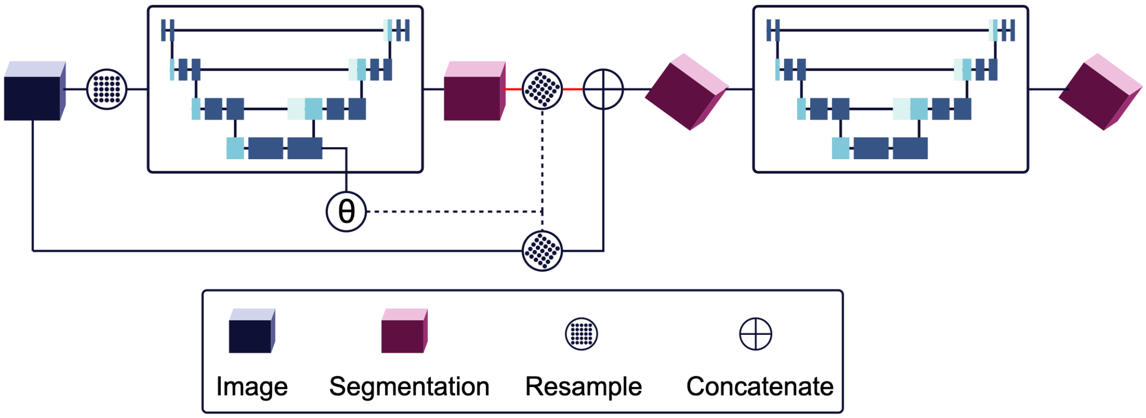


Figure 1: Network Architecture.

The proposed network architecture (Figure 1) consists of three end-to-end-trained modules: a coarse segmentation module, a transformation module, and a fine segmentation module.

2.2.1. Coarse Segmentation Module

The coarse segmentation module takes as input the axial CCTA volume downsampled to 3.0 mm isotropic with a $64 \times 64 \times 64$ matrix size and produces as output an equivalently sized multi-class segmentation (bloodpool, myocardium, and trabeculations). The architecture used is a volumetric attention residual U-Net with 4 downsampling/upsampling steps. The number of features produced by each convolution block is 32 in the highest resolution stage and is doubled at each downsampling step and halved at each upsampling step. The fundamental processing block is made up of a $3 \times 3 \times 3$ convolution, group normalization layer (Wu and He, 2018), and leaky rectified linear unit (ReLU) activation, applied twice at each stage. This is optionally converted to a residual block by element-wise addition of the input and the result of the last normalization layer, prior to applying the final activation; in practice, convolution and normalization layers are applied within the residual connection to match the feature lengths prior to summing. Traditional skip connections concatenate feature vectors from matching resolutions in the downsampling and upsampling paths. These can be converted to attention gates by first multiplying the downsampling path input by a multi-class saliency map learned from the downsampling and upsampling path inputs (Oktay et al., 2018). The result is passed into a final convolution to produce the raw logits corresponding to the background and foreground classes.

2.2.2. Transformation Module

The transformation module is responsible for learning the rotations corresponding to the standard cardiac planes, calculating the LV centroid, and resampling the input image and the coarse segmentation logits in the learned SAX coordinate system (to be passed as input to the fine segmentation module). The predicted rotations are represented as quaternions, a compact representation widely used in the graphics community due to several favorable mathematical properties. A matrix of quaternions is predicted as the output of one or more fully connected layers branched from the bottleneck of the coarse segmentation module. Note, however, that the rotations describing the standard cardiac planes (Q_{SAX}, Q_{2CH}, etc.) are the composite of (a) a shared baseline rotation (Q_{BLN}) orienting the long axis of the LV perpendicular to the plane of the image and (b) an additional rotation specific to each plane ($Q_{\Delta SAX}$, $Q_{\Delta 2CH}$, etc.). Therefore, rather than predicting the final rotations directly, the network is trained to predict the matrix of baseline rotation and the additional rotational offsets $[Q_{BLN}, Q_{\Delta SAX}, Q_{\Delta 2CH}, Q_{\Delta 3CH}, Q_{\Delta 4CH}]$. The LV centroid is estimated directly from the coarse segmentation prediction probabilities. Using the SAX rotation quaternion and LV centroid, the axial input image and coarse segmentation logits are resampled in the SAX coordinate system at 0.8 mm isotropic with a $128 \times 128 \times 192$ matrix size.

2.2.3. Fine Segmentation Module

Like the coarse segmentation module, the fine segmentation module is a volumetric attention residual U-Net, starting with 40 features in the highest resolution stage, but otherwise identical to the former.

2.3. Network Implementation and Training

2.3.1. Preprocessing and Augmentation

The training dataset was augmented at runtime by applying random rotations (100% probability, ±45° along each axis) and adding random Gaussian noise (50% probability, $\sigma \in [0, 100]$ HU). Note that variation in the predicted centroid and SAX quaternion consequently varies the volume provided to the fine segmentation module, resulting in additional implicit augmentation. Following augmentation, input images were clipped to the range $[-200, 600]$ Hounsfield units ("vascular windows") and normalized to the range $[0, 1]$.

2.3.2. Network Training

The coarse and fine segmentation modules are supervised using mean Jaccard loss across all classes (L_c and L_f, respectively). For the transformation module, we provide both *direct* supervision of the predicted quaternions $[Q_{BLN}, Q_{\Delta SAX}, Q_{\Delta 2CH}, Q_{\Delta 3CH}, Q_{\Delta 4CH}]$ and *indirect* supervision of the composite quaternions $[Q_{SAX}, Q_{2CH}, Q_{3CH}, Q_{4CH}]$. The loss L_q is the sum of the mean squared errors between the ground truth and predicted quaternions for both direct and indirect rotations, which is mathematically closely related to the angle between the rotations they represent. The total network loss L_t is then given as a weighted sum of these losses:

$$L_t = \alpha_c L_c + \alpha_q L_q + \alpha_f L_f \tag{1}$$

We set $\alpha_c = \alpha_f = 1$ and $\alpha_q = 10/n_q$ where n_q is the total number of quaternions being supervised.

The network was trained with a batch size of 1 for 24 epochs using the Adam optimizer. Learning rate warmup was used with an initial rate of 10^{-6} and a target rate of 10^{-4}, achieved using a linear ramp over 3 epochs. After the third epoch, the learning rate was exponentially decayed with a multiplicative factor of 0.9.

2.3.3. Implementation Details

The network was implemented in python (version 3.12.6) using Monai (version 1.5.dev2446) and PyTorch (version 2.5.1+cu124). Conversion between quaternion and matrix rotation representations was performed using RoMa (version 1.5.0). Experiments were run on an Ubuntu workstation (version 24.04) with a 16 core Intel i7-13700K processor, 64 GB RAM, and a single NVIDIA GeForce RTX 4090 GPU with 24 GB memory. Please see the repository for additional details.[1]

3. Results

Results of the hyperparameter search and ablation experiments are presented in Table 1. Regarding the transformation module, the depth and width of the hidden layers branched from the coarse segmentation module bottleneck were varied. The version with two 128-feature hidden layers performed best in terms of centroid error (0.805 mm) as well as angle error for three of the four standard cardiac planes (9.126° SAX, 9.030° 3CH, and 8.840° 4CH). For the remaining metrics, performance was near to the best performing metric. For the 2CH plane, the angle error was 9.480° compared to 9.096°. For the fine segmentation module, Dice scores for the bloodpool, myocardium, and trabeculations were 0.955, 0.928, and 0.808 compared to 0.956, 0.930, and 0.814 for the best performing networks. For these reasons, the version with two 128-feature hidden layers was selected for subsequent experiments; representative segmentations and plane predictions are shown in Figure 2.

Ablation experiments were performed to explore the value of the attention gates, residual blocks, cascading, end-to-end training, and the hidden layers in the transformation module. Removing the attention gates degraded performance in terms of centroid error, the angle errors for all standard cardiac planes, and the bloodpool Dice; Dice was slightly higher for the

---

[1] `https://github.com/sudomakeinstall/2025-midl-ccta-plane-prediction`

myocardium (within rounding error, 0.928 for both) and trabeculations (0.813 versus 0.808). Removing the residual blocks degraded performance for all metrics. Removing cascading degraded performance in terms of angle errors for the SAX, 3CH, and 4CH planes; performance was slightly improved in terms of centroid error (0.787 mm vs 0.805 mm), angle error for the 2CH plane (9.420° versus 9.480°), bloodpool Dice (0.957 versus 0.955), myocardium Dice (0.931 versus 0.928), and trabeculation Dice (0.817 versus 0.808). Training the coarse segmentation, transformation, and fine segmentation modules separately degraded performance for all metrics. Removing all hidden layers from the transformation module resulted in slightly lower centroid error (0.801 mm versus 0.805 mm) but degraded performance for all other metrics.

4. Discussion and Conclusions

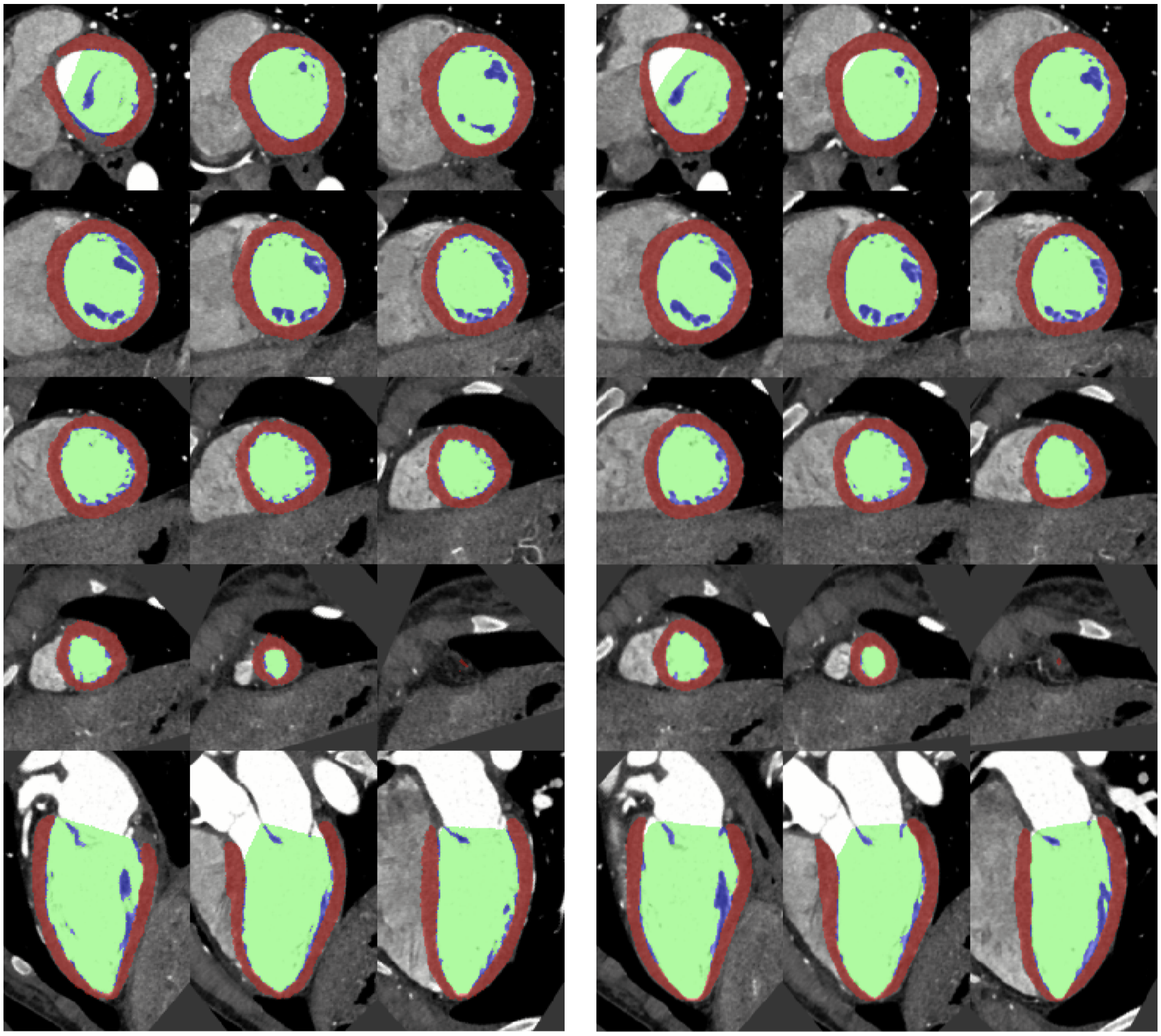
Here, we present a fully automated, volumetric, end-to-end trained network for simultaneous detection of standard cardiac planes (SAX, 2CH, 3CH, and 4CH) and comprehensive LV segmentation (bloodpool, myocardium, and trabeculations) in the predicted SAX coordinate system. The primary novelty of our network is in the transformation module, which learns the quaternions representing the standard cardiac planes, calculates the centroid from the coarse segmentation, and combines concepts from spatial transformer networks and cascading to resample the input image and coarse segmentation logits into the SAX coordinate system as input to the fine segmentation module. The network had high performance in terms of standard cardiac plane detection, with sub-millimeter centroid error and angle error $< 10°$ for all standard cardiac planes. The Dice scores achieved by our network are also high (0.955, 0.928, and 0.808 for the bloodpool, myocardium, and trabeculations, respectively), which is notable given the separate segmentation of the LV trabeculations, a high surface-area-to-volume structure. Note that segmentation of the LV trabeculations has value in the investigation of diagnoses such as LV non-compaction cardiomyopathy (Manohar et al., 2023) but is not typically included as a separate label in segmentation models.

This work has several limitations and areas for future improvement and validation. First, it would be useful to quantify intra- and inter-observer variability in standard cardiac plane angles in order to contextualize the angle errors observed in our network. Second, several potential improvements to the segmentation modules, particularly the use of transformer-based modules, have the potential to improve segmentation performance and should be investigated. Third, our intention in adding cascading (passing features from the coarse segmentation module to the fine segmentation module) was to improve segmentation performance, and yet ablation experiments instead demonstrated an improvement in the angle errors and slightly degraded segmentation performance; this paradoxical result is not fully explained by our experiments and is deserving of further investigation. Fourth, the dataset was obtained retrospectively from a single center; proposed network should undergo further validation in prospectively obtained, multicenter images.

This fully automated pipeline has the potential to replace current manual workflows, expediting the availability of standard cardiac planes and quantitative analysis for interpretation.

Table 1: Network performance, hyperparameter search, and ablation experiments. The reported evaluation metrics are the segmentation centroid, angle errors for the SAX, 2CH, 3CH, and 4CH planes, and Dice scores for the bloodpool (BP), myocardium (MC), and trabeculations (TB). In the first block of the table, the width and number of hidden layers in the transformation module are varied; for each evaluation metric, the best performing network is bolded. The network with two 128-feature hidden layers in the transformation module (highlighted in gray) is chosen as the proposed network and is used in subsequent experiments. In the second block of the table, ablation experiments are performed exploring the effect of the attention gates (Atn), residual blocks (Res), cascading (Csd), end-to-end training (E2E), and the transformation block hidden layers; metrics which improved relative to the proposed network are highlighted in green and those which worsened are highlighted in red.

Atn	**Res**	**Csd**	**E2E**	**Hdn**	**Centroid**	**SAX**	**2CH**	**3CH**	**4CH**	**BP**	**MC**	**TB**
Network Parameters					mm	Angle Error (°)				Dice		
Transformation Module Hyperparameter Search												
+	+	+	+	64	2.246	9.737	9.858	9.839	9.565	0.955	**0.930**	**0.814**
+	+	+	+	64-64	2.184	10.106	10.040	10.682	9.401	**0.956**	0.929	0.810
+	+	+	+	128	0.853	9.176	**9.096**	9.922	10.913	0.955	0.926	0.808
+	+	+	+	128-128	**0.805**	**9.126**	9.480	**9.030**	**8.840**	0.955	0.928	0.808
+	+	+	+	256	1.941	9.708	10.263	10.683	10.006	0.955	0.930	0.809
+	+	+	+	256-256	0.820	9.961	9.697	9.667	9.344	0.954	0.921	0.801
Ablation Experiments												
−	+	+	+	128-128	1.917	9.886	9.941	9.718	9.199	0.954	0.928	0.813
+	−	+	+	128-128	0.811	9.316	9.572	9.786	9.178	0.954	0.927	0.805
+	+	−	+	128-128	0.787	10.149	9.420	10.034	9.322	0.957	0.931	0.817
+	+	+	−	128-128	1.302	12.959	12.501	12.837	12.535	0.927	0.905	0.717
+	+	+	+	–	0.801	11.559	10.807	10.401	10.115	0.954	0.926	0.805



(a) Ground Truth (b) Prediction

Figure 2: Representative ground truth (left) and predicted (right) segmentations and standard cardiac planes for a patient in the testing partition. In each case, the SAX stack is shown in rows 1-4 (sliced from base to apex) and the long axis reformations are shown in row 5 (2CH, 3CH, and 4CH from left to right).

Acknowledgments

This study was supported by grants from the Radiological Society of North America (RR24-065; DV), the Etta K. Moskowitz Foundation (DV), the American Heart Association (AHA 24POST1187968; AM), and the National Institutes of Health (NHLBI R01 HL146754; KN).

References

Md Zahangir Alom, Chris Yakopcic, Mahmudul Hasan, Tarek M. Taha, and Vijayan K. Asari. Recurrent residual U-Net for medical image segmentation. *Journal of Medical Imaging*, 6(01):1, March 2019. ISSN 2329-4302. doi: 10.1117/1.JMI.6.1.014006. URL https://www.spiedigitallibrary.org/journals/journal-of-medical-imaging/volume-6/issue-01/014006/Recurrent-residual-U-Net-for-medical-image-segmentation/10.1117/1.JMI.6.1.014006.full.

Steffen Bruns, Jelmer M. Wolterink, Richard A. P. Takx, Robbert W. Van Hamersvelt, Dominika Suchá, Max A. Viergever, Tim Leiner, and Ivana Išgum. Deep learning from dual-energy information for whole-heart segmentation in dual-energy and single-energy non-contrast-enhanced cardiac CT. *Medical Physics*, 47(10):5048–5060, October 2020. ISSN 0094-2405, 2473-4209. doi: 10.1002/mp.14451. URL https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.14451.

Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation, May 2021. URL http://arxiv.org/abs/2105.05537. arXiv:2105.05537 [cs, eess].

Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L. Yuille, and Yuyin Zhou. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation, February 2021a. URL http://arxiv.org/abs/2102.04306. arXiv:2102.04306 [cs].

Zhennong Chen, Marzia Rigolli, Davis Marc Vigneault, Seth Kligerman, Lewis Hahn, Anna Narezkina, Amanda Craine, Katherine Lowe, and Francisco Contijoch. Automated cardiac volume assessment and cardiac long- and short-axis imaging plane prediction from electrocardiogram-gated computed tomography volumes enabled by deep learning. *European Heart Journal - Digital Health*, 2(2):311–322, 2021b. ISSN 2634-3916. doi: 10.1093/ehjdh/ztab033. URL https://academic.oup.com/ehjdh/article/2/2/311/6179804. author+an: 3=gras.

Martha Gulati, Phillip D. Levy, Debabrata Mukherjee, Ezra Amsterdam, Deepak L. Bhatt, Kim K. Birtcher, Ron Blankstein, Jack Boyd, Renee P. Bullock-Palmer, Theresa Conejo, Deborah B. Diercks, Federico Gentile, John P. Greenwood, Erik P. Hess, Steven M. Hollenberg, Wael A. Jaber, Hani Jneid, José A. Joglar, David A. Morrow, Robert E. O'Connor, Michael A. Ross, and Leslee J. Shaw. 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR Guideline for the Evaluation and Diagnosis of Chest Pain: Executive Summary: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation*, page CIR.0000000000001030, October 2021. ISSN 0009-7322, 1524-4539. doi: 10.1161/CIR.0000000000001030. URL https://www.ahajournals.org/doi/10.1161/CIR.0000000000001030.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, Las Vegas, NV, USA, June 2016. IEEE. ISBN 978-1-4673-8851-1. doi: 10.1109/CVPR.2016.90. URL http://ieeexplore.ieee.org/document/7780459/.

Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, Honolulu, HI, July 2017. IEEE. ISBN 978-1-5386-0457-1. doi: 10.1109/CVPR.2017.243. URL https://ieeexplore.ieee.org/document/8099726/.

Simon Jegou, Michal Drozdzal, David Vazquez, Adriana Romero, and Yoshua Bengio. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1175–1183, Honolulu, HI, USA, July 2017. IEEE. ISBN 978-1-5386-0733-6. doi: 10.1109/CVPRW.2017.156. URL http://ieeexplore.ieee.org/document/8014890/.

Bang Jun Guo, Xiuxiu He, Yang Lei, Joseph Harms, Tonghe Wang, Walter J. Curran, Tian Liu, Long Jiang Zhang, and Xiaofeng Yang. Automated left ventricular myocardium segmentation using 3D deeply supervised attention U-net for coronary computed tomography angiography; CT myocardium segmentation. *Medical Physics*, 47 (4):1775–1785, April 2020. ISSN 0094-2405, 2473-4209. doi: 10.1002/mp.14066. URL https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.14066.

Fanwei Kong, Nathan Wilson, and Shawn Shadden. A deep-learning approach for direct whole-heart mesh reconstruction. *Medical Image Analysis*, 74:102222, December 2021. ISSN 13618415. doi: 10.1016/j.media.2021.102222. URL https://linkinghub.elsevier.com/retrieve/pii/S136184152100267X.

Changling Li, Xiangfen Song, Hang Zhao, Li Feng, Tao Hu, Yuchen Zhang, Jun Jiang, Jianan Wang, Jianping Xiang, and Yong Sun. An 8-layer residual U-Net with deep supervision for segmentation of the left ventricle in cardiac CT angiography. *Computer Methods and Programs in Biomedicine*, 200:105876, March 2021. ISSN 01692607. doi: 10.1016/j.cmpb.2020.105876. URL https://linkinghub.elsevier.com/retrieve/pii/S0169260720317090.

Yu-Cheng Liu, Mohammad Shahid, Wannaporn Sarapugdi, Yong-Xiang Lin, Jyh-Cheng Chen, and Kai-Lung Hua. Cascaded atrous dual attention U-Net for tumor segmentation. *Multimedia Tools and Applications*, 80(20):30007–30031, August 2021. ISSN 1380-7501, 1573-7721. doi: 10.1007/s11042-020-10078-2. URL https://link.springer.com/10.1007/s11042-020-10078-2.

Ashish Manohar, Davis M. Vigneault, Deborah H. Kwon, Kadir Caliskan, Ricardo P. J. Budde, Alexander Hirsch, Seung-Pyo Lee, Whal Lee, Anjali Owens, Harold Litt, Francois Haddad, Gabriel Mistelbauer, Matthew Wheeler, Daniel Rubin, W. H. Wilson Tang, and Koen Nieman. Quantitative metrics of the LV trabeculated layer by cardiac CT and cardiac MRI in patients with suspected noncompaction cardiomyopathy. *European Radiology*, December 2023. ISSN 1432-1084. doi: 10.1007/s00330-023-10526-1. URL https://link.springer.com/10.1007/s00330-023-10526-1.

Yael Maxwell. Coronary CTA Reimbursement for US Hospitals to Double in 2025, November 2024. URL https://www.tctmd.com/news/coronary-cta-reimbursement-us-hospitals-double-2025.

Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In *CVPR*, pages 1–11, 2016. ISBN 978-1-5090-5407-7. doi: 10.1109/3DV.2016.79. URL http://arxiv.org/abs/1606.04797.

Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y. Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning Where to Look for the Pancreas, May 2018. URL http://arxiv.org/abs/1804.03999. arXiv:1804.03999 [cs].

Joyce Peper, Dominika Suchá, Martin Swaans, and Tim Leiner. Functional cardiac CT–Going beyond Anatomical Evaluation of Coronary Artery Disease with Cine CT, CT-FFR, CT Perfusion and Machine Learning. *The British Journal of Radiology*, 93(1113): 20200349, September 2020. ISSN 0007-1285, 1748-880X. doi: 10.1259/bjr.20200349. URL https://academic.oup.com/bjr/article/doi/10.1259/bjr.20200349/7452065.

Russell A. Reeves, Ethan J. Halpern, and Vijay M. Rao. Cardiac Imaging Trends from 2010 to 2019 in the Medicare Population. *Radiology: Cardiothoracic Imaging*, 3(5): e210156, October 2021. ISSN 2638-6135. doi: 10.1148/ryct.2021210156. URL http://pubs.rsna.org/doi/10.1148/ryct.2021210156.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *MICCAI*, pages 234–241. 2015. ISBN 978-3-319-24573-7. doi: 10.1007/978-3-319-24574-4_28. URL http://link.springer.com/10.1007/978-3-319-24574-4_28. ISSN: 16113349.

Sujith K. Seneviratne, Quynh A. Truong, Fabian Bamberg, Ian S. Rogers, Michael D. Shapiro, Christopher L. Schlett, Claudia U. Chae, Ricardo Cury, Suhny Abbara, Thomas J. Brady, John T. Nagurney, and Udo Hoffmann. Incremental Diagnostic Value of Regional Left Ventricular Function Over Coronary Assessment by Cardiac Computed Tomography for the Detection of Acute Coronary Syndrome in Patients With Acute Chest Pain: From the ROMICAT Trial. *Circulation: Cardiovascular Imaging*, 3(4):375–383, July 2010. ISSN 1941-9651, 1942-0080. doi: 10.1161/CIRCIMAGING.109.892638. URL https://www.ahajournals.org/doi/10.1161/CIRCIMAGING.109.892638.

Jing Wang, Shuyu Wang, Wei Liang, Nan Zhang, and Yan Zhang. The auto segmentation for cardiac structures using a dual-input deep learning network based on vision saliency and transformer. *Journal of Applied Clinical Medical Physics*, 23(5):e13597, May 2022. ISSN 1526-9914, 1526-9914. doi: 10.1002/acm2.13597. URL `https://aapm.onlinelibrary.wiley.com/doi/10.1002/acm2.13597`.

Wenbin Wu, Guanjun Liu, Kaiyi Liang, and Hui Zhou. Inner Cascaded U2-Net: An Improvement to Plain Cascaded U-Net. *Computer Modeling in Engineering & Sciences*, 134(2):1323–1335, 2023. ISSN 1526-1506. doi: 10.32604/cmes.2022.020428. URL `https://www.techscience.com/CMES/v134n2/49520`.

Yuxin Wu and Kaiming He. Group Normalization, June 2018. URL `http://arxiv.org/abs/1803.08494`. arXiv:1803.08494 [cs].

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Transactions on Medical Imaging*, 39(6):1856–1867, June 2020. ISSN 0278-0062, 1558-254X. doi: 10.1109/TMI.2019.2959609. URL `https://ieeexplore.ieee.org/document/8932614/`.