A Multimodal Contrastive Learning for Detecting Aortic Dissection on 3D Non-Contrast CT with Anatomy Simplification

Duoer Zhang[1], Wenbo Xiao[2], Chen Jiang[1,3], Yuxuan Qiu[1,3], Zhan Feng[2], Hong Wang[4], Yefeng Zheng[5], and Wentao Zhu[1,5⋆]

[1] Zhejiang Lab, Hangzhou, China
[2] The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China
[3] Hangzhou Institute for Advanced Study, UCAS, Hangzhou, China
[4] Xi'an Jiaotong University, Xi'an, China
[5] Westlake University, Hangzhou, China
[6] The College of Optical Science and Engineering, Zhejiang University, Hangzhou, China
zhuwentao.ee@gmail.com

Abstract. Accurate detection of aortic dissection (AD) in emergency settings is of significant importance, as misdiagnosis can significantly delay subsequent treatments and even endanger patients' lives. Currently, non-contrast CT scans are standard protocols in emergency departments for patients with chest pain, yet their ability to detect AD remains limited. We introduce a novel multimodal contrastive learning framework designed to learn discriminative features from both contrast-enhanced CT and corresponding diagnostic reports. These features are then aligned with non-contrast CT scans through a multimodal contrastive learning approach. Specifically, we first segment and straighten the aorta to effectively apply attention to the aortic area. Finally, the pre-trained encoder is fine-tuned for the tasks of AD detection and lumen segmentation using non-contrast CT scans. Our experiments, conducted on a test dataset comprising 239 subjects (127 with AD and 112 without), demonstrated that the proposed framework achieves an accuracy of 0.958, an F1-score of 0.969, and an AUC of 0.983 in AD detection. These results surpass those of six state-of-the-art classification models. In lumen segmentation experiments, the framework achieves an average DSC of 0.705, outperforming others. These findings indicate that our proposed framework not only outperforms existing AD detection methods but also holds the potential to accurately localize false lumen using non-contrast CT scans alone.

Keywords: Aortic Dissection · Classification · Segmentation · Multimodal Contrastive Learning.

⋆ Corresponding Author

1 Introduction

Aortic dissection (AD) is a critical and life-threatening vascular disease characterized by a tear in the inner layer of the aorta, leading to the formation of a false lumen within the aortic wall [21]. This condition can precipitate aortic rupture, resulting in fatal hemorrhaging. If patients were not treated in time, the 48-hour mortality rate could reach 30% [13]. It is concerning that approximately one-third of patients are misdiagnosed or experience diagnostic delays [17]. Especially in the emergency department, missing a diagnosis can be devastating for the patient and pose challenges to the health service provision [10].

Currently, contrast-enhanced computed tomography (CE-CT) is the gold standard for diagnosing AD. It provides radiologists with crucial information for making conclusive decisions [13,16]. However, CE-CT is expensive and time-consuming. Moreover, its use of contrast agents may cause severe allergic events and kidney failure [5,28]. In contrast, non-contrast CT (NC-CT) is widely available in most emergency rooms and can be performed rapidly [20]. However, NC-CT has limitations in diagnostic accuracy, especially in cases with nonspecific symptoms, which often lead to misdiagnoses in the clinic. Therefore, improving the sensitivity and accuracy of AD diagnosis on NC-CT in an emergency setting is clinically demanding.

Deep learning methods have shown outstanding performance in medical image analysis [6,19]. Several publications have been investigated to detect AD on NC-CT in recent years [2,4,8,18]. Xiong et al. [27] proposed a cascaded multi-task generation framework to simultaneously synthesize CE-CT, segment the lumens, and detect AD. However, due to the limitations of multi-task training, there is still potential to improve both accuracy and segmentation performance.

Cheng et al. [3] proposed a 3D full-resolution U-Net algorithm to segment the true and false lumens for AD detection on NC-CT with an accuracy of 0.938. But, it utilized CT with a slice thickness of 0.625 mm, resulting in a relatively high cost for clinical applications.

Despite the simplicity and direct mapping characteristics of end-to-end methods, some state-of-the-art works have proven that large image-text models have become a very promising alternative in the field of medical image analysis. Multimodal contrastive learning extends the capabilities of contrastive learning by integrating information from multiple modalities, such as images and text [22]. Zhang et al. [29] presented a ConVIRT framework for learning visual representations by exploiting the naturally occurring pairing of images and textual data. In medical imaging, this approach leverages the synergy between radiological images and associated textual reports, enabling models to learn rich, cross-modal representations.

Inspired by these works, we propose a novel 3D multimodal contrastive learning framework for AD detection and lumen segmentation using NC-CT. The framework primarily consists of several innovative components, including data preprocessing, multimodal contrastive learning and sub-task training. To better identify and quantify the global volumetric features of the aorta and the local characteristics of the primary tear, aortic straightening is performed be-

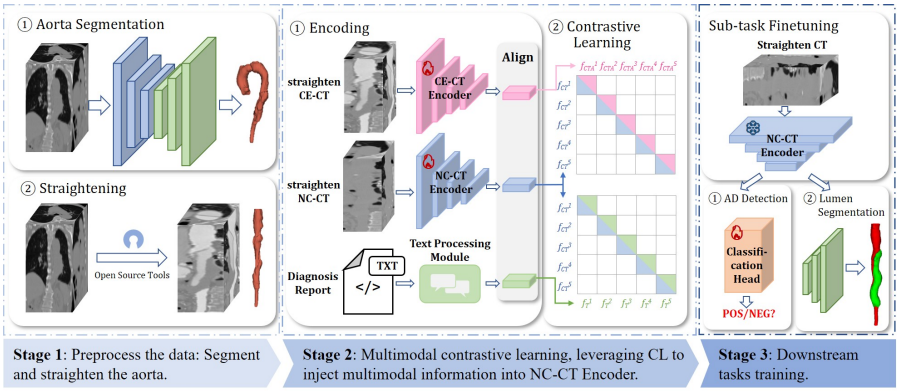

Fig. 1. An overview of our proposed multimodal contrastive learning framework. First, the aorta region is segmented and straightened. Subsequently, a multimodal contrastive learning model is trained to align the features of non-contrast CT scans with those derived from contrast-enhanced CT images and textual reports. Finally, the encoder is fine-tuned for the sub-tasks using non-contrast CT scans.

fore training [1]. Then, a pre-trained contrastive learning model is trained to enhance the encoder's feature representation by aligning features from CE-CT and textual descriptions to NC-CT. After, a specialized classification head and a decoder are trained on NC-CT alone using a fixed pre-trained encoder. This approach aims to provide a more accurate and efficient method for the diagnosis of AD and localization of false lumen in emergency settings, potentially leading to improved clinical decision-making and patient care.

The main innovations of the proposed method can be summarized as follows: 1) A novel multimodal contrastive learning architecture that integrates information from multiple modalities to enhance visual representations of NC-CT;2) Through the pre-trained NC-CT feature extraction encoder, different downstream tasks can be supported, such as AD detection and lumen segmentation, and achieve consistent performance improvements; 3) The proposed framework works effectively with 5-mm-thick NC-CT, presenting potential for clinical application in emergency settings.

2 Method

The 3D multimodal contrastive learning architecture overview is illustrated in Fig. 1, which has three main components: data preprocessing, a multimodal contrastive learning architecture, and sub-task networks. In the first step, we employ automatic methods to segment and straighten the aorta. In the second step, a multimodal contrastive learning model is utilized to train. In the last step, we fine-tune the pre-trained image encoder for AD classification and lumen segmentation using NC-CT.

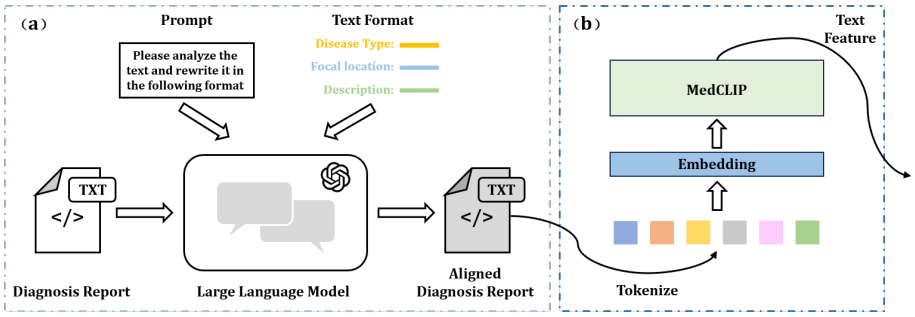

Fig. 2. Text processing module.

Aorta Segmentation and Anatomy Simplification. To reduce unrelated content noise, we extract the aorta for subsequent analysis. As shown in Stage 1 of Fig. 1, we employ an open-source deep learning segmentation tool, called TotalSegmentator [26] to segment the aorta from NC-CT. Then, we straighten the curved aorta to simplify the geometric complexity using automatic scripts from 3D Slicer (version 5.6.2, http://www.slicer.org).

Task Definition. The aorta dataset used in this study could be represented as (x_u, x_v, x_t), where x_u represents NC-CT images, x_v represents CE-CT images, and x_t represents textual reports which describe the diagnostic information in x_v. Our goal is to train a parameterized image encoder E_u by contrastive learning, which maps an image to a fixed-dimensional vector. Then we utilize E_u for the downstream tasks.

Feature Extraction. To preserve the modality-specific characteristics of NC-CT and CE-CT, we employ two independent 3D ResNet encoders (E_u and E_v) for feature extraction. The weight-sharing strategy, though generally used in multimodal learning, is not suitable here due to the significant distribution gap between NC-CT and CE-CT caused by the contrast agent. By decoupling the encoders, E_u can focus on capturing structural information from NC-CT, while E_v is optimized for functional information from CE-CT. This design ensures that the modality-specific features are well preserved and aligned in the embedding space.

Furthermore, we design a text processing module to better extract the features of diagnosis reports. As shown in Fig. 2(a), We first introduce a large language model (GPT-4o) to standardize the textual inputs. The aligned diagnostic report is then converted to text features via a tokenizer and MedCLIP [25], this process is displayed in Fig. 2(b).

We sample a set of multimodal data pairs $(\tilde{x}_u, \tilde{x}_v, \tilde{x}_t)$ from the dataset. The data of different modalities within a single pair is converted into embedding features. Then a linear transformation is applied to obtain the resulting features f_u, f_v and f_t, where $f_{(u,v,t)} \in \mathbb{R}^d$. At the training stage, N input data

pairs $[(\tilde{x}_u^1, \tilde{x}_v^1, \tilde{x}_t^1), (\tilde{x}_u^2, \tilde{x}_v^2, \tilde{x}_t^2) \cdots (\tilde{x}_u^N, \tilde{x}_v^N, \tilde{x}_t^N)]$ sampled from training dataset is encoded into $[(f_u^1, f_v^1, f_t^1), (f_u^2, f_v^2, f_t^2) \cdots (f_u^N, f_v^N, f_t^N)]$.

Training Loss Function. In the training process, a weight-aware multimodal contrastive loss $\mathcal{L}_{con}$ is proposed to maximize the similarity between aligned modalities:

$$\mathcal{L}_{con} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{(\alpha,\beta)\in\mathcal{M}} \varphi(\alpha, \beta) \cdot \left[\log \frac{e^{sim\langle f_\alpha^i, f_\beta^i \rangle / \tau}}{\sum_{k=1}^{N} e^{sim\langle f_\alpha^i, f_\beta^k \rangle / \tau}} \right] \tag{1}$$

where $\mathcal{M} = \{(u, v), (u, t)\}$, N represents the batch size, $sim\left\langle f_\alpha^i, f_\beta^i \right\rangle$ represents the cosine similarity, $\tau \in \mathbb{R}^+$ is a temperature hyperparameter, and $\varphi(\alpha, \beta)$ is a scalar that can be dynamically adjusted according to:

$$\varphi(\alpha, \beta) = \frac{1}{1 + \frac{1}{N} \sum_{i=1}^{N} sim\left\langle f_\alpha^i, f_\beta^i \right\rangle} \tag{2}$$

Furthermore, we propose a modality consistency loss $\mathcal{L}_{cns}$ to better integrate information from the three modalities. $\mathcal{L}_{cns}$ emphasizes the consistency of the three modalities in the feature representation space, ensuring the features from different modalities describing the same object are as similar as possible, thus promoting the fusion of multimodal information:

$$\mathcal{L}_{cns} = -\frac{1}{N} \sum_{i=1}^{N} \left[\log \frac{e^{\sum_{(\alpha,\beta)\in\mathcal{M}'} sim\langle f_\alpha^i, f_\beta^i \rangle}}{\sum_{k=1}^{N} e^{\sum_{(\alpha,\beta)\in\mathcal{M}'} sim\langle f_\alpha^i, f_\beta^k \rangle}} \right] \tag{3}$$

where $\mathcal{M}' = \{(u, v), (u, t), (v, t)\}$. By combining $\mathcal{L}_{con}$ and $\mathcal{L}_{cns}$ with weights w_1 and w_2 respectively, the overall training loss $\mathcal{L}$ could be expressed as:

$$\mathcal{L} = w_1 \cdot \mathcal{L}_{con} + w_2 \cdot \mathcal{L}_{cns} \tag{4}$$

Sub-task Fine-tuning. We continue to fine-tune a classifier and a decoder to improve the performance of the previous well-trained feature extraction encoder for classification and segmentation. As shown in Stage 3 of Fig. 1, sub-task networks use the frozen pre-trained NC-CT encoder to transfer the NC-CT volume into a group of features and continue to train two sub-networks for AD detection and true & false lumen segmentation. Note that the sub-task network is a lightweight CNN using a binary cross-entropy loss:

$$L = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \tag{5}$$

Table 1. Validation results of AD detection by different models.

Methods	ACC	Sens	Spec	Prec	F1-Score	AUC
ResNet-18 [9]	0.912	0.921	0.902	0.914	0.918	0.976
ResNet-50 [9]	0.904	0.898	0.911	0.919	0.908	0.955
DenseNet-121 [11]	0.818	0.928	0.686	0.779	0.847	0.914
Inception-V3 [23]	0.797	0.906	0.667	0.764	0.829	0.888
DeepLA [7]	0.838	0.872	0.797	0.836	0.854	0.913
MTGA [27]	0.916	**0.953**	0.875	0.896	0.924	0.973
Ours	**0.958**	0.951	**0.973**	**0.987**	**0.969**	**0.983**

3 Experiments

3.1 Experimental setup

Dataset. The dataset was collected from patients with both NC-CT scan and CE-CT scan with diagnostic reports from xxx Hospital. The training dataset consisted of 580 subjects (290 with AD and 290 without), and the test dataset consisted of 239 subjects (127 with AD and 112 without). We register the images using Elastix [14].

The output volume size is set to $80 \times 80 \times 128$. All the volumes were resampled to an average resolution of $1.0 \times 1.0 \times 5.0$ mm^3. The CT values of the NC-CT and CE-CT volumes are truncated to the range of $[0, 300]$ and $[0, 800]$ in the Hounsfield unit, respectively.

Implementation Details. The proposed framework is implemented using PyTorch (version 1.8) on an NVIDIA V100 GPU. The contrastive learning model is trained for 500 epochs using a cosine optimizer with an initial learning rate of 1×10^{-4}. The sub-task network is trained for 200 epochs using an Adam optimizer with an initial learning rate of 1×10^{-5}. w_1 and w_2 in eq. 4 are set to 1 and 0.01, respectively.

We compare the proposed framework with four state-of-the-art classification models, such as ResNet-18 [9], ResNet-50, DenseNet-121 [11] and Inception-V3 [23]. In addition, two AD detection frameworks, DeepLA [7] and MTGA [27] are included. For segmentation performance, three segmentation models are included, 3D UXNet [15], 3D nnU-Net [12] and SAM-Med3D [24].

To evaluate the performance of AD detection, we employ several metrics, including accuracy (ACC), sensitivity (Sens), specificity (Spec), precision (Prec), F1-score and Area Under the Curve (AUC). To evaluate the performance of lumen segmentation, we employ the Dice Similarity Coefficient and Jaccard Index.

3.2 Experimental results

AD Detection. The key purpose of the proposed method is to classify AD and non-AD using NC-CT. All the classification tasks are evaluated on our test

Table 2. Results of lumen segmentation on NC-CT by different models.

	nnUNet [12]	SAMM3D [24]	UXNet [15]	Ours	NoS
	Dice Similarity Coefficient				
Normal	0.823	0.823	0.848	**0.866**	0.796
True	0.653	0.618	0.661	**0.672**	0.620
False	0.493	0.508	0.571	**0.577**	0.538
Avg	0.656	0.650	0.693	**0.705**	0.651
	Jaccard Index				
Normal	0.700	0.699	0.738	**0.765**	0.671
True	0.494	0.459	0.504	**0.517**	0.464
False	0.350	0.360	0.415	**0.424**	0.395
Avg	0.515	0.506	0.552	**0.569**	0.510

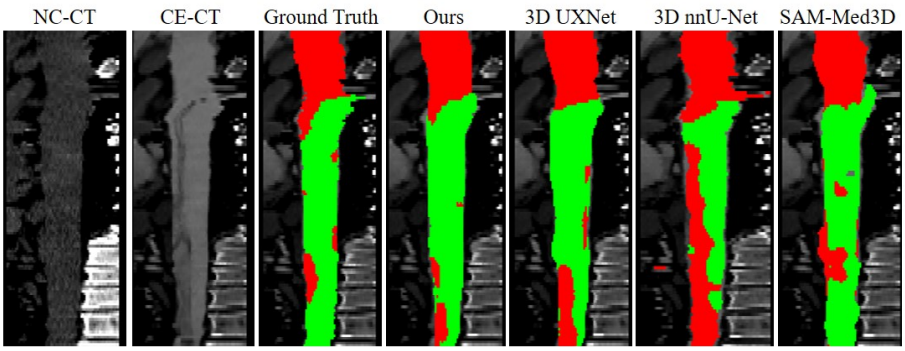

Fig. 3. Example segmentation results of true & false lumens by different models.

dataset. Table 1 shows the performance of our proposed method and benchmark methods on AD detection. Our method achieves an accuracy of 0.958, a Spec of 0.973, a Prec of 0.987, an F1-score of 0.969, and an AUC of 0.983. These results outperform those of four SOTA classification models and two previously published AD detection frameworks.

The superior performance of the proposed method can be attributed to its unique multimodal contrastive learning architecture. By integrating information from multiple modalities in the pre-training process, the encoder can capture more comprehensive and discriminative features from NC-CT to enhance classification accuracy significantly. Thus, it has the potential to enable earlier and more cost-effective diagnosis of AD in emergency settings.

True and False Lumen Segmentation. An additional sub-task is to segment the true and false lumens of AD patients which represents the most clinically concerning hazard, aortic tear. The quantitative segmentation results are presented in Table 2 by our method and three SOTA methods. We also compared the results obtained by our method on the dataset without anatomy simplifi-

Table 3. Ablation results of multimodal contrastive learning method.

Methods	ACC	Sens	Spec	Prec	F1-Score	AUC
NCCT-CECT	0.791	0.795	0.786	0.808	0.801	0.897
NCCT-TEXT	0.895	0.945	0.839	0.870	0.906	0.964
NCCT-CECT-TEXT (NoS)	0.899	0.874	0.929	0.933	0.902	0.968
NCCT-CECT-TEXT (ResNet-18)	0.937	0.945	0.929	0.938	0.941	0.975
NCCT-CECT-TEXT (Ours)	**0.958**	**0.951**	**0.973**	**0.987**	**0.969**	**0.983**

cation, which is denoted as NoS(non-straightened aorta data) in the table. Our method outperforms others with a DSC of 0.866 for lumen without dissection and 0.672 & 0.577 for true & false lumens with dissection, respectively. The segmentation results demonstrate a more accurate location of the false lumen of AD patients, providing more insights for subsequent clinical treatments. An example of segmentation results is shown in Fig. 3, red for true lumen and green for false lumen. The proposed framework's segmented true and false lumen masks were more accurate than others.

Ablation Study. The ablation results are summarized in Table 3 with three main parts. First, we split the proposed multimodal framework into two independent sub-frameworks to separately verify the efficiency of image-text and image-image compared to our proposed image-image-text learning. Second, we verify the performance of the non-straightened aorta data. We also verify the performance of ResNet-50 encoder compared to a smaller model like ResNet-18. Our proposed method achieves the best performance, highlighting the benefits of integrating different modalities and aorta geometry simplification.

4 Discussion and Conclusion

In this study, we proposed a multimodal contrastive learning framework designed to improve the performance of AD detection and lumen segmentation using NC-CT. The image encoder was pre-trained to extract features from CE-CT and corresponding textual reports through multimodal contrastive learning, and then finetuned a classifier and a decoder for AD detection and true & false lumen segmentation. Our experimental results demonstrated that this framework significantly outperforms current SOTA models in terms of AD detection accuracy and segmentation performance based on NC-CT.

However, our proposed method still has limitations, such as only using the CT information of the aorta area and discarding other content that can assist in diagnosis, and the dataset size is still limited for large models which may limit the performance. We will further improve it in subsequent work.

Despite this, the proposed methodology holds substantial promise for minimizing AD misdiagnosis in emergency settings through the automatic workflow with widely available non-contrast CT.

Acknowledgments. This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 62401516), and the National Natural Science Foundation of China (Grant No. 62427807).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, D., Zhang, X., Mei, Y., Liao, F., Xu, H., Li, Z., Xiao, Q., Guo, W., Zhang, H., Yan, T., et al.: Multi-stage learning for segmentation of aortic dissections using a prior aortic anatomy simplification. Medical Image Analysis **69**, 101931 (2021)
2. Chen, H., Yan, S., Xie, M., Huang, J.: Application of cascaded GAN based on ct scan in the diagnosis of aortic dissection. Computer Methods and Programs in Biomedicine **226**, 107130 (2022)
3. Cheng, Z., Zhao, L., Yan, J., Zhang, H., Lin, S., Yin, L., Peng, C., Ma, X., Xie, G., Sun, L.: A deep learning algorithm for the detection of aortic dissection on non-contrast-enhanced computed tomography via the identification and segmentation of the true and false lumens of the aorta. Quantitative Imaging in Medicine and Surgery **14**(10), 7365 (2024)
4. Dong, F., Song, J., Chen, B., Xie, X., Cheng, J., Song, J., Huang, Q.: Improved detection of aortic dissection in non-contrast-enhanced chest CT using an attention-based deep learning model. Heliyon **10**(2), e24547 (2024)
5. El-Abd, Y.J., Hagspiel, K.D.: Review of imaging with focus on new techniques in aortic dissection. Techniques in Vascular and Interventional Radiology **24**(2), 100748 (2021)
6. Erickson, B.J., Korfiatis, P., Akkus, Z., Kline, T.L.: Machine learning for medical imaging. Radiographics **37**(2), 505–515 (2017)
7. Hata, A., Yanagawa, M., Yamagata, K., Suzuki, Y., Kido, S., Kawata, A., Doi, S., Yoshida, Y., Miyata, T., Tsubamoto, M., Kikuchi, N., Tomiyama, N.: Deep learning algorithm for detection of aortic dissection on non-contrast-enhanced CT. European Radiology **31**, 1151–1159 (2020)
8. Hata, A., Yanagawa, M., Yamagata, K., Suzuki, Y., Kido, S., Kawata, A., Doi, S., Yoshida, Y., Miyata, T., Tsubamoto, M., et al.: Deep learning algorithm for detection of aortic dissection on non-contrast-enhanced CT. European Radiology **31**, 1151–1159 (2021)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 770–778 (2015)
10. Howard, D.P., Banerjee, A., Fairhead, J.F., Perkins, J., Silver, L.E., Rothwell, P.M.: Population-based study of incidence and outcome of acute aortic dissection and premorbid risk factor control: 10-year results from the Oxford vascular study. Circulation **127**(20), 2031–2037 (2013)
11. Huang, G., Liu, Z., Weinberger, K.Q.: Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 2261–2269 (2016)
12. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**, 203 – 211 (2020)

13. Kicska, G.A., Koweek, L.M.H., Ghoshhajra, B.B., Beache, G.M., Brown, R.K., Davis, A.M., Hsu, J.Y., Khosa, F., Kligerman, S.J., Litmanovich, D., et al.: ACR Appropriateness Rriteria® suspected acute aortic syndrome. Journal of the American College of Radiology **18**(11), S474–S481 (2021)
14. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity-based medical image registration. IEEE Transactions on Medical Imaging **29**(1), 196–205 (2009)
15. Lee, H.H., Bao, S., Huo, Y., Landman, B.A.: 3d ux-net: A large kernel volumetric convnet modernizing hierarchical transformer for medical image segmentation. ArXiv **abs/2209.15076** (2022)
16. LePage, M.A., Quint, L.E., Sonnad, S.S., Deeb, G.M., Williams, D.M.: Aortic dissection: CT features that distinguish true lumen from false lumen. American Journal of Roentgenology **177**(1), 207–211 (2001)
17. Lovatt, S., Wong, C.W., Schwarz, K., Borovac, J.A., Lo, T., Gunning, M., Phan, T., Patwala, A., Barker, D., Mallen, C.D., et al.: Misdiagnosis of aortic dissection: a systematic review of the literature. The American Journal of Emergency Medicine **53**, 16–22 (2022)
18. Lyu, J., Fu, Y., Yang, M., Xiong, Y., Duan, Q., Duan, C., Wang, X., Xing, X., Zhang, D., Lin, J., et al.: Generative adversarial network–based noncontrast CT angiography for aorta and carotid arteries. Radiology **309**(2), e230681 (2023)
19. Niederer, S.A., Lumens, J., Trayanova, N.A.: Computational models in cardiology. Nature Reviews Cardiology **16**(2), 100–111 (2019)
20. Nienaber, C.A., Clough, R.E.: Management of acute aortic dissection. The Lancet **385**(9970), 800–811 (2015)
21. Nienaber, C.A., Clough, R.E., Sakalihasan, N., Suzuki, T., Gibbs, R., Mussa, F., Jenkins, M.P., Thompson, M.M., Evangelista, A., Yeh, J.S., et al.: Aortic dissection. Nature Reviews Disease Primers **2**(1), 1–18 (2016)
22. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International Conference on Machine Learning. pp. 8748–8763. PmLR (2021)
23. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 2818–2826 (2015)
24. Wang, H., Guo, S., Ye, J., Deng, Z., Cheng, J., Li, T., Chen, J., Su, Y., Huang, Z., Shen, Y., et al.: Sam-med3d: towards general-purpose segmentation models for volumetric medical images. arXiv preprint arXiv:2310.15161 (2023)
25. Wang, Z., Wu, Z., Agarwal, D., Sun, J.: MedCLIP: Contrastive learning from unpaired medical images and text. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. vol. 2022, p. 3876 (2022)
26. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., et al.: TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. Radiology: Artificial Intelligence **5**(5), e230024 (2023)
27. Xiong, X., Ding, Y., Sun, C., Zhang, Z., Guan, X., Zhang, T., Chen, H., Liu, H., Cheng, Z., Zhao, L., et al.: A cascaded multi-task generative framework for detecting aortic dissection on 3-D non-contrast-enhanced computed tomography. IEEE Journal of Biomedical and Health Informatics **26**(10), 5177–5188 (2022)
28. Yu, Y.T., Ren, X.S., An, Y.Q., Yin, W.H., Zhang, J., Wang, X., Lu, B.: Changes in the renal artery and renal volume and predictors of renal atrophy in patients

with complicated type B aortic dissection after thoracic endovascular aortic repair. Quantitative Imaging in Medicine and Surgery **12**(11), 5198 (2022)
29. Zhang, Y., Jiang, H., Miura, Y., Manning, C.D., Langlotz, C.P.: Contrastive learning of medical visual representations from paired images and text. In: Machine Learning for Healthcare Conference. pp. 2–25. PMLR (2022)