A Causality-Inspired Model for Intima-Media Thickening Assessment in Ultrasound Videos

Shuo Gao[†1], Meng Yang[†3], Jun Xue[3], Yang Chen[2], Jingyang Zhang[2(✉)], and Guangquan Zhou[1(✉)]

[1]School of Biological Science and Medical Engineering, Southeast University, Nanjing, China
guangquan.zhou@seu.edu.cn
[2]School of Computer Science and Engineering, Southeast University, Nanjing, China
zjysjtu1994@gmail.com
[3]Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing, China

Abstract. Carotid atherosclerosis represents a significant health risk, with its early diagnosis primarily dependent on ultrasound-based assessments of carotid intima-media thickening. However, during carotid ultrasound screening, significant view variations cause style shifts, impairing content cues related to thickening, such as lumen anatomy, which introduces spurious correlations that hinder assessment. Therefore, we propose a novel causal-inspired method for assessing carotid intima-media thickening in frame-wise ultrasound videos, which focuses on two aspects: eliminating spurious correlations caused by style and enhancing causal content correlations. Specifically, we introduce a novel Spurious Correlation Elimination (SCE) module to remove non-causal style effects by enforcing prediction invariance with style perturbations. Simultaneously, we propose a Causal Equivalence Consolidation (CEC) module to strengthen causal content correlation through adversarial optimization during content randomization. Simultaneously, we design a Causal Transition Augmentation (CTA) module to ensure smooth causal flow by integrating an auxiliary pathway with text prompts and connecting it through contrastive learning. The experimental results on our in-house carotid ultrasound video dataset achieved an accuracy of 86.93%, demonstrating the superior performance of the proposed method. Code is available at https://github.com/xielaobanyy/causal-imt.

Keywords: Intima-Media Thickening · Ultrasound · Causality Analysis.

1 Introduction

Carotid atherosclerosis is a widespread and complex cardiovascular disease that poses a notable global public health threat [1]. Intima-Media Thickening (IMT)

[†] Equal contribution

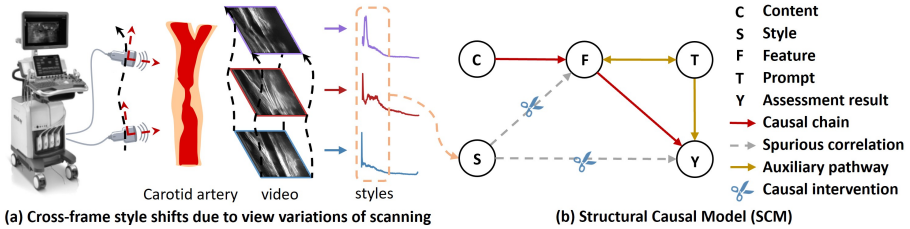

Fig. 1. The causal mechanism for intima-media assessment: (a) Cross-frame style shifts due to view variations of scanning; (b) Structural Causal Model (SCM).

acts as a critical indicator for screening carotid atherosclerosis [16], where ultrasound emerges as the preferred modality for imaging owing to its non-invasiveness and accessibility [17]. In clinical practice, the physician maneuvers the ultrasound probe along the carotid artery to collect a complete ultrasound video for tracking intima-media changes. This video contains numerous frames that reveal the spatial distribution of thickening along the carotid artery [5], providing valuable insights for characterizing vascular health and designing personalized treatment. Therefore, it is desirable to perform IMT assessment in a frame-wise manner for the carotid ultrasound video, leading to fine-grained cues for thickening position.

For frame-wise IMT assessment in carotid ultrasound videos, existing methods [9,21] typically treat it as a frame-level event detection task [2], often relying on off-the-shelf paradigms from the field of generic video processing. For example, the spatiotemporal aggregation module originally designed for generic object tracking [3] is directly transferred to the carotid ultrasound video for capturing dynamic IMT features [20]. As these existing methods are rooted for the general video analysis with only mild view variations across frames [6], their efficacy on carotid ultrasound videos for IMT assessment heavily relies on a strict scanning protocol [22], i.e., holding the ultrasound probe with a fixed position and static orientation to keep cross-frame view stability. However, this protocol contradicts clinical practice, where the physician skillfully maneuvers the ultrasound probe, dynamically scanning along the carotid artery while adjusting its orientation to optimize tissue penetration with clear intima-media [12], as shown in Fig. 1(a). In this way, substantial view variations would occur across frames with discrepancies in echo intensity, causing style shifts that corrupt IMT-related content cues, e.g., lumen anatomy. Hence, IMT assessment for these view-varying frames tends to be misguided by spurious correlations to style shifts, while ignoring causal correlations to content cues.

To address this issue, we construct a Structural Causal Model (SCM) for IMT assessment, as shown in Fig. 1(b), to identify the non-causal impact pathway of spurious correlations and explore an ideal causal chain. Theoretically, the causal content C should be the only endogenous parent that determines IMT assessment Y by deriving the feature F as an intermediate result, thereby forming a causal chain $C \rightarrow F \rightarrow Y$. However, view-varying frames contains diverse styles S that misguide feature extraction and, finally, IMT assessment through

$S \rightarrow F \rightarrow Y$, opening a backdoor path with spurious correlation $S \dashrightarrow Y$. Based on the above causality mechanism, the elimination of spurious correlation can be achieved by blocking the non-causal path, i.e., $S \not\rightarrow F \rightarrow Y$, shielding prediction from influence of style shifts. Moreover, causal correlations can be enhanced via the causal chain $C \rightarrow F \rightarrow Y$ with two aspects: 1) ensuring causal equivalence between C and Y for a direct impact consolidation; 2) improving causal transition by augmenting an auxiliary pathway $C \rightarrow F \leftrightarrow T \rightarrow Y$, where T prompts Y and recalibrate F to smoothly transfer causal impact through the causal chain.

In this paper, we present, to our knowledge, *the first causality-inspired model* for IMT assessment in ultrasound videos. This model eliminates spurious style correlations across view-varying frames while enhancing the causal content correlations for accurate assessment. Specifically, our contributions are three folds: 1) We propose a Spurious Correlation Elimination (SCE) module that cut-offs non-causal style impacts by enforcing prediction invariance with simulated style perturbations; 2) We develop a Causal Equivalence Consolidation (CEC) module to enhance the direct causal content correlation, via adversarial optimization on assessment predictions under content randomization; and 3) We design a Causal Transition Augmentation (CTA) module for a smooth causal impact flow, where an auxiliary causal pathway is formed using text prompts with chain-of-thought guidance and further involved into the causal chain by contrastive learning. We evaluate our method on a in-house dataset of carotid ultrasound videos, showing its highest accuracy and clear advantages for frame-wise IMT assessment.

2 Methodology

As illustrated in Fig. 2, the frames of a carotid ultrasound video are fed into the proposed causality-inspired model for frame-wise IMT assessment. Specifically, the style shifts across these view-varying frames are identified as spurious correlations to be eliminated, while causal content factors are enhanced via equivalence consolidation and transition augmentation.

2.1 Spurious Correlation Elimination (SCE)

In clinical practice, physicians flexibly manipulate the ultrasound probe for carotid scanning, leading to varying views across frames in a ultrasound video. It would cause cross-frame shifted styles S, which corrupts the extracted features F and finally misguides the IMT assessment C via a spurious correlation ($S \rightarrow F \rightarrow Y$). To eliminate such spurious correlation, we apply a causal intervention on F by a do-operator [11], which simplifies the correlation analysis between S and Y with F conditional fixed:

$$P(Y|do(F)) = \sum\nolimits_S P(Y|do(F), S) \cdot P(S|do(F)) = \sum\nolimits_S P(Y|F, S) \cdot P(S), \quad (1)$$

which depicts the cumulative effect of multiple S on Y. It implies that, to eliminate their spurious correlations, it is crucial to disentangle the correlations be-

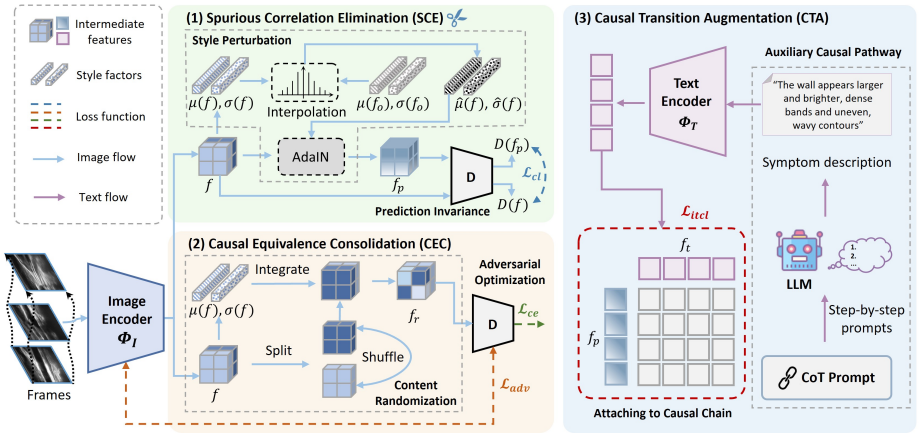

Fig. 2. The architecture of the proposed method. Specifically, we apply Spurious Correlation Elimination (SCE) module to remove non-causal style effects by enforcing prediction invariance with style perturbations (Sect. 2.1), then we design Causal Equivalence Consolidation (CEC) module to strengthen causal content correlation by adversarial optimization during content randomization (Sect. 2.2). Furthermore, we design a Causal Transition Augmentation (CTA) module to enhance causal effects by integrating an auxiliary causal pathway and linking it by contrastive learning (Sect. 2.3).

tween Y and a sufficiently large set of S. To this end, we simulate style perturbations with a highly diverse distribution and enforce prediction invariance over them, facilitating potential spurious correlations can be thoroughly eliminated.

Style Perturbation. Style factor is typically represented using feature statistic, i.e., mean μ and standard deviation σ [14]. To ensure a wide perturbation range for style, we leverage shifted styles $\mu(f_o)$ and $\sigma(f_o)$ from view-varying frames and further augment them via random style interpolation, leading to a perturbed style $\hat{\mu}(f)$ and $\hat{\sigma}(f)$:

$$\hat{\mu}(f) = \lambda \cdot \mu(f) + (1-\lambda) \cdot \mu(f_o),\ \hat{\sigma}(f) = \lambda \cdot \sigma(f) + (1-\lambda) \cdot \sigma(f_o), \tag{2}$$

where $\lambda \sim \text{Uniform}(0, 1)$ is a random interpolation weight. After acquiring this diverse style, we further transfer it by Adaptive Instance Normalization (AdaIN) [10] to obtain style-perturbed features f_p based on original features f:

$$f_p = \hat{\sigma}(f) \cdot \left(\frac{f - \mu(f)}{\sigma(f)} \right) + \hat{\mu}(f). \tag{3}$$

Prediction Invariance. Eventually, we design a consistency loss $\mathcal{L}_{\text{cl}}$ using KL divergence D_{KL} to ensure the invariance of predictions between f and f_p:

$$\min_{\Phi_I,\, D} \mathcal{L}_{\text{cl}} = D_{KL}[D(f)||D(f_p)] + D_{KL}[D(f_p)||D(f)], \tag{4}$$

$$\min_{D} \mathcal{L}_{\text{ce}} = -y \log\left(D\left(f\right)\right). \tag{5}$$

where $D(f)$ and $D(f_p)$ represent the corresponding predicted distributions, and y represents the labels for thickening or non-thickening.

2.2 Causal Equivalence Consolidation (CEC)

In the idea causal chain, the assessment prediction Y is primarily determined by the content C via the intermediate feature F, i.e., $C \rightarrow F \rightarrow Y$. It motivates that the prediction results should be strongly correlated with the content factor, i.e., with high causal equivalence. To achieve this, we propose a Causal Equivalence Consolidation (CEC) module to strengthen the causal content correlation, which randomizes content factors and performs adversarial optimization on assessment predictions to enforce their causal equivalence towards content.

Content Randomization. As the content factor is typically defined as the channel order of the feature map [7], we randomize the content by independently splitting each feature channel and shuffling them, achieving features with random content $\mathcal{R}(f)$. Moreover, to avoid potential style degeneration during this content randomization [4], we integrate the original style factors $\mu(f)$ and $\sigma(f)$ into $\mathcal{R}(f)$ for style stability.

$$f_r = \mu(f) + \mathcal{R}(f) \cdot \sqrt{\sigma(f)^2 + \epsilon}. \tag{6}$$

Adversarial Optimization. To ensure causal equivalence between content and assessment predictions, we employ adversarial optimization on these assessment predictions to align their distribution with even randomized contents. Specifically, we minimize the adversarial loss $\mathcal{L}_{\text{adv}}$ to encourage the image encoder Φ_I to fool the classifier D, making it use content-randomized feature f_r to generate low-confidence predictions $D(f_r)$ with uniform probability $U = 1/2$:

$$\min_{\Phi_I} \mathcal{L}_{\text{adv}} = -U \log \left(D \left(f_r \right) \right). \tag{7}$$

Intuitively, randomized contents are enforced to produce meaningless assessment, enhancing the causal sensitivity to the content and ensuring causal equivalence.

2.3 Causal Transition Augmentation (CTA)

Besides causal quaivalence consolidation, another aspect for causal correlation enhancement is to design a Causal Transition Augmentation (CTA) mechanism, where an auxiliary pathway is attached upon the causal chain to avoid the degraded causal propagation, i.e., $C \rightarrow F \leftrightarrow T \rightarrow Y$. For this goal, we construct an auxiliary causal pathway using text prompts guided by chain-of-thought for fine-grained IMT symptoms. This pathway is then attached to the original causal chain using contrastive learning for alignment.

Auxiliary Causal Pathway Text conveys abundant information that complements images [15], serving as a valuable data resource for constructing an auxiliary causal pathway. However, naive text prompts are not well-suited for IMT assessment, as it is a highly complex task that typically demands a step-by-step

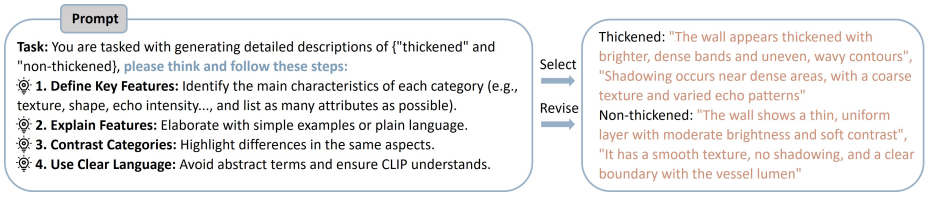

Fig. 3. Generation of intima-media symptom descriptions using step-type prompts.

logical process to define key symptoms. Motivated by this, we leverage the chain of thought [23] mechanism to guide the generation of fine-grained text prompts, as shown in Fig. 3. Such text prompts with symptom description establish an auxiliary pathway that augments the original image-based causal chain.

Attaching to Causal Chain. Since the auxiliary causal pathway is constructed using text prompts, we adpot contrastive learning [26] to align these text prompts with images into a shared feature space, attaching it to the original image-based causal chain. Specifically, text prompts are fed into the text encoder Φ_T to derive the text feature f_t, which is aligned to the image feature f via a contrastive loss:

$$\min_{\Phi_I} \mathcal{L}_{\text{itcl}} = -\log \frac{\exp\left(f \cdot f_t^+/\tau\right)}{\exp\left(f \cdot f_t^+/\tau\right) + \sum_{i \in \mathcal{N}} \exp\left(f \cdot f_t^{(i)}/\tau\right)}. \tag{8}$$

where f_t^+ is the embedding of the positive symptom prompt, $\mathcal{N}$ represents the set of negative samples and $f_t^{(i)}$ represents the negative ones. $(\cdot)$ denotes the cosine similarity between two embeddings. τ is a temperature scaling parameter.

2.4 Overall Loss

The total loss of our proposed causality-inspired model consists of several parts, where α_1 , α_2, and α_3 are trade-off parameters.

$$\mathcal{L}_{\text{total}} = \min_{\Phi_I} \mathcal{L}_{\text{itcl}} + \min_{\Phi_I, D} \alpha_1 \mathcal{L}_{\text{cl}} + \min_{D} \alpha_2 \mathcal{L}_{\text{ce}} + \min_{\Phi_I} \alpha_3 \mathcal{L}_{\text{adv}}. \tag{9}$$

3 Experiments

Datasets. In the IMT assessment task, we used a carotid ultrasound video dataset comprising 120 videos, including 30 thickened and 90 non-thickened videos. All frames were cropped to 506×477, resized to 224×224, and categorized into two types: thickened and non-thickened. Each frame is paired with a fine-grained textual description of its category and additional frame data for style information. The dataset was split into 60%, 20%, and 20% for training, validation, and testing, respectively. For this dataset, we adopted Accuracy, Sensitivity, Sensitivity, Precision and F1-score as evaluation metrics.

Table 1. The quantitative evaluation demonstrates the superiority of our method.

Method	Accuracy(%)	Sensitivity(%)	Precision(%)	F1-Score(%)
ResNet-50 [8]	78.14	32.35	76.64	77.30
APCNet [21]	77.60	23.53	74.65	75.83
DCCNet [25]	70.49	8.82	67.49	68.91
LSMD [20]	81.42	21.46	40.71	44.88
TSM [13]	77.77	14.84	85.19	77.78
MVFNet [24]	75.00	85.71	89.28	85.89
Proposed	**86.93**	**88.89**	**91.10**	**88.10**

Table 2. The ablation results of the proposed module.

SCE	CEC	CTA	Accuracy(%)	Sensitivity(%)	Precision(%)	F1-Score(%)
✓	✓		84.65	21.33	71.67	77.62
	✓	✓	76.70	51.85	81.56	78.62
✓		✓	85.23	88.89	90.52	86.69
✓	✓	✓	**86.93**	**88.89**	**91.10**	**88.10**

Implementation Details. We employed the pre-trained CLIP (ViT-B/16) model for image and text encoding [18]. It was trained using SGD Optimizer with learning rate 6×10^{-5}, batch size 16, and epoch number 100. Beside, We empirically set the overall loss weight coefficients α_1, α_2, and α_3 to 0.5, 0.1, and 0.1, respectively.

Comparison with State-of-the-Arts. We compare our method with existing approaches, such as **ResNet** [8], **APCNet**[21], **DCCNet**[25], **LSMD** [20], **TSM** [13] and **MVFNet** [24]. Competing models are retrained on our datasets using the training codes provided by their respective authors. The results indicate that our model shows superior performance in four key aspects (Table 1). Proposed method achieve superior performance in all experiments, surpassing the next best method, MVFNet, with average improvements of 11.93%, 3.18%, 1.81%, and 2.21% in accuracy, sensitivity, precision, and f1-score, respectively. Notably, due to the limited number of positive thickening cases, most models pose significant long-tail challenges during training. For example, the sensitivity for positive thickening in ResNet, APCNet, and DCCNet is 32.35, 23.53, and 8.82, respectively. However, our proposed method effectively addresses this issue. This success can be attributed to its ability to eliminate the spurious correlations caused by style shifts and enhance causal correlations to content, allowing the model to rely on content information to make decisions across different frames. Fig. 4 shows the visualization results of Grad-CAM [19], which shows that our proposed model focuses on the content cues of the intima-media, such as bifurcation locations and intima-media structure (last row). Furthermore, by incorporating customized intima-media symptom texts, this method captures

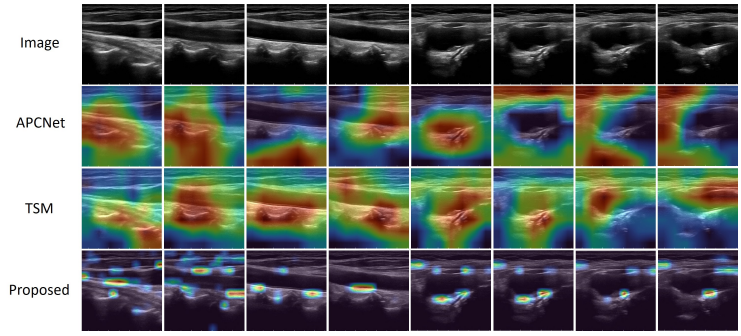

Fig. 4. Grad-CAM visualization results for different methods. The first row shows the original video frames, with the first four representing non-thickened frames and the last four representing thickened frames.

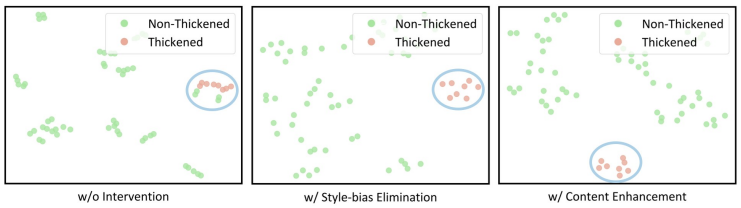

Fig. 5. t-SNE visualization in a batch with and without intervention.

fine-grained information related to thickening, enabling it to distinguish subtle differences between the two categories.

Ablation Study. Table 2 summarizes the ablation study for each module on the intima-media dataset. The results show that both the SCE and CEC significantly improve prediction performance. Specifically, incorporating SCE leads to improvements across all performance metrics, with accuracy and precision increasing by 10.23% and 9.54%, respectively. Furthermore, adding CEC further boosts model accuracy by 1.7%. Fig. 5 shows the t-SNE results of image feature distributions with and without causal intervention. The two right subfigures illustrate the model's enhanced thickened judgment after style intervention and content enhancement, with samples from the same batch clearly distinguished. Furthermore, compared to descriptions generated by general prompt-based methods (*Large, irregular zones dominate the image, with uneven borders and rough lines*), CTA demonstrates more robust performance in handling imbalanced class cases.

4 Conclusion

The assessment of IMT plays a vital role in the diagnosis and treatment of carotid atherosclerosis. To achieve this, we propose an assessment method based

on carotid ultrasound videos from a causal perspective. The method eliminates interference from style variations between video frames through causal modeling. Additionally, it enhances robustness by incorporating fine-grained symptom descriptions. This approach offers a new perspective for IMT assessment and contributes to advancing the application of causal learning in cardiovascular disease research. In the future, we will conduct more comprehensive experiments to validate the robustness of this method.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (U22A2023, 62325112, 62371121), the National Key R & D Program of China (2023YFC2411700, 2023YFC2411705), the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2024-RC320-02), the Research Start-up Grant of Southeast University (4009002412), the Civil Space Technology Pre-research Foundation (D010101), and the Jiangsu Key Research and Development Program (BE2022827).

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

References

1. Bir, S.C., Kelley, R.E.: Carotid atherosclerotic disease: A systematic review of pathogenesis and management. Brain circulation **8**(3), 127–136 (2022)
2. Chen, M., Wei, F., Li, C., Cai, D.: Frame-wise action representations for long videos via sequence contrastive learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13801–13810 (2022)
3. Cheng, H.K., Schwing, A.G.: Xmem: Long-term video object segmentation with an atkinson-shiffrin memory model. In: European Conference on Computer Vision. pp. 640–658. Springer (2022)
4. Deng, Y., Tang, F., Dong, W., Ma, C., Pan, X., Wang, L., Xu, C.: Stytr2: Image style transfer with transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11326–11336 (2022)
5. Espeland, M.A., Hoen, H., Byington, R., Howard, G., Riley, W.A., Furberg, C.D.: Spatial distribution of carotid intimal-medial thickness as measured by b-mode ultrasonography. Stroke **25**(9), 1812–1819 (1994)
6. Fasching, J., Walczak, N., Morellas, V., Papanikolopoulos, N.: Classification of motor stereotypies in video. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 4894–4900. IEEE (2015)
7. Gatys, L.A., Ecker, A.S., Bethge, M., Hertzmann, A., Shechtman, E.: Controlling perceptual factors in neural style transfer. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3985–3993 (2017)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
9. He, L., Yang, Z., Wang, Y., Chen, W., Diao, L., Wang, Y., Yuan, W., Li, X., Zhang, Y., He, Y., et al.: A deep learning algorithm to identify carotid plaques and assess their stability. Frontiers in Artificial Intelligence **7**, 1321884 (2024)

10. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE international conference on computer vision. pp. 1501–1510 (2017)
11. Jiao, L., Wang, Y., Liu, X., Li, L., Liu, F., Ma, W., Guo, Y., Chen, P., Yang, S., Hou, B.: Causal inference meets deep learning: A comprehensive survey. Research **7**, 0467 (2024)
12. Landwehr, P., Schulte, O., Voshage, G.: Ultrasound examination of carotid and vertebral arteries. European radiology **11**, 1521–1534 (2001)
13. Lin, J., Gan, C., Han, S.: Tsm: Temporal shift module for efficient video understanding. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 7083–7093 (2019)
14. Liu, Y., Qin, G., Chen, H., Cheng, Z., Yang, X.: Causality-inspired invariant representation learning for text-based person retrieval. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 14052–14060 (2024)
15. Long, S., Han, S.C., Wan, X., Poon, J.: Gradual: Graph-based dual-modal representation for image-text matching. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 3459–3468 (2022)
16. Nezu, T., Hosomi, N., Aoki, S., Matsumoto, M.: Carotid intima-media thickness for atherosclerosis. Journal of atherosclerosis and thrombosis **23**(1), 18–31 (2016)
17. Polak, J.F., O'Leary, D.H.: Carotid intima-media thickness as surrogate for and predictor of cvd. Global heart **11**(3), 295–312 (2016)
18. 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)
19. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp. 618–626 (2017)
20. Shan, C., Zhang, Y., Liu, C., Jin, Z., Cheng, H., Chen, Y., Yao, J., Luo, S.: Lsmd: Long-short memory-based detection network for carotid artery detection in b-mode ultrasound video streams. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control (2024)
21. Singh, S., Jain, P.K., Sharma, N., Pohit, M., Roy, S.: Atherosclerotic plaque classification in carotid ultrasound images using machine learning and explainable deep learning. Intelligent Medicine **4**(2), 83–95 (2024)
22. Touboul, P.J., Grobbee, D.E., Ruijter, H.d.: Assessment of subclinical atherosclerosis by carotid intima media thickness: technical issues. European journal of preventive cardiology **19**(2_suppl), 18–24 (2012)
23. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V., Zhou, D., et al.: Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems **35**, 24824–24837 (2022)
24. Wu, W., He, D., Lin, T., Li, F., Gan, C., Ding, E.: Mvfnet: Multi-view fusion network for efficient video recognition. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 2943–2951 (2021)
25. Yang, J., Li, X., Guo, Y., Song, P., Lv, T., Zhang, Y., Cui, Y.: Automated classification of coronary plaque on intravascular ultrasound by deep classifier cascades. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control (2024)
26. 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)