VasoMIM: Vascular Anatomy-aware Masked Image Modeling for Vessel Segmentation

AAAI-26 / IAAI-26 / EAAI-26
JANUARY 20–27, 2026 | SINGAPORE

Dexing Huang, Xiao-Hu Zhou*, Mei-Jiang Gui, Xiao-Liang Xie, Shi-Qi Liu, Shuangyi Wang, Rui-Ze Ma, Nu-Fang Xiao, Zeng-Guang Hou*

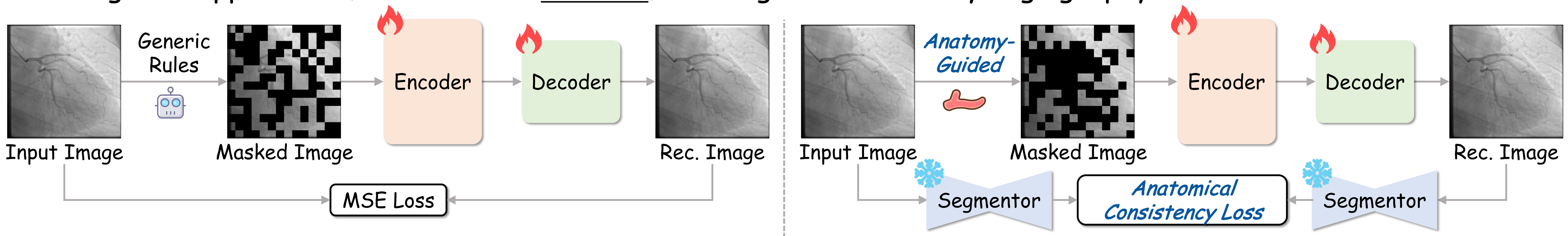
Institute of Automation, Chinese Academy of Sciences

{huangdexing2022, xiaohu.zhou, zengguang.hou}@ia.ac.cn

arXiv Contact

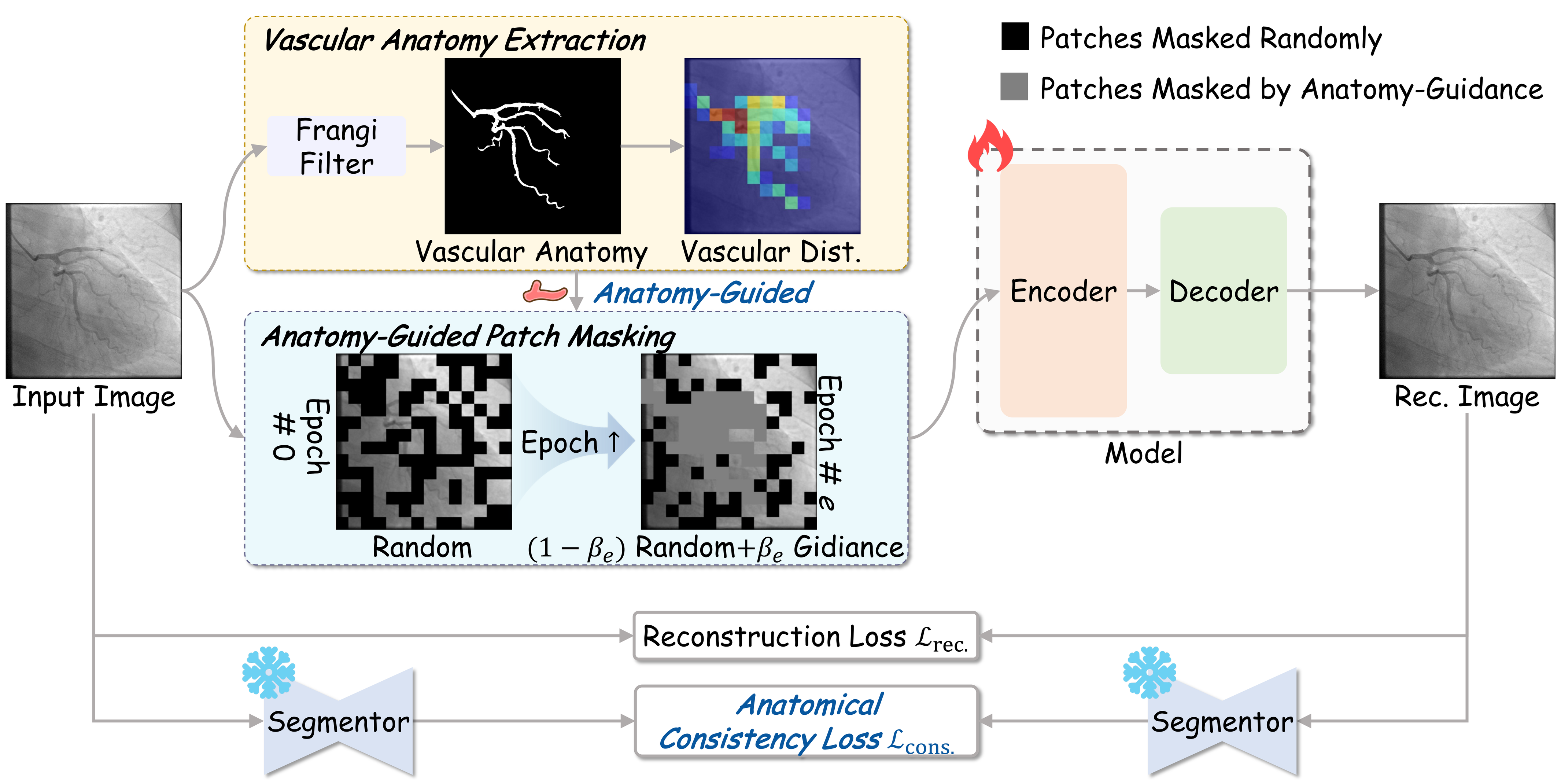
I. Motivation & Contributions

- **Background:** Vessel segmentation is critical for clinical applications. However, the prohibitive cost of obtaining annotations hinders fully-supervised approaches. Self-supervised pre-training (SSP) overcomes this bottleneck by exploiting large-scale unlabeled data, learning robust representations that enhance segmentation performance.
- **Challenge:** The potential of SSP for X-ray angiogram analysis has not been explored. The challenges stem from two aspects: domain-specific pre-training techniques and large-scale data. Existing masked image modeling (MIM) lacks explicit anatomical awareness, specifically within the masking strategies and reconstruction objectives.
- **Solution:** We propose VasoMIM, a vascular anatomy-aware framework designed to enhance vascular representation learning. To support this, we establish XA-20K, the largest-scale X-ray angiography dataset to date.



Conventional MIM Framework *vs.* Vascular Anatomy-aware MIM Framework (Ours)

II. Methods



- **Vascular Anatomy Extraction** ⇒ Frangi Filter (*MICCAI*, 1998)
- **Anatomy-Guided Masking Strategy** ⇒ Focuses on vascular regions.

Vascular Distribution $f(m_i) = \frac{\sum_{j=1}^{P^2} \mathbb{I}(m_{ij}=1)}{\sum_{i,j=1}^{N,P^2} \mathbb{I}(m_{ij}=1)}$

Weak2Strong Guidance $\beta_e = \beta_0 + \frac{e}{E}(\beta_E - \beta_0)$

- **Anatomical Consistency Loss** ⇒ Learns more discriminative vascular representations.

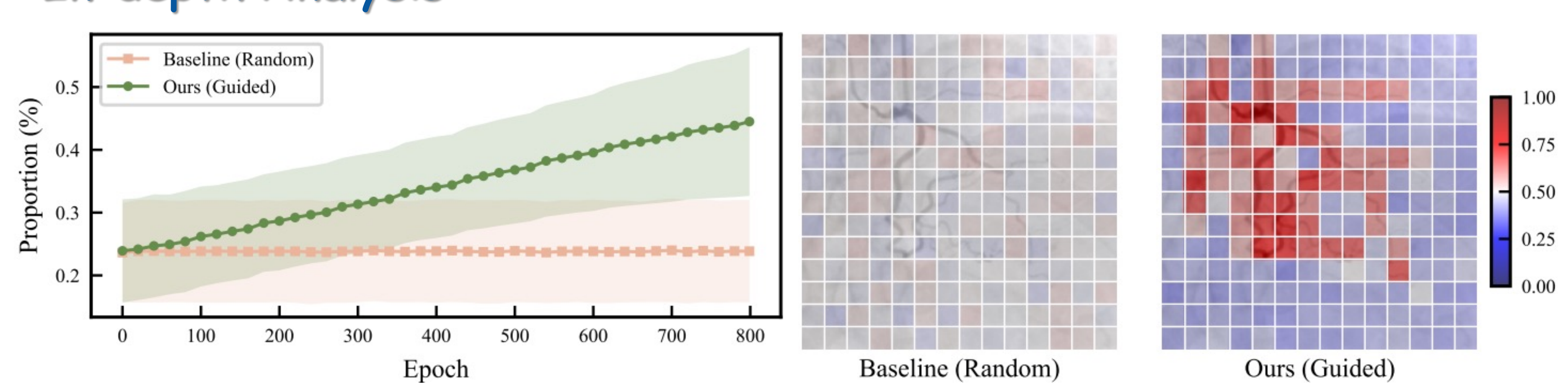
$$\mathcal{L}_{\text{cons.}} = \mathcal{L}(\mathcal{S}(I), \mathcal{S}(I'))$$

III. Main Results

Comparison with SOTAs on Three Benchmarks

Method	ARCADE DSC (%)	ARCADE clDice (%)	CAXF DSC (%)	CAXF clDice (%)	XCAV DSC (%)	XCAV clDice (%)
Traditional						
Frangi Filter (Frangi et al. 1998)	41.30	40.91	64.01	65.73	58.46	57.15
From Scratch						
U-Net (Ronneberger et al. 2015)	58.27$\pm$1.33	59.70$\pm$1.40	78.72$\pm$0.74	82.68$\pm$0.87	68.63$\pm$2.80	63.47$\pm$3.33
Contrastive Learning						
MoCo v3 (Chen, Xie, and He 2021)	60.99$\pm$0.30	62.68$\pm$0.18	77.76$\pm$0.51	80.91$\pm$0.31	70.85$\pm$0.34	63.97$\pm$0.71
DINO (Caron et al. 2021)	65.86$\pm$0.49	67.84$\pm$0.52	80.13$\pm$0.53	82.90$\pm$0.51	72.28$\pm$0.96	66.36$\pm$1.17
Masked Image Modeling						
MAE (He et al. 2022)	68.17$\pm$0.29	69.89$\pm$0.22	83.53$\pm$0.14	87.37$\pm$0.21	76.43$\pm$0.17	72.58$\pm$0.49
SimMIM (Xie et al. 2022)	66.92$\pm$0.43	68.93$\pm$0.71	82.24$\pm$0.34	85.77$\pm$0.17	75.10$\pm$0.36	69.98$\pm$0.42
AMT (Liu, Gui, and Luo 2023)	68.15$\pm$0.23	69.77$\pm$0.38	83.47$\pm$0.09	87.40$\pm$0.04	76.51$\pm$0.20	72.60$\pm$0.44
DeblurringMIM† (Kang et al. 2024)	<u>68.60</u>$\pm$0.44	70.21$\pm$0.37	<u>83.85</u>$\pm$0.09	<u>87.78</u>$\pm$0.20	<u>77.02</u>$\pm$0.08	<u>73.58</u>$\pm$0.19
CrossMAE (Fu et al. 2025)	62.40$\pm$0.33	64.23$\pm$0.27	80.07$\pm$0.13	83.45$\pm$0.19	72.25$\pm$0.24	65.94$\pm$0.15
HPM (Wang et al. 2025)	66.82$\pm$0.28	68.49$\pm$0.41	82.61$\pm$0.21	86.18$\pm$0.10	75.48$\pm$0.19	70.79$\pm$0.26
CheXWorld† (Yue et al. 2025)	67.95$\pm$0.26	<u>70.31</u>$\pm$0.48	80.64$\pm$0.31	82.65$\pm$0.31	73.74$\pm$0.24	67.13$\pm$0.32
VasoMIM	**68.85**$\pm$0.47	**70.56**$\pm$0.36	**84.49**$\pm$0.17	**88.33**$\pm$0.09	**77.52**$\pm$0.26	**74.18**$\pm$0.34

In-depth Analysis



Our strategy clearly favors masking patches rich in vascular anatomy.

Setting	SS ($\times 10^{-2}$) ↑	CHI ↑	DBI ↓
w/o $\mathcal{L}_{\text{cons.}}$	−4.19	17.11	25.32
w/ $\mathcal{L}_{\text{cons.}}$	**0.54**	**607.24**	**4.03**

SS: Silhouette Score; CHI: Calinski-Harabasz Index; DBI: Davies-Bouldin Index.

Superior clustering metrics.
Highly discriminative representations.

Ablations

Guidance	$\mathcal{L}_{\text{cons.}}$	ARCADE	CAXF
–	–	68.00	83.15
–	✓	68.45	84.03
✓	–	68.30	83.96
✓	✓	**68.85**	**84.49**

Different Components

Case	β_0	β_E	ARCADE	CAXF
Random	0	0	68.45	84.03
Weak-to-Strong	0	0.5	**68.85**	**84.49**
	0	1	68.52	84.24
	1	1	65.36	81.17
Strong-to-Weak	0.5	0	67.81	83.41

Different Guidance

IV. Conclusion & Future

- We propose a domain-specific pre-training framework (VasoMIM) and a dataset (XA-20k) to facilitate vessel segmentation.
- Future efforts will focus on curating larger-scale datasets and assessing VasoMIM's generalizability on diverse downstream applications, e.g., stenosis detection.