/TemplateVersion (2026.1)

www.xueshuxiangzi.com

VasoMIM：血管解剖感知掩码图像建模用于血管分割

De-Xing Huang[1,2], Xiao-Hu Zhou[1,2*], Mei-Jiang Gui[1,2], Xiao-Liang Xie[1,2], Shi-Qi Liu[1,2], Shuang-Yi Wang[1,2], Tian-Yu Xiang[1,2], Rui-Ze Ma[1,3], Nu-Fang Xiao[1,4], Zeng-Guang Hou[1,2*]

[1]State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences
[3] School of Artificial Intelligence, Taiyuan University of Technology
[4] School of Computer Science and Engineering, Hunan University of Science and Technology
Emails: { huangdexing2022, xiaohu.zhou, zengguang.hou } @ia.ac.cn

Abstract

在 X 射线血管造影中准确分割血管对于众多临床应用至关重要。然而，标注数据的缺乏构成了一个重大挑战，这推动了自监督学习（SSL）方法的采用，例如利用大规模无标签数据来学习可迁移表示的掩码图像建模（MIM）。不幸的是，传统的 MIM 往往由于血管和背景像素之间严重的类别不平衡而无法捕捉血管解剖结构，导致血管表征较弱。为了解决这一问题，我们引入了 Vascular anat_o_my-aware _M_asked _I_mage _M_odeling（VasoMIM），这是一个为 X 射线血管造影量身定制的新型 MIM 框架，该框架在预训练过程中显式整合了解剖学知识。具体而言，它包含两个互补组件：解剖引导掩蔽策略和解剖一致性损失。前者优先掩蔽含有血管的图块，使模型专注于重构血管相关区域。后者则在原始图像和重构图像之间强制保持血管语义的一致性，从而提高血管表示的辨别能力。实证结果表明，VasoMIM 在三个数据集上均达到了最新的性能。这些发现凸显了其促进 X 射线血管造影分析的潜力。

Project Page — https://dxhuang-casia.github.io/VasoMIM

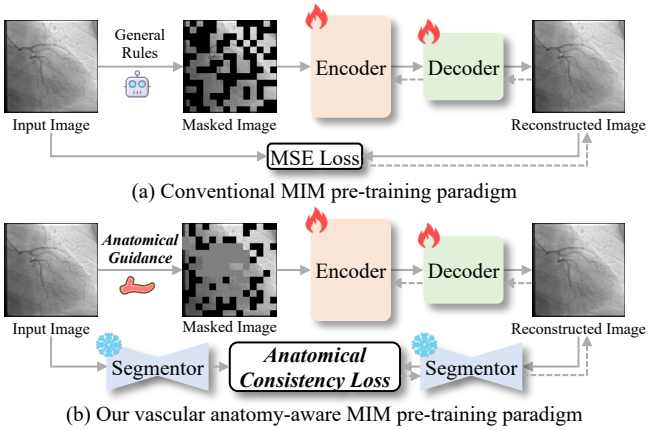


Figure 1: 传统 MIM 和 VasoMIM 的比较。(a) 传统 MIM 根据一般规则进行掩码处理，并通过最小化像素级损失来学习重建图块。(b) VasoMIM 通过血管解剖学来指导图块掩码，并在重建过程中强制解剖一致性，使模型能够学习更丰富的血管表示。深灰色图块是与血管相关的区域。

介绍

心血管疾病（CVDs）构成了全球健康危机，并且仍然是全球范围内死亡的主要原因（Vaduganathan et al. 2022）。X 光血管造影被认为是诊断 CVDs（Kheiri et al. 2022）、规划治疗（Members et al. 2022）和指导术中程序的金标准（Huang et al. 2025）。然而，由于低对比度、运动伪影以及重叠的解剖结构（Huang et al. 2024），放射科医生常常难以在 X 光血管造影图像中准确划定血管。因此，迫切需要自动的血管分割方法。

在过去的十年中，已经提出了许多血管分割算法（Ronneberger, Fischer, and Brox 2015; Huang et al. 2024; Wu et al. 2025），帮助减轻放射科医生的工作负担。然而，训练高性能模型需要具有像素级标签的大规模 X 光血管造影数据集，而生产这些注释仍然是劳动密集型、耗时的，并依赖于专业领域知识（Esteva et al. 2019）。自监督学习（SSL）通过从大量未标注数据中学习可泛化表示提供了一种吸引人的替代方案，从而提升下游性能（Gui et al. 2024）。特别是，掩码图像建模（MIM）（He et al. 2022; Fu et al. 2025; Zhuang et al. 2025b; Tang et al. 2025）在自然和医学图像分析中取得了显著的成功，其通过训练模型重建被掩盖的图像块，如图 1（a）所示。

然而，由于血管和背景像素之间极端的类别不平衡，适应 MIM 到 X 射线血管造影仍然具有挑战性。我们将这一困难归因于 MIM 框架中在两个关键方面缺乏明确的解剖意识。首先，在遮挡过程中，含有血管的图块比仅含背景的图块更容易被忽略。目前的遮挡策略通常基于一般规则，可以分为与数据无关和数据自适应的（Hinojosa, Liu, and Ghanem 2024）。前者包括随机遮挡（He et al. 2022）、块状遮挡（Bao et al. 2022）、均匀遮挡（Li et al. 2022）等。后者则涉及基于特定反馈设计遮挡策略，例如注意力图（Kakogeorgiou et al. 2022; Liu, Gui, and Luo 2023）、损失预测（Wang et al. 2025）、奖励函数（Xu et al. 2025）等。然而，这些方

*Corresponding Authors

www.xueshuxiangzi.com

法都不足以专注于血管解剖，导致很少含有血管的图块被遮挡，从而阻碍了血管表示的有效学习。其次，像素级重构损失在重构过程中未能保持语义一致性。大多数 MIM 方法最小化原始和重构图块之间的均方误差（MSE），但这种像素级目标忽略了血管解剖，因此未能鼓励区分性血管表示的学习。

为了解决这些挑战，我们引入了 Vas cular anat omy-aware M asked I mage M odeling (VasoMIM)，这是一种新的为 X 射线血管造影图量身定制的 MIM 范式，如图 1 (b) 所示。VasoMIM 的核心理念是将血管解剖知识注入 MIM。为了解决第一个挑战，我们引入了一种解剖引导的掩蔽策略，该策略偏重于包含血管的块，指导模型重建与解剖相关的区域。对于第二个挑战，我们提出了一种解剖一致性损失，强制原始和重建的血管造影图中的血管解剖保持一致，从而鼓励模型学习更多区分性的血管表示。这引发了如何从 X 射线血管造影图中提取血管解剖的问题。为了解决这个问题，我们应用了 Frangi 滤波器 (Frangi et al. 1998)，以无监督的方式获得血管分割掩码。为了将该滤波器整合到端到端 SSL 中，我们事先使用它产生的伪标签训练一个分割器。

总之，我们的主要贡献如下：

- 提出了一种新的血管解剖结构感知 MIM 框架，VasoMIM，以增强模型对 X 射线血管造影图中血管内容的理解能力。
- 提出了两个互补的组件：(1) 一种解剖引导的掩蔽策略，优先掩蔽包含血管的区域，引导模型关注血管区域；(2) 一种解剖一致性损失，确保原始图像和重建图像之间的语义一致性，从而提高血管表示的可辨识性。
- 广泛的实验证明了将解剖学知识集成到 MIM 框架中的好处。同时，VasoMIM 在血管分割任务中始终优于现有的 SSL 替代方案 [1]。

相关工作

医学影像中的自监督学习

自监督学习（SSL）(Chen et al. 2020; Chen, Xie, and He 2021; Caron et al. 2021; He et al. 2022; Wang et al. 2025) 提供了一种切实可行的解决方案，以缓解医学影像中标注稀缺的问题。现有的方法可以分为两种主要范式：基于对比学习（CL）和基于掩码图像建模（MIM）。CL 方法旨在在特征空间中将正样本对拉近，而将负样本对推远，比如 C 2 L (Zhou et al. 2020)、MICLe (Azizi et al. 2021)、VoCo (Wu, Zhuang, and Chen 2024)、RAD-DINO (Pérez-García et al. 2025) 等。然而，由于这些方法侧重于图像级表征，它们在分割 (Chaitanya et al. 2020) 等密集预测任务上可能表现不佳。MIM 方法训练模型以重建被掩盖的图像块，从而产生更细粒度的表征 (Yuan et al. 2023)。训练目标的选择至关重要。大多数方法使用像素级重建损失 (Kang et al. 2024; Li et al. 2024; Xu et al. 2025)，而有些则结合对比损失以改善 3D 医学图像中的多视图对齐 (Zhuang et al. 2025a,b)。然而，这些目标不能确保重建图像的语义一致性，常常导致血管表征较弱。我们的工作通过引入语义级解剖一致性损失来解决这一限制。

[1] VasoMIM 在图像去噪任务中也表现出卓越的性能（详见补充材料）。

在 MIM 中，由于图像的高冗余性，掩蔽策略的设计至关重要。大多数方法采用高比例的随机掩蔽（例如，75% ）。为了创造更具挑战性的预文本任务，许多精心设计的掩蔽策略被提出。AMT 根据注意力图进行掩蔽。SemMAE 首先学习图像的语义部分，并使用部分分割结果指导补丁掩蔽。HAP 利用人体结构的先验知识指导面向人类感知预训练的掩蔽过程。HPM、AnatoMask 和 AHM 等方法分别通过损失预测器、自蒸馏框架和策略网络来识别难以重建的补丁，并优先掩蔽这些困难的补丁。然而，这些方法都无法明确将血管解剖学纳入掩蔽过程中，而这是引导模型聚焦于血管区域的关键。

血管分割

传统的血管分割方法，比如 Frangi 滤波器 (Frangi et al. 1998)、活动轮廓 (Taghizadeh Dehkordi et al. 2014) 和图切割 (Wang et al. 2020)，依赖于手工设计的特征，这些特征在 X 射线血管造影图像广泛的变化性中表现不佳。深度学习的出现显著促进了这一领域的发展。早期基于 CNN 的架构如 U-Net (Ronneberger, Fischer, and Brox 2015) 及其变体 (Zhou et al. 2019; Li et al. 2020) 捕获了层次化特征，但在其局部感受野上受到限制。最近的基于 transformer 和 Mamba 的模型 (Chen et al. 2024; Hatamizadeh et al. 2021; Ruan, Li, and Xiang 2024; Wang et al. 2024) 通过整合全局上下文解决了这个限制，从而改善了复杂血管结构的描绘。然而，这些深度模型对数据需求极高，标注过的血管造影图像的稀缺性仍然限制了它们的进步 (Zhou et al. 2021)。在这项工作中，我们利用 MIM 从大规模无标注数据中学习可推广的特征表示，从而提高现有分割器的性能。

方法

整体框架如图 2 所示。首先，我们使用 Frangi 滤波器 (Frangi et al. 1998) 从 X 射线血管造影中提取血管解剖结构。接下来，利用该解剖学知识来指导掩模过程。最后，将解剖学一致性损失合并到训练目标中，以捕获具有辨别性的血管特征。

血管解剖结构提取

Frangi 滤波器通过突出类似血管的特征，以无监督的方式有效增强血管解剖结构。按照 (Wu et al. 2025) 中的实施，我们的方法分三个阶段进行。

多尺度 Hessian 血管特性。给定一个 X 射线血管造影图像 $I \in \mathbb{R}^{C\times H\times W}$，我们在尺度 $\sigma \in \{1,2,3,4\}$ 使用高斯核 $G(\sigma)$ 对其进行平滑，并计算多尺度 Hessian $H_\sigma(i) = \nabla^2 [I * G(\sigma)](i)$，其中 C、H 和 W 分别表示图像的通道数、高度和宽度。在每个像素 i，我们获得 $H_\sigma(i)$ 的特征值 $|\lambda_1| \le |\lambda_2|$，并定义血管特性响应

$$V_\sigma(i) = \begin{cases} |\lambda_1(i)|, & \lambda_2(i) < 0, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

最终的血管性图 $V(i)$ 是所有尺度上的最大响应，捕捉不同直径的血管。

自适应百分位阈值法。随后，我们对血管特征图 V 应用自适应阈值法，在其第 α 个百分位值 $T = \text{Percentile}_\alpha [V(i)]$ 处创建一个二值图 $\hat{B}(i) = \mathbb{I}[V(i) \ge T]$。这种方法有效地过滤掉了低强度噪声，同时保留了显著的血管解剖结构。

www.xueshuxiangzi.com

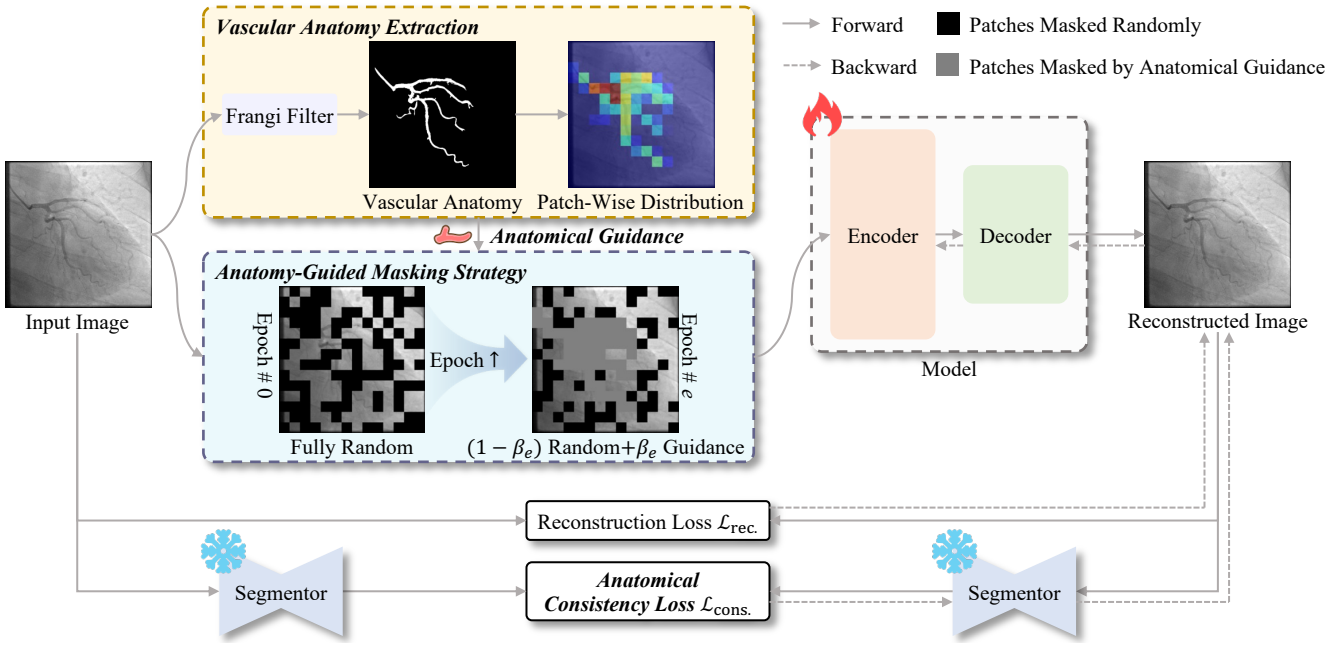


Figure 2: VasoMIM 的整体框架。在预训练期间，每个 X 光血管造影图首先通过 Frangi 滤波器处理，以提取其血管解剖结构。从这个解剖结构中，我们推导出分块的血管解剖分布 f 来指导遮罩过程。最后，通过最小化 $\mathcal{L}_{\text{train}}$ 来优化模型，$\mathcal{L}_{\text{train}}$ 是标准像素重建损失 $\mathcal{L}_{\text{rec.}}$ 和设计的解剖一致性损失 $\mathcal{L}_{\text{cons.}}$ 的结合。

种子区域生长。最后，基于最高的血管性强度值，自动识别出一个最优种子像素 $s = \arg\max_i V(i)$ 。然后将区域生长算法 (Adams and Bischof 1994) 应用于二值阈值图 $\hat{B}$ ，迭代地扩展区域以包含具有高血管性值的邻近像素。此过程产生一个二值掩码 $B \in \mathbb{R}^{1\times H\times W}$ ，可以准确地描绘血管解剖结构。

尽管阈值 α （默认情况下为 92 ）会影响提取的血管解剖结构的质量，但补充材料中提供的消融实验表明，VasoMIM 对 α 的变化具有鲁棒性。

与自然图像相比，X 光血管造影片显示出更大的空间冗余性，因为血管仅占每张图像的一小部分。现有的基于规则的一般遮盖策略缺乏任何解剖学意识，因此更频繁地遮盖仅包含背景的区域。结果，预训练主要基于背景内容，而不是学习血管表示。

为了解决这个问题，我们引入了一种解剖结构引导的掩蔽策略。我们的核心思想是包含血管的块更具信息量，应该以更高的概率被掩蔽。形式上，定义了一个块状血管解剖分布 f 。在预训练过程中，X 射线血管造影图 I 及其对应的掩膜 B 分别分割成不重叠的块 $x \in \mathbb{R}^{N\times\left(P^2C\right)}$ 和 $m \in \mathbb{R}^{N\times P^2}$ ，其中 P 是块的大小，$N = HW/P^2$ 是序列长度。让 $m_i \in \{0,1\}^{P^2}$ 为 m 的第 i 个块。然后，$f(m_i)$ 定义如下：

$$f\left(m_i\right) = \frac{\sum_{j=1}^{P^2} \mathbb{I}\left(m_{ij}=1\right)}{\sum_{i,j=1}^{N,P^2} \mathbb{I}\left(m_{ij}=1\right)}, \quad i = 1, 2, \cdots, N \qquad (2)$$

，其中 $\mathbb{I}(\cdot)$ 是指示函数。

弱到强的解剖引导。通过利用局部的血管解剖分布，我们执行解剖感知的局部掩码。然而，在训练的早期阶段，掩盖太多含有血管的局部区域可能会削弱模型重建那些富含血管内容的局部区域的能力。为此，我们采用了一种从弱到强的解剖引导策略。如图 2 所示，对于特定的训练纪元 e ，掩盖区域中的 β_e 是根据 f 采样的，其余的 $1-\beta_e$ 是随机选择的。在预训练期间，β_e 线性增加。其中，E 是最大预训练纪元，而 $\beta_0, \beta_E \in [0,1]$ 是超参数。在这种掩码策略下，$\beta_e \gamma N$ 的局部区域通过解剖引导掩盖，其余的 $(1-\beta_e)\gamma N$ 局部区域是随机掩盖的，其中 γ 表示掩盖率。

这种策略确保模型最初遇到的是血管和背景的均衡混合，然后逐步集中于血管区域，从而提高其学习血管表示的能力。

在传统的 MIM 中，重建目标通常最小化被遮盖和重建的图块之间的均方误差（MSE）。然而，这种损失未能保持语义一致性，也没有促使模型学习对于下游任务至关重要的区分性血管表示。为了克服这一限制，我们引入了一个解剖一致性损失，它明确指导模型在重建过程中保留血管的解剖结构：

其中，I' 是重建的 X 射线血管造影图像。$\mathcal{L}(\cdot,\cdot)$ 是一个抽象的度量函数，默认使用交叉熵损失。这里，$\mathcal{S}(\cdot)$ 代表用于提取血管解剖结构的分割器。尽管 Frangi 滤波器提供了高质量的血管掩码，但其不可微分性阻止了直接集成到端到端的预训练过程中。为了解决这个问题，我们在 Frangi 滤波器生成的伪标签上训练了一个轻量级的 UNeXt-S 网络（ ~ 0.3 M）(Valanarasu and Patel 2022) ，然后在后续的预训练过程中冻结其权重。

训练目标。除了解剖一致性损失外，我们还包括了像素级重建损失（即，MSE 损失），遵循传统的 MIM 方法。整体训练损失由以下公式给出：

$$\mathcal{L}_{\text{train}} = \mathcal{L}_{\text{rec.}} + \mathcal{L}_{\text{cons.}} \qquad (3)$$

www.xueshuxiangzi.com

结果

数据集

表 ?? 汇总了用于预训练和血管分割的数据集。更多详细描述可以在补充材料中找到。

预训练。我们从五个公开数据集中收集了一个含有 20,650 冠状动脉 X 光血管造影图的语料库：ARCADE (Popov et al. 2024)、CADICA (Jiménez-Partinen et al. 2024)、Stenosis (Danilov et al. 2021)、SYNTAX (Mahmoudi et al. 2025) 和 XCAD (Ma et al. 2021)。

血管分割。VasoMIM 在三个基准上进行评估，包括一个域内数据集 ARCADE 和两个域外数据集 CAXF (Li et al. 2020) 和 XCAV (Wu et al. 2025)。注意，用于血管分割的 ARCADE 图像不包含在预训练集中。

实现细节

预训练。我们的实现基于 MAE (He et al. 2022)。默认情况下，ViT-B/16 (Dosovitskiy et al. 2021) 被用作骨干网络，遵循之前的工作。VasoMIM 使用 AdamW 优化器 (Loshchilov and Hutter 2019)，批量大小为 256，输入分辨率为 224×224，在 NVIDIA A6000 GPU 上进行 800 轮次的预训练。

血管分割。我们采用 U-Net (Ronneberger, Fischer, and Brox 2015) 作为分割解码器，并在 ARCADE、CAXF 和 XCAV 上进行端到端微调，每个数据集训练 500 个 epoch，使用调整为 224×224 的输入图像。优化使用 AdamW 进行，初始学习率为 $1e^{-4}$，权重衰减为 0.05。采用具有 $T\_max = 500$ 的余弦退火调度。所有实验均在 NVIDIA A6000 GPU 上进行。

评估指标。采纳 Dice 相似系数 (DSC) 和中心线 Dice (clDice) (Shit et al. 2021) 来评估血管分割的性能。与 DSC 相比，clDice 通过测量预测和真实血管中心线之间的重叠更好地捕捉拓扑正确性。

我们将 VasoMIM 与基于 CL 和基于 MIM 的 SSL 方法进行比较。所有的基线模型使用其官方实现，在表 ?? 中的血管造影数据集上进行预训练。每个模型使用三个随机种子进行微调，我们将所有指标报告为“mean±std”。

域内数据集 (ARCADE)。当从零开始训练时，U-Net 在 ARCADE 上达到 58.27% 的 DSC 和 59.70% 的 clDice。通过在大规模无标签数据上进行预训练，我们的 VasoMIM 将 DSC 性能提升了 10.58%，将 clDice 性能提升了 10.86%，分别达到 68.85% 的 DSC 和 70.56% 的 clDice，明显超过了领先的 SSL 基线。与 Frangi 滤波器相比，VasoMIM 在 DSC 上获得了 27.55% 的绝对增益，在 clDice 上获得了 29.65% 的绝对增益，强调了我们具解剖意识预训练在提高分割性能中的重要作用。

域外数据集 (CAXF 和 XCAV)。我们进一步在域外数据集上进行实验。我们的 VasoMIM 一直表现最佳，分别在 CAXF 和 XCAV 上达到 84.49% 和 77.52% 的 DSC。值得注意的是，它在 CAXF 上比最佳基线高出 +0.64% 的 DSC，在 XCAV 上高出 +0.50% 的 DSC。这个性能差距甚至比在域内数据集 (即 +0.25% DSC) 上还要大，突显出 VasoMIM 的强泛化能力和鲁棒性。

定性结果。从三个数据集中选择的几个具有代表性的案例展示在图 3 中。VasoMIM 能够生成精确的血管分割结果，即使是非常细或模糊的血管也不例外。由于篇幅限制，更多的定性结果在补充材料中提供。

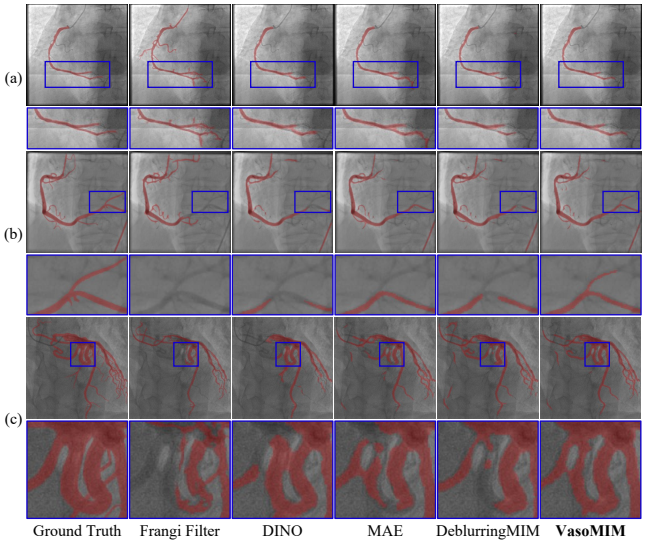


Figure 3: 针对 (a) ARCADE，(b) CAXF 和 (c) XCAV 的定性结果。细节在蓝色框内放大显示。

对 ARCADE 和 CAXF 进行了消融研究，默认设置在 orange 中突出显示。每个模型使用三个随机种子进行微调。表 2 显示了逐步将每个提议的组件添加到基础模型（即 MAE（$\gamma = 0.5$））后的结果。

解剖引导掩膜策略的有效性。我们观察到，仅仅添加解剖引导的掩膜策略就能显著提升性能（在 ARCADE 上为 +0.30% DSC，在 CAXF 上为 +0.81% DSC）。为了更好地理解这种改善，我们测量了预训练过程中被掩膜的图像块中含有血管的比例。如图 4 (a) 所示，我们的策略掩膜了更多含有血管的图像块，而基线方法则几乎固定（且低）比例地掩膜含有血管的图像块，无论图像内容如何。我们进一步在图 4 (b) 中展示了一个例子，其中每个图像块在预训练期间的掩膜比率以颜色表示。我们的策略明显倾向于掩膜富含血管解剖结构的图像块，而基线方法则没有显示这样的偏好。

解剖一致性损失的作用。在基线中添加解剖一致性损失 $\mathcal{L}_{\text{cons.}}$ 也能带来显著的提升（例如，在 CAXF 上 83.15% → 84.03% DSC）。这一改进表明 $\mathcal{L}_{\text{cons.}}$ 使模型能够学习到更多有区分性的血管表示。为了进一步验证 $\mathcal{L}_{\text{cons.}}$ 的影响，我们首先将补丁分为含有血管的组和仅包含背景的组，然后使用 UMAP (Healy and McInnes 2024) 将使用和不使用 $\mathcal{L}_{\text{cons.}}$ 预训练的模型的表示投射到低维空间，并评估这些表示的聚类效果。如表 3 所示，使用 $\mathcal{L}_{\text{cons.}}$ 预训练的模型比不使用 $\mathcal{L}_{\text{cons.}}$ 预训练的模型实现了更高的 SS 和 CHI，以及更低的 DBI，这表明血管表示的聚类更紧凑且分离更好。

深入分析

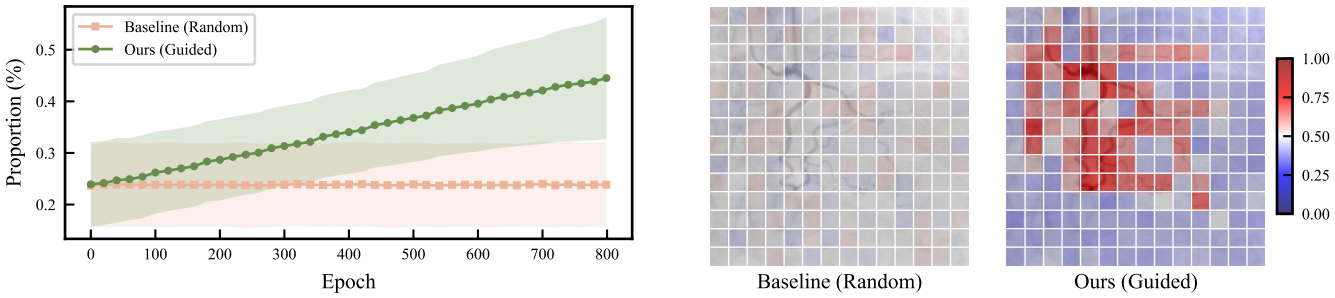
接下来我们对提出的 VasoMIM 进行详细分析。默认设置在 orange 中突出显示。所有结果在三个随机种子上取平均。

遮罩率。如图 5 所示，我们观察到一个中等遮罩率（即 0.5）能带来更好的性能。这个结果与之前的工作有些不同，例如，MAE 采用了更高的比例 $\gamma = 0.75$。我们假设这种差异源于 X 光血管造影图像的独特特征。与包含不同尺寸物体的自然图像不同，X 光血管造影中的

www.xueshuxiangzi.com

Method	ARCADE		CAXF		XCAV	
	DSC (%)	clDice (%)	DSC (%)	clDice (%)	DSC (%)	clDice (%)
Traditional						
Frangi Filter (Frangi et al. 1998) [MICCAI '98]	41.30	40.91	64.01	65.73	58.46	57.15
From Scratch						
U-Net (Ronneberger, Fischer, and Brox 2015) [MICCAI '15]	58.27 ±1.33	59.70 ±1.40	78.72 ±0.74	82.68 ±0.87	68.63 ±2.80	63.47 ±3.33
Contrastive Learning						
MoCo v3 (Chen, Xie, and He 2021) [ICCV '21]	60.99 ±0.30	62.68 ±0.18	77.76 ±0.51	80.91 ±0.31	70.85 ±0.34	63.97 ±0.71
DINO (Caron et al. 2021) [ICCV '21]	65.86 ±0.49	67.84 ±0.52	80.13 ±0.53	82.90 ±0.51	72.28 ±0.96	66.36 ±1.17
Masked Image Modeling						
MAE (He et al. 2022) [CVPR '22]	68.17 ±0.29	69.89 ±0.22	83.53 ±0.14	87.37 ±0.21	76.43 ±0.17	72.58 ±0.49
SimMIM (Xie et al. 2022) [CVPR '22]	66.92 ±0.43	68.93 ±0.71	82.24 ±0.34	85.77 ±0.17	75.10 ±0.36	69.98 ±0.42
AMT (Liu, Gui, and Luo 2023) [AAAI '23]	68.15 ±0.23	69.77 ±0.38	83.47 ±0.09	87.40 ±0.04	76.51 ±0.20	72.60 ±0.44
DeblurringMIM † (Kang et al. 2024) [MedIA '24]	68.60	70.21 ±0.37	83.85	87.78	77.02	73.58
CrossMAE (Fu et al. 2025) [TMLR '25]	62.40 ±0.33	64.23 ±0.27	80.07 ±0.13	83.45 ±0.19	72.25 ±0.24	65.94 ±0.15
HPM (Wang et al. 2025) [TPAMI '25]	66.82 ±0.28	68.49 ±0.41	82.61 ±0.21	86.18 ±0.10	75.48 ±0.19	70.79 ±0.26
CheXWorld † (Yue et al. 2025) [CVPR '25]	67.95 ±0.26	70.31	80.64 ±0.31	82.65 ±0.31	73.74 ±0.24	67.13 ±0.32
VasoMIM [Ours]	68.85	70.56	84.49	88.33	77.52	74.18

Table 1: 对 ARCADE、CAXF 和 XCAV 的当前最佳方法进行比较。所有方法均使用其官方代码进行重新实现。最好的和第二好的结果分别用 red 和 yellow 进行标记。结果以“三个随机种子的“ mean ± std ”形式”报告，Frangi 滤波器除外。† 表示模型专注于医学影像。



(a) (b)

Figure 4: 一些解剖学引导遮盖策略的证据。(a) 预训练期间被遮盖的图块中包含血管的图块比例。(b) 预训练过程中图块的遮盖比率，即 $\frac{1}{E}\sum_{j=1}^{E}\mathbb{I}(\text{Patch } x\_i \text{ is masked in epoch } j)$ 。

Guidance	$\mathcal{L}_{\text{cons.}}$	ARCADE		CAXF	
		DSC (%)	clDice (%)	DSC (%)	clDice (%)
–	–	68.00	69.78	83.15	86.41
–	✓	68.45 ↑ 0.45	70.18 ↑ 0.40	84.03 ↑ 0.88	87.71 ↑ 1.30
✓	–	68.30 ↑ 0.30	69.94 ↑ 0.16	83.96 ↑ 0.81	87.61 ↑ 1.20
✓	✓	**68.85 ↑ 0.85**	**70.56 ↑ 0.78**	**84.49 ↑ 1.34**	**88.33 ↑ 1.92**

Table 2: 提出的解剖学引导掩蔽策略和解剖一致性损失的消融研究。

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.

Table 3: XCAD 上的聚类指标结果。

血管仅占图像的一小部分。使用较大的遮罩率容易遮住仅含有很少信息内容的背景补丁。这可能限制模型学习有用的血管表征的能力。

不同的遮蔽策略。我们评估了几种遮蔽策略，其结果汇总在表 4 中。通过改变 β_0 和 β_E 构建了不同的策略。例如，设定 $\beta_0 = \beta_E = 0$ 会产生一个完全随机的遮蔽策略，而 $\beta_0 = \beta_E = 1$ 表示所有 γN 补丁都纯粹基于解剖学指导进行遮蔽。有趣的是，简单地增加解剖学指导的程度并不总是能带来更好的性能。显然，在遮蔽过程中保持一定程度的随机性是必不可少的。具体来说，配置 $\beta_0 = 0$ ，$\beta_E = 0.5$ 达到了最佳效果，平衡了更强的解剖学指导（$\beta_0 = \beta_E = 1$ ）和更大的随机性（$\beta_0 = \beta_E = 0$ ）。这个结果相当直观。激进地遮蔽具有最高解剖学相关性的补丁导致大多数含有血管的补丁被遮蔽。在这种情况下，模型被迫仅从几乎没有语义提示的背景补丁中重建血管解剖结构。

从强到弱的解剖指导。我们进一步研究了一种反向方法，在预训练期间采用了一种从强到弱的策略。如表 5

www.xueshuxiangzi.com

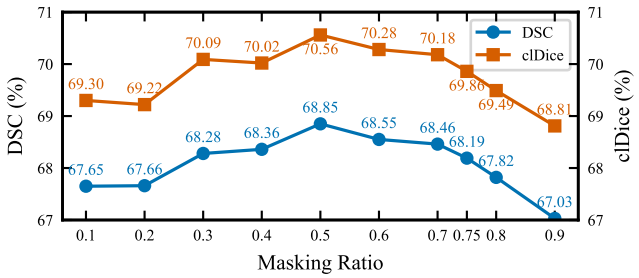


Figure 5: 对 ARCADE 上掩码比例 γ 的深入分析。在我们的默认设置中，γ 被设定为 0.5 。

Case	Guidance	Randomness	β_0	β_E	DSC (%)	
					ARCADE	CAXF
Random	weak	strong	0	0	68.45	84.03
Anatomy-Guided	↓	↓	0	0.5	**68.85** ↑ **0.40**	**84.49** ↑ **0.46**
			0	1	68.52 ↑ 0.07	84.24 ↑ 0.21
	strong	weak	1	1	65.36 ↓ 3.09	81.17 ↓ 2.86

Table 4: 不同掩码策略的影响。随着 β_E 的增加，解剖指导变得更强，而随机性减小。

Case	Manner	β_0	β_E	DSC (%)	
				ARCADE	CAXF
Random	-	0	0	68.45	84.03
Anatomy-Guided	Weak-to-Strong	0	0.5	**68.85** ↑ **0.40**	**84.49** ↑ **0.46**
	Strong-to-Weak	0.5	0	67.81 ↓ 0.64	83.41 ↓ 0.62

Table 5: 弱到强与强到弱解剖引导的比较。

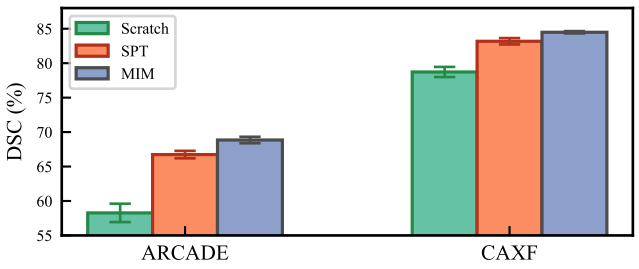


Figure 6: SPT 与 MIM 相比。我们的 MIM（即 VasoMIM）在 ARCADE 和 CAXF 上分别提供平均 +2.11% 和 +1.31% 的 DSC 提升。

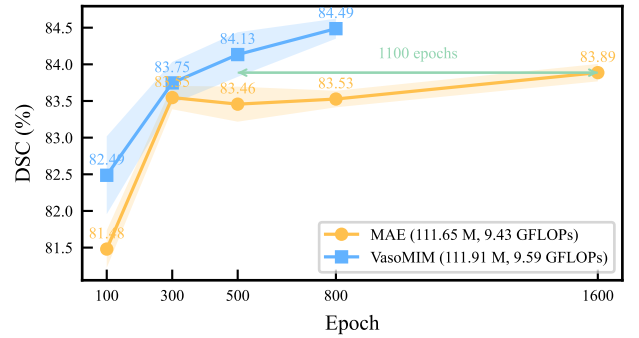


Figure 7: 在 CAXF 上进行不同训练轮次预训练的模型进行微调。GFLOPs 在 NVIDIA A6000 GPU 上评估，使用单个 224×224 RGB 图像（带掩码）作为输入。

Method	Ratio (%)		
	25	50	100
Scratch [MICCAI '15]	38.09	50.67	58.27
MAE [CVPR '22]	56.58	63.04	68.17
VasoMIM [Ours]	**56.93**	**63.64**	**68.85**

Table 6: 在 ARCADE 上用 25% 、50% 和 100% 训练数据进行微调。DSC 在此表中报告。

所示，这种方法导致了明显的性能下降，甚至与使用随机掩蔽策略相比。这一发现强调了我们从弱到强掩蔽设计的重要性。

基于分割器的预训练与 MIM。我们使用相同的分割器（即 U-Net）比较了基于分割器的预训练（SPT）方法和我们的 MIM。在 SPT 中，分割器通过在预训练数据集上由 Frangi 滤波器生成的伪标签以全监督的方式进行预训练，然后在下游数据集上进行微调。图 6 显示，两种预训练范式都显著优于从头开始训练的模型。此外，MIM 进一步将 ARCADE 上的 DSC 从 66.74% 提升到 68.85% （+2.11% ），并将 CAXF 上的 DSC 从 83.18% 提升到 84.49% （+1.31% ）。总体而言，这些结果表明我们的 MIM 学习到了比 SPT 更加稳健和可迁移的表示。

预训练和数据效率。我们发现，在资源有限的情况下，VasoMIM 的优势更加明显，无论是更少的预训练周期还是用于微调的带注释数据较少。首先，我们微调已经在 100 、300 、500 和 800 周期内预训练的 MAE 和 VasoMIM 模型，然后在 CAXF 上测试。如图 7 所示，VasoMIM 在每个预训练时长上均优于 MAE。值得注意的是，在仅经过 300 轮预训练（$\sim$ 6 小时）后，VasoMIM 在 CAXF 上达到 83.75% 的 DSC，略高于 MAE 的 83.53% ，而 MAE 需要 800 周期（$\sim$ 12 小时）预训练。即使我们将 MAE 的预训练扩展到 1,600 周期（$\sim$ 24 小时）以获得进一步的性能提升，VasoMIM 在仅 500 周期（$\sim$ 10 小时）预训练的情况下仍表现更好。接下来，为测试数据效率，我们使用训练标签的 25% 、50% 或 100% 来微调在 ARCADE 上的模型。表 6 显示 VasoMIM 在所有情况下均优于 MAE。

在本文中，我们提出了 VasoMIM，这是一种针对 X 射线血管造影图像专门设计的血管解剖结构感知的 MIM 框架。通过将解剖学指导整合到遮蔽策略中，我们的模型能够关注于与血管相关的区域。此外，所提出的解剖一致性损失显著提高了学习到的血管表示的可辨别性。大量实验表明，所提出的框架在三个基准上表现出优越的性能。这项工作为将解剖知识整合到 SSL 框架中提供了新的视角。

References

Adams, R.; and Bischof, L. 1994. Seeded region growing. IEEE TPAMI, 16(6): 641–647.

Azizi, S.; et al. 2021. Big self-supervised models advance medical image classification. In Proc. CVPR, 3478–3488.

Bao, H.; Dong, L.; Piao, S.; and Wei, F. 2022. BEIT:

www.xueshuxiangzi.com

BERT pre-training of image transformers. In Proc. ICLR.

Buades, A.; Coll, B.; and Morel, J.-M. 2005. A non-local algorithm for image denoising. In Proc. CVPR, volume 2, 60–65.

Caron, M.; et al. 2021. Emerging properties in self-supervised vision transformers. In Proc. ICCV, 9650–9660.

Chaitanya, K.; Erdil, E.; Karani, N.; and Konukoglu, E. 2020. Contrastive learning of global and local features for medical image segmentation with limited annotations. In Proc. NeurIPS, volume 33, 12546–12558.

Chen, J.; et al. 2024. TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers. MedIA, 97: 103280.

Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. In Proc. ICML, 1597–1607.

Chen, X.; Xie, S.; and He, K. 2021. An empirical study of training self-supervised vision transformers. In Proc. CVPR, 9640–9649.

Danilov, V. V.; et al. 2021. Real-time coronary artery stenosis detection based on modern neural networks. Sci. Rep., 11(1): 7582.

Dosovitskiy, A.; et al. 2021. An image is worth 16x16 words: Transformers for image recognition at scale. In Proc. ICLR.

Esteva, A.; et al. 2019. A guide to deep learning in healthcare. Nat. Med., 25(1): 24–29.

Frangi, A. F.; Niessen, W. J.; Vincken, K. L.; and Viergever, M. A. 1998. Multiscale vessel enhancement filtering. In Proc. MICCAI, 130–137.

Fu, L.; et al. 2025. Rethinking patch dependence for masked autoencoders. TMLR.

Goyal, P.; et al. 2017. Accurate, large minibatch SGD: Training ImageNet in 1 hour. arXiv.

Gui, J.; et al. 2024. A survey on self-supervised learning: Algorithms, applications, and future trends. IEEE TPAMI, 46(12): 9052–9071.

Harbron, R.; et al. 2016. Patient radiation doses in paediatric interventional cardiology procedures: A review. J. Radiol. Prot., 36(4): R131.

Hatamizadeh, A.; Nath, V.; Tang, Y.; Yang, D.; Roth, H. R.; and Xu, D. 2021. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In Proc. MICCAIW, 272–284.

He, K.; Chen, X.; Xie, S.; Li, Y.; Dollár, P.; and Girshick, R. 2022. Masked autoencoders are scalable vision learners. In Proc. CVPR, 16000–16009.

Healy, J.; and McInnes, L. 2024. Uniform manifold approximation and projection. Nat. Rev. Methods Primers, 4(1): 82.

Hinojosa, C.; Liu, S.; and Ghanem, B. 2024. ColorMAE: Exploring data-independent masking strategies in masked autoencoders. In Proc. ECCV, 432–449.

Huang, D.-X.; et al. 2024. SPIRONet: Spatial-frequency learning and topological channel interaction network for vessel segmentation. arXiv.

Huang, D.-X.; et al. 2025. Real-time 2D/3D registration via CNN regression and centroid alignment. IEEE TASE, 22: 85–98.

Jiménez-Partinen, A.; et al. 2024. CADICA: A new dataset for coronary artery disease detection by using invasive coronary angiography. Expert Syst., 41(12): e13708.

Kakogeorgiou, I.; et al. 2022. What to hide from your students: Attention-guided masked image modeling. In Proc. ECCV, 300–318.

Kang, Q.; et al. 2024. Deblurring masked image modeling for ultrasound image analysis. MedIA, 97: 103256.

Kheiri, B.; Simpson, T. F.; Osman, M.; German, D. M.; Fuss, C. S.; and Ferencik, M. 2022. Computed tomography vs invasive coronary angiography in patients with suspected coronary artery disease: A meta-analysis. JACC: Cardiovasc. Imaging, 15(12): 2147–2149.

Kruzhilov, I.; et al. 2025. CoronaryDominance: Angiogram dataset for coronary dominance classification. Sci. Data, 12(1): 341.

Li, R.-Q.; Bian, G.-B.; Zhou, X.-H.; Xie, X.; Ni, Z.-L.; and Hou, Z. 2020. CAU-net: A novel convolutional neural network for coronary artery segmentation in digital substraction angiography. In Proc. ICONIP, 185–196.

Li, X.; Wang, W.; Yang, L.; and Yang, J. 2022. Uniform masking: Enabling MAE pre-training for pyramid-based vision transformers with locality. arXiv.

Li, Y.; Luan, T.; Wu, Y.; Pan, S.; Chen, Y.; and Yang, X. 2024. AnatoMask: Enhancing medical image segmentation with reconstruction-guided self-masking. In Proc. ECCV, 146–163.

Liu, Z.; Gui, J.; and Luo, H. 2023. Good helper is around you: Attention-driven masked image modeling. In Proc. AAAI, volume 37, 1799–1807.

Loshchilov, I.; and Hutter, F. 2019. Decoupled weight decay regularization. In Proc. ICLR.

Luo, Y.; Majoe, S.; Kui, J.; Qi, H.; Pushparajah, K.; and Rhode, K. 2020. Ultra-dense denoising network: Application to cardiac catheter-based X-ray procedures. IEEE TBME, 68(9): 2626–2636.

Ma, Y.; et al. 2021. Self-supervised vessel segmentation via adversarial learning. In Proc. ICCV, 7536–7545.

Mahmoudi, S. S.; et al. 2025. X-ray coronary angiogram images and SYNTAX score to develop machine-learning algorithms for CHD diagnosis. Sci. Data, 12(1): 471.

Members, W. C.; et al. 2022. 2021 ACC/AHA/SCAI guideline for coronary artery revascularization: A report of the American College of Cardiology/American Heart Association Joint Committee on clinical practice guidelines. JACC, 79(2): e21–e129.

Pérez-García, F.; et al. 2025. Exploring scalable medical image encoders beyond text supervision. Nat. Mach. Intell., 7: 119–130.

www.xueshuxiangzi.com

Popov, M.; et al. 2024. Dataset for automatic region-based coronary artery disease diagnostics using X-ray angiography images. Sci. Data, 11(1): 20.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-Net: Convolutional networks for biomedical image segmentation. In Proc. MICCAI, 234–241.

Ruan, J.; Li, J.; and Xiang, S. 2024. VM-UNet: Vision Mamba UNet for medical image segmentation. arXiv.

Shit, S.; et al. 2021. clDice: A novel topology-preserving loss function for tubular structure segmentation. In Proc. CVPR, 16560–16569.

Taghizadeh Dehkordi, M.; Doost Hoseini, A. M.; Sadri, S.; and Soltanianzadeh, H. 2014. Local feature fitting active contour for segmenting vessels in angiograms. IET CV, 8(3): 161–170.

Tang, F.; et al. 2025. MambaMIM: Pre-training Mamba with state space token interpolation and its application to medical image segmentation. MedIA, 103606.

Vaduganathan, M.; Mensah, G. A.; Turco, J. V.; Fuster, V.; and Roth, G. A. 2022. The global burden of cardiovascular diseases and risk: A compass for future health. JACC, 80(25): 2361–2371.

Valanarasu, J. M. J.; and Patel, V. M. 2022. UNeXt: MLP-based rapid medical image segmentation network. In Proc. MICCAI, 23–33.

Wang, C.; et al. 2020. Tensor-cut: A tensor-based graph-cut blood vessel segmentation method and its application to renal artery segmentation. MedIA, 60: 101623.

Wang, H.; et al. 2025. Bootstrap masked visual modeling via hard patch mining. IEEE TPAMI. DOI:10.1109/TPAMI.2025.3557001.

Wang, J.; Chen, J.; Chen, D.; and Wu, J. 2024. LKM-UNet: Large kernel vision Mamba UNet for medical image segmentation. In Proc. MICCAI, 360–370.

Wu, C.-H.; et al. 2025. DeNVeR: Deformable neural vessel representations for unsupervised video vessel segmentation. In Proc. CVPR, 15682–15692.

Wu, L.; Zhuang, J.; and Chen, H. 2024. VoCo: A simple-yet-effective volume contrastive learning framework for 3D medical image analysis. In Proc. CVPR, 22873–22882.

Xie, Z.; et al. 2022. SimMIM: A simple framework for masked image modeling. In Proc. CVPR, 9653–9663.

Xu, Z.; Liu, Y.; Xu, G.; and Lukasiewicz, T. 2025. Self-supervised medical image segmentation using deep reinforced adaptive masking. IEEE TMI, 44(1): 180–193.

Yuan, J.; et al. 2023. HAP: Structure-aware masked image modeling for human-centric perception. In Proc. NeurIPS, volume 36, 50597–50616.

Yue, Y.; Wang, Y.; Tao, C.; Liu, P.; Song, S.; and Huang, G. 2025. CheXWorld: Exploring image world modeling for radiograph representation learning. In Proc. CVPR, 20778–20788.

Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; and Zhang, L. 2017. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE TIP, 26(7): 3142–3155.

Zhou, H.-Y.; Yu, S.; Bian, C.; Hu, Y.; Ma, K.; and Zheng, Y. 2020. Comparing to learn: Surpassing imagenet pretraining on radiographs by comparing image representations. In Proc. MICCAI, 398–407.

Zhou, S. K.; et al. 2021. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proc. IEEE, 109(5): 820–838.

Zhou, Z.; Siddiquee, M. M. R.; Tajbakhsh, N.; and Liang, J. 2019. UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE TMI, 39(6): 1856–1867.

Zhuang, J.; Luo, L.; Wang, Q.; Wu, M.; Luo, L.; and Chen, H. 2025a. Advancing volumetric medical image segmentation via global-local masked autoencoders. IEEE TMI. DOI:10.1109/TMI.2025.3569782.

Zhuang, J.; et al. 2025b. MiM: Mask in mask self-supervised pre-training for 3D medical image analysis. IEEE TMI. DOI:10.1109/TMI.2025.3564382.

www.xueshuxiangzi.com

在本补充材料中，我们首先详细介绍了用于预训练、血管分割和图像去噪的数据集。接下来，我们介绍更多的实施细节。最后，我们提供了更多的定量和定性结果。

数据集描述

预训练。(1) ARCADE (Popov et al. 2024)：该数据集包括冠状动脉血管分割和动脉粥样硬化斑块定位的两个子集。每个子集包含 1,500 张图像，分为 1,000 张用于训练，200 张用于验证，300 张用于测试。每个子集仅使用 1,000 张训练图像进行预训练。(2) CADICA (Jiménez-Partinen et al. 2024)：它包括来自 42 名患者的冠状动脉 X 射线血管造影视频，帧尺寸为 512×512 像素，视频长度从 1 帧到 151 帧不等。我们从这些视频中选择 6,594 帧高质量图像进行预训练。(3) Stenosis (Danilov et al. 2021)：它提供来自 100 名患者的 X 射线血管造影视频，总帧数为 8,325 。我们使用来自 70 名患者的 7,492 个训练帧进行预训练。(4) SYNTAX (Mahmoudi et al. 2025)：该数据集包括来自 231 名患者的 2,943 张 X 射线血管造影片。所有图像均用于预训练。(5) XCAD (Ma et al. 2021)：该数据集提供 1,747 张 X 射线血管造影片，其中 1,621 张用于预训练。

血管分割。(1) ARCADE：我们在冠状动脉血管分割子集的验证集（200 张图像）上微调模型，并在其测试集（300 张图像）上报告性能。(2) XCAV (Wu et al. 2025)：该数据集由来自 59 名患者的 111 个血管造影视频组成。我们提取具有像素级血管注释的帧，并按患者将它们分为 175 个训练帧（47 名患者）和 46 个测试帧（12 名患者）。(3) CAXF (Li et al. 2020)：它由 36 个 X 光血管造影视频中的 538 张图像组成。根据原始协议 (Li et al. 2020)，从 24 个视频中使用 337 帧进行训练，剩下的来自 12 个视频的 201 张图像用于测试。

图像去噪。X 射线血管造影系统通常使用低剂量协议以减少患者的辐射暴露 (Harbron et al. 2016)，这引入了大量噪声。图像去噪算法可以通过抑制这些噪声来明确血管解剖结构 (Luo et al. 2020) 。由于没有公开数据集存在，我们从 CoronaryDominance (Kruzhilov et al. 2025) 的一个子集构建了一个数据集。从 249 位患者的血管造影视频中，我们手动选择每个视频中的一个高质量帧，生成 1,595 个帧。然后我们通过应用泊松噪声、高斯噪声和高斯模糊来模拟噪声。生成的数据集被分成 1,286 个训练帧（200 位患者）和 309 个测试帧（49 位患者）。为了评估，我们采用峰值信噪比（PSNR）和结构相似性指数（SSIM）作为指标。

实现细节

预训练。对于所有实验，我们使用表 1 中的数据集，其中包括用于预训练的 20 K X 射线血管造影图。大部分配置借用了 MAE (He et al. 2022) 。默认情况下，我们使用 ViT-B/ 16 (Dosovitskiy et al. 2021) 作为骨干网络，并对其进行 800 个周期的预训练。对于分割器，我们采用轻量级 UNeXt-S (Valanarasu and Patel 2022)，其参数仅有 0.3 M。我们采用线性学习率缩放规则 (Goyal et al. 2017)：$lr = lr_{\text{base}} \times \text{batch\_size}/256$ 。其他实现细节见表 S1 。

血管分割。U-Net (Ronneberger, Fischer, and Brox 2015) 用作分割解码器并在每个数据集上针对 500 次迭代进行微调。所有输入图像都调整为 224×224 的尺寸。我们使用 AdamW 优化器，初始学习率为 $1e^{-4}$ ，权重衰减为 0.05 。应用具有 $T\_\text{max} = 500$ 的余弦退火学习率调度。

Config	Value
Optimizer	AdamW
Base learning rate	$1.5e^{-4}$
Weight decay	0.05
Momentum	$\beta_1, \beta_2 = 0.9, 0.95$
Layer-wise lr decay	1.0
Batch size	256
Learning rate schedule	Cosine decay
Warmup epochs	40
Training epochs	800
Augmentation	RandomResizedCrop

Table S1: VasoMIM 的预训练配置。

Method	PSNR (dB)	SSIM
Traditional		
LSF (Buades, Coll, and Morel 2005) [CVPR '05]	28.36	0.6300
From Scratch		
DnCNN (Zhang et al. 2017) [TIP '17]	35.00	0.8822
Self-Supervised Pre-Training		
DINO (Caron et al. 2021) [ICCV '21]	35.08	0.8840
MAE (He et al. 2022) [CVPR '22]	35.11	0.8837
DeblurringMIM † (Kang et al. 2024) [MedIA '24]	35.10	0.8830
HPM (Wang et al. 2025) [TPAMI '25]	35.10	0.8835
VasoMIM [Ours]	35.14	0.8842

Table S2: 冠状动脉主导性的最新方法比较。所有方法都使用其官方代码库重新实现。最好和次好的结果分别以 red 和 yellow 突出显示。结果取三个随机种子的平均值。

在图像去噪的定量结果如表 S2 所示，VasoMIM 达到了领先的 PSNR 为 35.14 dB 和 SSIM 为 0.8842 ，超过了所有基线方法。这种一致的增益强调了我们框架学习的血管表示的优越可迁移性。

血管解剖质量的影响。Frangi 滤波器中的阈值 α 控制提取的血管解剖质量，如图 ?? 所示。随着 α 的增加，Frangi 滤波器的性能显著改善，而 VasoMIM 的性能仅略有提升，如表格 ?? 所示。这表明我们的解剖感知预训练框架对血管解剖质量具有高度的鲁棒性。

Setting	DSC (%)	
	ARCADE	CAXF
w/o $\mathcal{L}_{\text{cons.}}$	68.30	83.96
Joint	68.78 ↑ 0.48	84.06 ↑ 0.10
In Advance	**68.85 ↑ 0.55**	**84.49 ↑ 0.53**

Table S4: 用于解剖一致性损失（方程 (4)）的分割器 $\mathcal{S}$ 的不同训练设置比较。

www.xueshuxiangzi.com

Loss	DSC (%)	
	ARCADE	CAXF
–	68.30	83.96
$\mathcal{L}_{\text{w-rec.}}$	68.78 ↑ 0.48	84.14 ↑ 0.18
$\mathcal{L}_{\text{cons.}}$	**68.85 ↑ 0.55**	**84.49 ↑ 0.53**

Table S5: 解剖加权重建损失 $\mathcal{L}_{\text{w-rec.}}$ 与解剖一致性损失 $\mathcal{L}_{\text{cons.}}$ 。

掩膜比例对 CAXF 的影响。为了进一步研究掩膜比例对 CAXF 的敏感性，我们改变 γ 并在图 ?? 中展示结果。与 ARCADE 一致，VasoMIM 在中等比例（$0.3 \sim 0.5$ ）下表现最佳，因此我们默认采用 $\gamma = 0.5$ 。
提前训练与联合训练。在 VasoMIM 中，我们使用 Frangi 滤波器生成的伪标签提前训练分割器，并在 MIM 预训练期间保持其权重不变。我们还评估了一种联合训练的变体，其中分割器和 MIM 模型在预训练期间同时优化。正如表格 S4 所示，提前训练比联合训练带来更大的 DSC 增益，而联合训练将预训练时间从 15.5 增加到 20.5 小时。因此，提前训练分割器既更有效也更高效。
解剖加权重建损失与解剖一致性损失。在本节中，我们将我们解剖一致性损失 $\mathcal{L}_{\text{cons.}}$ 与解剖加权重建损失 $\mathcal{L}_{\text{w-rec.}}$ 进行比较，其中标准重建损失 $\mathcal{L}_{\text{rec.}}$ 被按补丁式血管分布 f 加权。该变体的总体训练损失为 $\mathcal{L}'_{\text{train}} = \mathcal{L}_{\text{rec.}} + \mathcal{L}_{\text{w-rec.}}$ 。如表 S5 所示，$\mathcal{L}_{\text{w-rec.}}$ 在 ARCADE 上提高了 DSC 0.48% ，在 CAXF 上提高了 0.18% ，而 $\mathcal{L}_{\text{cons.}}$ 分别获得了更大的增益 0.55% 和 0.53% 。这些结果表明 $\mathcal{L}_{\text{cons.}}$ 使模型能够学习更稳健的血管表示。

$\mathcal{L}$	DSC (%)	
	ARCADE	CAXF
–	68.30	83.96
L_1	68.81 ↑ 0.51	84.32 ↑ 0.36
Dice	67.97 ↓ 0.33	84.12 ↑ 0.16
Cross Entropy	**68.85 ↑ 0.55**	**84.49 ↑ 0.53**

Table S6: 对不同度量函数 $\mathcal{L}$ 的敏感性。

不同的度量函数 $\mathcal{L}$ 。我们研究了方程 (4) 中不同的度量函数，包括 L_1 损失、Dice 损失和 CE 损失。如表 S6 所示，我们发现 CE 损失是最佳选择，相较于基线，在 ARCADE 上提升了 DSC 0.55% ，在 CAXF 上提升了 0.53% 。
不同的解码器设计。根据 (He et al. 2022) ，我们研究了两个关键的解码器设计选择：深度和嵌入维度。结果总结在表 S7 中。与 MAE 及其变体（表 2）一致，具有 8 个解码器区块和 512 嵌入维度的配置是最佳选择。
模型复杂性。我们在表格 S8 中比较了 MAE 和我们的 VasoMIM 的模型复杂性。将 UNeXt-S (Valanarasu and Patel 2022) 整合到 VasoMIM 中仅增加了可以忽略不计的开销。考虑到图 7 中报告的训练效率提升，总体而言，VasoMIM 是更高效的选择。

Blocks	# params (M)	DSC (%)	
		ARCADE	CAXF
1	89.84 (1.25×)	**68.86**	84.38
2	92.99 (1.20×)	68.39	84.22
4	99.30 (1.13×)	68.11	83.85
8	111.91 (1.00×)	68.85	**84.49**
12	124.52 (0.90×)	68.75	84.31

Dim	# params (M)	DSC (%)	
		ARCADE	CAXF
128	87.68 (1.28×)	67.58	83.59
256	92.61 (1.21×)	66.68	82.47
512	111.91 (1.00×)	**68.85**	**84.49**
1024	188.25 (0.59×)	68.16	84.23

Table S7: 对不同解码器配置的敏感性。

Method	# params (M)	GFLOPs	Speed (s/epoch)
MAE	111.65 (1.00×)	9.43 (1.02×)	54 (1.30×)
VasoMIM	111.91 (1.00×)	9.59 (1.00×)	70 (1.00×)

Table S8: 模型复杂度。GFLOPs 和速度是在一台 NVIDIA A6000 GPU 上评估的。GFLOPs 使用单个 224×224 RGB 图像（带掩膜）作为输入进行计算。速度使用批量大小为 256 进行评估。

血管分割的定性结果。我们在 ARCADE、CAXF 和 XCAV 上展示了更多的定性结果。如图 S3 所示，VasoMIM 在减少误报的情况下实现了更精确的血管分割结果。
分块掩码比例的可视化。我们在预训练过程中展示了更多关于分块掩码比例的实例，清晰表明了 VasoMIM 优先掩盖包含血管的图块。

www.xueshuxiangzi.com

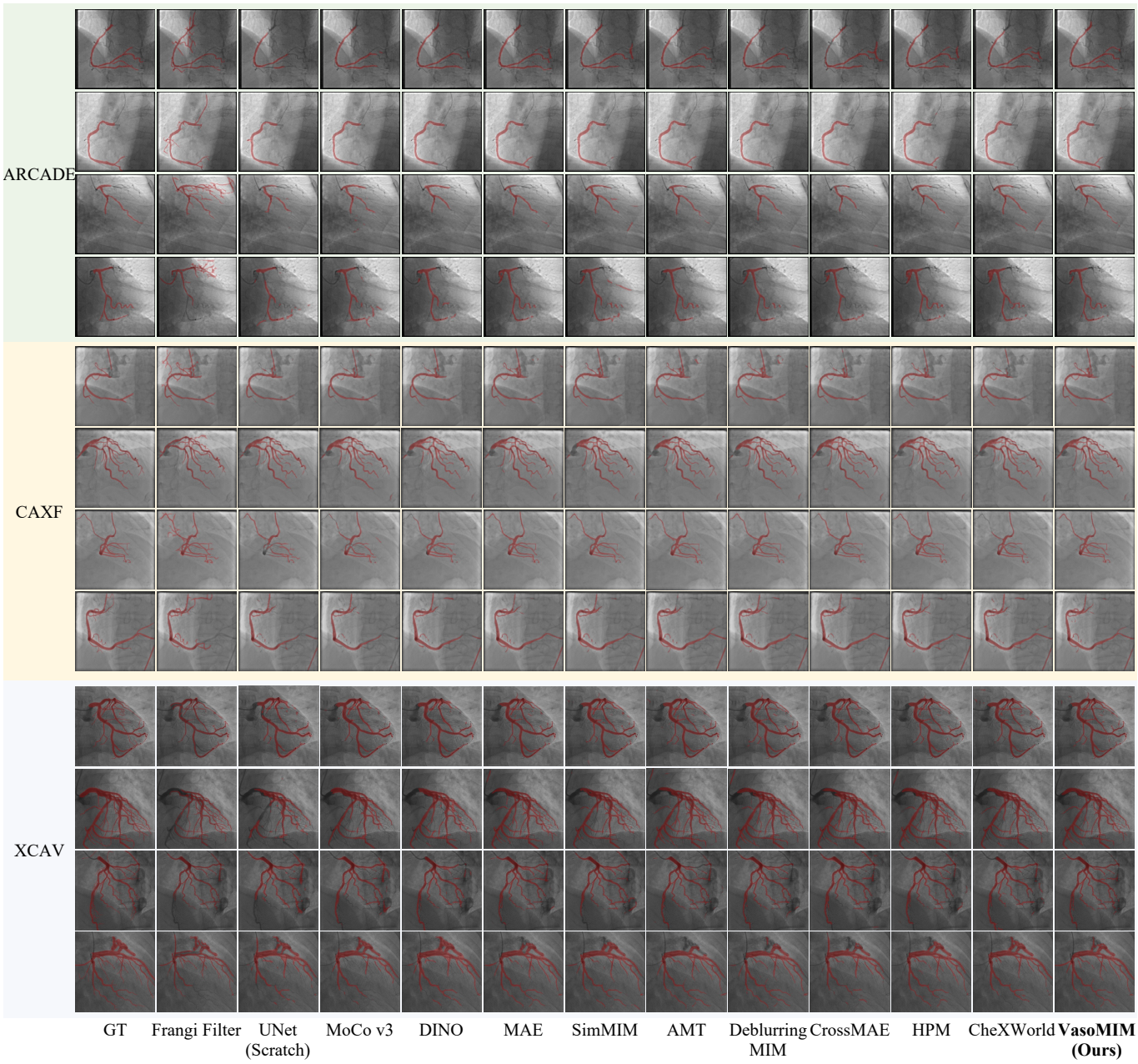


Figure S3：关于 ARCADE、CAXF 和 XCAV 的更多定性结果。

www.xueshuxiangzi.com

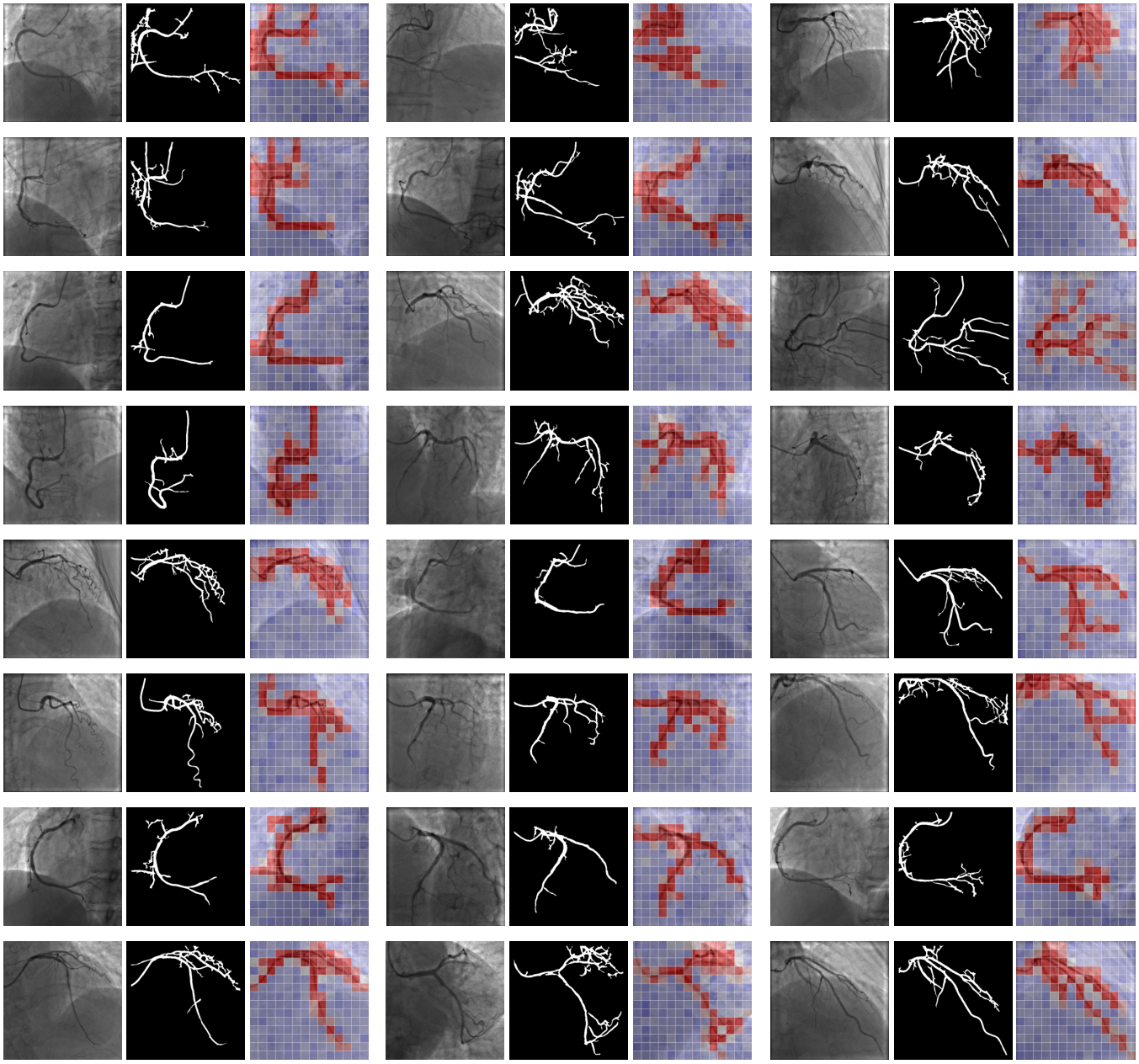


Figure 第四部分: 整个预训练过程中的块状屏蔽比例，即 $\frac{1}{E}\sum_{j=1}^{E}\mathbb{I}\left(\text{Patch } x\_i \text{ is masked in epoch } j\right)$ 。对于每种情况，我们展示输入图像、通过 Frangi 滤波器提取的血管解剖结构，以及预训练过程中块状屏蔽比例。红色表示更高的比例，蓝色表示相反。

www.xueshuxiangzi.com