

## 摘要

医学图像分割在医学诊断和治疗中起着关键作用，但传统方法依赖大量标注数据，存在耗时昂贵且隐私问题。本文复现了《Self-Supervised Medical Image Segmentation Using Deep Reinforced Adaptive Masking》一文，提出了一种自监督的医学图像分割方法，利用深度强化学习进行自适应掩码。通过自适应硬掩码策略，提高了分割模型的性能，并减少了对标注数据的依赖。实验结果表明，该方法对现有的医学图像数据集提高了分割精度和分类准确率。

**关键词：**医学图像分割; 自适应硬掩码

## 1 引言

医学图像分割在医学诊断和治疗中起着关键作用，特别是在皮肤镜图像中，准确分割病变区域对于后续的病变分类和诊断至关重要。然而，传统的图像分割方法通常依赖大量的标注数据，这不仅耗时昂贵，还需要专业知识进行标注，标注过程也容易受到隐私问题的影响，导致目前高质量的标注数据稀缺。

因此，为了解决标注数据稀缺的问题，需要选择能在没有或少量人工标注的情况下学习视觉表示的方法，如自监督学习的方法。同时，通过将掩膜图像的过程建模为强化学习问题，利用重建模型的反馈信号指导代理学习掩膜策略，可以选择更合适的掩膜位置和大小，从而提高分割模型的性能。根据以上依据，选择了 Self-Supervised Medical Image Segmentation Using Deep Reinforced Adaptive Masking 这篇论文 [2] 来进行复现。这篇论文提出了一种自监督的医学图像分割方法，利用深度强化学习进行自适应掩膜。

这篇论文通过自适应掩膜的策略，能够更精确地分割出病变区域，提高分割模型的性能。同时，利用自监督学习方法，可以减少对大量标注数据的依赖，降低数据标注的成本和时间，该技术的应用将推动医学图像分析领域的发展，为临床诊断和治疗提供更可靠的技术支持。

## 2 相关工作

### 2.1 图像分割网络 U-Net

U-Net 是一种用于生物医学图像分割的卷积神经网络 [3]，由于结构类似于字母“U”，因此得名。它的主要特点是通过对称的编码器-解码器结构，实现了高效的图像分割。在编码器部分，U-net 对图像通过多次卷积和池化操作逐步减少空间信息，同时增加特征信息；在解码器部分，U-net 通过上采样操作逐步恢复空间信息，并通过跳跃连接与编码器的对应层特征结合，从而实现精确的分割。

## 2.2 VGG16 模型

VGG16 是经典的深度卷积神经网络模型 [4]，由牛津大学视觉几何组提出，主要用于图像分类任务，它使用了 16 个加权层，包括 13 个卷积层和 3 个全连接层，主要特点是采用了许多 3x3 的小卷积核，并以深层网络的形式堆叠起来。这种设计使得模型能够捕捉到图像中的复杂特征。

## 2.3 通用图像分割网络 Segment Anything Model

Segment Anything Model(SAM) 模型 [1] 是由 Meta AI 推出的一种通用图像分割模型，可以处理各种图像分割任务。SAM 采用了类似于 U-Net 的编码器-解码器结构，在一个包含 1100 万张图像和 11 亿个掩码的数据集上进行训练，得到的大模型在各种分割任务上表现出色，具有强大的零样本性能。

# 3 本文方法

## 3.1 掩码图像建模

掩码图像建模 (Masked Image Modeling, MIM) 是一种用于计算机视觉的自监督学习方法。它的核心思想是将输入图像的一部分像素进行掩码处理，然后让模型根据未被掩码的部分来预测被掩码的部分。通过这种方式，模型可以学习到图像的局部和全局特征，并将其应用在下游的图像分割，图像分类等任务，从而提升其在各种视觉任务中的表现。

## 3.2 自适应硬掩码 Adaptive Hard Masking

传统的掩码图像建模方法通常通过随机选择图像中的一部分像素进行掩码处理，然后让模型根据未被掩码的部分来预测被掩码的部分。然而，与自然图像相比，医学图像中的信息分布较为不均匀，与医学图像相比，自然图像的信息通常是高度分散和多样化的，医学图像的重建损失要低得多，使用随机掩码策略可能不足以获得很好的医学图像分割性能。

该论文提出了使用自适应硬掩码 (Adaptive Hard Masking, AHM) 的自监督学习方法，该方法引入了一种自适应的掩码策略，能够根据图像内容的复杂度和重要性来动态调整掩码区域。它可以生成对模型来说更具挑战性的掩码，使模型能够更好地学习到复杂的特征。在两个医学图像数据集上进行的实验表明，AHM 用于医学图像分割时，优于其他先进的自监督学习方法。

## 3.3 具体流程

自适应硬掩码流程：

1. 将图像输入特征提取网络 (feature extraction network)，得到图像特征后将其输入到决策网络 (policy network) 与价值网络 (value network) 中，分别输出决策图 (policy map) 与价值图 (value map)。
2. 使用输出的决策图对原图像进行掩码，得到掩码图像后将图像分割网络 U-Net 作为重建网络，重建掩码图像得到进行掩码并重建后的图像。

3. 根据原图像与重建后图像的差异，计算重建的平均重建损失，并根据损失计算 reward。
4. 使用上述步骤得到的 reward，按照论文中使用的公式，计算各个网络更新使用的梯度后，更新这些网络的梯度。

具体流程图如下：

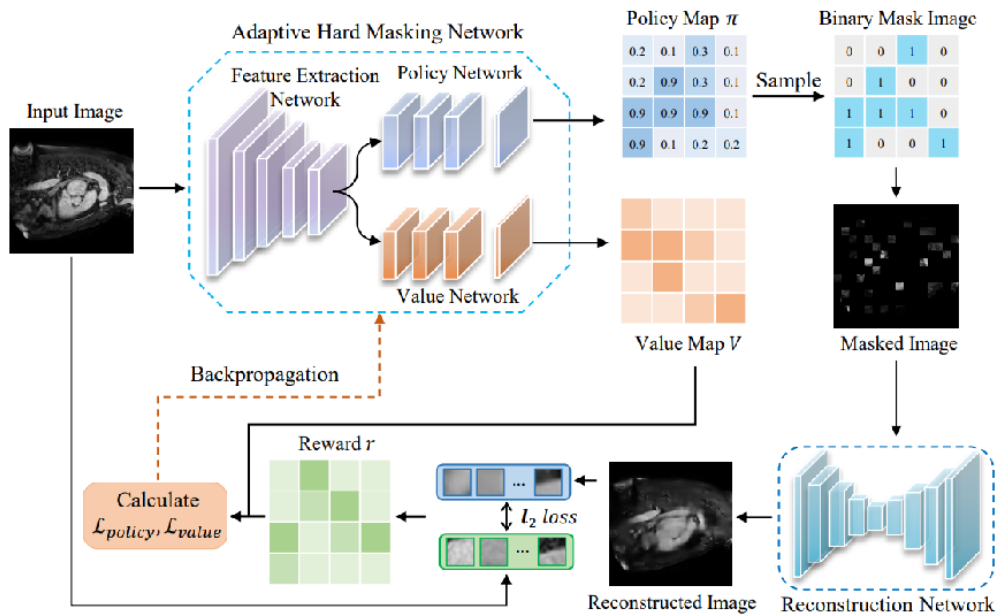


Fig. 3. The framework of the proposed AHM. Our AHM Architecture is based on A3C, where the adaptive hard masking network is used to predict regions with higher reconstruction loss. Our adaptive sampling procedure samples more patches from more discriminative regions and fewer patches from the low information or redundant regions.

图 1. 自适应硬掩码流程

## 4 复现细节

### 4.1 与已有开源代码对比

本文仅参考了 <https://github.com/milesial/Pytorch-UNet> 中在 pytorch 环境下使用 U-net，复现论文中使用自适应硬掩码训练 U-net 的改进工作未参考其他源代码。

### 4.2 实验环境搭建

我们的实验是使用 PyTorch 实现的并在 NVIDIA GeForce GTX 4060 GPU 上运行，在重建网络的预训练中，我们使用 Adam 优化器，初始学习率设置为  $2e-4$ ，批量为 12。对于迁移学习，用于下游分割任务的 U-Net 是使用 Adam 优化器以初始学习率进行训练设置为  $3e-4$ ，每 3 个 epoch 衰减 10%，批量大小为 4。

### 4.3 创新点

与原文不同，本文额外使用了医院中实际的病人数据集进行实验，这样真实的数据集更能体现论文中方法的实用性。

## 5 实验结果分析

分别使用标准的 U-net 与使用自适应硬掩码方法训练后的 U-net, 对数据集进行图像分割后, 与标准掩码计算 dice score, 得到使用自适应硬掩码方法的 U-net 的 dice score 有提升, 并且对于下游的图像分类任务, 分类的准确率也有提升。具体的结果如下图:

	直接训练U-Net分割	使用自适应硬掩码 后训练U-Net分割
分割dice score	0.6000	0.6345
分类准确率	76.29%	78.64%

图 2. 实验具体结果

## 6 总结与展望

本文复现并改进了基于深度强化学习的自监督医学图像分割方法, 通过自适应硬掩码策略, 显著提高了分割模型的性能, 减少了对标注数据的依赖。与标准 U-Net 相比, 采用自适应硬掩码的 U-Net 在医学图像分割任务中表现更佳, 为临床诊断和治疗提供了可靠的技术支持。

展望未来, 可以进一步探索自适应硬掩码策略在更多医学图像数据集上的应用, 并结合其他自监督学习方法, 进一步提升模型的泛化能力和实用性。此外, 考虑到医学图像的多样性和复杂性, 可以尝试引入多模态数据和多任务学习, 进一步优化分割模型的性能。希望该方法能够在实际临床环境中得到广泛应用, 为医学图像分析领域的发展贡献更多力量。

## 参考文献

- [1] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment Anything. *arXiv preprint arXiv:2304.02643*, 2023.
- [2] Yunxin Liu, Gang Xu, Thomas Lukasiewicz, and Zhenghua Xu. Self-Supervised Medical Image Segmentation Using Deep Reinforced Adaptive Masking. *IEEE Transactions on Medical Imaging*, 43(4):TBD, 2024.
- [3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv preprint arXiv:1505.04597*, 2015.
- [4] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv:1409.1556*, 2015.