

题目

Self-supervised Semantic Segmentation: Consistency over Transformation

摘要

这篇论文介绍了一种新颖的自监督算法S3-Net，用于医学图像分割。本文引入了Inception Large Kernel Attention (I-LKA) 模块来捕捉上下文信息，并结合了可变形卷积来处理病变变形，强调对仿射变换的不变性。该方法还包含了一个空间一致性损失项，以提高分割质量。该方法在皮肤病变和肺部器官分割任务中的表现优于当前最先进的方法，显示了其在处理数据集偏差、捕捉长距离依赖性和保持局部一致性方面的有效性。S3-Net方法以端到端方式迭代学习图像内容聚类的像素级特征表示。在皮肤损伤和肺器官分割任务上的实验结果表明，与SOTA方法相比，本文的方法具有优越的性能。

1 引言

由于在监督学习方法中依赖大量标注数据集且在医学成像中，标注数据稀缺且获取成本高，通常导致模型偏差。虽然迁移学习有所帮助，但它受限于标注数据的可用性。无监督方法和半监督算法提供了替代解决方案，但也有自身的局限性，例如数据偏差和在领域转换中缺乏鲁棒性。自监督学习（SSL）作为一种有前景的方法出现，它通过使用前置任务和下游任务来从未标记数据中学习。论文指出，现有的自监督方法存在局限性，特别是在捕捉长距离依赖性、处理病变变形以及对常见医学图像转换的不变性方面。为了解决这些问题，论文提出了一种新颖的自监督算法，该算法采用了Inception Large Kernel Attention (I-LKA) 模块和可变形卷积，强调空间一致性和对仿射变换的不变性。

2 相关工作

2.1 自监督学习SSL

SSL通过允许以无监督的方式从未标记的数据中提取语义和有效的表示，显示了视觉任务的显著优势。这种方法的根源在于，通过增强表示学习可以实现显著的性能改进。对比学习方法通过优化相似性约束来努力获得弹性表示，从而在给定数据集中区分相似(正)和不相似(负)对。

2.2 SGSCN和dual-branch Transformer network.

SGSCN该方法利用不同的损失函数对具有相似特征表示的空间连接图像像素进行分组。在迭代学习过程中，网络同时学习单个图像中每个像素的特征表示和聚类分配。此外，引入上下文感知一致性损失，通过加强属于同一簇的像素与其中心之间的空间接近性来增强图像区域的描绘。

dual-branch Transformer network旨在捕获不同尺度的全局上下文依赖关系和局部信息。考虑了尺度之间的语义依赖，产生监督信号以确保尺度间的一致性，并在每个尺度内强制空间稳定性损失以促进内容聚类。此外，引入了交叉熵损失函数，有效地建模聚类分布并改善聚类之间的

决策边界。

3 本文方法

3.1 本文方法概述

本文提出的方法的框架如图1所示。S3-Net集成了本地和远程依赖关系来完成特征嵌入过程。通过合并这些依赖项，我们的目标是捕获全面的全局信息，同时保留输入数据中固有的本地细节。为了在不需要手动注释的情况下实现有效的内容聚类，我们的方法结合了辅助模块和精心设计的数据驱动损失函数。通过利用数据中存在的固有结构和关系，我们的方法使模型能够以鲁棒性和准确性驱动分割任务。

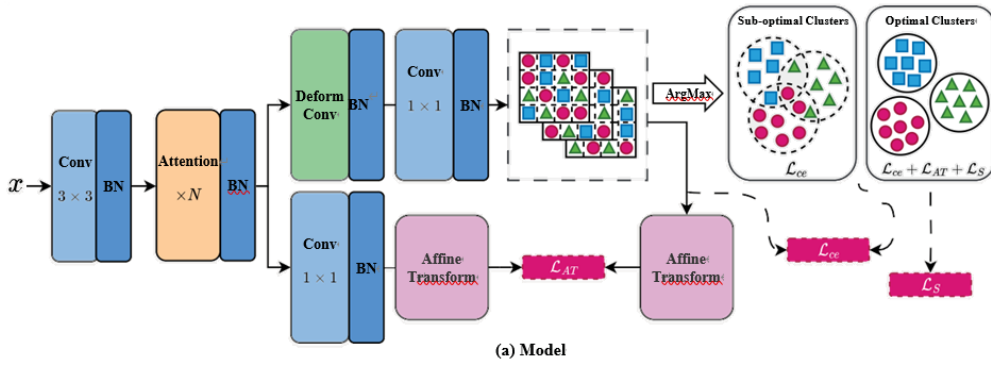


图 1. 方法示意图

3.2 特征提取模块

编码器架构包括三个块，它们共同将输入图像编码到潜在空间。第一个块采用顺序结构，包括一个 3×3 卷积层，然后是一个批处理归一化层。这种配置便于将输入图像嵌入到高维空间中。在随后的堆叠N块中，我们使用I-LKA模块来捕获本地和全局依赖项。后面是另一个 1×1 卷积层。这种设置确保通过跳过连接在像素级保存细粒度的空间信息，同时使用I-LKA模块指导网络捕获全局依赖关系。在最后一个块中，我们结合了一个可变形的卷积层，专门设计用于有效地模拟病变边界的变形，这在医学图像中很常见。这个额外的层增强了网络准确捕获和表示变形的能力，从而提高了整体性能。有图2。

3.2.1 DeformableConv 模块

通过定义可变形卷积层来增加网络的灵活性，使其能够适应输入数据的几何变形。在医学图像分析中，可变形卷积可以帮助模型更准确地定位和分割不规则形状的解剖结构。

3.2.2 I-LKA 模块

结合自注意力和大核卷积的优缺点提出 I-LKA 模块，能够在生成注意力图时维持低的计算复杂性和参数数量，同时提取特征空间内的长距离关系。使得模型在处理医学图像时有对全局信息的捕捉能力，同时还要保留局部细节。公式如下：

$$\text{Inc}(x) = \{(\text{DW-Conv}(F(x))) \mid r \in N\} \quad (1)$$

$$\text{Attention} = \text{Conv } 1 \times 1 (\text{DW-D-Conv}(\text{Inc}(x))) \quad (2)$$

$$\text{Output} = \text{Attention} \otimes F(x) \quad (3)$$

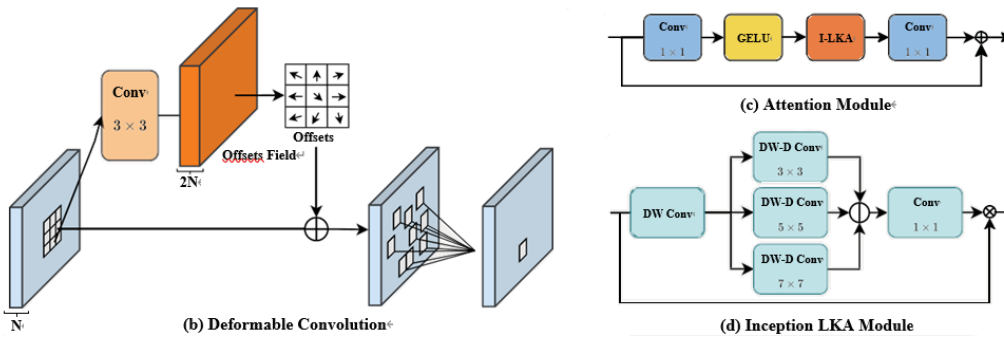


图 2. 模块示意图

3.3 损失函数定义

本文在训练过程中最终损失函数包括三个不同的损失项

3.3.1 交叉熵损失 (Cross-entropy loss)

在训练阶段，我们采用迭代方法来最小化交叉熵损失，该交叉熵损衡量软预测图与分割图之间的差异。通过优化这个损失函数，我们的网络学会产生准确而有意义的分割结果。公式如下：

$$L_{ce}(S, Y) = -\frac{1}{H \times W} \sum_{i=1}^{H \times W} \sum_{j=1}^K Y_{i,j} \log(S_{i,j}) \quad (1)$$

3.3.2 仿射变换损失 (Affine transformation loss)

该损失函数旨在增强网络对仿射变换（如平移、旋转、缩放）的不变性。通过最小化网络对仿射变换特征的预测掩码和变换后的真实掩码之间的差异，提高网络的鲁棒性和对变化的处理能力。公式如下：

$$L_{AT}(Y', Y^a) = -\frac{1}{H \times W} \sum_{i=1}^{H \times W} \sum_{j=1}^K Y_{i,j}^a \log(Y'_{i,j}) \quad (2)$$

3.3.3 空间一致性损失 (Spatial consistency loss)

此损失函数侧重于促进图像中邻近像素之间的空间一致性。通过计算图像边缘信息，最小化基于边缘信息的成对差异，从而提高分割结果的空间连贯性和视觉一致性。公式如下：

$$L_S = \sum_{i,j} (|(X_{i,j} - Y_{i,j}) - Z_{i,j}| + |(X_{i,j} - XY_{i,j}) - Z_{i,j}| + |(Y_{i,j} - XY_{i,j}) - Z_{i,j}|) \quad (3)$$

最后引入加权因子 λ_1 、 λ_2 和 λ_3 控制每个损失项的相对重要性：

$$L_{joint} = \lambda_1 L_{ce} + \lambda_2 L_{AT} + \lambda_3 L_S \quad (4)$$

4 复现细节

4.1 与已有开源代码对比

本文是在原论文已有开源代码的基础上进行改进的，主要是参考了源代码的整体架构，从而对算法细节进行了一些改进与创新，主要的创新包括：增加仿射变化的参数学习，将空间一致性损失改为边缘细化损失。

4.2 实验环境搭建

为了学习可训练的参数，我们采用SGD优化，迭代地最小化总体损失函数，最多迭代50次。SGD优化的学习率为0.36，动量为0.9。实验是在单个4060 GPU上使用PyTorch库进行的。使用了PH2数据集，该数据集由200张黑素细胞病变的RGB图像组成。

4.3 性能评估指标

采用Dice (DSC) 评分、XOR指标和hamoud距离 (HM) 作为评估指标。根据中[1]提出的方法，我们只考虑与ground truth (GT) map重叠程度最高的聚类作为评估我们方法的目标类预测。在我们的评估中，DSC分数作为预测目标区域与GT之间一致性的指标。更高的DSC分数反映了更好的表现。相反，HM和XOR度量的是预测目标和GT图之间的差异。因此，HM和XOR值越低，

性能越好。

4.4 创新点

4.4.1 仿射变化的参数学习

构建一个局部化网络用来从输入特征图中学习空间变换所需的参数，它由两个卷积层、两个最大池化层和两个 ReLU 激活函数层组成，后面跟着一个自适应平均池化层。然后构建一个全连接层序列，由两个线性层中间插一个 ReLU 激活函数组成，将特征最终压缩成6维仿射变换所需的参数。可学习的仿射变换可以帮助模型自适应地校正因尺度、旋转或位置的不同产生的变化，使得模型能够更准确地识别对象，增加模型对旋转、缩放、平移等的鲁棒性。

4.4.2 引入边缘细化损失函数

将空间一致性损失改为边缘细化损失。边缘细化损失函数的目的是最小化分割映射和它的下采样和随后的上采样对应之间的差异，从而产生一个边缘映射。它可以帮助模型捕捉到更为细致的结构，尤其是在边缘区域，得以更准确地捕捉细小结构的边界。公式如下：

$$L_{Boundary} = \sum_{i,j} |(Down - Up - Y)_{i,j} - Y_{i,j}| \quad (1)$$

5 实验结果分析

经过多次实验，最终对权重因子 λ_1 、 λ_2 和 λ_3 进行了优化，分别设置为3.5、0.5和2。另外设置 α 为3，下采样因子 β 为16。将改进后的模型运用到PH2数据集进行测试评估，输出图与Ground truth分别进行DICE, HM, XOR评估，在评估方法时，只考虑与Ground truth重叠程度最高的聚类作为目标类预测。改进后的模型经过评估后的如表所示，DICE有所提升，HM有所下降，比原模型表现要好，但XOR上升不如原模型。

Methods	DICE ↑	HM ↓	XOR ↓
k-means	71.3	130.8	41.3
S3	88.0	20.4	22.0
S3-improved	88.47	18.81	28.87

表1. 原论文的评估结果和修改结果对比

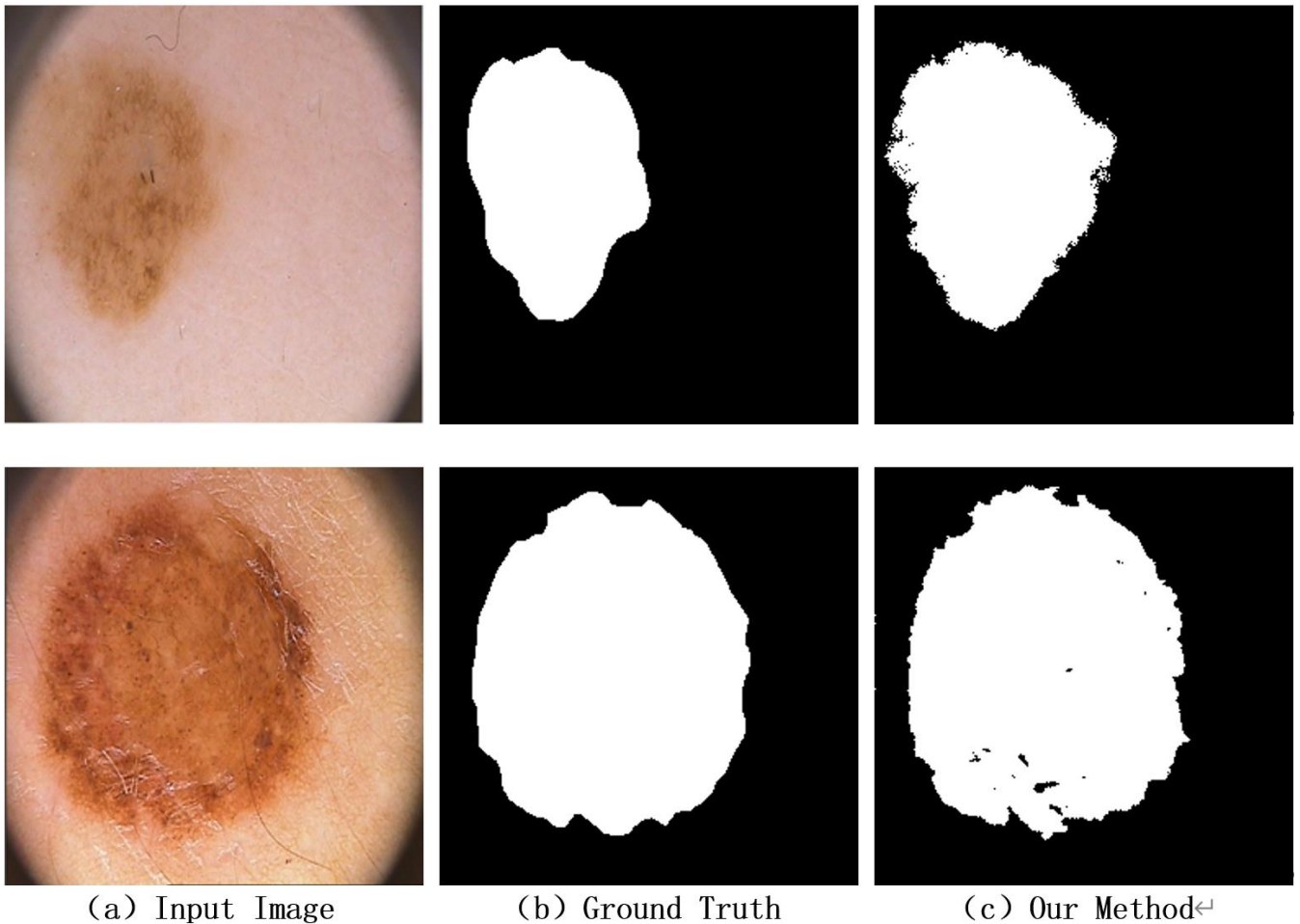


图3. 分割结果图

6 总结与展望

本通过本课程学习到的知识和参考学习了网络上的相关资料，复现本篇论文并对其进行适当的改进。复现论文的过程中让我对图像的分割有了更深刻的理解，尤其在对论文的模型结构进行改进的时候，从提出改进方法，构建并训练网络和调整参数，最终达到一个较为理想的效果，真正的提升了自己的技术水平。

当然，本篇论文的模型还有可以改进的地方，还存在着一些难以排除的干扰，希望通过接下来的学习可以找到更好的改进办法取得更好的分割效果，能创造出属于自己的新东西，真正步入科研的正轨。

参考文献

- [1] Euijoon Ahn, Dagan Feng, and Jinman Kim. A spatial guided self-supervised clustering network for medical image segmentation. In Medical Image Computing and Computer Assisted Intervention – MICCAI 2021: 24th International Conference, Strasbourg, France, September 27 – October 1, 2021, Proceedings, Part I 24, pages 379 – 388. Springer, 2021.
- [2] Abdulrahman Gharawi, Mohammad D Alahmadi, and Lakshmi Ramaswamy, “Self-supervised skin lesion segmentation: An annotation-free approach,” *Mathematics*, vol. 11, no. 18, pp. 3805, 2023.
- [3] Sanaz Karimijafarbigloo, Reza Azad, Amirhossein Kazerouni, and Dorit Merhof, “Ms-former: Multi-scale self-guided transformer for medical image segmentation,” in *Medical Imaging with Deep Learning*, 2023.
- [4] Mohammad Reza Hosseinzadeh Taher, Fatemeh Haghighi, Michael B Gotway, and Jianming Liang, “Caid: Context-aware instance discrimination for self-supervised learning in medical imaging,” in *International Conference on Medical Imaging with Deep Learning*. PMLR, 2022, pp. 535 – 551.
- [5] Yuting He, Guanyu Yang, Rongjun Ge, Yang Chen, Jean-Louis Coatrieux, Boyu Wang, and Shuo Li, “Geometric visual similarity learning in 3d medical image self-supervised pre-training,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 9538 – 9547.
- [6] Xiaohong Huang, Zhifang Deng, Dandan Li, and Xueguang Yuan, “Missformer: An effective medical image segmentation transformer,” *arXiv preprint arXiv:2109.07162*, 2021.
- [7] Shanghua Gao, Zhong-Yu Li, Ming-Hsuan Yang, MingMing Cheng, Junwei Han, and Philip Torr. Large-scale unsupervised semantic segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 2022.
- [8] Reza Azad, Ehsan Khodapanah Aghdam, Amelie Rauland, Yiwei Jia, Atlas Haddadi Avval, Afshin Bozorgpour, Sanaz Karimijafarbigloo, Joseph Paul Cohen, Ehsan Adeli, and Dorit Merhof. Medical image segmentation review: The success of u-net. *arXiv preprint arXiv:2211.14830*, 2022.