

基于动态扫描机制的 Vision RWKV 让医学影像分割轻量而高效

sora*, Zhendong Li*,
*Ningxia University
Email: kyochilian@gmail.com
*Ningxia University
Email: lizhendong@nxu.edu.cn

Abstract—基于 Transformer 的分割框架是当前医学影像分割的主流方法，能够构建全局关系。然而，在需要高精度与高分辨率的医学图像分割中，Transformer 计算复杂度高，应用有限。最近的 RWKV 模型降低了空间聚合复杂度并具备全局处理能力，但其在视觉领域存在长程建模羸弱的问题，同时循环嵌套机制在分割领域中的解码性能不足。为了解决这些问题，我们提出了一种基于 Vision RWKV 的分割框架---，该框架在具有较低复杂度的同时具备了全局建模能力。为了增强 RWKV 的长程空间连续性，我们设计了一种动态扫描机制，该机制将图像分为全局 block 与局部 block，根据不同层次的编码器动态地调整顺序和方向，将全局与局部特征融合，大大提升了模型的精度。同时，我们在解码器中构建了上采样模块---，该模块改良了 RWKV 的循环嵌套机制，提升了重建多尺度特征的能力。我们还采用多种方法进一步降低---的计算复杂度，实现模型的轻量化。(LRFormer 低分辨率自注意力, 池化 QKV 等等)实验表明，我们的方法在多个数据集取得优异性能，同时计算复杂度与计算速度显著增强，使其在高精度分割的同时具备轻量化特性。代码开源至: <https://github.com/shepherdxu/SCI-winning>。

Index Terms—computer vision, semantic segmentation, RWKV, Transformer, lightweight

I. INTRODUCTION

Semantic Segmentation 是计算机视觉中的一项重要任务，医学影像分割是关键应用之一，它将复杂的医学图像分割成不同的区域，使器官、病灶及注意区域清晰可见，对于医学辅助诊断与治疗具有重要意义。相较于人工，使用计算机视觉技术进行辅助诊断不仅提高了诊断效率，还提升了诊断精度，因此许多方法被提出用于医学影像分割任务。传统的 ML (机器学习) 方法在研究中被广泛应用，如基于图割 (Graph Cut) 的方法 [?], 随机森林 (Random Forest) [?] 等。近年来，随着 FCN (Full Convolution Network) 的提出 [?], 深度学习方法成为图像语义分割的主流方法。随着 Unet [?] 的提出，基于深度学习的分割方法占据了主导地位，通过上下采样，以及创新性地跳跃连接，使其成为了语义分割的主流框架。

写好文章介绍。首先写语义分割，然后写医学影像分割，然后写方法迭代。

neural networks (DNNs) have achieved remarkable success in various AI tasks. However, their enormous computational cost hinders deployment on resource-limited devices. Pruning redundant parameters is one of the most popular compression techniques [1].

II. RELATED WORK

A. Magnitude-based Pruning

Han *et al.* [1] proposed to remove weights with small absolute values.

B. Dynamic Sparse Training

Most recently, DST [2] allows the sparse topology to evolve during training.

III. METHODOLOGY

Let $\mathcal{W} = \{W_l\}_{l=1}^L$ denote the weights of an L -layer network. Our goal is to find a sparse mask M_l for each layer such that the remaining weights $W_l \odot M_l$ retain accuracy.

A. Dynamic Growth Criterion

We use the gradient magnitude as the saliency score:

$$s_{ij}^{(l)} = \left| \frac{\partial \mathcal{L}}{\partial w_{ij}^{(l)}} \right|. \quad (1)$$

IV. EXPERIMENTS

We evaluate DST on CIFAR-10/100 and ImageNet with ResNet-50.

TABLE I
TOP-1 ACCURACY (%) ON CIFAR-10 UNDER DIFFERENT SPARSITIES.

Method	90% sparsity	95% sparsity
Baseline	93.5	91.2
Magnitude [1]	92.8	89.7
DST (ours)	94.1	92.3

V. CONCLUSION

We presented a simple yet effective DST framework that dynamically adjusts sparse connectivity during training. Future work includes extending DST to transformer architectures.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant 62XXXXXX.

REFERENCES

- [1] S. Han, J. Pool, J. Tran, and W. Dally, “Learning both weights and connections for efficient neural network,” *Proc. NIPS*, 2015.
- [2] U. Evci, T. Gale, J. Menick, P. S. Castro, and E. Elsen, “Rigging the lottery: Making all tickets winners,” *Proc. ICML*, 2020.