首届《遥感学报》青年学术论坛暨第六届全国遥感地理信息研究生论坛

Linear Attention meets Semantic Segmentation

李睿
武汉大学遥感信息工程学院

遥感学报

中国科学院空天信息创新研究院
Aerospace Information Research Institute,Chinese Academy of Sciences

西南交通大学 SWJTU
Faculty of Geosciences and Environmental Engineering
地球科学与环境工程学院

Introduction

- **Context matters for semantic segmentation**
  Per-pixel classification is often ambiguous in the presence of only local information. But the task becomes much simpler if contextual information, from the whole image, is available

- **Attention is effective for context extraction**
  With strong capabilities to capture long-range dependencies, dot-product attention mechanisms have been applied in vision and NLP tasks

- **Capturing context information is consuming**
  Utilization of the dot-product attention mechanism often comes with significant memory and computational costs, which increases quadratically with the size of the input over space and time

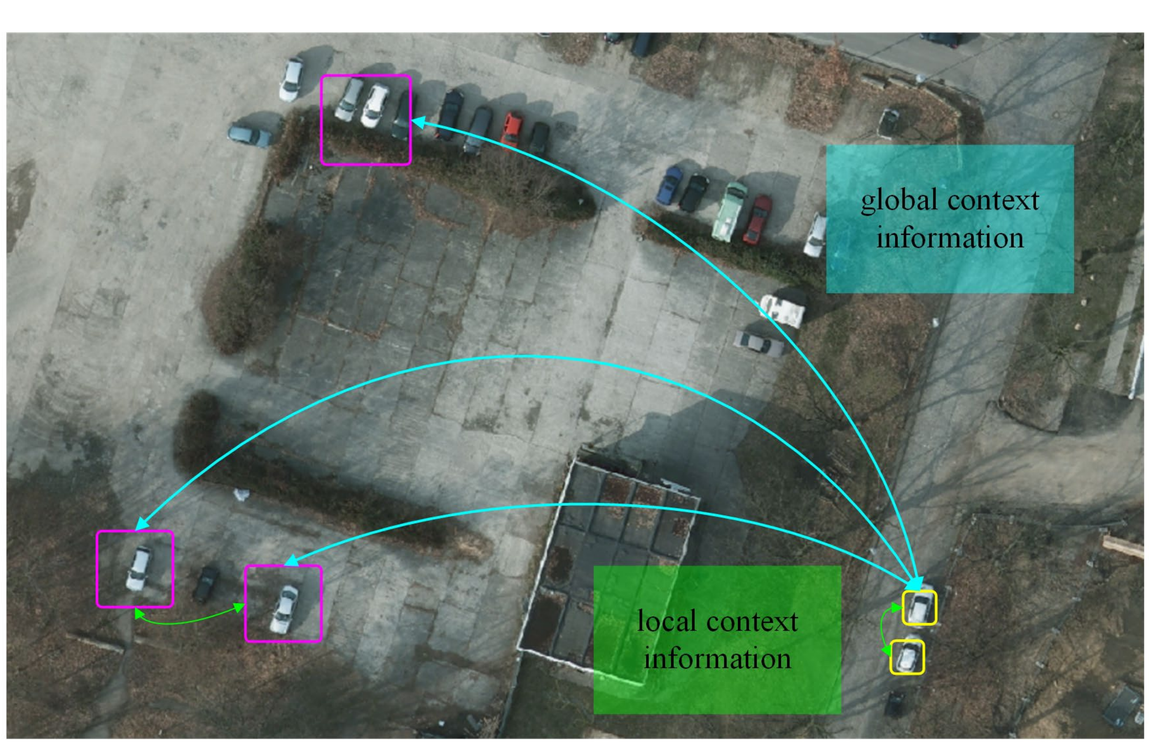

Fig. 1. Illustration of global and local contextual information.

Methodology

- **Dot-product attention mechanism**

$$D(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \rho(\boldsymbol{Q}\boldsymbol{K}^T)\boldsymbol{V}.$$

where

$$\begin{cases} \boldsymbol{Q} = \boldsymbol{X}\boldsymbol{W}_q \in \mathbb{R}^{N\times D_k}; \\ \boldsymbol{K} = \boldsymbol{X}\boldsymbol{W}_k \in \mathbb{R}^{N\times D_k}; \\ \boldsymbol{V} = \boldsymbol{X}\boldsymbol{W}_v \in \mathbb{R}^{N\times D_v}. \end{cases}$$

and

$$\rho(\boldsymbol{Q}\boldsymbol{K}^T) = softmax_{row}(\boldsymbol{Q}\boldsymbol{K}^T)$$

Replace the softmax function by its Taylor expansion

$$e^{\boldsymbol{q}_i{}^T\cdot\boldsymbol{k}_j} \approx 1 + \boldsymbol{q}_i{}^T\cdot\boldsymbol{k}_j$$

- **Linear attention mechanism**

$$D(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \frac{\sum_j \boldsymbol{V}_{i,j} + \left(\frac{\boldsymbol{Q}}{\|\boldsymbol{Q}\|_2}\right)\left(\left(\frac{\boldsymbol{K}}{\|\boldsymbol{K}\|_2}\right)^T\boldsymbol{V}\right)}{N + \left(\frac{\boldsymbol{Q}}{\|\boldsymbol{Q}\|_2}\right)\sum_j\left(\frac{\boldsymbol{K}}{\|\boldsymbol{K}\|_2}\right)^T_{i,j}}$$

Results

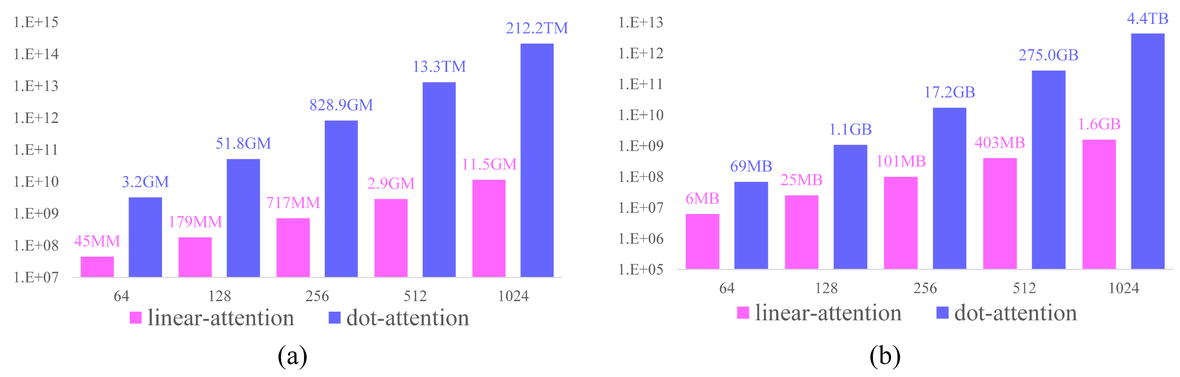

Fig 2. Comparison between the (a) computational and (b) memory requirements of the linear attention mechanism and dot-product attention mechanism.

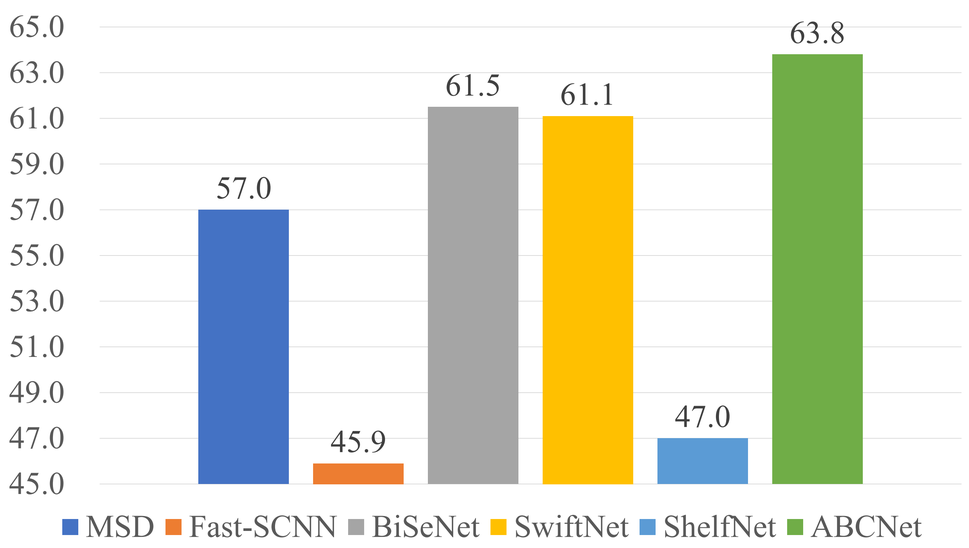

Fig 3. The mIoU of different methods on the UAVid dataset..

Conclusion

The proposed linear attention mechanism is an effective and efficient method which balances the global context and resource consumption well. Based on the linear attention mechanism, the proposed ABCNet achieves a comparative result on UAVid dataset.

References

[1] Li, Rui, et al. "Multistage Attention ResU-Net for Semantic Segmentation of Fine-Resolution Remote Sensing Images." IEEE **GRSL** (2021).
[2] Li, Rui, et al. "Multiattention network for semantic segmentation of fine-resolution remote sensing images." IEEE **TGRS** (2021).
[3] Li, Rui, et al. "ABCNet: Attentive bilateral contextual network for efficient semantic segmentation of Fine-Resolution remotely sensed imagery." **ISPRS P&RS**, 181 (2021): 84-98.

Home Page
个人主页