



Attention-based Multi-Context Guiding

for Few-Shot Semantic Segmentation

Tao Hu, Pengwan Yang, Chilong Zhang, Gang Yu, Yadong Mu, Cees G.M. Snoek
 University of Amsterdam, Peking University, Tsinghua University, Megvii Inc.(Face++)



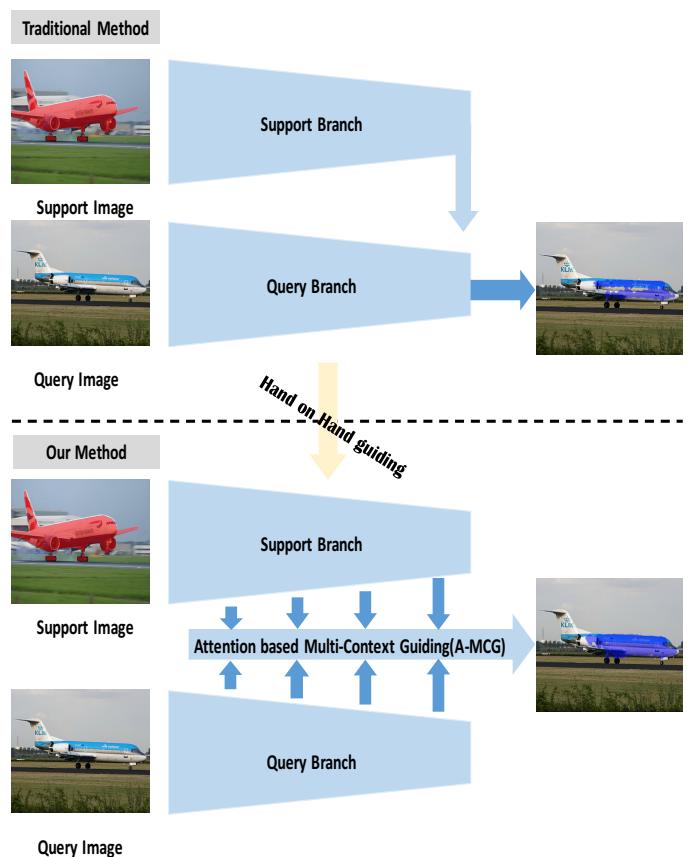
清华大学
Tsinghua University

Face++ 旷视

Task Introduction

Baby can recognize object by only few "labeled" images. How can a system realize segmentation of a target class based on few pixel-level semantic annotations?

Motivation



➤ Multi-Context, Attention, Multi-shot fusion

Contribution

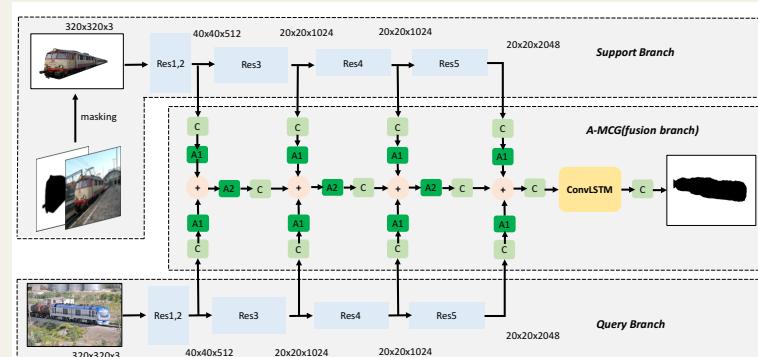
(1). We first propose a Multi-Context Guiding structure to fuse the small-to-large scale context features between support branch and query branch to globally guide the query branch segmentation.

(2). We introduce a Residual Attention Module(Wang et al. 2017) in our MCG network to realize the attention mechanism in few-shot learning of segmentation.

(3). We embed the Conv-LSTM (Xingjian et al. 2015) module into the end of our network to better merge the feature map from support set in multi-shot semantic segmentation.

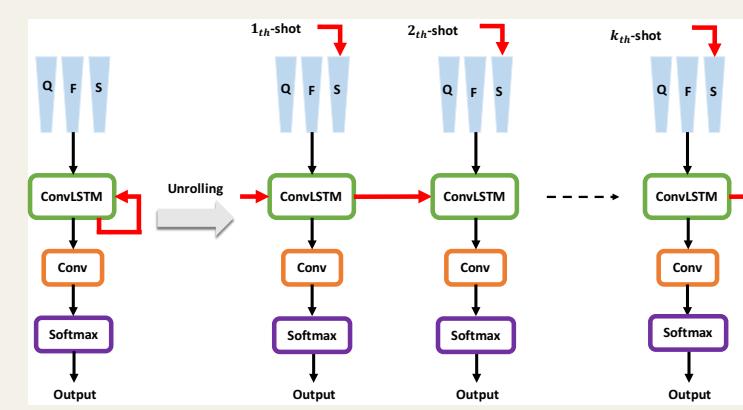
(4). Compared with previous methods, our A-MCG reaches state-of-the-art 61.2%, 62.2% measured in mIoU in 1-shot and 5-shot setting.

Network Structure



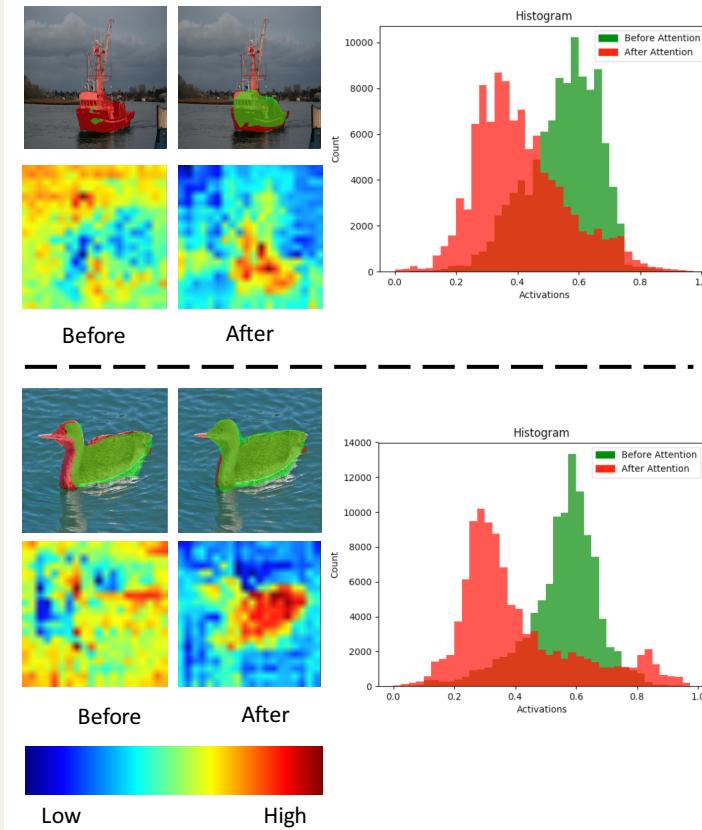
➤ C means Conv.
 ➤ A1, A2 denotes two kinds of attention.

Unrolling ConvLSTM



➤ Q: Query. F: Fusion. S: Support.

Attention Visualization



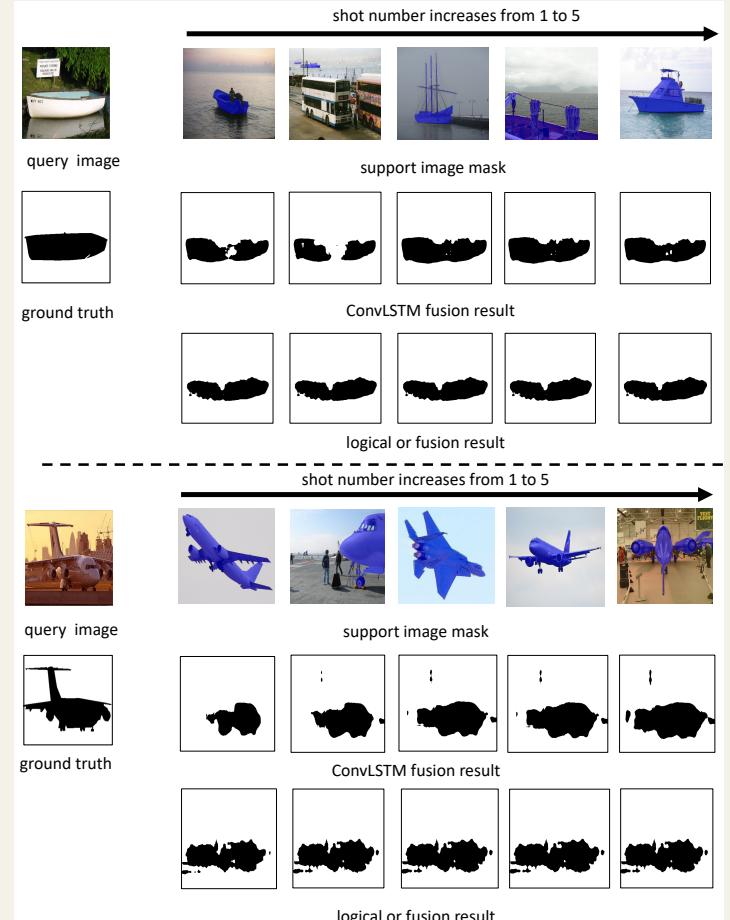
➤ Validate Attention Mechanism qualitatively and quantitatively.

Multi-shot fusion by ConvLSTM

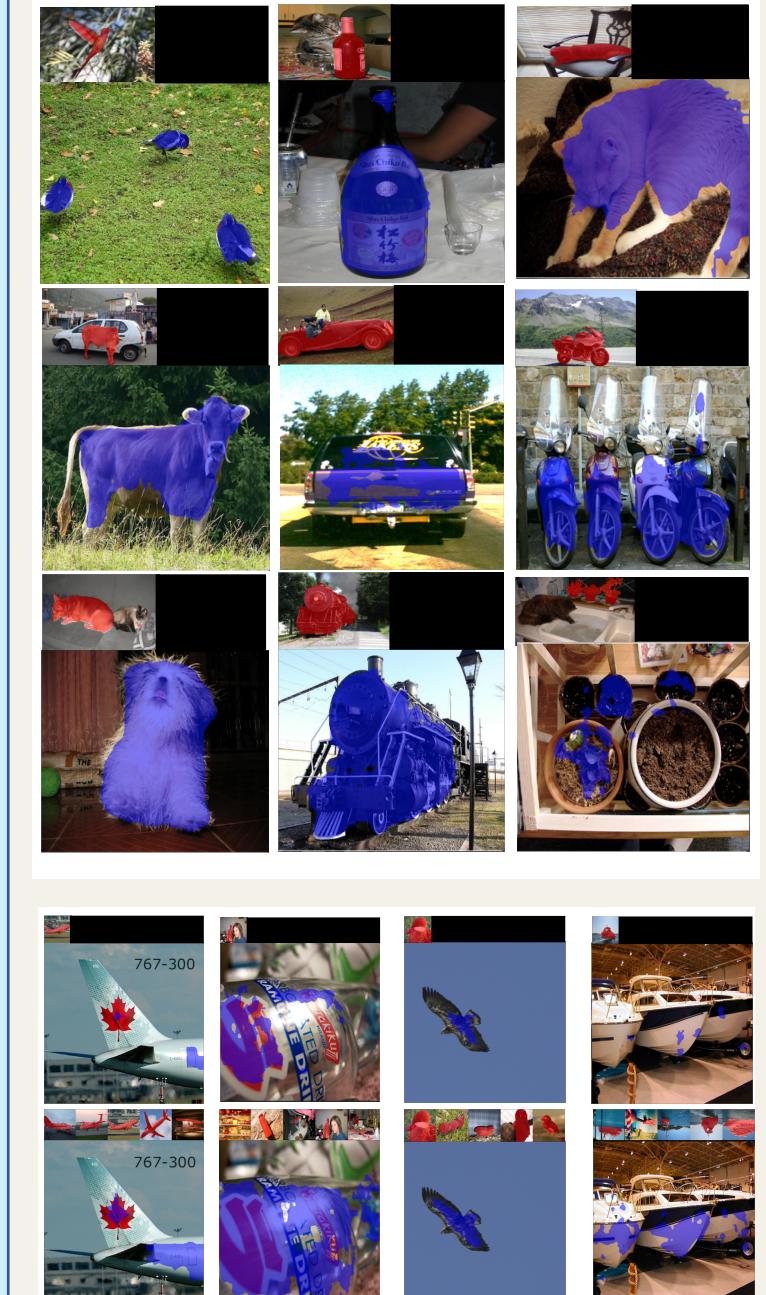


➤ Multi-shot fusion works better than "logical or".
 ➤ K-loss performs better than 1-loss ConvLSTM.

Gradually Increasing Shot-Number



Result Visualization



Contact Information

