ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10096973

PROGRESSIVE META-POOLING LEARNING FOR LIGHTWEIGHT IMAGE CLASSIFICATION MODEL

Peijie Dong[1], *Xin Niu*[1,*], *Zhiliang Tian*[1,*], *Lujun Li*[2], *Xiaodong Wang*[1]
Zimian Wei[1], *Hengyue Pan*[1], *Dongsheng Li*[1]

[1] School of Computer, National University of Defense Technology, Changsha 410073, China
[2] Chinese Academy of Sciences, Beijing, China

ABSTRACT

Practical networks for edge devices adopt shallow depth and small convolutional kernels to save memory and computational cost, which leads to a restricted receptive field. Conventional efficient learning methods focus on lightweight convolution designs, ignoring the role of the receptive field in neural network design. In this paper, we propose the Meta-Pooling framework to make the receptive field learnable for a lightweight network, which consists of parameterized pooling-based operations. Specifically, we introduce a parameterized spatial enhancer, which is composed of pooling operations to provide versatile receptive fields for each layer of a lightweight model. Then, we present a Progressive Meta-Pooling Learning (PMPL) strategy for the parameterized spatial enhancer to acquire a suitable receptive field size. The results on the ImageNet dataset demonstrate that MobileNetV2 using Meta-Pooling achieves top1 accuracy of 74.6%, which outperforms MobileNetV2 by 2.3%.

***Index Terms*—** meta learning, receptive field, lightweight network, pooling operation

1. INTRODUCTION

Lightweight networks tailor specifically for mobile or resource-constrained platforms, with fewer parameters and memory consumption. MobileNet[1] and ShuffleNetv2[2] utilize factorization to make the standard convolution light-weight and computationally friendly. However, the lightweight networks are spatially local, a.k.a. the receptive field is limited.

The receptive field determines how the network aggregates spatial contexts and has a significant impact on the performance of the lightweight model. Evidence shows that a large receptive field is necessary for high-level recognition tasks with diminishing rewards[3, 4]. Conventional approaches adopt small kernel stacking [5, 6], pooling-based methods[7, 8, 9], and special convolution operations [10, 11, 12] to enlarge the receptive field. Among them, spatial pooling is playing an important role, which maintains

*Corresponding author

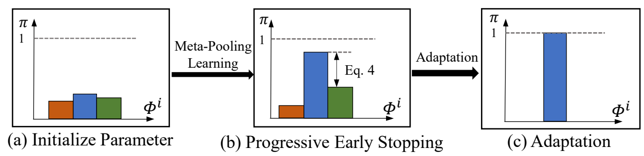

Fig. 1. Overview of the progressive meta-pooling learning. (a) Initialize the meta-parameter α^i. (b) Progressive early stopping when Eq.4 is satisfied. (c) Adapt the learned spatial enhancers.

the most important activation for the most distinguishing features. Besides the average pooling and max pooling, recent pooling operations mainly focus on how to preserve the local details when down-sampling the feature map. We explore the potential of pooling operations to aggregate spatial information and expand the receptive field. Other methods [13, 14] resort to attention mechanism to expand the receptive field from the perspective of attention. Most of the attention-based methods are plugged into a network in a fixed manner and the receptive field is predetermined. However, as revealed in the CoAtNet[15], having a larger receptive field does not necessarily means a better performance while having an appropriate local receptive field can improve the generalization ability of the model.

This study is to determine a suitable receptive field for each layer of a lightweight network in a data-driven way. To make the receptive field learnable, we resort to meta-learning approaches. Meta-learning aims to learn knowledge shared across multiple tasks, and quickly adapt the model to new unseen tasks. Model-Agnostic Meta-Learning[16] (MAML) has been spotlighted due to its simplicity and generality. To date, otsuzuki [17] proposes to use meta-learning to learn L_p pooing as well as a binary mask to determine the kernel shape of pooling layers for character recognition, but neglects to investigate receptive fields. We apply the concept of MAML to receptive field search by assuming the multiple tasks in MAML as learning for multiple receptive fields. MAML minimizes the expectation of loss across tasks while we minimize the expectation of loss across various receptive fields.

In this paper, we present a meta-learning-based approach