# 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
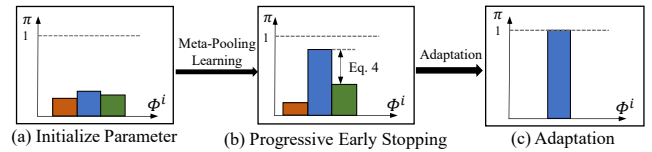
---

*Corresponding author



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

the most important activation for the most distinguing 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.

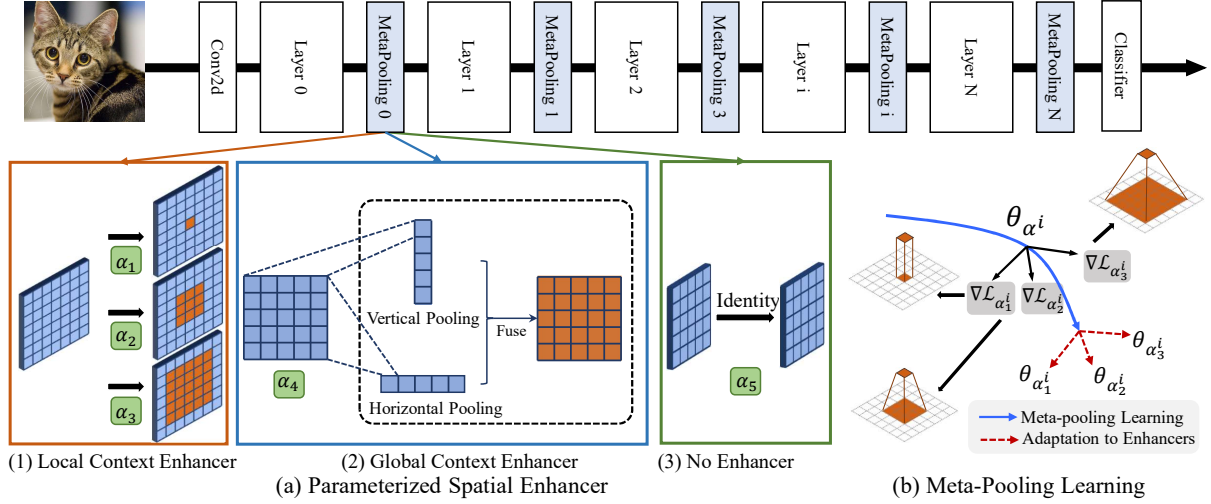In this paper, we present a meta-learning-based approach

**Fig. 2**. Overview of the proposed Meta-Pooling. We insert the independent Meta-Pooling after each layer of a lightweight network. (a) Parameterized spatial enhancer encompasses three receptive field enhancers, a.k.a. local, global, and no enhancer. The meta-parameter $\alpha$ is adopted to measure the importance of different enhancers. (b) Diagram of our meta-pooling learning algorithm, which optimizes for a representation $\theta_{\alpha^i}$ that can adapt to the enhancers of the $i$-th layer.

called Meta-Pooling to make the receptive field of lightweight networks learnable and enhance the feature discriminability and robustness. As shown in Fig.2(a), our proposed Meta-Pooling consists of a parameterized spatial enhancer, including a local context enhancer, a global context enhancer, and no enhancer, which provides the lightweight model with diverse sizes of receptive fields. We employ a Progressive Meta-Pooling Learning (PMPL) strategy with early stopping to learn the distribution of parameterized spatial enhancers, as shown in Fig.2(b) and Fig.1.

Our contributions are summarized as follows:

(1) We present a Meta-Pooling framework to learn the versatile receptive fields for the lightweight network, which is composed of parameterized spatial enhancers.

(2) We propose Progressive Meta-Pooling Learning with an early stopping strategy that allows the model to gradually determine the receptive fields.

(3) We demonstrate that Meta-Pooling outperforms the baseline by 2.3% and exhibits a remarkable performance compared with pooling and attention-based methods.

## 2. METHODOLOGY

We describe the parameterized spatial enhancers in Sec.2.1, which provides the lightweight model variety of receptive field patterns. We then introduce a meta-learning-based approach to make the receptive field learnable with progressive early stopping in Sec.2.2.

### 2.1. Parameterized Spatial Enhancer

Parameterized spatial enhancer can provide a variety of receptive fields for the lightweight models. As shown in Fig.2, it includes a global context enhancer, a local context enhancer,

and no enhancer. The global and local context enhancer can enlarge the receptive field while no enhancer represents the skip connection and can maintain the receptive field.

**Global Context Enhancer.** Global Context Enhancer enables the lightweight model to capture an informative long-range context from horizontal and vertical directions. To capture long-range interactions spatially with less computational burden, we factorize the global pooling from the vertical and horizontal directions, inspired by CoordAttention[9] and Strip Pooling[8]. We present a spatial separable pooling operation, which is capable to extract global context economically. Specifically, given the input data $X \in \mathbb{R}^{C \times H \times W}$, two spatial extents of pooling kernels $(H, 1)$ or $(1, W)$ are adopted to encode horizontal and vertical direction, respectively. The information encoded from the horizontal directions can be formulated as: $z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i)$. Similarly, the information aggregated from the vertical direction can be written as: $z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w)$. Then, we use point-wise convolution to refine the knowledge from the direction-aware feature map and apply Sigmoid to get the spatial activation. The final output of spatial separable pooling is obtained by refining the original feature map with the spatial activation.

**Local Context Enhancer.** Local Context Enhancer compensate for the lack of local context by introducing a slightly larger kernel. It includes a series of pooling operations with a large kernel, including average pooling and max pooling. In order to provide a more hierarchical receptive field expansion option to the lightweight model, the max pooling and average pooling with diverse kernel sizes $k = \{3, 5, 7\}$ are utilized to gather local contexts.

**Parameterized Spatial Enhancer.** To adaptively determine the receptive field, we parameterize the above spatial enhancers to form the Meta-Pooling and make it learnable.

**Algorithm 1:** Progressive Meta-Pooling Learning.

**Input:** layer depth $L$, model weight $\theta$,
meta-parameter $\{\alpha^i\}^L$, learning rate $\zeta$ for the
inner loop and $\eta$ for the outer loop, early
stopping threshold $\rho$, loss on meta dataset
$\mathcal{L}_{mt}$, loss on training dataset $\mathcal{L}_{tr}$

**while** not all fixed **do**

  Update meta-parmaeters $\alpha$ by
  $\theta^{'} \leftarrow \theta - \zeta \nabla_\theta \mathcal{L}_{tr}(\theta, \alpha)$;
  $\alpha \leftarrow \alpha - \eta \nabla_\alpha \mathcal{L}_{mt}(\theta^{'}, \alpha)$;
  Update model weight $\theta$ by $\theta \leftarrow \theta - \zeta \nabla_\theta \mathcal{L}_{tr}(\theta, \alpha)$
  **foreach** $l \leftarrow 0$ *to* $L-1$ **do**
    $\pi^l \leftarrow$ Softmax$(\alpha^l)$
    $\pi^l_k \leftarrow \max\{\pi^l\}$
    diff $\leftarrow 0$
    **foreach** $m \leftarrow 0$ *to* $|\pi|$ **do**
      diff $\leftarrow \min\left(\text{diff}, \pi^l_k - \pi^l_m\right)$
    **end**
    **if** diff $\geq \rho$ **then**
      Fix the distribution by $\arg\max \pi^l$;
    **end**
  **end**

**end**

---

**Table 1.** Comparision with pooling-based methods on ImageNet. All methods are based on MobileNetV2 [1]. The symbol $\star$ indicates that the PyTorch implementation is not available.

| | Top1(%) | Lat.(ms) | F.(M) | P.(M) |
|---|---|---|---|---|
| Skip | 71.34 | 7.37 | 320.24 | 3.50 |
| MaxPool | 71.35 | 7.59 | 320.81 | 3.50 |
| AveragePool | 71.68 | 7.70 | 320.81 | 3.50 |
| DPP [18]$\star$ | 70.15 | - | - | - |
| LiftDownPool [19] $\star$ | 73.91 | - | - | - |
| GatedPool [20] | 71.06 | 10.35 | 320.24 | 3.73 |
| GaussPool [21] | 72.87 | 16.72 | 321.04 | 3.85 |
| BlurPool [22] | 69.42 | 8.23 | 320.24 | 3.50 |
| Ours | **74.60** | 16.97 | 322.12 | 3.55 |

model can be viewed as MAML of each spatial enhancer. Given $\mathcal{T}$ as the given dataset, our meta-pooling learning approach is formulated as follows:

$$\alpha^i = \arg\min_{\alpha^i} \mathbb{E}_{\beta \sim p(\beta|\pi)}\left[\mathcal{L}_\tau\left(f^{i+1:}_{\theta^*(\alpha^i)}, \alpha^i\right)\right]$$
$$s.t. \quad \tau = p\left(f^{0:i-1}_{\theta^*}(x), y|\{x, y\} \sim \mathcal{T}\right) \quad (2)$$
$$\theta^*(\alpha^i) = \arg\min_{\theta^*} \mathbb{E}[\mathcal{L}_\mathcal{T}(\theta, \alpha^i)]$$

where $\theta$ denotes the weights of the lightweight model and $f^{i:j}$ is a composition from $i$-th layer to $j$-th layer. $f^{0:i-1}(x)$ means the intermediate distribution from the last layer, and the distribution is determined by all the previous layers from $p(f)$. We approximate the bi-level objective through first-order model-agnostic meta-learning as follows:

$$\theta^{'} \leftarrow \theta - \zeta \nabla_\theta \mathcal{L}_{tr}\left(\theta, \alpha^i\right)$$
$$\alpha^i \leftarrow \alpha^i - \eta \sum_i \nabla_{\alpha^i} \mathcal{L}_{mt}\left(\theta^{'}, \alpha^i\right) \quad (3)$$

where $\zeta$ and $\eta$ are the learning rate of $\theta$ and $\alpha^i$. Denoted by $\mathcal{L}_{mt}$ and $\mathcal{L}_{tr}$ is the loss on the meta dataset and training dataset, respectively. The meta dataset is created by separating the original training dataset.

To reduce the computational cost, we propose the progressive early stopping mechanism to progressively discrete the lightweight network by layer through the distribution of parameter $\pi^i$, as shown in Fig.1. Specifically, a threshold $\rho$ is set to determine whether to stop searching for the $i$-th layer and the condition is as follows:

$$\min\left\{\pi^i_k - \pi^i_m, \forall m \neq k | \pi^i_k = \max\left\{\boldsymbol{\pi^i}\right\}\right\} \geq \rho \quad (4)$$

Once the above condition is met on any layer, the receptive field operation with the highest probability would be fixed, whose parameters would stop updating in the subsequent training. After all of the meta-parameters are fixed, we adapt the learned spatial enhancer to the original task and train the lightweight network from scratch. The iterative procedure is outlined in Alg.1.

Specifically, the meta-parameters $\alpha^i = \{\alpha^i_1, \alpha^i_2, .., \alpha^i_n\}$ are introduced to measure the importance of different spatial enhancers $\Phi^i = \{\beta^i_1, \beta^i_2, ..., \beta^i_n\}$, where $\alpha^i_n$ means the $n$-th parameter of the $i$-th layer and $\beta^i_j$ means the $j$-th receptive field expansion operations of the $i$-th layer. By meta-parameters, we relax the categorical choice of a particular receptive field and make the learning process end-to-end.

## 2.2. Progressive Meta-Pooling Learning

We introduce a Progressive Meta-Pooling Learning (PMPL) algorithm to optimize the parameterized spatial enhancers. Moreover, a progressive early stopping mechanism is adopted to accelerate the training process. The PMPL allows the layers of the model at various depths $d = \{d^0, d^1, .., d^n\}$ to automatically measure the importance of different spatial enhancers.

Given the parameterized spatial enhancers $\Phi^i$, we employ categorical distribution $\beta \sim p(\beta|\pi)$ with probabilities $\pi^i$ for each depth $d^i$. We apply continuous relaxation to meta-parameter $\alpha^i$, and the probabilities are computed as a Softmax over $\alpha^i$ defining the importance for the corresponding spatial enhancer $\Phi^i$ of the $i$-th layer by:

$$\bar{\beta}^i(x) = \sum_{k=0}^{|\Phi^i|} \pi^i_k \beta_k(x), \pi^i_j = \frac{\exp\left(\alpha^i_j\right)}{\sum_{k=0}^{|\Phi|} \exp\left(\alpha^i_k\right)} \quad (1)$$

where $x$ and $\bar{\beta}^i$ are the input and mixed output of the $i$-th layer. In this way, learning the receptive field for a lightweight

**Table 2**. Comparison of attention-based methods under width multipliers of $\{1.0, 0.75, 0.5\}$.

|  | Param.(M) | M-Adds(M) | Top-1 Acc(%) |
|---|---|---|---|
| MobileNetV2 | 3.50 | 300 | 72.3 |
| +SE[13] | 3.89 | 300 | $73.5_{+1.2}$ |
| +CBAM[14] | 3.89 | 300 | $73.6_{+1.3}$ |
| +Meta-Pooling(Ours) | 3.91 | 320 | $\mathbf{74.6}_{+2.3}$ |
| MobileNetV2-0.75 | 2.5 | 200 | 69.9 |
| +SE[13] | 2.86 | 210 | $71.5_{+1.6}$ |
| +CBAM[14] | 2.86 | 210 | $71.5_{+1.6}$ |
| +Meta-Pooling(Ours) | 2.86 | 210 | $\mathbf{71.8}_{+1.8}$ |
| MobileNetV2-0.5 | 2.0 | 100 | 65.4 |
| +SE[13] | 2.1 | 100 | $66.4_{+1.0}$ |
| +CBAM[14] | 2.1 | 100 | $66.4_{+1.0}$ |
| +Meta-Pooling(Ours) | 2.1 | 100 | $\mathbf{66.9}_{+1.5}$ |

## 3. EXPERIMENTS

We first introduce the experiment setup in Sec.3.1 and then carry out various experiments based on MobileNetV2 in Sec.3.2. Finally, we perform an ablation study to verify the contribution of our proposed Meta-Pooling in Sec.3.3.

### 3.1. Experiment Setup

Our experiments are implemented with PyTorch. The SGD optimizer with a momentum of 0.9 and weight decay of 4e-5 is adopted to train the lightweight models, and the Adam optimizer is utilized to train the meta-parameters with a weight decay of 1e-3. The cosine learning schedule is adopted for SGD optimizer. We use 4 NVIDIA GPUs for training and the batch size is set to 128. The threshold $\rho$ of progressive early stopping is 0.2. We first train for 30 epochs by PMPL and retrain for 200 epochs when the receptive fields of all layers are fixed. The learning rate of $\zeta$ is 0.05 and $\eta$ is 3e-4. We use the same data augmentation as in MobileNetV2[1]. All of the experiments are conducted on ImageNet dataset[23].

### 3.2. Results

Using the progressive early stopping strategy enables the model to converge quickly within 30 epochs. We notice that deeper layers tend to employ Global Context Enhancer, which reveals the lack of receptive field in lightweight networks. We further compared Meta-Pooling with other methods.

**Pooling-based Methods.** We compared Meta-Pooling with a range of pooling-based approaches[22, 18, 20, 21, 19], as shown in the Tab.1. Meta-Pooling achieves 3% higher accuracy than max pooling and average pooling. The pooling-based operations can well preserve detailed information and position shift-invariant but fail to explore larger receptive fields.

**Attention-based Methods.** We compare our Meta-Pooling with other lightweight attention methods, such as SE[13], CBAM[14], etc. As shown in Tab.4, the SE and CBAM improve the performance by 1%, while our Meta-Pooling shows

**Table 3**. Ablation study of enhancers and training strategy. "No" refers to skip connection. "Local" refers to an average pooling layer with a kernel size of 3. "Global" refers to spatial separable pooling. "Random" refers to a random selection from three enhancers. "PMPL" refers to the progressive meta-pooling learning strategy with early stopping.

| No | Local | Global | Random | PMPL | Acc(%) |
|---|---|---|---|---|---|
| ✓ | - | - | - | - | 72.3 |
| - | ✓ | - | - | - | $71.6_{-0.7}$ |
| - | - | ✓ | - | - | $74.3_{+2.0}$ |
| ✓ | - | ✓ | ✓ | - | $73.2_{+0.9}$ |
| ✓ | ✓ | ✓ | ✓ | - | $73.4_{+1.1}$ |
| ✓ | ✓ | ✓ | - | ✓ | $74.6_{+2.3}$ |

**Table 4**. Comparison with attention-based methods on ImageNet dataset.

|  | Params | GFLOPs | Top-1(%) | Top-5 (%) |
|---|---|---|---|---|
| MobileNetV2[1] | 3.50M | 0.32 | 72.3 | 90.0 |
| +SE[13] | 3.56M | 0.32 | 73.5 | 90.5 |
| +CBAM[14] | 3.57M | 0.32 | 73.6 | 90.5 |
| +SK[24] | 5.28M | 0.35 | 74.1 | 91.8 |
| +CoordAttention[9] | 3.95M | 0.33 | 74.3 | **91.9** |
| +Meta-Pooling(Ours) | 3.91M | 0.32 | **74.6** | 91.8 |

acceptable improvements by 2.3%, which is comparable to CoordAttention [9].

**Different Width Multiplier.** We adopt three weight multipliers, which are $\{1.0, 0.75, 0.5\}$. As shown in Tab.2, Meta-Pooling achieves better trade-off compared with SE and CBAM under different width multipliers. However, we found that the diminishing improvement brought by Meta-Pooling as the model becomes smaller, which may be due to the difficulty of optimizing the large receptive field for smaller models.

### 3.3. Ablation Study

The ablation study of the spatial enhancers and learning strategy is shown in Tab.3, which exhibits the effectiveness of our approach. We observe that a larger receptive field can indeed improve the performance of lightweight networks, which . For example, simply using global can achieve a 2% improvement over the baseline. Besides, our proposed PMPL strategy has an improvement of 1.2% compared with the random selection of spatial enhancers.

## 4. CONCLUSION

We present Meta-Pooling to make the receptive field learnable for lightweight networks. By introducing the parameterized spatial enhancer, the Meta-Pooling can model receptive fields of diverse sizes. The Progressive Meta-Pooling Learning strategy with early stopping is employed to optimize the parameterized spatial enhancer. The results demonstrate that our Meta-Pooling makes up for the lack of receptive field and achieve remarkable performance.

# 5. REFERENCES

[1] Mark Sandler, Andrew G. Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, 2018.

[2] Ningning Ma, Xiangyu Zhang, Haitao Zheng, and Jian Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *ECCV*, 2018.

[3] Zhuang Liu, Hanzi Mao, Chaozheng Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie, "A convnet for the 2020s," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11966–11976, 2022.

[4] Xiaohan Ding, X. Zhang, Yi Zhou, Jungong Han, Guiguang Ding, and Jian Sun, "Scaling up your kernels to 31×31: Revisiting large kernel design in cnns," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11953–11965, 2022.

[5] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition.," in *ICLR*, Yoshua Bengio and Yann LeCun, Eds., 2015.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.

[7] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, pp. 1904–1916, 2015.

[8] Qibin Hou, Li Zhang, Ming-Ming Cheng, and Jiashi Feng, "Strip pooling: Rethinking spatial pooling for scene parsing," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4002–4011, 2020.

[9] Qibin Hou, Daquan Zhou, and Jiashi Feng, "Coordinate attention for efficient mobile network design," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13708–13717, 2021.

[10] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei, "Deformable convolutional networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 764–773, 2017.

[11] Mingxing Tan and Quoc V Le, "Mixconv: Mixed depthwise convolutional kernels," *arXiv preprint arXiv:1907.09595*, 2019.

[12] Liu Xiaolong, Li Lujun, Li Chao, and Anbang Yao, "Norm: Knowledge distillation via n-to-one representation matching," in *ICLR*, 2023.

[13] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu, "Squeeze-and-excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, pp. 2011–2023, 2020.

[14] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In-So Kweon, "Cbam: Convolutional block attention module," in *ECCV*, 2018.

[15] Zihang Dai, Hanxiao Liu, Quoc V Le, and Mingxing Tan, "Coatnet: Marrying convolution and attention for all data sizes," *Advances in Neural Information Processing Systems*, vol. 34, pp. 3965–3977, 2021.

[16] Chelsea Finn, Pieter Abbeel, and Sergey Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1126–1135.

[17] Takato Otsuzuki, Heon Song, Seiichi Uchida, and Hideaki Hayashi, "Meta-learning of pooling layers for character recognition," *ArXiv*, vol. abs/2103.09528, 2021.

[18] Faraz Saeedan, Nicolas Weber, Michael Goesele, and Stefan Roth, "Detail-preserving pooling in deep networks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9108–9116, 2018.

[19] Jiaojiao Zhao and Cees G. M. Snoek, "Liftpool: Bidirectional convnet pooling," *ArXiv*, vol. abs/2104.00996, 2021.

[20] Chen-Yu Lee, Patrick W. Gallagher, and Zhuowen Tu, "Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree," in *AISTATS*, 2016.

[21] Takumi Kobayashi, "Gaussian-based pooling for convolutional neural networks," in *NeurIPS*, 2019.

[22] Richard Zhang, "Making convolutional networks shift-invariant again," *ArXiv*, vol. abs/1904.11486, 2019.

[23] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1106–1114, 2012.

[24] Xiang Li, Wenhai Wang, Xiaolin Hu, and Jian Yang, "Selective kernel networks," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 510–519, 2019.