DynConv-ResNet: Enhancing ResNet with Dynamic Convolution

Vikrant Bhati, Kamal Sherawat
vikrant@vt.edu, kamals@vt.edu

1 Problem Statement

Convolutional Neural Networks (CNNs) have been successful in computer vision tasks due to their ability to extract hierarchical features. However, traditional convolution layers are static and apply the same convolutional kernels to every input regardless of context. This static feature of Convolutional Neural Networks (CNNs) [LeCun et al., 1998] in architectures like ResNet representation, can limit the adaptability and generalization ability of CNNs—especially for tasks that require some level of context sensitivity. To improve their performance, networks are generally made deeper or wider which introduces significant computational costs.

We aim to address this issue using Dynamic Convolutional Neural Networks (DCNNs) [Chen et al., 2020] which solve this issue by adjusting convolutional kernels to each input, making DCNNs more context-aware and enabling adaptable hierarchical feature extraction. One form of DCNN that we plan to use is Omni-Dimensional Dynamic Convolution (ODConv) [Shi et al., 2022], which applies learnable attention in four dimensions—input channels, output channels, spatial positions, and kernel groups. In this project, we plan to extend the ResNet model to use dynamic convolutional layers where traditional CNN layers will be replaced with ODConv layers to compare its performance against a baseline ResNet while performing tasks such as image classification, image segmentation, and analyzing time-series data using real-world datasets.

2 Approach

2.1 Baselines:

We are planning to use the ResNet model proposed by He et al. [2016] as the baseline to compare against our modified ResNet model with integrated ODConv layers.

2.2 Model Architecture:

Integrate ODConv into a standard ResNet architecture by replacing static convolutions in selected layers with ODConv modules, as proposed by [Shi et al., 2022].

2.3 Tasks and Datasets:

- **Image Classification:** Train both baseline ResNet and ODConv-ResNet on data sets such as ImageNet [Deng et al., 2009].
- **Image Segmentation:** Use COCO [Lin et al., 2014] or Pascal VOC [Everingham et al., 2010] data sets to assess performance in image segmentation tasks.
- **Real-World Application:** Explore the use of ODConv in time series or cross-domain visual tasks to evaluate generalization.

2.4 Evaluation Metrics:

We will evaluate the modifier ResNet model and the baseline model against metrics such as accuracy, mean intersection of the union (mIoU for segmentation) and computational cost (FLOPs, inference time).

3 Expected Experimental Results

We expect the modified convolutional ResNet (ODConv-ResNet) to have greater accuracy, flexibility, and robustness over the baseline ResNet model. The greatest improvements can be expected in tasks such as segmentation, where flexible spatial adaptation is important. Additionally, the ODConv layer is designed to be computationally efficient, so we expect similar inference times with the gain in accuracy.

4 Potential Contributions

- A complete empirical evaluation of ODConv integrated in ResNet for multiple vision tasks.
- Benchmarks showcasing ODConv's performance and efficiency advantages over standard CNNs.
- Initial investigation into application area beyond vision, such as spatio-temporal data or time-series analysis.

References

Yuntao Chen, Zihang Dai, Meng Chen, Li Yuan, Zichao Liu, Han Hu, Jifeng Guo, and Yizhou Wu. 2020. Dynamic convolution: Attention over convolution kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11030–11039.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. IEEE.

Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. 2010. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on* *computer vision and pattern recognition (CVPR)*, pages 770–778.

Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer.

Weihao Shi, Kai Han, Jianyuan Guo, Yehui Tang, Chunjing Xu, and Yunhe Wang. 2022. Omni-dimensional dynamic convolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9020–9030.