Sequentially Aggregated Convolutional Networks

Yiwen Huang*[1] Pinglai Ou*[2] Rihui Wu*[3] Ziyong Feng[4]
[1]Huazhong University of Science and Technology [2]Virginia Tech [3]University of Sydney [4]DeepGlint

IEEE 2019 International Conference on Computer Vision
ICCV 2019 Seoul, Korea

Problem and Motivation

Ever since the groundbreaking success of ResNet [1], **shortcut connections** have been regarded as a key component of modern deep networks. **The aggregation nature of shortcut connections was however often neglected.** *What could we learn from this perspective and how would it help us design new network architectures?*

Related Work

- **Ensembles** of relatively **shallow networks** [2, 3].
- Aggregated transformations [4, 5, 6].

Contribution

- **Aggregation based** convolutional layers.
- **New interpretation** of DenseNet-like architectures.
- **Windowed aggregation**.
- **New SOTA results** on ImageNet [7]. *(compared to other models of similar complexity)*

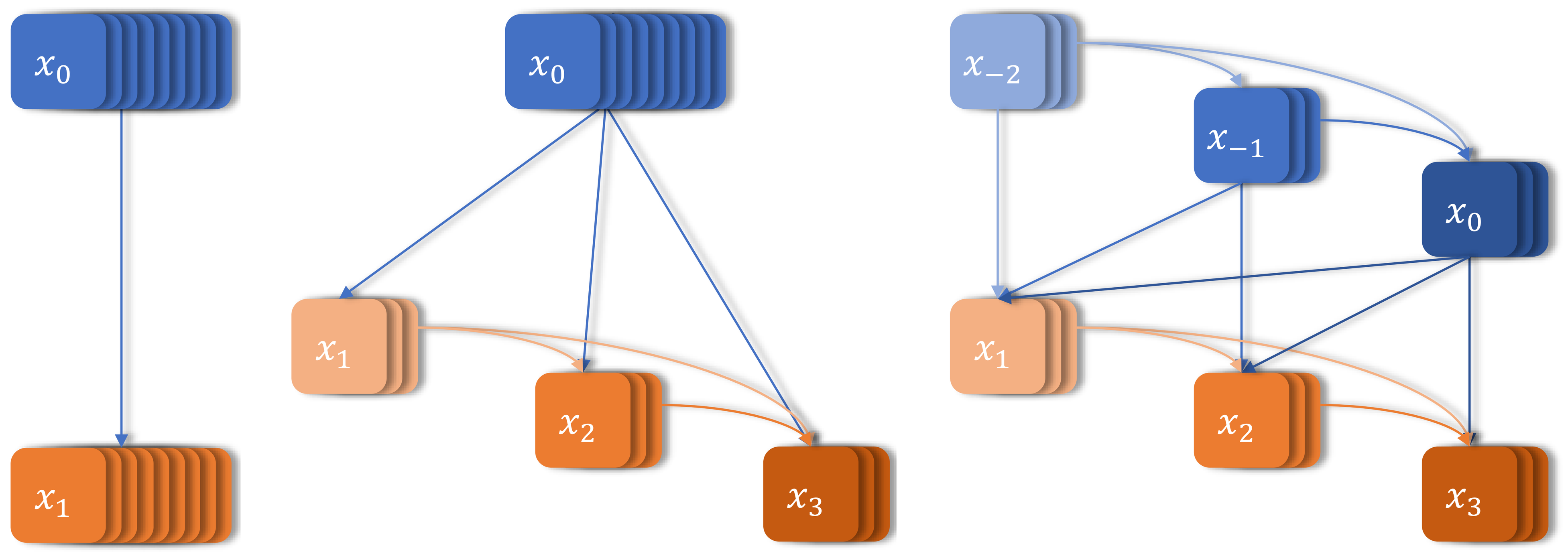
Paper, Code and Data:

https://github.com/GroupOfAlchemists/SeqConv

References

[1] K. He et al. Deep residual learning for image recognition. In: CVPR. 2016.
[2] A. Veit et al. Residual networks behave like ensembles of relatively shallow networks. In: NIPS. 2016.
[3] G. Huang et al. Deep networks with stochastic depth. In: ECCV. 2016.
[4] S. Xie et al. Aggregated residual transformations for deep neural networks. In: CVPR. 2017.
[5] G. Huang et al. Densely connected convolutional networks. In: CVPR. 2017.
[6] H. Liu et al. DARTS: Differentiable Architecture Search. In: ICLR. 2019.
[7] J. Deng et al. Imagenet: a large-scale hierarchical image database. In: CVPR. 2009.
[8] K. He et al. Identity mappings in deep residual networks. In: ECCV. 2016.

Method



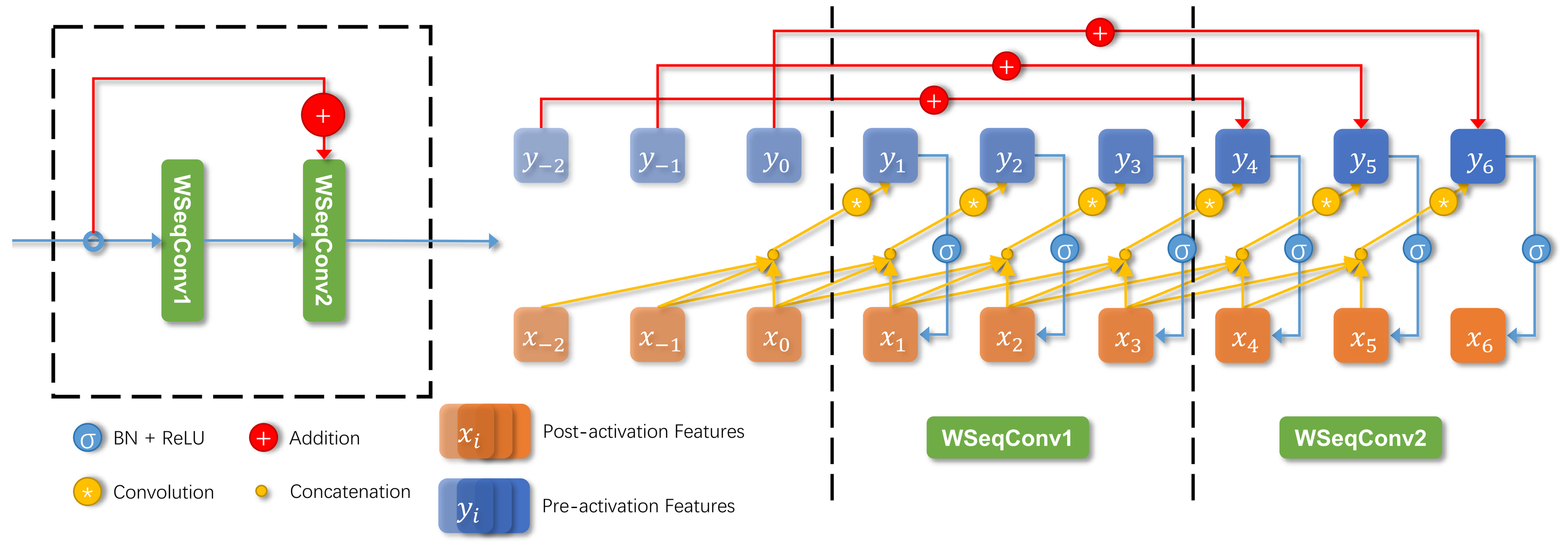
A regular wide convolutional layer (left), a SeqConv layer (middle) and its windowed variant (right).

- **Regular Convolutional Layer**: $$x_1 = F(x_0) \quad (1)$$
- **Basic SeqConv Layer**: $$x_i = F_i([x_0, x_1, \dots, x_{i-1}]) \quad (2)$$
- **Windowed SeqConv Layer**: $$x_i = F_i([x_{i-g'}, x_{i-g'+1}, \dots, x_{i-1}]) \quad (3)$$

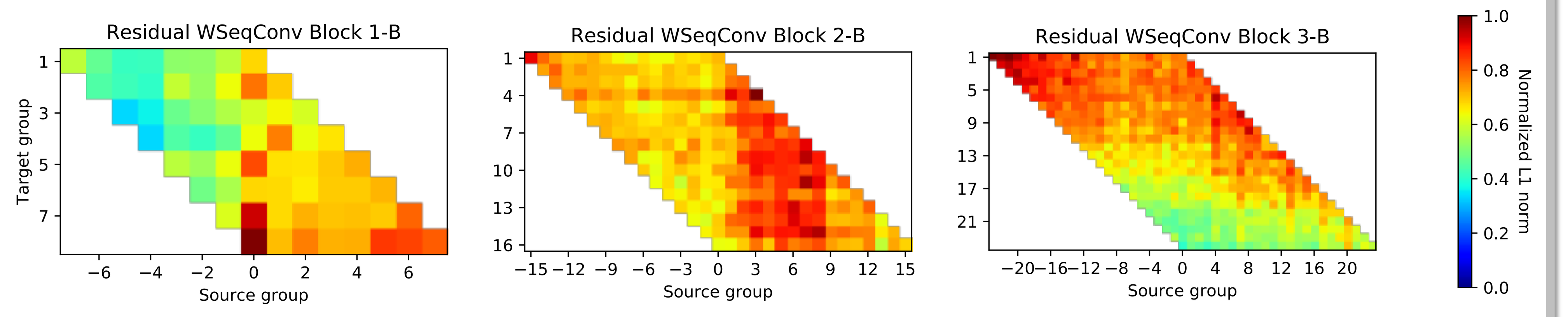
Eq.3 is equivalent to **applying a sliding rectangular window ϕ** across the channel dimension on Eq.2.

Application

SeqConv could be handily integrated into ***any* backbone network**, the following illustration shows a pre-activation residual block [8] with windowed SeqConv layers.



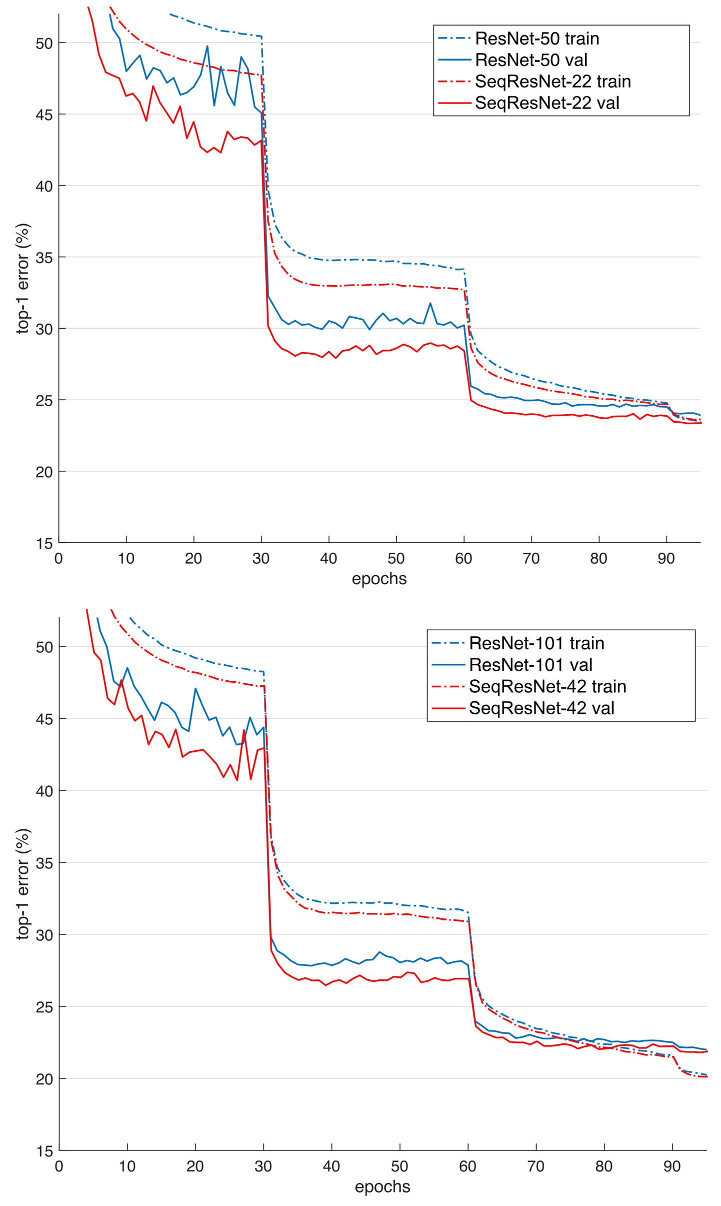
Efficiency of Windowed Aggregation



Weight visualization on trained WSeqConv layers indicates that the windowed aggregation mechanism has ***much higher* overall utilization on aggregated feature groups** than dense aggregation.

Experiments

By simply **replacing each convolutional layer** in ResNet [8] and ResNeXt [4], our SeqConv based models achieve ***SOTA* results** on ImageNet [7] classification.



	# params.	FLOPs	top-1 err	top-5 err
ResNet-101 [1]	44.5M	7.34G	22.44	6.21
DenseNet-264 [5]	33.3M	5.52G	22.15	6.12
ResNet-152	60.2M	10.82G	22.16	6.16
SeqResNet-B42	25.6M	5.33G	**22.06**	**5.98**
SeqResNeXt-24	26.2M	4.32G	**21.92**	**5.82**
ResNet-50	25.6M	3.86G	24.01	7.02
DenseNet-169	14.1M	3.22G	23.80	6.85
SeqResNet-B22	11.8M	2.73G	**23.67**	**6.78**
ResNet-50 *	25.6M	3.86G	23.9	-
SeqResNet-B22 *	11.8M	2.73G	**23.35**	6.68
ResNeXt-50 [4] *	25.0M	4.00G	22.2	-
ResNet-101 *	44.5M	7.34G	22.0	-
SeqResNet-B42 *	25.6M	5.33G	**21.75**	5.89
SeqResNeXt-24 *	26.2M	4.32G	**21.50**	5.73