SUPPRESSION HELPS: LATERAL INHIBITION-INSPIRED CONVOLUTIONAL NEURAL NETWORK FOR IMAGE CLASSIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Convolutional neural networks (CNNs) have become powerful and popular tools since deep learning emerged for image classification in the computer vision field. For better recognition, the dimensions of depth and width have been explored, leading to convolutional neural networks with more layers and more channels. In addition to these factors, neurobiology also suggests the widely existing lateral inhibition (e.g., Mach band effect), which increases the contrast of nearby neuron excitation in the lateral direction, to help recognition. However, such an important mechanism has not been well explored in modern convolutional neural networks. In this paper, we explicitly explore the filter dimension in the lateral direction and propose our lateral inhibition-inspired (LI) design. Our naive design incorporates the low-pass filter, while eliminating the central weight to mimic the inhibition strength decay. The inhibition value is computed from the filtering result of the input, with a simple learnable weight parameter per channel for multiplication to decide the strength. Then the inhibition value is subtracted from the input as suppression, which could increase the contrast to help recognition. We also suggest an alternative using depthwise convolution, as a general form. Our design could work on both the plain convolution and the convolutional block with residual connection, while being compatible with existing modules. Without any channel attention along the channel dimension, the preliminary results demonstrate an absolute improvement of 3.68% and 0.69% over AlexNet and ResNet-18, respectively, in the ImageNet data set, with little increase in parameters, indicating the merits of our design to help feature learning for image classification.

1 INTRODUCTION

In recent years, convolutional neural networks (CNNs) (Hinton et al., 2012; Simonyan & Zisserman, 2015; Szegedy et al., 2015; He et al., 2016) have become powerful and popular tools since deep learning emerged for image classification in the computer vision field. They have archived record-breaking performance versus traditional methods (Quinlan, 1986; Cortes & Vapnik, 1995) with handcrafted features (Dalal & Triggs, 2005; Lowe, 1999) on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Deng et al., 2009). Today, convolutional neural networks still have unique merit, as they have been studied the most, along with the fact that convolution is clear to understand with a strong connection to human visual perception and image processing, which makes them good models for feature learning research.

Different factors have been considered for better recognition performance of convolutional neural networks. VGGNet (Simonyan & Zisserman, 2015) applies a small convolution kernel size of 3×3 for increased depth of network, while ResNet (He et al., 2016) introduces deep residual learning to make training very deep network feasible. The success of such networks indicates that depth dimension is crucial for recognition performance. Wide Residual Networks (Zagoruyko & Komodakis, 2016), on the other hand, demonstrate width as another dimension important to improved performance. In addition to these factors, neurobiology also suggests the widely existing lateral inhibition (e.g., Mach band effect, which is shown in Fig. 1), which increases the contrast of nearby neuron excitation in the lateral direction, as a practical way to help feature learning for better recognition.

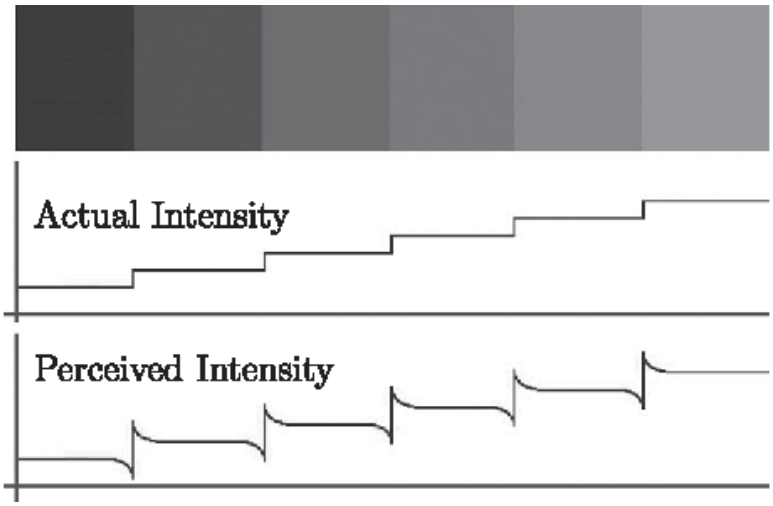

Figure 1: Illustration forMach band effect. Because of lateral inhibition, the actual perception is different from real input, which creates larger contrast close to the boundary, while dark area gets darker and bright area gets brighter.

However, such an important mechanism has not been well explored in the design of modern convolutional neural networks. Although there have been attempts to explore this property, the most influential one (Cao et al., 2018) in recent years relies on fixed rules to compute lateral inhibition using the pre-trained VGGNet model for attention map and weakly supervised saliency detection. Proper incorporation and modeling of inhibition is still needed during the training process for modern convolutional neural networks on large datasets to serve as backbone models.

In this paper, we particularly apply the lateral inhibition mechanism in modern convolutional neural networks, which is inspired by the findings of neurobiology. We explicitly explore the filter dimension in the lateral direction and propose our lateral inhibition-inspired (LI) design. Our naive design incorporates the low-pass filter, while eliminating the central weight to mimic the inhibition strength decay. The inhibition value is computed from the filtering result of the input, with a simple learnable weight parameter per channel for multiplication to decide the strength. Then the inhibition value is subtracted from the input as suppression, which could increase the contrast to help recognition. We also suggest an alternative using depthwise convolution, as a general form. Our design could work on both the plain convolution and the convolutional block with residual connection, while being compatible with existing modules. Without any channel attention along the channel dimension, the preliminary results demonstrate an absolute improvement of 3.68% and 0.69% over AlexNet and ResNet-18, respectively, in the ImageNet data set, with little increase in parameters, indicating the merits of our design to help feature learning for image classification.

Our main contributions can be summarized as:

- To our best knowledge, we are the first to incorporate lateral inhibition into modern convolutional neural networks such as ResNet with little increase in parameters.
- We archive obvious improvements on ImageNet dataset with our lateral inhibition design.
- We propose a novel design using a modified Gaussian low-pass filter for lateral inhibition decay, as well as a more flexible design using depthwise convolution.
- We visualize our learned filter for lateral inhibition in both designs and agree with the neurobiology findings, which could potentially lead to improved network design.

2 Related Work

Lateral Inhibition Convolutional Neural Networks. Cao et al. (2018) propose their LICNN with fixed implementation for lateral inhibition computation uisng pre-trained VGGNet, as top-down step

for attention map and weakly supervised saliency detection. Their inhibition vlaue consists of an average term from the neighborhood and another term only from a higher activation nearby, and this inhibition formula including term coefficients are intuitively selected and unlearnable. The average term calculate the mean vaue within predefined neighborhood (7×7) to protect the inhibition from a high response zone. The other term only takes suppression from any neighbor with higher activation and decay that suppression with distance. After lateral inhibition, nodes with activation below zero will not be considered for another forward feed. This could result in less noise and background removal. Ullah et al. (2021) propose their lateral inhibition design for remote sensing application, with multiple filters in different sizes (3×3 and 5×5) and perform subtraction using the small filter and the large filter for suppression. The limitation is it requires adding more filters to the network for computation to archive lateral inhibition.

3 METHOD

Our method aims to explicitly incorporate the lateral inhibition mechanism into the modern convolutional neural network, so that suppression could be applied from active nodes to neighbors in the lateral direction, increasing contrast and sharpness for more accurate perception, as neurobiology indicates. It is involved in training stage, not as post-processing using pre-trained models.

Our method takes two steps: first generating the lateral inhibition value based on the input and then decreasing that value from the input. We propose naive inhibition to mimic neurobiology findings, and a more flexible alternative as a general form.

3.1 INHIBITION DESIGN

3.1.1 NAIVE DESIGN FOR INHIBITION

We design our naive inhibition according to neurobiology findings. It follows that inhibition of an active neuron decreases as distance goes further from that neuron in the lateral direction. For implementation, we take the exsiting Gaussian low-pass filter with weights decay from the center as the fundation, and eliminate the central weight since no suppression comes from itself. Here, we denote the function of the lateral inhibition value as *LI*, the Gaussian low-pass filter as *G*, the central weight of *G* as *C*, the convolution as *Conv*, the learnable weight *W* with a single parameter per channel for multiplication to decide the inhibition strength, and the input features as *x* before lateral inhibition.

The weights for the 2D Gaussian low-pass filter (*G*) are calculated using Equation 1 (with normalization). Here *u* and *v* denote the spatial coordinates regarding to the center of the filter, and σ denotes the standard deviation which decides the speed of decay (we set filter size as 3 and σ as 1):

$$G(u, v) = \frac{1}{2\pi\sigma^2} e^{-\frac{u^2+v^2}{2\sigma^2}} \tag{1}$$

The lateral inhibition value for the channel *i* is calculated using Equation 2, by eliminating the central weight value C of the Gaussian low-pass filter *G* first, to determine no suppression coefficient for itself, then multiplication with the learnable weight for the channel *i*.

$$LI(x_i) = (Conv(G, x_i) - C * x_i) * W_i \tag{2}$$

3.1.2 FLEXIBLE DESIGN FOR INHIBITION

We also propose a more flexible alternative with depthwise convolution to replace the modified Gaussian low-pass filter as well as the learnable weight for multiplication. Here, we denote the depthwise convolution as *DW*. Therefore, the filter from the depthwise convolution at channel *i* will decide both the shape and the strength of the lateral inhibition function for that channel *i*, as shown in Equation 3. This is a general form.

$$LI(x_i) = Conv(DW_i, x_i) \tag{3}$$

Although we do not manually control how to learn it, visualization of the filter coefficient from the depthwise convolution indicates such inhibition has been learned.

3.2 Output after Inhibition

We denote the output function after inhibition as O, then the output of channel i is the input in channel i with the inhibition value subtracted, as shown in Equation 4.

$$O(x_i) = x_i - LI(x_i) \tag{4}$$

3.3 Inhibition Place

Our lateral inhibition works on plain convolution as well as convolutional blocks with residual connection. For plain convolution, we place it right after that convolution, as shown in Fig. For convolutional blocks with residual connection, we place it on the input flow right after the identity saving operation, as shown in Fig.

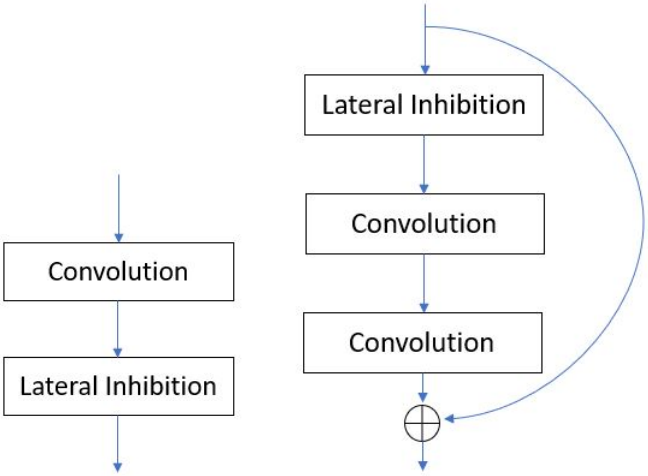

Figure 2: Place for Lateral Inhibition. It is right after the plain convolution, as shown on the left side, or on the input right after the summation of the identity function, as shown on the right side.

4 Experiments

Dataset. We use the ImageNet-1k data set (Russakovsky et al., 2015) to evaluate our proposed methods. ImageNet-1k dataset consists of 1k object categories, with 1.2 million images for training and 50k images for validation.

Experiment Settings. We use the Pytorch (Paszke et al., 2019) deep learning framework and follow the official training example. We adopt standard settings: image crop size of 224 $\times$ 224, initial learning rate of 0.1 (which decays every 30 epochs by a factor of 10), momentum of 0.9, and weight decay of 0.0001. We use standard data augmentation from the training example, which includes random horizontal flipping, with random scale and ratio change in the range of [0.08, 1.0] and [0.75, 1.33], respectively. We train each model with batch size of 128 and batch accumulation of 2, for 100 epochs on 4 GPUs.

Results on ImageNet. Table 1 shows the results of AlexNet (Hinton et al., 2012) and our naive lateral inhibition-inspired (LI) design. This naive design uses the low-pass filter while eliminating the central weight for inhibition decay. It is very clear that our design brings a significant improvement with an absolute value of 3.68% on this large dataset. This demonstrates the effectiveness of our

Table 1: Top-1 accuracy of AlexNet models on ImageNet dataset. LI denotes our lateral inhibition-inspired (LI) design, with a little increase in parameter. Note AlexNet contains only plain convolution.

Method	Top-1 Acc (%)	Params.	Improv. (abs)
AlexNet	51.18	61.10M	-
+LI (our naive low-pass)	54.86	61.10M	3.68

Table 2: Top-1 accuracy of ResNet18 models on ImageNet dataset. LI denotes our lateral inhibition-inspired (LI) design, with a little increase in parameter. Channel attention (SE) serves as a reference.

Method	Top-1 Acc (%)	Params.	Improv. (abs)
ResNet18 (baseline)	70.30	11.69M	-
+SE (Hu et al., 2018)	71.24	11.78M	0.94
+LI (our naive low-pass)	70.99	11.69M	0.69
+LI (our depth conv)	**71.53**	11.70M	**1.23**

lateral inhibition, which obviously benefits feature learning on this network with plain convolution to a large extent.

Table 2 shows the results of ResNet (He et al., 2016) and our lateral inhibition-inspired (LI) design in both naive and flexible versions. With tiny parameter increase (only one parameter per channel for multiplication), our naive design archives an absolute improvement of 0.69% on this large dataset. This is impressive given the reference improvement of 0.94% from Squeeze-and-Excitation (SE) channel attention (Hu et al., 2018), which requires much more parameter increase. Surprisingly, our flexible design has an absolute boost of 1.23%, with a negligible increase in parameter. Here the performance and parameter increase are compared with the baseline ResNet.

Inhibition Filter Weight. We manually check the lateral inhibition filter weight. For our naive design, mostly positive for the value to be multiplied with the modified Gaussian low-pass filter. That weight value mostly is in range of (0, 2), with the setting of filter size of 3 and standard deviation of 1, we could see the inhibition coefficients are positive at the surrounding but negative at the center, given Equation 2. Similarly in the flexible design, we could also spot such pattern in even numbered layers. A classic pattern is provided below in Equation 5, and the inhibition value needs to be substracted from the original input. This pattern is also agreed with neurobiology findings.

$$E = \begin{bmatrix} 0.0582 & 0.1878 & 0.1253 \\ 0.2821 & -0.2401 & 0.2503 \\ 0.1138 & 0.2972 & 0.1011 \end{bmatrix} \tag{5}$$

5 Conclusion

In this paper, we propose our lateral inhibition-inspired design on convolutional neural networks for image classification. We provide a naive design which is based on a Gaussian low-pass filter for inhibition decay and a more flexible alternative using depthwise convolution as a general form. Our design works on both plain convolution and convolutional blocks with residual connection. Preliminary results using AlexNet and ResNet demonstrate obvious improvements with little increase in parameter on ImageNet dataset for image classification. We hope that our research could inspire researchers to reconsider the value of this track for better feature learning.

REFERENCES

Chunshui Cao, Yongzhen Huang, Zilei Wang, Liang Wang, Ninglong Xu, and Tieniu Tan. Lateral inhibition-inspired convolutional neural network for visual attention and saliency detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pp. 886–893. IEEE, 2005.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

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

Geoffrey E Hinton, Alex Krizhevsky, and Ilya Sutskever. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1106–1114, 2012.

Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, 2018.

David G Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pp. 1150–1157. Ieee, 1999.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, and Others. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, 32:8026–8037, 2019.

J Ross Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, and Others. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *Proceedings of 3rd International Conference on Learning Representations*, 2015.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

Ihsan Ullah, Sean Reilly, and Michael G Madden. Enhancing Semantic Segmentation of Aerial Images with Inhibitory Neurons. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 5451–5458. IEEE, 2021.

Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In *Proceedings of the British Machine Vision Conference*, 2016.

A APPENDIX

You may include other additional sections here.