Enhancing Semantic Segmentation of Aerial Images with Inhibitory Neurons

[1'2]Ihsan Ullah, [2]Sean Reilly, and [2]**Michael G Madden**
[1]CeADAR Ireland's Center for applied AI, Technological University Dublin
[2]Machine Learning & Data Mining Research Group, School of Computer Science, National University of Ireland Galway

ICPR 2020 25th International Conference on Pattern Recognition, Milan, Italy 10 | 15 January 2021

NUI Galway OÉ Gaillimh

Abstract

In a Convolutional Neural Network, each neuron in the output feature map takes input from the neurons in its receptive field. This receptive field concept plays a vital role in today's deep neural networks. However, inspired by neuro-biological research, it has been proposed to add inhibitory neurons outside the receptive field, which may enhance the performance of neural network models. In this paper, we begin with deep network architectures such as VGG and ResNet, and propose an approach to add lateral inhibition in each output neuron to reduce its impact on its neighbours, both in fine-tuning pre-trained models and training from scratch. Our experiments show that notable improvements upon prior baseline deep models can be achieved. A key feature of our approach is that it is easy to add to baseline models; it can be adopted in any model containing convolution layers, and we demonstrate its value in applications including object recognition and semantic segmentation of aerial images, where we show state-of-the-art result on the Aeroscape dataset. On semantic segmentation tasks, our enhancement achieves 13.43% higher mIoU than a single baseline model on a single source (the Aeroscape dataset) and 7.03% higher the previous state-of-the art result, which was an ensemble model trained on multiple sources. Our experiments illustrate the potential impact of using inhibitory neurons in deep learning models, and they also show better results than the baseline models that have standard convolutional layers.

Challenge & Objectives

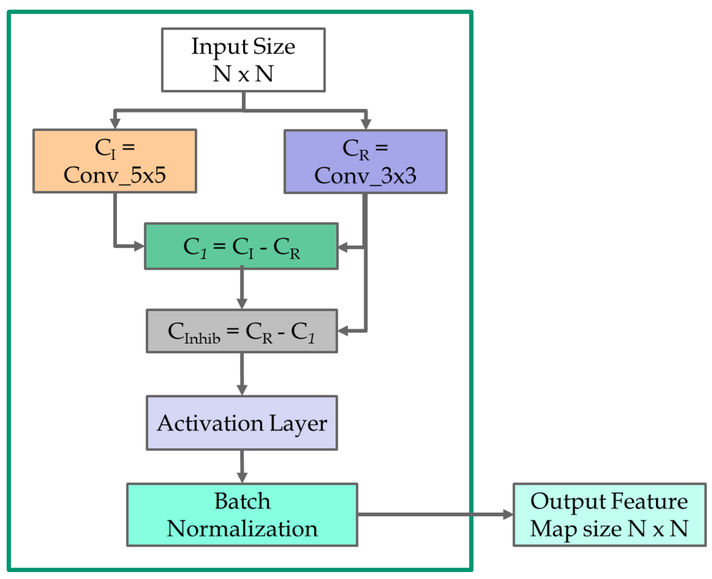
- The challenge:
  - Deep models like VGG and ResNet have millions of parameters
  - Semantic Segmentation needs large amount of labelled data to train: **for hazardous crime scene situations, such data is hard to obtain**
- Fine-tuning and transfer learning can help
  - Make use of models that were trained on less specialized datasets
- Therefore, our objectives are:
  - Develop an approach to semantic segmentation that works better than current state of the art
  - Do not deeply change structures of existing networks
  - Improve speed

Background

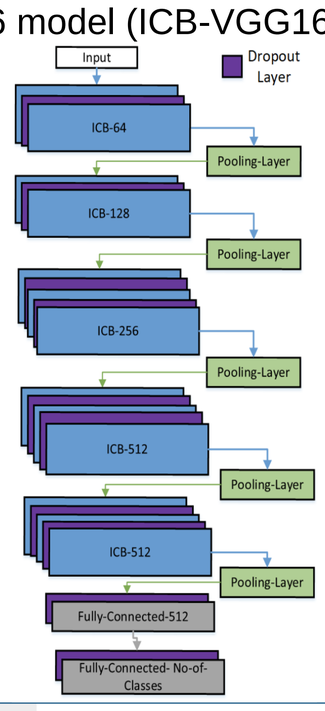
- Lateral Inhibition
  - Capacity of an excited neuron to reduce the activity of its neighbours [1-3]
  - Such neurons are called inhibitory neurons, located immediately outside the receptive field of a neuron
  - Result in an inhibitory field surrounding the receptive field
  - Inhibitory field can improve texture analysis
- LIPNET [4-5]
  - lateral inhibition used in a neural network for image classification and segmentation
  - Bottom-up approach that removes the neighbouring neuron's impact by **subtracting** it
- Others - Contour detection [1], Image segmentation [4], Face detection [5], Recently: deep CNN for saliency detection [6]

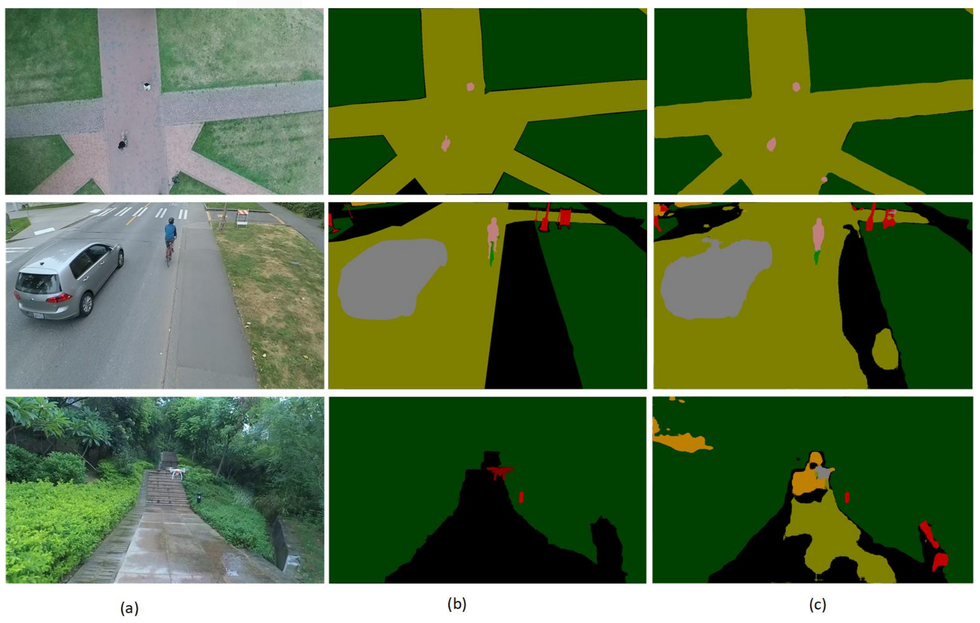
Proposed Inhibitory Convolutional Block

- Inhibitory Convolutional Block (ICB)
- Applicable on all models where we have Convolutional Layer



- Example: Enhanced inhibitory-VGG16 model (ICB-VGG16)





(a) Original Image (b) Ground Truth (c) Predicted

Model	mIoU
SS-Inhi-VGG16-FT1	46
FCN-Imagenet-8s [9]	47.56
SS-VGG16-FFT	51
FCN-Ensemble-SingleSource [9]	53.08
FCN-Ensemble-MultiSource [9]	57.08
SS-ICB-VGG16-FFT	**64.43**

- Comparison of our ICB model vs state-of-the-art on Aeroscape Dataset
- SS, ICB, FT1, and FFT stand for Semantic Segmentation, Inhibitory, Fine tuned on 1 layer, and Fully FT (All layers)

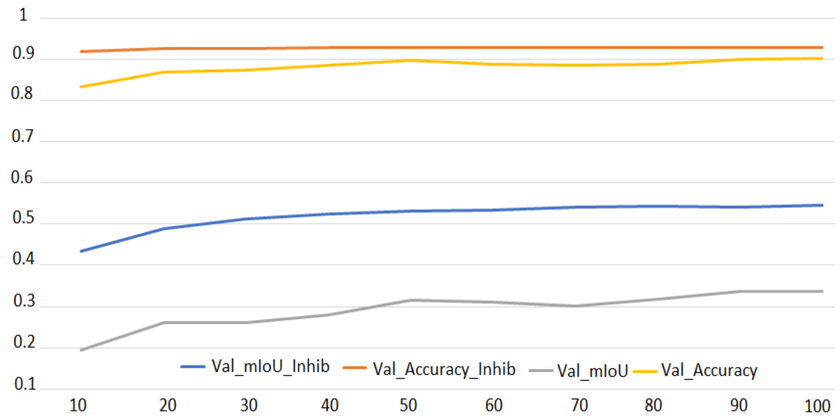
Results

Name of Model	Layers	Accuracy
BL-VGG16*	16	93.43
ICB-VGG16	**16**	**93.62**
BL-ResNet-BS32*	20	90.31
BL-ResNet-BS200*	20	90.6
ICB-ResNet-BS200*	**20**	**91.63**
ICB-ResNet-DO-0.3	20	92.21
ICB-ResNet-DO-0.3-Frac	**20**	**92.91**
ResNet [8]	110	93.57

CIFAR-10 Results, ICB and BL represent inhibition and Baseline, respectively.

Name of Model	Layers	Accuracy
BL-VGG16*	16	69.88
ICB-VGG16	**16**	**71.28**
BL-ResNet-Aver*	20	66.22
ResNet-DO-0.3-Aver	20	66.85
ICB-ResNet-DO-0.3-Aver	20	66.96
ICB-ResNet-DO-0.3-Frac	**20**	**69.48**

CIFAR-100 Results - ICB, Aver, and Frac represent inhibitory model, average, and fractional, respectively



Conclusions

- A general mechanism that can be applied to modify existing CNN architectures in order to introduce inhibitory neurons.
- Facilitates model re-use for a new application that are implemented in Keras and Tensorflow libraries,
- Achieved strong results on object recognition and state-of-the-art results for the aerial view semantic segmentation problem.
- Model re-use and fast learning both lead to lower GPU utilization, reducing energy consumption for the model training process, with consequent environmental benefits

Acknowledgment

We acknowledge funding from **European Union's Horizon 2020** Research and Innovation Programme, under Grant Agreement No. 700264.