

Recurrent Convolutional Neural Network for Object Recognition

Xuewen Yang

July 2 2018

Abstract

In recent years, the convolutional neural network (CNN) has achieved great success in many computer vision tasks. Partially inspired by neuroscience, CNN shares many properties with the visual system of the brain. A prominent difference is that CNN is typically a feed-forward architecture while in the visual system recurrent connections are abundant. Inspired by this fact, we propose a recurrent CNN (RCNN) for object recognition by incorporating recurrent connections into each convolutional layer. Though the input is static, the activities of RCNN units evolve over time so that the activity of each unit is modulated by the activities of its neighboring units. This property enhances the ability of the model to integrate the context information, which is important for object recognition. Like other recurrent neural networks, unfolding the RCNN through time can result in an arbitrarily deep network with a fixed number of parameters. Furthermore, the unfolded network has multiple paths, which can facilitate the learning process. The model is tested on four benchmark object recognition datasets: CIFAR-10, CIFAR-100, MNIST and SVHN. With fewer trainable parameters, RCNN outperforms the state-of-the-art models on all of these datasets. Increasing the number of parameters leads to even better performance. These results demonstrate the advantage of the recurrent structure over purely feed-forward structure for object recognition.

1. Introduction

The past few years have witnessed the bloom of convolutional neural network (CNN) in computer vision. Over many benchmark datasets CNN has substantially advanced the state-of-the-art accuracies of object recognition [3] [1] [4]. For example, after training on 1.2 million images from ImageNet, CNN has achieved better performance than handcraft features by a significant margin in classifying objects into 1000 categories. Furthermore, the pretrained CNN features on this dataset have been transferred to other datasets to achieve remarkable results.

CNN is a type of artificial neural network, which o-



Figure 1. Importance of context for object recognition. Without the context (face), it is hard to recognize the black curve in the middle area as a nose.

riginates from neuroscience dating back to the proposal of the first artificial neuron in 1943. The context is important for object recognition (Figure 1). A feed-forward model can only capture the context (e.g., the face in Figure 1) in higher layers where units have larger RFs, but this information cannot modulate the activities of units in lower layers responsible for recognizing smaller objects (e.g., the nose in Figure 1). To utilize this information, one strategy is to use top-down (or feedback) connections to propagate it downwards, which is adopted in the convolutional deep belief networks (CDBN). In this study, the author takes a different strategy, that is, use recurrent connections within the same layer of deep learning models. It is expected that, equipped with context modulation ability, these lateral connections may boost the performance of deep learning models. In the paper, the author presents a recurrent CNN for static object recognition. The architecture is illustrated in Figure 2, where both feed-forward and recurrent connections have local connectivity and shared weights among different locations. This architecture is very similar to the recurrent multilayer perceptron (RMLP) which is often used for dynamic control [2] [5] (Figure 2). The main difference is that the full connections in RMLP are replaced by shared local connections, just as the difference between MLP and CNN. For this reason, the proposed model is called the recurrent convolutional neural network (RCNN).

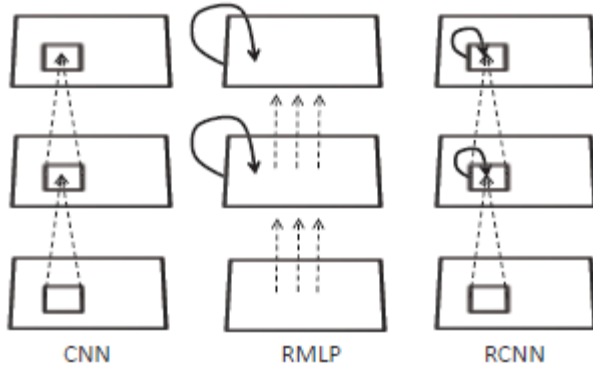


Figure 2. Illustration of the architectures of CNN, RMLP and RCNN. For each model two hidden layers are shown.

2. Conclusion

The author proposed a recurrent convolutional neural network (RCNN) for (static) object recognition. The basic idea was to add recurrent connections within every convolutional layer of the feed-forward CNN. This structure enabled the units to be modulated by other units in the same layer, which enhanced the capability of the CNN to capture statistical regularities in the context of the object. The recurrent connections increased the depth of the original CNN while kept the number of parameters constant by weight sharing between layers. Experimental results demonstrated the advantage of RCNN over CNN for object recognition. Over four benchmark datasets, with fewer parameters RCNN outperformed the state-of-the-art models. Increasing the number of parameters led to even better performance. This work shows that it is possible to boost the performance of CNN by incorporating more facts of the brain. It would be interesting to see other fact of the brain to be integrated into deep learning models in future.

References

- [1] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *Computer Science*, 2014. 1
- [2] B. Fernandez, A. G. Parlos, and W. K. Tsai. Nonlinear dynamic system identification using artificial neural networks (ANNs). In *IJCNN*, 1990. 1
- [3] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *NIPS*, 2012. 1
- [4] M. Lin, Q. Chen, and S. Yan. Network in network. *Computer Science*, 2013. 1
- [5] G. V. Puskorius and L. A. Feldkamp. Neurocontrol of nonlinear dynamical systems with kalman filter trained recurrent networks. *IEEE Transactions on Neural Networks*, 1994. 1