Classifying Ship Images using Deep Convolutional Neural Network

Duc-Cuong Dao
Hanoi University of Science and Technology
Hanoi, Vietnam
duccuong.hust@gmail.com

Hua Xiaohui
Shanghai Jiaotong University
Shanghai, China
sophiexhh@gmail.com

Olivier Morére
Agency for Science, Technology and Reasearch
Singapore
olivier.morere@gmail.com

Antoine Veillard
Agency for Science, Technology and Reasearch
Singapore
antoine.veillard@gmail.com

ABSTRACT

In this work we investigate the architecture, performance of one of the first popularized Convolutional Neural Networks - AlexNet in ship image classification task. The main point distinguishing our work from the existing is that we focus on classificationt task for ship images which is a promising application in marine industries, habour management and malitary defense. [TODO: add more]

Keywords

Deep learning, convolutional neural networks, image classification

1. INTRODUCTION

It is needless to say how important of image classification/recognition is in the field of computer vision - image recognition is essential for bridging the huge semantic gap between an image, which is simply a scatter of pixels to untrained computers, and the object it presents. Therefore, there have been extensive research efforts on developing effecitive visual object classifiers. .... feature extraction. In traditional approaches, image features (e.g., SIFT [Lowe, 2004]) are carefully hand-crafted (i.e., fixed by the engineers). This introduces a serious drawback when it comes to recoginizing/classifying natural patterns because natural features exist in variety of forms, shapes, directions, etc. ... will requires an enormously large amount of engineered features which turns out to be impossible in reality.

.... Deep-learning methods address the problem of learning hierarchical representations of features. They allow the computer to read raw data (e.g., image, text, speech, etc.) and *automatically* discover the representations of those data with multiple levels of abstractions that needed for recognition/classification tasks.

SoICT 2015 Hue, Vietnam

2. RELATED WORK

Starting with LeNet-5 [LeCun et al., 1990], ConvNets have typically had a standard structure - stacked convolutional layers (optionally followed by contrast normalization and max-pooling) are followed by one or more fully-connected layers. Variants of this basic design are prevalent in the image classification literature and have yielded the best results to-date on MNIST, CIFAR and most notably on the ImageNet classification challenge. For larger datasets such as Imagenet, the recent trend has been to increase the number of layer and layer size, while using dropout [Srivastava et al., 2014] to address the problem of overfitting.

Despite the concerns that max-pooling layers result in loss of accurate spatial information, the same convolutional network archiecture as [Tompson et al., 2014] has aslo been sucessfully employed for localization, object detection and human post estimation. Inspire by a neuroscience model of the primate visual cortex, .. use a series of fixed Gabor filters of different sizes in order to handle multiple scales, similarly to the Incepttion model. However, contrary to the fixed 2-layer model of [15], all filters in the Inception model are learned. Furthermore, Inception layers are repeated many time, leading to a 22-layer deep model in the case of the GoogLeNet model.

Network-in-Network is an approach proposed by [Lin et al., 2013] in order to increase the respresentational power of neural newtworks. When appliead to convolutional layers, the method could be viewd as addition 1×1 convolutional layers follwed typically by the rectified linear activation.

3. BACKGROUND

In this section, we present a rough overview of deep learning and its hallmark achievements in various types of machine learning problems. After that, we provide a brief introduction to a typical ConvNet named AlexNet [Krizhevsky et al., 2012] and its training strategy. Although ConvNets can be trained using unsupervised learning, we concentrate on the supervised learning strategy of training as our task (image classification) falls into this family.

3.1 Deep Learning

Deep-learning methods allow a machine to read raw data (e.g., pixel values of an image) and *automatically* discover the representations needed for dectection or classification

[Yann LeCun, 2015].

Deep-learing models are composed of multiple processing layers, each can be considered as a non-linear feature transformation. Higher layers represents higher level of abstraction.

3.2 Convolutional Neural Networks

ConvNets are hierarchical neural networks whose convolutional layers alternate with subsampling layers, reminiscent of simple and complex ceels in the primary visual cortex. ConvNets vary in how convolutional and subsample layer are realized and how they are trained.

3.2.1 Convolutional Layer

A convolutional layer is parametrized by the size and the number of maps, kernel sizes, skipping factors and the connection table. Each layer has M maps of equal size (M_x, M_y). A kernel of size (K_x, K_y) is shifted over the valid region of the input image (i.e. the kernel has be completely inside the image). The skipping factors S_x and S_y define how many pixels the filter/kernel skips in x- and y-direction between subsequent convolutions. The size of the output maps is then defined as:

$$M_x^n = \frac{M_x^{n-1} - K_x^n}{S_x^n + 1} + 1; M_y^n = \frac{M_y^{n-1} - K_y^n}{S_y^n + 1} + 1 \qquad (1)$$

where index n indicates the layer. Each map in layer L^n is connected to at most M^{n-1}. Neurons of a given map share their weights but have different receptive fields.

3.2.2 Pooling Layer

The biggest architectural difference between our implementation and the ConvNet of [leCun et al] is the use of a max-pooling layer instead of a sub-smapling layer. No such layer is used by [Simard et al..] who simply skips nearby pixels prior to convolution, instead of pooling or averaging.

3.2.3 Classification Layer

4. EXPERIMENTS

In this section we give a detailed description of all the experiments we performed. We trained the AlexNet with the same architecture described in [Krizhevsky et al., 2012]. [...]

4.1 The Dataset

We used a crawler to collect ship images from different sources along with their meta information (name, category, etc.).

4.2 Training AlexNet

AlexNet introduced by [Krizhevsky et al., 2012] was the first work that popularized ConvNets in Computer Vision. It had shown outstanding performance on the ImageNet ILSVRC challenge in 2012 with top-5 error of 16% (TODO: add top-1) compared to runner-up with 26% error. The network has 60 million parameters and 500,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and two globally connected layers with a final 1000-way softmax. The detail architecture of AlexNet is depicted in Fig. 1

4.3 Results

Table 1: Detail descriptions of VGGNet's layers

Layer	Type	Maps and Neurons	Kernel
0	input		

5. CONCLUSIONS

6. ACKNOWLEDGMENTS

We would like to thank Assoc. Prof., Stéphane Bressan at School of Computing, National University of Singapore and Dr. Antoine Veillard at Image Persuasive Lab (IPAL), Insitute of Infocomm Research (I2R), A*STAR for insightful supervision and valuable discussions throughout the work of this paper.

References

A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel. Handwritten digit recognition with a back-propagation network. In D. Touretzky, editor, *Advances in Neural Information Processing Systems 2*, pages 396–404. Morgan-Kaufmann, 1990.

M. Lin, Q. Chen, and S. Yan. Network in network. *CoRR*, abs/1312.4400, 2013. URL `http://arxiv.org/abs/1312.4400`.

D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60 (2):91–110, Nov. 2004. ISSN 0920-5691. doi: 10.1023/B:VISI.0000029664.99615.94.

N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014. URL `http://jmlr.org/papers/v15/srivastava14a.html`.

J. Tompson, A. Jain, Y. LeCun, and C. Bregler. Joint training of a convolutional network and a graphical model for human pose estimation. *CoRR*, abs/1406.2984, 2014. URL `http://arxiv.org/abs/1406.2984`.

G. H. Yann LeCun, Yoshua Bengio. Deep learing. *Nature Insight Review*, 521(5):436–444, May 2015.