



Computer Vision Crew

김도윤(12기/산업경영공학부14) / 조장

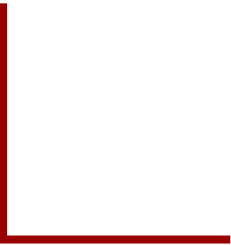
권형근(11기/컴퓨터학과16)

최동재(12기/산업경영공학부16)

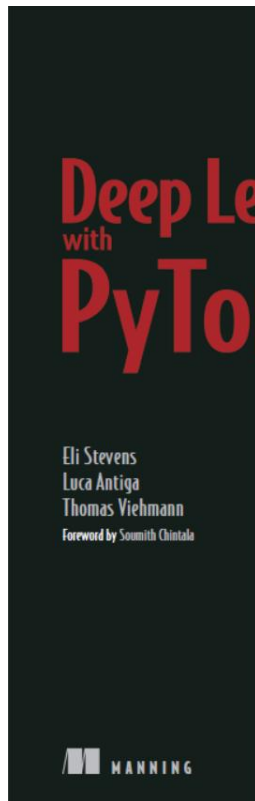




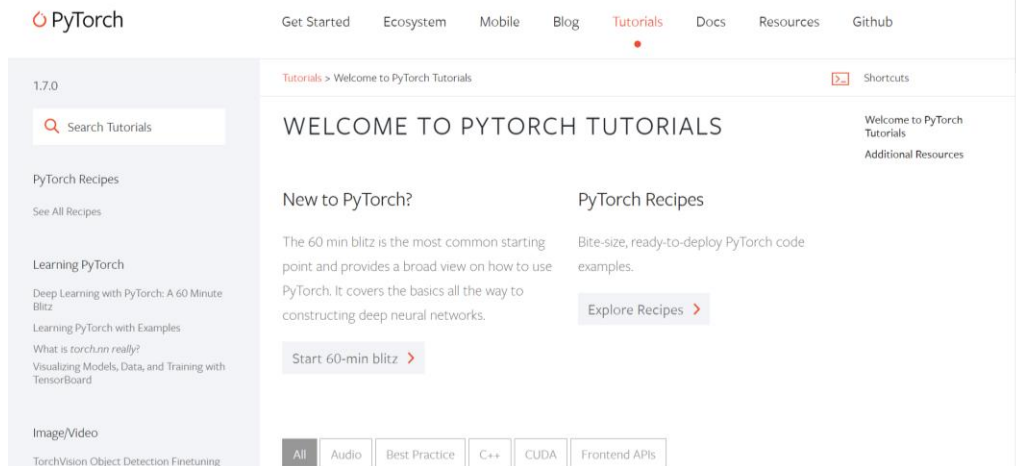
Contents

1. Deeplearning with PyTorch
 2. Paper Review
 3. Computer Vision Project
- 

1. Deep Learning with PyTorch



- 딥러닝 모델 구축을 위한 필수 모듈, PyTorch
- 이미지/영상, 자연어, 음성 데이터 처리 최적화
- PyTorch 공식 홈페이지에서 직접 제공
- PyTorch 공식 Tutorial 병행



2. Paper Review

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet ILSVRC-2012 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used **data augmentation** and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called **"dropout"** that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.



1 Introduction

Current approaches to object recognition make essential use of machine learning methods. To improve their performance, we can collect larger datasets, learn more powerful models, and use better techniques for preventing overfitting. Until recently, datasets of labeled images were relatively small — on the order of tens of thousands of Caltech-101 [6], Caltech-101/256 [9], and CIFAR-10/100 [12]. Simple recognition tasks can be solved quite well with datasets of this size, especially if they are augmented with **label-preserving transformations**. For example, the current-best error rate on the MNIST digit-recognition task (c.0.3%) approaches human performance [4]. But objects in realistic settings exhibit considerable variability, so to learn to recognize them it is necessary to use much larger training sets. And indeed, the shortcomings of small image datasets have been widely recognized (e.g., Pinto et al. [23]), but it has only recently become possible to collect labeled datasets with millions of images. **The new larger datasets include** **LargeNIST [23]**, which consists of hundreds of thousands of **fully-segmented images**, and **ImageNet [6]**, which consists of over 15 million labeled high-resolution images in over 22,000 categories.

To learn about thousands of objects from millions of images, **we need a model with a large learning capacity**. However, the immense complexity of the object recognition task means that this problem cannot be specified even by a dataset as large as ImageNet, so **our model should also have lots of prior knowledge**, to compensate for all the data we don't have. Convolutional neural networks (CNNs) combine one such class of models [16, 11, 13, 18, 15, 22, 26]. Their capacity can be controlled by **varying the depth and breadth**, and they also make strong and mostly correct assumptions about the nature of images (namely, stationarity of statistics and locality of pixel dependencies). Thus, compared to standard feedforward neural networks with similarly-sized layers, CNNs have much fewer connections and parameters and so they are easier to train, while their theoretically-best performance is likely to be only slightly worse.

Published as a conference paper at ICLR 2015

VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION

Karen Simonyan^{*} & Andrew Zisserman^{*}
Visual Geometry Group, Department of Engineering Science, University of Oxford
{karen, andy}@robots.ox.ac.uk

ABSTRACT

In this work we investigate the effect of the convolutional network depth on its accuracy in the large-scale image recognition setting. Our main contribution is a thorough evaluation of networks of increasing depth using an architecture with very small (3×3) convolutional filters, which shows that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 16-19 weight layers. These findings were the basis of our ImageNet Challenge 2014 submission, where our team secured the first and the second places in the localization and classification tracks respectively. We also show that our representations **generalize well to other datasets**, where they achieve state-of-the-art results. We have made our two best-performing ConvNet models publicly available to facilitate further research on the use of deep visual representations in computer vision.

1 INTRODUCTION

Convolutional networks (ConvNets) have recently enjoyed a great success in large-scale image and video recognition (Krizhevsky et al. 2012, Oquab & Fupria, 2013, Simonyan et al. 2014, **Simonyan & Zisserman 2015**), which has become possible due to the large public image repositories, such as ImageNet (Ding et al. 2009), and high-performance computing systems, such as GPUs or large-scale distributed clusters (Lian et al. 2012). In particular, an important role in the advance of deep visual-recognition architectures has been played by the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al. 2014), which has served as a testbed for a few generations of large-scale image recognition systems, from high-dimensional shallow feature encoders (Poggio et al. 2010) (the winner of ILSVRC-2011) to deep ConvNets (Krizhevsky et al. 2012) (the winner of ILSVRC-2012).

With ConvNets becoming more of a commodity in the computer vision field, a number of attempts have been made to improve the original architecture of Krizhevsky et al. (2012) in a bid to achieve the state-of-the-art performance in the Net-performing submissions to the ILSVRC-2013 (Oquab & Fupria, 2013, Simonyan et al. 2014) **involving smaller receptive window size and smaller stride of the first convolutional layer**. Another line of improvement dealt with training and testing the networks densely over the whole image and over multiple scales (Girshick et al. 2014, He et al. 2014). In this paper, we address another important aspect of ConvNet architecture design — **its depth**. To this end we fix other parameters of the architecture, and steadily increase the **depth of the network by adding more convolutional layers**, which is feasible due to the use of very large (2×2) convolution filters in all layers.

As a result, we come up with significantly more accurate ConvNet architectures, which not only achieve the state-of-the-art accuracy on ILSVRC classification and localization tasks, but are also **applicable to other image recognition datasets, where they achieve excellent performance even when used as a part of a multi-stage pipeline (e.g. deep feature classification by linear SVM without fine-tuning)**. We have released our two best-performing models^(*) to facilitate further research.

The rest of the paper is organized as follows. In Sect. 2 we describe our ConvNet configurations. The details of the image classification training and evaluation are then presented in Sect. 3, and the ^{*}former affiliation: Google DeepMind ^{*}current affiliation: University of Oxford and Google DeepMind http://www.robots.ox.ac.uk/~vgg/research/very_deep/

- 유명 CNN, Object Detection 모델, GANs 관련 논문 리뷰
- Stanford University cs231n 강의 영상 참고

3. Computer Vision Project

Search

Dataset

FER2018

NowYMS • updated 2 years ago (Version 1)

Data Tasks Notebooks (10) Discussion Activity Metadata

Download (287 MB) New Notebook

Usability 2.9 License Data files © Original Authors Tags No tags yet

Data Explorer

28713 MB

README

fer2013.bib

fer2013.csv

ferSubmission.csv

Summary

4 files

4 columns

< README (476 B)

If you use this dataset in your research work, please cite

"Challenges in Representation Learning: A report on three machine learning contests." I Goodfellow, D Erhan, PL Carrier, A Courville, M Mirza, B Hamner, W Cukierski, Y Tang, DH Lee, Y Zhou, C Ramaliah, F Feng, R Li, X Wang, D Athanasakis, J Shave-Taylor, M Mikolov, J Park, R Ionescu, M Popescu, C Grozea, J Bergeret, J Xie, L Romaszko, B Xu, Z Chuang, and Y. Bengio. arXiv 2013.

See fer2013.bib for a bibtex entry.

- PyTorch를 이용한 프로젝트 진행 예정 (주제 미정)
e.g. 표정에 따른 감정 판별, 상품의 브랜드 마크 탐지 등



Thank you