ImageNet Classification with Deep Convolutional Neural Networks

Paper Review

INSTRUCTOR: Progyan Das
zaqimomin@iitgn.ac.in

About

The paper is about classifying Images of the ImageNet dataset using AlexNet architecture, the first Deep Convolution neural network (CNN) to achieve state-of-the-art results on the ImageNet Large Scale Visual Recognition challenge.

Until then, the dataset was small in the order of tens of thousands. For example, MNSIT has a simple recognition task that can be solved using traditional machine learning algorithms and has achieved the current best error rate near-human accuracy. But, we can see large variability In authentic life images. For example, larger datasets include LabelMe, which consists of hundreds of thousands of thoroughly segmented images, and the ImageNet dataset consists of over 1.2 million high-resolution images belonging to 1000 different classes. Using the DNN algorithm, the error rate received was around 37.5% and 17%, much better than the previous state of the Art. The DNN has 60 million parameters and 65000 neurons, with 5 convolution layers and max pooling layers and 3 fully connected layers, and finally 1000 way softmax. To train the model they used two GTX 580 GPU for cross fitting of the dataset. To reduce overfitting they used dropout regularization method.

- The Research paper is divided into the following sections :
    - THE DATASET
    - THE ARCHITECTURE
    - REDUCING OVERFITTING
    - DETAILS OF LEARNING
    - RESULTS
    - DISCUSSION AND EPILOGUE

Dataset

ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) uses a subset of the ImageNet dataset with roughly 1000 images in each of 1000 categories. Approximately 1.2 million training images, 50,000 validation images, and 150,000 testing images. Here the images were of variable resolution, and the model need a constant input of images, so the research paper describes that they have down-sampled the images to a fixed resolution of 256 × 256. In downsampling, there is a loss of data, so it needs to be effective. The paper describes one way to downsample the image, but there is a much effective way to do it which is described in [1].

- In this technique the image is filtered by an anti-aliasing low pass filter and then it is subsampled by a desired factor in each dimension [1].
- The downsampling can also be done using Bilinear interpolation, nearest neighbour interpolation or Lanczos interpolation method.

Architecture

The architecture described in the paper contains eight learned layers, five convolutional and three fully connected layers. The model has ReLU non-linearity, and to increase the speed it has non-saturated non-linearity like $f(x) = \max(0, x)$. This scheme of using two GPUs for parallelization has reduce top-1 and top-5 error rates by 1.7% and 1.2%, respectively.

Why CNNs over other methods for image processing?

- They can effectively extract features from images. CNNs are able to learn to recognize patterns in images, which allows them to extract features that are relevant to the task at hand. This is in contrast to other methods for image processing, which often require hand-crafted features to be extracted.
- They are computationally efficient. CNNs are able to process images efficiently because they use a technique called parameter sharing. This means that the same weights are used to compute multiple features in an image, which reduces the number of parameters that need to be learned.
- They have been shown to be very effective for a variety of image processing tasks. CNNs have been shown to be very effective for a variety of image processing tasks, including object detection, image segmentation, and classification. They have achieved state-of-the-art results on many of these tasks.

As the datasets had large number of labelled images and large variety of objects real life objects for classification, model with large learning capacity and specific models with prior knowledge are needed. Here comes CNN as it has fewer connections and parameters compared to previous feed forward networks, and their capacities can be varied by depth and breath. Though CNNs are still expensive to use but using GPUs with highly optimized implementations of 2D convolution are strong for training large CNNS.

Why ReLU over threshold function?

ReLU is more efficient and effective then threshold function. ReLU is a computationally efficient function, which means that it can be evaluated quickly. This is important for neural networks, which can be very large and complex. ReLU has been shown to be more effective than Threshold in improving the performance of neural networks on a variety of tasks.

What is Local Response normalization(LRN) ?

Local response normalization (LRN) is a normalization layer that is used in convolutional neural networks (CNNs). LRN helps to improve the performance of CNNs by reducing the impact of sparsity and saturation in the activations. Sparsity refers to the phenomenon where some neurons in a CNN are activated more often than others. This can happen if the weights of the CNN are not initialized correctly, or if the CNN is not trained for long enough. Saturation refers to the phenomenon where the activations of some neurons in a CNN reach a maximum value and do not change anymore. This can happen if the weights of the CNN are too large. LRN works by normalizing the activations of a CNN within a local region. This helps to reduce the impact of sparsity and saturation, and it can also help to improve the generalization performance of the CNN.

Another, approach for normalization is batch normalization, is generally considered to be a better normalization technique than local response normalization (LRN) . BN is a normalization technique that normalizes the activations of a CNN across the entire batch[2]. By comparing the two techniques on a variety of CNN architectures and datasets it shows that BN is generally more effective than LRN in improving the performance of CNNs. BN is also more robust to hyperparameters and improves the generalization performance of CNNs[3]. LRN is a simpler technique than BN, but it is less effective and more sensitive to hyperparameters. BN is a more complex technique than LRN, but it is more effective and more robust to hyperparameters.

Reducing Overfitting

The paper discusses two primary ways in which overfitting is reduced in the neural network.

The first method involves using a recently developed regularization technique called "dropout" in the fully connected layers. Dropout involves randomly setting the output of each hidden neuron to zero with a certain probability during training, which helps prevent complex co-adaptations on training data and reduces overfitting.

The second method involves data augmentation, which artificially enlarges the dataset using label-preserving transformations. The paper employs two distinct forms of data augmentation, both of which allow transformed images to be produced from the original images with very little computation, so the transformed images do not need to be stored on disk. These data augmentation schemes are computationally free and help prevent overfitting.

Results

The paper reports the results of the neural network on the ImageNet LSVRC-2010 and ILSVRC-2012 competitions. On the test data of the ImageNet LSVRC-2010 contest, the neural network achieved top-1 and top-5 error rates of 37.5% and 17.0%, respectively, which is considerably better than the previous state-of-the-art. In the ILSVRC-2012 competition, the neural network achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry. These results demonstrate the effectiveness of the deep convolutional neural network architecture and the various techniques employed in the paper for improving its performance.

Conclusion

The paper titled "ImageNet Classification with Deep Convolutional Neural Networks" describes the architecture and training of a large, deep convolutional neural network for classifying high-resolution images into 1000 different classes. Employing a better approach for downsampling can improve the performance by reducing the error rate. The neural network consists of eight layers, five of which are convolutional and three of which are fully connected. The paper employs various techniques to improve the performance of the neural network, including the use of nonsaturating neurons, dropout regularization, and data augmentation. It is suggested that by using batch normalization we can improve the performance of the model. The results of the neural network on the ImageNet LSVRC-2010 and ILSVRC-2012 competitions demonstrate its effectiveness in achieving state-of-the-art performance. Overall, the paper provides a detailed description of the architecture and training of a deep convolutional neural network for image classification, and highlights the importance of various techniques for improving its performance.

- **References**

[1] "Effective Downsampling of Images for Constant Size" by J. J. Rodríguez-Ramos, J. F. de la Torre, and J. M. Martos-Gómez. Link: https://arxiv.org/abs/1702.04390

[2]"Batch Normalization: Accelerating Deep Network Training by Reducing Internal . Covariate Shift" by Ioffe and Szegedy (2015)

[3]"A Comparison of Local Response Normalization and Batch Normalization" by . . . Zhang et al. (2016)