LaTeX Author Guidelines for CVPR Proceedings

First Author
Institution1
Institution1 address
firstauthor@i1.org

Second Author
Institution2
First line of institution2 address
secondauthor@i2.org

Abstract

The ABSTRACT is to be in fully-justified italicized text, at the top of the left-hand column, below the author and affiliation information. Use the word "Abstract" as the title, in 12-point Times, boldface type, centered relative to the column, initially capitalized. The abstract is to be in 10-point, single-spaced type. Leave two blank lines after the Abstract, then begin the main text. Look at previous CVPR abstracts to get a feel for style and length.

1. Introduction

Please follow the steps outlined below when submitting your manuscript to the IEEE Computer Society Press. This style guide now has several important modifications (for example, you are no longer warned against the use of sticky tape to attach your artwork to the paper), so all authors should read this new version.

2. Background

Neural networks can be used for image classification. With the appropriate back propagation (BP) algorithm and the right loss fiction the network eventually will be able to conjure weights that encapsulate some pattern that can differentiate between different classes. neural networks do have a disadvantage since the image have to be flattened to be used in neural network all the spatial features have been lost and the network will have to be bigger to obtain high performance. But with convolutional neural network (CNN) the spatial features can be maintained and work with CNN. and that ability to gives CNN the upper hand comparing it to a neural network [3].

Higher performance in terms of accuracy can be achieved by using a ResNet architecture. the more complex the pattern the more complex the model needs to be for it to approximate the pattern. in deep learning a more complex model equates (not always but generally true) to a model with more parameters. the number of filters and the size can be increased the meet the demand complexity of the data. unfortunately in bigger CNN models gradients often get smaller and smaller as the algorithm progresses down to the lower layers. As a result, the Gradient Descent update leaves the lower layer connection weights virtually unchanged, and training never converges to a good solution. in that instances that's called gradient vanishing and when the opposite happens its gradient exploding problem [1]. Such problems can be mitigated using different activation functions with using clever initialization techniques such as Glorot and He Initialization. Alternatively with the introduction of ResNet architecture allowed for training of large and deep CNN models by using the residual block [2].

3. Methodology

3.1. CNN

The main idea behind CNN is using kernels as a combination of estimators. the BP algorithm shall find the right combination of filters or kernels that can abstract and learn high to low level features from the data. to illustrate the idea clearly in figure 1 the input image have been blurred by applying a blur kernel. the input image can be considered 2D matrix assuming the image has been normalized by multiplying each 3*3 subset of with a 3*3 kernel as illustrated in figure 2.

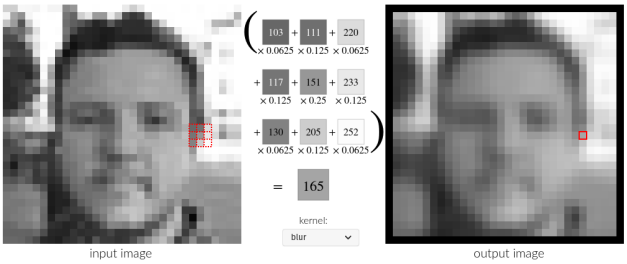

Figure 1. Blur kernel

In CNN the values for the kernels are unknown that's what the BP algorithm is trying to obtain. Setting the size of the kernel or the filters and the numbers of the kernels

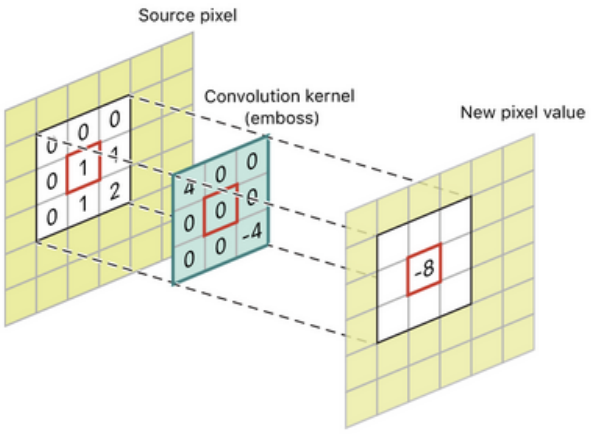

Figure 2. applying a kernel

the BP algorithm using Gradient Descent like algorithm will hopefully converge into set of filters that can detect features then using neural network at the of the CNN network for classification.

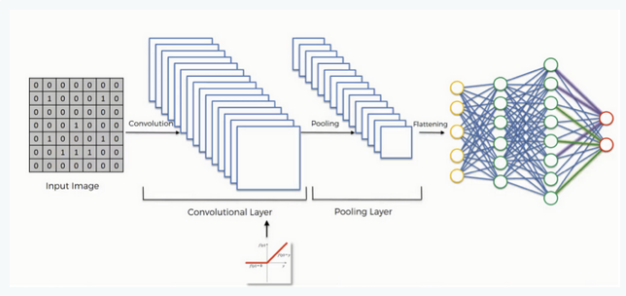

Figure 3. CNN Network

3.2. VGGNet

The VGGNet was developed by Simonyan and Zisserman [5]. The VGG network is famous for its simplicity. it consists of (3*3) convolution blocks stacked on top of each others and as you traverse down the network stack the number of filters increases in figure 4.

Training VGG network is extremely difficult and tedious. The original authors had to utilize pre-training to train smaller size of VGG and then use that as a backbone for bigger VGG networks. that was the only way at that time to train VGG network assuming using random initialization for the weights. but Xavier and Yoshua [1] & Dmytro and Jiri Matas [4] demonstrated that such training scheme isn't a must. Using an appropriate initialization technique and swapping to the non-saturating activation function can eliminate the need for pre-training VGG or deep neural network in general.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 4. VGG architectures

3.2.1 Mini VGG Implementation

The original VGGNet was designed for ImageNet competition. the image size was 224 * 224 which is different from CIFAR-10 images. my implementation is similar to VGG11 as seen if Figure 4 with smaller number of filters for each VGG block. other layer such as drop out and batch normalization have been utilized to combat over-fitting and stabilize the network.

References

[1] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. volume 9 of *Proceedings of Machine Learning Research*, pages 249–256, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010. JMLR Workshop and Conference Proceedings.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[3] Yann LeCun, Patrick Haffner, Léon Bottou, and Yoshua Bengio. Object recognition with gradient-based learning, 1999. International Workshop on Shape, Contour and Grouping in Computer Vision ; Conference date: 26-05-1998 Through 29-05-1998.

[4] Dmytro Mishkin and Jiri Matas. All you need is a good init, 2015.

[5] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2014.