

Dynamical Change of the Perceiving Properties of Neural Networks as Training with Noise and Its Impact on Pattern Recognition

Roman Nemkov

Department of Information Systems & Technologies, North-Caucasus Federal University, 2, Kulakov Prospect, Stavropol, Russian Federation

nemkov.roman@yandex.ru

Abstract. General parameters of convolutional networks (kernels) are set in the learning process. Also in addition to the method of training the quantity of information that is passed through the kernel influences the quality of setting. This quantity of information depends on the size of training sample and the concentration of receptive fields. You can increase the concentration of the receptive fields for a fixed training set size due to the multilayer coating of arbitrary maps with fields of different types, that will be equivalent to the use of noisy training sample. This can increase the networks performance in the test.

Keywords: convolutional networks, training with noise, different types of receptive fields.

1 Introduction

To date, problem of invariance is the main and yet unsolved problem of pattern recognition: the same object may have substantially different from each other external characteristic (shape, colour, texture, etc.) as well as it may be differently displayed on the retina (view from different angles), which greatly complicates its classification. This global problem within the neural network technologies can be solved by to create big training sets [1]. If creating such sets are difficult then this sets are expanded by the addition of noise. [4–6]. A neural network can be regarded as a pyramidal hierarchical graph then noise can be created by changing communications between nodes in a graph [2, 3] or changing perceiving properties in nodes of a graph [8]. Convolutional neural networks (CNNs) have three perceiving properties in a node-neuron: a receptive field (RF), an activation function and a method for producing a weighted sum (simple weighted sum or higher-order polynomial). A change of RFs is the easiest and most promising way of the three. The same pattern can be differently perceived by changing the RFs, that leads to the creation of noise. The influence of noise (which was created due to changing the shape of the fields) is reinvestigated on the pattern recognition in this article.

2 The Generation of Noise by Changing the Receptive Fields

Training with noise in the context of gradient descent for a neural network can be written as

$$\frac{\partial E}{\partial w} + \varepsilon = \nabla, \quad (1)$$

where $\frac{\partial E}{\partial w}$ is the gradient vector from the network's weights, ε is the additional noisy component corrects the gradient vector. The gradient vector after a correction can't exactly point to the local minimum, but training with noise has two benefits: generalization ability increases and local minimums can be better overcome during the gradient descent. When you change perceiving properties in the nodes of networks you have the same training with noise, where ε can be explicitly written:

$$\varepsilon = \frac{\partial E}{\partial w}(\text{new_perception}) - \frac{\partial E}{\partial w}(\text{standart_perception}), \quad (2)$$

where $\frac{\partial E}{\partial w}(\text{new_perception})$ is the gradient vector from weights (the perceiving properties have been already changed), $\frac{\partial E}{\partial w}(\text{standart_perception})$ is the gradient vector from weights with the standart perceiving properties (RFs have square form).

The changing of perceiving properties in the nodes of a network occur due to changing the shape of the RFs. Each element of RF has neighbors. The neighbors are elements located in the one or two discrete steps from current element. Thus, the value of current element (within RF) can be replaced by the neighboring value. If you do this operation for all elements of RF then weighted sum will be changed, hence output of neuron will be also changed. The replacement is shown on the Fig. 1.

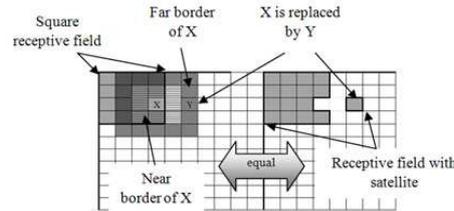


Fig. 1. The replacement of neuron-pixel X by Y with the help of changing the shape of the receptive field (the receptive field with the satellite).

The map (which is the input for the current neurons) receives another covering of RFs, but the kernels of convolution layer (which passed through themselves

this new covering) remain the same. The quantity of information affecting the kernels increases and the kernels can extract the best invariant features. This process is shown in Fig. 2.

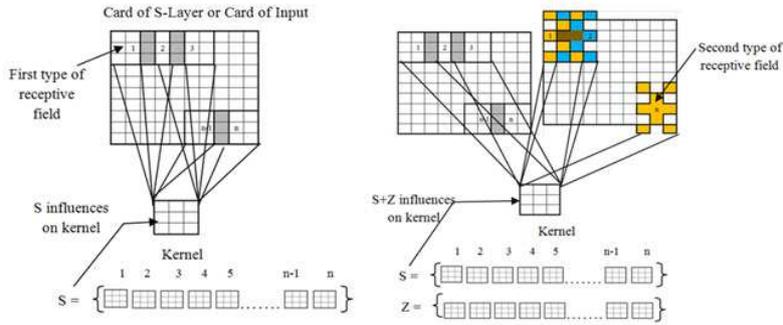


Fig. 2. Discrete perception of information with the help of convolution layer (C-Layer) (Left). The same perception, but with the help of C-Layer with different receptive fields. The pattern will be perceived by the first type of receptive field in the first stroke. In the second stroke the same pattern will be perceived by the second type of receptive field (Right).

This technology can expand the training set by the patterns which are created depending on where (how far away from the current element) elements of RFs take the information for a replacement. Let all elements of RFs are replaced then additional training sets, which will be obtained by this technology, is shown in Fig. 3.

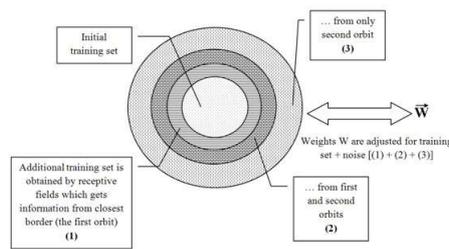


Fig. 3. Additional sets which will be obtained by the changing of RFs.

Any convolutional layers may have their coverings, hence the change of perception can be on the different layers. If CNN has three convolutional layers

then the quantity of combinations (or “refracting prisms”) will be $2^3-1=7$ (one “prism” is a standard perception). The unique pattern will be obtained within the frames of each scheme-“prism”. A strategy is also important for a marking the particular layer using RFs. There are two opposing strategies: the RF is chosen by random way and is superimposed on desired location or the same type of RF with a specific index is superimposed on all desired locations. The second strategy can model the primitive affine transformations if the RF simulates a shift for all its elements.

Patterns need create with the combinations of all “refracting prisms”, with using the both strategies, with using the RFs which are fully updated for maximum coverage of any of the three additional sets.

3 Experiment

MNIST was chosen for experiments with noise. This is due to the fact, that the most schemes of the creating of noise have been tested in this set. The architecture of CNN is shown in Fig. 4.

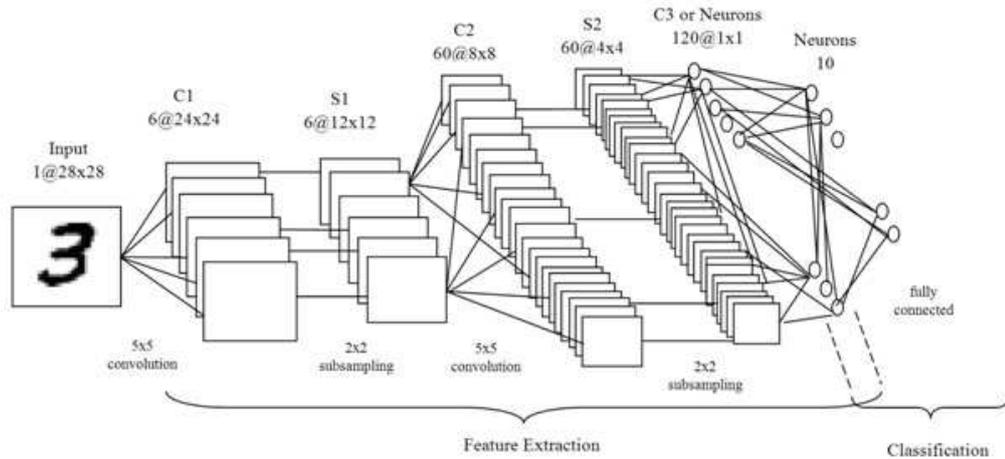


Fig. 4. The architecture of CNN for the work with MNIST set.

The simplest algorithm of gradient descent (without momentum, weights decay and other tricks) has been used for maximum simplicity and repeatability of the experiment. The initial value of the learning rate (η) is equal to 0.005, after every 100 epochs new value are obtained from the old value by multiplying by 0.3. Error function is mean-square error (MSE). The pattern is recognized if the error on the output layer does not exceed the value of 0.001. Pools of RFs for convolutional layers are shown in Figure 5.

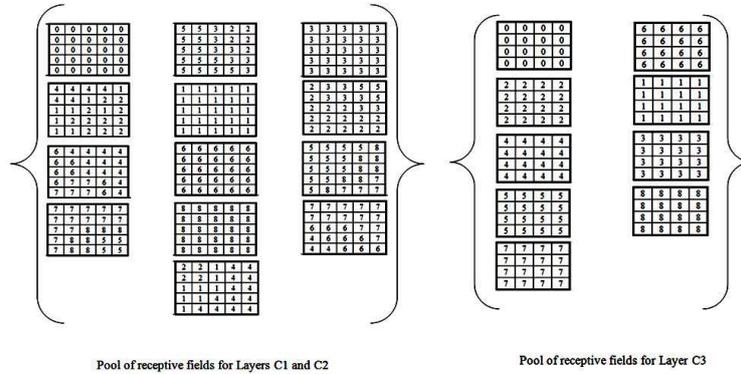


Fig. 5. Pools of RFs for convolutional layers.

The RF with arbitrary index is set in the proper position by the strategy of markup. Geometrical interpretation of the index for shift is shown in Fig. 6.

9	10	11	12	13
14	1	2	3	15
16	4	0	5	17
18	6	7	8	19
20	21	22	23	24

Fig. 6. The geometric interpretation of index for the shift of element of RF.

Thus, the noise from the first additional training set (Fig. 3, set (1)) was used. Comparative results are given in Table 1.

Table 1. The comparison between different learning algorithms

Algorithm	Distortion	Error	Ref.
2 layer MLP (MSE)	affine	1.6%	[4]
SVM	affine	1.4%	[5]
Tangent dist.	affine+thick	1.1%	[4]
LeNet-5 (MSE)	affine	0.8%	[4]
2 layer MLP (MSE)	elastic	0.9%	[6]
CNN (MSE)	this distortions	1.2%	this paper
Best result	elastic	0.23%	[7]

This is a good result that has been achieved without the involvement of additional noise from the sets (2) or (3) (Fig. 3). Research has shown that the

change of perceiving properties in the nodes of CNN can effectively expand the training set and reduce the error of generalization. Also, this technology is easy compatible with the elastic distortions and dropconnects or dropouts.

References

- [1] Dean, J., Corrado, G.S., Monga, R., Chen, K., Devin, M., Le, Q.V., Mao, M.Z., Ranzato, M.A., Senior, A., Tucker, P., Yang, K., Ng, A. Y.: Large Scale Distributed Deep Networks. NIPS, 2012.
- [2] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Improving Neural Networks by Preventing Co-adaptation of Feature Detectors. CoRR, 2012.
- [3] Wan, L., Zeiler, M. D., Zhang, S., LeCun, Y., Fergus, R.: Regularization of Neural Networks using DropConnect. ICML 3, volume 28 of JMLR Proceedings, page 1058-1062.
- [4] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based Learning Applied to Document Recognition. Proceedings of the IEEE. v. 86, pp. 2278-2324, 1998.
- [5] Decoste, D., Scholkopf, B.: Training Invariant Support Vector Machines. Machine Learning Journal. vol 46, No 1-3, 2002.
- [6] Simard, P., Steinkraus, D., Platt, J.C.: Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. ICDAR, page 958-962. IEEE Computer Society, 2003.
- [7] Ciresan, D., Meier, U., Schmidhuber, J.: Multicolumn Deep Neural Networks for Image Classification. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), CVPR '12, pages 3642-3649, Washington, DC, USA, 2012. IEEE Computer Society. ISBN 978-1-4673-1226-4.
- [8] Nemkov, R., Mezentseva, O.: The Use of Convolutional Neural Networks with Non-specific Receptive Fields. The 4-th International Scientific Conference " Applied Natural Science 2013", Novy Smokovec, High Tatras, Slovak Republic, Oktober 2-4, 2013, p.148.