FEATURE EXTRACTION FROM IMAGES WITH USE OF CONVOLUTIONAL NEURAL NETWORKS: APPLICATION TO SECURING PERSONAL DATA

T. R. Sokhin, K. V. Mironov

Department of Computer Science and Robotics
Ufa State Aviation Technical University
Ufa, Russia
e-mail: qwinpin@gmail.com, : mironovconst@gmail.com

Abstract**:** *The task of recognizing the bank cards on images is considered in this paper. This task is motivated by premoderation of images in social networks in order to avoid leakage of personal data. An example of such data is the number of bank card. Detection algorithm based on convolutional neural networks is applied for this task. A number of experiments were conducted to develop the most optimal neural network architecture, which took into account the speed and accuracy of recognition. In the course of the experiments, the recognition accuracy of 91% and the processing speed of 25 images per second were achieved.*

1. Introduction s

With the development of network technologies in recent years, the task of monitoring content has become increasingly important. Large companies often use traffic monitoring systems to avoid sending sensitive corporate information over open communication channels. In private life, it is also important to keep track of the content that users put on publicly available resources, such as social networks. The main threat in this case is the unwanted publication of personal data. In this paper, the threat of accidental publication of a bank card image in a social network is considered. In 2016, a study was conducted on the Visa payment system at Newcastle University, which concluded that the application of a distributed bruteforce attack allows to receive a CVV card in a matter of seconds may be sufficient to steal from it all available funds.

Thus, the ability to verify photos for having bank cards on them could improve the security of personal data in social networks. Such a check can be built into the content premoderation system. In this paper, we propose an algorithm for recognizing bank cards in photographs.

2. Related works

One of the simplest methods, which allows to significantly reduce the complexity of the image recognition problem is the method of contour analysis. A contour is a curve that describes the boundary of an object in an image. Using this approach assumes that the contour contains enough information about the shape of the object, while internal points are not taken into account. Examining only the contours of objects allows you to transition from the image space to the contour space, which significantly reduces the complexity of algorithms and calculations. The main advantage of contour analysis is the invariance with respect to rotation, scale and contour displacement on the image under test. It is best suited for finding an object of some given shape [2,3].

The next one is a relatively simple method of a comparison with a template (template matching). This method is used to find areas of images that are most similar to some given pattern. The size of the template should be smaller than the size of the image being scanned. Search for a template is made by sequentially moving it along the tested image and assessing the match of each new area with the template. Based on the results of such check, the area that has the highest matching coefficient is selected. In fact, this is the percentage of matches between the area of the picture and the template.

Searching by the template does not allow to say with certainty whether the original object was found, since this is a probabilistic characteristic, depending on the scale, viewing angles, picture rotations, and the presence of physical interference. Also, false triggering of the algorithm is possible, when the object is not actually found, but there are some common details between the template and the area in the image under test. Of course, such a situation can be avoided by checking the value of the matching coefficient (so that it is not less than some boundary limit), but this does not always work properly due to the reasons described above [2]. The disadvantages of this method include the time of operation, which increases depending on the size of the incoming image.

The idea of searching by key points is that the most significant points are highlighted in the image, which will be preserved even if the size, noise, or lighting would change. This method also has several disadvantages: if one of the elements in the image to be examined is absent or will be greatly changed in the list of key points, the image will not be recognized. In this regard, try to introduce as many similar points as possible, which greatly increases the algorithmic complexity of the method, and, consequently, the recognition time in general.

In the method of Viola-Jones [4], a scanning window approach is used: the image is scanned by the search window (the so-called scan window), and then the classifier is applied to each position. The system of training and selection of the most significant features is fully automated and does not require human intervention, but, despite the considerable accuracy and ease of use, the method has a number of disadvantages that are significant for the algorithm being developed: a relatively long image processing, because it requires a sliding window method across the image many times, with different window scales, and a strong dependence on the size of the image fed to the input.

Neural networks can also be used to recognize objects in images. Now there are many tasks solved with neural networks. It often has no reason – like many young technologies, NN are becoming mainstream. But it is indispensable in area of computer vision. However, not all kinds of NN are useful here; actually, we need to use the principles of visual system of mammals: the reaction to certain details of so-called features that help in object recognition. That NN are called convolutional.

Convolutional neural networks are one of the most recently popular models of deep learning, used primarily for image recognition. The concept of convolutional networks is built on three main ideas [6]:

1) local receptive fields - the recognition of an item in an image should primarily be affected by its immediate environment, while pixels located in another part of the image are most likely not related to this element and not contain information that would help to identify it correctly;

2) shared weights - the same object can be found in any part of the image, and for its search in all parts of the image, the same pattern (set of weights) would be applied;

3) subsampling - when compared with a pattern, not the exact value for a given pixel or area of pixels is taken into account, but its aggregation in some neighborhood, for example, the average or maximum value.

From the mathematical point of view, the basis of convolutional neural networks is the operation of matrix convolution, which

consists in the elementwise multiplication of the matrix, which is a small portion of the original image (for example, 7 * 7 pixels) with a matrix of the same size, called the convolution kernel, and the subsequent summation of the obtained values . In this case, the core of the reconciliation is essentially a certain pattern, and the number resulting from the summation characterizes the degree of similarity of the given area of the image to this template. Accordingly, each layer of the convolutional network consists of a number of patterns, and the task of learning the network is to select the correct values in these templates - so that they reflect the most significant characteristics of the original images. In this case, each pattern is matched sequentially with all parts of the image - it is in this that the idea of dividing the weights finds expression. Layers of this type in a convolutional network are called convolution layers. In addition to convolution layers, there are sub-sampling layers in convolutional networks or sub-sampling, which replace small areas of the image with one number, thereby simultaneously reducing the size of the sample for the next layer and making the network more resistant to small changes in the data. In the last layers of a convolutional network, one or several fully connected layers are usually used, trained to perform a direct classification of objects.

Before 2012, at which CNN demonstrated their abilities at the ImageNet competition [5], this technology was not in demand, because of weak database and expensive hardware. But now, we have high computing power per dollar and millions of databases in free access. The methods of contour analysis [2] (the object presented in the form of the exterior outlines), template matching (finding small parts of an image which a template image) in general and Haar-like features in particular are far behind [3].

3. Architecture

In order to study the technologies and to design information security system, a model of CNN capable to recognize image of a bankcard has been developed. In cases of publication of those in social networks owner is deprived of support even when his money is stolen. Detection of such images allow to take measures to prevent event like that or another threats. To work with NN used TensorFlow with TFLearn API [7], CPU first-generation i7 (only because OpenCL still are not supported enough) and not data of card from 2000 images.

There are first-priority principles for building NN:

•Avoid representational bottlenecks early in network. The size of image was chosen 200x200.

•Increasing the activations per tile allows for more disentangled features. Higher dimensional representations are easier to process locally within a network.

•Spatial aggregation can be done over lower dimensional embeddings without much or any loss in representational power. That promotes faster training without loss in quality, but practice shows decrease in efficiency with dimension reduction on early layers.

•Balance the width and depth of the network. It is more important with high-performance system.

Consideration of several typical architectures led to the ideas of the Inception v3 network: the use of convolutions of various sizes on one source data and their subsequent integration into a single feature map. In the early layers, the correlated neurons concentrate in local areas, which means that if several neurons in the same coordinate can learn about the same thing, then in the tensor after the first layer, their activation will be in a small region near some point. The greatest number of such correlations can be obtained with convolutions 1 to 1. Slightly less with 3 by 3 and even smaller with convolutions 5 by 5. This allows to extract features of the image at various levels of detail while preserving the original image.

Also used the idea of using 1 by 1 filter isolate the core of feature on large area (a convolution of 5 by 5 or 3 by 3 pixels) in the image.

We know convolution of 5 by 5 can be replaced by two convolutions of 3 by 3, which reduces the number of network parameters, and hence the learning and networking time. For example, convolution 5 by 5 contains 25 parameters. If we replace it with a stack of two layers with the contractions 3 by 3, we get the same mapping, but the number of parameters will be less - 18, which is 22% less. Next, we can proceed to asymmetric convolution 1xn and nx1 type, but this factorization does not work well on early layers, it gives very good results on medium grid-sizes (On m×m feature maps, where m ranges between 12 and 20), so it is useless in this case. For reducing overfitting in neural networks in final layers we used dropout [8] - at each stage of training, some nodes are ignored with a probability of 0.5.

3. Experimental results

Two different structures of neural networks were developed. One of them used only successive convolutional layers, terminating in two fully connected layers in fig. 1. The second network used layers from several parallel convolutions and merged the results in fig. 2.

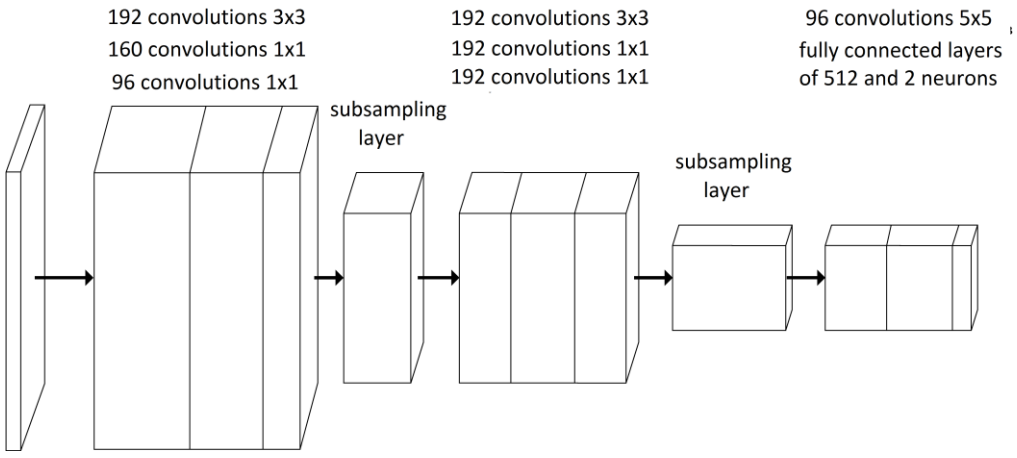

Fig. 1. *"Classical" architecture of the CNN applied for the recognition of bank cards.*

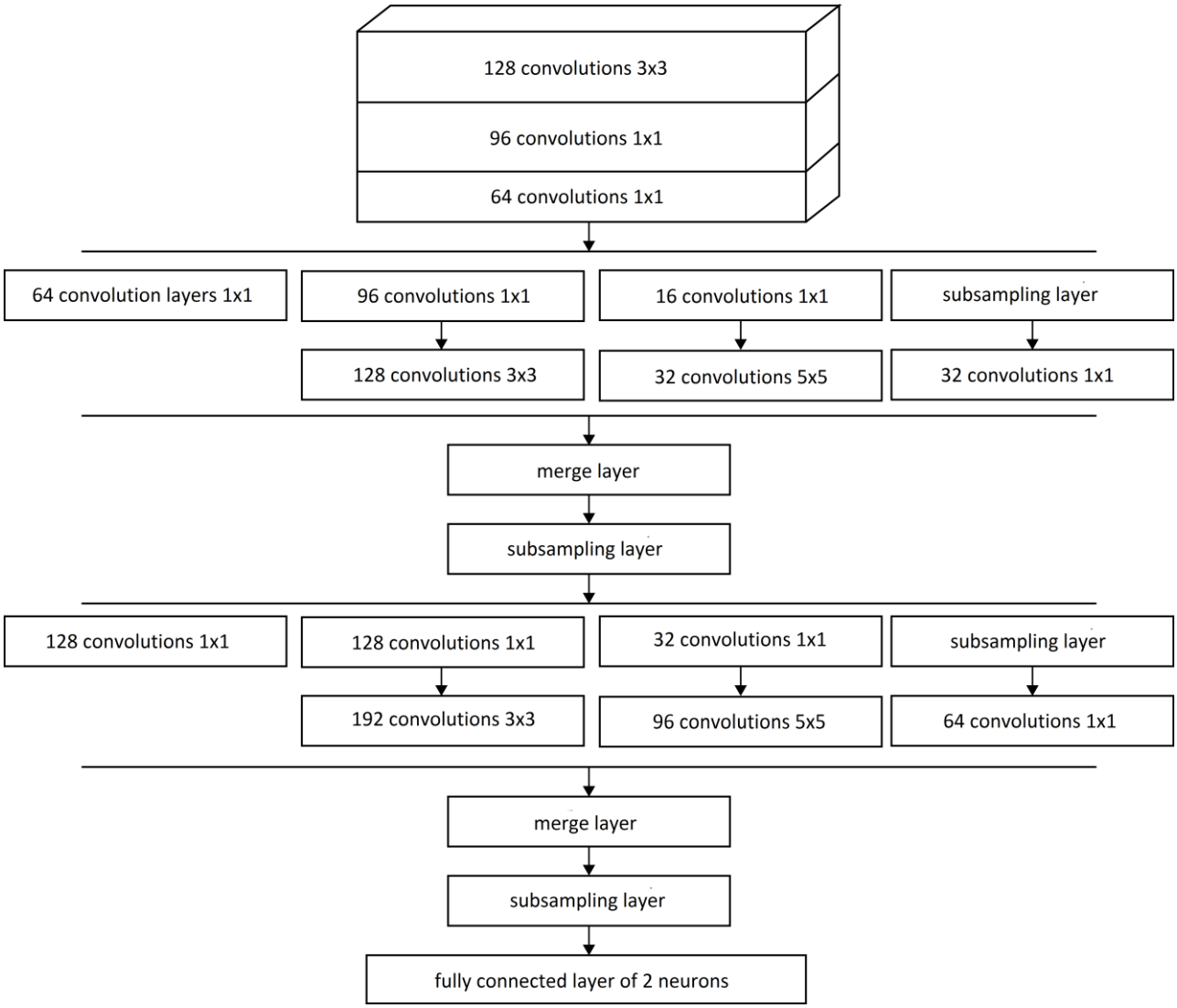

Fig. 2. *CNN based on "Inception" model*

The training sample of images includes 2000 photos in jpg format. The test sample images has 1000 photos in jpg format. The input of the network is supplied with images of 32 by 32 pixels in 3 color RGB channels. At the output of CNN, matrices containing the characteristics of the original image are obtained, then these matrices are transferred to the fully connected layers - perceptrons - from 512 and 2 neurons, whose output indicates the neural network's confidence in the object's belonging to the image of one of the classes. In this case it is "Is a bank card" and "Is not a bank card".

In order to achieve maximum efficiency, the training was conducted with the following parameters:

- Learning rate;

- Dropout percentage - the percentage of training images that will not be used to train the weights of the neural network.

Turning to the work of Jan LeCun in 1988 [6], which proposes the idea of convolution and pulling using the method of learning the backward propagation of an error, we obtained a classical "flat" convolutional neural network, depicted in Figure 1. The final accuracy of the neural network is 70.89 %. This neural network demonstrates good results, however, there are signs of retraining during testing - low accuracy of detection of bank cards with designs that differ from given ones.

The best result is 79,64% with recognize speed about 21 image per second with structure in fig. 2. The decision to use neural networks, which consist of elements whose functionality is analogous to most functions of a biological neuron, first originated in the 30s of the last century (Pitts, Rashevsky, Turing). However, only recently they have found practical application, including in the field of information security. The developed model includes the most successful solutions obtained from the existing experimental database of researchers from around the world. Based on these techniques, the best architecture of the convolutional neural network and training parameters were empirically obtained, allowing to achieve a high result, taking into account the existing limitations on processing speed and processing power. The already existing practice of introducing neural networks into social networks, other resources on the Internet demonstrates that the ratio of implementation costs and support to the benefits obtained is extremely positive.

The best result of training obtained with the Inception model, is 91.35% with an average processing time of 1000 images in 47 seconds, which is almost 1.5 times better than the results of the previous model. This model can be used to process a large number of images with high accuracy.

7. Conclusion

The task of protecting personal data has one of the highest values in the field of information security at the current time. Bank cards, which are so common among all segments of the population, have many vulnerabilities that allow them to gain access to funds through them. Together with the possibility of an easy access to the Internet and social networks in particular, even for those who do not have sufficient knowledge of the simplest precautions, in order to ensure the security of their personal data, this becomes a wide-spread problem.

It is possible to protect this vector of the distribution of personal data through the participation of the social networks themselves, which is why the possibility of premodering the content that the user publishes is offered. To enable fast enough moderation of images with minimal increase in equipment costs and corresponding costs, it was necessary to select an image recognition algorithm that combines high speed with accuracy of recognition. As a result, it was decided to use neural networks. Convolutional neural networks are used to work with images.

Due to the fact that there is no full theory of architecture development and training of neural networks, a significant part of the research work relied on the method of experimental study, during which the most optimal architecture of a convolutional neural network was developed, which ensures the accuracy of bank cards recognition on the image in 91% and speed recognition of 1000 images in 47 seconds.

Acknowledgements

The research work is supported by the Russian Fund for Basic Research, grant #16-07-00243.

References

1. M. A. Ali, B. Arief, M. Emms, A. van Moorsel. "Does the Online Card Payment Landscape Unwittingly Facilitate Fraud?" IEEE Security & Privacy, vol. 15, No. 2, pp.78 to 86, 2016.
2. ArealIdea "Analysis of Computer Vision Algorithms" Web-page. https://arealidea.ru/articles/stati-i-publikatsii/analiz-algoritmov-kompyuternogo-zreniya-poiska-obektov-i-sravneniya-izobrazheniy/ visited on 30.06.2018 (in Russian).
3. I. O. Sakovich, Yu. S. Belov. "Survey on basic methods of contour analysis for highlighting the contours of moving objects" Engineering Journal: Science and Innovation, No. 12, 2014 URL: http://engjournal.ru/catalog/it/hidden/1280.html visited on 30.06.2018 (in Russian).
4. P. Viola, M. Jones "Rapid object detection using a boosted cascade of simple features" Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001.
5. ImageNet Large Scale Visual Recognition Competition, http://www.image-net.org/challenges/LSVRC/2017/ visited on 30.06.2018.
6. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel: Backpropagation Applied to Handwritten Zip Code Recognition, Neural Computation, 1(4):541-551, Winter 1989.
7. Tensor Flow: An open source machine learning framework for everyone, https://www.tensorflow.org/ visited on 30.06.2018.
8. Cireşan D.C. "Deep big multilayer perceptrons for digit recognition" – Springer Berlin Heidelberg, 2012. – c. 581-589.