Classification of breast cancer cytological specimen using convolutional neural network

Michał Żejmo¹, Marek Kowal¹, Józef Korbicz¹ and Roman Monczak²

¹University of Zielona Góra, Institute of Control and Computation Engineering, 65-516 Szafrana 2, Zielona Góra, Poland
²Department of Pathomorphology, Regional Hospital in Zielona Góra, Zyty 26, 65-046 Zielona Góra, Poland
E-mail: {M.Zejmo,M.Kowal,J.Korbicz,R.Monczak}@issi.uz.zgora.pl

Abstract. The paper presents a deep learning approach for automatic classification of breast tumors based on fine needle cytology. The main aim of the system is to distinguish benign from malignant cases based on microscopic images. Experiment was carried out on cytological samples derived from 50 patients (25 benign cases + 25 malignant cases) diagnosed in Regional Hospital in Zielona Góra. To classify microscopic images, we used convolutional neural networks (CNN) of two types: GoogLeNet and AlexNet. Due to the very large size of images of cytological specimen (on average 200000 × 100000 pixels), they were divided into smaller patches of size 256 × 256 pixels. Breast cancer classification usually is based on morphometric features of nuclei. Therefore, training and validation patches were selected using Support Vector Machine (SVM) so that suitable amount of cell material was depicted. Neural classifiers were tuned using GPU accelerated implementation of gradient descent algorithm. Training error was defined as a cross-entropy classification loss. Classification accuracy was defined as the percentage ratio of successfully classified validation patches to the total number of validation patches. The best accuracy rate of 83% was obtained by GoogLeNet model. We observed that more misclassified patches belong to malignant cases.

1. Introduction
Breast cancer is the most common cancer among women worldwide. In 2012 nearly 1.7 million new cases were diagnosed (second most common cancer overall) and 522 000 cases of death (fifth cause of death from cancer overall). This represents about 12% of all new cancer cases and 25% of all cancers in women [1]. In Poland there were 17 000 cases and 5574 resulted in death (National Cancer Registry in Poland, 2016 [2]).

Breast cancer is mostly diagnosed by so called ”triple-test” containing three medical examinations usually occurring in order: palpation, ultrasonography or mammography imaging and fine needle biopsy (FNB) [3]. Here, we concentrate on automatic analysis of results of FNB examination. In FNB method, cytological material for examination is collected directly from the tumor using specialised needle. Next, the material is fixed and stained. Finally, slide glass with cellular material is examined by the pathologist under the microscope. Unfortunately, detection of cancer cells is not an easy task. Novice pathologists must spend a lot of time observing thousands of specimens in order to gain experience and knowledge to become professional specialist. Moreover, analysis of entire slide is time-consuming process, even for experienced...
pathologists. A number of scientific groups conduct an intensive research in the area of Computer-Aided Cytology (CAC) [4, 5, 6, 7, 8, 9, 10, 11, 12]. Typically, CAC systems process the images in order to segment nuclei, then different features of nuclei are extracted and finally samples are classified [13, 14, 15]. Usually, different machine learning techniques are employed to fulfill these tasks. Unfortunately, microscopic images of tissue are challenging for automatic analysis. Cells have tendency to clump and generate nests, thus, it is hard to extract isolated nuclei. Furthermore, nuclei features used to discriminate malignant from benign cases are not standardised and are usually chosen based on subjective experiences. Therefore, most proposed approaches lack generality and satisfactory accuracy is only achieved for images acquired from specific source.

In recent years, there have been a lot of successes in complex object recognition and classification using CNN models. Their architecture is inspired biologically by the organization of human visual cortex. The ability to learn features invariant to translation, rotation and shifting is their great advantage. Recent studies in the field of digital pathology show the high efficiency of CNN in the diagnosis of breast cancer. Spanhol et al. applied AlexNet CNN model to classify breast cancer histopathological images [16]. They were learning CNN using training patches generated by different strategies. They obtained best classification accuracy 89.6% for images with 40x enlargement and patches of size 64x64 pixels selected in a random way from original image. Hatipoglu and Bilgin applied CNN to distinguish cellular and non-cellular structures in breast cancer histopathological images [17]. Neural network was used to classify small patches of size 9x9 and it reached 86.88% accuracy. In 2016, International Symposium on Biomedical Imaging (ISBI) held the grand challenge Camelyon16 to evaluate systems for the automated detection of metastatic breast cancer in histopathological images of sentinel lymph node biopsies. Wang et al. won this competition using GoogLeNet CNN model [18]. Their system reached 98.4% patch classification accuracy. Cruz-Roa et al. applied CNN to automatic detection of invasive ductal carcinoma tissue regions in histological images of breast cancer [19]. The method was evaluated over a data set consisting of 162 patients. They compared CNN with random forest classifier in terms of F-measure and balanced accuracy. CNN yielded better results (71.80%, 84.23%) than random forest (67.53%, 78.74%).

The approaches mentioned above deal with the automatic analysis of histopathological images of breast cancer. We focus on the classification of cytological images of breast cancer. Cytological biopsies are much less invasive than histological, thus they are less traumatic for patients. Unfortunately, they are substantially harder to analyse because cytological smears have more chaotic and complex structure.

Our approach is based on the classification of small patches of cytological image in order to detect malignant cells. At the beginning, pathologist has marked the presence of benign or malignant cells in some areas of virtual slides. Next, annotated fragments were divided into small overlapping patches with a size of 256 × 256 pixels. Our studies show that the size of patches have influence on classification results. Overall, our experience is that the larger size of patches, the smaller classification error. However, the size of the patches is limited by the size of RAM accessible on GPU unit. In our case the maximum size of patches was limited to 256 × 256 pixels. Patches with the larger size had to be downsized by changing resolution, what deteriorate classification results. Patches with smaller size contain less information on the distribution of nuclei in the sample and therefore results of classification was also not satisfactory.

We formed training and validation data sets in such a way that all patches coming from a particular patient can belong either to training or to validation data set. In addition, each patch has been verified using SVM classifier in terms of how much cellular material represents. If the patch did not contain cellular material above the threshold, it was not included in neither the training nor the validation collection. To implement and tune CNN models, we used Caffe framework and NVIDIA Digits software. Networks were learned using three different gradient
descent based optimisation algorithms. We explored the accuracy of GoogLeNet and AlexNet architecture for three different collections of patches. These collections of patches differ in cellular coverage degree.

The remainder of this paper is organised as follows. In Section 2, materials and methods are described. Section 3 gives the description of experiments and results. Concluding remarks are given in Section 4.

2. Material and Methods

2.1. Study data set

Cytological images, which we were using in this study came from patients of Regional Hospital in Zielona Góra, Poland. Cellular material was acquired from affected tissue using 0.5-mm-diameter needle under the control of an ultrasonograph. Next, the material was fixed with Cellfix (Shadon) fixative spray and dyed with hematoxylin and eosin (h+e). The time between preparation of smears and their preservation in fixative spray never exceeded three seconds.

In this way, 50 specimens were prepared for 50 patients with breast cancer. Half of the cases are malignant, half are benign. All of the cancers were histologically confirmed, and all of the patients who carried benign disease were either biopsied or followed for a year. Digitization of glass slides was done using Olympus VS120 Virtual Microscopy System. The system consists of a 2/3” CCD camera and 40x lens with giving together 0.172 µm resolution. The average size of slides from the system is approximately 200000 × 100000 pixels. All the scans were prepared using Extended Focal Imaging (EFI) system. EFI allows for extended focal depths impossible to obtain using optics alone. Scans with EFI are performed by scanning a preparation several times with the different positions of focus plane. Final frame is formed by selecting only sharp focus regions from each frame. Finally, scanned slides are saved on a hard disk using Virtual Slide Image (VSI) format, dedicated to store very large images. Pathologists can browse virtual slides on a computer screen using special software. Pathologist from Hospital in Zielona Góra identified regions of malignant and benign changes in order to prepare images for training CNN. For each virtual slide 11 regions of interests (ROI) representing malignant or benign cells were selected. Next, each ROI was converted to 8 bit/channel RGB TIFF image of size 1583 × 828 pixels. Compression to TIFF was done by lossless LZW algorithm. Fig. 1 is showing sample results of selection procedure. Example of ROI selected by pathologist from virtual slide is shown in Fig. 2. On the whole, pathologist has selected 550 ROIs (275 malignant + 275 benign).

2.2. Training and validation data sets

The collection of 50 virtual slides was randomly split into training and validation set. Training set consist of 30 virtual slides (15 benign and 15 malignant) and validation set consist of 20 virtual slides (10 benign and 10 malignant). Therefore, we obtained 330 ROI images for training and 220 ROIs for validation. Next, we generated for every ROI a collection of overlapping patches of size 256 × 256 pixels. Sliding window was moving horizontally from left to the right in order to gather a single patch at every 32 pixels. Procedure was repeated for subsequent lines with the vertical step equal to 32 pixels. From each image 697 patches were extracted. From a diagnostic point of view only the patches which illustrate cellular material are useful for training and validation process. That’s why we decided to remove patches that has a low cellular coverage ratio. Such coefficient was determined for every patch using SVM classifier. It was responsible for labelling pixels as cell (nuclei and cytoplasm) or other (background or blood) based on their RGB colour. To train SVM, several manually labelled images were used. Finally, we created three collections of patches with cellular coverage ratio equal or greater than 50%, 75% and 90% respectively. The sizes of training and validation collections for different cellular coverage ratios are presented in Table 1. Figure 3 shows sample patches with cellular coverage ratio equal or greater than 90%.
2.3. Convolutional Neural Networks
Nowadays, CNN become very popular technique of machine learning, which is often used to image recognition [20]. Idea of CNN was inspired biologically by processing the image in the visual cortex of animals. One of the most famous example of CNN application was LeNet network that was successfully applied for handwritten digits recognition [21]. CNN is usually built of many convolutional layers connected to each other by pooling layers. At the top of this structure, reside fully connected layers similar to feed-forward layers from multilayer perceptron (MLP). Figure 4 shows the sample structure of CNN model.
**Table 1.** Collections of patches

| Cellular coverage | Training set [amount] | Validation set [amount] |
|-------------------|-----------------------|------------------------|
|                   | Benign | Malignant | Benign | Malignant |
| 50%               | 62229  | 52225     | 29822  | 28981     |
| 75%               | 47053  | 23974     | 17654  | 14204     |
| 90%               | 34189  | 10107     | 10214  | 6644      |

**Figure 3.** Examples of patches used for training and validation of CNN (left: benign cells and right: malignant cells).

**Convolutional layer** is composed of learnable filters and is a core part of CNN. It is responsible for processing input image in order to extract features describing input image. Parameters of filters are weights of the network tuned during the training. With such approach, network is able to learn features from examples during training procedure. Kernels of filters are spatially small and rectangular, but extends through the full depth of the input image due to operation of convolution. For example, filter in a first layer of a CNN might have a kernel of size 5x5x3 (5 pixels width and height, and depth 3 because the RGB colour channels). Each kernel is sliding across the width and height of the input and compute dot products between the kernel and the fragment of the image cutted by the window of the same size as kernel and centred at the actual position of the kernel. CNN is usually composed of many convolutional layers. Shallow convolutional layers extract low level features and deeper layers are able to extract high level features.

**Pooling layer** is used to reduce progressively the spatial size of the input and to reduce the amount of tunable parameters. In this layer small rectangular patches of feature maps from convolution layer are downsampled to single value. Downsampling is usually done by max or averaging operator. The pooling layer provides a form of translation invariance. Pooling layers are inserted periodically between the convolutional layers.

**Fully-connected layer** is at the top of CNN. The aim of this layer is capture complex
relationship between high-level features. This layer mixes output of pooling layer to feature vector. Thus, fully connected layers are one-dimensional, so there can be no convolutional or pooling layer after.

In addition CNN can have ReLU (Rectified Linear Units) layer and LRN (Local Response Normalisation) layer. ReLU layer consist of neurons with the activation function \( f(x) = \max(0, x) \). Such neurons increase the nonlinear properties of the decision function without affecting the receptive fields of the convolution layer. The LRN layer performs a kind of local normalisation over defined input area. Each input value is divided by \( (1 + \alpha n \sum_i x_i^2)^\beta \), where \( n \) is the size of each local region, \( \alpha \) and \( \beta \) are arbitrarily chosen parameters and the sum is taken over the region centred at the point, which is normalised.

**Figure 4.** Architecture of sample CNN model.

A lot of different CNN architectures have been already tested and described in the scientific literature [22, 23, 24, 25, 26, 27, 20, 29]. They vary in the layer configuration and depth of structure. In our studies we used two well known architectures: AlexNet [20] and GoogLeNet [29]. These structures have proven their usefulness in many complex image recognition tasks.

AlexNet belongs to the class of deep structures. It is made up of 5 convolutional layers with ReLU activation functions, 3 max-pooling layers and 3 fully connected layers with ReLU neurons. Moreover, in the first two convolutional layers, outputs are normalized using LRN technique. In fully connected part of the network a dropout technique was used to reduce the validation error [30]. This technique assumes that each hidden neuron can set its output value to 0 for one epoch with probability 0.5. Therefore, the neurons which are dropped out do not contribute to the output and are not tuned during backpropagation step. In this way, different structures of the network are sampled during training procedure. At validation step, all the neurons participate in the output, however each individual output of the neuron is multiplied by 0.5.

GoogLeNet belongs to the class of deep CNN architecture which originally consists of 22 layers [29]. To make such architecture computationally efficient, its founders organized the structure of the network in the form of Inception modules [29]. The main concept of the Inception structure is to consider how an optimal local sparse structure of CNN can be approximated by easily available dense components. Generally GoogLeNet is a network consisting of Inception modules stacked upon each other, with occasional max-pooling layers. All layers use ReLU activation function.

2.4. GPU computation

The process of learning CNN is extremely computationally expensive, due to huge amount of calculations on arrays of data and weights. Standard CPU is optimised for single-threaded performance, therefore computing efficiency of CPUs is completely insufficient for training CNN with complex structure and huge amount of training data. Best solution to meet this problem
is to use GPU. To train and simulate our CNN models we used deep learning framework Caffe\(^1\). For data, results and models management we decided to use NVIDIA DIGITS\(^2\) system. Mentioned software use CUDA\(^3\) (Compute Unified Device Architecture) and CuDNN (CUDA Deep Neural Network) library for massively parallel computations to shorten drastically the time needed to learn our CNN models. All experiments were done using GeForce GTX TITAN X with 12GB of RAM.

3. Results
Two CNN architectures GoogLeNet and AlexNet were compared in terms of classification accuracy of breast cancer cytological images. Each architecture was trained several times for different training data and different training algorithm. Three different training collections of images were prepared. Subsequent training collections contain images with cellular coverage ratio equal 50%, 75% and 90% respectively. To train CNN, we have used 3 different gradient descent algorithms: stochastic gradient descent (SGD), adaptive gradient (ADA) and Nesterov’s accelerated gradient (NAG). Both architectures were learned and validated with each data set and each training algorithm. In total, each CNN model was trained 9 times. Each training took 10 epochs. Due to GPU memory limitations, we used mini-batch version of backpropagation algorithm. Thus, weights of the network were updated after the presentation of each mini-batch of training data. For AlexNet architecture, we choose mini-batch of size 450 samples and for GoogLeNet mini-batch of size 100 samples. After each training epoch, network was validated to control the overfitting problem. Learning rate was initialized with a value \(lr = 0.01\), and was dropping by a constant factor 10 when objective function begins to reach an evident plateau. Objective function is defined as a cross-entropy classification loss:

\[
L = -\frac{1}{N} \sum_{i=1}^{N} \log \left( \frac{e^{x_{i,l}}}{\sum_{j=1}^{K} e^{x_{i,j}}} \right)
\]

where \(N\) is the size of the mini-batch, \(K\) is the number of network outputs, \(x_{i,j}\) is the \(j\)-th output of the network for \(i\)-th sample and \(l\) is the label of the correct ground truth class for \(i\)-th sample. Classification accuracy for validation collection was measured as the percentage ratio of successfully classified patches to the total number of patches.

Table 2 shows summarised results of conducted research. For collection of patches with 50% cellular coverage ratio, both CNN architectures gave similar accuracy close to 80%, regardless of the training method. For 75% and 90% cellular coverage ratios GoogLeNet achieve better results. The worst classification accuracy was achieved using ADA training algorithm. Best accuracy 83% was obtained by GoogLeNet model for patch collection with 90% cellular coverage ratio using SGD training method. Figure 5 shows the process of training of the best GoogLeNet model and best AlexNet model. From these figures, we can observe that learning process converged to a local minimum. However, we can also observe that relatively quickly appears the effect of overfitting. This problem can be caused by low number of training samples. We plan to extend training dataset in future work.

Table 3 presents detailed classification results for the best GoogLeNet model, which were obtained for validation data set. For each case, the percentage of properly classified patches was determined. Most of the cases were correctly diagnosed. But, for benign cases, patient no. 10 was apparently classified incorrectly as malignant. For malignant cases, 2 patients no. 19 and 20 were incorrectly diagnosed as benign. Diagnosis for the patient no. 19 is uncertain, on the border between benign and malignant class. Incorrectly classified patches originating

---

\(^1\) http://caffe.berkeleyvision.org/
\(^2\) https://developer.nvidia.com/digits
\(^3\) http://www.nvidia.com/CUDA
Table 2. Patch classification accuracy for different CNN models, trained with different learning algorithms and using different training collections

| Minimum cellular coverage ratio | 50% | 75% | 90% |
|---------------------------------|-----|-----|-----|
| CNN model | Grad. | Acc. | CNN model | Grad. | Acc. | CNN model | Grad. | Acc. |
| AlexNet | SGD | 80% | AlexNet | SGD | 74% | AlexNet | SGD | 80% |
| | ADA | 78% | | ADA | 55% | | ADA | 61% |
| | NAG | 77% | | NAG | 74% | | NAG | 74% |
| GoogLeNet | SGD | 80% | GoogLeNet | SGD | 82% | GoogLeNet | SGD | 83% |
| | ADA | 80% | | ADA | 75% | | ADA | 73% |
| | NAG | 76% | | NAG | 82% | | NAG | 79% |

Table 3. Detailed classification results for the best GoogLeNet model

| Ground truth diagnosis | Patient no. | Total number of patches | Percentage of well classified patches |
|------------------------|-------------|-------------------------|--------------------------------------|
| Benign                 | 1           | 1675                    | 100%                                 |
|                        | 2           | 1180                    | 100%                                 |
|                        | 3           | 232                     | 100%                                 |
|                        | 4           | 142                     | 100%                                 |
|                        | 5           | 731                     | 99.2%                                |
|                        | 6           | 336                     | 91.7%                                |
|                        | 7           | 2467                    | 88.3%                                |
|                        | 8           | 449                     | 82.4%                                |
|                        | 9           | 2236                    | 61.5%                                |
|                        | 10          | 766                     | 0%                                   |
| Malignant              | 11          | 152                     | 100%                                 |
|                        | 12          | 1453                    | 94.8%                                |
|                        | 13          | 2557                    | 91.5%                                |
|                        | 14          | 88                      | 90.9%                                |
|                        | 15          | 543                     | 74.4%                                |
|                        | 16          | 980                     | 69.9%                                |
|                        | 17          | 31                      | 67.7%                                |
|                        | 18          | 256                     | 66.8%                                |
|                        | 19          | 474                     | 49.2%                                |
|                        | 20          | 110                     | 37.3%                                |

from wrongly classified patients were reviewed by the pathologist. He classified most of them as "difficult cases". It was noted that some misclassified malignant patches are difficult to diagnose due to the lack of the information originating from the rest of the specimen. Sometimes the context in which patch is located is just as important as the information included in the patch. Finally, we must also take into account the fact that even the pathologists do not always agree in their diagnoses.

4. Conclusions
Content of cytological images is highly complex and stochastic, so their analysis is difficult in an automated way. To tackle this issue, we applied a deep learning approach to classify cytological images coming from fine needle biopsy of breast cancer. Proposed approach is based...
Figure 5. Top figure shows learning process of GoogLeNet CNN and bottom shows learning process of AlexNet CNN. Line labeled as 1 denote classification accuracy for validation data, lines with labels 2 and 3 denote loss function for validation and training data, respectively.

on processing small patches selected from cytological images. Patches are processed by CNN. It gives as his output information whether the tumor is benign or malignant. The best patch classification accuracy reached 83%. Such result was obtained by GoogLeNet architecture. Other authors have already reported better results for breast cancer classification using CNN but their studies are related to histopathological images. We treat obtained results as starting point to further research aimed at improving the classification accuracy.

Analysis of the results indicates that CNN does less well with malignant cases than with
benign cases. We explain this fact by far fewer number of malignant patches than benign patches in the training set. In addition, the pathologist drew our attention to the fact that some malignant patches are very similar to benign and to deal with this problem we should use larger patches. Further investigations will be conducted in two directions: firstly, to prepare richer collection of training images and secondly to develop the structure of CNN which will be able to combine local and global information from the cytological image.

Acknowledgments
The research was supported by National Science Centre, Poland (2015/17/B/ST7/03704).

References
[1] Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin D M, Forman D and Bray F 2013 Globocan 2012 v1.0, cancer incidence and mortality worldwide: Iarc cancerbase no. 11. accessed on: 13/09/2016 URL [2] Wojciechowska U and Didkowska J 2013 Accessed on: 13/09/2016 URL [3] Underwood J 1987 Introduction to Biopsy Interpretation and Surgical Pathology (London: Springer-Verlag) ISBN 9780387174952
[4] Jeleń L, Fevens T and Krzyżak A 2010 Int. J. Appl. Math and Comp. Sci. 18 75–83
[5] Filipczuk P, Kowal M and Obschowicz A 2013 Proc. 35th Int. Conf. IEEE Engineering in Medicine and Biology Society, EMBC 2013 (Osaka, Japan)
[6] Kowal M, Filipczuk P, Obuchowicz A, Korbiez J and Monczak R 2013 Computers in Biology and Medicine 43 1563–1572
[7] George Y, Elbagoury B, Zayed H and Roushdy M 2012 Int. J. Computer Science Issues 9 247–256
[8] Malek J, Sebri A, Mabrouk S, Torki K and Tourni R 2009 J. Signal Processing Systems 55 49–66
[9] Mu T and Nandi A 2014 Proc. INTERACT 2014. The 39th IEEE International Conference on Advanced Information Networking and Applications, AINA 2014, Proceedings, 2014 pp 686–690
[10] Niwas I, Palanisamy P and Sujathan K 2010 Proc. IEEE Symp. Industrial Electronics & Applications, ISIEA 2010 p 686–690
[11] Plissiti M and Nikolou C 2012 IEEE Trans. Image Processing 21 4568–4580
[12] Soliman O and Hassanien A 2012 Rough Sets and Current Trends in Computing (Lecture Notes in Computer Science) vol 7413 ed Yao J, Yang Y, Slowiński R, Greco S, Li H, Mitra S and Polkowski L (Springer, Berlin Heidelberg) pp 106–115
[13] Buyssens P, Elmoataz A and Lézoray O 2013 Computer Vision – ACCV 2012: 11th Asian Conference on Computer Vision, Daejeon, Korea, November 5–9, 2012, Revised Selected Papers, Part II ed Lee K M, Matsusita Y, Rehg J M and Hu Z (Berlin, Heidelberg: Springer Berlin Heidelberg) pp 342–352 ISBN 978-3-642-37444-9 URL [14] Kowal M and Filipczuk P 2014 Int. J. Appl. Math and Comp. Sci. 24 19–31
[15] Mazurek P and Oszustowska-Mazurek D 2014 Int. J. Appl. Math and Comp. Sci. 24 49–63
[16] Spanhol F A, Oliveira L, Petitjean C and Heutte L 2016 Proc. Int. Joint Conf. on Neural Networks (IJCNN 2016) (Vancouver, Canada) pp 123–128
[17] Hatipoglu N and Bilgin G 2014 Proc. 4th Int. Conf. on Image Processing Theory, Tools and Applications (IPTA) (Paris, France) pp 1–6
[18] Wang D, Khosa A, Gargerya R, Irshad H and Beck A H 2016 CoRR abs/1606.05718 [19] Cruz-Roa A, Basavanahally A, Gonzalez F, Gilmore H, Feldman M, Ganesan S, Shih N, Tomaszewski J and Madabhushi A 2014 Proc. SPIE Medical Imaging vol 9034 (San Diego, USA) pp 904103–904103–15
[20] Krizhevsky A, Sutskever I and Hinton G E 2012 Proc. 26th Conf. on Neural Information Processing Systems (NIPS) (Lake Tahoe, USA) pp 1097–1105
[21] LeCun Y, Bottou L, Bengio Y and Haffner P 1998 Proceedings of the IEEE 86 2278–2324
[22] LeCun Y, Huang F J and Bottou L Proc. IEEE Conf. Computer Vision and Pattern Recognition, CVPR 2004 vol 2 (IEEE) pp II–97
[23] Jarrett K, Kavukcuoglu K, Lecun Y et al. 2009 Proc. 12th IEEE Int. Conf. on Computer Vision, ICCV (Kyoto, Japan) pp 2146–2153
[24] Krizhevsky A and Hinton G 2010 Convolutional deep belief networks on cifar-10 URL [25] Lee H, Grosse R, Ranganath R and Ng A Y 2009 Proc. 26th Int. Conf Machine Learning (Montreal, Canada) pp 609–616
[26] LeCun Y, Boser B, Denker J S, Henderson D, Howard R E, Hubbard W and Jackel L 1990 ed Touretzky
D S (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.) chap Handwritten Digit Recognition with a Back-propagation Network, pp 396–404 ISBN 1-55860-100-7

[27] Pinto N, Doukhan D, DiCarlo J and Cox D 2009 PLoS Comput Biol 5 e1000579

[28] Turaga S C, Murray J F, Jain V, Roth F, Briggman M H K, Denk W and Seung H S 2010 Neural computation 22 511–538

[29] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A 2015 Proc. IEEE Conf. Computer Vision and Pattern Recognition, CVPR (Boston, USA) pp 1–9

[30] Srivastava N, Hinton G E, Krizhevsky A, Sutskever I and Salakhutdinov R 2015 J. Machine Learning Research 15 1929–1958

[31] Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S and Darrell T 2014 Proc. 22nd ACM Int. Conf. on Multimedia, MM '14 (Orlando, USA) pp 675–678 ISBN 978-1-4503-3063-3