Application of the transfer learning to the medical images texture classification task

M Privalov[1,*], *M* Stupina[1]

[1]Don State Technical University, Rostov-on-Don, Russia

Abstract. This study is conducted to determine effectiveness and perspectives of application of the transfer learning approach to the medical images classification task. There are a lot of medical studies that involve image acquisition, such as X-Ray radiography, ultrasonic scanning, computer tomography (CT), magnetic resonance imaging (MRI) etc. Besides those medical procedures there are different operations that use medical images processing including but not limited to digital radiograph reconstruction (DRR), radiotherapy planning, brachy therapy planning. All those tasks could be effectively performed with help of software capable to perform segmentation, classification and object recognition. Those capabilities are naturally depend on neural classifiers. Presented work investigates different approaches to solving image classification task with neural networks, specifically, using pre-processing for feature extraction and end-to-end application of convolutional neural networks (CNN). Due to requirement of significantly big datasets and large computing power CNNs sometimes may appear difficult to train, so our results pay attention to application of transfer learning technique that can potentially relax requirements to classifier training. The conclusions of this study state that transfer learning can be effectively used for classification tasks, especially texture classification.

1 Introduction

At the present time we can see sweeping development of the computer technologies and methods of the artificial intelligence which enter into all realms of the human activity. One of the most popular directions is application of the digital image processing and recognition methods that are being used in different systems such as medical diagnostics systems, video surveillance, security and others. Those methods are usually based on image features calculation which could be a color, shape or texture followed by decision making based on those feature values. Texture is an important image feature that plays key role in solving variety of tasks in medical image processing, flaw detection, object recognition, motion tracking [1].

For example, in medical diagnostics there is a number of tasks related to detection of diffuse lesions of organs where shape of the object has secondary priority or cannot be precisely detected. For example, when searching tumors on computer tomographic (CT) images there could be difficult to detect contours just using brightness features because

* Corresponding author: maxim.privalov@gmail.com

healthy tissue and tissue that is the result of malignant proliferation are very similar. In analysis of different structures texture may point to object state and on machining quality analysis made for processed surface texture may help to understand if the quality level is in conformance to standards. Thus, changes in texture or its non-uniformness can point to the flaw existence. According to that, texture processing and classification is actual task.

2 Modern state of the texture classification

Texture analysis tasks are including texture segmentation, classification and synthesis. But in medical diagnostics and flaw detection the most used task is texture classification. First attempts of the texture description were made in 1962 and after that texture classification methods were actively developed, especially since 1980s of the last century. At that time popularity of the texture classification methods that were based on features extraction using different approaches was highly growing. In [2] we can see classification of those approaches which have become classic but when we attempt to classify those methods using higher level of representation we can define two big methods groups according to [3]:

1. Methods that are based on filtering.
2. Methods that are based on statistical modelling.

We can attribute to the first group those texture classification methods that based on Laws features [3] that imply application of the filters bank to the image fragment inside some aperture. Further this group of methods was widened by Gabor filters, wavelet analysis methods based on Daubechies and Mallat wavelets, wavelet pyramids.

Second group consists of methods based on statistical analysis of brightness distribution, Markov random field models and fractal image features.

Despite big number of features that could be extracted from the image that can help increasing classification accuracy it is often necessary to solve several problems performing texture analysis:

– providing invariance to the overall brightness level, point of view, image view angle and its scale;
– problem-oriented tuning of the classifier.

Invariance is usually achieved by applying additional transforms. For example, to provide invariance to the texture rotation angles it is often good to use the log-polar transform [4] which in combination with wavelet transform could give quite stable features for classification.

But the second problem related to the application of statistical and filtering based methods has more difficult solution. If we look onto the gray level co-occurence matrix features proposed by Haralick [5] we can see that approach suggests texture description by statistical features calculated from the matrices built using four displacement vectors with angles 0, 45, 90 and 135 degrees. Co-occurrence matrix contains in its cells frequencies of the corresponding grey levels that co-occur divided by the vector having specific norm and angle. Those matrices are further used as source data for GLDS features calculation (fig. 1):

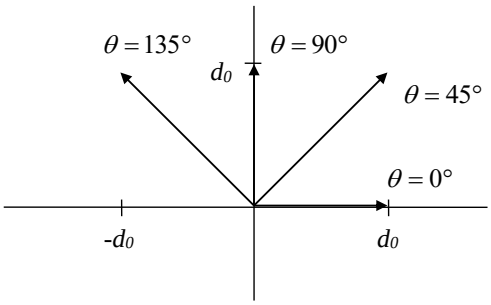


Fig. 1. Displacement vectors for GLDS features calculation.

As we can see from the figure there is a features parameter d_0 and its selection is conditioned by texture features. Coarse grained textures lead to larger d_0 size selection while fine grained ones lead to smaller d_0 size (in other words, size d_0 of the displacement vector depends on the texture base element size, so called "texton" [6]). In other words we have an additional hyperparameter and necessity to tune it to the different tasks. Moreover, when solving multiclass texture classification the dependency on hyperparameter d_0 may worsen classification error for cases when it is necessary to process textures having large variety of the base element size that can lead to the big differences of the texture base elements from the displacement vector sizes. However there are modern and promising methods in the area of machine learning that can help with solving mentioned problems.

Modern approaches that could automate model training and thus reduce the number of hyperparameters are Bag of Words approach and deep convolutional neural networks [3].

Bag of Words approach is based on the similar method applied to the natural language processing tasks and consists of following steps:

1. Selection of N regions of interest on the fixed grid.
2. Calculation of the local region representation as feature set x_i using size D for different angles, scales and intensities.
3. Generation of code words that describe texture using any clustering method, for example, k-means.
4. Transformation of the texture descriptors into encoded embeddings.
5. Aggregation of the features using union operation with averaging, or maximum calculation, or building the spatial pyramid.
6. Texture classification using support vector machine, random forest or feedforward neural network.

This approach is better adapted to the recognition task of different textures but still strongly dependent on quality features set that should be calculated on the step 2.

The most perspective approach is one using deep neural networks, especially convolutional neural networks (CNN). Deep neural network is different from the shallow neural network by the number of layer that in common is greater than 2. Because the every layer implements the set of classifiers that distinguish between textures using some function describing decision boundary, such classifier superposition can implement very complex discrimination surface. Convolutional neural networks are even more effective in such task because of the principle of their operation.

Convolutional layer of the neural network performs source data convolution with the filters set having aperture size f and number of components corresponding to the number of components in the input data. The most popular variant is 2D discrete convolution (fig. 2), but for the dimensionality changes there is also 1x1 convolution layer while for the

volumetric data we may use 3D convolutions or convolutions with even bigger dimensionality.

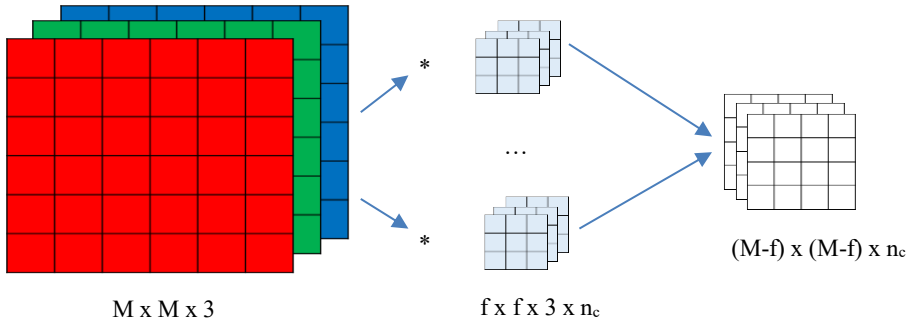


Fig. 2. An example of colour input image processing using convolutional layer.

Besides convolutional layers there are pooling layer and dropout layer that are used for result data dimensionality reduction and overfitting elimination. Pooling layers can combine input data replacing group of their values by their maximum value, mean value or more complex combination. Dropout layers are randomly excluding weights blocking transfer of previous layer values to the next layer inputs. Those layers are usually combined with convolutional layers so different authors may use term 'convolutional layer' for the union of the filter layer and combining layer. An example of the simple convolutional network used for handwritten digits recognition and called LeNet-5 [7] is shown on the fig. 3. The network consists of two convolutional layers with the filter size f=5 and stride s=1 followed by combining layer with the filter size f=2 and stride s=2. Then processing is performed in the two dense layer with 120 and 84 neurons that have ReLU activation function followed by the softmax layer with 10 outputs corresponding to 10 digits.

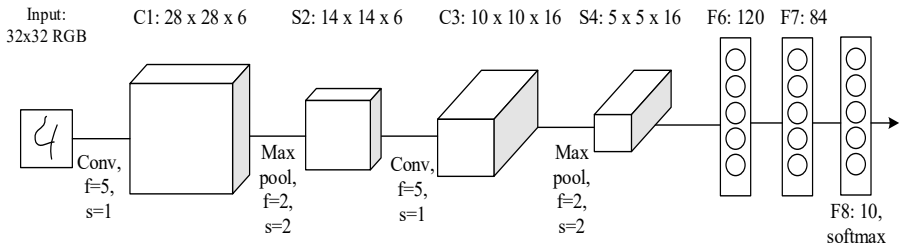


Fig. 3. Convolutional network LeNet-5 for handwritten digits classification.

Specific features of those networks are following: the number of filters per layer is usually growing in direction from input to top of the network, on contrary filter size is reducing, the network in most cases is finalized by several dense layers on top of it. Final layer type depends on the classification task: for the binary classification it is usually the sigmoid and for the multiclass classification it is softmax function.

3 Texture classification using deep neural networks

The main idea laying beneath application of CNNs to the texture classification is that filter coefficients used in convolution and defining which image features will be extracted by the network are parameters that will be searched during neural network training by backpropagation algorithm. Thus, CNN application may reduce number of the

hyperparameters and in such way significantly reduce the amount of work required to build the textures classifier for some specific task from the problem domain. Using texture classification task specifics it is possible to also apply data preprocessing with purpose of texture features extraction that are difficult to get from the image using convolution only.

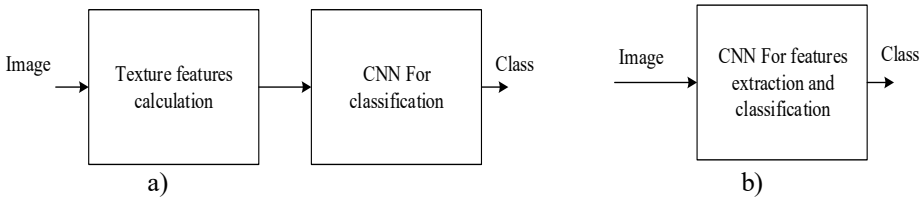
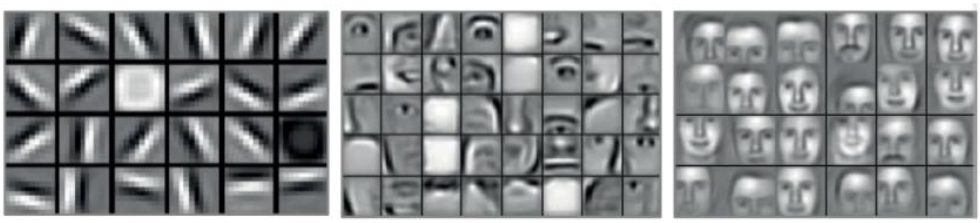


Fig. 4. Texture classification using convolutional neural network: a) directly; b) with pre-processing.

Depending on the task that is being solved training data availability may differ. In some cases it is easy to gather data in large amounts but in others this is not true and it is necessary to spend significant efforts to build the dataset, and when the number of images is still low it also requires to use data augmentation. Also quite often case is having the training data with skewed classes: for example, when solving the task of medical diagnostics it is not easy to gather representative data set of norm because in the medical establishments that are not equipped with automated packaging and archiving computer system long term storage is usually used only for pathological studies. In those cases transfer learning may help. The main idea is that large convolutional neural networks that were trained on the big image sets with purpose of solving some competition task such as ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [8] are able to detect quite big amount of features including texture. During forward propagation low-level and mid-level features such as intensity changes present in different directions and for different scales, lines, object corners are detected by earlier or middle layers. High-level features such as macro parts of different objects are detected in later layers (fig. 5).



a) low-level features b) mid-level features c) high-level features

Fig. 5. Features with different level of detail that are detected by convolutional neural network trained in face recognition.

Idea of the transfer learning lays in attempt to use features already extracted by the pre-trained network and apply them to solution of the other classification task. Implementation of this approach suggests that output layer is removed from source network and replaced by the new classifier that will process detected features in terms of the current task. Sometimes it is possible to remove not only output layer, but several top layers of the pre-trained neural network. When training modified neural network the weights of the part that isn't removed are usually blocked, in other words training is applied only to the new top layers. In case when the classification accuracy is not enough it is possible allow training of all network layers and thus not only prepare new classifier but tune lower layers to the task (fig. 6). In some cases structure of the neural network may be left unchanged but the last layers will be trained using the new training dataset.

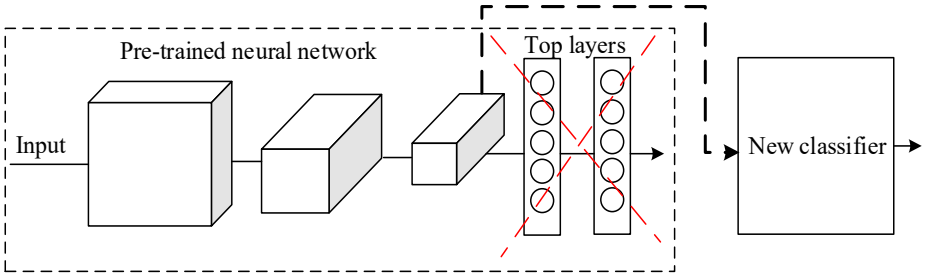


Fig. 6. Texture classification using pre-trained convolutional neural network and transfer learning.

Candidates that are good in solving task using this method and well positioned in IILSVRC: AlexNet [9], ResNet [10], GoogleNet [11] and similar. As final classifier it is possible to use support vector machine method (SVM), small feed-forward neural network with the dense layer, decision tree combined with principal component analysis (PCA) and others. For comparing texture classification results we used approach described in [2] that suggests Mallat wavelets decomposition [12] feature with shallow Elman neural network. With purpose to extract image texture features we used 9 features calculated over 2 levels of wavelet decomposition including data from the low frequency domain along with high frequency features.

Another idea proposed in current article is to use wavelet transform result space as the source input for CNN. We make those two levels of the transform and combine them in image like shown on fig. 7.

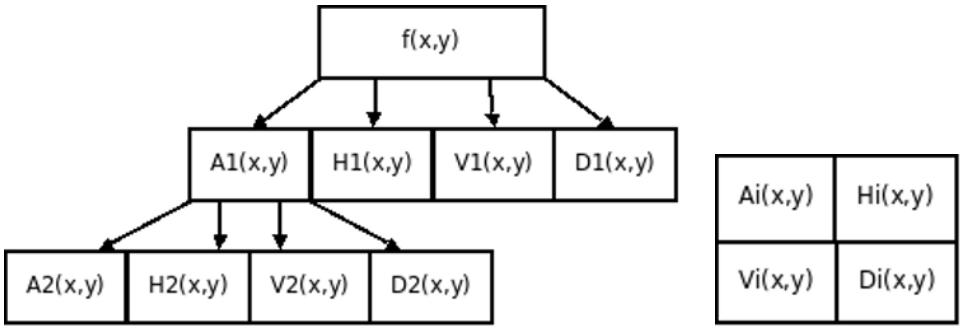


Fig. 7. Two levels of the wavelet transform (a) and combining of them into 2D image (b).

Here A_i – low frequency domain from the wavelet decomposition, H_i, V_i, D_i are correspondingly high frequency domains. Second level of the transform is combined into 2D image in the same way replacing A_i part.

Accuracy of those approaches was compared to the results achieved with deep CNNs: AlexNet, VGGVD (variant of the VGG-16 neural network) [13] and method based on specialized texture features calculated using Gabor wavelets. In experiments output layer of the AlexNet and VGGVD was replaced by the SVM classifier.

In order to allow adequate comparison of the results achieved when using pre-trained and specialized neural networks we use Brodatz textures album as the benchmark, which has detailed description in [3] and many other papers. Classification accuracy comparison for the mentioned approaches is listed in table 1.

Table 1. Texture classification results comparison.

Method	Way of texture representation	Classifier	Accuracy, %
Mallat wavelets and Elman neural network	9 features from 2 levels of the wavelet decomposition	Elman neural network	90.2
Mallat wavelts and VGGVD	2D image made from 2 levels of the wavelet decomposition	Feed-forward dense layer on top	96.3
LBP [14]	LBP, Bag of Words method	Feed-forward neural network	90.7
ScatNet	Gabor wavelets	PCA and decision tree	84.5
AlexNet	Using of features extracted by the pre-trained neural networks	Support vector machine	98.2
VGGVD			98.7

Using proposed approach performed training of the neural network in classification of kidney tumor tissues on CT images. For that purpose were gathered 3 series of CT images containing confirmed healthy tissues and malignant. Series sizes were 131 and 42 fragments with malignant textures from different apparatus and 132 fragments for healthy kidney tissues. To increase the number of samples data augmentation was also used. Visually we can see that tumor was detected and application of the transfer learning allowed to decrease errors out of the organ position.

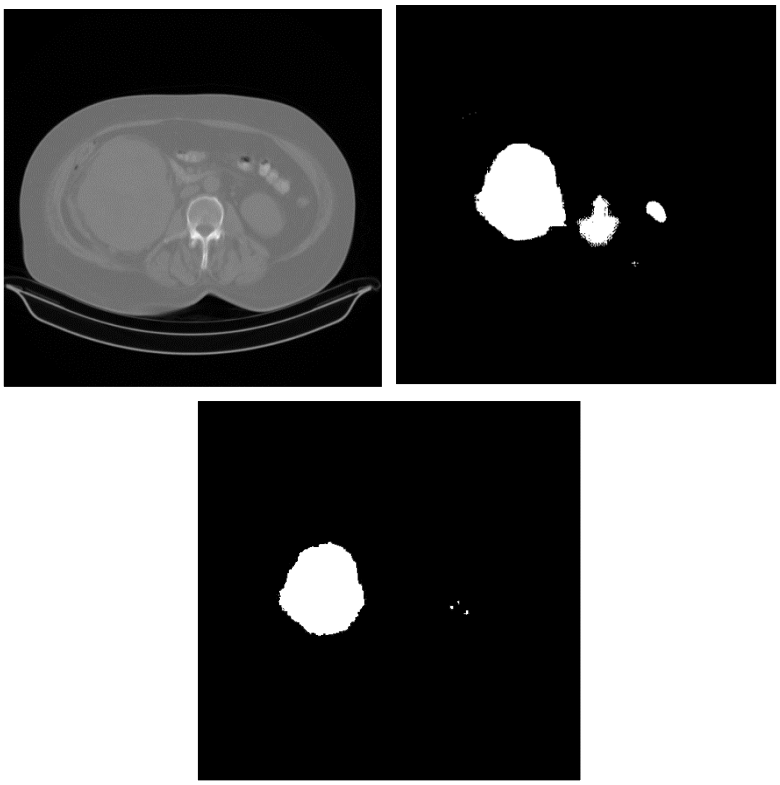


Fig. 8. Bottom view of the CT study slice of the patient having right kidney tumor (left), binarization result of the wavelets plus Elman network (middle) and wavelets plus CNN (right).

4 Result analysis and conclusions

As we can see from the results achieved by the authors [3], usage of the pre-trained deep neural networks, such as AlexNet, VGGVD, has shown classification accuracy more than 98%. This can be explained by very complex models. For example, AlexNet has 8 and VGGVD has 16 layers having weights and they were trained on very large datasets containing from 15 to 50 million of images representing 1000 classes. Usage of such neural networks, especially in conditions of lack of the computing power and not very large training datasets from the problem domain, could be very promising. But it is also clearly visible from the experiment with using wavelet transform as the input image that combination of the CNNs and specialized texture description algorithms can produce good results that are competitive and thus will form the direction of further research.

References

1. Fekri-Ershad Shervan A 2019 *New Benchmark Dataset for Texture Image Analysis and Surface Defect Detection* 1-7 doi: 10.13140/RG.2.2.33612.46722.
2. Skobtsov Y A, Privalov M V, Kudryashov A G 2010 *Kherson National Technical University* (Kherson: KNTU) **2(38)** 103-109
3. Li Liu et al 2019 *International Journal of Computer Vision Springer* **127** 74-109
4. C Pun, M Lee 2003 *IEEE trans PAMI (Pattern Analysis and Machine Intelligence)* **25(5)** 590-602
5. Privalov M V 2006 *Kherson National Technical University* (Kherson: KNTU) **1(24)** 316-321
6. Julesz B, Bergen J 1983 *The Bell System Technical Journal* **62(6)** 1619–1645
7. Guangfen Wei, Gang Li, Jie Zhao, Aixiang He 2019 *Sensors* **19(1)**
8. Olga Russakovsky, Jia Deng, et al 2015 *International Journal of Computer Vision* **115(3)** 211-252
9. Krizhevsky A, Sutskever I, Hinton G 2012 *NIPS* 1097–1105
10. He K, Zhang X, Ren S, Sun J 2016 *CVPR* 770–778
11. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A 2015 *CVPR* 1–9
12. Mallat S, Wavelet A 2009 *Tour of Signal Processing The Sparse way* (Elsiver) p 805
13. Simonyan K, Zisserman A 2015 *Very deep convolutional networks for large-scale image recognition, International conference on learning representation*
14. Pietikäinen M, Ojala T, Xu Z 2000 *Pattern Recognition* **33(1)** 43–52