Open
Full Length Article

ORIGINAL ARTICLE

Image Classification using AlexNet with SVM Classifier and Transfer Learning

Fareed Ahmed Jokhio[1], Atiya Jokhio[2]

[1,2] Department of Computer Systems Engineering, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah

ABSTRACT

Since the last decade, Deep Learning has paved the direction for demanding and wide-ranging applications in almost everyday life and attained quite good performance on a diversity of complex problems such as face recognition, motion detection, health informatics and many more. This paper discusses the performance of state-of-the-art pre-trained Convolutional Neural Networks (CNNs) model. We have analyzed the accuracy and training time of AlexNet (already trained on millions of images) while using it with Support Vector Machines (SVM) classifier and under Transfer Learning (TL). Since training CNN from scratch is time-consuming in case of real problems and is compute-intensive so the image features extraction using already trained AlexNet network reduces effort and computation resources. The learned features are used to train a supervised SVM classifier and transfer learning (fine-tuning) of layers is achieved by retraining in the last few layers of pretrained network for image classification. The classification accuracy while Transfer Learning is obtained greater and more stable than SVM classifier.

Keywords: Deep Learning, Convolutional Neural Networks (CNNs), Image Classification, Pretrained Networks, Transfer Learning

Author`s Contribution	Address of Correspondence	Article info.
[1,2,4] *Manuscript writing, Data Collection Data analysis, interpretation, Conception, synthesis, planning of research, and discussion*	*Fareed Ahmed Jokhio* *fajokhio@quest.edu.pk*	*Received: Jan 03, 2019* *Accepted: June 27, 2019* *Published: June 30, 2019*

Cite this article: *Jokhio FA, Jokhio A. Image Classification Using AlexNet with SVM Classifier and Transfer Learning. J. Inf. commun. technol. robot. appl.2019; 10(1):44-51.*

Funding Source: *Nil*
Conflict of Interest: *Nil*

INTRODUCTION

Image Classification plays a vital role in computer vision and is used to narrow down the gap between computer vision and human vision. Several traditional classification techniques have been proposed since the last few years for image classification [1, 2]. Furthermore, the training and testing phase precedes different steps like Image preprocessing, image features extraction, training a classifier with learned features and utilizing a trained classifier for making tactical decisions. Though these traditional systems lead to huge contributions to image classification their performance is still unlike satisfactory, maybe due to improper expression of image semantics due to traditional feature extraction. Most of the existing methods employ local features, for example, Scale Invariant Feature Transform (SIFT) [3, 4], Histogram of Oriented Gradients (HOG) [5], Binary Robust Invariant Scalable Key points (BRISK) [6] to generate image representations. Although heterogeneous characteristic in images like scale perspective, complicated background, and illumination, leads to difficulty in representing a high-level concept from low features extracted which in turn limits the performance of traditional techniques and considers still as challenging task in the field of computer vision.

Alternatively, Deep Learning concept applied to processing of images has been revolutionized in various tasks of computer vision as image classification [7], representation or recognition and achieved good performance to become the best solution in many complex problems [8,9], in health informatics or medical image analysis [10] like image registration, denoising, segmentation.

Extending the structure of traditional Neural Networks (NNs), Convolutional Neural Network is the furthermost commonly used architecture in deep learning [11]. The deep CNN has testified a strengthen ability for image classification on several data sets such as ImageNet. Recently, significant work showed that CNN models pretrained on large data sets e.g., ImageNet [12], can be directly transferred to extract CNN visual features for visual recognition tasks [13,14] and learning of image features directly from CNNs, eliminate the need of manual feature extraction[15].

This work uses an already trained CNN network with its two different scenarios i.e. with SVM classifier and Transfer Learning for image classification. This paper is organized as follows: Section 1 provides a brief introduction and background study. Section 2 reviews the prior work related to CNN methods. Section 3 describes the methodology of pretrained network tasks for image classification in detail. Section 4 provides implementation of AlexNet with SVM classifier and Transfer Learning. Section 5 demonstrates results and discusses those results. Finally, section 6 presents the concluding statements based on experimental results.

LITERATURE REVIEW

The creativity and innovation make technology to its maximum peak having an idea in the field of Artificial Neural Networks (ANNs) that intimates the logical structure of the brain by considering classification as the foremost active application of locality. By keeping in a view of revolutionizing the future of Artificial Intelligence (AI), deep learning theory showed its improved performance [16, 17]. More recently, supervised deep learning showed great success in computer vision due to the availability of the massive amount of labeled training data as well as computing power [18]. With the addition of more hidden layers between the input and output layers in neural network architecture, deep learning networks, especially convolutional neural networks has gained interest in recent years for its improved and best solutions of classification [19]. CNN is evolved from the early neural network architecture Neocognitron [20]. By doing CNN convolution and sub-sampling layer, simple and complex cells in Neocognitron have been replaced and by reducing the number of parameters due to weight sharing in CNN, boost the computing capability of network.

With the development of technology, CNN showed successful results in the image classification problems. For example, for handwritten numeral recognition, Lecun [21] proposed his first CNN structure in the 1990s called LeNet. The Convolutional connections and back propagation algorithms were introduced due to derivation of many features from LeNet architecture in modern deep CNNs. Further, CNN's become popular in 2012, when CNN so-called as AlexNet, outperformed many previous designed methods on the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [22].Since then, Szegedy in 2015 [23] proposed a deep convolutional neural network based on 22 layers architecture on top of ImageNet for classification and detection. By keeping the computational power constant, the depth and width of the network were increased and based on Hebbian principle and multi-scale processing, the quality was optimized. Later, G.Huang [24] presented novel architectures by introducing direct connections between two layers with same feature-map size, like Resnets and DensNets which makes CNN's deeper and more powerful and achieves state-of-the-art results. Inspired by these improved performances while working on layers in CNN architecture, we fine-tune the layers of already trained models for image classification tasks.

RESEARCH METHODOLOGY

The proposed research methodology is based on pretrained CNN model with its two different tasks for image classification. The pictorial form of designed methodology is shown in figure 01.

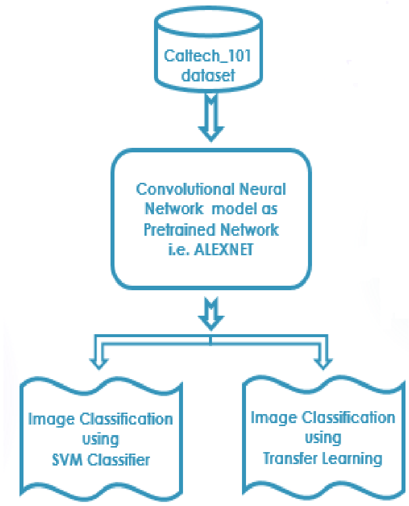

Figure 1.Designed Methodology

The designed methodology contains following steps:

Step 1: The image dataset is given to CNN pretrained model.

Step 2: The features are extracted using AlexNet.

Step 3 (i): To use learned features to train a classifier.

Step 3 (ii): Fine-Tuning the layers of AlexNet.

Step 4: Achieve image classification.

3.1. Convolutional Neural Network

It is a feed-forward neural network, uses grid-like topology for image features learning to perform classification directly from images.

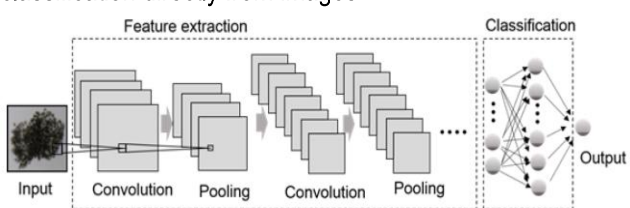

Figure 2. A Convolutional Neural Architecture [25]

Based on basic deep learning architecture discussed in [26, 27]. The input image is provided in the form of an array of pixels. The CNN performs a sequence of Convolutional, Rectified Linear unit/ activations and pooling operations on the hidden layers as step of image feature extraction. After learning the best features, the architecture performs classification.

3.2. AlexNet: A Pretrained CNN Model

This network is already trained on millions of images and can classify images into 100 object categories [28]. It is used for image feature extraction. AlexNet consists of 25 layers and its architecture is based on the basic architecture discussed in section 3.1. The names of AlexNet layers are shown in figure 02. Each layer has its own hyperparameters, weights, and biases.

1	'data'	Image Input
2	'conv1'	Convolution
3	'relu1'	ReLU
4	'norm1'	Cross Channel Normalization
5	'pool1'	Max Pooling
6	'conv2'	Convolution
7	'relu2'	ReLU
8	'norm2'	Cross Channel Normalization
9	'pool2'	Max Pooling
10	'conv3'	Convolution
11	'relu3'	ReLU
12	'conv4'	Convolution
13	'relu4'	ReLU
14	'conv5'	Convolution
15	'relu5'	ReLU
16	'pool5'	Max Pooling
17	'fc6'	Fully Connected
18	'relu6'	ReLU
19	'drop6'	Dropout
20	'fc7'	Fully Connected
21	'relu7'	ReLU
22	'drop7'	Dropout
23	'fc8'	Fully Connected
24	'prob'	Softmax
25	'output'	Classification Output

Figure 3. AlexNet Layers

3.3. SVM Classifier

The SVM Classifier is an efficient and popular supervised algorithm for image classification. It implements N classes as a one-against-all strategy, which is equivalent to a combination of N binary SVM classifiers. Against all other classes, it makes a single class by discriminating every classifier [29, 30]. For the classification process, a multiclass SVM classifier is trained using the image features obtained from pretrained network.

3.4. Transfer Learning

Transfer learning uses a pre-trained neural network. This technique allows the detachment of the last outer layers (classification layer) trained on a specific dataset, uses the remaining arrangement to retrain and get new weights related to interested classes [31, 32, 33].Transfer learning while training its newly assigned layers, plots the training progress plot which shows the training details with validation accuracy of pretrained network.

IMPLEMENTATION

The AlexNet is implemented as discussed below.

4.1. AlexNet with SVM classifier

The required number of categories is selected from the Caltech_101 object categories dataset. Various experiments are performed with 10, 20, 30, 40 and 50 categories. The AlexNet is used for feature extraction. The dataset is split into training and testing phase with 50-50 ratio in specified number of categories and images. All the images of dataset are resized with the input size of AlexNet network i.e. 227-by-227 and conversion of any grayscale image in dataset into RGB is performed. Finally, by using learned features, training of a multiclass SVM Classifier is done for image classification.

4.2. AlexNet with Transfer Learning

In these experiments, images are taken from Caltech_101 dataset. In experiments different number of categories is selected. The number of categories ranges from 10 to 50. The AlexNet is loaded for feature extraction. The Final Layers of AlexNet that are “Fully connected layer named as fc8”, “Softmax layer named as prob “and “Classification Output layer named as output” are replaced with new layers for adapting a new task or dataset. The dataset is split into training and testing phase with 50-50 ratio in specified number of categories and images. All the images of dataset are resized with the input size of AlexNet network i.e. 227-by-227 and conversion of any grayscale image in dataset into RGB is performed. Then, training of fine-tuned network by setting some training options is performed to classify a new dataset. Finally, images are classified by using the trained network.

RESULTS AND DISCUSSIONS

5.1. Hardware/Software

Experiments are executed on Intel Core ™ i7- CPU @ 2.70GHz (4 CPUs) machine running on Microsoft Windows 10. The tool used for executing techniques and writing codes in MATLAB R2018a using many packages i.e. Image Processing Toolbox, Deep Learning Toolbox team, Deep Learning Toolbox Model for AlexNet Network.

5.2. Results Plots of SVM Classifier and Transfer Learning

Once the classifier is trained e.g. SVM or after completion of training progress under Transfer Learning, the classification of the validation set, which corresponds the 50% of the total assigned images category is plotted as confusion matrix. It shows the performance of a classifier on a test data set. If a class is mislabeled as the other class between multiple classes, we can easily identify it from a confusion matrix.

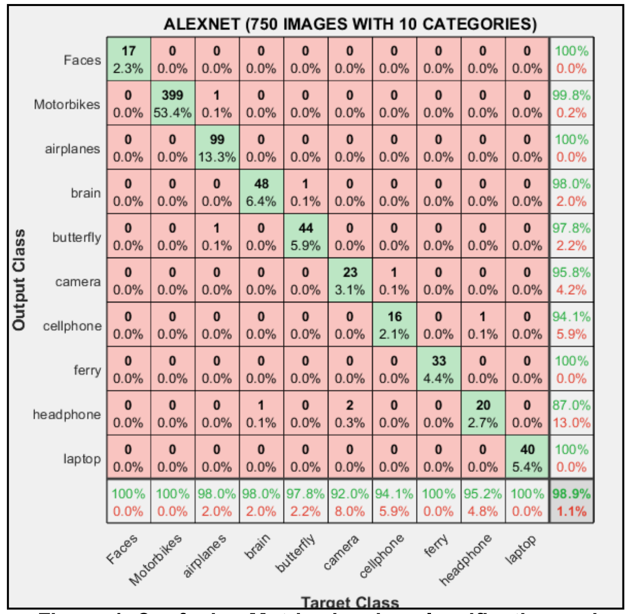

Figure 4. Confusion Matrix showing classification and misclassification of each category

The total of 1500 images is divided into 10 categories/classes. We have used the 50-50 ratio that means half images are taken as a training set of images while remaining half is used for testing set of images. The confusion matrix plot shows 50% (750 images) which are the tested images, classified after features extracted from AlexNet.

Transfer learning while training its newly assigned layers, plots the training progress plot which elaborates the training details with validation accuracy of pretrained network for the assigned number of images and categories. The Final Layers of AlexNet are replaced with new layers for adapting a new task or dataset.

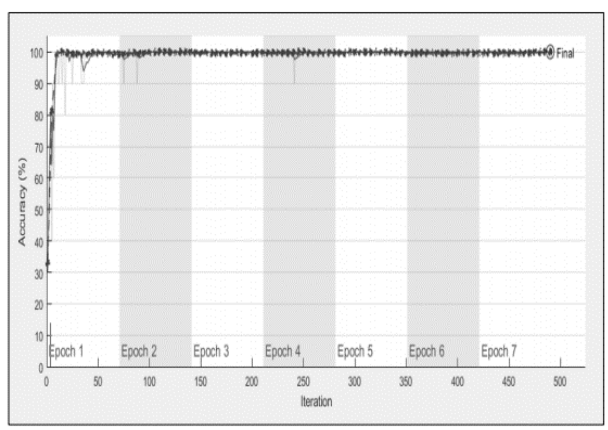

Figure 5. Accuracy during training with transfer learning

It is the accuracy plot that consists of iterations and epoch. The figure shows that the percentage of accuracy in the starting i.e. at epoch 1 is less. But with the completion of more epochs like 2, 3 and up to 7, the percentage of accuracy increases, and the plot becomes smooth.

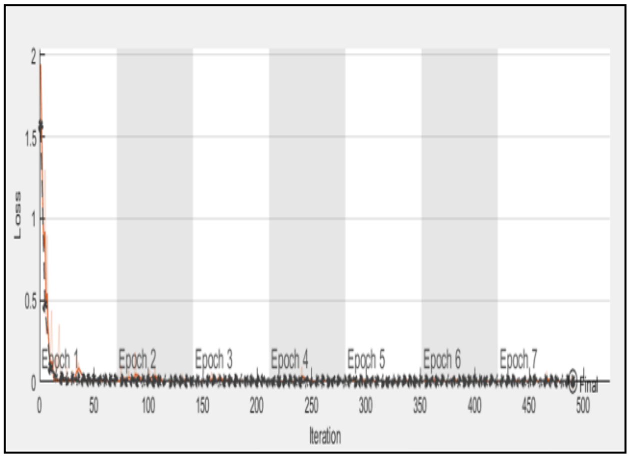

Figure 6. Training progress showing the loss while performing Transfer Learning

The error rate/ loss is high at the start, but with the completion of epochs it becomes less with jerks and becomes almost smooth when the accuracy approximates to 100 percent.

Results	
Validation accuracy:	100.00%
Training finished:	Reached final iteration
Training Time	
Start time:	27-Jan-2019 18:52:26
Elapsed time:	205 min 43 sec
Training Cycle	
Epoch:	7 of 7
Iteration:	490 of 490
Iterations per epoch:	70
Maximum iterations:	490
Validation	
Frequency:	5 iterations
Patience:	Inf
Other Information	
Hardware resource:	Single CPU
Learning rate schedule:	Constant
Learning rate:	0.0001

Figure 7. Training progress results summary displayed after completion of training.

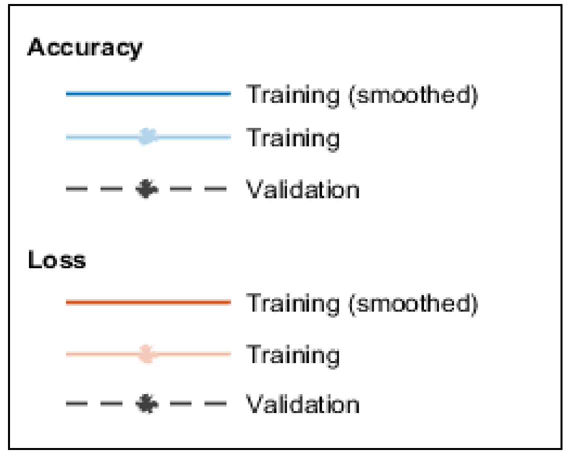

Figure 8. The description of accuracy and loss lines occurred while performing training as shown in Figure05 and Figure06

5.3. Performance of AlexNet with SVM Classifier and Transfer Learning

The comparative analysis of AlexNet after feature extraction with SVM classifier and Transfer Learning in terms of accuracy is plotted below:

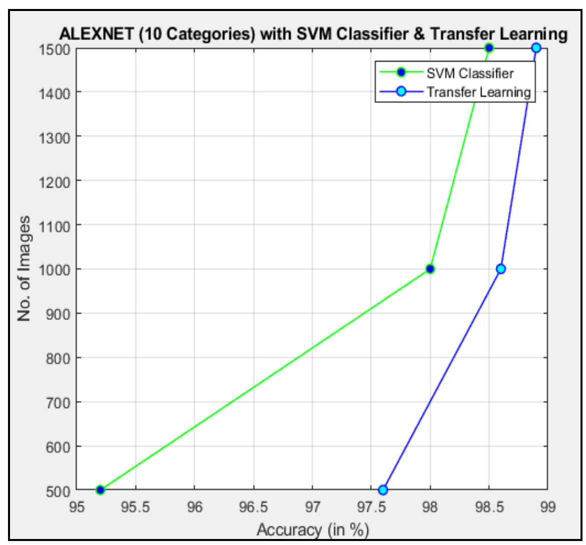

Figure 9. AlexNet accuracy plot of 10 categories with SVM Classifier and Transfer Learning

The experiment is carried out by changing the number of images like 500, 1000 and 1500 with keeping the number of categories constant. The increase in accuracy is achieved by increasing the number of images with constant number of categories. As a result, Transfer Learning has obtained greater accuracy than SVM classifier.

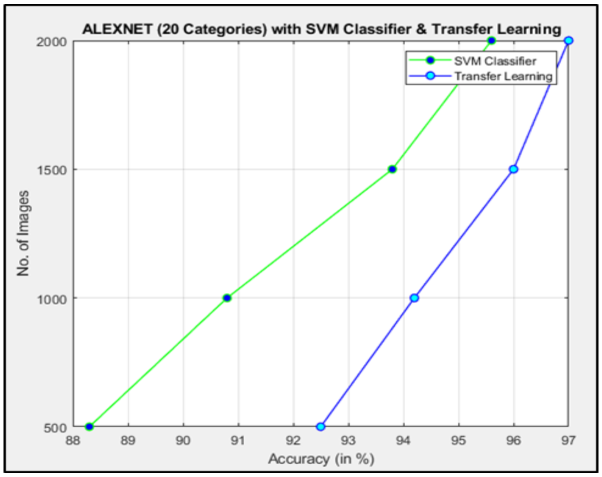

Figure 10. AlexNet accuracy plot of 20 categories with SVM Classifier and Transfer Learning.

The number of images is changed and taken up to 2000 with 20 categories. Initially, with 500 images accuracy of SVM classifier is 88% while accuracy of AlexNet with transfer learning is 92.5%. As shown in figure 10, as the number of training images is increased, the accuracy is also increased. SVM classifier goes from 95% to 98.5% with increase in images. Transfer learning goes from 97.5 to approximately 99% and achieved greater accuracy than SVM classifier.

The performance of AlexNet with transfer learning is much better as compared with SVM classifier. However, training time required to train AlexNet with transfer learning is much higher than the training time required for training the SVM classifier. Table 01 shows accuracy and training time for all experiments performed during this research.

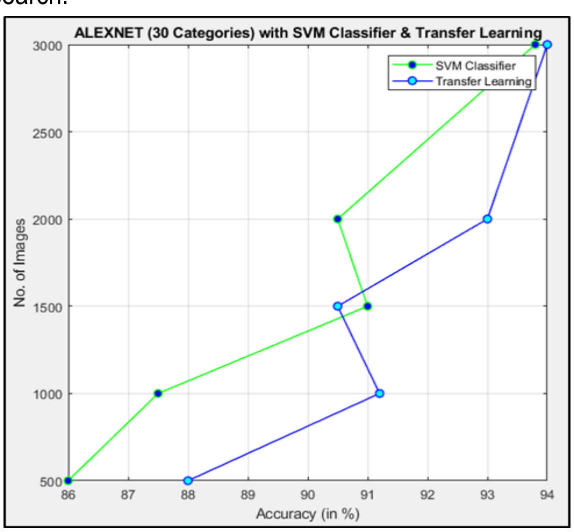

Figure 11: AlexNet accuracy plot of 30 categories with SVM Classifier and Transfer Learning.

The experiment is executed with changing the number of images and categories by taking the maximum of 3000 images with intervals on 30 categories and analyzed the abrupt increase in accuracy with SVM classifier and Transfer Learning.

The evaluation of the performance for the proposed methodology was measured by executing many experiments. The maximum of 4000 images with intervals was taken with 40 categories, 5000 images with 50 categories so that more analysis of proposed methodology can be analyzed. The few graphs of comparison of AlexNet with SVM classifier and Transfer Learning are discussed above. For more classification all results of accuracy with changing number of images and categories are summarized in table 01.

In the experiments, to show the effectiveness of the proposed image classification method, we mainly compare the results of pretrained CNN model in its two different tasks. And the results are reported in Table01. We can clearly see the difference in accuracy of SVM classifier and Transfer Learning and analyzed that as the number of images increases so the accuracy also increases in each category. The effect on accuracy i.e. decreasing is obtained by changing the number of categories from 10 to 20, 20 to 30 and so on with keeping images constant. For example, the accuracy while taken 1000 images in 10, 20, 30, 40 and 50 have been decreased in both scenarios i.e. with SVM and Transfer Learning.

Furthermore, the training time of pretrained network while performing experiments of the proposed method has also been noted and mentioned in Table 01. We can clearly see the difference in training time of SVM classifier and Transfer Learning and analyzed the increase in training time with increase in number of images. The greater change in training time, from minutes to hours is obtained between these both scenarios.

Table 01. Accuracy and Training Time results					
No. of Categories	No. of Images	Accuracy (In %)		Training Time	
		SVM Classifier	Transfer Learning	SVM Classifier (in Mints)	Transfer Learning (in Hrs.)
10	500	95.2	97.6	0.3	0.1
	1000	98	98.6	0.6	0.5
	1500	98.5	99	01	01
20	500	88.3	92.5	0.3	0.2
	1000	90.8	94.2	0.6	0.4
	1500	93.8	96	01	01
	2000	95.6	97	1.5	1.7
30	500	86	88	0.5	0.1
	1000	87.5	91.2	01	0.5
	1500	91	90.5	1.2	01
	2000	90.5	93	1.5	1.6
	3000	93.8	94	3.6	3.2
40	500	74.2	82	0.5	0.1
	1000	81.2	87.5	01	0.5
	1500	86.7	90	1.3	01
	2000	88.2	90	02	1.7
	3000	91.5	93.5	3.1	3.3
	4000	92.8	94.5	04	5.7
50	1000	82.3	84.6	1.1	0.5
	1500	84	87	1.8	01
	2000	86	90	2.8	1.5
	3000	88.4	90.2	3.7	3.3
	4000	90.5	91.3	06	5.8
	5000	92.6	94	09	8.9

CONCLUSION

In this paper, we classify images using pretrained CNN model (already trained on more than a million images) as Feature Extraction and Transfer Learning and achieved the improved state-of-the-art results of classification. AlexNet with SVM classifier achieved good accuracy with quite less training time that is in seconds to minutes. Transfer learning has also obtained quite a good classification accuracy results, greater and more stable in accuracy than SVM classifier but it takes training time in minutes to hours depending upon the number of images specified. The benefit of pretrained CNNs model is the elimination of laborious feature-engineering task that provides easiness in learning newly assigned task to network and can benefit large scale applications.

REFERENCES

1. Lu. Dongsheng, Weng Q, “A survey of image classification methods and techniques for improving classification performance” International Journal of. Remote Sensing, (2007), 28(5): 823-870.
2. Pooja Kamavisdar, Sonam Saluja, Sonu Kumari Agrawal, "A survey on image classification approaches and techniques" International journal of advanced research in computer and communication engineering, (2013) Vol 2.1, pp. 1005-1009.
3. Ajay Sahay and Anupama Gautam.,” Comparison between SIFT and SURF image forgery algorithms” International Journal of Computer Applications (0975-8887), (2017) Vol.164- No 2, pp.9-11.
4. Tao He, Yong Wei, Zhijun Liu, Guorong Qing, Defen Zhang, “Content-based image retrieval method based on SIFT features” IEEE International Conference on Intelligent Transportation, Big Data & Smart City. (2018) pp. 649- 652.
5. N. Dalal, B. Triggs, “Histograms of oriented gradients for human detection” IEEE computer society conference on computer vision and pattern recognition CVPR, (2005), pp. 886-893.
6. Stefan Leutenegger, Margarita Chli, and Roland Y. Siegwart, "BRISK: binary robust invariant scalable keypoints" computer vision IEEE international conference on computer vision (ICCV), (2011) pp. 2548-2555.
7. Vinayak Bharadi, Misbah Naimuddin, Arusa Irfan and Nikita Narayan, “Image Classification using Deep Learning” International Journal of Engineering Research & Technology (IJERT), (2017) Vol. 6 Issue 11, pp.17-19.
8. S. J. Sujit, R. E. Gabr, I. Coronado, M. Robinson, S. Datta, and P. A. Narayana, "Automated Image Quality Evaluation of Structural Brain Magnetic Resonance Images using Deep Convolutional Neural Networks," 2018 9th Cairo International Biomedical Engineering Conference (CIBEC), Cairo, Egypt, 2018, pp. 33-36
9. Wan-jie Liang, Hong Zhang, Gu-feng Zhang & Hong-Xin Cao, “Rice Blast disease Recognition using a Deep Convolutional Neural Network”. Nature- Scientific Report. (2019) pp. 1-10.
10. Gaobo Liang, Huichao Hong, Weifang Xie, Lixin Zheng. “Combining Convolutional Neural Network with Recursive Neural Network for Blood Cell Image Classification. IEEE Access. (2018), Vol.6
11. Patrick C Gray, Abram B. Fleishman, David J. Klein, Matthew W. McKown, Vanessa S. Bezy, Kenneth J. Lohmann, David W. Johnston, “A Convolutional Neural Network for Detecting Sea Turtles in Drone Imagery.“ Methods in Ecology and Evolution. (2018), pp. 1-11.
12. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li. Fei-Fei. , “ImageNet: A large-scale hierarchical image database” IEEE Conference on Computer Vision and Pattern Recognition, (2009), pp. 248–255.

13. S. He et al., "Combining Deep Learning with Traditional Features for Classification and Segmentation of Pathological Images of Breast Cancer," 2018 11th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 2018, pp. 3-6.
14. Y. Zhang, X. Sun, H. Sun, Z. Zhang, W. Diao and K. Fu, "High-Resolution SAR Image Classification with Deeper Convolutional Neural Network," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, 2018, pp. 2374-2377.
15. Gurnani, Ayesha &Mavani, Viraj., “Flower Categorization using Deep Convolutional Neural Networks” CoRR, (2017), pp.4321-4324.
16. Y. LeCun, Y. Bengio, and G. Hinton., “Deep learning,” nature, (2015) vol. 521, no. 7553, pp. 436.
17. J. Schmidhuber. , “Deep learning in neural networks: An overview”, Neural Networks., (2015) vol. 61, pp. 85-117.
18. S. Basu, et al. “Learning sparse feature representations using probabilistic quadtrees and deep belief nets” Neural Processing Letters, (2017), vol. 45, no. 3, pp. 855-867.
19. Jiuxiang Gu, et al., “Recent Advances in Convolutional Neural Networks” CoRR, (2017), pp. 187-332.
20. Fukushima, et al. “Neocognitron: A neural network model for a mechanism of visual pattern recognition. IEEE Tran. Syst. Man Cybern. SMC-13(5), (1983), 826-834.
21. Y. LeCun et al., “Handwritten digit recognition with a back-propagation network” in Proceedings Advanced Neural Information Processing System, (1990) pp. 396-404.
22. A. Krizhevsky, et al., “Imagenet classification with deep convolutional neural networks” Advances in neural information processing systems, (2012), pp. 1097-1105.
23. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D.Anguelov, D. Erhan, V. Van-houcke, and A. Rabinovich.” Going deeper with convolutions. In Computer Vision and Pattern Recognition (CVPR), (2015), pp. 1-9.
24. G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger., “Densely connected convolutional networks" in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), (2017), pp. 2261-2269.
25. “Convolutional Neural Network” [online]. Available: https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html.[Last Accessed: 30-Feb-2019].
26. V. Sze. et.al., “Efficient Processing of Deep neural networks: A Tutorial and Survey” Proceedings of the IEEE, (2017), Vol. 105, No. 12, pp. 2295-2329.
27. Heba Mohsen, A. El-Dahshan, M. El-Horbaty, AbdelBadeeh. , “Classification using deep learning neural network for brain tumors” ScienceDirect Future Computing and Informatics Journal (2018). pp. 68-71.
28. “Pretrained Convolutional Neural Network”[online].Available:https://www.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html. [Last Accessed]: 15-Feb-2019
29. Christopher J.C. Burges. , “A tutorial On Support Vector Machine for Pattern Recognition” Journal Data Mining and Knowledge Discovery, (1998), Vol.2 Issue 2, pp. 121-167.
30. S. Zhong, C. Chang, and Y. Zhang, "Iterative Support Vector Machine for Hyperspectral Image Classification," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, 2018, pp. 3309-3312.
31. Marcio Nicolau, et. al., “Fusarium Damaged Kernels Detection Using Transfer Learning on Deep Neural Network Architecture”, (2018) arXiv preprint arXiv:1802.00030v1.
32. C. Li and Y. Bai, "Fire Flame Image Detection Based on Transfer Learning," 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), Nanjing, China, 2018, pp. 370-374.
33. S. Y. Jo, N. Ahn, Y. Lee and S. Kang, "Transfer Learning-based Vehicle Classification," 2018 International SoC Design Conference (ISOCC), Daegu, Korea (South), 2018, pp. 127-128.