International Journal of Pure and Applied Mathematics
Volume 118 No. 7 2018, 185-190
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue
ijpam.eu

A Survey of Deep Convolutional Neural Network Applications in Image Processing

R Aarthi[1]
Dept of Computer Science and Engineering
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India
[1]r_aarthi@cb.amrita.edu

S Harini[2]
Dept of Computer Science and Engineering
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India
[2]CB.EN.U4CSE15316@cb.students.amrita.edu

***Abstract*—In the area of computer vision, artificial deep neural networks have won plentiful contests in pattern recognition and machine learning.Conventional architectures used for dealing computer vision problems are heavily under control on user features. But the new deep learning techniques have provided a substitute for automatically learning problem-related features. Thereforethe understanding of what type of deep networks are suitable for a given problem set is a challenging task.**

Keywords—Convolutional Deep Belief Network(CBBN),Shared Hidden Layer(SHL),ImageNet,Pooling.

I. Introduction

In computer vision, many problems are solved using machine learning techniques. Image classification and object recognition have generally been solved by using feature detection algorithm like SIFT[42], SURF[25], HoG[26] etc. These were normally followed by learning or classification algorithms like Support Vector Machines (SVM[41]). Subsequently, the result of these algorithms relies upon the features used. This implies that advances in computer vision algorithms were purely based on the available better group of features. These components of features began to turn out more complexity in result with difficulty with coming up better and more intricate features. There are two stages to be taken after in computer vision domain.One is feature design and second is designing of learning algorithms,both of which were to a great extent autonomous.

An Artificial Neural Network[43] is a data processing archetype that is influenced by the working of biological sensory systems, such as the brain process information. The key component of this paradigm is the novel structure of the information handling framework. It is made out of an extensive number of very interconnected handling components (neurons) working as one to tackle particular issues. Learning procedure of ANN is like humans, learning from a number of examples. An ANN is designed for a particular application, for example, pattern recognition or data classification, through a learning procedure. Learning in this frameworks includes changes in accordance with the brief associations that exist between the neurons.

A Convolutional Neural Network (CNN) consist of more than one convolutional layers with a subsampling layer and after that took after by at least one completely associated layers as in a standard multilayer system. All kinds of 2D data such as image or speech signal can be processed with the architecture of the convolutional neural network. This is obtained by weight, local connection, and pooling.Neural network systems, with their amazing capacity to get significance from entangled or loose information, can be utilized to concentrate designs and recognize patterns that are too unpredictable to ever be seen by either people or other computational strategies.

II. Literature

A.Classification

CireşAn, Dan, Ueli Meier, Jonathan Masci, and Jürgen Schmidhuber[7] explains about the implementation of an automated system that can be capable to distinguish the German traffic sign. The effective use of parameterized GPU for the implementation of DCNN [9]increased the efficiency up to 99.46%.Instead of using one DCNN for the entire process, they combined various DNN that trained under different conditions into multiple columns. And hence it achieves the highest accuracy even in the case of illumination and contrast changes. This method can be adopted for recognizing any other traffic signs. The building block of their work is Multi-column deep neural network(MCD)[10]. It replaces the convolutions with max pooling[11] layers in the traditional architecture. According to the architecture, preprocessed the input image by n different preprocessors and a random number of segments prepared on every section input. The averaging individual foresight of each DNN gives the final prediction. Each DNN module contains a group of convolutional and max-pooling layers. Convolutional layer performs the 2D convolution and the results are passed through the activation function. The use of max pooling layer in place of subsampling layer differentiates DNN from CNN. The size of the max pooling and convolutional kernels are chosen in a way that the output maps to a 1D feature vector. Dataset is preprocessed well before training. The random limited esteem for using translation, rotation, scaling is selected such a way that ±10% of the original size of the image and ±5 degree for rotation. The different DNN column is allocated with different weight. One input can be processed in a different way. Finally, the result obtaining from each DNN is averaged. This will reduce the average error. They used a system with core i7–950 (3.33 GHz), 24 GB DDR3, and four graphics cards of type GTX 580 and 9 DNN layer structure for implementing this work. Which can be prepared by using simple gradient descend. Each layer contains 25 net for 5 different data. The output achieves the recognition rate of 99.46%.Comparing with individual DNN it is a drastic achievement. Combining the individual DNN increases the accuracy as well as increases the robustness to various type of noise.

Imagenet is an image database composed by the world net echelons in which each node is portrayed by hundreds of images. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton[4] explains about the scope of DCNN in the field of big data. Here CNN is used to classify the 1.2 million of the high-quality image into more than 1000 categories[12].It states they could achieve the lowest error rate among existing method. The effective management of CPU system reduces the time required for executing. To improve the performance of the classifier in a large dataset, learn the more powerful model and find the best method to reduce the overfitting[13]. convolutional neural networks are the best choice due to its efficiency in dealing with large datasets and quality to reduce overfitting. It is a very improved GPU usage of 2D convolution[14].

Zou, Yuexian, Lei Li, Yi Wang, Jiasheng Yu, Yi Li, and W. J. Deng[3] is based on the wireless capsule endoscopy images by using the ability of powerful learning capability o DCNN it is used to separate the high semantic features from WCE images. It has achieved the 95% of accuracy by training the system with millions of images with various parameters. A wireless miniature encapsulated video camera is used to capture images inside the stomach. The camera takes 50000 to 60000 images per 8 hours of study. It is noninvasive, painless and disposable. By using these images the DCNN system classifies the digestive organs into the stomach, small intestine and colon. There exist several methods to classify these organs-by using DCT and PCA [15], using Receiving Operating Characteristic (ROC) curve analysis[16], classification using boundary based on energy function [17]. But this work concentrates on the capacity of the DCNN for classification.

The input is n number of 2D images of size n1 x n2.Similarly, the output data is images of size m1 x m2.The input is converted into output by applying convolution kernal[18]. For each location of the input, the filter detects the feature value. For each input map, the pooling function is applied. Maximum pooling calculates the maximum value overall a small neighborhood. But average pooling computes the average of the neighborhood pixels. In fully connected layer the output unit is connected to the input unit like a traditional neural network. Sigmoid or tanh function is used as the activation function. The final layer is acting as a classifier. Stochastic gradient descent method is used to minimize the error between actual output and obtained output. The input is 3 channel 96x96x3.first and second convolution layer 96 filter of size 5x5.But the third convolution layer has 196 filters. For the pooling layers, they used max pooling method of size 3x3.Initially, the weight of layer is set to zero mean Gaussian distribution with standard distribution 0.01.but the weight of fully connected layer is 0.1.They used mini-batch stochastic gradient descent method for the training process with a weight of 0.01. The momentum and weight decay are initialized to 0.9 and 0.004. The experiment is done in two steps. At first step, the dataset is applied to the svm[41],scsvm for a comparative study. The next step it s used the DCNN and verified that their method achieved the highest accuracy.

B.Object recognition

Due to the reasons of variation in the expression, occlusion, background face verification is a challenging problem.Huang, Rongbing, Fangnian Lang, and Changming Shu [8] Addressing the major problems like face verification and similarity representation, which apply for a regularization to learn nonlinear metric learning with CNN. Architecture can extend to recognize with dataset other than low datset[19]. DConvNet has many advantages, which can learn optimal shift-invariant local feature detectors[20] and frame representations that do not change with geometric distortion. Due to this reasons, DconvNet is used to implement nonlinear metric learning. The architecture is a combination of similarity distance metric learning and deep convolutional neural network.

The first level is to detect the face from the given image. Perform the dimensionality reduction on it.PCA[21] is suitable for this application.Next step is the process of checking the similarity between available image sets.The dataset has been pre-processed and stored in a lower dimension.By using distance metric like Euclidean distance, Manhattan distance we can able to find the image with minimum distance value[22].But these are mainly concentrated on the discrimination of the metric and do not have the robustness to intra-personal variations.

Bai, Jinfeng, Zhineng Chen, Bailan Feng, and Bo Xu [2] Explained about the technique used to extract the character from the image by using DCNN. The architecture is built upon shared hidden layers[23].It is implemented by using natural feature extraction method. Softmax layer is followed by this layer and it is language dependent[24]. By analyzing the universal features and the specific features, the system will finalize what the character is.This work explains about shared hidden layer deep convolutional neural network for image character recognition. The first few layers of the hidden layers work as universal feature extractor for every language. But the final layer is language dependent. Extracting character from a video or image is a challenging problem due to blur, uneven illumination, complex background, distortion etc. Due to the resistive nature of deep convolutional neural network towards distortion, scale and shift it gives a good result on character recognition compared to another method. The SHL-CNN architecture has the task of recognizing the English characters and Chinese character.

C.Quality enhancement

Nguyen, Kien, Clinton Fookes, and SridhaSridharan[5] explained the use of both supervised and unsupervised learning technique together to improve the learning capability. The architecture explains that replacing the layers of the DCNN with the trained unsupervised learning technique. For each step, the accuracy is estimated and compared with the existing systems. Found that it increases the recognition and classification accuracy and training computational expense. The major element of a recognition system is feature detection. The arrival of subjective feature detection methods such as SURF[25] and HOG[26] increases the accuracy of the learning algorithm.The deep learning architecture allows the system to learn feature themselves depending on the data.The deep learning architecture is classified into two-supervised and unsupervised learning. Here they are explaining about Convolutional Deep Belief Network and Deep convolutional neural network. This work enhances the learning capability of DCNN[27] by using CDBN[28]. Replace the earliest layer of the filter from the DCNN with corresponding layers from the CDBN. This reduces the computational expense and achieves competitive classification performance. And also reduces the training time from 7 days to 5 days. We can also try with other learning approaches.

The architecture that Dong, Chao, Chen Change Loy, Kaiming He, and Xiaoou Tang [6]explained is to increase the resolution of the image by applying convolutional neural network. Taking a low-resolution image as an input image and giving output as a higher resolution image using end to end mapping between lower and higher resolution images. The lightweight CNN structure explains the condition of workmanship of renovation quality and achieves better performance for practical good users. Image super-resolution is one of the classical problems of computer vision. An image that captured by a camera may have noise. Some cases that we cannot go for retaking the image. So the need for image super-resolution arises. There exist several methods for image super-resolution

- *Based on internal similarities[29,30,31]*
- *Mapping function between low and high resolution pair[32,33,34,35]*
- *Sparse coding method[36,39,40]*

In the proposed method it is the duty of the hidden layer to learn the dictionaries or manifold to model the patches. The patch extraction and aggregation is drafted in convolutional layers,
so are associated with the optimization. The whole network is named as a super-resolution convolutional neural network (SRCNN).

The input is a lower resolution image. The first step is to upscale the image into the desired size using bilinear or bicubic interpolation method. The mapping from the lower resolution image to higher resolution image is performed by the hidden layer of the neural network. In Patch extraction and portrayal removes the covering patches from the low-determination picture and speaks to each patch as a high-dimensional vector. These vectors contain an arrangement of highlight maps, of which the number equivalents to the dimensionality of the vectors. The non-direct mapping will nonlinearly map every high-dimensional vector onto another high-dimensional vector. Each mapped vector is theoretically the portrayal of a high-determination patch. These vectors incorporate another arrangement of highlight maps. Recreation operation adds up to the above high-determination patch savvy portrayals to deliver the last high-determination image. Contrasting the SRCNN and other methods, it is lightweight and has super execution. To excerpt more information we have to change the inner structure of the hidden layer.

Xu, Li, Jimmy SJ Ren, Ce Liu, and JiayaJia[1] presented a work to make use of the DCNN to resolve the deconvolution problem. By using these system problems like noise, saturation, camera artifact problems can be resolved. The first venture of the calculation is to find out the type of degradation and then apply the hierarchical structure. The hierarchy includes two trained supervised submodules with proper initialization. In problems like image and video processing, the main source of input is image or video captured by the camera. It must contain unavoidable artifacts such as noise, saturation, compression artifacts. Already there exists some model to remove those artifacts[37].The problem of generative model is that it has strong assumption on a type of noise.

The proposed calculation in light of the physical and mathematical qualities of the noise. Using convolutional neural network trained the system in such a way that will work for any type of noise without the prior knowledge. The input can give to the system without any preprocessing. Hidden layer works in the context of pseudo inverse deconvolution[38].It enables a practical system and, more importantly, provides an empirically effective strategy to initialize the weights in the network, which otherwise cannot be easily obtained in the conventional random-initialization training procedure. The final step is to perform deconvolution to retain the original image. The first step in the network architecture is to identify the need of using the exact kernel structure for the deconvolution application along with the convolutional neural network. It is suitable to use the one-dimensional kernels to avoid the variation in the size of weight. The second step is the process of supervised pre-training.Third step initialization of traditional neural network comes. Outliers rejection deconvolution can be used to reduce the noise. For that, another neural network is using.

III.CONCLUSION

In this paper, a survey based on an application of deep convolutional neural network is presented. This work will help to familiarize the application of neural network in detail. Application of convolutional neural network is broadly classified into three-object detection, classification, and quality enhancement. The multicolumn deep neural network is used for the traffic sign recognition system. But for image deconvolution two network modules are concatenated for improving the quality of an image. Verification of face is done by using DeconvNet package that contains DCNN architecture. A single image is directly fed into the neural network for super-resolution of an image.It will work on the luminous color space. To classify the images of large dataset like image-net, a neural network is well performed. Also possible to club the supervised learning technique like CDBN with deep architecture to enhance the learning. Using the coarse feature extraction capability of the shared hidden layer, it is used for the character recognition. Using the higher semantic image features, DCNN can be used for the real-time classification digestive organs from wireless sensor endoscopy images.

ACKNOWLEDGMENT

The authors gratefully acknowledge the contribution of LolithGopalan(M.Tech Batch 2015-2017) for his work on the original version of this document.

REFERENCES

[1] Xu, Li, Jimmy SJ Ren, Ce Liu, and Jiaya Jia. "Deep convolutional neural network for image deconvolution." In Advances in Neural Information Processing Systems, pp. 1790-1798. 2014.

[2] Bai, Jinfeng, Zhineng Chen, Bailan Feng, and Bo Xu. "Image character recognition using deep convolutional neural network learned from different languages." In 2014 IEEE International Conference on Image Processing (ICIP), pp. 2560-2564. IEEE, 2014.

[3] Zou, Yuexian, Lei Li, Yi Wang, Jiasheng Yu, Yi Li, and W. J. Deng. "Classifying digestive organs in wireless capsule endoscopy images based on the deep convolutional neural network." In 2015 IEEE International Conference on Digital Signal Processing (DSP), pp. 1274-1278. IEEE, 2015.

[4] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." In Advances in neural information processing systems, pp. 1097-1105. 2012.

[5] Nguyen, Kien, Clinton Fookes, and Sridha Sridharan. "Improving deep convolutional neural networks with unsupervised feature learning." In Image Processing (ICIP), 2015 IEEE International Conference on, pp. 2270-2274. IEEE, 2015.

[6] Dong, Chao, Chen Change Loy, Kaiming He, and Xiaoou Tang. "Learning a deep convolutional network for image super-resolution." In European Conference on Computer Vision, pp. 184-199. Springer International Publishing, 2014.

[7] CireşAn, Dan, Ueli Meier, Jonathan Masci, and Jürgen Schmidhuber. "Multi-column deep neural network for traffic sign classification." Neural Networks 32 (2012): 333-338.

[8] Huang, Rongbing, Fangnian Lang, and Changming Shu. "Nonlinear metric learning with a deep convolutional neural network for face verification." In Chinese Conference on Biometric Recognition, pp. 78-87. Springer International Publishing, 2015.

[9] Ciresan, D. C., Meier, U., Masci, J., Gambardella, L. M., & Schmidhuber, J. (2011a). Flexible, high performance convolutional neural networks for image classification. In International joint conference on artificial intelligence (pp. 1237–1242). AAAI Press.

[10] Meier, U., Ciresan, D. C., Gambardella, L. M., & Schmidhuber, J. (2011). Better digit recognition with a committee of simple neural nets. In International conference on document analysis and recognition (pp. 1135–1139). IEEE.

[11] Scherer, D., Müller, A., & Behnke, S. (2010). Evaluation of pooling operations in convolutional architectures for object recognition. In International conference on artificial neural networks (pp. 82–91). Springer.

[12] A. Berg, J. Deng, and L. Fei-Fei. Large scale visual recognition challenge 2010. www.image- net.org/challenges. 2010.

[13] K. Jarrett, K. Kavukcuoglu, M. A. Ranzato, and Y. LeCun. What is the best multi-stage architecture for object recognition? In International Conference on Computer Vision, pages 2146–2153. IEEE, 2009.

[14] Cope, Ben. "Implementation of 2D Convolution on FPGA, GPU and CPU."*Imperial College Report* (2006): 2-5.

[15] Jeff Berens, Michal Mackiewicz and Duncan Bell. "Stomach, intestine, and colon tissue discriminators for wireless capsule endoscopy images", Proc. SPIE 5747, Medical Imaging 2005: Image Processing, 283.

[16] Vilariño, F., L. I. Kuncheva, et al., "ROC curves and video analysis optimization in intestinal capsule endoscopy." Pattern Recognition Letters 27(8): 875-881, 2006.

[17] Lee, J., J. Oh, et al., "Automatic classification of digestive organs in wireless capsule endoscopy videos." Proceedings of the 2007 ACM symposium on Applied computing, ACM.

[18] LeCun, Y., K. Kavukcuoglu, et al., "Convolutional networks and applications in vision." Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on, 2010.

[19] Nguyen, H.V., Bai, L.: Cosine similarity metric learning for face verification. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part II. LNCS, vol. 6493, pp. 709–720. Springer, Heidelberg (2011).

[20] D. Lowe, “Distinctive image features from scale invariant keypoints,” International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[21] Palghamol, Tanuj N., and Shilpa P. Metkar. "Constant dimensionality reduction for large databases using localized PCA with an application to face recognition." In *Image Information Processing (ICIIP), 2013 IEEE Second International Conference on*, pp. 560-565. IEEE, 2013.

[22] Dokmanic, Ivan, et al. "Euclidean distance matrices: essential theory, algorithms, and applications." *IEEE Signal Processing Magazine* 32.6 (2015): 12-30.

[23] Huang, Jui-Ting, et al. "Cross-language knowledge transfer using a multilingual deep neural network with shared hidden layers." *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013.

[24] Le, Hai-Son, et al. "Structured output layer neural network language models for speech recognition." *IEEE Transactions on Audio, Speech, and Language Processing* 21.1 (2013): 197-206.

[25] Luo, Yong, and Yuanzhi Chen. "Robust matching algorithm based on SURF." Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2015 12th International Computer Conference on. IEEE, 2015.

[26] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference, June 2005, vol. 1, pp. 886–893 vol. 1.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in Neural Information Processing Systems 25, pp. 1097–1105. Curran Associates, Inc.2012.

[28] H. Lee, R. Grosse, R. Ranganath, and A. Ng, “Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations,” in Proceedings of the 26th Annual International Conference on Machine Learning, New York, NY, USA, 2009, ICML’09, pp. 609–616, ACM.

[29] Freedman, G., Fattal, R.: Image and video upscaling from local self-examples. TOG 30(2), 12 (2011).

[30] Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: ICCV. pp. 349–356 (2009).

[31] Yang, J., Lin, Z., Cohen, S.: Fast image super-resolution based on in-place example regression.In: CVPR. pp. 1059–1066 (2013)

[32] Bevilacqua, M., Roumy, A., Guillemot, C., Morel, M.L.A.: Low-complexity single image super-resolution based on nonnegative neighbor embedding. In: BMVC (2012).

[33] Jia, K., Wang, X., Tang, X.: Image transformation based on learning dictionaries across image spaces. TPAMI 35(2), 367–380 (2013).

[34] Timofte, R., De Smet, V., Van Gool, L.: Anchored neighborhood regression for fast example-based super-resolution. In: ICCV. pp. 1920–1927 (2013).

[35] Yang, J., Wang, Z., Lin, Z., Cohen, S., Huang, T.: Coupled dictionary training for image super-resolution. TIP 21(8), 3467–3478 (2012).

[36] Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse rep- resentation. TIP 19(11), 2861–2873 (2010).

[37] Štruc, Vitomir, et al. "Removing illumination artifacts from face images using the nuisance attribute projection." *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010.

[38] Westwick, David T., and R. E. Kearney. "Identification of physiological systems using pseudo-inverse based deconvolution." *Engineering in Medicine and Biology Society, 1995., IEEE 17th Annual Conference*. Vol. 2. IEEE, 1995.

[39] Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse representations. In: Curves and Surfaces, pp. 711–730. Springer (2012)

[40] Yang, Jianchao, et al. "Linear spatial pyramid matching using sparse coding for image classification." *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009.

[41] Sachin, R., Sowmya, V., Govind, D. "Dependency of various color and intensity planes on CNN based image classification", Advances in Intelligent Systems And Computing Vol. 678, pp. 167-177 (2018)

[42] Pathinarupothi, R.K, Dhara Prathap, J. Rangan, E.S., Gopalakrishnan, A.E, Vinaykumar, R, Soman, K.P.,"Single Sensor Techniques for Sleep Apnea Diagnosis Using Deep Learning", Proceedings - 2017 IEEE International Conference on Healthcare Informatics, pp. 524-529(2017)

[43] Raj, R.A., Haresh, V., Sini Raj, P. "Comparitive study of emotion detection using multilevel HMM and convolution neural networks from real-time videos", International Journal of Pure and Applied Mathematics, Vol.114, Issue 11 Special Issue, Pages 71-81(2017)