ICIC Express Letters
Volume **12**, Number **10**, October **2018**

ICIC International ©2018 ISSN 1881-803X
pp. 1017–1023

CLASSIFICATION OF SHAPE IMAGES USING K-MEANS CLUSTERING AND DEEP LEARNING

Swe Zar Maw[1], Thi Thi Zin[1], Mitsuhiro Yokota[1] and Ei Phyo Min[2]

[1]Graduate School of Engineering
University of Miyazaki
1-1 Gakuen Kibanadai Nishi, Miyazaki 889-2192, Japan
swezarmaw@gmail.com; { thithi; t0b210u }@cc.miyazaki-u.ac.jp

[2]University of Computer Studies, Yangon
No.(4), Main Road, Shwe Pyi Thar Township, Yangon, Myanmar
eiphyo.ucsy@gmail.com

Received February 2018; accepted May 2018

Abstract. *In this day and age, deep learning becomes attractive method for learning multilevel features and representation of data. In this work, we propose the new idea for image classification by combining K-means clustering algorithm and deep learning. Firstly, we handle K-means clustering algorithm for image preprocessing step because there are many variations in foreground and background of input images. Image preprocessing using K-means clustering can upgrade the achievement of the accuracy. We apply two dimensional deep convolutional neural network in order to classify shape, color, size and location of images in the BabyAIImageandQuestion Datasets into different classes. We trained 10,000 images that can properly classify near 100% and tested 5,000 images for each sub dataset in BabyAIImageandQuestion Datasets that can deliver outstanding performance. The goal of our research is to develop a system that targets for the visual ability of children which includes visual acuity, tracking, color perception, depth perception, and object recognition by effectively applying the deep learning algorithm. We also hope that our proposed method can provide very effective real world application.*
Keywords: Image classification, K-means clustering, Deep convolutional neural network, Deep learning

1. **Introduction.** Artificial Intelligence (AI) has been applied as a prior technique for decades and then advances in machine learning and deep learning are widely used in our time. Deep learning is an amazing technique which will continue to do so in the near future because it can verify one of the best techniques to beat the state-of-the-art performances [1]. One of the useful utilization of deep learning is image classification/object recognition. Human beings can easily identify and classify any images by their experience and knowledge. However, it was difficult for computers. Image classification is one of the widespread cases in image processing. To perform image classification, we mainly use the BabyAIImageandQuestion Datasets [2] to process our research. In this dataset, there are five different sub datasets (shape, color, size, horizontal location and vertical location). There are 10,000 images in each training set and 5,000 images in each testing dataset. The size of images is 32×32 and all images are grayscale colors. All images own different foreground and background grayscale colors. To solve this problem, we apply K-means clustering algorithm [3] in image preprocessing step in order to get binary image from grayscale image. After getting binary images, we implement image classification by using two dimensional deep convolutional neural network.

We need to cluster only two groups (foreground and background) in this dataset. Therefore, we use K-means clustering algorithm because it can yield compact clusters and can

DOI: 10.24507/icicel.12.10.1017

process Deep Convolutional Neural Networks (DCNNs) faster than other techniques. DCNNs trained by backpropagation can satisfy well on image classification tasks with a large amount of training images and many categories. The deep features extracted by using DCNNs can achieve state-of-the-art performance on the classification. These facts are the reasons of combining K-means clustering algorithm and deep learning to represent image classification. The objective of this paper is to classify shape, color, size and location of images as well as ask the questions to the children and give the answer to verify that it corrects or not.

This paper is organized as follows. Section 2 presents the problem statement and preliminaries of our process. In Section 3, we explain the detailed control design of our research. The main experimental results and performance accuracy are clearly described in Section 4. We conclude our approach with suggestion on future works and improvements in Section 5.

2. **Problem Statement and Preliminaries.** In this section, we express the problem statement of our process compared with some prior related works and preliminaries of our research. One of the main problem statements of image classification is extracting deep features from input images to classify into different classes. The practices, complication and promises of image classification have been solved by using machine learning techniques. In recent years, deep learning greatly becomes a popular technique to improve the performance of image classification by training into neural networks that imitates the brain of human.

The numerous number of papers concerned with image classification and deep learning have been published in recent years. The images are described as histograms of keywords by using the methods based on Bags of Visual Words (BoW). There were some of the first methods used in the challenges of image classification problems [4]. These methods were upgraded by the approximate global geometric correspondence which partitions the image into increasingly fine sub-regions and computes histograms of local features found inside each sub-region through the use of spatial pyramids [5]. In 2010, Lin et al. [6] used a basic deep learning algorithm, and achieved the best performance in the ImageNet: Large Scale Visual Recognition Challenge (ILSVRC). Perronnin et al. [7] acquired the state-of-the-art image classification result in 2011 by employing the Fisher kernel in order to extract higher order statistics, a previously unexplored alternative due to the zero order statistics base for BoW methods. The model proposed in [8] by Krizhevsky et al. won the ILSVRC 2012, which expresses a break point in large scale object recognition, since a large convolutional neural network had a well performing on natural image classification.

Further improvements and different architectures of convolutional neural networks had been created since 2013. A new visualization strategy is introduced by Zeiler and Fergus [9] to provide additional understanding on the behavior and function of intermediate feature layers. GoogLeNet [10] and VGG [11] used deeper architectures and reached a better performance level.

In our work, we use a new dataset, BabyAIImageandQuestion Datasets that aimed for preschooler. We firstly handle K-means clustering algorithm in image preprocessing steps. After implementing preprocessing steps, we perform two dimensional deep convolutional neural network for image classification. We can train our network with very fast training times compared to standard deep network approaches without using graphic cards (GPUs). We accomplish the great performance accuracy on these datasets.

3. **Control Design.** Firstly, we preprocess the images from dataset in order to use as input images because there are many variations such as intensity value, size, rotation and translation in images. Before performing the classification, we implement the clustering

on the intensity level of input images to distinguish background and object. In this paper, we use K-means clustering by using K value as 2 (2 class).

K-means is the famous "clustering" algorithm. It maps the feature value of each pixel or small region into feature space, generates cluster in the feature space and then remaps the pixel or small region into image plain and generate connected region. The K-means algorithm is as follows:

$$J(V) = \sum_{i=1}^{c_i} \sum_{j=1}^{c} (\|x_i - v_i\|)^2$$

where $\|x_i - v_i\|$ is the Euclidean distance between x_i and v_j

c_i: the number of data points in the ith cluster.

c: the number of cluster centers.

For given K, the data is classified into K clusters.

i. Set the initial K cluster centers.
ii. Calculate the distance between each data and cluster center and classify the data into nearest cluster.
iii. Re-calculate the cluster center as the average of classified data using the following equation.

$$v_i = \left(\frac{1}{c_i}\right) \sum_{j=1}^{c_i} x_i$$

where c_i represents the number of data points in ith cluster.

iv. Repeat 2 and 3 until no data has changed the belonging cluster.

Figure 1 shows our overall system flowchart. The size of input images is 32×32 pixels. Table 1 expresses the layers involved in the 2D DCNN architecture in order to process our work.

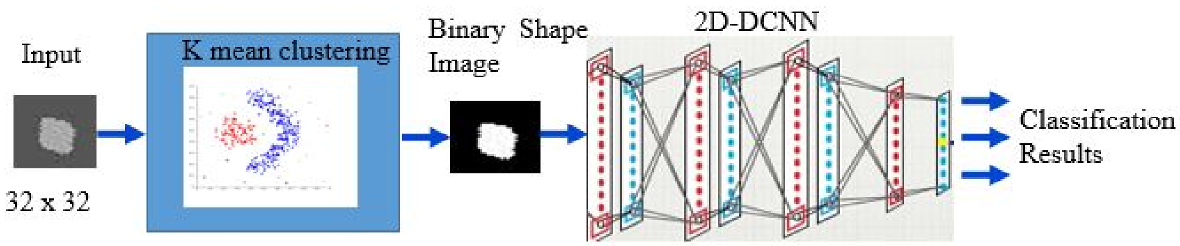


FIGURE 1. Overall system flowchart

TABLE 1. 2D DCNN architecture

Layer	Type	Feature Maps	Kernel Size
1	Convolution	6	5×5
2	Max Pooling	6	2×2
3	Convolution	12	5×5
4	Max Pooling	12	2×2
5	Convolution	18	2×2
6	Max Pooling	18	2×2

By applying K-means clustering, we get binary images that are trained and tested by 2 dimensional deep convolutional neural network to classify different classes. After doing the clustering, we perform the image classification using deep learning. We build two dimensional deep convolutional neural network architecture. In our architecture, there are 3 convolutional layers and max pooling layers, 1 fully connected layer and 1 output layer. Weights are randomly initialized and weight updating by stochastic gradient descent is performed until first convolution layer.

Figure 2 demonstrates the detailed construction of the convolutional neural network. In hidden layer 1, we use six (5×5 Kernel) matrics for convolution and then six (2×2 Kernel) strides are used for max-pooling. Each size of feature map for hidden layer 1 is 14×14 pixels. In hidden layer 2, twelve (5×5 Kernel) matrics and (2×2 Kernel) strides are applied again. Each size of feature map is 5×5 pixels. In hidden layer 3, we use 18 (2×2 Kernel) matrics for convolution and then 18 (2×2 Kernel) strides are used for max-pooling. Each size of output images is 2×2 pixels. In every hidden layer, we operate the sigmoid activation function after convolution layer. A sigmoid function is real-valued and differentiable, having a non-negative or non-positive first derivative, one local minimum, and one local maximum.

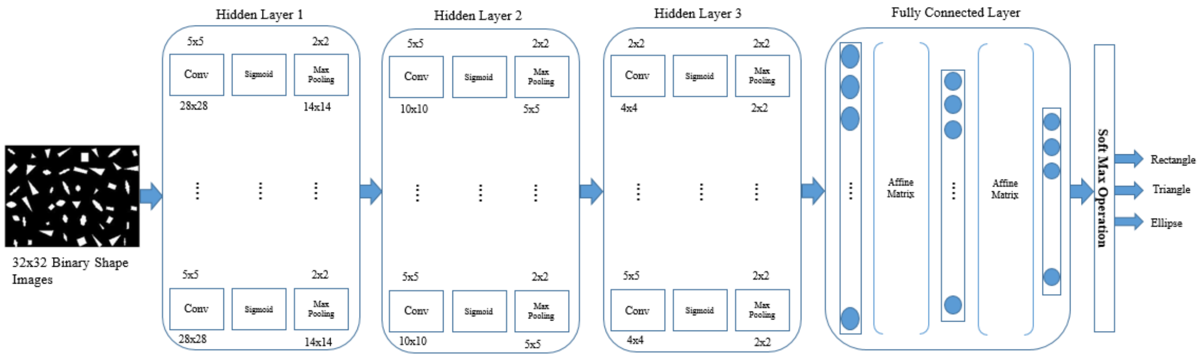


FIGURE 2. Two dimensional deep convolutional neural network

A sigmoid function is real-valued and differentiable, having a non-negative or non-positive first derivative, one local minimum, and one local maximum. We can evaluate the sigmoid function as follows:

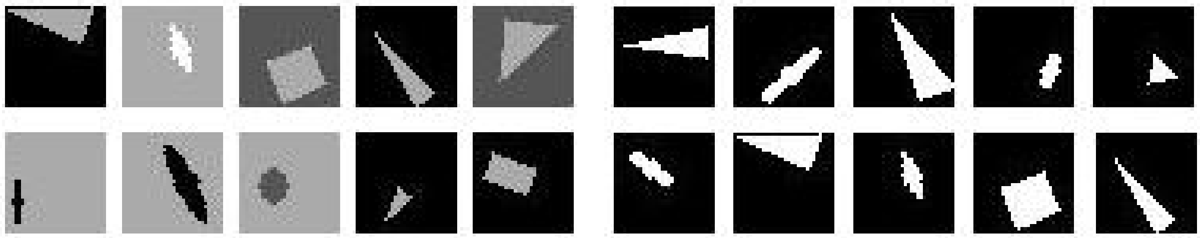
$$f(x) = \frac{1}{1 + e^{-x}}$$

where x = weight matrix $*$ input matrix + bias.

After that, we spread all output results into 5184 neurons for fully connected layer. We multiply these neurons by 100×5184 random affine matrix. The result of two matrices multiplicaition becomes 100 nodes. Then, 100 nodes are multiplied by 10×100 random matrix again. We put the final result 10 nodes into the Soft Max Operation for classification. The output results from soft max operation can provide us to consider the classes of the input images.

4. **Main Results.** We use BabyAIImageandQuestion Datasets to implement our proposed method. There are five sub datasets in BabyAIImageandQuestion Datasets. They are color, shape, size, horizontal location and vertical location. For each dataset, 10,000 images are used for training and 5,000 images are used for testing. All images are 32×32 size and grayscale images. The original images and binary images are shown in Figure 3.

The 10,000 shape images are performed by using K-means clustering to achieve binary images. However, 9747 images are correctly extracted. The accuracy of preprocessing for training dataset is 97.47%. As testing images, we handle 5,000 images and correctly extract 4858 images. The accuracy is 97.16%. For shape classification, we implement 9747 clustered training images to classify three categories (triangle, rectangle and ellipse). We get 98.5% as the accuracy. And, 4861 clustered testing images are employed and the accuracy of testing images is 97.5%. Finally, we can consider the overall accuracy is 98%. Tables 2(a) and 2(b) express the experimental results of shape image background extraction and classification results.



(a) (b)

FIGURE 3. (a) Some original images from training dataset and (b) some binary images after applying K-means clustering

TABLE 2. The experimental results for shape image classification

(a) Background extraction results

Training Dataset (#image)	Testing Dataset (#image)
Data – 10000	Data – 5000
Correctly Extracted – 9747	Correctly Extracted – 4858
Accuracy – 97.47%	Accuracy – 97.16%

(b) Classification results

Training Dataset (#image)	Testing Dataset (#image)
Data – 9747	Data – 4858
Accuracy – 98.5%	Accuracy – 97.5%

TABLE 3. The experimental results for color image classification

(a) Background extraction results

Training Dataset (#image)	Testing Dataset (#image)
Data – 10000	Data – 5000
Correctly Extracted – 9671	Correctly Extracted – 4630
Accuracy – 96.71%	Accuracy – 92.6%

(b) Classification results

Training Dataset	Testing Dataset (#image)
Data – 9671	Data – 4630
Accuracy – 96.85%	Accuracy – 96.79%

Before doing the color image classification, the 10,000 color training are extracted to get binary image by using K-means clustering. 9671 images are correctly extracted. The training accuracy is 96.71%. For testing images, we can extract correctly 4630 images over 5000 images. We get 83.54% as the accuracy. To implement color classification, 9671 clustered training images are identified into four classes (black, white, light gray and dark gray). We get 96.85% as the accuracy. And, 4630 clustered testing images are employed and the accuracy of testing images is 96.79%. Finally, we can consider the overall accuracy is 97%. The experimental results of color image classification are shown in Tables 3(a) and 3(b). Tables 4(a) and 4(b) demonstrate the experimental results of size image background subtraction and classification.

Tables 5(a) and 5(b) illustrate the experimental results for background extraction and classification results for horizontal location image classification.

TABLE 4. The experimental results for size image classification

(a) Background extraction results

Training Dataset (#image)	Testing Dataset (#image)
Data – 10000	Data – 5000
Correctly Extracted – 9853	Correctly Extracted – 4911
Accuracy – 98.53%	Accuracy – 98.22%

(b) Classification results

Training Dataset	Testing Dataset (#image)
Data – 9671	Data – 4630
Accuracy – 96.85%	Accuracy – 96.79%

TABLE 5. The experimental results for horizontal location image classification

(a) Background extraction results

Training Dataset (#image)	Testing Dataset (#image)
Data – 10000	Data – 5000
Correctly Extracted – 9965	Correctly Extracted – 4987
Accuracy – 99.65%	Accuracy – 99.47%

(b) Classification results

Training Dataset	Testing Dataset (#image)
Data – 9965	Data – 4987
Accuracy – 93.42%	Accuracy – 92.34%

TABLE 6. The experimental results for vertical location image classification

(a) Background extraction results

Training Dataset (#image)	Testing Dataset (#image)
Data – 10000	Data – 5000
Correctly Extracted – 9974	Correctly Extracted – 4982
Accuracy – 99.74%	Accuracy – 99.64%

(b) Classification results

Training Dataset	Testing Dataset (#image)
Data – 9974	Data – 4982
Accuracy – 95.28%	Accuracy – 96.47%

Table 6 displays the experimental results of vertical location image classification. Tables 6(a) and 6(b) are for background extraction and classification results.

5. **Conclusion.** In this paper, we implemented the K-means clustering for image preprocessing and it can provide to get high accuracy for image classification. Deep learning can support to train huge amount of data at the same time. Moreover, it can also do achieve the satisfied results for classification. In future, we would like to apply our proposed method to other deep learning datasets and compare the accuracy results with other image classification techniques. We hope that we can receive high accuracy for these datasets.

Acknowledgment. We would like to express our sincere gratitude to our senior Ms. Cho Nilar Phyo for her good intention and for supporting to our research.

REFERENCES

[1] Y. LeCun, Y. Bengio and G. Hinton, Deep learning, *Nature*, vol.521, no.7553, pp.436-444, 2015.
[2] O. Breuleux and J. Louradour, *DumitruErthun, BabyAIImageandQuestion Datasets*, image_and_question_data_1obj.tar.gz, 2008.
[3] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu, An efficient k-means clustering algorithm: Analysis and implementation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.24, no.7, pp.881-892, 2002.
[4] G. Csurka, C. R. Dance, L. Fan, J. Willamowski and C. Bray, Visual categorization with bags of keypoints, *Proc. of the ECCV International Workshop on Statistical Learning in Computer Vision*, pp.59-74, 2004.
[5] S. Lazebnik, C. Schmid and J. Ponce, Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol.2, pp.2169-2178, 2006.
[6] Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, L. Cao and T. Huang, Large-scale image classification: Fast feature extraction and SVM training, *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.1689-1696, 2011.
[7] F. Perronnin, Y. Liu, J. Sanchez and H. Poirier, Large-scale image retrieval with compressed Fisher vectors, *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.3384-3391, 2010.
[8] A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems*, pp.1-9, 2012.
[9] M. D. Zeiler and R. Fergus, Visualizing and understanding convolutional networks, *Proc. of the ECCV International Workshop on Statistical Learning in Computer Vision*, pp.818-833, 2014.
[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, Going deeper with convolutions, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-9, 2015.
[11] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *Proc. of the ICLR*, pp.1-14, 2015.