

ORIENTATION AND SCALE BASED WEIGHTS INITIALIZATION SCHEME FOR DEEP CONVOLUTIONAL NEURAL NETWORKS

AZIZI ABDULLAH
WONG EN TING

ABSTRACT

Image classification is generally about the understanding of information in the images concerned. The more the system able to understand the image contains, the more effective it will be in classifying desired images. Recent work has shown that the convolutional neural network (CNN) paradigm is useful for obtaining more accurate image classification results. A crucial component in the CNN is the convolution filters which consist of a series of predefined filter weight initialization values. The filter weights are then automatically learned by the neural network throughout the back-propagation training algorithm. However, most initialization schemes used in the deep convolutional neural networks are mainly to deal with vanishing gradient problems. Thus, selecting optimal weights are crucial to improve convergence and minimize the complexity which can enhance the generalization performance. One possible solution is to replace the standard weights with parameterized filters that proven to be efficient in extracting useful features such as Gabor filter bank. The Gabor filter bank is popular due to its ability in dealing with spatial transformation, especially on edges and texture information of different scales and directions. Thus, in this paper, we investigate the effect of utilizing Gabor and convolutional filters on small size kernels of deep VGG-16 architecture. The standard VGG-16 filter is replaced with the Gabor filter bank to obtain uniform distribution at all layers of the network. The result shows that the orientation and scale weights initialization scheme outperforms the standard filter weights on an image classification problem.

Keywords: Convolutional neural network, Gabor filter, Weight Initializer, Machine learning

INTRODUCTION

The need for an efficient image recognition system is crucial due to its vast application potential such as in biomedical, commerce and agricultural. In the literature, deep convolutional neural network (DCNN) models have shown its remarkably performance in various computer recognition tasks. One of the main issues in CNNs is that it is heavily reliant on large datasets for training. The large dataset provides some advantages in constructing a diverse classifier. Thus, the need for distinctive feature extractor techniques to examine information in the dataset is crucial. In CNN, the convolutional layer is responsible for extracting local features for image descriptions. It contains a set of filter weights used for image processing tasks such as texture analysis, edge detection and image enhancement. In many feature extraction tasks, it is useful to apply some parameterized filter attributes such as orientation and scale for feature improvement.

In CNN, the most expensive part is the learning image features from convolutional filters. The filter weight initialization schemes can be divided into three categories such as random, statistical distribution, and via unsupervised feature learning (Goodfellow, et al., 2015). However, most of the schemes used in CNN is statistical and supervised learning. For example, in (Glorot, et al, 2010) used a function that estimates the standard deviation of hyper-parameters initialization for a CNN layer based on layer input and output size. In (He, et al,

2015) proposed normal or uniform distribution functions of hyper-parameters initialization based on the neural network topology and type of activation functions. However, these methods investigated mainly on the weight initializers while training deep CNNs. They evaluate loss against the true labels and back-propagated through the network by generating the gradients for all layers, which is known as vanishing gradient problem (Alekseev, et al, 2010 and Daniel, 2018). It shows that the selection of weight initializers are crucial for improving convergence and minimize training complexity to improve the model generalization performance. In (Coates, et al, 2011) proposed an unsupervised learning algorithm namely the k-means clustering to learn centroid as a convolutional kernel. In this approach, a set of small image patches are constructed and the clustering algorithm is applied to group similar cluster together. But in (Chan, et al., 2015) proposed to use principal component analysis (PCA) to learn multi-stage filter banks for deep learning. Learning features with an unsupervised scheme is one possible solution to reduce the limitation of hand-crafted features. However, one simple and inexpensive solution to choose the weight initializers is random filters. It also works reasonably well in CNNs as reported in (Jarrett, et al, 2009 and Pinto, et al., 2011).

Another plausible way to choose the weight initializers is to use features that resemble the convolutional kernels from supervised training. For example, in (Krizhevsky, et al, 2012) proposed to train filters on ImageNet for a deep neural network. When the filters were visualized, some filter coefficients resemble the Gabor basis functions or filters (Gabor, 1946). Thus, we have taken this as a motivation to replace the VGG-16 layers with the Gabor filter bank to increase the accuracy for classifying images. In image processing, Gabor filter can be classified as one of the anisotropic filtering techniques that is capable in dealing with spatial transformation. The approach is widely used to extract robust image representation and useful features like edges and contours within an image. It is believed that the filter bank capable of modulating over an image to extract useful features from different angles and scales. Furthermore, they are claimed by many computer vision researchers to be closely mimicking the human visual system for encoding natural images. There are many papers discussed the use of Gabor filters within their networks, such as in (Luan, et al, 2018; Yao, et al., 2016; Sarwar, et al., 2017; Alekseev, et al., 2019). The proposed networks have shown promising results in terms of speed and accuracy on various dataset benchmarks.

Following this, some studies have been made in recent years to incorporate the Gabor filter bank in DCNN. In (Luan, et al., 2018), a set of learn-able filters that trained on ImageNet dataset known as Alexnet were replaced by Gabor filters to extract useful image features in the DCNN. The experimental results demonstrate that Alexnet using Gabor filters having the lowest error rate of 0.42, whereas Alexnet the learn-able filters having the error rate of 0.73. Based on this research, it can be observed that filter design will affect the model accuracy. The results show some improvement especially in recognizing objects in which the scale and rotation changes occur frequently. Most of the works on the Gabor filters are focused on employing the Gabor function in the first or shallow layer of the network. In (Alekseev, et al, 2019) proposed to use the parameters of Gabor functions in the first layer of the Alexnet deep convolutional neural networks. The experimental results shown that, applying Gabor layer significantly outperforms the standard Alexnet's filters with better convergence rate as the number of hyper-parameters is reduced. Besides, they found that the Alexnet's convolutional filters are often redundant and some of them are similar to Gabor filter bank which is used for feature extraction. In (Yao, et al., 2016) they simply employ Gabor filters to generate Gabor features and uses them as input to a CNN, and (Sarwar, et al, 2017) only fixes the first or second convolution layer by Gabor filters, which mainly aims at reducing the training complexity of CNNs. In addition, (Glorot, et al., 2010) demonstrated that the filter weights initialization are important to extract useful features from images in DCNNs. In (Pham, 2019), proposed to

replace the first layer of a CNN with the Gabor filter to increase speed and accuracy for classifying images. They have created two simple 5-layer AlexNet CNNs and employ grid-search and random-search for initializing the Gabor filter weights. Thus, selecting the optimal Gabor filter bank is not an easy task, and is made more difficult and challenging by having some parameters that need to be considered in the Gabor filter bank.

In this paper, we propose the replacement of VGG-16 with the Gabor filters. In contrast to previous works, the study aims to investigate the effectiveness of a small kernel or filter size of the Gabor orientation and scale parameters in the deep convolutional neural network. We argue that, by applying the filter in the VGG-16 deep layers may allow the algorithm to learn more complex features in understanding images as well. By replacing with the Gabor filters may overcome the vanishing gradient problem that could reduce the training complexity and improving convergence of conjugate gradients, which can enhance the learning models. For experiment, the number of filters, kernel size and layers are retained to show the depth of the network. We have tested the method on a collection of images taken from ImageNet (Deng, et al, 2009) and Google image website.

The contribution of this paper. (1) We compare the standard VGG-16 filters and the proposed orientation and scale of Gabor filters for image classification (2) we demonstrate the effectiveness of the Gabor filter to obtain features for images. Standard VGG-16 filter as a feature detector is not the best to describe image objects, but efficient in selecting specific properties of Gabor filters such as orientation and scale can be. The rest of the paper is organized as follows, Section 2 describes fundamental principles of deep neural network and its architecture. Section 3 describes our parameters of Gabor functions for the weight initialization in convolutional layers. Experimental results on the fruit dataset are shown and discussed in Section 4. Section 5 concludes the paper.

RELATED WORK

With recent advancements in deep learning, computer vision problems such as object classification and detection can be solved more effectively. The deep learning solutions have shown significant performance improvements in many benchmark datasets.

DEEP NEURAL NETWORKS

Deep architectures have been shown to be effective in image classification. Researchers have reported that the scheme is extensively been studied and achieve impressive results in many computer vision benchmark datasets such as in MNIST (LeCun, et al, 2010), CIFAR (Krizhevsky, et al., 2009) and ImageNet (Krizhevsky, 2012). In general, deep learning is one of the most effective supervised machine learning algorithms for solving complex problems. It learns by using a hierarchical structure that extracts different relational or spatial features from different layers that perform different tasks. For example, it uses the lowest layer of this scheme as a feature detector to detect simple patterns. After that, these patterns have become inputs into deeper, following, layers that form more complex representations of the input data. There are several methods of learning deep architecture. One of the most popularly used in computer vision is the convolution neural networks (CNNs). The CNN is a data-driven algorithm that learns robust spatial and robust representation of visual features from training data. The generated visual features are then learned and used across the whole image for image classification. The scheme is generally known for its invariant properties towards many conditions such as shift, translation, rotation, and scale. Thus, if an object shifted or translated across the input image, the network still can detect it. Such features gave ease and reliability

and very useful for image classification tasks.

In the literature, to get a good classification result, the deep learning network trains using a huge number of different image data. Typically, deep convolutional neural network algorithms may take days or even weeks to train on very large datasets to construct a strong model for making accurate decision. However, there exists a solution that is able to reduce the training time by using the pre-trained models as the main source architecture for solving related problems, which is known as transfer learning (Ventura, et al, 2007 and Schuldt, et al., 2004). In this scheme, a selected neural network model trains on a training dataset that is similar to the one that has been under investigation. After that, the model is used directly to classify new instances of other similar problems. The scheme has some benefits for deployment such as can reduce time for training time and enhance the classifier model for classification. For example, (Ren, et al., 2015) uses a transfer learning scheme for object detection on the Pascal VOC dataset challenge, (Long, et al, 2015) for semantic segmentation and (Krizhevsky, et al, 2012 and Simonyan, et al., 2015) for other image classification problems.

DEEP LEARNING ARCHITECTURE FOR IMAGE CLASSIFICATION

An object classification system needs effective visual image descriptors and machine learning algorithms to obtain good performance results. In general, a descriptor is used to represent the visual image information, and after that a machine-learning is applied to learn this visual information. There exist many common feature descriptors to describe images. However, the selected descriptors should be informative and meaningful to describe visual object such as in the human visual system.

In general, object classification is a task in computer vision that involves identifying the presence of one or more intended features in a given test image. It is a challenging problem and consists of two crucial processes namely feature extraction and object classification. The feature extraction extracts the most distinctive and important features of visual information and transform it into a vector data format. Then, the vector features are used to construct a model by using a machine learning algorithm. In CNN, the important features are automatically extracted and learned from a series of images via at different convolutional layers. A simple CNN may contain three main layers for building a model for classifying objects such as convolutional, pooling, and full-connected layers. Each of these layers has different hyper-parameters that can be optimized to perform different tasks. For example, convolutional layers are the layers that construct the visual local feature map. The characteristic of the layer is controlled by series of important hyper-parameters such as the kernel number and the size of convolutional kernels for filtering the original image. The next layer is the pooling layers, which typically used to reduce the dimensionality of the deep network by a sub-sampling operation. The most popular function used in this step is max pooling or average pooling which takes the maximum or the average value in a filter region respectively. The last layer is fully connected layers, which used to flatten the image features before classification. One major challenge in using CNNs is how to identify the optimum way in designing model architectures that best utilize these important CNN layers. Following this, many deep learning algorithms have been proposed to enhance and modernize the deep learning architectures resulted in the rapid advancement in the state-of-the-art for solving computer vision tasks such as object detection and localization (Redmon, et al, 2018; Liu, et al., 2016; Ren, et al., 2015).

The first successful of CNN architectures as reported in the literature is LeNet-5 (LeCun, et al., 1988). The system was developed for use in a handwritten character recognition problem. The proposed CNN architecture gives an accuracy of 99.2% on the MNIST dataset. Fig. 1 shows the LeNet-5 CNN for handwritten character recognition. The proposed LeNet-5

architecture consists of seven deep layers with an input grayscale image of dimension size 32×32 . The model proposes three main types of layers in a convolutional neural network namely the conventional layer, pooling layer, and fully connected layer. It starts by extracting image patterns by convolving the input image followed by an average pooling layer. This process is repeated two and a half times before the output feature maps are flattened and become an input to fully connected layers for a final prediction. LeNet-5 model uses a small number of filters i.e. 6 filters with the size of 5×5 pixels in the first layer. Next, after pooling, the number of filters used for convolution increased to 16 filters with the same filter size to extract more distinctive image features. Following the success of the LeNet-5, a great effort has been made to develop more complex deep learning architecture such as AlexNet (Krizhevsky, et al., 2012), VGG (Simonyan, et al., 2015), GoogLeNet (Szegedy, et al., 2015) and etc. In this case, the modern architectures show the trend of increasing the number of filters and the depth of the network in the proposed approaches. However, in this paper, we use VGG-16 due to easily accessible pre-trained models available in Keras (Gulli et al., 2017) and become robust benchmarks for comparison (Dickerson, et al., 2017).

VGG-16 ARCHITECTURE

The method was proposed by Simonyan et al., who showed that the small size of convolutional filters and the more network depth can improve model performance. In general, VGG-16 is motivated by the success of AlexNet CNN architecture (Krizhevsky, et al., 2012). The model has demonstrated remarkable success in the 2014 ImageNet large scale visual recognition challenge 2014 (ILSVRC-2014) (Russakovsky, et al., 2015). Besides, the algorithm is known in having better visual representation due the depth of the network layers and smaller filter dimensions GoogLeNet (Sultan, et al., 2018). The main differences between these schemes are (a) the use of a large number of a small filter of the size 3×3 and 1×1 with the stride of one for convolutions. (b) it has 16 weighted layers i.e. 13 in the convolutional layers and three in the fully connected layers. The number of 16 layers shown to be generalizable in the ILSVRC- 2014 dataset. (c) it uses a very large number of filters used for convolutions. In this model, the number of filters increases with the depth of the CNN architecture from 64 filters at the start and increases through 128, 256, and 512 filters for the following layers of the mode. (d) the max-pooling layers of size 2×2 are used in the most convolutional layers for image sub-sampling. Fig. 1 shows the VGG-16 architecture.

THE PROPOSED METHOD

GABOR FILTERS

The convolutional layer is one of the core building blocks in DCNNs. It consists of a set of kernels (a.k.a filters) and the main task of this layer is to extract some salient features of visual images such as edges and textures information. One popularly used kernel is Gaussian distribution or also known as Glorot/Xavier (Glorot, et al., 2010), random (Pinto, et al., 2011 and Jarrett, et al., 2009) and via unsupervised learning algorithm (Chan, et al., 2015 and Coates, et al., 2011). However, using the Gaussian type initialization weights in the models still have some limitations to accurately classifying objects due some limitations to address such as spatial frequency structural from images. Besides, using learn-able filters via unsupervised learning is also possible. But, some studies have shown that the filters are often redundant and approximating or similar to a series of anisotropic filtering i.e. Gabor filter bank (Krizhevsky, et al., 2012). Thus, we argue that it is possible to learn the deeper VGG-16 deep learning architecture using this filter type? The question of how the best practice principles could be

operationalized to the parametrized of Gabor filter in the deep learning convolutional layer is still open.

In the literature, Gabor filters provide a local representation by utilizing different scales and orientations for compact image representation. It claims to be efficient in preserving the spatial information relations

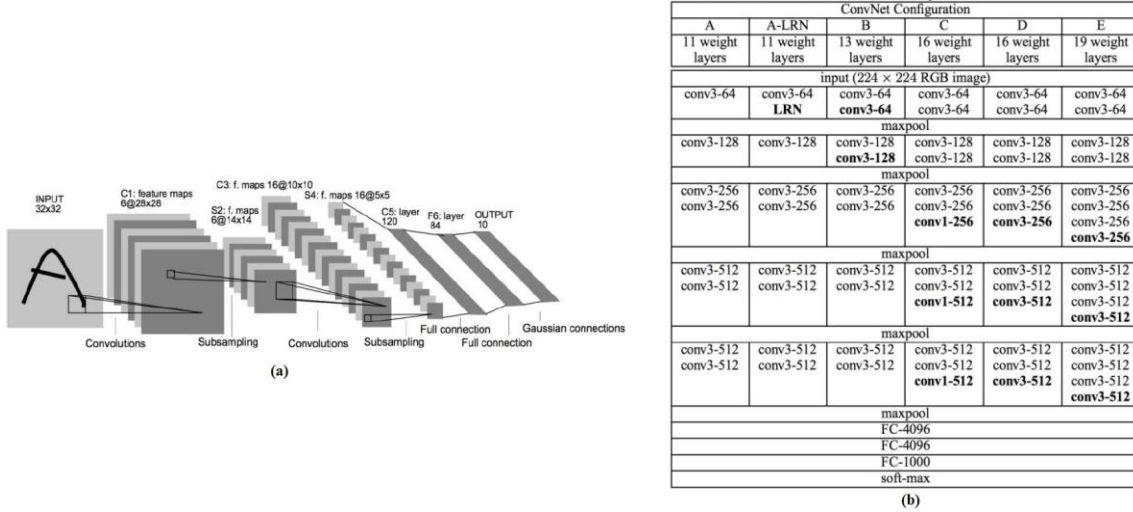


FIGURE 1. Left side shows the LeNet-5 Convolutional Neural Network. Right side shows the VGG-16 Convolutional Neural Network (taken from the 2014 paper).

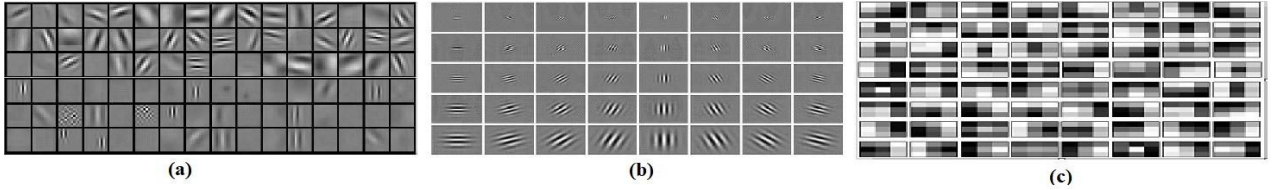


FIGURE 2. (a) Illustrates Alexnet filters. (b) Shows Gabor filters. (c) Shows the standard VGG filters of size 3 x 3. Some of learned AlexNet and VGG filters are similar to Gabor filters. This shows that the Gabor filters may contain some importance of hyper-parameters initializing weights of deep neural networks. Thus, it could help in reducing the convergence problem and makes it easier to train a very deep CNN architecture.

which are important intrinsic properties of images. Besides, these features are widely used for image description and classification (Eqlouss, et al., 2012 and Abdullah, et al., 2009). Moreover, its spectral features are widely used in various pattern analysis application (Huang, et al., 2004 and Ayad, et al., 2019). The Gabor filters are based on a sinusoidal plane wave with particular frequency and orientation. In this work, the real product of a sinusoid and a Gaussian is used as follows:

$$g(x, y; \lambda, \theta, \phi, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \phi\right) \quad (1)$$

where

$$x' = x \cos \theta + y \sin \theta \quad (2)$$

$$y' = -x \sin \theta + y \cos \theta. \quad (3)$$

In Equation (1), λ represents the wavelength of the sinusoidal factor, θ represents the orientation of the normal to the parallel stripes of a Gabor function, ψ is the phase offset, σ is

the standard deviation of the Gaussian envelope and γ is the spatial aspect ratio, and specifies the ellipticity of the Gabor function. Fig. 2 shows the visualization of learn-able filters i.e. Alexnet and Gabor filters. Some of the learned Alexnet filters resemble and are very similar to Gabor filter bank.

GABOR WEIGHTS INITIALIZER

We propose to use a bank of Gabor filters in VGG-16 deep learning architecture. In contrast to this work in (Alekseev, et al., 2019), we propose to apply the Gabor filters in all of VGG-16 layers. We believe, by initializing at all layers will result to similar and uniform initialization values that can reduce the training complexity and convergence rate which will improve the learning process. In this experiment, the parameters of λ , θ , and γ are

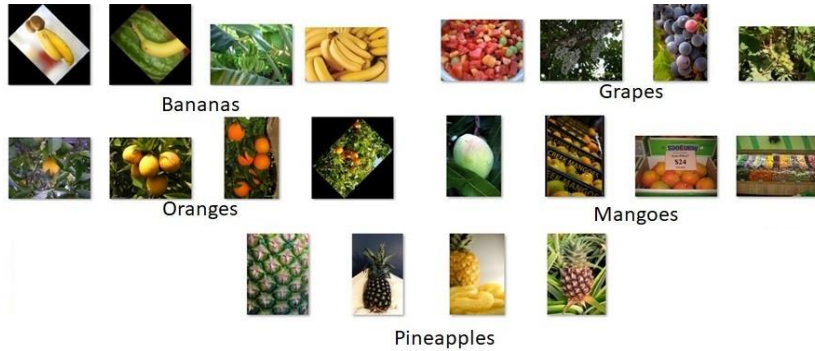


FIGURE 3. Image examples with ground truth for different groups namely banana, orange, grape, mango, and pineapple respectively.

retained because these values are not influencing the scale and orientation of the filters concerned in this paper. We set these parameters to 1.0, π , and 0.3 for sinusoidal wavelength factor, the phase offset, and the Gaussian standard deviation, respectively. The Gabor filter size is set as 3 by 3 to maintain the VGG-16 architecture. We have tried several values to determine the best interval of the scale (σ) and orientation (θ) values. In this work, we found that the uniform distribution of scales in step of 0.05 and 10.0 for orientations are the optimal values and will be used in all experiments. Table 1 shows Gabor filters with different weight initialization values of theta and sigma for each convolutional block using a proposed uniform distribution scheme.

TABLE 1. Gabor filters weight initialization values

Number of convolutional filters	Orientation (θ)	Scale (σ)
64	{10,20,30,...,80}	{0.05,0.10,0.15,...,0.40}
128	{10,20,30,...,160}	{0.05,0.10,0.15,...,0.40}
256	{10,20,30,...,320}	{0.05,0.10,0.15,...,0.40}
512	{10,20,30,...,320}	{0.05,0.10,0.15,...,0.80}

EXPERIMENTS AND RESULTS

EXPERIMENTAL SETUP

In this paper, we used the Google Colaboratory (a.k.a. Colab) for experiments. The Colab is a cloud service based on Jupyter Notebook for disseminating machine learning education and

research. It offers the maximum lifetime of VM is 12 hours, and idle time out after 90 minutes. We used the python language and Keras deep learning framework, which is already installed in the system for creating the VGG-16 model. The model is implemented on an Intel(R) Xeon(R) CPU @ 2.20GHz machine with Tesla K80 GPU of 12GB RAM. After the model is trained on the training and the validation data sets, we evaluated the model using the test data.

DATASET

We have developed a dataset for fruit classification. The dataset contains a collection of 3,750 images taken from the deep-fruit dataset (Sa, et al., 2016) and collected from the Google website. A total of 3000 images are used for training (600 images for each class) and 750 images are used for testing (150 images for each class). These images are all in JPEG with different sizes and were categorized into 5 different classes namely banana, orange, grape, mango, and pineapple. In this dataset, there is only one interest object category for each image. Fig. 3 shows the ground truth for different groups in the dataset.

EVALUATION METHOD

To evaluate the proposed approach, we have used the common evaluation measure, namely accuracy. We used it because it is standardized and will enable us to compare our proposed algorithms with other systems. For evaluating Gabor based VGG-16 architecture, we compute the average accuracy performance on the queries. In this case, we have used a total of 10 runs taken from 10 different train and test images randomly. Each accuracy (ACC) is measured by dividing the number of correct classifications by the total number of test samples. The average accuracy is computed as follows:

$$Average\ Accuracy = \frac{\sum_{n=1}^{10} ACC_n}{10} \quad (4)$$

Table 2 shows the average classification results on the standard VGG-16 and the proposed approach. The results shown that the proposed approach outperforms the standard VGG-16 architecture with accuracy of 79.62%. The results shown that the uniform distribution of scale (sigma) and orientation (theta) parameters can enhance deep feature representations.

TABLE 2. The average classification results on 10 runs

	VGG-16 (%)	Gabor VGG-16 (%)
Average	77.45 ± 7.32	79.62 ± 6.51

CONCLUSION

The Gabor filter bank is useful in extracting good image features. The Gabor VGG-16, which incorporates orientation and scale parameters, shows slightly higher performance on the fruit dataset. It scores the best performance of 79.62% whereas the standard filters score 77.45%. The marginally higher results may be due to the weights initialization of Gabor filters that can reduce the training complexity and improve the classifier performance. Another reason is that the Gabor kernel size is small (3 x 3) in representing the spatial texture information of different scales and orientations, resulting in less distinctive features. In future work, we want to test on

more benchmark datasets, with the different kernel sizes and layers and more runs for significant test.

ACKNOWLEDGEMENT

This work has been supported by the Malaysia's Ministry of Higher Education Fundamental Research Grant FRGS/1/2019/ICT02/UKM/02/8.

REFERENCES

- Abdullah, A., Veltkamp C. R., and Wiering, A. M. 2009. Spatial pyramids and two-layer stacking svm classifiers for image categorization: A comparative study. *International Joint Conference on Neural Networks*. 5–12.
- Alekseev, S. A., and Bobe., 2019. A. Gabornet: Gabor filters with learnable parameters in deep convolutional neural network. *International Conference on Engineering and Telecommunication (EnT)*. 1–4.
- Ayad, H., Abdullah, S. H. N. S, and Abdullah, A. 2019. Visual object categorization based on gabor filter generalization via k-means clustering. *International Journal of Engineering and Technology*. vol. 8.
- Chan, H. T., Jia, K., Gao, S., Lu, J., Zeng, Z., and Ma, Y. 2015. Pcanet: A simple deep learning baseline for image classification?. *IEEE Transactions on Image Processing*. vol. 24. 5017–5032.
- Coates, A., Ng, A. Y., and Lee, H. 2011. An analysis of single-layer networks in unsupervised feature learning. In *AISTATS*.
- Daniel G., 2018. Hyper-parameters in Action! Part II — Weight Initializers <https://towardsdatascience.com/hyper-parameters-in-action-part-ii-weight-initializers-35ae1a28404>. June 19.
- Deng, J., Dong, W., Socher, R., Li, J. L, Li, K., and Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition*. 248–255.
- Dickerson, L. N., 2017. Refining bounding-box regression for object localization.
- Eqlouss, A. A. E., Abdullah, A., and Abdullah, S. H. N. S, 2012. Ensembles of gradient based descriptors with derivative filters for visual object categorization. in *RiTA*.
- Gabor, D., 1946. Theory of communication. *Journal of the Institution of Electrical Engineers - Part I: General*. vol. 94. 58–58.
- Glorot, X., and Bengio, Y. 2010. Understanding the difficulty of training deep feedforward neural networks. in *AISTATS*.
- Goodfellow, G. I., Bengio, Y., and Courville, C. A. 2015. Deep learning. *Nature*. vol. 521. 436–444.
- Gulli, A., and Pal, S., 2017. Deep learning with Keras. Packt Publishing Ltd.
- He, K., Zhang, X., Ren, S., and Sun, J. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *IEEE International Conference on Computer Vision (ICCV)*. 1026–1034.
- Huang, L.L., Shimizu, A., and Kobatake, H. 2004. Classification-based face detection using gabor filter features. *Sixth IEEE International Conference on Automatic Face and Gesture Recognition. Proceedings*. 397–402.
- Jarrett, K., Kavukcuoglu, K., Ranzato, M., and LeCun, Y. 2009. What is the best multi-stage architecture for object recognition? *IEEE 12th International Conference on Computer Vision*. 2146–2153.
- Krizhevsky, A., Sutskever, I., and Hinton, E. G. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*.
- Krizhevsky, A., Nair, V., and Hinton, G. 2009. Learning multiple layers of features from tiny images. Technical Report Available: <http://www.cs.toronto.edu/~kriz/cifar.html>
- LeCun, Y., and Cortes, C., 2010. MNIST handwritten digit database.

- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P., 1988. Gradient-based learning applied to document recognition. in *Proceedings of the IEEE*. 2278–2324.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, E, S., Fu, Y. C., and Berg, C. A., 2016. Ssd: Single shot multibox detector. in *ECCV*.
- Long, J., Shelhamer, E., and Darrell, T., 2015. Fully convolutional networks for semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 3431–3440.
- Luan, S., Chen, C., Zhang, B., Han, J., and Liu, J. 2018. Gabor convolutional networks. *IEEE Winter Conference on Applications of Computer Vision (WACV)*. 1254–1262.
- Pham, L., 2019. Gabor filter initialization and parameterization strategies in convolutional neural networks. *Electronic Thesis and Dissertation Repository*.
- Pinto, N., Stone, Z., Zickler, E. T., and Cox, D., 2011. Scaling up biologically-inspired computer vision: A case study in unconstrained face recognition on facebook. *CVPR 2011 WORKSHOPS*. 35–42.
- Redmon, J., and Farhadi, A., 2018. Yolov3: An incremental improvement. *ArXiv*, vol. abs/1804.02767.
- Ren, S., He, K., Girshick, B. R., and Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. vol. 39. 1137–1149.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, C. A., and Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*. vol. 115. no. 3. 211–252.
- Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., and McCool, C., 2016. Deepfruits: A fruit detection system using deep neural networks. *Sensors (Basel, Switzerland)*, vol. 16.
- Sarwar, S. S., Panda, P., and Roy, K. 2017. Gabor filter assisted energy efficient fast learning convolutional neural networks. *IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*. 1–6.
- Schuldt, C., Laptev, I., and Caputo, B., 2004. Recognizing human actions: a local svm approach. *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*. vol. 3. 32–36.
- Simonyan, K., and Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. *CoRR*. vol. abs/1409.1556.
- Sultana, F., Sufian, A., and Dutta, P., 2018. Advancements in image classification using convolutional neural network. *Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*. 122–129.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., 2015. Going deeper with convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1–9.
- Ventura, D., and Warnick, S., 2007. A theoretical foundation for inductive transfer.
- Yao, H., Chuyi, L., Dan, H., and Wei-yu, Y., 2016. Gabor feature based convolutional neural network for object recognition in natural scene. *3rd International Conference on Information Science and Control Engineering (ICISCE)*. 386–390.

Azizi Abdullah

Wong En Ting

Fakulti Teknologi & Sains Maklumat,

Universiti Kebangsaan Malaysia.

azizia@ukm.edu.my, a160554@siswa.ukm.edu.my