Understanding how noise affects the accuracy of CNN image classification

Yixian Lau
School of Computing
Asia Pacific University of Technology and Innovation (APU)
Kuala Lumpur, Malaysia
ansonlau1325@gmail.com

Weisheng Sim
School of Computing
Asia Pacific University of Technology and Innovation (APU)
Kuala Lumpur, Malaysia
simws3003@gmail.com

Kiansing Chew
School of Computing
Asia Pacific University of Technology and Innovation (APU)
Kuala Lumpur, Malaysia
jeremychew1022@gmail.com

Yongxiang Ng
School of Computing
Asia Pacific University of Technology and Innovation (APU)
Kuala Lumpur, Malaysia
yongxiangng04@gmail.com

Zailan Arabee Abdul Salam
School of Computing
Asia Pacific University of Technology and Innovation (APU)
Kuala Lumpur, Malaysia
zailan@apu.edu.my

***Abstract*— Image classification by using Convolutional Neural Network (CNN) has been a popular method in image processing. This is because this model produces high accuracy of output with efficient amount of work. There are a lot of factors that could affect the performance of CNN in image processing such as image resolution, size of data, quality of image, kernel size and so on. The purpose of this research is to test the accuracy of CNN in training and testing with and without the presence of noise augmentation.**

***Keywords*— Noise, convolutional neural network, accuracy**

I. INTRODUCTION

Image Classification is a supervised learning. The image classification will need a large set of data and separate into two different set images, one for the training and the others is for testing. The training set images will used for the machine learning to learn and recognize the characteristic of the labelled object and the characteristic will be used to classify the images. For an example, the images in the training set are dog and cat, the machine will learn capture the properties and special features of the animal. After the learning process of the machines, the test set images will be applied for the machines to classify. The images in the testing set are not labelled and the machines will need to use the characteristic that he learns from the training set to classify the testing set images.

The image classification is widespread in our life. For instance, the autonomous vehicle, the search engines and so on. The most popular neural network to do image classification is the Convolutional Neural Network (CNN). For each input images, the CNN will pass it through the convolutional layer with filters, Pooling, fully connected layer (FC) and applying the SoftMax function to classify the object using the probabilities. The convolutional layer is the first layer that using a small squares of input data such as image matrix and filter or kernel to maintain the relationship of the pixels. In the pooling layer, the parameters will be reduced when the images are too large. There are three type of pooling which are max pooling, average pooling and sum pooling. The last layer is fully connected layer, FC layer is going to convert the matrix into the vector and combine them into a model. At the last, applying the SoftMax of others activation function to classify the output.[5]

There are several factors to affect the performance of the image classification such as different location of the training images and testing images. The degradation image like underwater images, motion-blurred images, fish-eye images and more will also be factors that affecting the accuracy of the image classification. Not only that the noise of the images will influence the performance of the image classification.

Since the born of camera in the world, some of the images that captured at the night will have noise. There are various of ways to lower the noise in the images. For example, the aperture of the camera is to let more light to entering the camera, if the aperture is smaller, the lesser noise will be. The (CCD/CMOS) sensor is the sensor of accepting the light and transform the light into electrons. Usually, the CCD sensor is better than the CMOS sensor that create higher quality and lower noise images. The third way is the photographer can tune the camera photography parameter such as longer shutter speed, higher ISO value, and more. The A.I algorithm also can improve the images by reducing the noise. However, no matter how small the aperture is, which sensor are using, how the photographer tuning the photography parameter or how A.I algorithm reduce the noise, the images that captured at the night will still have a lot of noise.

Other than night captured photos, quality of the images also one of the factors in causing noise in the images. In most cases, the videos or images that captured by the surveillance camera is filled with huge amount of gaussian noise as they usually equipped with lower end camera sensor which leads to failure in obtaining great exposure for the captured photos and videos. Therefore, noise should never be neglected as a research topic as there are still lots of noisy images surround us.

From this, we are going to investigate how is the noise of the images will affect the performance of the image classification. We are assuming that the noise will affect the accuracy of the image classification to decreases. Due to the images on the website cannot have both clear and noise images, we will use the clear images dataset and apply varying degree of noise on the images for the image classification to train and test on it.

In this research investigation, we are going to use the python as our programming language, and we will run it on the Kaggle platform with the GPU accelerator in order to run the experiment faster and more stable.

II. RELATED WORKS

In the actual cases, each of the images input will differ to each other. For example, sometimes the dataset will be in large scale and even contain in many categories. Yet sometimes do not. The quality of image might also get affected together with the image resolution. In addition, there are also images that are taken based on the situation. For instance, taking a picture of a moving train will degrade the quality of the image hence making it more challenging to be identified.

When dealing with large datasets and categories, Luo et al., [1] present with his ResNet-32 model to experiment with dealing the increasing size of datasets and dataset categories. There are 2 ways of experimenting which is increasing the number of categories for each dataset while other is increasing the number of datasets in a category. From the experiment, when the categories of dataset increase, the accuracy will decrease. Not only that, when increasing the number of datasets in a category, the accuracy also increases.

In testing how image resolution affects the performance of Convolutional Neural Network, Kannojia and Jaiswal [2] uses MNIST and CIFAR-10 in their experiment. They tend to rescale the images and use the original image to train the model and test on altered resolution image and vice versa. It is found that the accuracy of both methods has dropped. Besides that, training with altered image resolution tend to prevent from losing more accuracy.

Dodge and Karam [3] use Caffe, GoogleNet, VGG-CNN-S and VGG-16 as their model to experiment with their dataset. They experiment by altering the image quality which means changing normal clear images by adding many kinds of image distortions. This includes blurring the image, adding noise, using JFEG and JPEG2000 image compression. From these models, due to the obvious changing of image quality by the blur and noise, it causes all these models' accuracy to drop.

There are also researchers like Pei et al. [4] using degraded images to test for CNN image classification. They choose the degraded image categories like underwater images, fish-eye images, foggy images and motion-blurred images. The images will be tested for the accuracy on the CNN image classification. Besides that, they produced more and different degradation of images and conduct a quantitative evaluation. The degraded images will be tested by existing degradation removal algorithm. From the result, the accuracy decreases when the level between the degradation gap is too high. Besides that, the degradation removal algorithm is not able to do much enhancement as it only produces visual -pleasant images instead but not increasing the feature classification.

III. THE SIMPLE CNN MODEL FOR IMAGE CLASSIFICATION

In this research study, we adopted the Keras Simple CNN Model for Image Classification and modify parameters of specific layer to accustomed and stabilized results from additional image filters that we had applied on training and testing sets. If these modification and scaling does not apply or tuned, adopting the original parameters of layer would result in false accuracy and occurrence of overfitting within simulation results. In order to prevent overfitting, our research team had made efforts on researching the original layers available and the ones that would help stabilize and handle the additional filters that would be added to the image dataset.

In the following section, simple description would provide for better understanding on the layers of the Keras Convolutional Neural Network that had been adopted by previous research. The base architecture of the model is consisting of 7 main layers (19 specific layers in total) as shown in the Table 1. Each layer of the model is based on the application programming interface – Keras which provides us the basis needed for a convolution neural network construction on a Python programming language and run on Kaggle platform with GPU accelerator provided.

TABLE I. THE LAYERS OF KERAS CNN MODEL

Model Layers Type	Layers
Image Input Layer	1
Convolutional (2 Dimension) Layer	3
Batch Normalization Layer	4
Max Pooling Layer	3
Dropout Layer	4
Fully Connected Layer (Dense)	1
Flatten Layer Dense Layer	1 2
Output Layer	1

First and foremost, an image input layer is needed to input image data taken from the dataset into the first layer of the network. Following on, two-dimension Convolutional Layer are available with different numbers of filter and same kernel size are set upon the 3 levels of layers. The purpose of filtering each image with different size filters was to allow the network to extract specific textures and features of an image. Each values of the filters are processed and assigned by the network and with each different sizes of filter will allow the network to capture more than one characteristic and properties of each image given. A Max Pooling Layer will be followed on after Convolutional Layer, the purpose was to summarize and reduce the size of parameters values processed in the Convolutional Layer into size of 2 x 2.

Due to having multiple layers of Convolutional Layers and constant reassigning values within the filters during the progress, there would be value variability such as generalization error and shifts. This would be where Batch Normalization comes in to normalize and standardized the values into mini batches. Dropout Layer another vital role to decrease the risk of overfitting the network. It randomly removes values based on the dropout rate parameter settings. The optimize range for CNN are from 0.5 to 0.8 as probability of dropping out values. The lower the rate, the more values are being remove from the nodes of the network.

Finally, a fully connected layers are needed to finalize the classification of the images with the help of Flatten and Dense Layers. Flatten Layer concludes the two-dimension into one-dimension array like as input for the fully connected layers. The first Dense Layer will be assigned with ReLU activation function which is also implemented in all convolutional layers, whereas the second Dense Layer is using Softmax activation function.

Parameters on Dropout rate are altered to avoid overfitting the network after applying noise augmentation on the images. The optimal results achieved for dropout rate are 0.5 for all level of layers. The following Table II shows the architecture of the network's model.

IV. EXPERIMENT

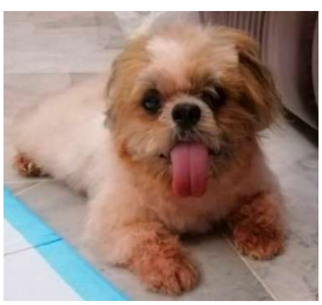
A) Additional Noise Filter on Image

Additional Noise are added with the data augmentation as our research topic. Each different level deviation of noise would be applied to the image for both or either training or testing sets. Table III below shows a sample illustration on the levels of noise applied to an image:

TABLE II. ARCHITECTURE OF CNN MODEL

Layers	Parameters	Activation
Conv2D	Filter Size: 32, Kernel: (3 x 3)	ReLU
BatchNormalization	-	
MaxPooling	Pool Size: (2 x 2)	
Dropout	DropOut Rate: 0.5	
Conv2D	Filter Size: 64, Kernel: (3 x 3)	ReLU
BatchNormalization	-	
MaxPooling	Pool Size: (2 x 2)	
Dropout	DropOut Rate: 0.5	
Conv2D	Filter Size: 128, Kernel: (3 x 3)	ReLU
BatchNormalization	-	
MaxPooling	Pool Size: (2 x 2)	
Dropout	DropOut Rate: 0.5	
Flatten	-	
Dense	Units: 512	ReLU
BatchNormalization	-	
Dropout	DropOut Rate: 0.5	
Dense	Units: 2	Softmax

TABLE III. NOISE APPLIED IMAGE

Noise Deviation Level		
0	10	20
		
30	40	50
		

B) Experiment Result

There are two methods of conducting the experiment. First, we train the CNN model with noise images and we conduct the model testing set with noise images. The alternative method is training the model with noise images and test the model with original images. In these 2 experiments, we will increase the level of noise from 10 to 50. Each of the levels will be conducted twice to achieve consistency and their average would be recorded as well.

After running twice of each level of noise, we will calculate the prediction result by taking the correct classification prediction and divided the numbers of tested images in percentage. The result of each experiment is being grouped and recorded as shown in Table IV and Table V. Val_Acc represents the validation accuracy of the CNN model when testing the classification of cat and dog images. Meanwhile, Acc represents the accuracy when training the model.

From Fig 1., we can know that the average of validation accuracy and the average of training accuracy decreases each time the noise level being increased. Thus, we can know that the when dataset is trained with higher noises, the CNN model still unable to correctly classify the difference between the cat and dog image even it is tested with noise dataset. This is due to the reason that the image was too blur for the model to correctly extract the feature to categorize it.

From Fig 2., the average of validation accuracy and the average of training accuracy both decreases when the level of noise increases each time. Which means that noise will affect the classification of the cat and dog image even it is tested with original images. The reason might be the kernel of CNN is unable to sense the feature that contains in the original image. Which means that the model is more used to the noised image feature each time the level increases. Therefore, both accuracy in the experiment decreases.

Table IV: Summary Results of Training and Testing Model with Noise Dataset

	Image Classification 1st Run Attempt			Image Classification 2nd Run Attempt			Average Statistic Result		
Noise Level	Prediction Result (%)	Val. Acc.	Acc.	Prediction Result (%)	Val. Acc.	Acc.	Average Prediction Results (%)	Average Val. Acc.	Average Acc.
10	83.3333	0.8475	0.8260	61.1111	0.8395	0.8280	72.2222	0.8435	0.8270
20	61.1111	0.8323	0.8230	55.5556	0.7799	0.8110	58.3333	0.8061	0.8170
30	27.7778	0.7494	0.7990	33.3333	0.8175	0.8150	30.5556	0.7835	0.8070
40	38.8889	0.6425	0.7840	50.0000	0.6195	0.7770	44.4444	0.6310	0.7805
50	55.5556	0.5448	0.7650	50.0000	0.5501	0.7610	52.7778	0.5475	0.7630

Table V: Summary Results of Only Training Model with Noise Dataset

	Image Classification 1st Run Attempt			Image Classification 2nd Run Attempt			Average Statistic Result		
Noise Level	Prediction Result (%)	Val. Acc.	Acc.	Prediction Result (%)	Val. Acc.	Acc.	Average Prediction Results (%)	Average Val. Acc.	Average Acc.
10	44.4444	0.8784	0.8424	77.7778	0.8475	0.8216	61.1111	0.8630	0.8320
20	44.4444	0.8550	0.8319	38.8889	0.8269	0.8216	41.6667	0.8410	0.8268
30	50.0000	0.8415	0.8230	38.8888	0.7823	0.8110	44.4444	0.8119	0.8170
40	77.7778	0.6132	0.7804	50.0000	0.6142	0.7819	63.8889	0.6137	0.7812
50	88.8889	0.5872	0.7690	66.6667	0.5765	0.7702	77.7778	0.5819	0.7696

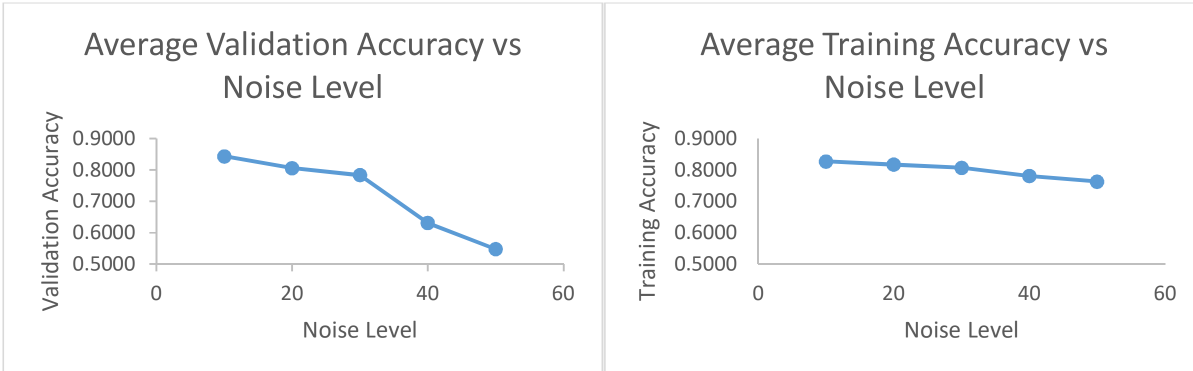


Fig. 1. The average of validation accuracy and training accuracy of different level of noise in training and testing model with noise image.

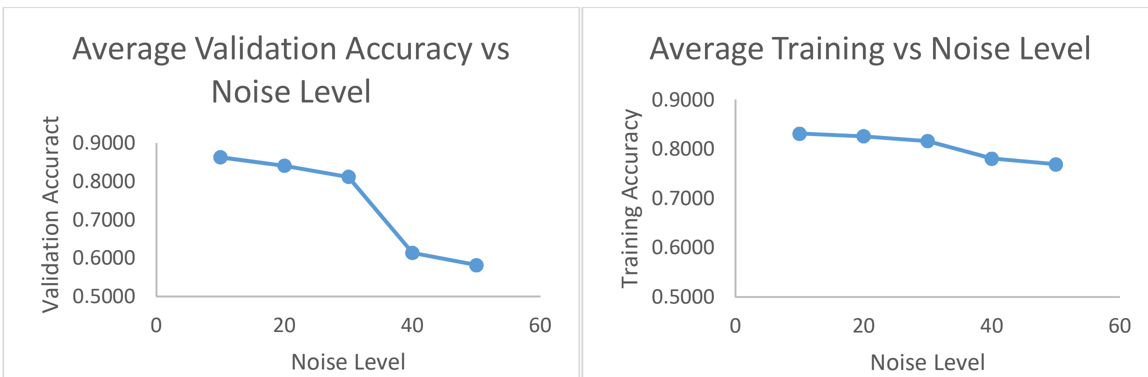


Fig. 2. The average of validation accuracy and training accuracy of different level of noise in training model with noise image only

V. CONCLUSION

We are considering the validation accuracy and the training accuracy as the measurement of our experiment. Each level of noise is run twice for both experiments. The result for each level of noise is calculated by calculating the average score for both tests in both experiments. The reason of doing so is to reduce the possible bias from the result.

From the result, we can conclude that the accuracy of CNN image classification is highly affected by the level of noise. When the level of noise is increased, the accuracy will decrease. It is also significant to notice that the accuracy of CNN image classifier tends to have largest drop in both experiments when the noise varies from standard deviation of 30 to 40. Although the images are still highly recognizable via human eyes, but it seems to have a huge downgrade on the machine capability for recognizing the images correctly.

When comparing the accuracy for both experiments when the noise level is increasing, we can notice that the accuracy degradation is slightly lower when the classifier is train and test with both same level of noisy images compared to train with noisy image and test with original images. By associating this result with the previous research result done by Dodge and Karam[3], it does proof that the accuracy is indeed affected by the level of difference of noise between training dataset and testing dataset. Therefore, it is recommended to train the image classifier with a higher noise images when it is planned to use for classifying noisy images. However, doing so will lead to having a poor accuracy in identifying high quality images. Since our research are mainly focus on a single type of CNN architecture, it is highly recommending the future studies to replicate and focus their research by using various CNN architectures and different number of categories for better generalization purpose.

REFERENCES

[1] C. Luo, X. Li, J.Yin, J. He, Denggao and J. Zhou, Journal of Software, “How Does the Data Set and the Number of Categories Affect CNN-Based Image Classification Performance,” 2019, pp. 168-181.

[2] S.P.Kannojia and G. Jaiswal, International Journal of Computer Sciences and Engineering, “Effects of Varying Resolution on Performance of CNN based Image Classification An Experimental Study,”2018, vol.6, pp.451-456.

[3] S. Dodge and L. Karam, “ Understanding How Image Quality Affects Deep Neural Networks,”2016, pp.1-6.

[4] Y. Pei, Y. Huang, Q. Zou, H. Zang, X. Zhang and S. Wang, ArXiv, “Effects of Image Degradations to CNN-based Image Classification,” 2018, pp. 2-10.

[5] Prabhu, “Understanding of Convolutional Neural Network (CNN) — Deep Learning,” 2018, Available: https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148.