# Super Resolution Based Augmentation for Image Classification on Small Data Set

**Thitiporn khantong[1] and Parinya Sanguansat[2]**

[1,2]Engineering and Technology, Panyapiwat Institute of Management,
Nonthaburi, Thailand
E-mail: thitiporn.khantong@gmail.com, parinyasan@pim.ac.th

Received: August 7, 2020 / Revised: October 9, 2020 / Accepted: October 9, 2020

*Abstract*—**For image classification to be more accurate, image files with high resolution were to be put up for test. Practically, the set of images obtained for classification might be low resolution files. In this case, it was more difficult to do image classification. The result was that the accuracy rate was low. It was unable to be applied for work. We studied to find a solution to fix this problem. The answer to the problem was to do data augmentation which was consisted of Affine Transformation and Super Resolution. Affine Transformation are all linear transformations which can be represented by a matrix and combined into a single overall including rotation, scaling, translation and shearing where all points in an object are transformed in the same way.**

**Image Super Resolution made to increase the high resolution and quality for image. This method has been shown that it outperformed the basic interpolation method. In this research, we studied whether Super Resolution is able to enhance the analyze model of the image and then transformed image is more responsible to image classification process. It can be seen that the methods applied were useful to accurately and precisely identify low resolution image files. Besides, we adopted a transfer learning method which was designed and developed for the task. It has been reused the first starting point form model in the second task could help us classify the type of image more accurately and precisely.**

*Index Terms*—**Image Classification, Data Augmentation, Transfer Learning, Image Super Resolution**

## I. Introduction

Up to this day, Computer Vision is well recognized by common people since it works to process images so that the computer would understand the image or be able to distinguish objects as close as to what a human brain does. There are 4 types of Computer Vision including image classification, object detection, semantic segmentation and instance segmentation. Each type has its own function. The one matched to our experiment is Image Classification. For the past few years, doing image classification has become an interesting subject - its objective was to classify the type or character of image precisely with minimum error. The experiment was separated into 2 sections. One was to teach the computer to test the prepared testing data for classification before inputting the actual set of data for classification. It found that we could well classify the dataset with the big size or a lot number of images in the testing set and actual set. What we have found to be an obstacle to this experiment was that we could not know if how many actual images were obtained. In certain case, there might be the situation that the actual data obtained was consisted of small-sized image or less number of images. After conducting classification, the outcome was unable to be adopted to actual work because of many factors. Therefore, we suggested to use Data Augmentation and Image Super Resolution were applied for image classification in degree that to increase quality for images to its highest. This would also help us to do classification easier and increase the level of accuracy in image classification. The technique of image augmentation or affine transform was applied for minimizing the loss and diversity of data because the more testing data it obtained, the more accurate it would be. The basic techniques were consisted of Flip, Mirroring, Scale Image, Translation, Crop and Rotation. Meanwhile, using Image Super Resolution to enhance resolution and quality of image has been found to be more eligible and efficiently for classification according to the research. After the dataset was managed, Transfer Learning was additionally applied to do classification so that the dataset could be classified faster. Besides, Convolutional Neural Network was also used for classification.

## II. Objective

- To categorize images from a small dataset into image types and image sizes.
- To propose a solution for small dataset classification.
- To develop a Data Augmentation Technique to be more efficiently and accurate.
- To use the limited dataset for image classification even if normally it requires a big dataset to do so.
- To increase the quantity for images and enhance quality for small images files so that it can be used flawlessly.

## III. Literature Review

Before we conducted this research, we had been researching similar work with the corresponding objective including problems. The subject we studied more was related literature with the details as follows: For image classification, mort of the research concerned about detection and object categorization segmentation. These techniques were basic techniques adopted to the research. For data augmentation, we included many forms such as image super resolution, using of GAN and Affine Transform in both basic form and applied form for 3D transformation which transform image from Transform Version to Original Version. The technique of Affine Transform jointly applied with other Classification technique.

## IV. Related Work

To do image classification, the model's resolution must be high enough so that it can be classified easier. It also prevents overfitting. When model trained with small images training dataset, the trained model tends to overly fit to the samples in training set. Moreover, there was another problem that images in a small training set were usually incomplete and low quality it could not be classified. It also worsened the results in generalization. However, The Data Augmentation have been widely to used avoid this overfitting problem by enlarging the size of images training dataset which increased the number of date by Crop, Translation, Scale Image, Mirroring, Flop and Rotation. Even though the image could be adjusted and improved, the existing hyper parameter did not help improve the image resolution as the date was small and it was possibly that the image with low quality would be mixed up in the same set. Therefore, we adopted the technique of Image super resolution to help fixing this problem. Image super-resolution methods attempted to recover a high-resolution image from one or more low-resolution input images. This disadvantage for image super resolution in some cases was that the image would be distorted from the original. Eventually, there would be a mistake in classification process.

## V. Techniques

### A. *Image Super Resolution*

Image super-resolution methods attempt to recover a high-resolution image from one or more low-resolution input images. Super-resolution methods can be classified in two main families: Multi-image Super Resolution methods, and Single Image Super Resolution methods (SISR). Multi-image Super Resolution methods attempt to use several low-resolution images of the same scene to determine new details in the high-resolution image, In addition, these methods tend to be limited to small increases in resolution [1], [2].

### B. *Data Augmentation*

Data augmentation is a strategy that enables practitioners to significantly increase the diversity of data available for training models, without actually collecting new data. Data augmentation techniques such as cropping, padding, and horizontal flipping are commonly used to train large neural networks [3].

### C. *Transfer Learning*

Transfer learning is a machine learning method where a model develop for a task is reused as the starting point for a model on a second task [4].

### D. *Computer vision*

Computer vision is the field of computer science that focuses on replicating parts of the complexity of the human vision system and enabling computers to identify and process objects in images and videos in the same way that humans do. Until recently, computer vision only worked in limited capacity [5].

### E. *Convolutional Neural Network*

The structure of artificial nerve has consisted of an input layer, a hidden layer, and an output layer, or commonly called multi-layer perceptron (MLP). However, the research presented by Le Cun et al. and has added convolutional calculation into such neural network – it is called Convolutional Neural Network (CNN). CNN is consisted of a pooling layer and a fully-connected layer which is a hidden layer and an output layer appeared in the artificial neural nets. CNN is able to extract special characteristics of photo and second-class category. The outstanding point of CNN is different from machine learning which categorize only type of information or organize them into groups [6], [7], [8].

- VGGNet.

VGGNet is a deep CNN structure since there are 16 layers by using Kernel of 3x3 in convolution. VGGNet-16 is consisted of 5 groups of convolution. The first Conv or Conv1 is consisted of Conv1{1,2} and there are 64 Kernel in each layer. Conv2 is consisted of Conv2{1,2} which has 128 Kernels. Conv3 is consisted of Conv3{1,2,3} which has 256 Kernels. Conv4 is consisted of Conv4{1,2,3} which has 512 Kernels. Conv5 is consisted of Conv5{1,2,3} which has 512 Kernels. Conv1-5 is followed by max pooling. Then the information is linked to fully-connected layer with 4,096 nodes. In the final layer, it is designed to have 1,000 nodes. The outcome is calculated by a software [9].

## VI. Methodology

The goal of the project was to determine whether Image Super-resolution and Data Augmentation can be used as a pre-processing step to improve classification accuracy. Experiments were carried out by using general image classification datasets.

### *A. Image Super – Resolution*

Image super-resolution methods attempt to recover a high-resolution image from one or more low-resolution input images. Super Resolution methods attempt to use several low-resolution images of the same scene to determine new details in the high-resolution image, In addition, these methods tend to be limited to small increases in resolution [1], [2].

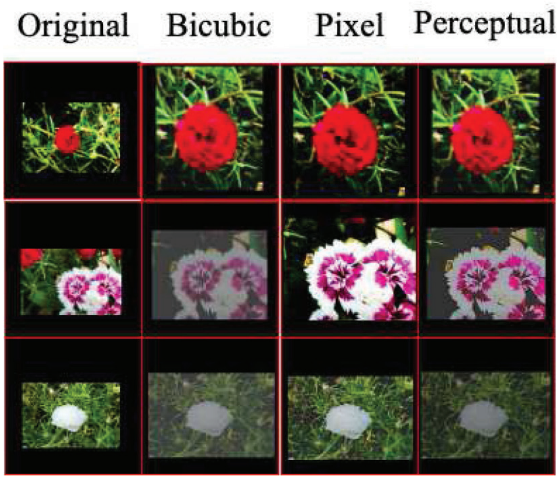


Fig. 1. Images are shown before versions transformed with up sampling method.

### *B. Convolutional neural networks for Image Super-Resolution*

First way to know if SISR is using deep learning can be traced back to [10], where three layers of CNN named SRCNN is made to solve the problem in a very effective way. This simple but light structure have three portion as: (1) Patch extraction layer (2) Non-linear mapping (3) Reconstruction they directly consider a convolutional neural network which is an end-to-end mapping function between photos with different resolution. The reason why this is a surprising invention is that they are made to match the easy in mind, and yet give great accurate and quickness even compared the model example-based method.

*1) Patch extrtaction and representation*

This operation extracts (overlapping) patches from the low-resolution image Y and represents each patch as a high-dimensional vector. These vectors comprise a set of feature maps, of which the number equals to the dimensionality of the vectors.

This a operation extract patches form low-resolution images and shown each patch a high-dimension:

$$F_1(Y) = \max(0, W_1 * Y + B_1)$$

W1 and B1 represent the filters and biases respectively, and '*' denotes the convolution operation.

*2) Non-linear mapping*

This operation non-linearly maps each high-dimensional vector onto another high-dimensional vector. Each mapped vector is conceptually the representation of a high-resolution patch. These vectors comprise another set of feature maps.

This operational nonlinearly maps each high-dimensional vector onto another high-dimensional vector. Each mapped vector is conceptually the representation of a high-resolution patch:

$$F_2(Y) = \max(0, W_2 * F_1(Y) + B_2)$$

W2 contains n2 filters of size n1 ×f2 ×f2, and B2 is n2-dimensional. Each of the output n2-dimension vectors is conceptually a shown of a high-resolution patch that will be used for reconstruction.

*3) Reconstruction*

This operation sums up the above high-resolution patch-wise shown create the last high-resolution image. The image is expected to have a similar to ground X axis.

In the last methods, the predicted overlapping high-resolution patches are often averaged to produce the final high-resolution image :

$$F(Y) = W_3 * F_2(Y) + B_3$$

W3 corresponds to c filters of a size n2 × f3 × f3, and B3 is a c-dimensional vector. If the representations of the high-resolution.

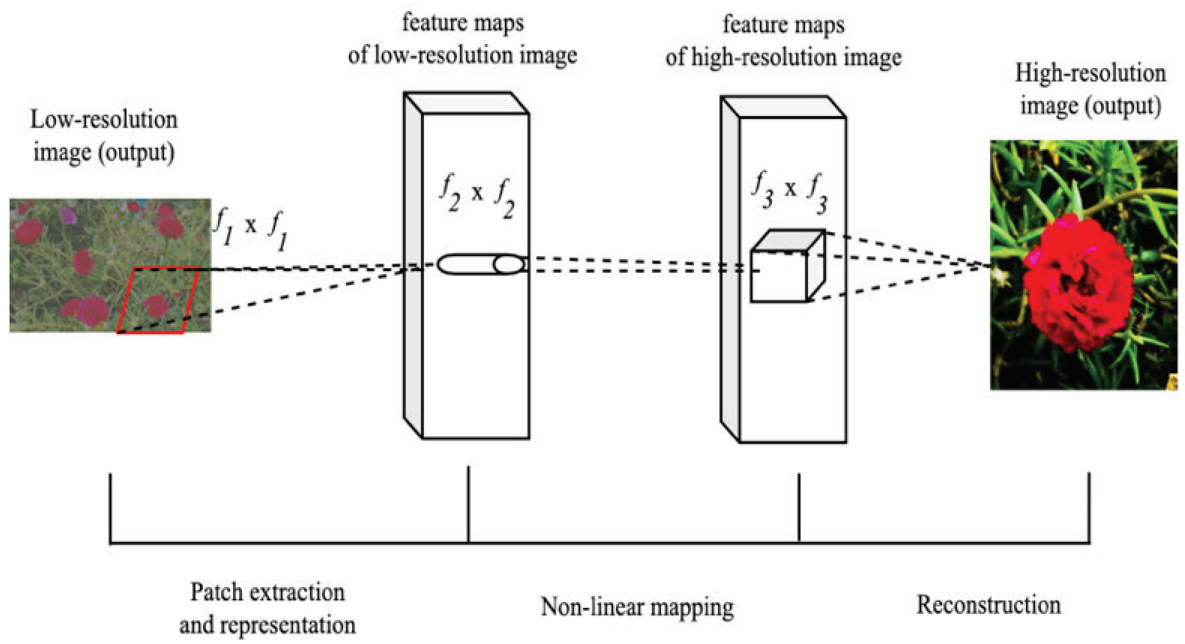


Fig. 2. CNN for Image super-resolution classification

*C. Affine Image transformation*

In the first method, we are going to define and apply a set of geometrical transformations called affine transform. We will present basic affine transform translation, rotation, scale, shearing.

*1) Translation*

A Translation is a function move every points in a fixed distance in Fig. 3 (a)

$$T = [T_x, T_y]$$

$T_x$ Distance from the horizontal axis

$T_y$ Distance from the horizontal axis

*2) Scaling*

A Scaling is a linear transformation that set of points up or down in the same all directions. Zoom in on the x and y axis in Fig. 3 (b)

$$x' = x \cdot S_x$$
$$y' = y \cdot S_y$$

$S_x$ = Scale factors form x axis

$S_y$ = Scale factors form y axis

*3) Rotation*

A Rotation is a circular transformation that a set of points about the origin. in Fig. 3 (c)

$$x' = x\cos\emptyset - y\ \sin\emptyset$$
$$y' = y\sin\emptyset - x\ \cos\emptyset$$

*4) Shearing*

A Shearing is a function offsets a set of points a distance proportional to x and y. The twisting of the x axis 2. The twisting of the axis of the image causing the point to move (x) to (y) after that the twisting of the x axis causes the point to be moved (x, y) to (x′, y′) in Fig. 3 (d)

$$x' = x$$
$$y' = Shy.x + y$$
$$x' = Shx.y + x$$
$$y' = y$$

Shx = Shearing factor form x axis

Shy = Shearing factor form y axis


Fig. 3. Affine transform: Translation, Scaling, Rotation, and Shearing

*D. Datasets*

*1) Kaggle flowers*

The dataset have 4242 images flowers. This dataset have portioned five classes (1) chamomile, (2) tulip, (3) rose, (4) sunflower, and (5) dandelion. For each class there have 800 images colour in per 1 class. Images not high resolution and this size 320x240 pixels. Photos are not reduced to single size, that have other proportion. in order to test the effects and affine transform of super-resolution on classification performance [11].

*2) Oxford_flowers 102*

Dataset is a consistent of 102 flower categories commonly occurring in the United Kingdom. Each class consists of between 40 and 258 images. The images have large scale, colour, pose and light variations. In addition, there are categories that have large variations within the category and several very similar categories in order to train and test the effects of super-resolution and affine transform on classification performance [12].

*3) Oxford_flowers 17*

Datasets is a consistent of 17 category flower dataset with 80 images for each class. The flowers chosen are some common flowers in the UK. The images have large scale, colour, pose and light variations and there are also classes with large variations of images in order to train the effects of super-resolution and affine transform on classification performance [13].

*E. Experiments*

We use a standard per-pixel loss function to train a set of up scaling filters. We will refer to this as the pixel loss method. The second approach was a standard bicubic interpolation, mainly to be used as a control. We trained network for 100 epochs, and a learning rate of 1e-3 using an Adam optimizer.

We hypothesize that using a transform factor which scales in addition, scaling, translation, rotation and shearing were applied to the training data for all trials. The network was trained to upscale consist of. We trained and tested a standard VGG-net classifier which is composed of a standard VGG net with a few layers added simply to account for the larger image sizes (128x128 of 96x96 as opposed to 32x32). A 16-layer VGG network was used for classifying the original 32x32 data, and a network with 4 additional layers was built from that to fit the 96x96 up sampled data.

In Fig. 12, it explains about the data in a training set which is divided into 2 sets including: Kaggle Flowers which contains 4,242 photos and Oxford_flowers102 which contains 102 categories.

The dataset includes in a training set can be portioned a validation set and a testing set. Each per them include 10 images for one class, totaling 6, 149 images per one set.

In Fig. 13, it explains about the data in the test set which applied 1 set (Oxford_flower17). It contains 17 categories. In each class, it contains 80 images.

Oxford_flowers102 which contains 102 categories. This testing set include for remaining 1,020 images (minimum 20 in one class).

In experiment, I am going to use examples of VGG 16 pretrained model throughout this guide. We can use other pretrained models similarly.

The main pattern was consisted of CNN which we used VGG NET and Transfer Learning. In the subsection, we divided into 3 patterns including: (1) Classification via Transfer Learning and Convolutional neural network without using Data Augmentation and Image Super-Resolution, (2) doing Classification via Transfer Learning Convolutional neural network using Data Augmentation but not using Image Super-resolution, (3) doing Classification via Transfer Learning Convolution neural network using Image Super-resolution but not using Data Augmentation, and (4) doing Image Classification via CNN, Transfer Learning and using Data Augmentation and Image Super-resolution. From the experiment.

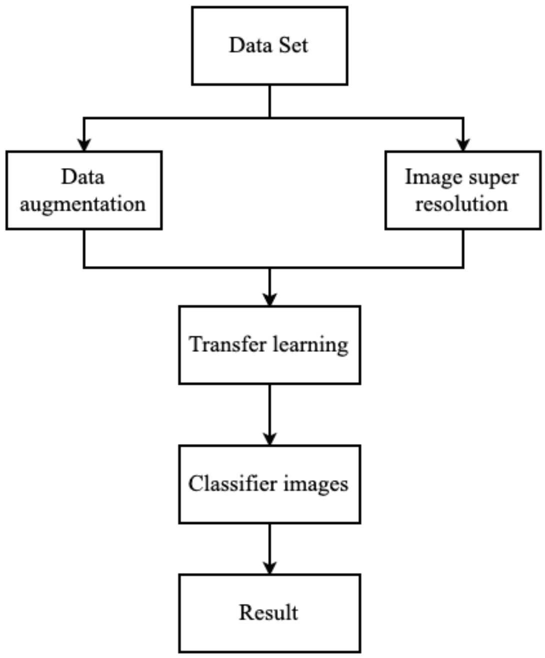


Fig. 7. Overall Methods of Image Classification

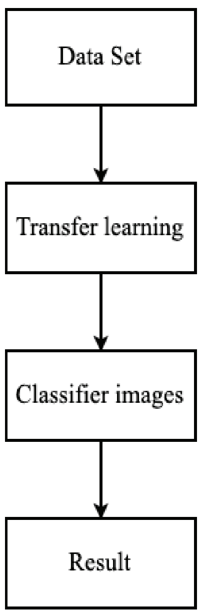


Fig. 8. Methods - 1 of Image Classification

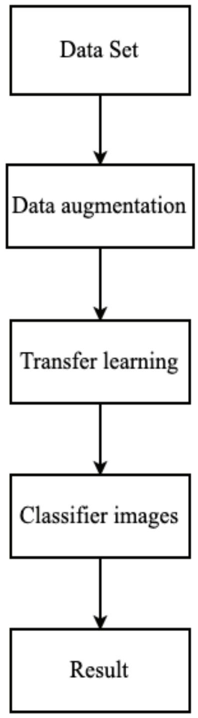


Fig. 9. Methods - 2 of Image Classification

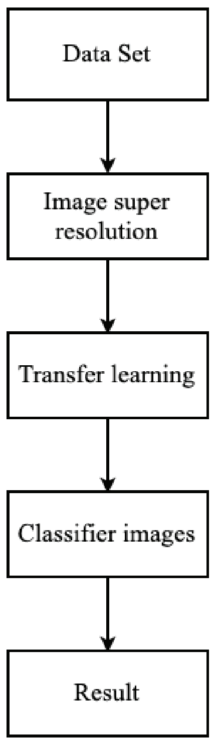


Fig. 10. Methods - 3 of Image Classificat

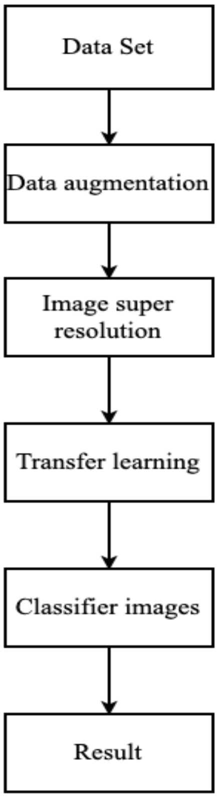


Fig. 11. Methods - 4 of Image Classification

## VII. Result

A variety of image super-resolution, Data augmentation techniques, CNN, transfer learning and image classification. Experiment have proceeded involving general and small Image datasets.

Fig. 12, we compare our results train dataset form train with methods 1- 4 on VGGNeTs for fine-grained classification in basic dataset. Fig. 13, we compare our results test dataset form train with method 1- 4 on VGGNeTs for fine-grained classification in small data set. Evaluation on effect of methods 1- 4 which will see the process on methods 4 have a high accuracy more than methods 1-3 that method have to recovery high resolution details for images. In our experiments, we used Macbook Pro with Google Colab .

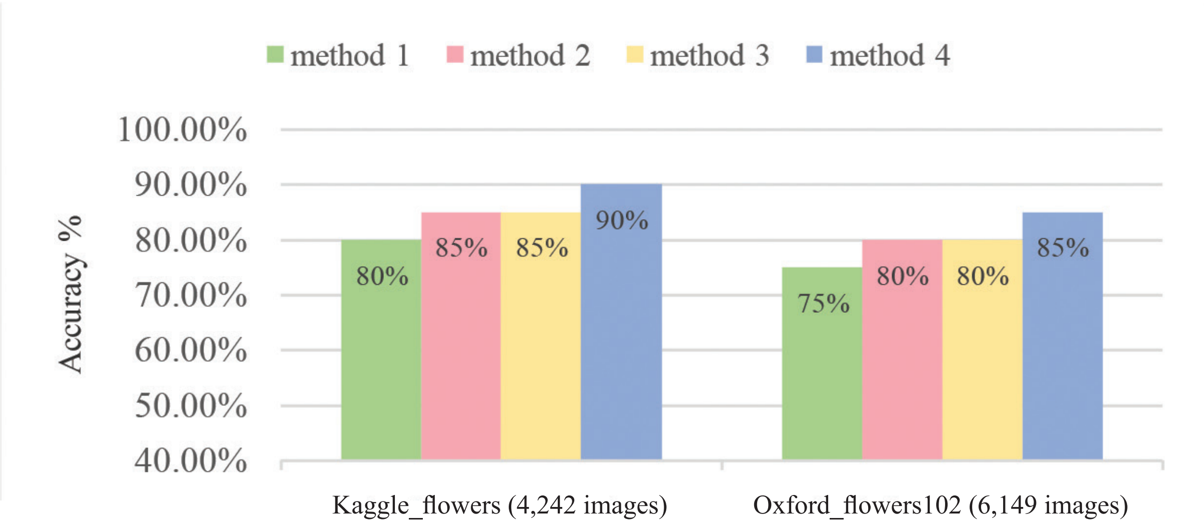


Fig. 12. Accuracy result of train methods compare result basic dataset for train with 1- 4 methods on vggnets

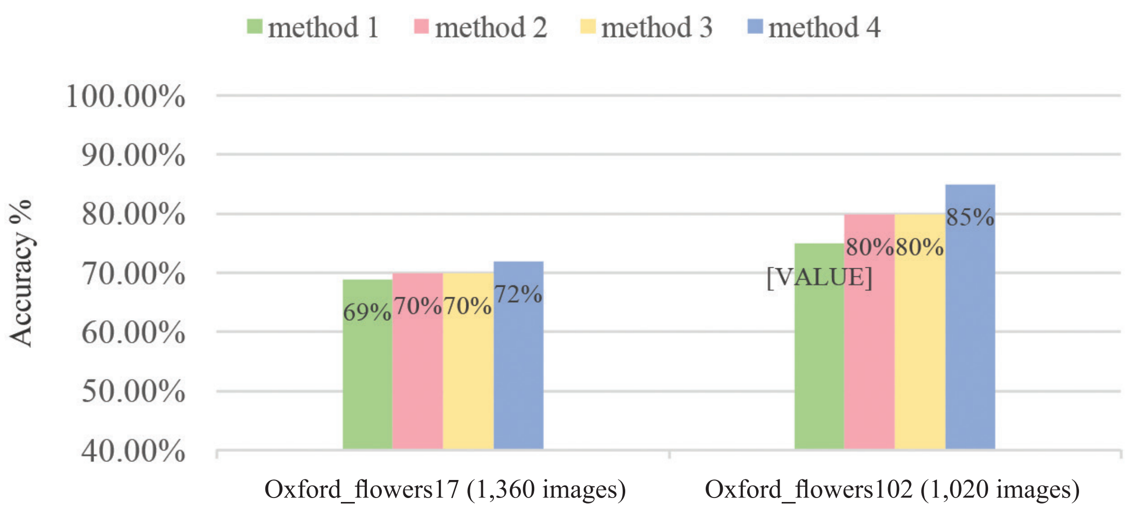


Fig. 13. Accuracy result of test methods compare result basic dataset for test with 1-4 methods on vggnets
Note: The effect of the presented methods depends on dataset

## VIII. Conclusion

We proposed that applying the new and former techniques to create an image enhancing process with the Image Classification gave us a better outcome that was corresponding with our need. However, it was not the best outcome we expected since we had incomplete images. The problem we have found from adopting several techniques together was that in each technique had its own complication. Therefore, the outcome was inaccurate. For example, the image was twisted from the original. In an overall, adopting several techniques together provided a better outcome but it could not fix all problems. In the future, these techniques should be further developed and used as a basic technique for classification with the limited data. Even though it has been through data augmentation process, some images were unusable–the dataset was unable to be adopted for work. In the future, we may develop these techniques to be more efficiently in case that we have limited dataset which is required to be used.

## References

[1] G. Daniel, S. Bagon, and M. Irani, "Super-Resolution from a Single Image," in *Proc. IEEE 12th International Computer Vision,* 2009, pp. 349-356.

[2] S. Borman and R. Stevenson, "Super-Resolution from Image Sequences-a Review," in *Proc. Midwest Symposium on Circuits and Systems*, 1998, pp. 374-378.

[3] C. Dong, C. Loy, K. He et al., "Image Super-Resolution Using Deep Convolutional Networks," in *Proc. IEEE transactions on Pattern Analysis and Machine Intelligence*, 2016, pp. 295-307.

[4] J. Krause, H. Jin, J. Yang et al., "Fine-Grained Recognition without Part Annotations," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5546-5555.

[5] J. Yang, J. Wright, T. S. Huang et al., " Image Super-Resolution Via Sparse Representation." in *Proc. IEEE transactions on image processing*, 2010, pp. 2861-2873.

[6] J. Lemley, S. Bazrafkan, and P. Corcoran, "Smart Augmentation Learning an Optimal Data Augmentation Strategy," *IEEE Access*, vol. 5, pp. 5858-5869, Mar. 2017.

[7] Z. Gao, E. Edirisinghe, and S. Chesnokov, "Image Super-Resolution Using Cnn Optimised by Self-Feature Loss," in *Proc. International Conference on Image Processing, ICIP,* 2019, pp. 2816-2820.

[8] J. Shijie, W. Ping, J. Peiyi et al., "Research on Data Augmentation for Image Classification Based on Convolution Neural Networks," in *Proc. 2017 Chinese Automation Congress (CAC)*, 2017, pp. 4165-4170.

[9] L. Shao, F. Zhu, and X. Li, "Transfer learning for visual categorization: A survey," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 26, no. 5, pp. 1019-1034, Jul. 2015.

[10] M. Paulin, J. Revaud, Z. Harchaoui et al., "Transformation Pursuit for Image Classification," in *Proc. 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3646-3653.

[11] K. Parthy. C S231n Convolutional Neural Networks for Visual Recognition. (2018). Couvse cs 23In. Redwood City, CA: Stanford University.

[12] S. Saha, "A Comprehensive Guide to Convolutional Neural Networks–the ELI5 way," *Data Science,* vol. 1, no. 1, pp. 1-11, 2018.

[13] Z. Cui, H. Chang, S. Shan et al., "Deep network cascade for image super-resolution," in *Proc. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, pp. 49-64.

[14] M. Aharon, Elad, M. Elad, and A.Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," in *Proc. IEEE Transactions on Signal Processing*, 2006, pp. 4311-4322.

[15] Kaggle. (2020, June). *Flower classification2019* [Online]. Available: https://www.kaggle.com/c/flower-classification-2019/overview

[16] M. E. Nilsback and A. Zisserman. (2020, Jun). *102 Category Flower Dataset* [Online]. Available: http://www.robots.ox.ac.uk/~vgg/data/flowers/102/index.html

**Thitiporn Khantong** is a senior student at Faculty of Engineering And Technology, Panyapiwat Institute of Management. She graduated B.Eng. in Computer Engineering from Bangkok University in 2015.

**Parinya Sanguansat** is Associate Professor in Electrical Engineering and head of Computer Engineering and Artificial Intelligence at Panyapiwat Institute of Management (PIM), Thailand. He graduated B.Eng., M.Eng. and Ph.D. in Electrical Engineering from Chulalongkorn University. His research areas include Machine Learning, Image processing and Computer Vision. He got many research grants from both private and public organizations. He has written several books about Machine Learning and MATLAB programming.