Multiple Classification of Flower Images Using Transfer Learning

Emine CENGİL
Computer Engineering Department
Firat University
Elazig, Turkey
ecengil@firat.edu.tr

Ahmet ÇINAR
Computer Engineering Department
Firat University
Elazig, Turkey
acinar@firat.edu.tr

***Abstract*—Deep learning technologies have been successful in many fields in recent years. Image classification problem is one of the areas where the use of the results is successful. The study draws attention to the use of pretrained models in problem solving. With the approach called transfer learning, frequently used pretrained deep learning models such as Alexnet, Googlenet, VGG16, DenseNet and ResNet are used for image classification. The results show that the models used achieve acceptable performance rates while the highest performance is achieved with the VGG16 model.**

Keywords — Deep Learning; Convolutional Neural Networks; Transfer Learning; Image Classification.

I. INTRODUCTION

In recent times, considerable improvement has been made in machine learning with the information obtained from the studies of the neurologists and academic environment. The concept of artificial intelligence, which starts with the idea of teaching the intelligence to the machines which separates people from other living things and is the most valuable thing that has been possessed, has evolved with the development of new techniques introduced by scientific areas. These techniques lately mentioned deep learning is often used widely.

Deep learning methods are based on neural networks and energy based models. Deep learning uses many processing unit layers for feature extraction. Each accomplished layer gets the output from the former layer as input [1].

In deep learning, the first layers extract more common features such as edge information, while the next layers extract more specific features. High-level features are reproduced from low-level features to constitute a hierarchical presentation. This drawing represents learning of multiple levels corresponding to dissimilar levels of abstraction [2]. Deep learning is basically based on learning from the representation of the data.

Deep learning methods are the algorithms used to solve many machine learning problem. Image classification area, is one of the areas where use of deep learning. Image classification is a subject of image processing that is quite extensive. It is a vital technology that many people use in every aspect of human life. Classification is the duration of supplying that images that do not belong to any class are included in their classes. [3].

Classification can be binary or multiple. There are many studies conducted in the literature for image classification. CNN architecture is the most preferred method. This was due to Imagenet [4] competition organized in 2012. A. Krizevsky et al. [5] participated in the competition with the model they created using convolutional neural networks and showed the best performance result of recent times.

Y. Seo et al. [6], utilized the Convolutional Neural Network (CNN) to classify the apparel image in the fashion business. In the study, it was pointed out that the clothing classification may be difficult due to the lack of tagged image data for various categories of clothing and for each category. Therefore, they have recommended GoogLeNet architecture pre-training in the ImageNet dataset and fine-tuning the fine-grained data set based on design features.

N. Hnoohom et al. [7], presented a forecast model for classifying fast food images in Thailand. The model was trained on natural images (GoogLeNet dataset) and used a deep learning process that was fine tuned to create the predictive Thailand fast-food model.

Y. A. L. Alsabahi et al. [8], used transfer learning based on Inception V3 model to classify DR images. They used the weight of the Inception V3 model trained in the ImageNet dataset and fine-tuned their data sets.

In the study, multiple classification problem was investigated. For this purpose, the training process has been realized by using pretrained models and kaggle flowers recognition dataset [9]. While the process was carried out with the matlab program, the Deep Network Design toolbox was used for the designing of the model. The results of the experiments with the pretrained CNN architectures used in most machine learning problems are shown.

The organization of this study is as follows; part 2 describes the structure of convolutional neural networks. Part 3 is the proposed method, which includes the data used, the transfer learning approach and the experimental results. Finally, section 4 describes the results of the study.

II. CONVOLUTIONAL NEURAL NETWORK

Convolutional neural networks are one of the deep learning architectures. While deep learning has achieved success in almost every field in which it is applied, convolutional neural networks have been most common in solving image classification problems. CNN is based on learning a hierarchy of properties. Convolutional neural networks (CNN) are multi-layered structures.

CNN is feed-forward and is a very influential method for detection. The network structure is simple; has fewer training parameters. CNN is fed forward and is a very effective method for detection. On the other hand, the complexity of the network model and the number of weights are decreased. CNN, as in traditional neural networks, has a loss function as a softmax in the last layer [10].

A. Convolution Layer

Convolutional neural networks are named after the convolutional process. The layer in which the convolution process is applied is called the convolution layer. The purpose of applying this layer is to extract the features of the processed data. Convolutional neural networks provide very good performance in feature extraction. This is because the layers of convolution that it contains.

Filtering process is applied to input image in convolution layer. In a convolutional neural network, there are usually multiple layers of convolution. In the first layers more general features of the image are extracted. Edge information is one of these features. There are many filters used to obtain edge information. The sudden color transitions on the image indicate that there is a edge where the transition occurs. In the later layers of the network, more specific features are extracted from the edge information. Filtering is the mathematical process of several matrices. The results are not the same as the filters used indicate the difference of their digital contents. The network is able to learn these values in the process of education [11].

The filter size applied to the image during convolution is generally 3x3, 5x5, 7x7 or 11x11 size filters. The larger the filter size applied to the image, the smaller the result image. This situation causes loss of information. For this reason, while the convolution process is generally applied in CNN models, small size filters such as 3x3 are preferred. Besides, it is not always good to use small size filters. There may also be situations where the use of large filters is advantageous. Filters operate the feature extraction business. Thus, the filter size varies according to the network used and the desired process [12].

B. Activation Layer

Deep learning is generally used to solve non-linear problems. This is because deep learning is more successful in nonlinear problems than other methods. The values obtained after the matrix multiplication in the convolution layer are linear. Activation functions are used to convert values to non-linearity.

In the CNN model, a non-linear layer is applied after each convolution layer. This layer is the activation layer [13]. Tanh, Sigmoid and ReLU are some of the most preferred activation functions in CNN models. Relu, which is one of the activation functions applied in order to gain non-linearity to the layer output, is used more frequently in deep learning problems than other functions [14].

C. Pooling Layer

In convolutional networks architecture, pooling layer is usually applied after convolution layers. The pooling layer is used to decrease the size of the image. The size of the image may be considerably large after the first convolution. So pooling should be done. The pooling process is the technique of filtering the kernel. There is more than one pooling method. The most common of these are; maximum, minimum and average pooling. The depth dimension does not change after pooling [15].

D. Fully Connected Layer

In convolutional neural network architectures, layers of convolution, activation and pooling are followed by a fully connected layer. This layer depends on all neurons of the previous layer. The fully connected layer is similar to the multi-layer perceptron [13]. The fully connected layer is the higher-level representation of the input image, the output resulting from the previously applied convolution, activation and pooling layers. These layers are not expected to make classification estimates. At this point, the fully connected layer is used to classify the input image based on the training set by looking at the features.

III. PROPOSED METHOD

In the proposed method, classification process was carried out by means of convolutional neural network models with the transfer learning approach. Deep Network Design that is a Matlab tool used for design of widely used models.

The study was performed under the same conditions using the NVIDIA Geforce Gtx 950M Gpu with AlexNet, VGG16, DenseNet201, googleNet and ResNet models. Five-class Kaggle Flowers data were used for the classification process.

A. Dataset

The dataset consists of five classes including chamomile, tulip, rose, sunflower and dandelion. Each class contains approximately 800 photos. Photos are not high resolution and are 320x420 pixels. A number of the dataset images are as shown in Figure 1.

Figure 1. Images from flowers dataset

Pictures are reduced to a single format and have different proportions. Data collection is based on scraped data from flickr, google images and yandex images [9].

B. Transfer Learning

While convolutional neural network architecture is used to solve machine learning problems, the network can be started with random values for training and can be done by transfer learning process. It is not a widely used method to train the network from scratch. This is because the amount of time and data required for training is high.

Transfer learning is a technique of learning a machine where a pretrained model is intended to be redefined in a second relevant task [17].

In this study, the convolutional neural network models, which were pretrained with multiple gpus for 2 to 3 weeks using Imagenet dataset containing millions of data, were re-trained with flowers dataset. Imagenet data consists of 1000 classes. Our data is composed of five classes. The class information in the fully connected layer, which is the last layer for the classification operation, is set to five. The training process was carried out in this way.

C. Experimantal Results

Since the introduction of deep learning, many models have been developed by academic and scientific areas to be used in problems. Model design was made by trial and error or by using adaptive methods. The models, which were shown to be high by using the models created by the designer, have been the focus of attention of the researchers and used these ready models in their problems. In this way some models have been widely used and still continue to be used. Models used in the study; AlexNet, ResNet, VGG16, GoogleNet and DenseNet201.

AlexNet: The network consists of 11 layers. The high number of layers in the network has a positive effect on feature extraction. In addition, the number of parameters has a negative effect on speed. The first layer of AlexNet is the convolution layer. The convolution layer follows maximum pooling and normalization layers. The final layer is the softmax which makes the classification process [5].

VGG16: The VGGNet [19] is the abbreviation for Very Deep Convolutional Networks. The network has addressed important aspects of CNN architectural design. The depth of this structure makes it very deep and abstract.

In the structure where one of the data increasing methods is used, Relu activation function is used after each convolution layer. The stack is trained with gradient descent. Small filters of 3×3 size were used in all layers.

ResNet: In ILSVRC 2015, a new architecture called Residual Neural Network was presented by Kaiming et al. Such links are also cognoscible as gated units or gate repeating units, and are similar to the last successful layers applied to RNNs. By the agency of to this technique, they can train a very deep-layered network while having a lower complexity than the VGGN [19].

GoogleNet: The winner of ILSVRC 2014 was GoogleNet (Inception). GoogleNet, designed by researchers from the Google group, has performed very close to human performance. The training lasted for several days. The network used a CNN inspired by LeNET-5 [20]. Here comes the name of the name. A new element was applied to an initial module.

All convolutions, including those in the Inception modules, use rectified linear activation. This module is based on a number of very small convolutions to decrease the number of parameters. The architecture consisting of 22 layers reduced the number of parameters, which are 60 million, to 4 million.

DenseNet: The structure has been developed by linking each layer to all other layers in the style of forward feed, adopting the fact that there are shorter connections between the intersecting networks near the input and those close to the output, which can be deeper, more accurate and efficient to train. For each layer, the property maps of all previous layers as input, their feature maps are also used as input to all subsequent layers. DenseNets achieves significant improvements to the latest technology in most areas, requiring less memory and calculation to achieve higher performance [21].

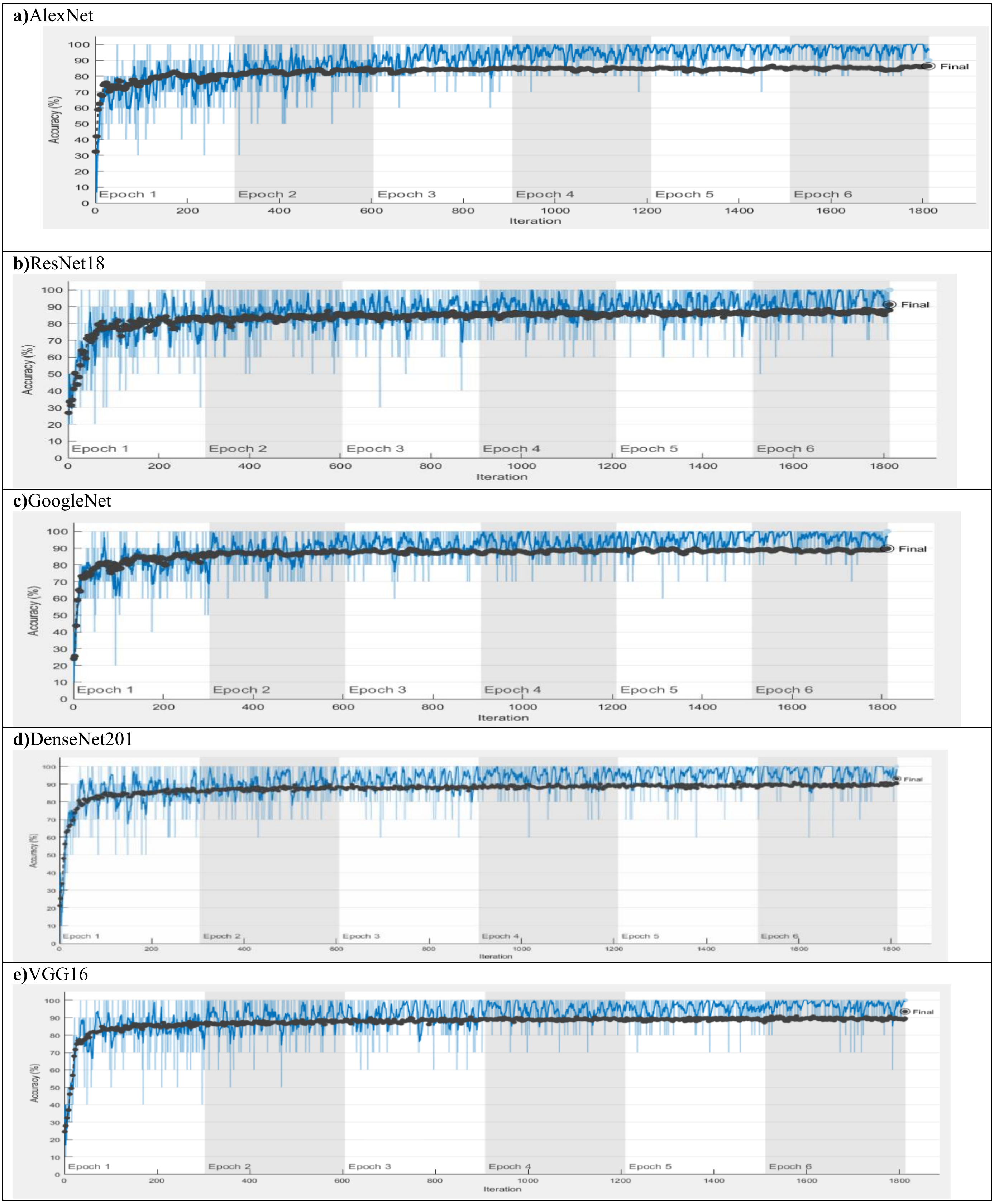

Figure 2. Accuracy graphics of Models a)AlexNet b)ResNet18 c)GoogleNet d)DenseNet201 e)VGG16

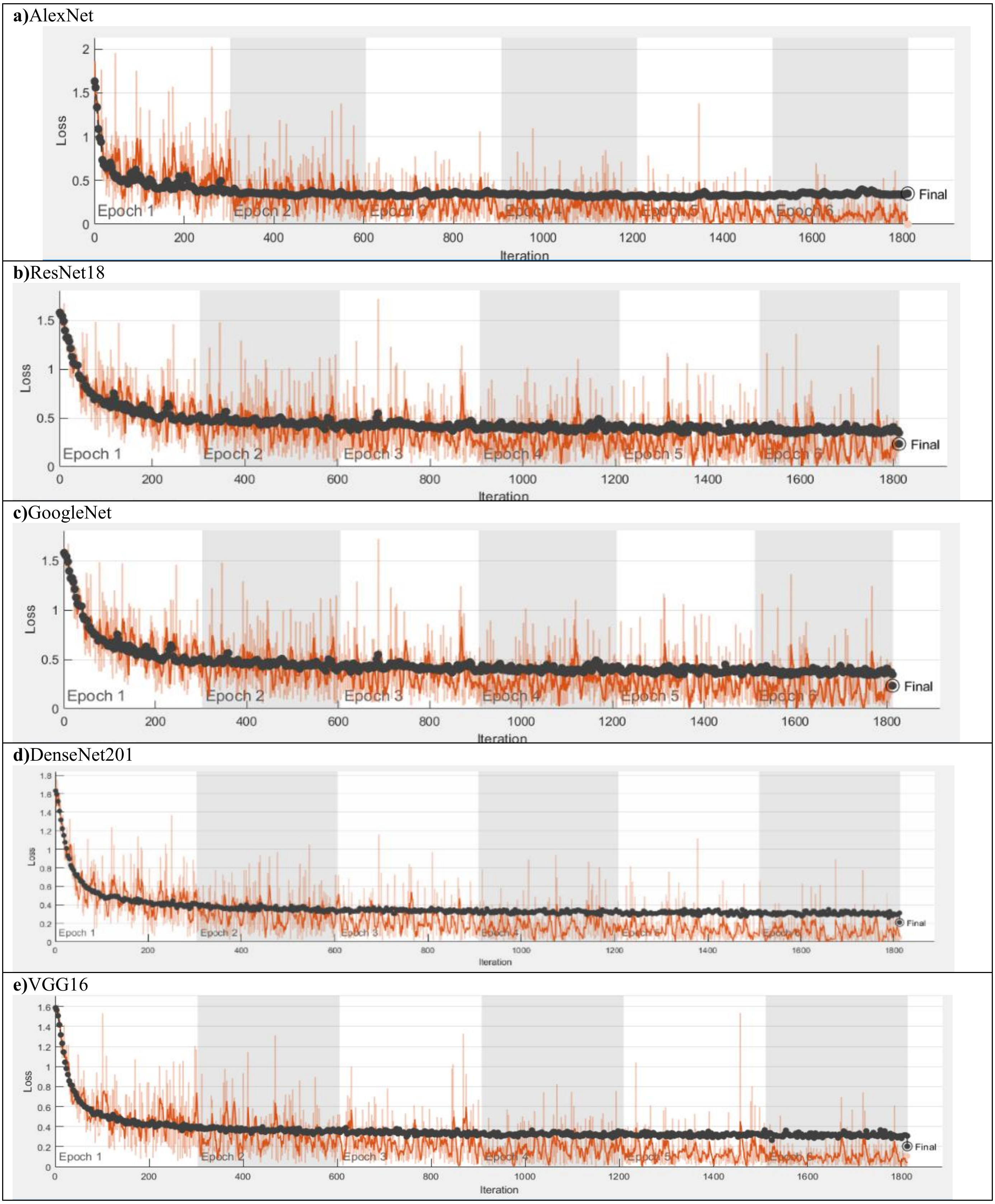

Figure 3. Loss graphics of Models

In the study, the pretrained models were tested by making them suitable for the classification task. Each model was operated with a maximum of 1812 iterations and a single GPU. The operating performances of the models are as in Table I. Validation accuracy as a model performance criterion has been shown. According to the results, the highest validation accuracy was realized with the VGG16 network. When the operating times are considered, DenseNet201 has the longest working time due to the number of parameters The accuracy and loss graphs of the models are as in figure 2 and figure 3.

TABLE I. EXPERIMANTAL RESULTS

Models	**Validation Accuracy**	**Elepsad time**	**Hardware Resource**	**Max. Iterations**
AlexNet	86.28	108 min. 49 sec.	Single GPU	1812
Resnet18	91.29	186 min. 40 sec.	Single GPU	1812
GoogleNet	89.75	127 min. 22 sec.	Single GPU	1812
DenseNet201	93.06	479 min. 21 sec.	Single GPU	1812
VGG16	93.52	467 min. 37 sec.	Single GPU	1812

IV. CONCLUSION

In this study, the importance of transfer learning approach is emphasized. Initialization of networks used for the classification task with pretrained networks rather than random weights is advantageous in terms of time and accuracy performance criteria.

The results of the five CNN models, which are widely used in image classification processes, are shown. According to the results, multi-classification work showed acceptable accuracy performance for each model while highest accuracy was seen with VGG16 network.

REFERENCES

[1] L. Deng and D. Yu, “Deep Learning: Methods and Applications,” Found. Trends® Signal Process., vol. 7, no. 3–4, pp. 197–387, 2014. Douangnoulack, Phonethep, and Veera Boonjing. "Building Minimal Classification Rules for Breast Cancer Diagnosis." *2018 10th International Conference on Knowledge and Smart Technology (KST)*. IEEE, 2018.

[2] Y. Bengio, “Learning Deep Architectures for AI,” Found. trends® Mach. Learn., vol. 2, no. 1, pp. 1–127, 2009. Aksebzeci, Bekir Hakan, and Ömer Kayaalti. "Computer-aided classification of breast cancer histopathological images." *Medical Technologies National Congress (TIPTEKNO), 2017*. IEEE, 2017.

[3] M. Sonka, V. Hlavac, and R. Boyle. Image processing, analysis, and machine vision. Cengage Learning, 2014.

[4] O. Russakovsky, J. Deng, H. Su, J., Krause, S. Satheesh, S. Ma,... and A. C. Berg,. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), 211-252, 2015.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton ImageNet classification with deep convolutional neural networks. In NIPS, pp. 1106–1114, 2012.

[6] Seo, Yian, and Kyung-shik Shin. "Image classification of fine-grained fashion image based on style using pre-trained convolutional neural network." *2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA)*. IEEE, 2018.

[7] Hnoohom, Narit, and Sumeth Yuenyong. "Thai fast food image classification using deep learning." *2018 International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON)*. IEEE, 2018.

[8]Alsabahi, Y. A. L., Lei Fan, and Xiaoyi Feng. "Image Classification Method in DR Image Based on Transfer Learning." *2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, 2018.

[9]https://www.kaggle.com/alxmamaev/flowers-recognition

[10] Ma W, Lu J., An Equivalence of Fully Connected Layer and Convolutional Layer. arXiv preprint arXiv:1712.01252. 2017.

[11] Accelerate Machine Learning with the cuDNN Deep Neural Network Library, https://devblogs.nvidia.com/accelerate-machine-learning-cudnn-deep-neural-network-library/

[12] Liu T, Fang S, Zhao Y, Wang P, Zhang J., Implementation of training convolutional neural networks. arXiv preprint arXiv:1506.01195. 2015.

[13] Pasi KG, Naik SR., Effect of parameter variations on accuracy of Convolutional Neural Network. InComputing, Analytics and Security Trends (CAST), International Conference on 2016 Dec 19 (pp. 398-403). IEEE.

[14] http://cs231n.github.io/neural-networks-1/ CS231n: Convolutional Neural Networks for Visual Recognition

[15] Lee CY, Gallagher PW, Tu Z., “Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree”, InArtificial Intelligence and Statistics (pp. 464-472), 2016.

[16] Ian Goodfellow Yoshua Bengio and AAron Courville, Deep Learning , page 526,2016.

[17] Guo Z, Chen Q, Wu G, Xu Y, Shibasaki R, Shao X., Village Building Identification Based on Ensemble Convolutional Neural Networks. Sensors.;17(11):2487, 2017.

[18] He K, Zhang X, Ren S, Sun J., Deep residual learning for image recognition. InProceedings of the IEEE conference on computer vision and pattern recognition, (pp. 770-778), 2016.

[19] LeCun Y,. LeNet-5, convolutional neural networks. URL: http://yann. lecun. com/exdb/lenet. 2015:20.

[20] Huang, Gao, et al. "Densely connected convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.