



Classification of Animals with Different Deep Learning Models

Özkan İNİK^{a*}, Bülent TURAN^b

^{a,b}Department of Computer Engineering, Gaziosmanpaşa University, Tokat/ Turkey.

^aozkan.inik@gop.edu.tr

^bbulent.turank@gop.edu.tr

*Corresponding author

ABSTRACT: The purpose of this study is that using different deep learning models for classification of 14 different animals. Deep Learning, an area of artificial intelligence, has been used in a wide range of recent years. Especially, it using in advanced level of image processing, voice recognition and natural language processing fields. One of the most important reasons for using a large field in image analysis is that it performs the feature extraction itself on the image and gives high accuracy results. It performs learning by creating at different levels representations for each image. Unlike other machine learning methods, there is no need of an expert for feature extraction on the images. Convolution Neural Network (CNN), which is the basic architecture of deep learning models, consists of different layers. These are Convolution Layer, ReLu Layer, Pooling Layer and Full Connected Layer. Deep learning models are designed using different numbers of these layers. AlexNet and VggNet models are used for classified of 14 different animals. These animals are Horse, Camel, Cow, Goat, Sheep, Wolf, Dog, Cat, Deer, Pig, Bear, Leopard, Elephant and Kangaroo respectively. Animals that are most likely to encounter when during driving road were selected. Because thinking this work to be a preliminary work for the control of autonomous vehicle driving. The images of animals are collected in color (RGB) on the internet. In order to increase the data diversity, images were also taken from the ready data sets. A total of 150 images were collected with 125 training and 25 test data for each animal. Two different data sets have been created, with each image having dimensions of 224x224 and 227x227. As a result of the study, the classification of the animals was realized with %91.2 accuracy with VggNet and %67.65 with AlexNet. The high error rate in AlexNet is due to the small number of layers in the network and the high selection of parameter values. For example, the filter size in the convolution layer in AlexNet architecture is 11x11 and the number of stride is 4. This situation causes data loss in transferring the information to the next layer. In contrast, VggNet has a filter size of 3x3 and a number of steps of 1, there is no data loss in the transfer to the next layer.

Keywords : AlexNet, CNN, Classification of Animals, Deep Learning, VggNet

1. Introduction

Image classification is the process of assigning one or more labels to an image according to based on a content. This is the standard supervised learning problems. The purpose of this problem is to train the system with a training set consisting of labeled images and then guessing image label when a new image is given. In the past, large-scale image classification in the field of computer vision and machine learning has received intense interest (Bengio et al., 2010; Deng et al., 2010; Deng et al., 2011; Lin et al., 2011; Rohrbach et al., 2011; Sánchez and Perronnin, 2011). Studies on image classification have gained a different dimension with the introduction of Deep Learning. Deep Learning is a subdivision of artificial intelligence and became popular in 2012. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC-2012) competition was won by AlexNet (Krizhevsky et al., 2012), a deep learning model, which has been a tremendous success in object classification. It achieved a winning top-5 test error rate of %15.3, compared to %26.2 achieved by the second-best entry.

Deep learning models (Bengio et al., 2013; Girshick, 2015; Girshick et al., 2014; He et al., 2016; Krizhevsky et al., 2012; Le, 2013; Ren et al., 2017; Simonyan and Zisserman, 2014; Szegedy et al., 2015; Zeiler and Fergus, 2014) discover the feature of images from raw data. In general, the first layers of these architectures consist of the convolution (Jarrett et al., 2009; LeCun et al., 1990; LeCun et al., 2004), next pooling layer, next fully connected layer and the latest classification layer. Because of their high performance in image classification, deep nets are used for voice recognition (Amodei et al., 2016; Bahdanau et al., 2016; Graves et al., 2013; Hinton et al., 2012), natural language processing (Hermann et al., 2015; Jozefowicz et al., 2016; Lample et al., 2016; Luong et al., 2015), robotics (Lenz et al., 2015; Levine et al., 2016) and object detection (Long et al., 2015; Redmon et al., 2016; Ren et al., 2015).

In the case of autonomous vehicle driving, it is very important for the vehicle to perceive the objects in the outdoor environment. In particular, in order to avoid crashing into a possible object while driving, it is necessary to perceive the object and maneuver according to the situation. In this study, 14 different groups of animals were classified. In this way, it was tried to detect the animals that would be in front of the vehicles while driving. Thus, object identification and classification, which are the basic logic of autonomous control systems, have been done.

The structure of this study is as follows: Section 2 describes VggNet, AlexNet and Data Set. In Section 3, the experimental work done with VggNet and AlexNet model is explained. Finally, Section 4 describes the conclusion.

2. Material and Methods

A. AlexNet

The first study of Deep Learning relies on Yann LeCun's study of document recognition (Lecun et al., 1998). However, Deep Learning has been widely heard with ImageNet competition at 2012. Because, for the first time in 2012, the error rate in the visual object classification has dropped sharply (Figure 1). When we look at Figure 1, the error rate in 2011 was 26.2%, but in 2012 it dropped to 15.3%. This success has been achieved with AlexNet (Krizhevsky et al., 2012).

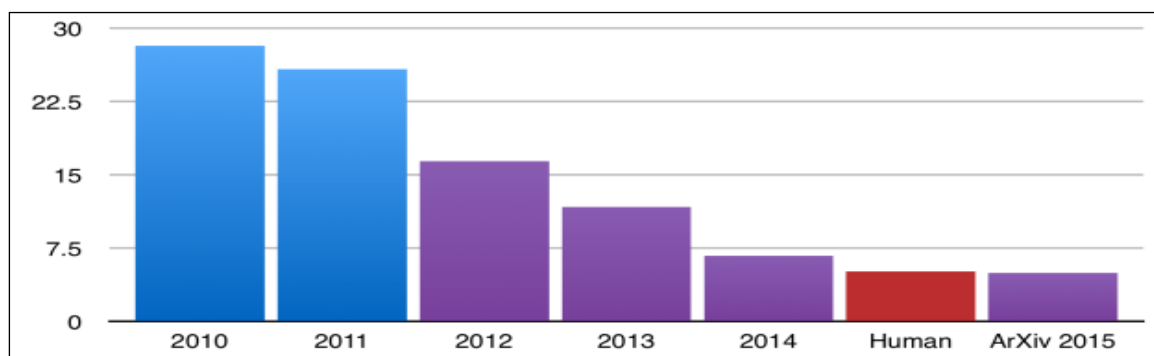


Figure 1: The Top-5 Error Rate according to Years. Error rate has been rapidly reducing since the introduction of deep neural networks in 2012 (devblogs.nvidia.com, 2016).

The AlexNet model is given in Figure 2. This model is designed to classify 1000 objects. The filter size is 11x11 and the number of stride is 4. The first convolution layer has 96 filters. The input layer image size is 224x224x3.

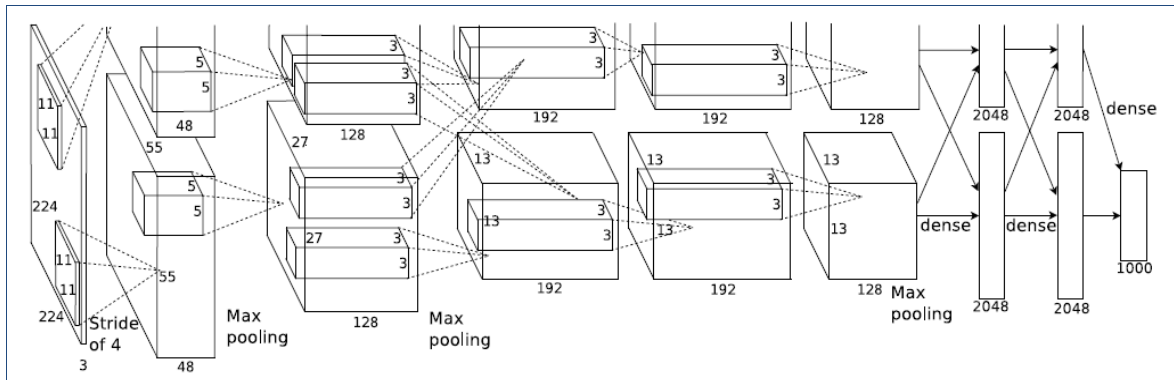


Figure 2. Layer Structure of AlexNet Model(Krizhevsky et al., 2012).

B. VggNet

VggNet is a deep learning model designed by Karen Simonyan and Andrew Zisserman (Simonyan and Zisserman, 2014). The model has been the winner ImageNet competition with a 7.3% error rate at the 2014. Its basic layer structure is similar to the AlexNet(Krizhevsky et al., 2012) model. VggNet model is important because of it is deeper than previous models. Thus, VggNet has reinforced the idea that deep learning networks should have a deep layer of network for the study of the hierarchical representation of visual data. The model filter size is 3x3 and the stride is 1. The loss of information has been removed from the previous layer to the next layer with the number of stride being 1. The model is given in Figure 3. It has 6 different architecture of VggNet in Figure 4. The configuration D (VggNet-16) produced best result.

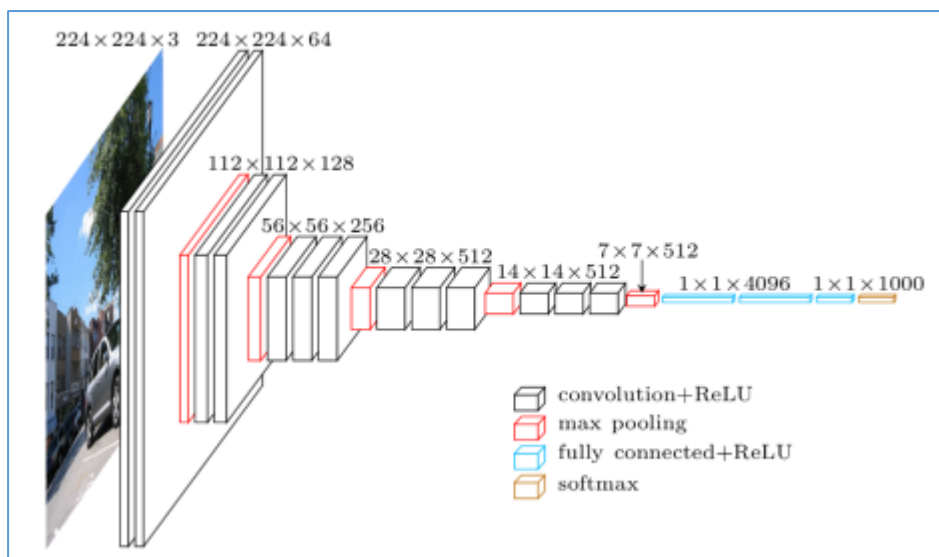


Figure 3. VggNet Model Architecture(Heuritech, 2018).

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 4. Different Architecture of VggNet Model (Deshpande, 2018).

C. Data Set

In this study, a new dataset was created for the classification of 14 different animals. These animals are; Horse, Camel, Cow, Goat, Sheep, Wolf, Dog, Cat, Deer, Pig, Bear, Leopard, Elephant, Kangaroo. Images of each animal are collected on the internet in the form of color (RGB). In order to increase the diversity of data, images were taken from STL-10 dataset(Coates et al., 2011). A total of 150 images were collected for each animal. The dataset consists of a total of 2100 images. For training %90 of these images were used and the remaining %10 were used for testing. Two different data sets have been created, with each input image size of 224x224 and 227x227. The classes of the prepared dataset are given in Figure 5.











































Class Num.	Class Name	Images			Class Num.	Class Name	Images		
1	Horse				8	Kangaroo			
2	Bear				9	Goat			
3	Camel				10	Cat			
4	Pig				11	Dog			
5	Elephant				12	Sheep			
6	Deer				13	Wolf			
7	Cow				14	Leopard			

Figure 5. The Classes of Dataset

The flow diagram of the studies of paper is given in Figure 6. Each step in Figure 6 is given below.

1. First of all the pictures of all the animals collected from the internet
2. Image pre-processing such as cropping, shifting, mirroring on the collected images was performed.
3. Create datasets in 224x224x3 and 227x227x3 image dimensions
4. Update AlexNet and VggNet models according to the number of classes
5. These models were trained and tested

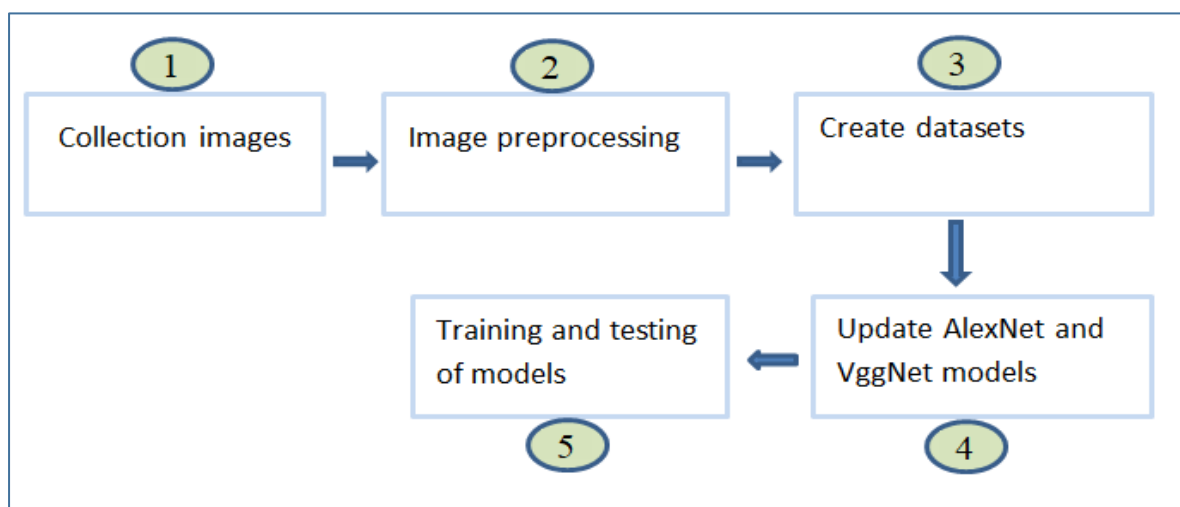


Figure 6. The Flow Diagram of the Method

Table 3. Classes in the Confusion Matrix

Num.	Class	Num.	Class	Num.	Class	Num.	Class
1	Horse	5	Elephant	9	Goat	13	Wolf
2	Bear	6	Deer	10	Cat	14	Leopard
3	Camel	7	Cow	11	Dog		
4	Pig	8	Kangaroo	12	Sheep		

4. Conclusion

In this study, VggNet model and AlexNet model from Deep Learning Models were used for animal classification. A new data set was created for the classification of animals. As a result of the study, the classification of the animals was realized with %91.2 accuracy with VggNet and %67.65 accuracy with AlexNet. The high error rate in AlexNet is due to the small number of layers in the network and the high selection of parameter values. For example, the filter size in the convolution layer in AlexNet architecture is 11x11 and the stride is 4. This number of stride causes data loss for next layers. In contrast, VggNet has 3x3 filter size and 1 stride, so there is no data loss in the next layer transfer.

Deep learning models are usually training with big datasets and results are obtained. However, in this study, it was seen that few data sets achieved a remarkable achievement.

Acknowledgment

This study was supported by Gaziosmanpaşa University under Scientific Research Projects. Project No: 2017/39 Project Name: " Derin Öğrenme İle Nesne Tanımlama".

References

- Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., Casper, J., Catanzaro, B., Cheng, Q., Chen, G., 2016. Deep speech 2: End-to-end speech recognition in english and mandarin, International Conference on Machine Learning, pp. 173-182.
- Bahdanau, D., Chorowski, J., Serdyuk, D., Brakel, P., Bengio, Y., 2016. End-to-end attention-based large vocabulary speech recognition, Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on. IEEE, pp. 4945-4949.
- Bengio, S., Weston, J., Grangier, D., 2010. Label embedding trees for large multi-class tasks, Advances in Neural Information Processing Systems, pp. 163-171.
- Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence 35, 1798-1828.
- Coates, A., Ng, A., Lee, H., 2011. An analysis of single-layer networks in unsupervised feature learning, Proceedings of the fourteenth international conference on artificial intelligence and statistics, pp. 215-223.
- Deng, J., Berg, A.C., Li, K., Fei-Fei, L., 2010. What does classifying more than 10,000 image categories tell us?, European conference on computer vision. Springer, pp. 71-84.
- Deng, J., Satheesh, S., Berg, A.C., Li, F., 2011. Fast and balanced: Efficient label tree learning for large scale object recognition, Advances in Neural Information Processing Systems, pp. 567-575.
- Deshpande, A., 2018. <https://adeshpande3.github.io/adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html>.
- devblogs.nvidia.com, 2016. Deep Learning for Julia.
- Girshick, R., 2015. Fast R-CNN. Ieee I Conf Comp Vis, 1440-1448.
- Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. 2014 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr), 580-587.
- Graves, A., Mohamed, A.-r., Hinton, G., 2013. Speech recognition with deep recurrent neural networks, Acoustics, speech and signal processing (icassp), 2013 ieee international conference on. IEEE, pp. 6645-6649.

- He, K.M., Zhang, X.Y., Ren, S.Q., Sun, J., 2016. Deep Residual Learning for Image Recognition. 2016 Ieee Conference on Computer Vision and Pattern Recognition (Cpvr), 770-778.
- Hermann, K.M., Kocisky, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., Blunsom, P., 2015. Teaching machines to read and comprehend, *Advances in Neural Information Processing Systems*, pp. 1693-1701.
- Heuritech, 2018. <https://blog.heuritech.com/2016/02/29/a-brief-report-of-the-heuritech-deep-learning-meetup-5/>.
- Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., 2012. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine* 29, 82-97.
- Jarrett, K., Kavukcuoglu, K., LeCun, Y., 2009. What is the best multi-stage architecture for object recognition?, *Computer Vision, 2009 IEEE 12th International Conference on. IEEE*, pp. 2146-2153.
- Jozefowicz, R., Vinyals, O., Schuster, M., Shazeer, N., Wu, Y., 2016. Exploring the limits of language modeling. *arXiv preprint arXiv:1602.02410*.
- Krizhevsky, A., Sutskever, I., Hinton, G., 2012. ImageNet classification with deep convolutional neural networks. In *NIPS'2012* . 23, 24, 27, 100, 200, 371, 456, 460.
- Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., Dyer, C., 2016. Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360*.
- Le, Q.V., 2013. Building high-level features using large scale unsupervised learning, *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE*, pp. 8595-8598.
- LeCun, Y., Boser, B.E., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W.E., Jackel, L.D., 1990. Handwritten digit recognition with a back-propagation network, *Advances in neural information processing systems*, pp. 396-404.
- Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86, 2278-2324.
- LeCun, Y., Huang, F.J., Bottou, L., 2004. Learning methods for generic object recognition with invariance to pose and lighting, *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. IEEE*, pp. II-104.
- Lenz, I., Lee, H., Saxena, A., 2015. Deep learning for detecting robotic grasps. *The International Journal of Robotics Research* 34, 705-724.
- Levine, S., Pastor, P., Krizhevsky, A., Ibarz, J., Quillen, D., 2016. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *The International Journal of Robotics Research*, 0278364917710318.
- Lin, Y., Lv, F., Zhu, S., Yang, M., Cour, T., Yu, K., Cao, L., Huang, T., 2011. Large-scale image classification: fast feature extraction and svm training, *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE*, pp. 1689-1696.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440.
- Luong, M.-T., Pham, H., Manning, C.D., 2015. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*.
- Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788.
- Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks, *Advances in neural information processing systems*, pp. 91-99.
- Ren, S.Q., He, K.M., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Ieee T Pattern Anal* 39, 1137-1149.
- Rohrbach, M., Stark, M., Schiele, B., 2011. Evaluating knowledge transfer and zero-shot learning in a large-scale setting, *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE*, pp. 1641-1648.
- Sánchez, J., Perronnin, F., 2011. High-dimensional signature compression for large-scale image classification, *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE*, pp. 1665-1672.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Szegedy, C., Liu, W., Jia, Y.Q., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going Deeper with Convolutions. *Proc Cvpr Ieee*, 1-9.
- Zeiler, M.D., Fergus, R., 2014. Visualizing and Understanding Convolutional Networks. *Computer Vision - Eccv 2014, Pt I* 8689, 818-833.