International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

Birds Identification System using Deep Learning

Prabhjot Kaur #1,Shikhar Gupta#2, Deepak Kumar#3, Kshitij Maurya#4, Vikram Singh#5

#Department of Computer Science & Engineering, Inderprastha Engineering College, Plot No 63, Site IV, Sahibabad, Ghaziabad, 201010, India

---

Abstract - *Birds species are disappearing or threatened with extinction today are in great numbers. Many of these species are rarely found and it is a difficult task to distinguish bird species, due to their different sizes, colors, and different views of humans at different angles.*

The difference between these animals should be noted as we know that it is easy for humans to see birds in pictures. In this study we focused on creating a model that can distinguish different species of birds using only their own image. For this, we used a convolutional neural network (DCNN) with ResNet-18 model. The accuracy of our model got improved using ResNet-18 transfer learning.

We used the Caltech-UCSD Birds 200 data set [CUB-200-2011] [1] for both training and testing purposes. To differentiate, using the Deep Convolutional Neural Network, DCNN training and testing is performed on Google Colab using the PyTorch Facebook library. The final results show that by using the transfer of learning we have achieved a test accuracy of 78.8 percent.

KeyWords: *Convolutional neural network(CNN), ResNet-18, Inception V3, VGGNet ,Transfer Learning.*

1. INTRODUCTION

Everyday life is often fast and busy and involves extracurricular activities. Bird watching is a hobby that can give you a break from everyday life and can also encourage you to be strong in the face of daily challenges. It can also provide health benefits and the joy of enjoying nature. Many people visit bird sanctuaries to see different species of birds or observe their beautiful feathers while observing the differences between the bird species and their characteristics. Understanding such differences between species can enhance our knowledge of rare birds and their ecosystem. However, because of spectator boundaries such as location, distance, and equipment, the bird's appearance is based on several factors.

As we all know, birds play a key role in maintaining environmental harmony. The presence of different species of birds in the natural environment is also important for a variety of environmental reasons. Exploring birds can tell us a lot about the world and the environment and can help capture important information about nature. Environmental scientists often use birds as an indicator of climate change as birds are sensitive to environmental changes and can be used to understand the ecosystem around us. Various real-world applications rely on birds, such as land pollution monitoring. The presence of bird species in the natural environment is also important for many natural reasons. This is another area where our classification system can be used.

But collecting and gathering information about birds requires a great deal of human effort and is a very expensive process. Therefore, a strong approach should be in place that can provide comprehensive bird information processing and serve as an important tool for professionals, agencies, and so on. Similarly, the composition of the species of birds plays an important role in determining the type of bird. Ornithologists (scientists who study birds) have been struggling to distinguish between different species of birds for decades. To distinguish the species of birds they must learn all the details of the birds, such as their climate, genetics, distribution, etc. Even professional birdwatchers sometimes disagree with the species of birds that represent this bird.

This is where machine learning and artificial intelligence work. Self-driving cars, Siri, etc., are some of the examples of machine learning and artificial intelligence used in the real world. So, why not use some practical artificial intelligence in the field of bird hunting. Machine learning, artificial intelligence, and mathematical design can help to identify many species of birds using their pictures. Today, image classification has become quite popular and has become one of the main fields of machine learning and in-depth learning. Identifying birds that create an image is a challenging task due to problems such as different bird species that vary greatly in shape and appearance, background image, lighting conditions in photos, and extreme variation in posture. In our project, we have tried to classify images of different species of birds using Deep Convolutional Neural Networks (DCNN) and transfer learning.

2. LITERATURE REVIEW

The studies [2,3] have focused on transfer learning techniques for classification. Currently, CNNs are capable

of analyzing birds from different angles and positions and have an accuracy of about 85%.

Recently, there has been some development in visual categorization methods for species identification [4] which extends its application in many domains.[5] involves the approach for generic object recognition.

In [6] Fagerlund and Herma developed signal processing techniques using Mel-Frequency Cepstral Coefficients (MFCCs). They achieved an accuracy of 71.3%.

Viches et al [7] considered that the identification of distinctive features is crucial so used data mining algorithms like ID3, J4.8, and Naïve Bayes.

[8] employs a global decision tree with Support Vector Machines (SVM) classifier in each node to separate two species. A. Marini [9] identified the species using both visual and vocal features of the bird using the CUB 200-2011 dataset and took the audio samples from the Xeno-Canto dataset.

In [12] Zhang et al. and Bilinear CNN by Lin et al. employed a two-stage pipeline of part detection and part-based object classification.

[12] involves training an individual CNN based on unique properties by detecting the location of parts of the object making it more efficient than R-CNN. On the other hand, Lin et al. [13] proposed a bilinear model having two streams with interchangeable roles.

The authors of the paper [14] developed a model to detect almost 200 types of objects (birds) with an accuracy of 71.5%. They used CUB-200-2011 dataset. In [15] the authors have used a region-based CNN (faster CNN) method that works, on the detection of birds based on the full image. The results show that the faster CNN works well as compared to other CNN models. In [16] work is done to improve faster CNN by using convolution and filter pruning techniques. The results obtained have improved the results to a great extent.

In [17], the cascaded convolutional neural network that does object detection is introduced. It uses three-level deep CNN, which does the object detection in a coarse to excellent manner. In[18], the authors have used CNN with the transfer learning model. They worked on the detection of the areas where the snow has been seen and for this CCTV footage was taken from the Japan website, which is publicly available.

In [19] the authors have used transfer learning to provide coarseness to the system and proved that a granular network is more effective. For this, they used the image dataset of COCO. In [20], they evaluated the R-CNN and SSD for evaluation of the manga (Japanese Comic) objects. The authors have shown that the fast RCNN works well for the character face and text detection. In [21] 360-degree panoramic images are used and post-processing is applied in the model to finetune the overall result.

In [22], the authors worked on the detection of harmful objects with the tensor flow API and have used the faster R-CNN algorithm for the experiment.

PROBLEM STATEMENT

We know that many species of birds are extinct or endangered and many of the birds not seen today were easily identified before that which is why this topic came to our minds. Humans cannot see each species of bird because of a certain degree of similarity. Identifying bird species can be a challenge for humans which is why we use algorithms to perform tasks for us.

The database was launched in 2010 and contains 6000 photographs of 200 classes of birds. Related to this have been additional label data including binding boxes, incorrect classification, and additional attributes. This was revised in 2011, to add more images, bringing the total number of images in the database to about 12,000. Available attributes have also been updated to include 15 parts, 312 binary attributes, and a box that includes each image. For the most part in this series, we will simply use pictures and class labels to develop and train bird prediction networks.

PROBLEM OBJECTIVE

In this project, we have used the deep convolutional neural network to classify various images of birds. The bird species classifier is trained and tested using the Caltech UCSD Birds 200 dataset. We have used Deep Convolutional Neural Networks (DCNN) on an image that is converted into a greyscale format to generate an autograph using the PyTorch library, where multiple nodes will be generated and then these different nodes are compared with the testing dataset, and the scores are saved in a sheet. After that, we analyzed this score sheet we used the algorithm with the highest prediction accuracy in our desktop application.

The major objectives of this model are:

- To recognize the birds using Deep Convolutional Neural Networks and image morphology.
- To develop a real-time application that can be used to detect various species of birds.
- To display all the information related to the bird.

3. PROPOSED METHODOLOGY

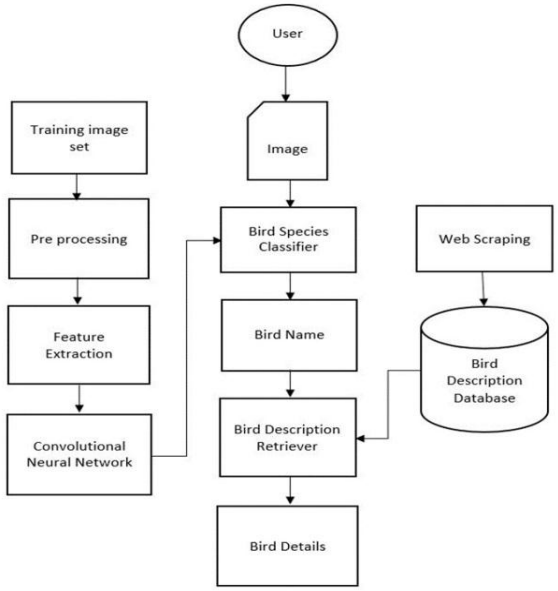

Fig.1

The block diagram given above represents our approach to the problem. The description of each block of the diagram is given below:

• Image: In this step, the input image is taken from the user to feed to the trained system.

• Training Image set: This belongs to the dataset which is used to train the model. It can contain around 7000 images of 200 species of birds which are used to train the CNN.

• Pre-processing: This step aims to process the image according to the algorithm's needs. It aims to remove the unwanted distortions and enhance an image so that our model can benefit from this improved data to work on.

• Feature Extraction: It is a process of dimensionality reduction by which a set of raw data is reduced to data that can be used for processing, alternatively, we can say that it aims to take out the meaningful data from the image without losing any information.

• Convolutional Neural Network: As mentioned above, it is a deep learning model that is used as an image classifier to classify images based on their features.

• Web scraping: It is a technique that is used to extract data from websites. The data taken from the websites is stored in a database which is later used to extract all the information about the birds.

• Output: Output will contain all the data related to the species of the bird.

Working

a) The image of the bird will be given by the user to identify the name and other details of the species of the birds .
b) Image will be transformed and analyzed using the Deep convolution Neural network technique and the class of bird will be predicted.
c) This class will be used to extract the details of the birds.
d) Name and other details will be displayed on the screen of the user.

4. PROJECT MODULE

There are three modules in our project which is used to develop the application for bird detection. They are:

- Bird Species Classifier
- Bird Description Database
- Desktop Application

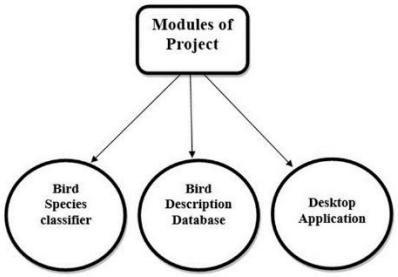

Fig.2

- **BIRDS SPECIES CLASSIFIER** - Under this module, we are going to perform the operations which are needed by our application to classify the various birds. We are going to perform various steps like image preprocessing, feature extraction, applying a convolutional neural network (CNN) to the images, and then it will classify the image. In the image preprocessing step, we are going to improve the image by suppressing unwanted distortions, and/or also we are going to enhance some of the important image features so that our models can benefit from this improved data and can work on this data and provide us better accuracy. In this, we are going to perform steps like resizing the image, denoising the image, segmentation, morphology, etc. to process the image and after that, we feed it to the model. After this, we are going to extract some features from the image by performing the feature extraction step, which is nothing but reducing the dimensionality of

the image by extracting the meaningful features without losing any information. Now, at last, we are going to apply the CNN model to analyze our images and after it has been trained then it can classify the various species into different classes. As defined before, CNN is a deep learning algorithm that can take an input image, and differentiate them based on various aspects or objects in the image.

- **Bird Description Database** - Image Classification requires a large dataset of images with proper distinction among different classes. These classes are features of the individual images based on which they can be classified. The dataset used by us is Caltech UCSD birds-200, featured in 2011. It is the successor of the CUB 200 dataset and contains almost double the number of images per class. The CUB 200 dataset featured in 2010 has around 6000 images of 200 bird species whereas the data set which we used has 11,788 images of 200 bird species with a rate of 60 images per specie.

  Image organization in the dataset is done in form of subdirectories based on their species. Each image has 15 part locations and almost 322 binaryattribute labels. The dataset contains all the North American species of birds.

  We used this dataset because the size and the number of images in the dataset is optimal for our development as it contains a large number of images so that the model could be trained to provide good accuracy also it is not very large that will require high computing resource and will be very time consuming to train this model.

- **Desktop Application -** In this module of our project, we have developed a desktop application that takes input from the user and displays the output to the user on its screen. We have used the python's PyQt5 library for developing the user interface of the application. The IDE we used for development is PyCharm.

  PyCharm: PyCharm is a dedicated Python Integrated Development Environment (IDE). It provides a wide range of essential tools for Python developers.

  PyQT5: PyQt5 is a cross-platform GUI toolkit, a set of python bindings for QT version 5. It can be used to develop interactive desktop application with ease because of its tools and simplicity provided by this library. The UI of this application is made by the QtDesigner which is a tool provided by PyQt5 for designing front-end of the application faster and with ease.

5. RESULT ANALYSIS AND DISCUSSION

Training a new image classification model from scratch is a very difficult and very time taking task. It can take lots of time for configuring the different layers of the model and arranging these layers in such a way that they give us the highest accuracy possible. Also, it needs a lot of resources with lots of data for the training and testing purposes of the model. Therefore, it is difficult for every person to make his/her artificial intelligence model for every problem.

A solution to this problem is provided with the use of the transfer learning technique. In this transfer learning, a pre-trained model is applied to the current problem for classification purposes. A model is trained on some other problem that is similar to the problem which has to be solved and then that pre-trained model is applied to the problem. This helps in boosting the accuracy of the model and making it fit for the classification of images. Also, not much time and resources are needed for this type of training.

But choosing the right pre-trained model for image classification is a difficult task, just because every model can classify images into different classes but with different losses and different accuracies. Therefore to get the best fit for our project we tried and tested three different models.

The models we used are VGGNet, ResNet, and InceptionV3 models. These models are all trained on the ImageNet dataset and performed well in the ImageNet Large Scale Visual Recognition Challenge or ILSVRC. The dataset on which these are trained contained 1.2 million images of different objects and these were to be categorized into 1000 different classes. These are validated and tested on 50,000 and 100,000 images respectively. So we tried all these models separately on our dataset and used the one which helped us in scoring the best accuracy.

a) **Image Preprocessing -** Before feeding the data to the algorithms, we have to preprocess the input, so that we can get more accuracy and the model can also gain information from the data. We applied different preprocessing to test and train the dataset.

b) **Result Analysis of Different Models -** We practiced three different models for the training and testing purpose of our classification problem, how to control model complexity for each of them and what their advantages and disadvantages are. After finding the accuracies of all the three models we then chose the model which helped us in achieving the highest top-1 accuracy. So, we have to predict 200 different classes or species of birds with 11,788

images. We divided these 11,788 images into train and test folders. We trained our model on 5,990 images and then used 5,790 images for the testing purpose of our model.

1)VGGNet - VGGNet is a convolutional neural network structure proposed by Karen Simon and Andrew Zisserman in 2014. We used VGG-16 structures for partition purposes. Input to VGG based on conversion is a 224 * 224 RGB image. To preview the layer, here, we take a picture in RGB format with pixel values in the range 0–255 and subtract the average values, calculated throughout the ImageNet training set, from it.

The average values of the VGGNet algorithm are [0.485, 0.456, 0.406]. These values are used in 3 different image channels. These input images after pre-processing exceed the weight layers. There are a total of 13 flexible layers and 3 layers fully integrated into the VGG16 architecture. VGG has smaller filters (3 * 3) with more depth instead of larger filters. It ends up with the same reception field as if you had a single layer of 7 x 7 flexibility. uploading a model to our file is done below.

```
[8] model_ft=models.vgg16(pretrained=True)
    num_ftrs=model_ft.classifier._modules['6']
    model_ft.classifier._modules['6']=nn.Linear(4096,200)

    if use_gpu:
      model_ft=model_ft.cuda()

    criterion=nn.CrossEntropyLoss()

    optimizer_ft=optim.SGD(model_ft.parameters(), lr=0.0001, momentum=0.9)
```

The pre-trained VGG-16 is downloaded and since it is trained on the ImageNet dataset which contains 1000 output classes its output layer is configured according to those classes. So we manipulated the last layer of the vgg16 network according to our needs.

We configured its output such as that it can classify the images in only 200 classes. The loss criterion that we use for training our model is 'CrossEntropyLoss' as it combines LogSoftmax and NLLLoss in one single class. It is useful in classification problems. We have an SGD optimizer for our model with a learning rate of 0.0001 and a momentum of 0.9.

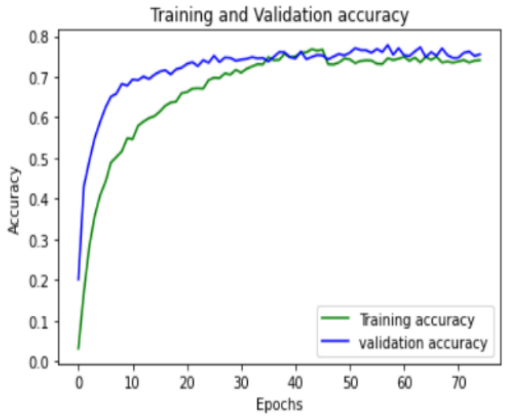

Fig.3

The above graph depicts the accuracy of the model with every epoch while the model is trained and tested after every epoch.

The top 1 and top 5 accuracies of the model are :

Top 1 – 76.12
Top 5 – 93.81

2) ResNet-18 - The full form of ResNet is Residual Network which means using skip connections or shortcuts. ResNet -18 is a CNN model that makes use of skip connections or shortcuts to jump over some layers and using these shortcuts helps it to build deeper neural networks. The 18 in the name represents the number of layers it contains, which means it is 18 layers deep neural network. ResNet solves a major problem in the neural

network field which is with the increase in network depth of the model the accuracy of the model is decreased and a degradation problem is exposed. ResNet solves this problem by using shortcuts to jump over some layers. It is also trained on the ImageNet dataset with 1000 classes. So we have to manipulate its last layer as well which is shown in the figure below:

```
[ ] model_ft=models.resnet18(pretrained=True)

    num_ftrs=model_ft.fc.in_features
    model_ft.fc=nn.Linear(num_ftrs,200)

    if use_gpu:
      model_ft=model_ft.cuda()

    criterion=nn.CrossEntropyLoss()

    optimizer_ft=optim.SGD(model_ft.parameters(), lr=0.0001, momentum=0.9)
```

As we can see in the figure, the last of the network is changed into a linear neural network with 200 classes. The loss we used here is CrossEntropyLoss with SGD optimizer which is configured with a momentum of 0.9 and with a learning rate of 0.0001. After training it for 75 epochs we got the following results

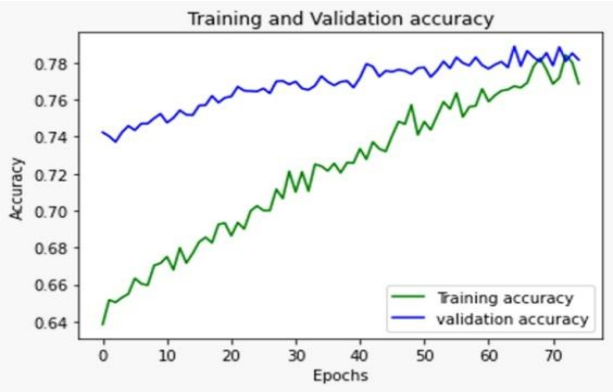

Fig.4

The top 1 and top 5 accuracies of the model are given below :

Top 1 – 79.23
Top 5 – 95.18

3) Inception V3 - Inception V3 is based on the original paper: "Rethinking the Inception Architecture for Computer Vision" by Szegedy, et. al. The model itself is made up of symmetric and asymmetric building blocks, including max pooling, average pooling, concatenations, dropouts, and fully connected layers. Batch normalization is used throughout the model and is applied in the activation inputs. The loss is computed via the Softmax function. In the Inception module, the network is a bit "wider" than "deeper". Filter with multiple sizes operates on the same level which makes the network a bit wider. The model is downloaded and manipulated in a below-mentioned way:

```
model_ft=models.inception_v3(pretrained=True)

num_ftrs=model_ft.AuxLogits.fc.in_features
model_ft.AuxLogits.fc=nn.Linear(num_ftrs,200)

num_ftrs=model_ft.fc.in_features
model_ft.fc=nn.Linear(num_ftrs,200)

if use_gpu:
  model_ft=model_ft.cuda()

criterion=nn.CrossEntropyLoss()

optimizer_ft=optim.SGD(model_ft.parameters(), lr=0.0001, momentum=0.9)
```

The model is pretrained and we the last layer is changed as per our needs.

After training the model for fifty epochs, we get the following results.

Fig.5

The top 1 and top 5 accuracies of the model are given below :

Top 1 – 77.35
Top 5 – 94.35

6. CONCLUSION

The application we made uses the concept of deep learning algorithms to address the problem of identification of bird species. It can classify the bird species only by using the images of the birds, and it uses those images to extract the features from the images. These features are then used by the classifier to predict the classes, or in our case species, of the images. The outcome evidently reveals that the techniques we used in our work reduces the time and space in prediction of the bird species.

We proposed an application that is capable of identifying different bird species. This type of application can be very useful for amateur bird watchers, who are starting in the field of bird watching and the researchers who study birds. Also, this application can be useful for environmentalists, as they use birds as an environmental indicator.

For developing such application we used Caltech UCSD Birds 200 dataset for the training and testing purpose of our model. We trained our model on Google's Colab platform using PyTorch library and used transfer learning to boost the accuracy of our model. Transfer learning technique helped us in achieving much higher

accuracy than we imagined. We chose pre-trained ResNet-18 architecture, after trying three different models, with transfer learning to boost the accuracy of our model. The model we trained using ResNet-18 was able to achieve an accuracy of 78.86 percent.

Also, in order to make our application lightweight, so that user does not have to download our whole model, we have uploaded our model and details of various birds on the cloud platform Heroku as an API. This helped us in reducing the size of our application by tremendous amounts. Also, these APIs can be used by any application or model from anywhere. We make requests to the APIs and it returns us the JSON objects in return which contain the complete information that we need.

7. FUTURE SCOPE

- **Improving Performance**:
  The performance of our bird species classifier can be improved by using some more advanced algorithm.

- **Adding Species:**
  Currently, we are able to classify only 200 species of birds which can be increased so that the model can classify more species.

- **Adding More Data:**
  The accuracy increases with the size of the dataset. Therefore,if we have more images per class accuracy will increase.

- **Making an IoS/Android Application:**
  An android or an IoS application can be made using the model in the future so that it can be much easier to carry and more people can have access to it easily.

8. REFERENCES

[1] Ciregan, Dan, Ueli Meier, and Jürgen Schmidhuber. "Multicolumn deep neural networks for image classification." In 2012 IEEE conference on computer vision and pattern recognition, pp. 3642-3649. IEEE, 2012.

[2] Branson, Steve, Grant Van Horn, Serge Belongie, and Pietro Perona. "Bird species categorization using pose normalized deep convolutional nets." arXiv preprint arXiv:1406.2952 (2014).

[3] Alter, Anne L., and Karen M. Wang. "An Exploration of Computer Vision Techniques for Bird Species Classification." (2017).

[4] Krause, J., Jin, H., Yang, J., and Fei-Fei, L. (2015). Finegrained recognition without part annotations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5546–5555.

[5] Gavves, E., Fernando, B., Snoek, C. G., Smeulders, A. W., and Tuytelaars, T. (2015). Local alignments for fine-grained categorization. International Journal of Computer Vision, 111(2):191–212.

[6] P. Somervuo, A. Harm¨ a and S. Fagerlund, "Parametric Representations ¨ of Bird Sounds for Automatic Species Recognition", IEEE Trans. Audio, Speech, Lang. Process., Vol.14, No.6, pp.2252–2263, November 2006.

[7] E. Vilches, I.A. Escolbar, E.E. Vallejo, and C.E. Taylor, "Data Mining Applied to Acoustic Bird Species Recognition", IEEE Int. Conf. on Patt. Recog., Hong Kong, China, pp.400–403, 2006.

[8] S. Fagerlund, "Bird Species Recognition Using Support Vector Machines", EUSASIP J. Adv. Signal Process., Vol.2007, pp.1– 8, 2007.

[9] Marini, Andreia & Facon, Jacques & Koerich, Alessandro. (2013). Bird Species Classification Based on Color Features. Proceedings - 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013. 4336-4341.

[10] O'Shea, Keiron & Nash, Ryan. (2015). 'An Introduction to Convolutional Neural Networks'. ArXiv e-prints.

[11] Hubel, David H., and Torsten N. Wiesel. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." The Journal of physiology 160, no. 1 (1962): 106

[12] N. Zhang, J. Donahue, R. Girshick, and T. Darrell. Partbased r-cnns for finegrained category detection. In Computer Vision–ECCV 2014, pages 834–849. Springer, 2014.

[13] T.-Y. Lin, A. RoyChowdhury, and S. Maji. Bilinear cnn models for fine grained visual recognition. In Proceedings of the IEEE International Conference on Computer Vision, pages 1449–1457, 2015.

[14]Ümit BUDAK, Abdulkadir ŞENGÜR, Aslı Başak DABAK and Musa ÇIBUK, Transfer Learning-Based Object Detection and Effect of Majority Voting on Classification Performance, 2019 International Artificial Intelligence and Data Processing Symposium (IDAP)

[15]. Wang, C., & Peng, Z. (2019). Design and Implementation of an Object Detection System Using Faster R-CNN. 2019 International Conference on Robots & Intelligent System (ICRIS). DOI:10.1109/icris.2019.00060

[16] Qian, L., Fu, Y., & Liu, T. (2018). An Efficient Model Compression Method for CNN Based Object Detection. 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS). DOI:10.1109/icsess.2018.8663809.

[17] Zhang, W., Li, J., & Qi, S. (2018). Object Detection in Aerial Images Based on Cascaded CNN. 2018 International Conference on Sensor Networks and Signal Processing (SNSP). DOI:10.1109/snsp.2018.00088

[18] Parintorn Pooyoi; Punyanuch Borwarnginn; Jason H. Haga; Worapan Kusakunniran…Snow Scene Segmentation Using CNN-Based Approach With Transfer Learning, 2019 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON).

[19] Wendi Cai; Jiadie Li; Zhongzhao Xie; Tao Zhao; Kang LU, Street Object Detection Based on Faster R-CNN…2018 37th Chinese Control Conference (CCC)… 10.23919/ChiCC.2018.8482613

[20] Hideaki Yanagisawa; Takuro Yamashita; Hiroshi Watanabe….A study on object detection method from manga images using CNN…. 2018 International Workshop on Advanced Image Technology (IWAIT)

[21] Yiming Zhang; Xiangyun Xiao; Xubo Yang…..Real-Time Object Detection for 360-Degree Panoramic Image Using CNN… 2017 International Conference on Virtual Reality and Visualization (ICVRV)