Identification of Bird Species Using a Deep Learning Technology

Aleena Varghese[1]
MTech Scholar, Dept. of computer science and engineering
Sahrdaya College of Engineering and Technology
Thrissur, India

Shyamkrishna K.[2]
Assistant Prof., Dept. of computer science and engineering
Sahrdaya College of Engineering and Technology
Thrissur, India

Dr. Rajeswari M.[3]
Prof., Dept. of computer science and engineering Sahrdaya College of Engineering and Technology Thrissur, India

***Abstract*—Numerous human beings go to bird sanctuaries to get a relief from stress by seeing various birds and enjoying the beauty of colours and traits of the birds. For ensuring biodiversity, species information is a fundamental factor. In many scenarios, it is very difficult to identify the species of the birds due to the similarities existing in between the intraclass and interclass varieties of bird species. Recently, with the help of convolutional neural network (CNN), many state-of-the art algorithms on image classification achieved remarkable successes. This paper has been discussed to develop a deep learning platform to recognize the bird species by its image. Convolutional neural network is used for both the classification and prediction processes. In order to improve the feature extraction, a skip connection oriented neural network model is being introduced. The proposed method achieved 99.00% of classification accuracy for the training image. By using the proposed method, the amateur bird watchers can easily identify the bird species from the captured bird image.**

***Index Terms*—Convolutional neural network, Bird species recognition, Image classification, Deep learning**

I. INTRODUCTION

Bird watching is one of the recreational activity that pro- vides more relaxation and enjoyment to the minds of human beings. Identification of bird species becomes challenging task due to the intraclass and interclass similarities existing in between them. Based on the physical characteristics, color, and shape the bird species are categorized into different classes. Due to the observer constraints such as location, distance, and equipment used to identify birds, recognizing birds with the naked eye of human being is based on basic characteristic features, and appropriate classification based on distinct features is often seen as tedious[1]. Nowadays a number of techniques are available to identify the species of birds. Deep learning is one of the emerging technology that can be used to recognize the birds. The convolutional neural network (CNN) is a category of deep learning neural networks. Convolutional neural networks constitute a big step forward in image recognition. They're maximum normally used to investigate visual imagery and are regularly operating behind the scene in image classification.

In the past, computer vision [2], [3] and its subcategory of recognition, which use strategies along with machine learning, had been notably researched to delineate the specific features of objects, consisting of veggies and fruits [4], landmarks [5], clothing [6], cars [7], plants [8], and birds [9], inside a selected cluster of scenes. The capability of convolutional neural networks to extract various features from captured im- age is utilized in many scenarios. In the past years, bird sound classification has received attention increasingly. Therefore, it is becoming ever more necessary to protect bird biodiversity, where monitoring bird population is the first step for the protection[10]. CNN is mainly designed to recognize visual features from images and it requires a minimum level of preprocessing.

In addition, the exponentially expanded amount of online data has gotten to be less demanding to gather as the learning information for the neural networks, and the refined infor- mation has been easily shared for the convolutional neural network learning. These intuitive driven to a convolutional neural network approach beyond the capabilities of the existing approaches. The convolutional neural network has become one of the leading architecture for most of the image recognition, classification, and prediction processes. In image recognition, video analysis, natural language processing, and drug discov- ery applications the convolutional neural network can be used for better performances. And the CNNs performance rates are progressing yearly.

In this proposed work we are making use of CNNs in the field of bird image recognition. From the captured image of the bird the CNN model could identify the species of that particu- lar bird and produces prediction results accordingly. The main objectives of the proposed work are to develop a deep learning model by making use of train and test colored images of birds in order to identify/ classify the bird species into particular classes of its species according to the classification results

Fig. 1. Caltech-UCSD Birds 200 dataset [26].

obtained from the skip connection oriented CNN model.

The rest of this paper is structured as follows. Section II briey reviews related approaches on bird species identification using convolutional neural networks. Section III describes about the dataset used for the processing. Section IV centered on the proposed bird species identification deep learning model and its features. The experimental results and analysis of the datasets are presented in Section V. Section VI provides conclusions and directions for the future study on bird species identification.

II. Related Work

The earlier approaches for the bird species identification used bird songs to identify the birds. The visual features

i.e. SIFT (Scale invariant feature transform) [11] from bird images and acoustic features both were used in the classical Machine learning algorithms to train a standard support vector machine(SVM) for classification. The fine-grained visual cate- gorization methods [12] have shown great and better results of image classification and within computer vision research they have become a promising approach. For the generic object recognition numerous techniques have been applied [13]. A few strategies applied local part learning that uses deformable part models and region-CNN for object recognition [14], bounding box generation, and distinctive parts selection for image recognition.

Some researches have centered on discriminative features based on the local traits of bird species [15], [17]. Simulta- neous detection and segmentation are used to localize score detections effectively [16]. Pose-normalization and model en- sembles [17] are also utilized to progress the execution of fine-grained detection by producing millions of key point sets through fully convolutional search. There are lots of strategies are available to distinguish the birds but they are costly and time consuming. There has been an increasing interest for automated acoustic monitoring of sound-emitting creatures, which may give reliable information on the presence/ absence of target species and on the common biodiversity status of an area in recent a long time[18].

Jie Xie et al. [10], investigated three types of time frequency representations (TFRs) such as Mel-spectrogram, harmonic component based spectrogram, and percussive component based spectrogram of bird sounds to characterize the differ- ent acoustic components of birds to identify particular bird species. Stavros Ntalampiras et al. [18], statistically analysed the similarities in between the audio signal of the bird and music genres rather than looking at the bird's audio signal alone. And for that they utilized the transfer learning tech- nology. Loris Nanni et al. [19], presented a combination of classifiers including AlexNet [20], GoogleNet [21], VGGNet [22], ResNet, and InceptionV3 to identify the bird species by processing its audio signal. The audio images such as spectrograms, ScatNet [23,24] scattering representations, and harmonic and percussion images are extracted from the bird's audio signal for the classification and prediction processes. In all these cases there might be a probability of occurring background noise such as environmental noise, sounds of insects while recording the bird's song. Thus it leads to the misclassification and reduces the probability to get accurate prediction results. The similarities existing among the songs of birds also provides inaccurate predictions about the species of the bird.

Jiaohua Qin et al. [25], replaced fixed size images with appropriately large size images as input to convolutional neural

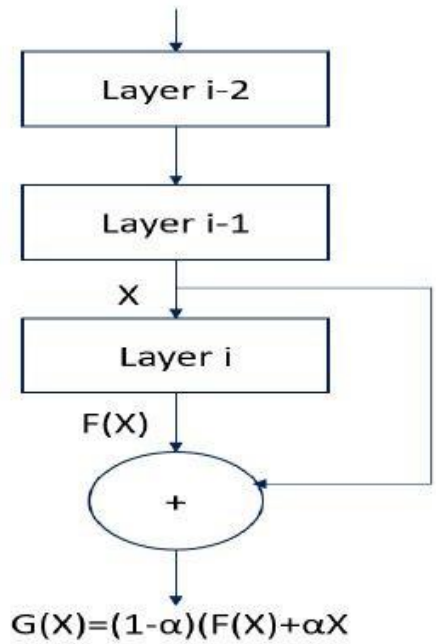

Fig. 2. Framework of skip connections [1].

network and few modules in that were replaced with an Inverted Residual Block module inorder to reduce the network parameters and computational cost . The convolution layer, Batch Normalization layer, ReLU activation function, Global Average Pooling, and Softmax were included to form this improved network. But they haven't specialized any animal dataset or any object dataset for the processes and focused mainly on general biological images. Plenty of techniques are available to classify the biological images into different categories and few of them are focused more on species identification. This proposed work also approached the species identification using convolutional neural network architecture.

III. DATASET

For the bird image recognition, it is required to have a solid dataset on which the identification system can be trained, tested, and validated. Thus we used one of the fine- grained biological image classification dataset Caltech-UCSD Birds 200–2011[26]. Caltech and UCSD have gathered data to produce this particular dataset and it is the extended version of the CUB-200 dataset. Figure 1 shows the Caltech-UCSD Birds 200 dataset. The dataset contains 11,788 images of 200 different categories of bird species. The dataset was splitted into training set, testing set, and validation set. It is very important to keep the testing set completely separate from the training set since it needed to be sure that the classification model will perform well in the real world scenarios. The pixel values are normalized in order to reduce the harshness, noise and disturbances in the images. Then it can be used for training the classification model. More than 60% of data allocated to the training set and rest of the data allocated to the testing set and validation set. Training set and validation set are randomly selected from dataset for the fine-tuning process.

IV. BIRD SPECIES IDENTIFICATION DEEP LEARNING FRAMEWORK

As of late, deep learning models have become the foremost well known tool for articial intelligence [27] and big data analysis. The rise of deep learning [28] algorithms has resulted in exceedingly complex cognitive tasks for computer vision and image recognition. The proposed deep learning model acquired to build this bird image classication system using the CNN framework is described as follows.

A. Building the CNN

The most popular neural network model being used for im- age classification problem is Convolutional Neural Networks (CNNs) . Local understanding of an image is good enough is the great idea behind CNNs. Having fewer parameters significantly moves forward the time it takes to learn as well as reduces the amount of data required to train the model is one of the practical advantage. The CNN model conguration for bird species identication utilized a stack of convolutional layers comprising an input layer, two Fully connected(FC) layers, and one final output softmax layer. Convolutional layers apply a convolution operation to the input and this passes the resulting information on to the next layer. Each convolutional layer comprised (a) 5*5 convolution, (b) Batch normalization(BN): It is a technique for training very deep neural networks that standardizes the inputs to a layer for each mini-batch. This has the impact of stabilizing the learning process and significantly reducing the number of training epochs required to train deep

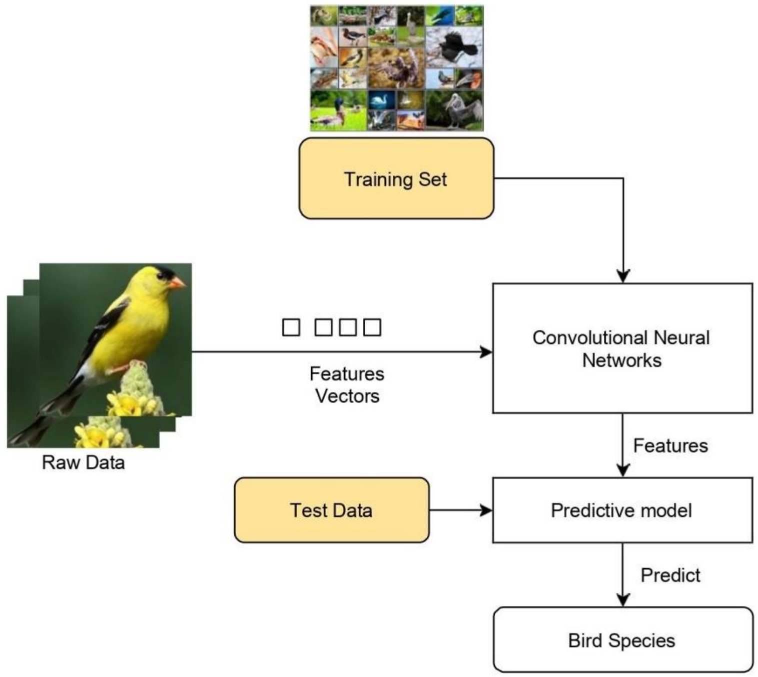

Fig. 3. Feature extraction paradigm.

networks, (c) Rectified linear unit(ReLU) activation: It is a piecewise linear function. If it is positive, then produces the input directly, otherwise, it will output zero, and (d) pooling layers: By combining the outputs of neuron clusters at one layer into a single neuron in the next layer the pooling layers reduces the dimensions of the data.

B. Skip Connection

In numerous convolutional architectures, skip connection is a standard module. Skip connection offers an alternative path for the gradient. It is tentatively approved that these extra ways are regularly advantageous for the model convergence. As the name suggests, these skip connections skip few layers within the neural network and feeds the output of one layer as the input to the next layers (instead of only the next one) in deep architectures. A transition from general to specific must have occurred somewhere in the network if first-layer features are general and last-layer features are specific [29]. We proposed adding skip connections [30] among corresponding convolu- tional layers, as shown in Figure 2 to address and quantify the degree to which a particular layer is general or specific. The skip layer connections should improve feature extraction through weighted summation of corresponding layers as follows[1]:

$$G(X) = (1 - \alpha)F(X) + \alpha X \quad (1)$$

where X is the input, G(X) is a linear combination of F(X) and X, F(X) is a function of input X, and α is a weight in the unit interval [0,1]. Result from the previous layer contributes less to overall performance than the layers preceding it if the weight α is greater than 0.5. And the result from the previous layer contributes more to the overall performance if the weight α is lesser than 0.5. By using the skip connections we can enable feature reusability and stabilize training and convergence and we can facilitate network training too.

C. Feature Extraction

This subsection describes the feature extraction process in identification of the bird species using bird images. Primary task is to extract features from raw input images, when extracting relevant and descriptive information for fine-grained object recognition [31][32]. To extract and learn features, CNNs apply a number of filters to the raw pixel data of an image. The result of applying the filters to an input image will be captured by the feature maps of a CNN. The feature map is the output of each layer. Since numerical representation facilitate processing and statistical analysis, various algorithms in machine learning require a numerical representation of objects. For both the classification and prediction processes we utilized the convolutional neural network.

In the feature extraction process, the feature vectors ex- tracted from the raw data (Image of a bird) is given to the CNN which is trained with the training dataset. The extracted features are then passed to the predictive model which compares the features with the test data. Then the model predicts the species of that particular bird in the image

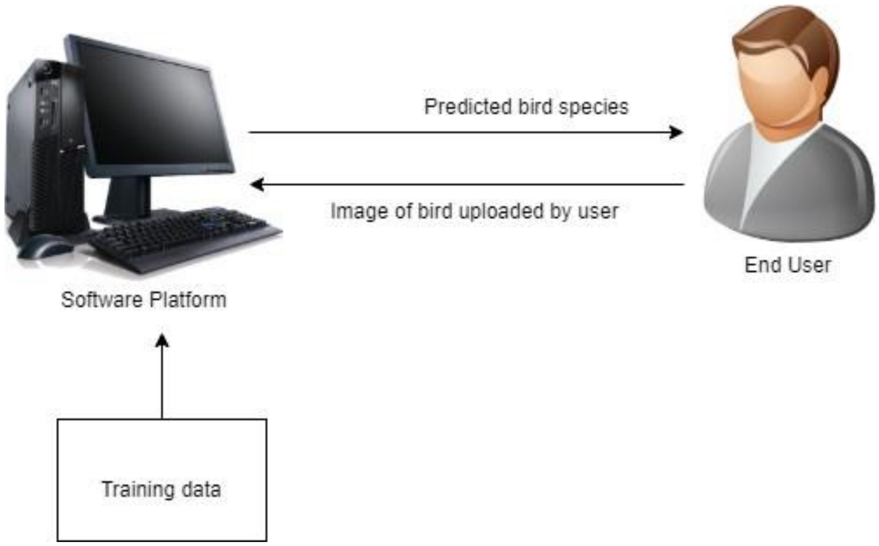

Fig. 4. Client-server architecture.

which is given as input. This feature extraction paradigm is illustrated in Figure 3. Using ReLU 5 and ReLU 6 the features were calculated. In order to find the object parts defined by bounding box coordinates and their dimensions (width and height) in the image, localization was used.

D. *System Implementation*

This subsection explains the implementation of the proposed deep learning platform to identify the species of birds. The image of the bird is given as the input to the system. The software system has a trained model with a set of data. The image of the bird gets converted into gray scale and then into matrix format. Various alignments from that image will be taken into consideration and each alignment will be given to the CNN for feature extraction. These extracted features are given to the CNN with trained model and then the resulting values gets compared. According to the compared values the classifier classifies the image into different categories. Then the predictive model predicts the species of that particular bird which is considered to be as the final result. The output layer of the network provides parts of the input image containing the bird.

In order to complete this task of species identification, a client-server architecture is established. An interface to allow a computer user to request services of the server and to display the results the server returns are provided by the client computers. A function or service to one or many clients, which initiate requests for such services are provided by the server component. The client-server architecture is shown in Figure 4. The server takes prediction requests for bird images from end-user device. If the image is a non-bird image then the system shows an error message to the end-user. If it is a bird image, then process it in the system using the proposed deep learning model. After an image has been predicted, the result send back to the end-user.

V. Results and Discussion

This section explains the details about the experimental results of the system used to identify the bird species. The bird and non-bird images can be differentiated and predicted using this proposed deep learning framework. The bird species identification system sends an error notification to upload only the images that contains a bird, when non-bird images are uploaded to the identification system. 100 bird images were uploaded from a mobile phone for preliminary testing in order to validate the effectiveness and efficiency of the proposed bird species identification system and to filter non- bird images uploaded to the system automatically. For the differentiation and classification of the images as true bird images, the proposed model achieved 100 percentage accuracy. The bird detection results are shown in Table 1.

TABLE I

. Prediction results of images uploaded from an end-user device.

Subjects	Predicted as Bird Image	Predicted as Non-bird Image
Image containing bird	100 %	0 %
Image containing non-bird	0 %	0 %

Compared the performance of the proposed framework with other two models such as convolutional neural network which is not oriented on skip connection method and Support vector machine(SVM). All the three models are compared using the same set of data. The learning rate of 0.00001 and 100 epochs are set for the performance comparison of the model. Used a linear kernel for the high dimensionality of the feature space for the support vector machine. The proposed CNN model which is oriented on skip connections achieved higher accuracy than the CNN model which is not oriented on skip connections, and SVM. This validated the effectiveness and efficiency of the presented model which is oriented on skip connections than the other existing models used to identify the bird species using images of the bird.

We developed an automatic deep learning model to classify the bird species by skip connection oriented CNN model in this study about species identification. Performed an empir- ical study by the skip connection architecture to evaluate the efficiency of using this architecture with the CNN. We used this skip connection along with the convolutional neural network to resolve the vanishing gradient problem. However, in this study, we are more focused on predicting the different species of birds more effcient and effective. The proposed skip connection oriented CNN model can predict the uploaded image of a bird as bird with 100% accuracy.

VI. CONCLUSION

The capabilities of deep learning can be utilized in the biological image classification. In this study, we developed a deep learning platform to identify the different bird species using the images of the birds. For this identification task, we have focused on the discriminative features of the birds such as belly, beak, and eyestripe. With an overall accuracy of 98.70% for the training dataset, the proposed bird species identification system could detect and differentiate uploaded images as birds. This skip connection oriented CNN model helps building applications that helps birdwatchers to identify the bird species by just capturing a picture of a bird and uploading it as input to this model. Numerous methods are available for biological image classification, but most of them are not specialized on the bird species identification. Thus, we specialized on the bird species identification and used images of birds to identify the bird species. From the experimental results it is clear that the proposed model outperforms other models which takes images of birds as input to the system.

REFERENCES

[1] Yo-Ping Huang and Haobijam Basanta, “Bird Image Retrieval and Recognition Using a Deep Learning Platform”, IEEE Access, Vol. 7, May 2019.

[2] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei- Fei, “ImageNet large scale visual recognition challenge,” Int. J. Comput. Vis., vol. 115, no. 3, pp. 211252, Dec. 2015.

[3] H. Yao, S. Zhang, Y. Zhang, J. Li, and Q. Tian, “Coarse-to-ne description for ne-grained visual categorization,” IEEE Trans. Image Process., vol. 25, no. 10, pp. 48584872, Oct. 2016.

[4] F. Garcia, J. Cervantes, A. Lopez, and M. Alvarado, “Fruit classication by extracting color chromaticity, shape and texture features: Towards an application for supermarkets,” IEEE Latin Amer. Trans., vol. 14, no. 7, pp. 34343443, Jul. 2016.

[5] L. Zhu, J. Shen, H. Jin, L. Xie, and R. Zheng, “Landmark classication with hierarchical multi-modal exemplar feature,” IEEE Trans. Multime- dia, vol. 17, no. 7, pp. 981993, Jul. 2015.

[6] X. Liang, L. Lin, W. Yang, P. Luo, J. Huang, and S. Yan, “Clothes co-parsing via joint image segmentation and labeling with application to clothing retrieval,” IEEE Trans. Multimedia, vol. 18, no. 6, pp. 11751186, Jun. 2016.

[7] Y.-P. Huang, L. Sithole, and T.-T. Lee, “Structure from motion technique for scene detection using autonomous drone navigation,” IEEE Trans. Syst., Man, Cybern., Syst., to be published.

[8] C. McCool, I. Sa, F. Dayoub, C. Lehnert, T. Perez, and B. Upcroft, “Visual detection of occluded crop: For automated harvesting,” in Proc. Int. Conf. Robot. Autom. (ICRA), Stockholm, Sweden, May 2016, pp. 25062512.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classication with deep convolutional neural networks,” in Proc. 25th Int. Conf. Advance Neural Inf. Process. Syst., Lake Tahoe, NV, USA, Dec. 2012, pp. 10971105.

[10] Jie Xie, Kai Hu, Mingying Zhu, Jinghu Yu, Qibing Zhu, “Investigation of Different CNN-Based Models for Improved Bird Sound Classification”, IEEE Access, vol. 7, pp. 175353-175361, 2019.

[11] Marini, A., Turatti, A. J., Britto, A. S., Koerich, A. L. (2015). Visual and acoustic identification of bird species. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2015).

[12] “Bird Species Categorization Using Pose Normalized Deep Convolu- tional Net” Steve Branson, Grant Van Horn, Serge Belongie ,Pietro Peron (2015).

[13] Li Liu, W. Ouyang, X. Wang, P. Fieguth, X. Liu, and M. Pietikäinen, “Deep learning for generic object detection: A survey,” Sep. 2018, arXiv:1809.02165. [Online]. Available: https://arxiv.org/abs/1809.02165

[14] K. Dhindsa, K. D. Gauder, K. A. Marszalek, B. Terpou, and S. Becker, “Progressive thresholding: Shaping and specicity in automated neurofeedback training,” IEEE IEEE Trans. Neural Syst. Rehabil. Eng., vol. 26, no. 12, pp. 22972305, Dec. 2018.

[15] C.-Y. Lee, A. Bhardwaj,W. Di, V. Jagadeesh, and R. Piramuthu, “Region-based discriminative feature pooling for scene text recognition,” in Proc. Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 40504057.

[16] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, “Simultaneous detection and segmentation,” in Proc. Eur. Conf. Comput. Vis., Jul. 2014, pp. 297312.

[17] S. Branson, G. V. Horn, S. Belongie, and P. Perona, “Bird species categorization using pose normalized deep convolutional nets,” in. Proc. Brit. Mach. Vis. Conf., Nottingham, U.K., Jun. 2014, pp. 114.

[18] Stavros Ntalampiras, “Bird species identification via transfer learning from music genres”, Ecological Informatics, Vol. 44, March 2018.

[19] Loris Nanni, Yandre M. G. Costa, Rafael L. Aguiar, Rafael B. Mangolin, Sheryl Brahnam and Carlos N. Silla Jr., “Ensemble of convolutional neural networks to improve animal audio classification”, EURASIP Journal on Audio, Speech, and Music Processing, Article number: 8(2020), May 2020.

[20] A. Krizhevsky, I. Sutskever, G. E. Hinton, “Imagenet classification with deep convolutional neural networks”, Advances in Neural Information Processing Systems, pp. 1097–1105, 2012.

[21] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, “Going deeper with convolutions”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9, 2015.

[22] K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, arXiv preprint (2014). arXiv:1409.1556.

[23] J. Bruna, S. Mallat, “Invariant scattering convolution networks”, IEEE Trans. Pattern Anal. Mach. Intell. 35(8), 1872–1886 (2013).

[24] L. Sifre, S. Mallat, “Combined scattering for rotation invariant texture analysis”, ESANN, vol. 44, pp. 68–81, 2012.

[25] Jiaohua Qin, Wenyan Pan, Xuyu Xiang, Yun Tan, Guimin Hou, “A biological image classification method based on improved CNN ”, Ecological Informatics, Vol. 58, July 2020.

[26] C. Wah et al. The Caltech-UCSD Birds-200-2011 Dataset. Tech. rep. CNS-TR-2011-001. California Institute of Technology, 2011.

[27] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, M. Ranzato, A. Senior, P. Tucker, K. Yang, Q. V. Le, and A. Y. Ng, “Large scale distributed deep networks,” in Proc. 25th Int. Conf. Adv. Neural Inf. Process. Syst., Lake Tahoe, NV, USA, Dec. 2012, pp. 12231231.

[28] L. Yang, A. M. MacEachren, P. Mitra, and T. Onorati, “Visually-enabled active deep learning for (Geo) text and image classication: A review,” Int. J. Geo-Inf., vol. 7, no. 2, p. 65, Feb. 2018.

[29] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks,” in Proc. Int. Conf. Advance Neural Inf. Process. Syst., Dec. 2014, pp. 33203328.

[30] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 770778.

[31] N. Zhang, J. Donahue, R. Girshick, and T. Darrell, “Part-based R-CNNs for ne-grained category detection,” in Proc. Int. Conf. Eur. Conf. Comput. Vis., Cham, Switzerland, Jul. 2014, pp. 834849.

[32] C. Huang, Z. He, G. Cao, and W. Cao, “Task-driven progressive part localization for ne-grained object recognition,” IEEE Trans. Multimedia, vol. 18, no. 12, pp. 23722383, Dec. 2016.