International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075, Volume-8, Issue-9S, July 2019

CBNWI-50: A Deep Learning Bird Dataset for Image Translation and Resolution Improvement using Generative Adversarial Network

Akanksha Sharma, Neeru Jindal

***Abstract*: Generative Adversarial Networks have gained prominence in a short span of time as they can synthesize images from latent noise by minimizing the adversarial cost function. New variants of GANs have been developed to perform specific tasks using state-of-the-art GAN models, like image translation, single image super resolution, segmentation, classification, style transfer etc. However, a combination of two GANs to perform two different applications in one model has been sparsely explored. Hence, this paper concatenates two GANs and aims to perform Image Translation using Cycle GAN model on bird images and improve their resolution using SRGAN.**

During the extensive survey, it is observed that most of the deep learning databases on Aves were built using the new world species (i.e. species found in North America). Hence, to bridge this gap, a new Ave database, 'Common Birds of North - Western India' (CBNWI-50), is also proposed in this work.

Index Terms*: *Generative Adversarial Networks, Indian-Subcontinent, Bird Dataset, Image Translation, Single Image Super Resolution

I. INTRODUCTION

Generative Adversarial Networks (GANs) were introduced to the world by Goodfellow et al. [1], since then, their popularity has increased exponentially in the field of artificial intelligence and machine learning. The basic idea behind GAN was to produce synthetic images from scratch using a latent noise vector. But over the years, new models of GAN have been developed for various applications, like image translation, single image super-resolution, etc.

In this paper, a novel model for image to image translation and resolution improvement has been proposed. The model uses existing CycleGAN [2] model for image to image translation and Super Resolution GAN [3] for resolution improvement of the translated images by concatenating the two GAN models.

The main contributions of this paper are:

1. A novel model for translation and resolution improvement for bird images using Cycle GAN and Super Resolution GAN.
2. To the best of the author’s knowledge, currently, there is no other deep learning dataset which is based on birds found in the Indian Subcontinent.

The paper has been divided in the following sections: Section II deals with models of Vanilla GAN, CycleGAN and SRGAN. Section III contains the basic block diagram of proposed model. Section IV consist information on database proposed in this paper. Section V presents results and discussion. Section VI proposes conclusion and future scope of this paper.

II. BASIC BUILDING BLOCKS OF GAN

A. Vanilla GAN

The basic block diagram of Generative Adversarial Network is shown in figure 1. It consists of two blocks called generator (G) and discriminator (D). The generator is provided with noise vectors as input and the first batch of fake data is generated. This fake data is then fed to the discriminator along with the training data. The discriminator is a simple classifier, which classifies the samples from the input as either original or fake. This is performed by allocating probabilities to the samples. A probability of ‘1’ means the sample is real, ‘0’ means the sample is forged. The information in form of the gradient is back propagated to generator network. This helps the generator to learn the features of the training dataset and in turn it generates images which match the statistical properties of the original images.
In the next step, both the generated data and the original data are fed to the discriminator, which makes the decision of whether the images are fake or forged, and the learning process goes on. The discriminator approximates the ratio of densities and then passes it to the generator in form of a gradient. The features are learned jointly alternating between the generator and the discriminator. In the beginning, the discriminator wins too easily, but as the training progresses the generator starts producing more realistic images.

B. Cycle GAN

Image to image translation is performed by learning the mapping between images using the training set of image pairs. Zhu et al [2] proposed a new method to translate images from one domain to another when paired samples are missing. A cycle consistency loss was introduced to enforce this as such type of mapping is highly under-constrained. The authors also used inverse mapping. Due to cycle consistency loss in action, this GAN was named CycleGAN, which performed the tasks of object transfiguration, photo

Revised Manuscript Received on July 09, 2019.
Akanksha Sharma, Electronics and Communication Department Thapar Institute of engineering and technology Patiala,India.
Neeru Jindal, Electronics and Communication Department Thapar Institute of engineering and technology Patiala,India

enhancement, style transfer, season transfer, etc. Detailed architecture is shown in figure 5.

Cycle GAN works on the principle of cyclic consistency loss. Cyclic consistency loss works on the principle that if an image 'X' has been converted from domain A to domain B to yield an image 'Y', then further translation of image 'Y' from domain B to domain A must yield the original image 'X'. The earlier models which performed image to image translations [9, 10], needed aligned or paired datasets to produce results, while Cycle GAN model was successful using only unpaired datasets. The model flow diagram operates in a cyclic form and producing improved results after each iteration for both the image domains. Flow diagram of CycleGAN is shown in figure 1.

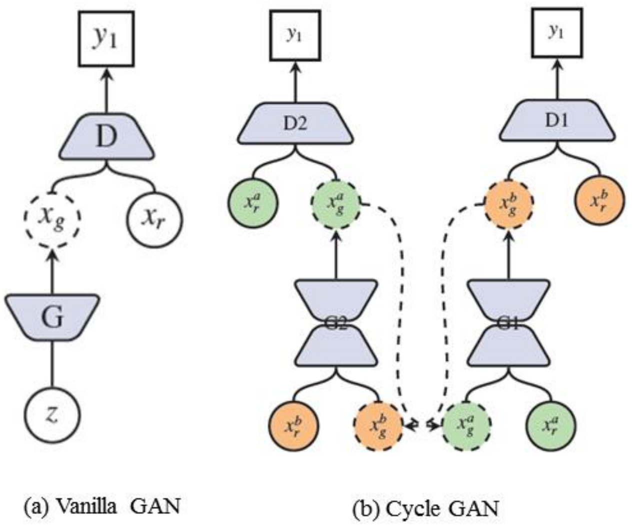

Figure 1 Basic flow diagrams of Vanilla GAN model and CycleGAN Model.

C. Super Resolution GAN

Although convolutional neural networks have been used for image resolution improvement and feature enhancement, the results obtained by GANs are better both quantitatively and qualitatively. Ledig et al. [3] aimed at recovering fine texture details while super resolving at large up-scaling factors. It is capable of inferring photo realistic natural images for generating up-scaling factors. To achieve these results, the authors used a perceptual loss function which consists of an adversarial loss and content loss. The adversarial loss discriminates between original image manifold and super resolved image. Content loss takes into account the perceptual similarity rather than pixel similarity. Content loss played a major role in super resolution, determining that ideal loss function depends on the application. Basic architecture and layer configuration of SRGAN are shown in figure 6.

.

III. PROPOSED MODEL

The proposed model uses Cycle GAN and SRGAN as Stage – I and Stage – II. The basic flow diagram of proposed model is shown in figure 2. The first step of the model is image pre-processing where the images are compressed to JPEG and resized to 143×143 pixels. Images for both the species go through same preprocessing steps. In stage – I, the images are fed to cycle GAN model and are randomly cropped to a size of 128×128 to obtain finer details. The output of the cycle GAN is the translated image which is then fed to SRGAN. In stage – II, a pre-trained model (trained on images from DIV2K and CBNWI-50) of SRGAN performs image super-resolution and improves the resolution between 2× and 8×, according to the desired output image size.

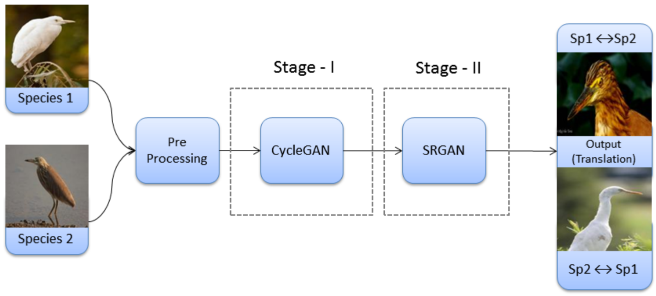

Figure 2 Basic Flow Diagram of the proposed model

IV. DATASETS

A. DIV2K

To train SRGAN model, DIV2K [4] dataset was used. This dataset contains 800 training images and 100 test images of 2K size. As super resolution on birds was to be performed, an additional 500 images of birds of 2K resolution were jumbled with DIV2K dataset images to fine-tune the SRGAN model to produce more texture on bird features like feather patterns and eyes. The additional 500 bird images were taken from CBNWI-50; a bird dataset proposed in this paper and is discussed in next part.

B. CBNWI – 50

Generative adversarial networks require a large number of images to train upon. Thus, large datasets are required. The complexity of dataset also plays a deciding factor for judging an algorithm's performance. Most of the datasets are accompanied with additional information called the annotations. This is complementary information and the data which comes with annotations is basically labeled data. Producing labeled data is a very tedious and grueling manual labor for the researches. Most of the deep learning datasets take years to make with contributions from multiple authors and subordinates. Most datasets contain small pictures with varying sizes and shapes, different illumination, colors, exposure, field of depth, etc. Diversity among the samples within the datasets is a must.

Table 1 Publically available deep learning datasets on Aves.

Name	No. of Species	Total Number of Images	Annotations/ Bounding Boxes
CUB-200 [5]	200	11,788	Yes
iNaturalist [6]	964	214,295	Yes

Birdsnap [7]	500	49,829	Yes
NA Birds [8]	400	48,000	Yes

In order to carry out translations on birds, publically available deep learning datasets, (listed in table 1), were insufficient as they contain very few images per species. Also, the number of images per species varies greatly. Interestingly, none of the above mentioned datasets cover the species found in Indian subcontinent. All the four datasets mentioned above cover only the continents of North America (New World species) and Europe. In order to build a dataset, data was collected in form of photographs of common birds found in North-Western India (predominantly Rajasthan). To increase the number of training images, various attacks were used. Details of the dataset are listed in table 2.

The dataset contains images of 50 bird species including local as well as migratory birds which can be easily sited in the north western states of Rajasthan, Haryana and Punjab. Majority of the data has been collected from the state of Rajasthan (India), within the 150 km radius of Jaipur City between 2016 and 2018. The data collection points include Man Sagar Lake (Jaipur), Chandlai Dam (Jaipur), Gatolav Lake (Dausa), Barkhera Jain Teerth Temple Lake (Jaipur), University of Rajasthan Campus (Jaipur), Thapar Technology Campus (Patiala), Ana Sagar Lake (Ajmer) and Sukhna Lake (Chandhigarh). A collage of some of species from dataset is shown in figure 3. Various attacks applied on the dataset images are shown in figure 4. A species wise list of bird images has been provided in appendix 1 at the end.

Table 2. Features of CBNWI-50

Camera used	Canon Powershot SX60 HS (16 MP)	
	Sony Cybershot DSC-H7 (8 MP)	
Photograph Format	RAW, JPEG	
States Included	Rajasthan, Haryana, Punjab	
Total Number of Species	50	
Total Number of Original Images	5,102	
Labeled Images	Species and attack-wise labeling on all images	
Annotations/BB	No	
Total Number of Images after applying Attacks – 35,714		
Attacks	Type	Parameters
Compression	Bicubic Compression	Image Size - 300×400
Noise Addition	Gaussian Noise	μ=0, σ = 0.009
Image Blurring	Average Filter	Filter Size = 5×5
Rotation	Counter Clock Wise (CCW)	10 Degree
Contrast and Brightness Adjustment	-	Lower Bound = 0.01
		Upper Bound = 0.90
Canvas Flipping	Horizontal Canvas Flip	-

Figure 3 Collage of some of the species from dataset.

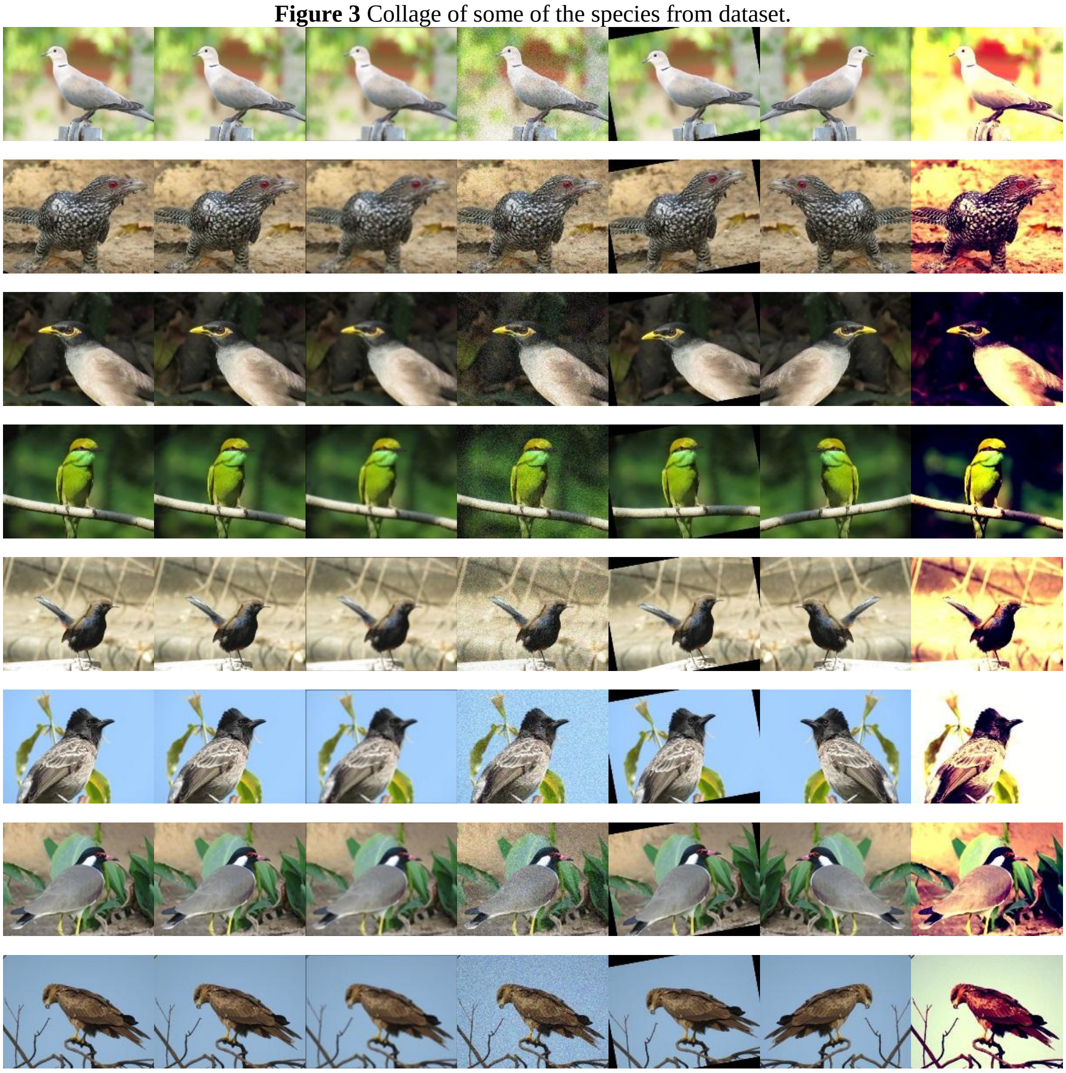

Figure 4 Dataset images after application of attacks. From Left to right (i) Original Image (ii) Compression (iii) Image Blurring (iv) Noise Addition (v) Rotation (vi) Canvas Flipping (vii) Contrast Adjustments

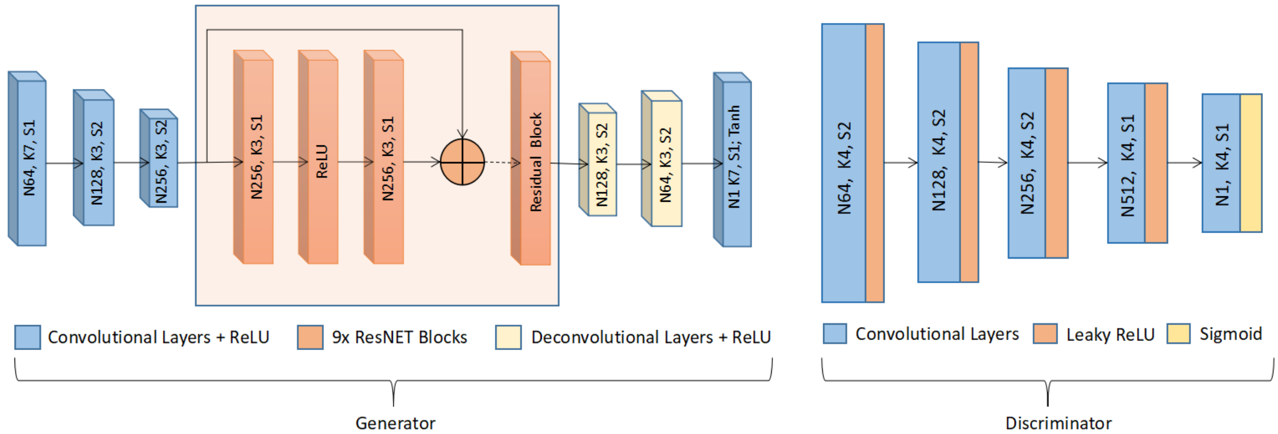

Figure 5 Architecture and layer details of Cycle GAN used in proposed model.

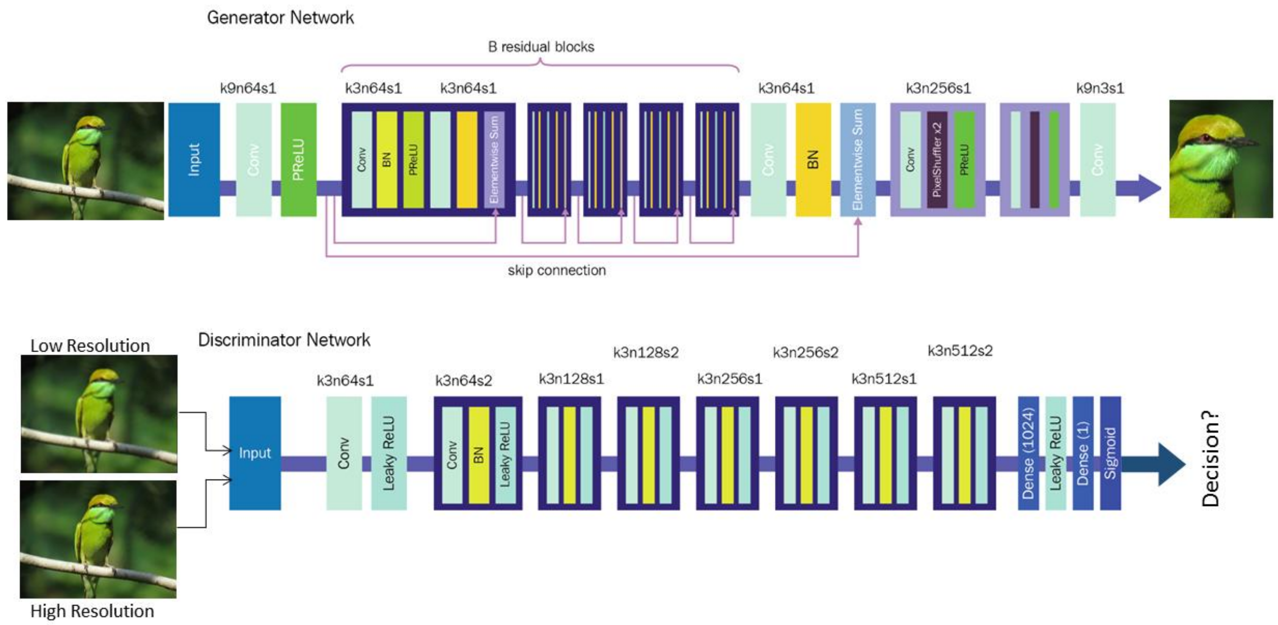

Figure 6 Architecture of SRGAN along with all the layer descriptions with corresponding kernel size (k), number of features (n) and stride (s) in each convolution layer.

V.RESULTS

A.Performance Evaluation Parameters

1) Stage – I

For stage- I, the process of translation was evaluated using human perception skills and in most of the simulations, convincing translated images were obtained.

2) Stage – II

For stage- II, three performance evaluation parameters were used, namely, PSNR (Peak Signal to Noise Ratio), SSIM (Structural Similarity Index) and MSE (Mean Square Error). These parameters were used in base model of SRGAN [3] and hence, the proposed work also used the same parameters for quality assessment of generated images.

MSE provides the mean squared error between the target value and estimated value in an experiment. It is always a positive number and thus, sometimes, it may lead to arbitrary results. MSE is inversely proportional to PSNR (Peak Signal to Noise Ratio). Thus, as MSE decreases, PSNR of the corresponding simulation should increase.

PSNR is the ratio of maximum power of a signal to the power of noise that corrupts it. It is a standard parameter for estimation of lossy image compression codec, where higher PSNR indicates better quality. For image quality estimation, it is calculated in dB by using the following mathematical relation.

While MSE and PSNR estimate absolute errors, **SSIM** measures change in structural information of an image including luminance and contrast. It is measured on a scale of 0 to 1, where a higher value of SSIM indicates that the two images are visually more similar.

B.Training Procedure

1) Stage- I

The images from CBNWI-50 were pre-processed, converted to JPEG and resized to a size of 143×143 using bicubic compression. For species which had less number of images, open source images were used from Flickr to expand the training dataset. The cycleGAN model uses ADAM optimizer with an initial learning rate of $2e^{-4}$. The learning rate is slowly declined to zero after 100 epochs. The training is carried out for a default number of epochs (200) but can be halted if convincing results are obtained earlier than expected.

2) Stage – II

As super resolution on birds was to be performed, an additional 500 images of birds of 2K resolution were jumbled with DIV2K dataset images to fine-tune the SRGAN model to produce more texture to bird features like feather patterns and eyes. The additional 500 bird images were from CBNWI-50

C. Simulation Design

In order to explore the limits to which we can perform style transfer in birds, three different types of simulations on different bird species were carried out. Scientific species classification is described as following; Kingdom (Animalia) > Phylum (Chordata) > Class (Aves) > Order > Family > Genus > Species. First experiment was performed for inter-species translations, second for inter-genus translations and third for inter-family translations. Inter – Order translations were also attempted, but they were not as successful as other translations due to vast morphological differences between birds of different orders.

D.Discussion

Human perception based evaluation is one of the best performance metric for translation tasks. It can be clearly observed that the translation of one bird species to another was successful in most cases. However, it should be noted that translated images for closely related bird species are much better when compared to inter – order translations. To understand the results better, a species classification tree has been presented in figure 7. Better translations are observed when translations are made on species of same genus, i.e. intra – genus (inter – species) translations are more successful. Intra – species translations can also be performed when male and female of the species are visibly different. However, it does not apply to those species where, male and female are of completely different size and features, like a peacock (*Pavo cristatus*, family - Phasianidae) and a peahen. Some intra family translations were also very successful, owing to the visual similarity of the said birds like Cattle Egret (*Bubulcus ibis*; family - Ardeidae) and Indian Pond Heron (*Ardeola grayii*; family - Ardeidae).

In the second stage of the simulation, emphasis was laid on super resolution of the translated image. The quantitative results for this analysis have been listed in table 3. The best results have been highlighted in all the segments. Figure 8 – 19 showcase various translations obtained and Figure 20 – 28 showcase various quantitative analysis graphs obtained for all three simulations.

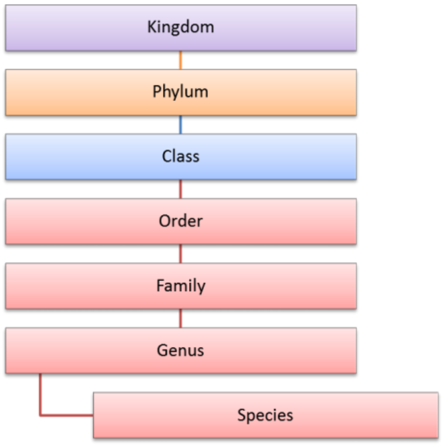

Figure 7 Species Classification

Inter Species Translations

Figure 8 Plum Headed Parakeet → Rose Ringed Parakeet

Inter Genus Translations

Figure 9 Cattle Egret → Indian Pond Heron

Figure 10 Indian Pond Heron → Cattle Egret

Figure 11 Yellow Footed Green Pigeon → Laughing Dove

Figure 12 Pied Kingfisher → White Throated Kingfisher

Figure 13 Yellow Footed Green Pigeon → Laughing Dove

Inter Family Translations

Figure 14 White Browed Wagtail → Oriental Magpie Robin

More Results

Figure 15 Yellow Footed Green Pigeon → Laughing Dove

Figure 16 Yellow Footed Green Pigeon → Laughing Dove

Figure 17 Cattle Egret → Indian Pond Heron

Figure 18 Pied Kingfisher → White Throated Kingfisher

Figure 19 Pied Kingfisher → White Throated Kingfisher

Quantitative Analysis Graphs for Inter – Species Translations

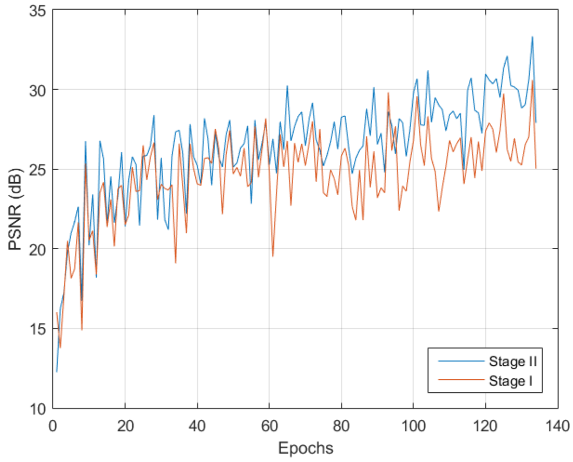

Figure 20 PSNR vs. Epochs plot for Inter - Species Translation (Plum Headed Parakeet to Rose Ringed Parakeet)

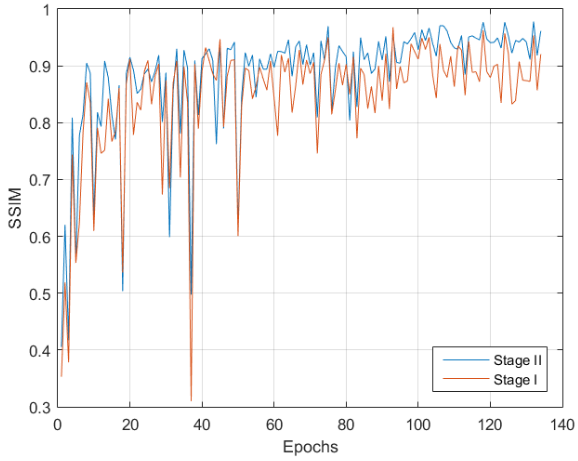

Figure 21 SSIM vs. Epochs plot for Inter - Species Translation (Plum Headed Parakeet to Rose Ringed Parakeet)

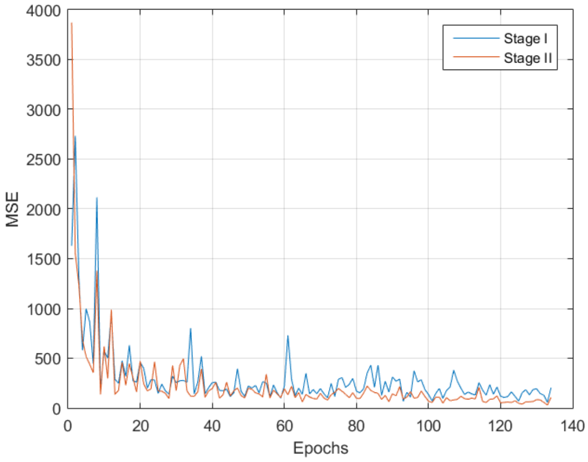

Figure 22 MSE vs. Epochs plot for Inter - Species Translation (Plum Headed Parakeet to Rose Ringed Parakeet)

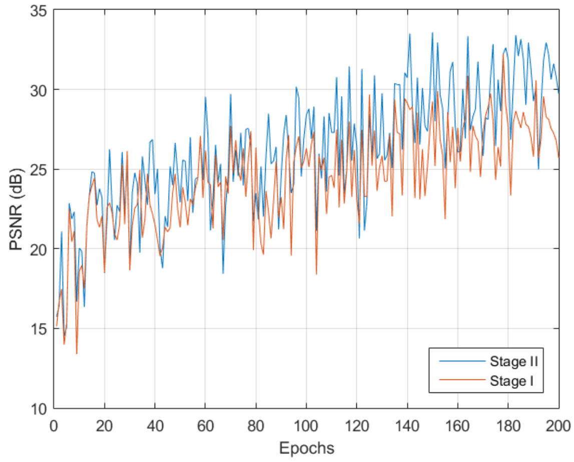

Figure 23 PSNR vs. Epochs plot for Inter - Genus Translation (Cattle Egret → Indian Pond Heron)

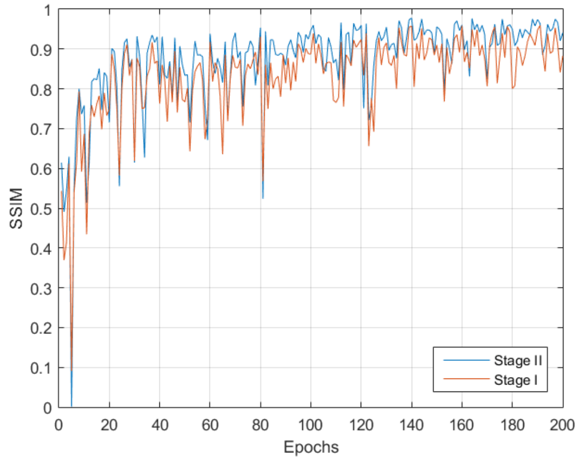

Figure 24 SSIM vs. Epochs plot for Inter - Genus Translation (Cattle Egret → Indian Pond Heron)

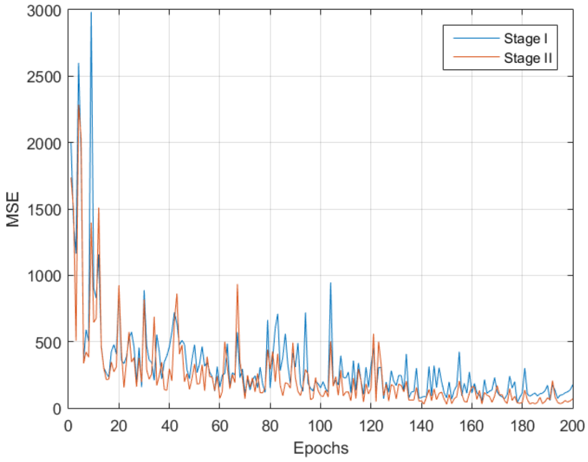

Figure 25 MSE vs. Epochs plot for Inter - Genus Translation (Cattle Egret → Indian Pond Heron)

Quantitative Analysis Graphs for Inter – Genus Species Translations

Quantitative Analysis Graphs for Inter – Family Translations

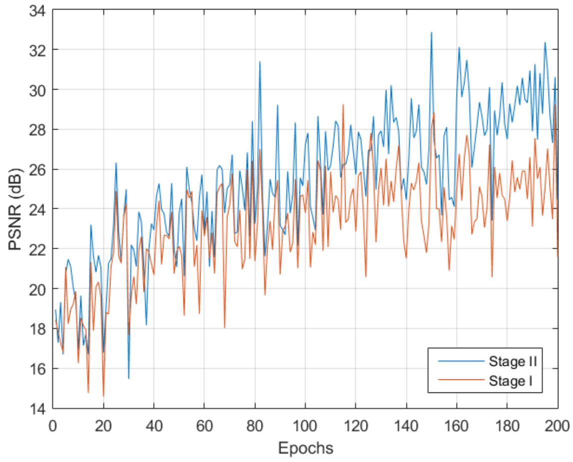

Figure 26 PSNR vs. Epochs plot for Inter - Family Translation (Oriental Magpie Robin → White Browed Wagtail)

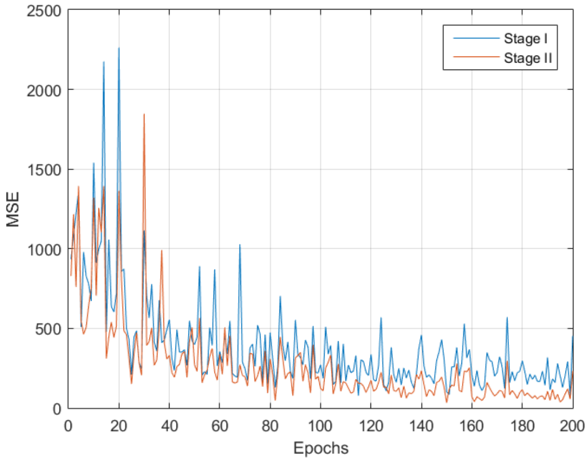

Figure 28 MSE vs. Epochs plot for Inter - Family Translation (Oriental Magpie Robin → White Browed Wagtail)

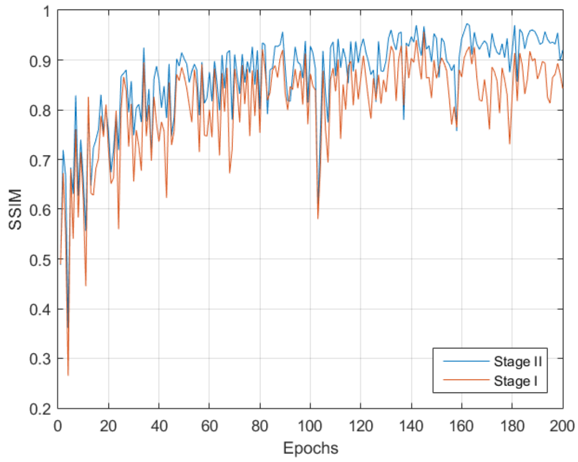

Figure 27 SSIM vs. Epochs plot for Inter - Family Translation (Oriental Magpie Robin → White Browed Wagtail)

Table 3 Quantitative Results obtained for all the 3 simulations carried out for Stage - I and Stage – II.

Translation	Order	Family	Genus	Species	PSNR (dB)	SSIM	MSE
Stage – I							
Inter – Species	Same	Same	Same	Different	30.5926	0.9679	56.7309
Inter – Genus	Same	Same	Different	Different	32.2363	0.9592	38.8550
Inter – Family	Same	Different	Different	Different	29.2560	0.9566	77.1763
Stage - II							
Inter – Species	Same	Same	Same	Different	**33.5918**	**09887**	30.1926
Inter – Genus	Same	Same	Different	Different	33.5855	0.9779	**28.4795**
Inter – Family	Same	Different	Different	Different	32.8863	0.9730	33.4541

VI. CONCLUSION AND FUTURE SCOPE

In this paper, a novel method for translation and resolution improvement of bird species is proposed. A new Aves database for Ave species found in north western part of India is also proposed. The dataset contains fifty bird species and more than five thousand labeled images. However, bounding boxes and annotations are yet to be completed.

For translation, we performed three simulations, intra-species, intra-genus and intra family. All of these translations were found to be successful. For resolution improvement of translated images, we used SRGAN which was pre-trained on DIV2K dataset and 2K bird images from CBNWI-50. Using SRGAN, an up-scaling factor between 2× and 8× can be achieved. A high PSNR value of 33.5918 was achieved along with SSIM value of 0.9778 and a minimum MSE value of 28.4795.

In future, a more versatile style transfer attempt on birds can be made using DualGAN [9] or DiscoGAN [10]. Also, CBNWI-50 can be expanded to more number of species. Bounding boxes and annotations can also be introduced to further enhance the complexity of dataset.

REFERENCES

[1]Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B.,

Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).

[2] Zhu, J.Y., Park, T., Isola, P. and Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).

[3] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W., 2017. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).

[4] Timofte, Radu and Agustsson, Eirikur and Van Gool, Luc and Yang, Ming-Hsuan and Zhang, Lei and Lim, Bee and others, 2017. NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results. In Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.

[5] Welinder, P., Branson, S., Mita, T., Wah, C., Schroff, F., Belongie, S., & Perona, P. (2010). Caltech-UCSD birds 200.

[6] Van Horn, G., Mac Aodha, O., Song, Y., Shepard, A., Adam, H., Perona, P., & Belongie, S. (2017). The inaturalist challenge 2017 dataset. arXiv preprint arXiv:1707.06642, 1(2).

[7] Berg, T., Liu, J., Woo Lee, S., Alexander, M. L., Jacobs, D. W., & Belhumeur, P. N. (2014). Birdsnap: Large-scale fine-grained visual categorization of birds. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2011-2018).

[8] Van Horn, G., Branson, S., Farrell, R., Haber, S., Barry, J., Ipeirotis, P., ... & Belongie, S. (2015). Building a bird recognition app and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 595-604).

[9] Yi, Z., Zhang, H., Tan, P. and Gong, M., 2017. Dualgan: Unsupervised dual learning for image-to-image translation. In Proceedings of the IEEE international conference on computer vision (pp. 2849-2857).

[10] Kim, T., Cha, M., Kim, H., Lee, J.K. and Kim, J., 2017, August. Learning to discover cross-domain relations with generative adversarial networks. In Proceedings of the 34th International Conference on Machine Learning-Volume 70 (pp. 1857-1865). JMLR. org.

AUTHORS PROFILE

Akanksha Sharma received her B. Tech. Degree in Electronics and Communication Engineering from Government Women Engineering College Ajmer, Rajasthan, India. She is currently pursuing her M. E. Degree in Electronics and Communication Engineering from Thapar Institute of Engineering and Technology, Patiala, Punjab, India.

Neeru Jindal received the B.Tech. and M.Tech. degree in Electronics and Communication in 2002 and 2007 respectively from Punjab Technical University, Jalandhar, Punjab, India. She has received Ph.D. degree in 2014 from the Department of Electronics and Communication Engineering, Thapar University, Patiala, Punjab, India. She has been involved in various research activities in the area of image and video processing. She holds the position of Assistant Professor in the department of Electronics and Communication Engineering, Thapar University Patiala, Punjab, India.

Appendix 1
A Species Wise List of Bird Images in CBNWI – 50

S. No.	Common Name	Original Images	Images with attacks	Total
1	Indian Pond Heron	176	1056	1232
2	Cattle Egret	212	1272	1484
3	Common Myna	252	1512	1764
4	Rose Ring Parakeet	193	1158	1351
5	Black Winged Stilt	165	990	1155
6	Rosy Starling	98	588	686
7	Black Drongo	83	498	581
8	Brahminy Myna	107	642	749
9	Lesser Cormorant	66	396	462
10	Eurasian Coot	51	306	357
11	Pied Kingfisher	59	354	413
12	Northern Shoveler	50	300	350
13	Common Teal	63	378	441
14	Purple Sunbird	178	1068	1246
15	Peacock	250	1500	1750
16	Common Tailor Bird	61	366	427
17	Roofus Treepie	77	462	539
18	Oriental Magpie Robin	182	1092	1274
19	Large Grey Babbler	100	600	700
20	Shikra	74	444	518
21	White Throated Kingfisher	165	990	1155
22	Spot Billed Duck	50	300	350
23	Laughing Dove	255	1530	1785
24	Eurasian Collared Dove	169	1014	1183
25	Rock Pigeon	150	900	1050
26	Yellow Footed Green Pigeon	50	300	350
27	Bank Myna	50	300	350
28	Pied Myna	50	300	350
29	Red Vented Bulbul	123	738	861
30	Common Crow	108	648	756
31	Red Wattled Lapwing	100	600	700
32	Indian Robin	100	600	700
33	Hoopoe	50	300	350
34	Spotted Owlet	50	300	350
35	Indian House Sparrow	100	600	700
36	Black Rumped Flame Back Woodpecker	50	300	350
37	White Browed Wagtail	50	300	350
38	Plum Headed Parakeet	50	300	350
39	Coopersmith Barbet	50	300	350
40	Greater Coucal	46	276	322
41	Black Kite	121	726	847
42	Green Bee Eater	134	804	938
43	Asian Koel	66	396	462

44	Jungle Babbler	147	882	1029
45	Indian Grey Hornbill	71	426	497
46	Common Moorhen	50	300	350
47	Purple Moorhen	50	300	350
48	Great White Pelican	50	300	350
49	Indian Roller	50	300	350
50	Common Sandpiper	50	300	350
	TOTAL	**5,102**	**30,612**	**35,714**