Enhancing Hazy Wildlife Imagery: AnimalHaze3k and IncepDehazeGan

Shivarth Rai
raishivarth@gmail.com

Tejeswar Pokuri
tejeswarpokuri3@gmail.com

Abstract

Atmospheric haze significantly degrades wildlife imagery, impeding computer vision applications critical for conservation, such as animal detection, tracking, and behavior analysis. To address this challenge, we introduce AnimalHaze3k–a synthetic dataset comprising of 3,477 hazy images generated from 1,159 clear wildlife photographs through a physics-based pipeline. Our novel IncepDehazeGan architecture combines inception blocks with residual skip connections in a GAN framework, achieving state-of-the-art performance (SSIM: 0.8914, PSNR: 20.54, and LPIPS: 0.1104), delivering 6.27% higher SSIM and 10.2% better PSNR than competing approaches. When applied to downstream detection tasks, dehazed images improved YOLOv11 detection mAP by 112% and IoU by 67%. These advances can provide ecologists with reliable tools for population monitoring and surveillance in challenging environmental conditions, demonstrating significant potential for enhancing wildlife conservation efforts through robust visual analytics. The AnimalHaze3k dataset can be publicly accessed at: https://shvrth.github.io/.

1. Introduction

Computer vision-based methods have become indispensable tools for wildlife conservation and ecological research, enabling non-invasive monitoring of animals, population tracking, and behavioral studies across diverse habitats [12][26]. These methods have become critical for conservation strategies, such as identifying animals through unique markings [3], detecting elusive and endangered species [23][10], and monitoring illegal poaching activities of protected animal species [3]. Their ability to process large volumes of data with high accuracy and consistency [20], and provide cost-effective solutions to real-time animal monitoring have made their use widespread [25][21]. However, environmental challenges like atmospheric haze greatly compromise image quality, undermining the effectiveness of these technologies [19].

Haze-induced degradation manifests through unnatural color shifts, decreased visibility, diminished contrast between subjects and backgrounds, and distorted depth perception [32][27]. Such quality deterioration directly impacts downstream conservation tasks—diminished visibility impedes movement tracking, low-contrast images complicate animal detection and population counting, and can obscure anatomical features challenging automation models for species identification.

Image dehazing aims to generate the latent haze-free image from the observed hazy image. The atmospheric scattering model [17][18] is the classical framework for modeling hazy image formation through light-particle interactions. This model describes hazy images as a per-pixel combination of attenuated scene radiance and atmospheric light interference:

$$I(x) = J(x) \cdot t(x) + A \cdot (1 - t(x)) \tag{1}$$

where x is the pixel index, $I(x)$ is observed hazy image, $J(x)$ is the clean scene radiance, A denotes the global atmospheric light, and $t(x)$ is the transmission matrix defined as:

$$t(x) = e^{-\beta \cdot d(x)} \tag{2}$$

where β represents the scattering coefficient of the atmosphere, and $d(x)$ denotes the depth information.

Recently, single image dehazing has made significant progress with the use of data-driven approaches utilizing synthetic image pairs to achieve superior performance. Initial CNN-based architectures [2][24][37] employed a decomposed estimation framework, separately predicting atmospheric light A and transmission maps $t(x)$, with supervision derived from synthetic transmission ground truth. Modern approaches [5][6][34] have pivoted toward direct prediction of latent haze-free images or their residual components relative to hazy inputs, capitalizing on pixel-level reconstruction to optimize perceptual quality. ViT-based methods [31][28] have outperformed CNN-based models through their use of global attention mechanisms.

While recent methodological advancements demonstrate considerable promise, their efficacy remains constrained by a critical bottleneck: the scarcity of paired real-world training data exhibiting authentic haze conditions. For the task of dehazing wildlife imagery, acquiring such datasets poses significant challenges due to environmental variability, and

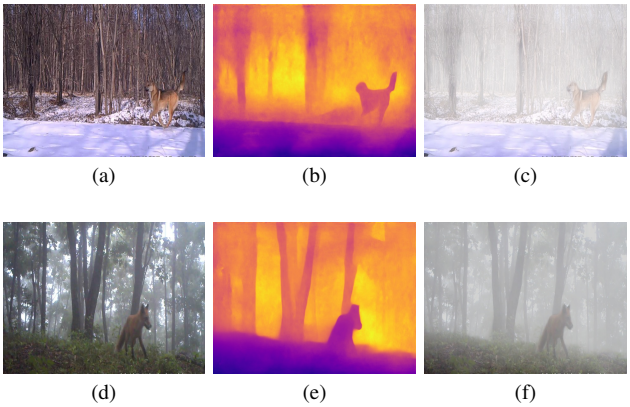


Figure 1. Samples from AnimalHaze3k dataset showing the ground truth, depth map used for synthetic haze generation and hazy image generated for a (a-c) Dog, and, (d-f) Red Fox.

the inherent difficulty of capturing the same haze-free and hazy scene in dynamic wildlife habitats. In response to these data acquisition challenges, contemporary dehazing approaches synthesize training pairs by artificially inducing haze in clean images. This process involves estimating scene depth—either leveraging existing depth maps from specialized datasets or deriving them algorithmically—and simulating haze formation through the atmospheric scattering model using Eq. (1).

This work aims to overcome the aforementioned hurdles and makes two-fold technical contributions:

- AnimalHaze3k dataset: A synthetic dataset of 3,477 hazy wildlife images generated from 1,159 real wildlife photographs via a physics-based pipeline, simulating diverse atmospheric haze conditions.
- IncepDehazeGan: A GAN architecture integrating inception blocks, residual skip connections, and a hybrid adversarial-L1 loss, achieving state-of-the-art dehazing on the proposed AnimalHaze3k dataset.

2. The AnimalHaze3k Dataset

The dataset used for acquiring clear images to generate synthetic hazy images is the Northeast Tiger and Leopard National Park (NTLNP) dataset[30]. This dataset contains 25,567 images of 17 species captured under both daylight and night-time conditions. Only daylight images were utilized in this study. The images were obtained through infrared camera traps deployed in China's Northeast Tiger and Leopard National Park between 2014 and 2020. The camera trap methodology ensures animals were observed in their natural, undisturbed habitats. The dataset features diverse environmental contexts including varied backgrounds, weather conditions, and seasonal variations for non-hibernating species. From this collection, we systematically selected 1,159 images representing 11 species. All images were pre-processed to remove temporal stamps (date and time) and were standardized to 640×480 pixel resolution.

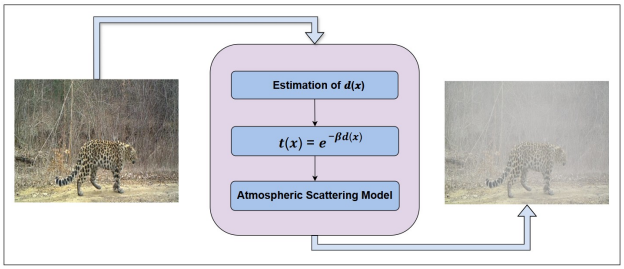


Figure 2. Data generation pipeline stages: (1) Depth estimation, (2) Transmission map calculation, (3) Synthetic haze generation.

The AnimalHaze3k data generation pipeline comprises three principal stages as illustrated in Fig. 2. In the first stage, a depth map for the input image is estimated using the HybridDepth[9] model. The reasons to choose HybridDepth are threefold: (1) HybridDepth is a state-of-art single image metric depth estimation pipeline. By fusing depth-from-focus (DFF) data with relative depth priors, HybridDepth achieves superior metric accuracy and generalization over models like ZoeDepth[1], DFV[35] and Depth Anything[36]. (2) HybridDepth maintains consistent depth estimations across different zoom levels - while Depth Anything and DFV can overestimate depth. (3) HybridDepth uses Depth Anything[36] as the relative depth estimator. Depth Anything has been trained on a vast amount of outdoor data that results in high quality synthetic hazy images generated. In the second stage of the data generation pipeline, transmission map, $t(x)$, is calculated using Eq. (2). Value of scattering coefficient β is uniformly sampled from [1.8, 3.0], in order to generate an arbitrary haze density for each input image and improve the diversity of the dataset. In the final stage, the synthetic hazy image is generated using Eqn.(2). Atmospheric ambient light, A, values are taken at random from discrete values [0.8, 0.85, 0.9, 0.95, 1] following established methodologies[15].

Each clear image generates three synthetic hazy variations through this pipeline, ensuring comprehensive coverage of potential atmospheric conditions. The training set contains 1,041 clear and 3,123 hazy images. The validation and test sets contain 59 clear and 177 hazy images each. The dataset thus contains paired clear and hazy images. Examples from the dataset are depicted in Fig. 1.

3. The IncepDehazeGan Model

IncepDehazeGan is a novel single-image dehazing GAN architecture. The IncepDehazeGan generator is a dense encoder-decoder network incorporating Inception

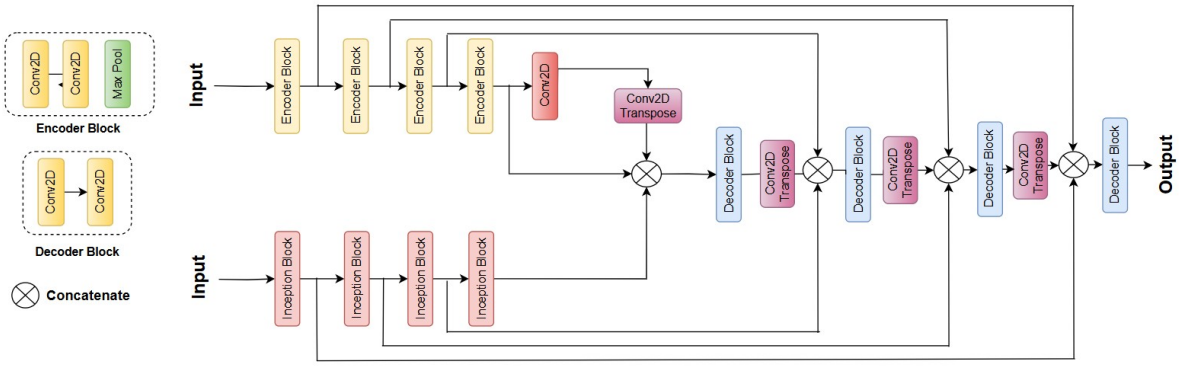


Figure 3. IncepDehazeGan Generator architecture

Blocks[29]. The encoder consists of two parallel processing sections: (1) a series of standard encoding blocks made up of two convolution layers followed by ReLU activation function and a max pooling layer, and (2) a series of Inception Blocks using four parallel convolutional layers of distinct kernel sizes $(1 \times 1, 1 \times 3, 3 \times 1, 3 \times 3)$.

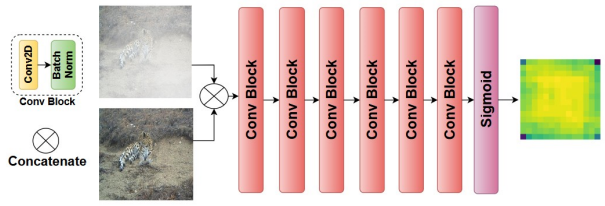


Figure 4. IncepDehazeGan Discriminator architecture

The decoder is a single series of blocks consisting of a transpose convolution layer preceded by two standard convolution layers. We use residual skip connections to share low-level feature maps learned in each encoder block with decoder blocks[8]. This multilayer feature fusion reduces information loss due to down sampling operations in the encoder section[16]. The overall generator architecture is illustrated by Fig. 3. For training the generator, we use a combination of adversarial loss and mean absolute error. The adversarial loss is computed using the output of the discriminator and the target labels of ones, encouraging the generator to produce realistic outputs that deceive the discriminator. This loss is defined as follows:

$$\mathcal{L}_{\text{Adv}}(x, y) = -\frac{1}{N}\sum_{i=1}^{N}\Big[y_i \log\left(\sigma(x_i)\right) + (1 - y_i)\log\left(1 - \sigma(x_i)\right)\Big] \tag{3}$$

Mean absolute error ($L1$ loss) is calculated between the generated image and the target image. This loss term helps to maintain the fidelity of the generated output relative to the ground truth and is defined as:

$$\mathcal{L}_{\text{L1}} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{4}$$

The net generator loss is a weighted sum of the adversarial loss and the L1 loss, where λ is a hyperparameter controlling the trade-off between these two losses. In our implementation, λ is set to 100. The net generator loss is thus expressed as:

$$\mathcal{L}_{\text{G}} = \mathcal{L}_{\text{Adv}} + \lambda \cdot \mathcal{L}_{\text{L1}} \tag{5}$$

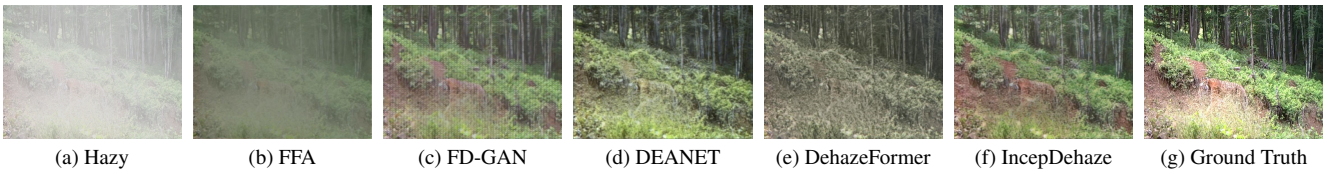
The IncepDehazeGan discriminator consists of six convolutional blocks followed by a final sigmoid activation. Each convolution block contains a convolutional layer with leaky ReLU activation and a batch normalization layer. Fig. 4 depicts the IncepDehazeGan discriminator architecture. This produces a single-channel matrix, encouraging the generator to produce realistic local features and textures across the image[13]. By focusing on small regions for each pixel, the discriminator also helps mitigate overfitting. For training the discriminator, we use Binary Cross Entropy to distinguish between ground truth and output produced by the generator.

4. Experimental Study

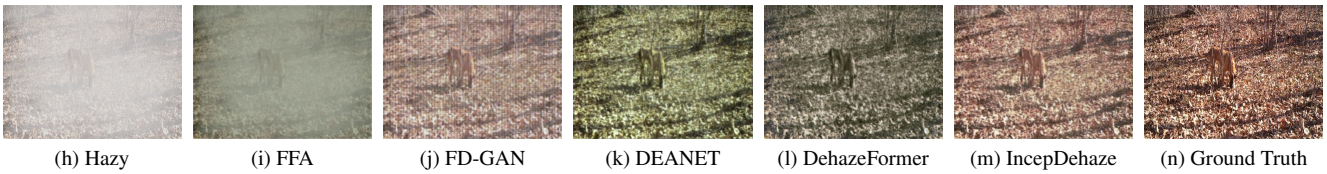
Experiments in this study were conducted on a NVIDIA Tesla P100 GPU, with 16GB memory and 3584 CUDA cores.

4.1. Single-Image Dehazing

We benchmark our AnimalHaze3k dataset using several state-of-the-art image dehazing models, FD-GAN[7], FFA-Net[22], DehazeFormer[28], DEA-Net[4] and the proposed IncepDehazeGan model. The models were trained for 50 epochs, with batch size of 4. We extensively evaluated performance using several metrics-SSIM[33], PSNR[11], FSIM[38] and LPIPS[39].

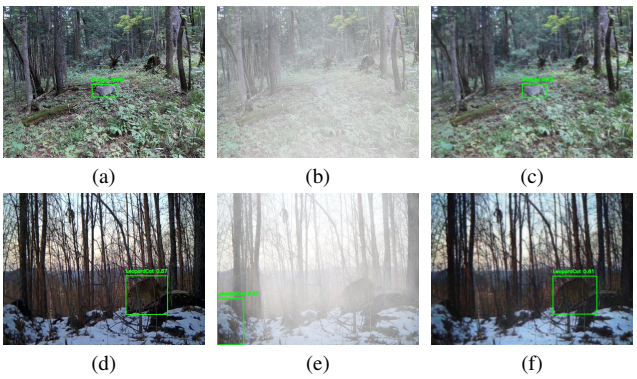


(a) Hazy (b) FFA (c) FD-GAN (d) DEANET (e) DehazeFormer (f) IncepDehaze (g) Ground Truth



(h) Hazy (i) FFA (j) FD-GAN (k) DEANET (l) DehazeFormer (m) IncepDehaze (n) Ground Truth

Figure 5. Qualitative comparison on dehazing results across SOTA models and IncepDehazeGan (our model).



(a) (b) (c) (d) (e) (f)

Figure 6. YOLOv11 animal detection on (a,d) ground truth, (b,e) hazy image, and (c,f) dehazed image

Model Name	SSIM ↑	PSNR ↑	FSIM ↑	LPIPS ↓
FFA[22]	0.5468	12.1842	0.6665	0.3610
FD GAN[7]	0.5580	17.5573	0.8314	0.1637
DEANET[4]	0.8303	18.6481	0.8936	0.2102
DehazeFormer[28]	0.8388	17.4550	0.8917	0.2375
IncepDehaze	**0.8914**	**20.5404**	**0.9363**	**0.1104**

Table 1. Performance of different dehazing methods on the AnimalHaze3k dataset.

Images	mAP ↑	mIoU ↑
Hazy	0.3216	0.4313
Dehazed Output	**0.6842**	**0.7201**

Table 2. Animal detection performance of YOLOv11 on hazy images and dehazed counterparts generated IncepdeHazeGan.

4.2. Animal Detection

To validate the efficacy of our approach for the downstream detection task, we train the YOLOv11[14] detection model on the NTLNP dataset and evaluate performance on hazy images from the AnimalHaze3k test set and the corresponding dehazed images generated by IncepdehazeGan. Performance was evaluated using mAP and IoU metrics.

Experimentation demonstrates the superiority of our approach at dehazing wildlife imagery. Achieving a remarkable SSIM of 0.8913 and PSNR of 20.54, our model surpasses existing SOTA methods (see Tab. 1). Our experiments on detection show a significant improvement in mAP (>112%) and mIoU (>67%) metrics when using dehazed images generated by IncepDehazeGan (see Tab. 2). This demonstrates the effectiveness of our model at providing high-quality data for all computer-vision tasks related to wildlife monitoring (see Fig. 6).

5. Conclusion

This work presents the AnimalHaze3k dataset, a synthetic single-image dehazing dataset addressing atmospheric degradation in wildlife imagery. For this purpose, we introduce the IncepDehazeGan model. Extensive experimental comparisons with current state-of-the-art models on various metrics (PSNR, SSIM, LPIPS and FSIM) confirm the effectiveness of our model in generating clearer images. Our experiment on the detection task demonstrates that IncepDehazeGan overcomes challenges posed by atmospheric haze and can provide higher quality data for tasks such as animal identification, population tracking, movement tracking, behavior pattern tracking and surveillance. Thus, it helps advance efforts for animal behavioral studies and conservation.

References

[1] Shariq Farooq Bhat, Reiner Birkl, Diana Wofk, Peter Wonka, and Matthias Müller. Zoedepth: Zero-shot transfer by combining relative and metric depth. *arXiv preprint arXiv:2302.12288*, 2023. 2

[2] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single

image haze removal. *IEEE transactions on image processing*, 25(11):5187–5198, 2016. 1

[3] Carl Chalmers, Paul Fergus, Serge Wich, and Aday Curbelo Montanez. Conservation ai: Live stream analysis for the detection of endangered species using convolutional neural networks and drone technology. *arXiv preprint arXiv:1910.07360*, 2019. 1

[4] Zixuan Chen, Zewei He, and Zhe-Ming Lu. Dea-net: Single image dehazing based on detail-enhanced convolution and content-guided attention, 2023. 3, 4

[5] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2157–2167, 2020. 1

[6] Jiangxin Dong and Jinshan Pan. Physics-based feature dehazing networks. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXX 16*, pages 188–204. Springer, 2020. 1

[7] Yu Dong, Yihao Liu, He Zhang, Shifeng Chen, and Yu Qiao. FD-GAN: generative adversarial networks with fusion-discriminator for single image dehazing. *CoRR*, abs/2001.06968, 2020. 3, 4

[8] Yang Du, Jun Li, Qinghong Sheng, Yuxin Zhu, Bo Wang, and Xiao Ling. Dehazing network: Asymmetric unet based on physical model. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–12, 2024. 3

[9] Ashkan Ganj, Hang Su, and Tian Guo. Hybriddepth: Robust metric depth fusion by leveraging depth from focus and single-image priors. *arXiv preprint arXiv:2407.18443*, 2024. 2

[10] Yan Guan, Yujie Lei, Yuhui Zhu, Tingxuan Li, Ying Xiang, Pengmei Dong, Rong Jiang, Jinwen Luo, Anqi Huang, Yumai Fan, et al. Face recognition of a lorisidae species based on computer vision. *Global Ecology and Conservation*, 45: e02511, 2023. 1

[11] Alain Horé and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369, 2010. 3

[12] Jin Hou, Yuxin He, Hongbo Yang, Thomas Connor, Jie Gao, Yujun Wang, Yichao Zeng, Jindong Zhang, Jinyan Huang, Bochuan Zheng, et al. Identification of animal individuals using deep learning: A case study of giant panda. *Biological Conservation*, 242:108414, 2020. 1

[13] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 3

[14] Rahima Khanam and Muhammad Hussain. Yolov11: An overview of the key architectural enhancements, 2024. 4

[15] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018. 2

[16] Chenhui Ma, Xiaodong Mu, and Dexuan Sha. Multi-layers feature fusion of convolutional neural network for scene classification of remote sensing. *IEEE Access*, 7:121685–121694, 2019. 3

[17] Srinivasa G Narasimhan and Shree K Nayar. Vision and the atmosphere. *International journal of computer vision*, 48: 233–254, 2002. 1

[18] Srinivasa G Narasimhan and Shree K Nayar. Interactive (de) weathering of an image using physical models. In *IEEE Workshop on color and photometric Methods in computer Vision*, page 1. France, 2003. 1

[19] Scott Newey, Paul Davidson, Sajid Nazir, Gorry Fairhurst, Fabio Verdicchio, R Justin Irvine, and René Van Der Wal. Limitations of recreational camera traps for wildlife management and conservation research: A practitioner's perspective. *Ambio*, 44:624–635, 2015. 1

[20] Mohammad Sadegh Norouzzadeh, Anh Nguyen, Margaret Kosmala, Alexandra Swanson, Meredith S Palmer, Craig Packer, and Jeff Clune. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, 115(25):E5716–E5725, 2018. 1

[21] Allan F O'Connell, James D Nichols, and K Ullas Karanth. *Camera traps in animal ecology: methods and analyses*. Springer, 2011. 1

[22] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. *CoRR*, abs/1911.07559, 2019. 3, 4

[23] Keni Ren, Gun Bernes, Mårten Hetta, and Johannes Karlsson. Tracking and analysing social interactions in dairy cattle with real-time locating system and machine learning. *Journal of Systems Architecture*, 116:102139, 2021. 1

[24] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 154–169. Springer, 2016. 1

[25] Lyes Saad Saoud, Loïc Lesobre, Enrico Sorato, Saud Al Qaydi, Yves Hingrat, Lakmal Seneviratne, and Irfan Hussain. Hubot: A biomimicking mobile robot for non-disruptive bird behavior study. *Ecological Informatics*, 85: 102939, 2025. 1

[26] Frank Schindler and Volker Steinhage. Identification of animals and recognition of their actions in wildlife videos using deep learning techniques. *Ecological Informatics*, 61: 101215, 2021. 1

[27] Dilbag Singh and Vijay Kumar. A comprehensive review of computational dehazing techniques. *Archives of Computational Methods in Engineering*, 26(5):1395–1413, 2019. 1

[28] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023. 1, 3, 4

[29] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014. 3

[30] Mengyu Tan, Wentao Chao, Jo-Ku Cheng, Mo Zhou, Yiwen Ma, Xinyi Jiang, Jianping Ge, Lian Yu, and Limin Feng. An-

imal detection and classification from camera trap images using different mainstream object detection architectures. *Animals*, 12(15):1976, 2022. 2

[31] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 1

[32] Wencheng Wang and Xiaohui Yuan. Recent advances in image dehazing. *IEEE/CAA Journal of Automatica Sinica*, 4 (3):410–436, 2017. 1

[33] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. 3

[34] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10551–10560, 2021. 1

[35] Fengting Yang, Xiaolei Huang, and Zihan Zhou. Deep depth from focus with differential focus volume. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12642–12651, 2022. 2

[36] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10371–10381, 2024. 2

[37] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3194–3203, 2018. 1

[38] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, 2011. 3

[39] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. *CoRR*, abs/1801.03924, 2018. 3