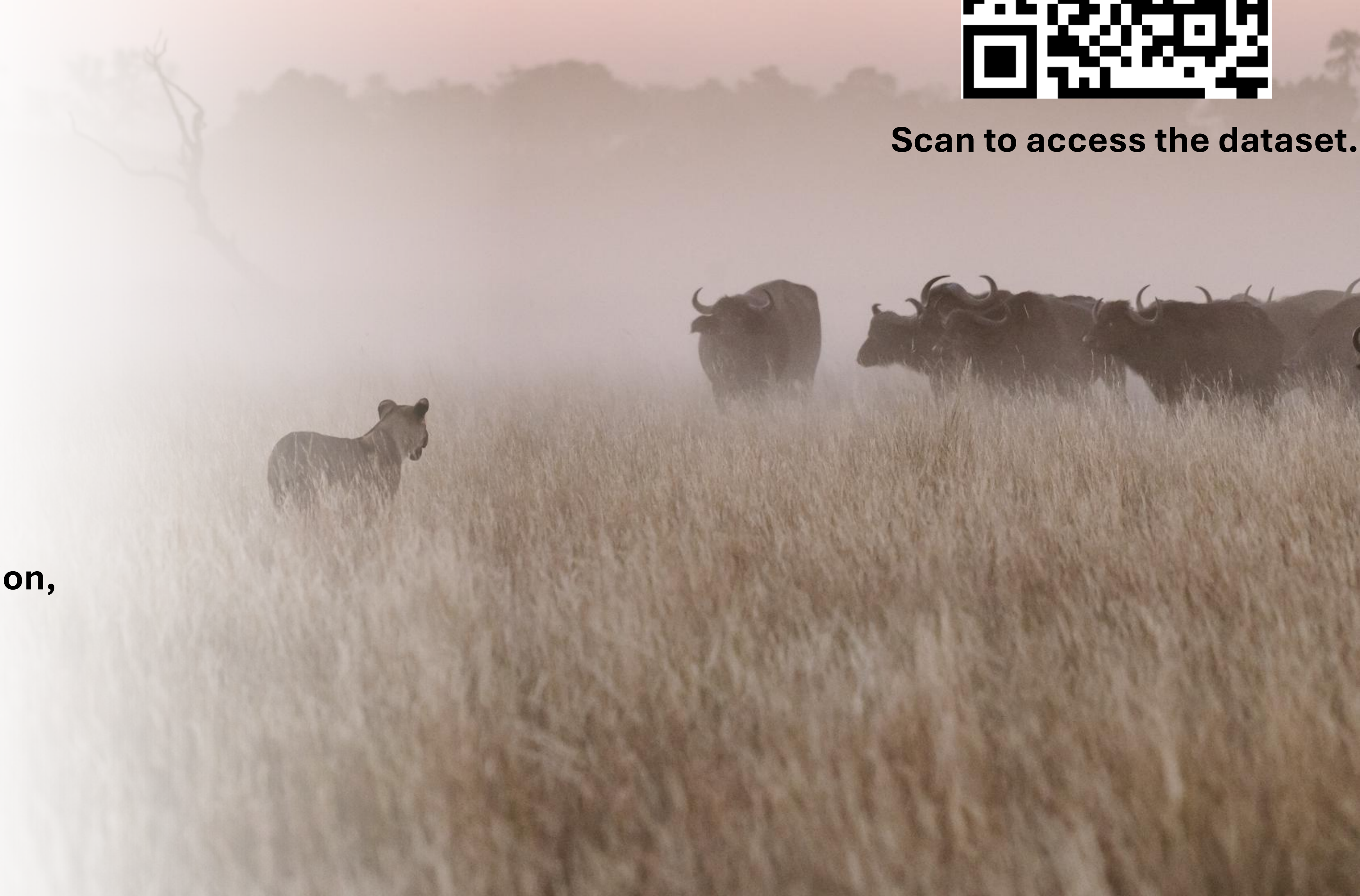
CVPR Nashville JUNE 11-15, 2025



Scan to access the dataset.

Enhancing Hazy Wildlife Imagery: AnimalHaze3k and IncepDehazeGan

Shivarth Rai, Tejeswar Pokuri
RUGVED Systems,
Manipal Academy of Higher Education,
Manipal, India

MANIPAL ACADEMY of HIGHER EDUCATION
(Deemed to be University under Section 3 of the UGC Act, 1956)
INSPIRED BY LIFE

R.U.G.V.E.D

Problem Statement

Animal image dehazing is the process of restoring clear, high-contrast wildlife images by removing atmospheric haze to improve visual quality and enhance computer vision performance.

Problem	Impact
Low Contrast & Detail	Missed detections, poor visibility
Hazy Input to CV Models	Reduced mAP, IoU, tracking accuracy
Poor Data Quality	Skewed behavior analysis & population stats
Color Distortion	Misclassification, reduced ecological value
Research Cost Increases	Manual reviews, misinformed field planning

Fig 1. Few Pairs of Hazy and Dehazed Images

Our Contributions

1) **AnimalHaze3K**

A novel synthetic wildlife dataset with 3,477 hazy images generated from 1,159 clear photographs using a physics-based haze simulation pipeline.

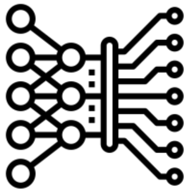
2) **IncepDehazeGan Architecture**

A GAN-based dehazing model integrating inception blocks and residual skip connections for multiscale feature extraction and minimal information loss (see Fig. 3).

Results

- Achieved a **SSIM of 0.8913** and **PSNR of 20.54**, our model surpasses existing SOTA methods
- Dehazed images improved YOLOv11 detection mAP by **112%** and IoU by **67%**
- Dataset size: **3,477 hazy images** from 1,159 clear

Dataset Creation Pipelines



Novel Model Architecture

Dataset Creation Pipeline

- **Image Selection and Processing :** 1,159 clear wildlife images were selected from the **NTLNP dataset[1]**, cleaned to remove timestamps, and resized to 640×480 to ensure uniformity.
- **Depth Map Estimation with HybridDepth[2] :** Each image underwent depth estimation using the **HybridDepth** model, chosen for its fusion of depth-from-focus (DFF) and relative depth priors, offering better metric accuracy and consistency across zoom levels.
- **Transmission Map Calculation :** Using the estimated depth d(x), a **transmission map t(x)** was computed via the atmospheric scattering model, where β (scattering coefficient) was uniformly sampled from [1.8, 3.0][3] to simulate variable haze densities.

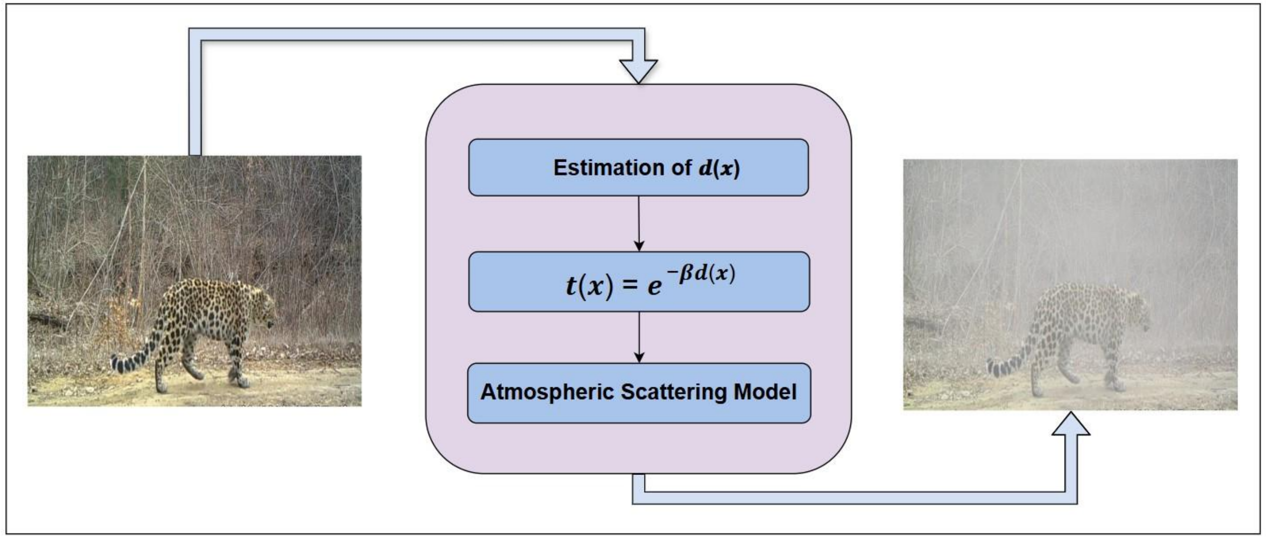


Fig 2. Dataset creation pipeline

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

Eqn 1.Transmision Map calculation

[1] Tan, Mengyu, et al. "Animal detection and classification from camera trap images using different mainstream object detection architectures." *Animals* 12.15 (2022): 1976.
[2] Ganj, Ashkan, Hang Su, and Tian Guo. "HybridDepth: Robust Metric Depth Fusion by Leveraging Depth from Focus and Single-Image Priors." *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2025.
[3] Li, Boyi, et al. "Benchmarking single-image dehazing and beyond." *IEEE Transactions on Image Processing* 28.1 (2018): 492-505.

Dataset Creation Pipeline

- **Synthetic Haze Generation :** The hazy image was synthesized with Eqn. 2, using randomly sampled ambient light values **$A \in [0.8, 0.85, 0.9, 0.95, 1.0]$[1]** to add photorealism.
- **Multiple Haze Variants per Image :** Each clear image generated **three unique hazy versions** with different combinations of β and A, ensuring diversity in atmospheric effects across the 3,477 hazy samples.

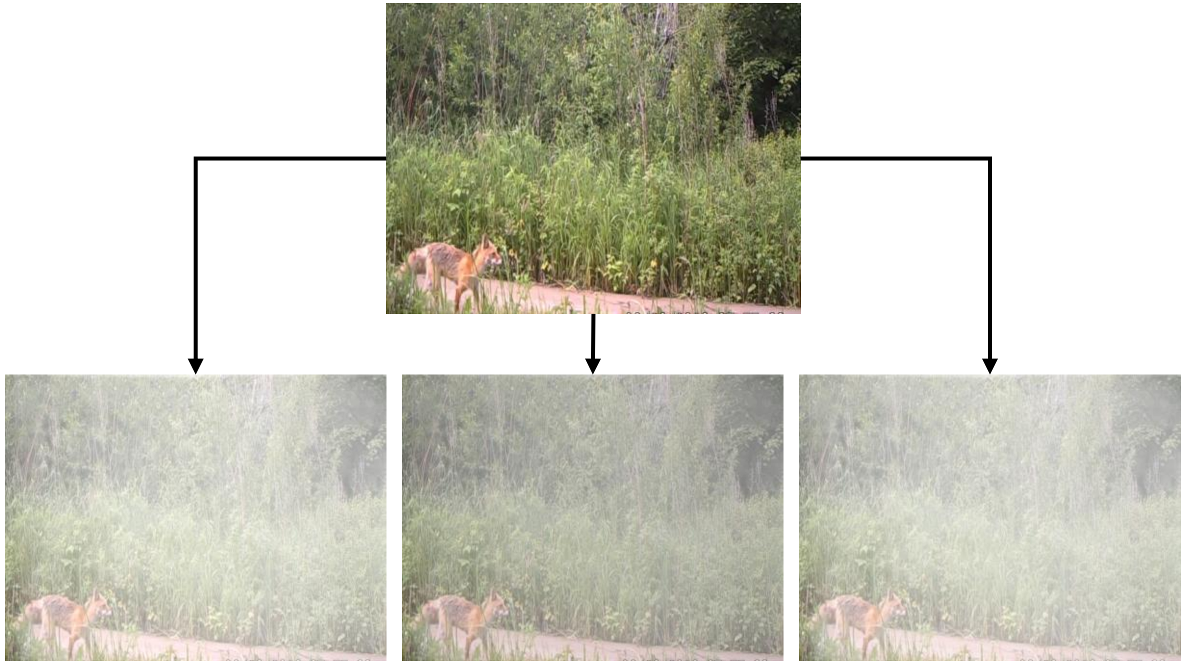


Fig 3. Sample Image of Input Image and generated Hazy Images

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

Eqn 2. Atmospheric Scattering Model

[1] Li, Boyi, et al. "Benchmarking single-image dehazing and beyond." IEEE Transactions on Image Processing 28.1 (2018): 492-505.

Model Architecture

- **Inception-Based Generator Design :** The generator uses a **dense encoder-decoder** structure with **Inception Blocks**[1], enabling **multi-scale feature extraction** through parallel convolutions (1×1, 1×3, 3×1, 3×3) - capturing both local details and global context.
- **Patch-Level based Gradient Calculations :** The discriminator doesn't calculate gradient based on single value instead, it assesses **patches**, encouraging the generator to **reproduce realistic local textures** and reduce over-smoothing.
- **Multi Level Feature Fusion :** The generator incorporates **multi-level feature fusion**, combining both low-level texture details and high-level semantic cues across layers.

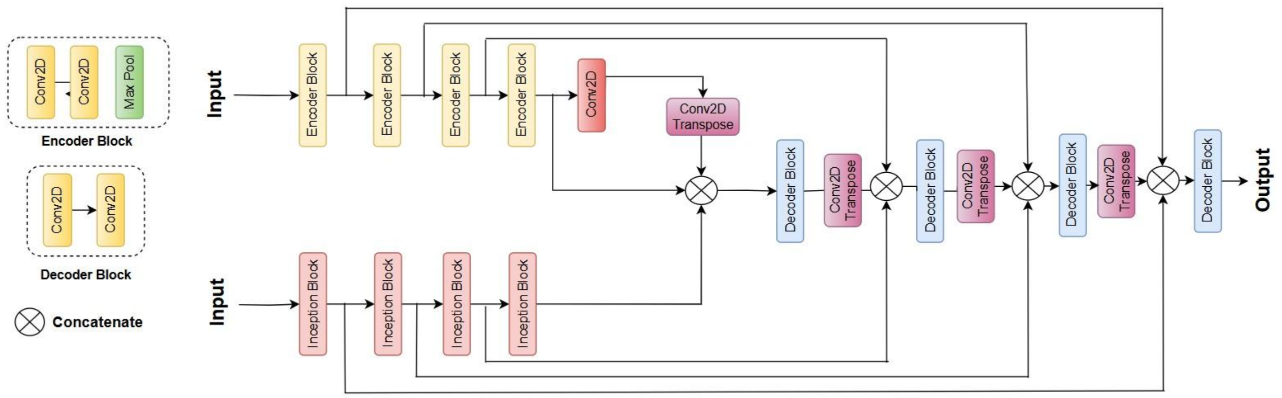


Fig 4. Generator Diagram

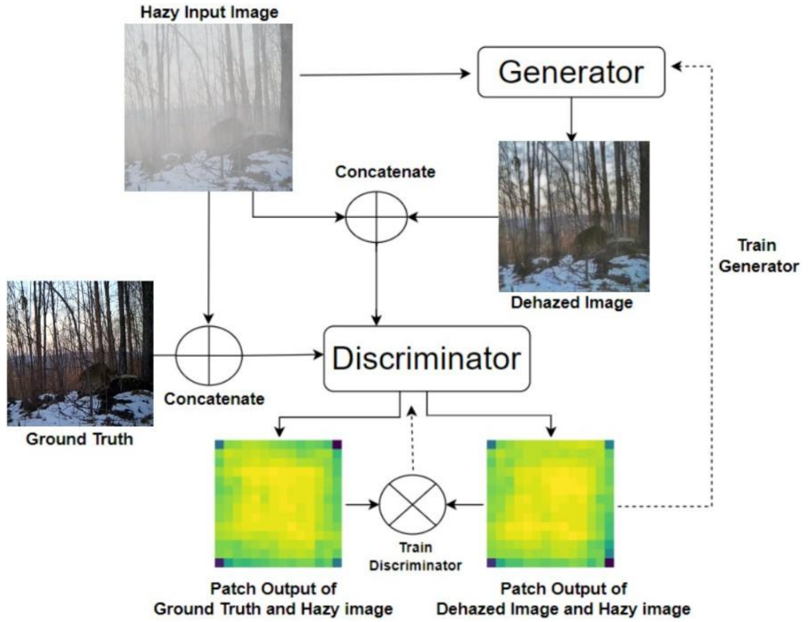


Fig 5. Training Pipeline

[1] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

Model Architecture

- **Residual Skip Connection :** Residual connections link encoder and decoder layers to fuse low-level features across scales, **reducing information loss** during down sampling and aiding structural consistency in dehazed outputs.
- **Hybrid Loss Function :** adversarial loss ensures realism and **L1 loss maintains pixel-wise fidelity** to the ground truth.

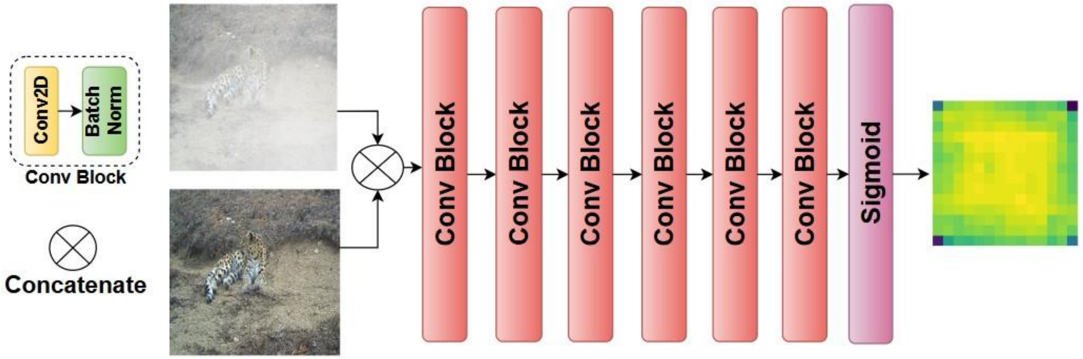


Fig 6. Discriminator Diagram

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

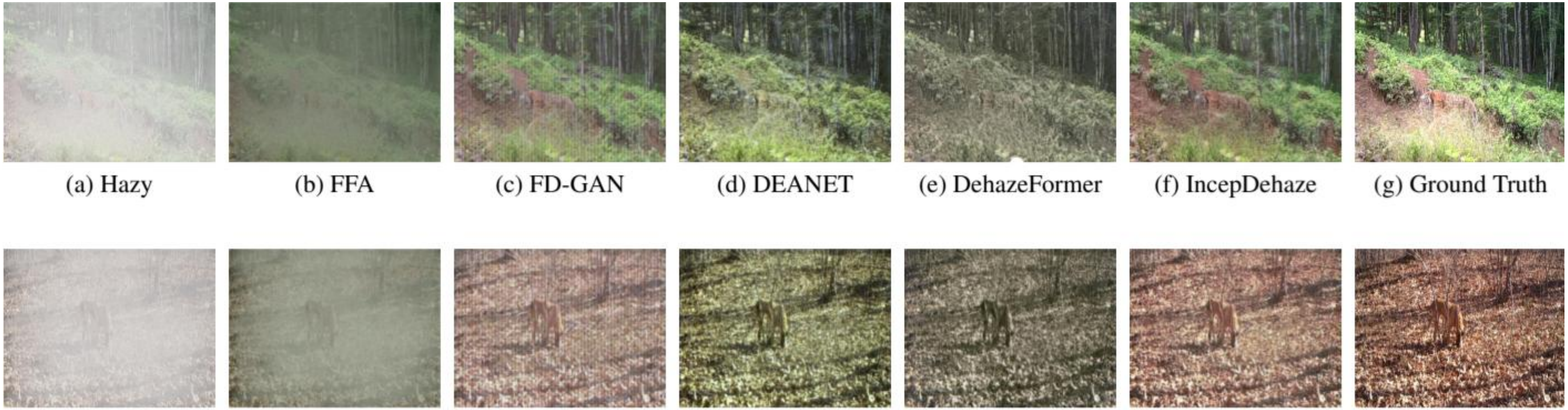
$$\mathcal{L}_{\mathrm{L1}} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

$$\mathcal{L}_{\mathrm{G}} = \mathcal{L}_{\mathrm{Adv}} + \lambda \cdot \mathcal{L}_{\mathrm{L1}}$$

Eqn 3. Training Loss Functions

Qualitative and Quantitative Analysis

Model Name	SSIM ↑	PSNR ↑	FSIM ↑	LPIPS ↓
FFA-Net[1]	0.5468	12.1842	0.6665	0.3610
FD-GAN[2]	0.5580	17.5573	0.8314	0.1637
DEA-Net[3]	0.8303	18.6481	0.8936	0.2102
DehazeFormer[4]	0.8388	17.4550	0.8917	0.2375
IncepDehazeGAN	**0.8914**	**20.5404**	**0.9363**	**0.1104**



(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

Fig 7. Qualitative comparison on dehazing results across SOTA models and IncepDehazeGAN (our model).

[1] Qin, Xu, et al. "FFA-Net: Feature fusion attention network for single image dehazing." *Proceedings of the AAAI conference on artificial intelligence*. Vol. 34. No. 07. 2020.

[2] Dong, Yu, et al. "FD-GAN: Generative adversarial networks with fusion-discriminator for single image dehazing." *Proceedings of the AAAI conference on artificial intelligence*. Vol. 34. No. 07. 2020.

[3] Chen, Zixuan, Zewei He, and Zhe-Ming Lu. "DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention." *IEEE Transactions on Image Processing* 33 (2024): 1002-1015.

[4] Song, Yuda, et al. "Vision transformers for single image dehazing." *IEEE Transactions on Image Processing* 32 (2023): 1927-1941.

Application Advantages of Dehazing

A. Improved Animal Detection

Tested using Yolov11[1] detection model.

Image Type	mAP ↑	mIoU ↑
Hazy	0.3216	0.4313
Dehazed Output	**0.6842**	**0.7201**

Fig 8. YOLOv11 animal detection on hazy image, dehazed image and ground truth respectively.

B. Qualitative comparison on Animal Segmentation

Segmentation performed using SegmentAnything[2] model.

Fig 9. SegmentAnything animal segmentation on hazy image, dehazed image and ground truth respectively.

[1] https://docs.ultralytics.com/models/yolo11
[2] https://segment-anything.com/demo

Our Contributions… In a Nutshell

AnimalHaze3k Dataset

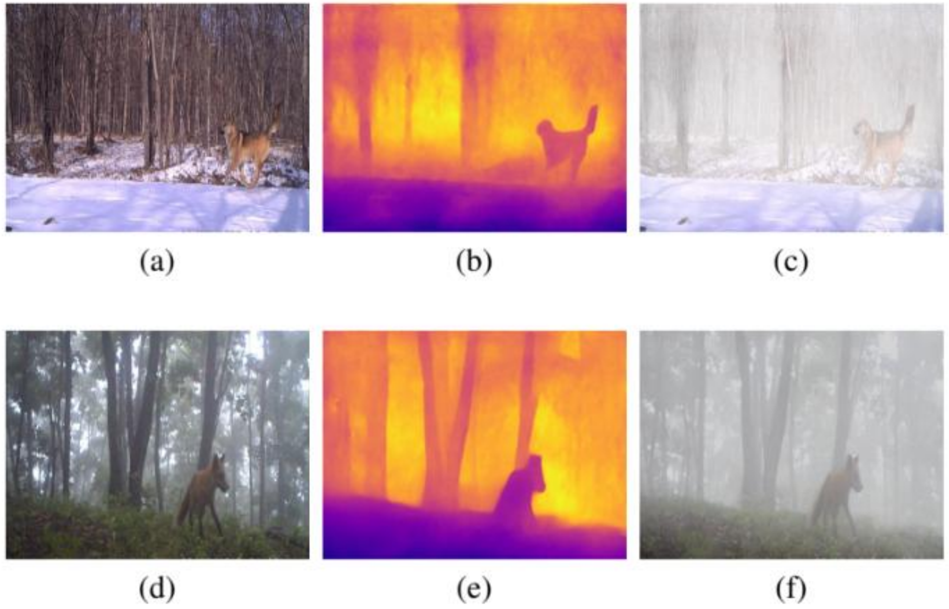


Fig 10. Samples from AnimalHaze3k dataset showing the ground truth, depth map used for synthetic haze generation and hazy image generated for a (a-c) Dog, and, (d-f) Red Fox.

IncepDehazeGan Model

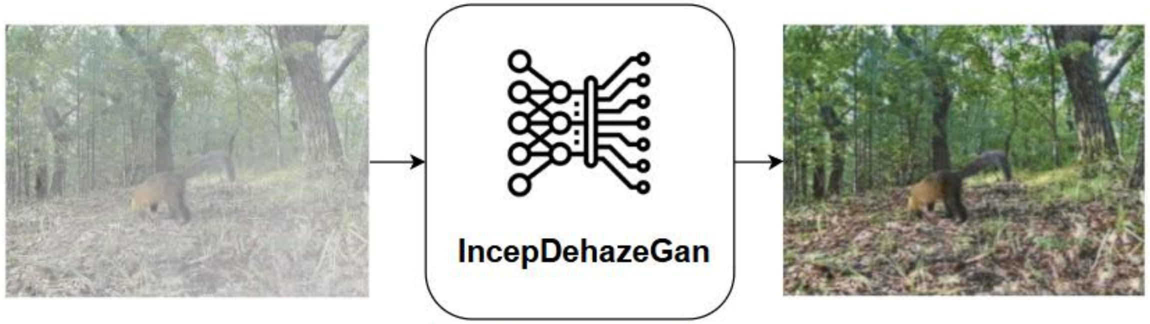


Fig. 11 Working of IncepDehazeGan

Fig 12. YOLOv11 animal detection on (a) ground truth, (b) hazy image, and (c) dehazed image.

Thank you for Listening!

Scan to access the dataset.

Contact us at:
Shivarth Rai {raishivarth@gmail.com}
Tejeswar Pokuri {tejeswar04@gmail.com}