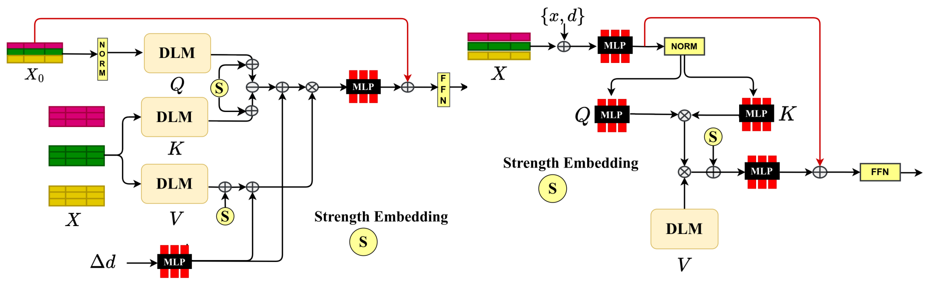
A More Details on Model Architecture



(a) View Transformer **(b)** Ray Transformer

Fig. 1: A more detailed representation of the DLM blocks in the View and Ray transformer of GAURA

In Figure 1, we elaborate on our approach, GAURA. As mentioned in the main paper, specifically, within the View Transformer, we substitute the conventional query, key, and value Multi-Layer Perceptrons (MLPs) with the Degradation Latent Module. This substitution is crucial as the inputs to these MLPs are degradation-dependent, necessitating specific degradation priors for effective restoration. However, in the ray transformer, we only tailor the value matrix to be degradation-specific, as the query-key attention mechanism inherently captures depth-information while lacking detailed appearance features crucial for restoration. Additionally, it is noteworthy that in the View Transformer, subtraction attention is employed due to computational constraints, whereas in the ray transformer, dot-product attention is utilized, as mentioned in [7].

B Synthetic Data Generation.

In this section, we provide a comprehensive overview of the generation process utilized to synthetically introduce corruption to clean images. For the *Low-Light* task, we implement the methodology outlined in [4]. Initially, the RGB image is converted to a Linear RGB scale, and all subsequent manipulations are performed within this color space. Subsequently, the image is appropriately downscaled, followed by the addition of Heteroscedastic noise. We vary the scaling factor within the range of 8 to 30 to encompass a broad spectrum of low-light scenarios. For the generation of *Haze*, we employ the straightforward Koschmieder model [3], which mimics image degradation caused by scattering and ambient light. The equation governing this model is provided as follows:

$$I(i) = J(i)T_B(i) + (1 - T_B(i))A \tag{1}$$

Table 1: Quantitative results on scenes containing *real-world* degradations - specifically low-light enhancement, motion blur removal, and dehazing. All the baselines compared against are all-in-one baselines which generalise both to scene and corruption. The best scores and second best scores are highlighted.

Models	Generalize to		Low-Light			Motion Blur			Haze		
	Scene	Corr.	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
GNT–(Airnet)	✗	✗	17.73	0.577	0.367	20.73	0.612	0.406	15.70	0.590	0.351
GNT–(PromptIR)	✗	✗	17.90	0.573	0.354	19.33	0.552	0.398	16.10	0.632	0.293
GNT–(DA-CLIP)	✓	✗	14.28	0.615	0.424	20.88	0.632	0.410	16.68	0.729	0.300
GNT–(AutoDIR)	✓	✓	11.64	0.570	0.436	20.20	0.601	0.404	14.29	0.678	0.316
Ours	✓	✓	19.91	0.736	0.352	22.12	0.712	0.346	16.82	0.759	0.288

In the equation, J, T_B, and A represent the scene radiance, backscatter transmission map, and global background light, respectively. We employ this equation for its simplicity, facilitating faster inference. Beta and global background light are set within the ranges of 1 to 5 and 125 to 200, respectively. To simulate motion blur and defocus blur, we utilize basic OpenCV functions. For *Motion Blurring*, we adjust the kernel size to modulate the intensity of blur while simultaneously varying the direction of blur to simulate motion in different directions. Accordingly, we set kernel values ranging from 2 to 6 and angles for direction ranging from 0 to 180 degrees. In *Defocus Blur*, we manipulate kernel values from 5 to 11 and randomly apply defocusing to either the foreground or background. It is important to note that we follow a similar approach as outlined in [5], which introduces varying degrees of blur in each captured image.

For simulating *Rain*, we adopt a similar methodology and utilize OpenCV to generate rain particles. Three key parameters are varied during the rain generation process: the intensity of rain, the size of the rain streaks, and the direction of rain. In the case of *Snow* generation, we utilize the imgaug library to introduce snowflakes onto the image. Unlike rain particles, we adjust a range of parameters including the direction of fall, density of the snowflakes, size of the snowflakes, and speed of the snowflakes. For all these corruptions, these parameters are carefully selected to encompass a wide range of variations, enhancing the transferability of synthetic data to real-world multi-view scenarios. We provide visualizations demonstrating the different intensities utilized in our training data across all corruption types in Figure 6.

Table 2: Ablation

Method	PSNR / SSIM / LPIPS
Simple Conditioning	21.34 / 0.689 / 0.394
ARM+DLM (Ours)	22.12 / 0.712 / 0.346

Table 3: Effect of the number of input views for low-light enhancement. Metrics are ordered as PSNR / SSIM / LPIPS

3D Restore	GAURA		
N.A.	3 views	6 views	10 views
17.64 / 0.736 / 0.415	19.18 / 0.697 / 0.387	19.48 / 0.720 / 0.364	19.91 / 0.738 / 0.352

Table 4: Quantitative results to measure view consistency. Metrics are ordered as RMSE / LPIPS

Method	Short Range Consistency↓		Long Range Consistency↓	
	Rain	Snow	Rain	Snow
GNT–(All-in-one) Restore	0.138 / 0.253	0.099 / 0.211	0.249 / 0.402	0.188 / 0.312
GAURA	0.115 / 0.215	0.083 / 0.189	0.204 / 0.333	0.153 / 0.299

C Results

Real-World Data. We provide a summary of the quantitative results against the GNT-all-in-one baselines in Table 1. In the main paper, we select the best-performing all-in-one model and present both qualitative and quantitative results accordingly. In this section, we delve into further details regarding the selected best method. For the *low-light* enhancement task, we utilize AirNet+GNT as the baseline. In the cases of *motion deblurring* and *dehazing* tasks, DA-CLIP+GNT serves as the baseline for comparison. Similarly, for rain and snow removal, we employ DA-CLIP+GNT as the comparative baseline for both qualitative and quantitative analyses. Regarding defocus, we utilize the state-of-the-art Single Pixel Defocus Deblur model [1]. We present a collection of results showcasing our method's performance on real-world data across several types of corruptions in Fig. 3

LLFF-Corrupted. Results pertaining to LLFF-Corrupted scenes are presented in Figure 5. We compare our method against two all-in-one approaches across the five degradations. It is evident that our method effectively restores the appearance details of the scenes while preserving their geometry simultaneously.

D Blind Restoration

Despite our method's current requirement for the degradation type as input, it is feasible to extend it to achieve complete independence from the degradation type. This form of restoration, without the need for specifying the degradation type, is referred to as Blind Restoration. To accomplish this, we propose training a convolutional network capable of taking degraded images as input and classifying the type of degradation. This network can be supervised using a cross-entropy loss function. For instance, utilizing a ResNet-18 backbone, we achieved 99.5%

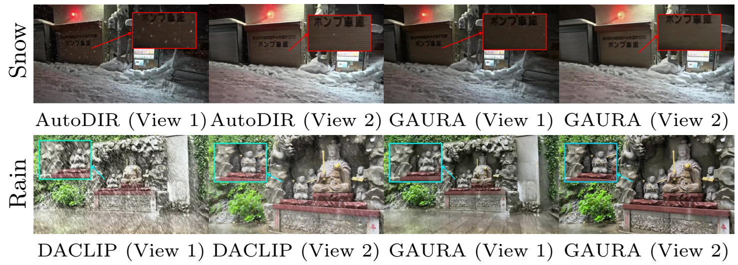


Fig. 2: We compare several restoration techniques' multi-view consistency against GAURA. We observe that in the rain and snow scene, the restoration from our model results in view consistent restoration, while the baseline restores the scene inconsistently.

accuracy in degradation type classification. Once trained, this network can be utilized to predict the degradation type directly from input images, rendering the restoration process independent of user input regarding degradation type.

E Additional Ablation Studies

E.1 Simple Conditioning vs ARM+DLM(Ours)

In Table 2, we compare our ARM+DLM module against a simpler variant that concatenates the latent conditioning to the inputs of the cross and self-attention blocks in the view and ray transformers respectively. Our proposed modules outperform the baseline across all metrics.

E.2 Effect of Number of Input Source Views

In Table 3, we measure the effect of the number of input views and observe a minimal drop in performance ($< 4\%$) with as little as 3 views. This indicates that our learned transformer modules are sufficiently robust to noisy epipolar input.

F Multi-View Consistency

In Table 4, we present the short range and long range consistency evaluated across the generated multi-views. We see that GAURA can render view-consistent clear images from arbitrary viewing angles, superior to other baselines. Along with quantitative result, we present qualitative results in Fig. 2 which clearly shows the superiority of GAURA over other baselines in terms of multi-view consistent restoration.

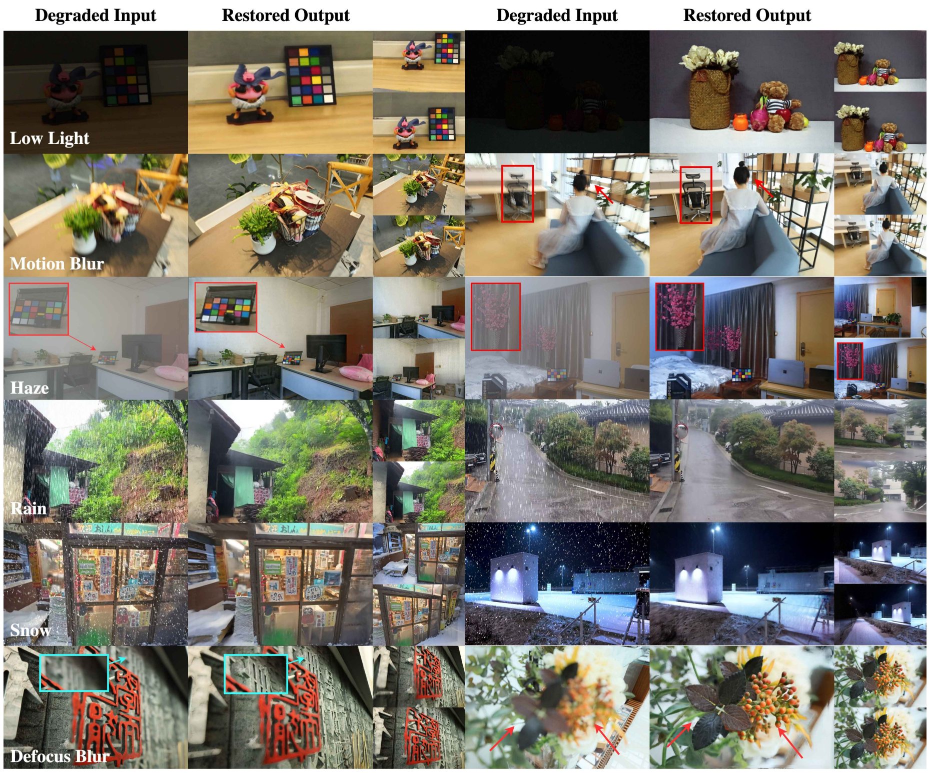


Fig. 3: Below are the gallery results showcasing our method's performance on various corruptions on real-world data. Each row presents results of our method on a single corruption across two scenes. The degraded image and its corresponding restored output are visualized side-by-side. Additionally, renders from two other novel viewpoints are provided on the right.

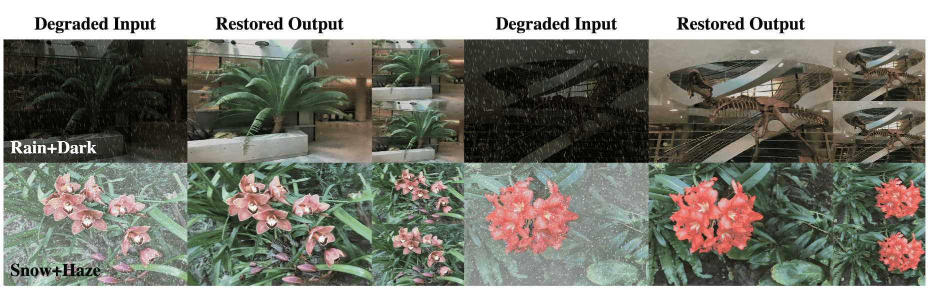


Fig. 4: Results on scenes corrupted with more than 1 degradations. We show results on scenes corrupted with Rain+Dark and Snow+Haze on the synthetically corrupted LLFF dataset.

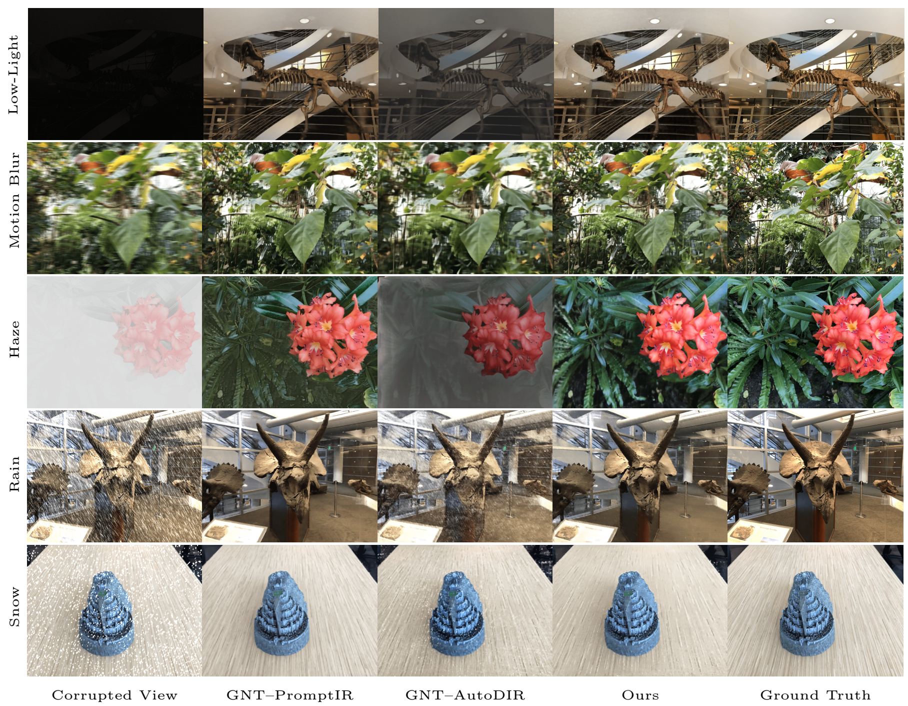


Fig. 5: Results on the Synthetic LLFF-Corrupted data are presented, comparing against GNT-PromptIR [6] and GNT-AutoDIR [2]. Qualitative comparisons across five corruptions are shown, illustrating our method's consistent ability to restore the scene across all degradation types.

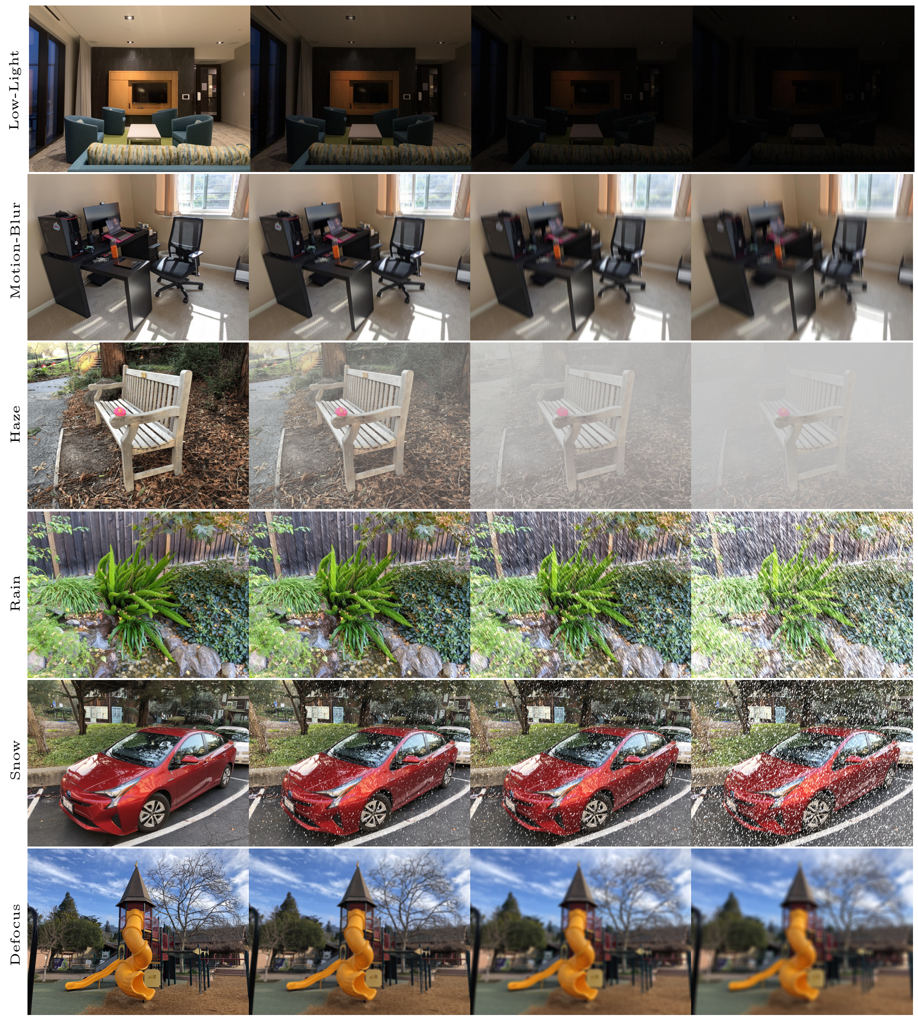


Fig. 6: We show visualisation of images degraded with varying intensities. The first column corresponds to the clean image, and starting from the second column, the images are degraded from low to high intensities. This approach enables us to capture a range of randomness in the degradation of the scene, facilitating easy generalization to real-world data.

References

1. Cui, Y., Tao, Y., Bing, Z., Ren, W., Gao, X., Cao, X., Huang, K., Knoll, A.: Selective frequency network for image restoration. In: The Eleventh International Conference on Learning Representations (2023) 3
2. Jiang, Y., Zhang, Z., Xue, T., Gu, J.: Autodir: Automatic all-in-one image restoration with latent diffusion. arXiv preprint arXiv:2310.10123 (2023) 6
3. Koschmieder, H.: Theorie der horizontalen sichtweite. Beitrage zur Physik der freien Atmosphare pp. 33–53 (1924) 1
4. Lamba, M., Kumar, M., Mitra, K.: Real-time restoration of dark stereo images. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 4914–4924 (2023) 1
5. Ma, L., Li, X., Liao, J., Zhang, Q., Wang, X., Wang, J., Sander, P.V.: Deblurnerf: Neural radiance fields from blurry images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12861–12870 (2022) 2
6. Potlapalli, V., Zamir, S.W., Khan, S., Khan, F.S.: Promptir: Prompting for all-in-one blind image restoration. arXiv preprint arXiv:2306.13090 (2023) 6
7. Varma, M., Wang, P., Chen, X., Chen, T., Venugopalan, S., Wang, Z.: Is attention all that nerf needs? In: The Eleventh International Conference on Learning Representations (2023) 1