A Appendix

A.1 Method Preliminaries

Image Restoration. In our image restoration module of FLARE, we employ the Swin Transformer (SwinIR) for enhancing low-quality input images $I_{LR} \in \mathbb{R}^{H\times W\times C_{in}}$ where H represents the image height, W signifies the image width, and C_{in} corresponds to the number of input channels [26]. This begins by extracting shallow features $F_0 \in \mathbb{R}^{H\times W\times C}$ through a 3x3 convolutional layer, which performs visual processing and feature mapping. Subsequently, deep features $F_{DF} \in \mathbb{R}^{H\times W\times C}$ are derived from F_0 using the deep feature extraction module. During image reconstruction, the high-quality image I_{HR} is reconstructed by combining shallow and deep features via the reconstruction module H_{REC} as $I_{HR} = H_{REC}(F_0 + F_{DF})$. The image restoration process is defined as,

$$I_{HR} = SwinIR(I_{LR}) = \mathcal{F}_{swin}(\mathcal{E}_{swin}(I_{LR})) + Loss \tag{7}$$

where $\mathcal{E}_{swin}(.)$ refers to the encoder that extracts features from the input image, while $\mathcal{F}_{swin}(.)$ represents the decoder responsible for reconstructing the output image from these features. The *Loss* function quantifies the difference between the input and output images produced by the SwinIR model as $Loss(I_{LR}, I_{HR})$.

Diffusion Models. Diffusion models offer multi-modal image generation through two distinct stages: the forward noise process and the reverse denoising process. In the forward process, a sequence $x_1, x_2, ..., x_T$ is generated from a starting point x_0, drawn from the distribution $p(x_0)$, by iteratively adding noise. This process relies on the equation $q(\mathbf{x}_t \mid \mathbf{x}_0) = N\left(\mathbf{x}_t; \alpha_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}\right)$. Here, each step introduces a noise component ε, sampled from the standard normal distribution. Conversely, the denoising process models the transition from x_t to x_{t-1} through the conditional probability $p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$. In this stage, the predicted statistics, $\hat{\mu}_\theta(\mathbf{x}_t), \hat{\Sigma}_\theta(\mathbf{x}_t)$, are determined, guided by a learnable parameter θ. Optimization occurs through a loss function $\ell^t_{\text{simple}}(\theta)$ [31] quantifying the difference between actual and predicted noise, leveraging a learnable neural network. The trained neural network, known as the predictor, can then generate an image $\hat{\mathbf{x}}_0$. We employed the UniDiffuser diffusion model [5]for handling various data types, allowing effective generation of image-text and image-image pairs without added complexity. UniDiffuser, designed for text-to-image tasks, is represented as,

$$\text{UniDiffuser}(prompt) = F(\mathcal{T}(prompt), \mathcal{G}(image)) \tag{8}$$

where $\mathcal{T}(prompt)$ and $\mathcal{G}(image)$ stand for the text encoder and image generator, respectively. The fusion module F efficiently combines features from the text encoder and image generator, focusing on producing perceptually realistic results.

A.2 Additional Experimental Details

Implementation Details. We implemented our double-stage FLARE method on a single NVIDIA A100 40GB GPU. In the first stage, we used SwinIR [26] (2021) for LR-to-HR image conversion due to its effectiveness in image restoration. In the second stage, we generated K=4 high-resolution images using standard augmentations such as RandomFlip and

ColorJitter. For weighted aggregation of the samples from $\tilde{D}^{HR}_{Aug} + \tilde{D}^{HR}_{T2I}$ (See Eq. 6), we employed α=0.50 and β=1.00. Finally, we trained the classifiers across all datasets and class settings for 25 epochs using the Adam optimizer with a learning rate of 2e-5.

Downstream Datasets. In-domain tasks represent generalization performance when training and testing are done on the same dataset, while out-of-domain tasks represent performance when testing is done on a downstream dataset different from the one used for training.

Baselines. To assess the effectiveness of our proposed approach, we compare it against two groups of methods in astronomical classification. The first group involves classification performed directly on extracted images D^{LR} (referred to as *Raw_LR*), while the second group involves augmented versions of the raw images D^{LR}_{Aug} (referred to as *Raw_Aug_LR*). For the *Raw_LR* group, our initial baseline is established by [8]. Moving to the *Raw_Aug_LR* group, we consider [29] as our second baseline, as it demonstrates the effectiveness of data augmentation techniques in image classification.

Classifiers. We utilized state-of-the-art pre-trained CNN models to train our custom cosmos dataset, including ResNet-50 [17], GoogleNet [38], DenseNet-121 [21], and ViT-B/16 [12], which were initially trained on the ImageNet dataset [11]. Our primary task involved classifying two distinct datasets: one for fine-grained data with 8 classes and the other for macro-level categories with 4 classes. Through these extensive experiments on various ConvNets, we aimed to assess how well these models could classify data in different situations, highlighting their flexibility and adaptability.

Metrics. We use two essential metrics for assessing the quality of HR images derived from LR versions. The first metric, Peak Signal-to-Noise Ratio (PSNR) [41], measures image noise, with higher values indicating improved image quality. The second metric, Multi-Scale Structural Similarity (MS-SSIM) [20], evaluates structural and textural fidelity, with higher MS-SSIM values signifying better preservation of intricate details in HR images compared to LR. Furthermore, for classification models, we employed standard metrics, including Accuracy, F1-Score, Precision, and Recall, to assess the performance of the classifiers across all dataset variants [2].

A.3 Additional Results

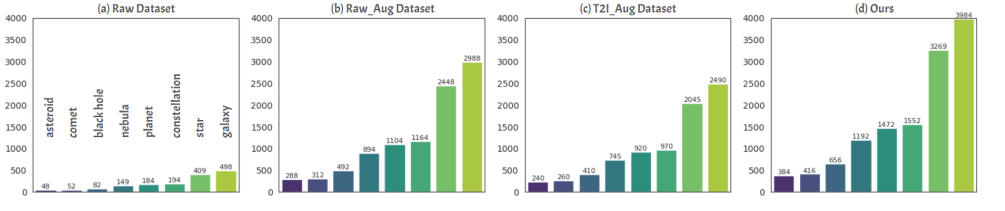

Figure 7: The original raw dataset (Raw_Aug), when transformed into our combined dataset using the FLARE approach, results in 7.8× increase in the number of samples. **Ours** represent the proposed **SpaceNet** dataset.

Table 5: Quantitative assessment of 4 classifiers across different methodologies for Macro and Fine-grained classes stating **in-domain** F1-Scores. **FLARE** indicate models trained on our **SpaceNet** dataset, where SpaceNet is combined with $\alpha = 0.5$ and $\beta = 1.0$ (Using Eq. 5 and 6). **Average** indicates average performance across all classifiers.

Data Type	Method	ResNet-50 [27]	GoogleNet [33]	DenseNet-121 [21]	ViT-B/16 [12]	Average
		F1-Score	F1-Score	F1-Score	F1-Score	F1-Score
Macro	Raw_LR [8] (bs.)	68.83(bs.)	67.38(bs.)	67.92(bs.)	72.00(bs.)	69.03(bs.)
	Raw_Aug_LR [28]	74.57(5.74) ↑	75.19(7.81) ↑	76.04(8.12) ↑	76.03(4.03) ↑	75.45(6.42) ↑
	Raw_Aug_HR	78.21(9.38) ↑	77.68(10.30) ↑	79.57(11.65) ↑	80.39(8.39) ↑	78.96(9.93) ↑
	T2I_Aug_HR	80.54(11.71) ↑	80.88(13.50) ↑	80.80(12.88) ↑	81.69(9.69) ↑	80.97(11.94) ↑
	FLARE (Ours)	**87.69(18.86)** ↑↑	**84.67(17.29)** ↑↑	**85.70(17.78)** ↑↑	**86.50(14.50)** ↑↑	**86.14(17.11)** ↑↑
Fine-grained	Raw_LR [8] (bs.)	55.80(bs.)	54.16(bs.)	55.62(bs.)	58.19(bs.)	55.94(bs.)
	Raw_Aug_LR [28]	66.58(10.78) ↑	67.21(13.05) ↑	67.42(11.80) ↑	66.43(8.24) ↑	66.91(10.97) ↑
	Raw_Aug_HR	70.53(14.73) ↑	69.89(15.73) ↑	72.53(16.91) ↑	72.44(14.25) ↑	71.35(15.41) ↑
	T2I_Aug_HR	77.19(21.39) ↑	76.05(21.89) ↑	77.06(21.44) ↑	77.02(18.83) ↑	76.83(20.89) ↑
	FLARE (Ours)	**83.58(27.78)** ↑↑	**81.07(26.91)** ↑↑	**83.60(27.98)** ↑↑	**82.58(24.39)** ↑↑	**82.71(26.77)** ↑↑

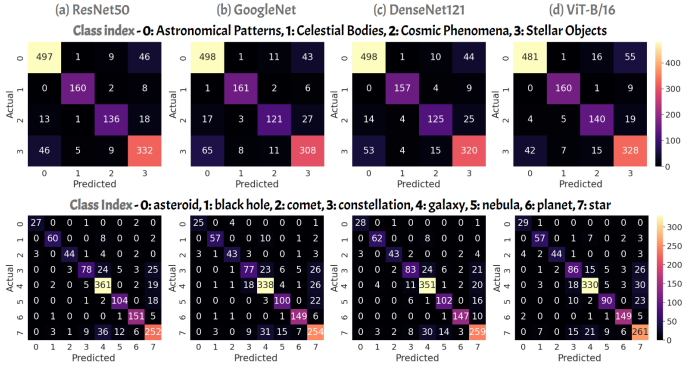

Figure 8: Confusion Matrix of 4 best models on our combined SpaceNet dataset

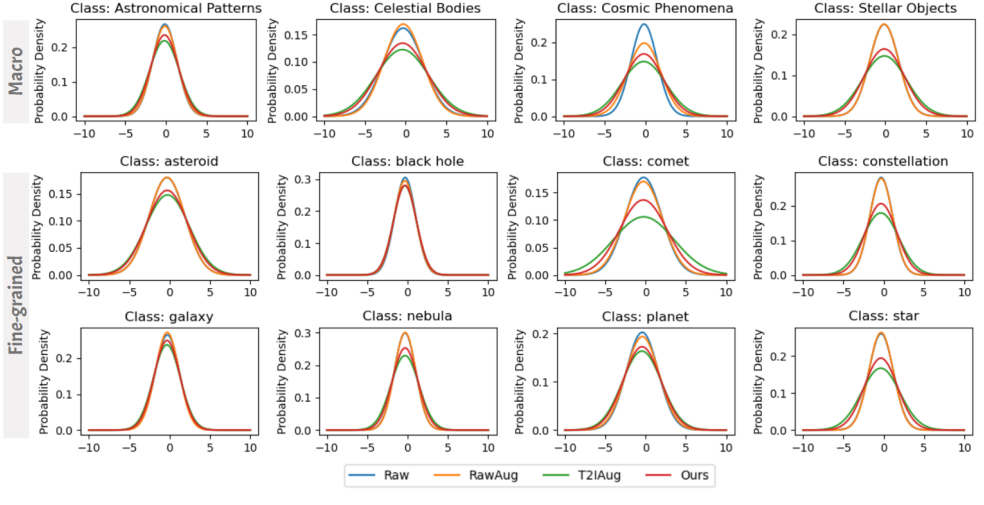

Figure 9: Normal distribution representing Mean and Variances across different dataset variants.

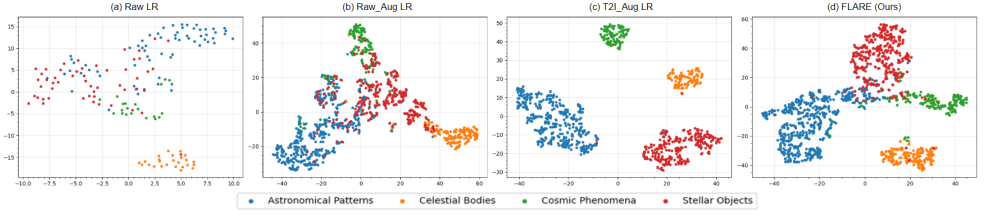

Figure 10: t-sne plots representing the distribution of features across different dataset variants. representing 4 macro classes.

Table 6: Notations describing different datasets.

Notation	Dataset Description
D^{LR}	Raw_LR, *i.e.*, Raw Lower Resolution images
D^{LR}_{Aug}	Raw_Aug_LR, *i.e.*, Raw Lower Resolution images and augmentations
D^{HR}_{Aug}	Raw_Aug_HR, *i.e.*, Raw Higher Resolution images
D^{HR}_{T2I}	T2I_Aug_HR, *i.e.*, Syntheic samples in Higher Resolution
$\tilde{D}^{HR}$	SpaceNet (Ours)