RARE-AWARE AUTOENCODING: RECONSTRUCTING SPATIALLY IMBALANCED DATA

Anonymous authors
Paper under double-blind review

ABSTRACT

Autoencoders are widely used for unsupervised and self-supervised learning when labels are scarce, but their reconstruction objectives can be biased by spatially imbalanced image data, favoring dominant background or frequent pixel patterns. This issue is common in scientific imaging including medical imaging, biology, physics, and astronomy, where relevant image content may appear sparsely or non-uniformly across spatial locations, and manual annotations are limited or expensive. We propose a reconstruction framework with two components: a self-information-based loss that upweights statistically uncommon pixel intensities conditioned on fixed spatial locations, and Sample Propagation, a replay strategy that re-exposes the model to hard-to-reconstruct samples across batches. Unlike most data-imbalance methods, which rely on task-specific annotations, our approach operates in the unsupervised setting. Nevertheless, we also adapt and benchmark methods from supervised learning. Experiments on a controlled simulated dataset and three real-world datasets spanning physical, biological, and astronomical domains show improved reconstruction over existing baselines, particularly in settings with strong spatial imbalance. Code and data will be released.

1 INTRODUCTION

We study small-data imaging problems in the natural sciences, where acquiring and annotating data is often expensive. In this regime, self-supervised and unsupervised learning are particularly useful because they reduce the dependence on manual labels. Autoencoders are a natural choice for such settings: they learn representations by reconstructing their inputs and therefore require no task-specific annotations. They are also widely used across denoising (Vincent et al., 2008), anomaly detection (An & Cho, 2015), latent dynamics modeling (Watter et al., 2015), generative modeling (Kingma & Welling, 2014), and clustering (Makhzani et al., 2015), making them relevant for unlabeled scientific imaging datasets (Vincent et al., 2008; Bengio et al., 2013).

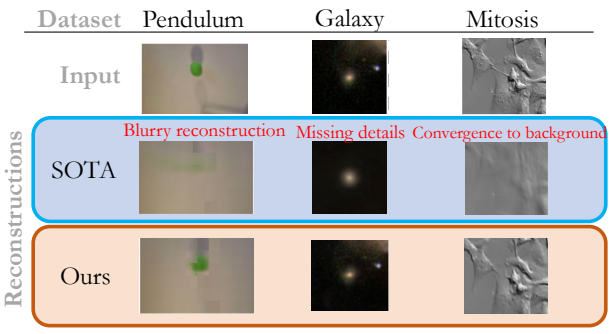


Figure 1: **Failure cases in image reconstruction under spatial imbalance.** *(Top)* Input frames. *(Middle)* Representative failure modes of baseline reconstruction methods induced by spatial imbalance, including blurry reconstructions, missing details, and convergence toward the background. *(Bottom)* Reconstructions obtained with our method, which mitigates these failure modes to preserve the underlying spatial structure and fine-scale details.

Here, we focus on autoencoder reconstruction when image content is unevenly distributed across fixed spatial locations in the training set. This setting is common in scientific imaging. For example, astronomical sources may appear with different orientations or offsets, cells may occupy different positions across microscopy fields of view, and objects in physical videos may visit some sensor locations much more often than others. Such variation can arise from acquisition constraints (Sun

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

et al., 2019; Ericsson et al., 2022; Zheng et al., 2017), data from a single source (Finn et al., 2017), or continuous streams such as videos (Srivastava et al., 2015; Zhang et al., 2018b).

We refer to this phenomenon as *spatial data imbalance*: a training condition in which fixed spatial locations receive unequal exposure to the range of pixel values or image content that the model is expected to reconstruct. This imbalance is especially problematic in small datasets, where location-wise rare, but important, pixel patterns may contribute little to the aggregate reconstruction objective. As a result, an autoencoder trained with standard losses can minimize average reconstruction error while failing to preserve rare spatial configurations (Fig. 1), often producing blurred images biased toward frequent appearances. Figure 2 illustrates this effect in a damped pendulum dataset: the object appears in every frame, but some spatial locations are visited much less frequently than others.

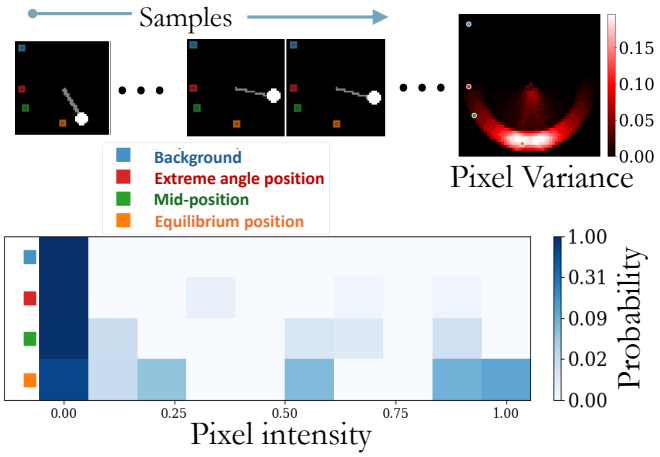


Figure 2: **Spatial data imbalance.** *Top:* Example frames from a damped pendulum dataset at different angles. Because the pendulum spends more time near equilibrium than at extreme angles, some spatial configurations appear much more frequently than others. *Top-right:* Per-pixel variance over the dataset, showing motion-induced spatial variability. *Bottom:* Selected-pixel intensity histograms show rare activations at extreme angles and more frequent changes near commonly visited regions.

Data imbalance is well studied in supervised learning, where it is usually defined over class labels and addressed through class reweighting, resampling, or augmentation (Zhang et al., 2023; Alvis & Seneviratne, 2024; Yang et al., 2022; Pintea et al., 2023). In unsupervised reconstruction, however, labels are unavailable and imbalance may arise at the level of location-wise pixel values rather than semantic classes, such that supervised balancing methods don't directly apply. Moreover, while spatial augmentations such as cropping, translation or oversampling can reduce foreground-background bias; they require information about which transformations or regions should be emphasized, and they do not explicitly estimate the rarity of pixel intensities per location. This motivates reconstruction objectives that can emphasize rare location-wise pixel patterns without prior knowledge.

We address spatial imbalance in autoencoder reconstruction with two complementary mechanisms inspired by supervised imbalance learning. First, analogous to class-reweighting (Cao et al., 2019; Lin et al., 2017), we introduce a location-conditioned self-information reconstruction loss (SI). The loss estimates the distribution of pixel intensities in the training batch and assigns larger weights to low-probability pixel values. Thus, the method emphasizes statistically uncommon intensity values conditioned on location, rather than relying on semantic labels or explicit object annotations. Second, inspired by hard-example and prioritized replay strategies (Schaul et al., 2016; Wu et al., 2016; Gu et al., 2020), we propose Sample Propagation (SPP), which re-exposes the model to hard-to-reconstruct samples across training batches. These mechanisms are architecture-agnostic and can be applied to different autoencoder variants without model modifications. Our main contributions are:

- We formulate spatial data imbalance as a reconstruction challenge for autoencoders.
- We propose a self-information reconstruction loss that emphasizes statistically uncommon pixel-intensity values at fixed spatial locations, reducing convergence toward mean appearances and blurred reconstructions.
- We introduce Sample Propagation (SPP), a hard-sample replay mechanism that increases exposure to difficult reconstruction cases during training.
- We benchmark data-balancing strategies and reconstruction losses for unsupervised autoencoder reconstruction across four AE architectures and three real-world datasets.

2 RELATED WORK

Image reconstruction in representation learning. Autoencoders depend on the latent space to capture relevant image structures (Srivastava et al., 2015; Hsieh et al., 2018; Jaques et al., 2020; Adiban et al., 2025). But when data distributions are biased or imbalanced, traditional loss functions often compress too aggressively or average out rare patterns. This weakens the learned representation and leads to uninformative reconstructions (Srivastava et al., 2015; Buda et al., 2018; Geirhos et al., 2020). Nevertheless, previous work often does not discuss spatial imbalance specifically.

Reconstruction challenges due to data statistics. In domains like natural sciences, data from a single video often exhibits a spatial capture bias and repetitive patterns (Jaques et al., 2020; Oakden-Rayner, 2020; Puyol-Antón et al., 2021; Garcia et al., 2025). This can lead models to prioritize frequent patterns while overlooking transient or spatially rare events, resulting in blur or averaging in regions with high pixel variance (Guo et al., 2023; Ercan et al., 2024; Yang et al., 2025). This imbalance leads to two main challenges. First is the averaged reconstructions, where the AE averages out or blurs details in locations with rare pixel values. Second is unbalanced samples, where the images with underrepresented rare locations are not often seen during training.

Image blurring. On image reconstruction, focused losses propose plug-in alternatives which do not require architectural changes. These include pixel-level (L1, L2), perceptual (VGG (Zhang et al., 2018a), Watson (Czolbe et al., 2020)), frequency-aware (FFL (Jiang et al., 2021)) and class-balanced (LDAM (Cao et al., 2019), Focal (Lin et al., 2017)) losses. Similar, Bredell et al. (Bredell et al., 2023) introduced a blur-aware loss (BEL) for variational autoencoders (VAEs) targeting high-frequency errors while preserving ELBO maximization. While these methods offer diverse strategies, they have not been compared and evaluated under sample imbalance. Here, we provide such an evaluation in comparison to our method, to properly formalize the spatial imbalance in AE reconstruction.

Weighted and adaptive reconstruction objectives. Several reconstruction methods assign non-uniform importance to pixels, patches, or tokens. PatchVAE (Gupta et al., 2020) emphasizes high-gradient patches in VAEs, Hard Patches Mining (Wang et al., 2023) selects difficult patches for masked image modeling, and AdaMAE (Bandara et al., 2023) adaptively selects tokens for masked autoencoder pretraining. These works are relevant precedents, but they target different objectives, architectures, or patch/token selection mechanisms that do not transfer directly across standard AE variants. We therefore focus our comparison on architecture-agnostic losses and balancing strategies applicable to AE reconstruction. Unlike prior adaptive reconstruction methods, our loss weights pixel-intensity values according to batch-level rarity conditioned on fixed spatial locations, targeting spatial imbalance rather than high-gradient, masked-token difficulty, or reconstruction error alone.

Imbalanced sample distributions. Data imbalance is extensively studied in supervised classification (Zhang et al., 2023; Pintea et al., 2023), typically through class reweighting, focal or margin-based losses (Lin et al., 2017; Cao et al., 2019), and resampling. Hard-example methods such as Online Hard Example Mining (Shrivastava et al., 2016), Prioritized Experience Replay (Schaul et al., 2016), hard-pixel mining (Wu et al., 2016; Gu et al., 2020), online batch selection (Loshchilov & Hutter, 2015), curriculum learning (Bengio et al., 2009), and selection via proxy (Coleman et al., 2020) further show that emphasizing difficult samples can improve learning. However, many of these methods rely on labels, supervised losses, or task-specific definitions of hardness, and their role in unsupervised AE reconstruction remains less explored. We study how such loss- and sample-level balancing ideas transfer to spatially imbalanced reconstruction, where imbalance appears in location-wise pixel statistics rather than class labels.

3 METHODS

Self-information loss (SI)**.** Standard reconstruction losses weight all pixels uniformly. Under spatial data imbalance, this makes the objective dominated by location-wise frequent pixel patterns, while rare pixel-intensity values may be averaged out. We therefore define a self-information loss that emphasizes statistically uncommon pixel-intensity values at fixed spatial locations.

Let $X, \hat{X} \in [0,1]^{B\times C\times H\times W}$ be a batch and its reconstruction. We flatten channel and spatial dimensions into locations $\ell \in \{1,\dots,L\}$, with $L = CHW$, so that $x_{b,\ell}$ denotes sample b at fixed location ℓ. For each location, we estimate the empirical batch-level intensity distribution with a J-bin

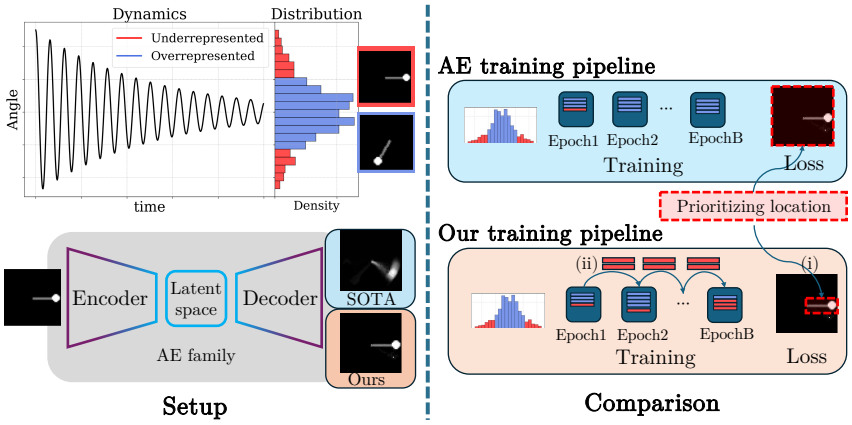


Figure 3: **Reconstruction under spatial data imbalance.** Illustration on a simulated pendulum dataset, where some pendulum positions occur frequently while others are rare. Black squares denote input or reconstructed frames. **Left-top:** The input distribution is imbalanced with frequent (blue) and rare (red) samples. **Left-bottom:** A standard AE backbone reconstructs frames through an encoder–latent–decoder pipeline. Baseline reconstructions (cyan), tend to produce smooth outputs biased toward frequent configurations. **Right:** Our method (orange), combines: (i) a *self-information loss* (dashed, red lines) that upweights statistically rare pixel-intensity values at fixed spatial locations, and (ii) *Sample Propagation* (red rectangles) that re-exposes the model to high-loss samples across batches. Together, these components improve reconstruction of rare pendulum configurations.

histogram. With $k_{b,\ell} = \lfloor x_{b,\ell}(J-1) \rfloor$, bin counts $c_{\ell,k} = \sum_{b=1}^{B} \mathbf{1}[k_{b,\ell} = k]$, and bin width $\delta = 1/J$, the density estimate and self-information score are

$$\widehat{p}_\ell(x_{b,\ell}) \approx \frac{c_{\ell,k}/B}{\delta} = \frac{J}{B} c_{\ell,k}, \qquad s_{b,\ell} = -\log(\widehat{p}_\ell(x_{b,\ell}) + \varepsilon). \tag{1}$$

High $s_{b,\ell}$ indicates that the pixel-intensity value $x_{b,\ell}$ is rare at location ℓ within the current batch.

We reshape s to (B, C, H, W) and apply per-image, per-channel min–max normalization,

$$\tilde{s}_{b,c,h,w} = \frac{s_{b,c,h,w} - \min_{h,w} s_{b,c,h,w}}{\max_{h,w} s_{b,c,h,w} - \min_{h,w} s_{b,c,h,w} + \varepsilon} \in [0,1]. \tag{2}$$

This normalization makes the weighting relative within each image and channel. Although min–max scaling is sensitive to extreme values, this behavior is aligned with our goal of emphasizing high-surprisal pixel values; we further avoid suppressing common pixels by adding a floor α. The final weight is $w_{b,c,h,w} = \tilde{s}_{b,c,h,w} + \alpha$, with $\alpha = 0.05$ and the self-information loss $\mathcal{L}_{\text{SI}}$ is given by:

$$\text{SI}_b = \frac{1}{CHW} \sum_{c,h,w} w_{b,c,h,w} |\hat{X}_{b,c,h,w} - X_{b,c,h,w}|, \quad \text{MSE}_b = \frac{1}{CHW} \sum_{c,h,w} (\hat{X}_{b,c,h,w} - X_{b,c,h,w})^2,$$
$$\ell_b = \text{MSE}_b + \lambda \, \text{SI}_b, \quad \mathcal{L}_{\text{SI}} = \tfrac{1}{B} \textstyle\sum_{b=1}^{B} \ell_b. \tag{3}$$

The MSE term preserves global reconstruction structure, while the weighted L1 term emphasizes statistically uncommon pixel-intensity values at fixed spatial locations; λ balances the two terms.

Sample Propagation (SPP). The self-information loss emphasizes statistically uncommon pixel-intensity values at fixed spatial locations within the current batch, but samples containing these rare location-wise patterns may still appear too infrequently to consistently influence training. We therefore introduce Sample Propagation (SPP), a hard-sample replay mechanism that propagates difficult reconstruction cases across consecutive batches.

Let B be the nominal batch size and M the *propagation memory size*, i.e., the maximum number of hard samples carried to the next training step. At step i, we concatenate the current batch X_i, with $|X_i| = B$, and the previous memory P_{i-1}, with $|P_{i-1}| \leq M$, forming the candidate pool $Q_i = X_i \cup P_{i-1}$. We compute the per-sample reconstruction loss ℓ_b for all $b \in Q_i$, rank samples by loss, use the B hardest samples $G_i = \text{TopB}(Q_i; \ell)$ for the gradient update, and store the M hardest samples $P_i = \text{TopM}(Q_i; \ell)$ for the next step. Thus, SPP increases the candidate pool from

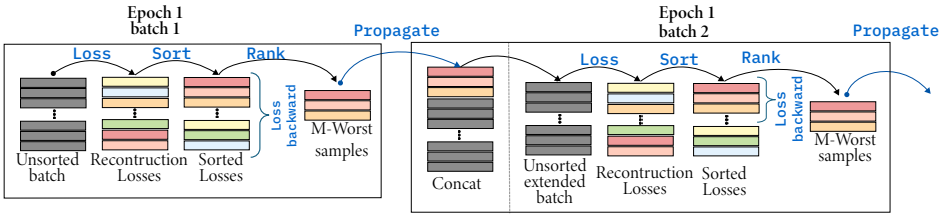


Figure 4: **Sample Propagation.** At step i, the current batch X_i is concatenated with the previous memory. Samples are ranked by reconstruction loss (red loss is higher); the top B samples are used for the update, and the top M samples are propagated to the next batch.

B to at most $B + M$, while keeping the number of gradient-contributing samples fixed at B. In the experiments, SPPk denotes $M = B/k$; for example, SPP4 propagates $B/4$ hard samples. Sample Propagation is illustrated in Fig. 4.

Hard-sample weighting. After selecting G_i, we optionally apply a focal-style weight. For $b \in G_i$, define $\ell_{\min} = \min_{b \in G_i} \ell_b$, $\ell_{\max} = \max_{b \in G_i} \ell_b$, and $f_b = 1 + \frac{\ell_b - \ell_{\min}}{\max(\ell_{\max} - \ell_{\min}, \varepsilon)}(B - 1)$. The final objective is $\mathcal{L}_{\text{SPP}} = \frac{1}{B} \sum_{b \in G_i} f_b \ell_b$. The self-information loss acts at the pixel/location level, whereas SPP and hard-sample weighting act at the sample level.

Training overhead is reported in Appendix A.6; inference cost is unchanged because the autoencoder architecture and parameter count remain fixed.

4 Experiments

We evaluate the proposed self-information loss SI and Sample Propagation (SPP) on autoencoder reconstruction under spatial data imbalance. Within each experiment, all methods are compared using the same backbone architecture and training protocol. We select shared training hyperparameters and tune method-specific parameters to obtain strong performance across baselines without favoring a single method. Architecture details, optimization settings, hyperparameter ranges, and selected values are reported in detail in the appendix.

We report Mean Squared Error (MSE↓), Peak Signal-to-Noise Ratio (PSNR↑), and Structural Similarity Index (SSIM↑), following prior work (Mustafa et al., 2022; Bredell et al., 2023). We evaluate four autoencoder families: a fully connected autoencoder (AE) (Chen & Guo, 2023), a variational autoencoder (VAE) (Kingma & Welling, 2014; Chen & Guo, 2023), a sparse autoencoder (SAE) (Chen & Guo, 2023), and a masked autoencoder (MAE) (He et al., 2022).

We compare against two groups of baselines. The first group modifies the reconstruction objective, including pixel losses (L1/L2), cross-entropy (CE), SSIM (Wang et al., 2004), perceptual losses based on VGG features (Zhang et al., 2018a), Watson loss (Czolbe et al., 2020), focal frequency loss (FFL) (Jiang et al., 2021), LDAM (Cao et al., 2019), and blur-aware loss (BEL) (Bredell et al., 2023). The second group modifies sample weighting or selection, including focal loss (Lin et al., 2017), curriculum learning (DCL) (Bengio et al., 2009), online batch selection (ONBS) (Loshchilov &

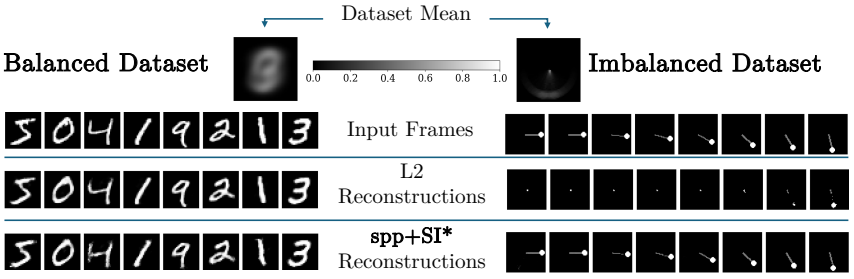
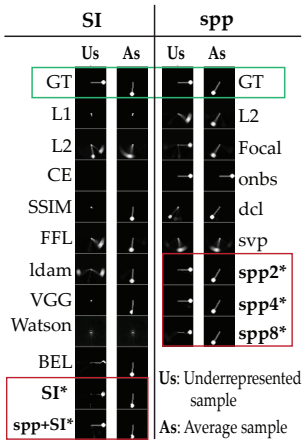


Figure 5: **Balanced and spatially imbalanced reconstruction settings.** MNIST (*left*) provides a reference case with relatively broad spatial coverage, while the damped pendulum dataset (*right*) induces spatial imbalance because some pendulum positions occur much less frequently than others. The standard MLP AE performs well on MNIST, but its reconstructions degrade for rare pendulum configurations. Our SPP + SI method improves reconstruction in the imbalanced setting.



Method	Baseline	MSE	PSNR	SSIM
SI	L1	$1.507e^{-2} \pm 3.60e^{-4}$	$1.822e^{1} \pm 1.04e^{-1}$	$8.644e^{-1} \pm 1.39e^{-2}$
	L2	$7.778e^{-3} \pm 5.30e^{-3}$	$2.251e^{1} \pm 3.87e^{0}$	$8.585e^{-1} \pm 7.18e^{-2}$
	CE	$2.446e^{-1} \pm 4.12e^{-4}$	$6.116e^{0} \pm 7.32e^{-3}$	$3.598e^{-2} \pm 4.05e^{-3}$
	SSIM	$6.265e^{-3} \pm 7.34e^{-3}$	$3.054e^{1} \pm 1.10e^{1}$	$9.537e^{-1} \pm 5.34e^{-2}$
	FFL	$3.124e^{-3} \pm 2.62e^{-3}$	$2.589e^{1} \pm 2.43e^{0}$	$9.259e^{-1} \pm 4.52e^{-2}$
	ldam	$1.640e^{-3} \pm 2.02e^{-3}$	$2.990e^{1} \pm 4.08e^{0}$	$9.473e^{-1} \pm 5.57e^{-2}$
	VGG	$6.559e^{-3} \pm 6.86e^{-3}$	$3.094e^{1} \pm 1.42e^{1}$	$9.553e^{-1} \pm 4.85e^{-2}$
	Watson	$1.572e^{-2} \pm 3.03e^{-4}$	$1.804e^{1} \pm 8.39e^{-2}$	$8.794e^{-1} \pm 1.17e^{-2}$
	BEL	$2.069e^{-4} \pm 8.24e^{-4}$	$\mathbf{5.002\,e^{1} \pm 8.84\,e^{0}}$	$\mathbf{9.966\,e^{-1} \pm 9.35\,e^{-3}}$
	SI*	$6.596e^{-4} \pm 1.74e^{-3}$	$4.234e^{1} \pm 8.93e^{0}$	$9.917e^{-1} \pm 1.84e^{-2}$
	spp+SI*	$\mathbf{1.547\,e^{-4} \pm 4.76\,e^{-4}}$	$4.533e^{1} \pm 6.69e^{0}$	$9.960e^{-1} \pm 6.81e^{-3}$
spp	L2	$1.735e^{-3} \pm 2.32e^{-3}$	$2.952e^{1} \pm 3.64e^{0}$	$9.553e^{-1} \pm 4.61e^{-2}$
	Focal	$1.624e^{-3} \pm 3.95e^{-4}$	$2.800e^{1} \pm 9.19e^{-1}$	$9.501e^{-1} \pm 2.62e^{-2}$
	onbs	$2.367e^{-2} \pm 2.58e^{-3}$	$1.631e^{1} \pm 8.48e^{-1}$	$8.027e^{-1} \pm 2.34e^{-2}$
	dcl	$2.318e^{-3} \pm 4.54e^{-3}$	$3.072e^{1} \pm 5.41e^{0}$	$9.458e^{-1} \pm 7.29e^{-2}$
	svp	$1.095e^{-2} \pm 4.05e^{-3}$	$1.989e^{1} \pm 1.55e^{0}$	$8.045e^{-1} \pm 4.09e^{-2}$
	spp2*	$\mathbf{2.282\,e^{-4} \pm 1.03\,e^{-3}}$	$\mathbf{4.620\,e^{1} \pm 7.24\,e^{0}}$	$\mathbf{9.956\,e^{-1} \pm 9.46\,e^{-3}}$
	spp4*	$2.228e^{-3} \pm 4.59e^{-3}$	$3.681e^{1} \pm 9.91e^{0}$	$9.870e^{-1} \pm 1.86e^{-2}$
	spp8*	$1.216e^{-3} \pm 3.32e^{-3}$	$4.376e^{1} \pm 1.09e^{1}$	$9.828e^{-1} \pm 4.45e^{-2}$

Figure 6: ***(Left)* Qualitative comparison on controlled data.** The first row shows the ground truth (GT). Rows compare baseline losses, the proposed self-information loss (SI*), the combined method (SPP + SI*), and SPP variants with different memory sizes. Columns show both underrepresented samples (Us) and average samples (As), illustrating behavior on rare and frequent configurations.

Table 1: ***(Right)* Reconstruction performance on controlled data.** We report MSE↓, PSNR↑, and SSIM↑ as mean±std over test samples. The top block compares reconstruction losses, and the bottom block sample-selection strategies. * denotes our methods; best mean value in each block is in bold.

Hutter, 2015), and selection via proxy (SVP) (Coleman et al., 2020). Together, these baselines cover reconstruction fidelity, perceptual quality, frequency emphasis, imbalance handling, and sample-level prioritization.

In practice, we found a fixed propagation memory M to be more stable than a dynamically varying memory size. We therefore evaluate SPP using fixed fractions of the batch: SPP2 uses $M = B/2$, SPP4 uses $M = B/4$, and SPP8 uses $M = B/8$. When a method is denoted simply as SPP + SI, we use SPP4, which represents a middle memory setting among the evaluated configurations.

4.1 Experiments on controlled data

We first use a controlled damped-pendulum dataset (Garcia et al., 2025). The pendulum appears in every frame, but damping makes some spatial configurations much more frequent than others, inducing location-wise imbalance. We also use MNIST as a reference setting with less pronounced spatial-location imbalance. Figure 5 compares a standard MLP autoencoder with our SPP + SI method. On the pendulum data, the standard AE tends to reconstruct frequent spatial configurations while failing to preserve rare ones. In contrast, SPP + SI better preserves rare configurations while maintaining good reconstruction on common samples.

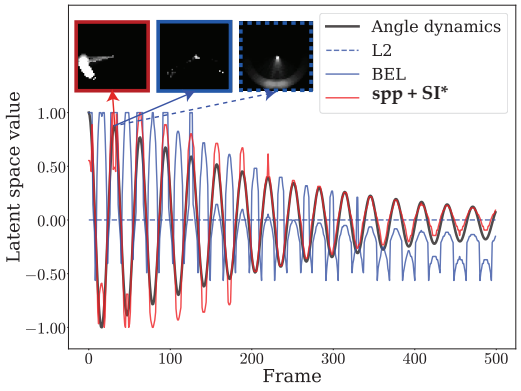
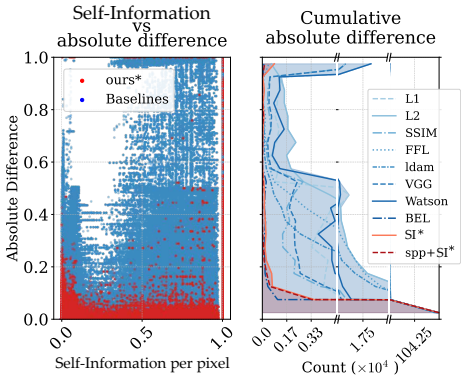
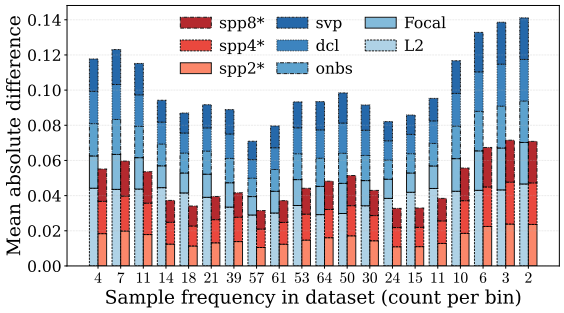


Figure 7: **Restricted latent-capacity comparison.** We constrain the AE latent space to one dimension and visualize the decoded trajectory across time. Under this bottleneck, L2 tends to collapse, BEL partially captures the oscillation, and SPP + SI better preserves the time-varying structure.

Figure 6 and Table 1 compare our method with the reconstruction and sample-level baselines introduced above. On the controlled pendulum data, the proposed methods improve reconstruction of rare configurations while preserving performance on frequent samples. The self-information loss reduces the tendency to average rare location-wise pixel patterns into frequent appearances, whereas SPP increases exposure to hard samples across batches. BEL performs strongly in this synthetic setting, likely because the images contain high-contrast edges; however, this advantage is less consistent on real data with more varied structures and textures.



(a) **Per-pixel error vs. SI.** Scatter-density plots show absolute reconstruction error against per-pixel self-information. Our methods concentrate more mass at low reconstruction error, for moderate- and high-self-information observations.



(b) **Error vs. sample frequency.** Mean absolute error is shown across binned sample-frequency groups. SPP variants reduce the error on low-frequency samples while keeping performance stable on frequent samples.

Figure 8: **Analysis on controlled data.** The self-information loss reduces reconstruction error for high self-information location-wise samples, while SPP reduces error on low-frequency samples.

To test whether the proposed loss acts on the intended failure mode, we analyze reconstruction error as a function of location-wise self-information in Fig. 8a. The proposed objective reduces error not only on average but also for high-self-information pixel values, where standard losses often leave larger residuals. Similarly, Fig. 8b groups samples by occurrence frequency and shows that SPP reduces the error gap between frequent and infrequent samples.

Finally, Fig. 7 visualizes a restricted 1D latent space. This experiment is not intended as a complete representation-learning evaluation, but as a diagnostic for whether the reconstruction objective preserves the dominant temporal structure under a severe bottleneck. With L2, the latent trajectory collapses toward an averaged solution; BEL partially captures the oscillatory signal; and SPP + SI better follows the temporal variation while producing more faithful reconstructions.

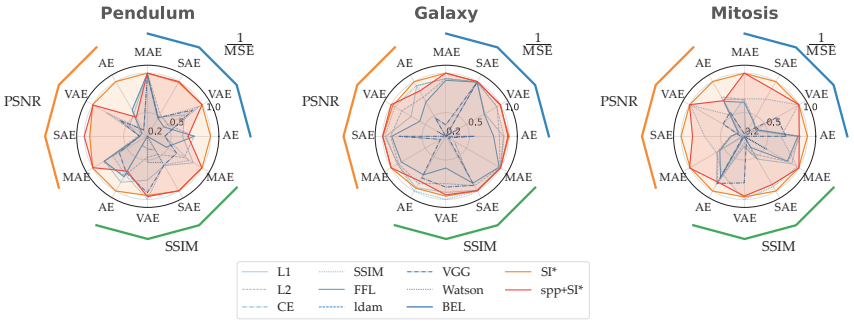
4.2 Real data experiments

We evaluate on three real datasets with different sources of spatial or event-frequency imbalance. Because the proposed self-information loss estimates location-conditioned intensity rarity, we use these datasets to test whether this mechanism remains useful beyond the controlled pendulum setting.

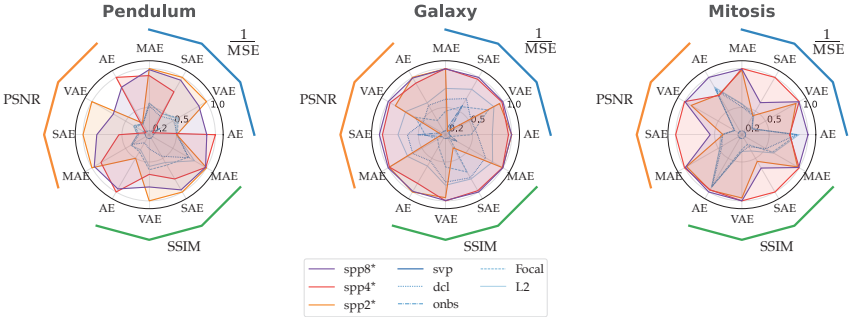
Pendulum. A real-world pendulum video (Garcia et al., 2025), where static background and frequently visited ball positions dominate the frame distribution, while some positions occur less often due to damping. *GalaxyZoo.* Galaxy images with diverse morphologies (Lintott et al., 2011), where centered objects and dark background dominate over infrequent off-centre structures, faint arms, or nearby sources like stars. *Mitosis.* A time-lapse microscopy sequence from the MOTChallenge CTMC-v1 dataset (Anjum & Gurari, 2020). Most frames contain common interphase cell appearances, while mitotic events are rare. Additional dataset details are reported in Appendix A.4.

Self-information loss. We evaluate the proposed SI loss and its combination with SPP across the datasets and backbones introduced above. Quantitative results are summarized in Fig. 9a and reported in detail in Appendix A.7, Tables 10- 12, together with higher resolution images.

Across the three datasets, self-information loss improves distortion metrics MSE and PSNR consistently, while remaining competitive in SSIM. The largest gains occur when the reconstruction objective is dominated by frequent background or repeated states. For example, in *Pendulum-SAE* and *Galaxy-AE*, our method substantially reduces MSE and improves PSNR, indicating that location-conditioned self-information reduces convergence toward oversmoothed appearances. Although some baselines obtain slightly higher SSIM in specific configurations, our method achieves better distortion metrics, suggesting a favorable trade-off between structural similarity and pixel-level fidelity. On *Mitosis-VAE*, the proposed loss also preserves subtle cellular structure while lowering reconstruction distortion, showing that the benefit is not limited to the synthetic pendulum setting. Qualitatively, Fig. 10a supports this interpretation: baseline losses often blur pendulum trajectories, smooth galaxy arms, and soften cellular boundaries, whereas our methods retain sharper traces, finer galaxy structure, and more defined cell morphology. We do not interpret these results as evidence



(a) **Reconstruction results comparing different losses.** The proposed self-information loss improves MSE and PSNR in most configurations and remains competitive in SSIM.



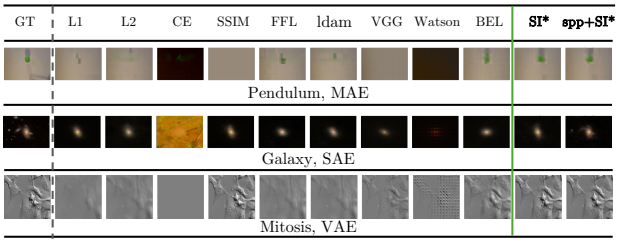
(b) **Reconstruction results comparing sample mining methods.** SPP variants improve average reconstruction quality in several settings and reduce sensitivity to low-frequency or hard-to-reconstruct samples.

Figure 9: **Real-data reconstruction performance across datasets and autoencoder families.** We summarize results on three real datasets and four autoencoder variants using normalized MSE, PSNR, and SSIM scores, with all axes oriented so that higher values indicate better performance. Thus, a larger radar area indicates stronger aggregate reconstruction performance across the reported metrics. Our methods, shown in shades of red, generally exhibit improved overall reconstruction quality. The radar plots provide a compact visual summary; complete raw numerical results are reported in Appendix Tables 10–14.

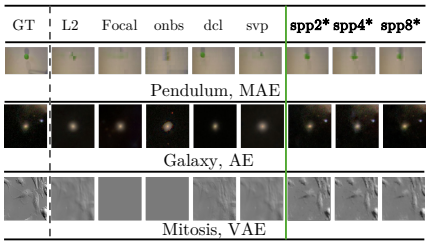
of semantic awareness; the method operates through location-conditioned intensity rarity, and the observed gains show that this signal is useful in spatially imbalanced reconstruction settings.

Sample propagation. We next evaluate Sample Propagation as a sample-level extension of the self-information loss. In all SPP experiments, the reconstruction objective uses the self-information loss as the backbone; SPP changes only the sample exposure by carrying hard-to-reconstruct samples across batches. Figure 9b summarizes the quantitative results, while Fig. 10b shows qualitative examples. Full tables and higher resolution images are provided in Appendix A.7.

Across datasets, SPP improves reconstruction in several configurations, but the memory size depends on the dataset and architecture. On *Pendulum*, larger memories were better, consistent with the controlled setting. On *Galaxy Zoo* and *Mitosis*, smaller memories such as SPP4/8 are often more



(a) **Reconstruction losses.** Standard losses often produce smoother reconstructions, while our methods preserve more visible structure in rare or high-detail regions.



(b) **Sample Propagation.** SPP often preserves sharper structures than baselines, especially for hard-to-reconstruct examples.

Figure 10: **Qualitative reconstruction comparisons.** Representative samples from Pendulum, Galaxy Zoo, and Mitosis are shown with ground truth and reconstructions. Samples are randomly drawn; for clarity, we show the architecture where the effect is most visually apparent.

stable, suggesting that replay is useful but excessive reuse of high-loss samples can reduce batch diversity, supporting the role of SPP as a controlled replay mechanism rather than simple oversampling.

Quantitatively, SPP mainly improves MSE and PSNR, indicating reduced distortion and less averaging toward frequent appearances. Qualitatively, Fig. 10b shows sharper pendulum traces, clearer galaxy structures, and better-defined cellular boundaries compared with focal, ONBS, DCL, and SVP baselines. These results suggest that the two components act at different levels: the self-information loss emphasizes rare location-wise observations within each image, while SPP increases the training exposure of samples that remain difficult to reconstruct.

4.3 DOWNSTREAM TASKS

Finally, we examine if the proposed training strategy transfers to downstream tasks. These experiments are intended as supporting evidence rather than a complete evaluation of task-specific performance.

We consider three settings: image generation following BEL (Bredell et al., 2023), anomaly detection following AESc (Collin & De Vleeschouwer, 2021), and pendulum parameter estimation following Delfys (Garcia et al., 2025). For generation, we report FID and LPIPS; for anomaly detection, image and pixel AUROC; and for parameter estimation, absolute error in frequency and damping.

Table 2: **Downstream task performance.** Preliminary evidence that **SPP + SI*** can benefit downstream tasks under the corresponding baseline protocols.

Task	Model	Metrics	
Generation	BEL	FID ↓ 280.5	LPIPS ↓ 0.3467 ± 0.0405
	SPP + SI*	**221.8**	$\mathbf{0.2826} \pm 0.0512$
Anomaly detection	AESc	Image AUROC ↑ 0.6017	Pixel AUROC ↑ 0.7853
	SPP + SI*	**0.6247**	**0.7875**
Pendulum parameter est.	Delfys	Abs. Err. freq. ↓ 0.0586	Abs. Err. damp. ↓ 0.064
	AE	1.0016	3.32
	SPP + SI*	**0.0516**	**0.002**

Table 2 shows that our method improves over the corresponding reference models in these settings. However, we treat these results as indicative rather than conclusive: downstream performance depends on task-specific protocols, encoder usage, and downstream tuning. The main evidence for the proposed method therefore remains the controlled and real-data reconstruction experiments above.

5 DISCUSSION

We studied spatial data imbalance in unsupervised autoencoder reconstruction, where fixed spatial locations receive unequal exposure to the range of appearances present in the training data. This imbalance can bias standard reconstruction losses toward frequent observations, leading to blurred or majority-like outputs for rare configurations. Our results show that reconstruction under such conditions can be improved without changing the model architecture, by simply combining location-conditioned self-information weighting with sample-level replay.

The self-information loss emphasizes low-probability intensity observations at fixed spatial locations, reducing the tendency of AEs to average out location-wise rare details. Sample Propagation (SPP) complements this by increasing the exposure of hard-to-reconstruct samples across batches. Together, these components improve reconstruction quality across controlled and real-world datasets, particularly in settings dominated by frequent background or repeated states. From a broader-impact perspective, improved unsupervised reconstruction may benefit scientific and medical imaging settings with limited annotations, but emphasizing statistically rare pixel patterns may also amplify rare noise or acquisition artifacts.

Limitations and future work. Our method estimates rarity from location-conditioned intensity statistics, and therefore does not directly model semantic structure. Rare structures whose intensities are not statistically uncommon may not be emphasized by the proposed loss. In addition, the best SPP memory size is dataset-dependent, and the current evaluation does not exhaustively cover all domain-specific reconstruction settings. Future work should study more robust normalization schemes, richer measures of spatial imbalance, and downstream protocols that directly test whether improved reconstruction yields better representations for task-specific scientific analysis.

AI DISCLOSURE

Generative AI tools were used only to improve grammar, readability, and wording, and to assist with basic LaTeX formatting. They were not used to generate references, develop research ideas, design experiments, interpret results, or write complete sections of the paper.

For code, AI tools were used only to help identify bugs in author-written code, not to generate the research code itself.

All AI-assisted content was reviewed by the authors, who take full responsibility for the final paper.

REFERENCES

Mohammad Adiban, Kalin Stefanov, Sabato Marco Siniscalchi, and Giampiero Salvi. S-hr-vqvae: Sequential hierarchical residual learning vector quantized variational autoencoder for video prediction. *IEEE Transactions on Multimedia*, 2025. URL https://doi.org/10.1109/tmm.2025.3535370.

Charika De Alvis and Suranga Seneviratne. A survey of deep long-tail classification advancements. *arXiv*, 04 2024. doi: 10.48550/arxiv.2404.15593. URL https://arxiv.org/abs/2404.15593.

Jinwon An and Sungzoon Cho. Variational autoencoder based anomaly detection using reconstruction probability. In *Special Lecture on IE*, volume 2, pp. 1–18, 2015.

Samreen Anjum and Danna Gurari. CTMC-v1: Cell Tracking with Mitosis Detection Dataset Challenge. https://motchallenge.net/data/CTMC-v1/, 2020. Accessed: 2025-04-29.

Wele Gedara Chaminda Bandara, Naman Patel, Ali Gholami, Mehdi Nikkhah, Motilal Agrawal, and Vishal M. Patel. Adamae: Adaptive masking for efficient spatiotemporal learning with masked autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pp. 41–48, 2009.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 35, pp. 1798–1828, 2013.

Gustav Bredell, Kyriakos Flouris, Krishna Chaitanya, Ertunc Erdil, and Ender Konukoglu. Explicitly minimizing the blur error of variational autoencoders. *Published as a conference paper at ICLR*, 2023. arXiv:2304.05939.

Mateusz Buda, Atsuto Maki, and Maciej A Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural networks*, 106:249–259, 2018.

Kaidi Cao, Colin Wei, Adrien Gaidon, Nuno Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Shuangshuang Chen and Wei Guo. Auto-encoders in deep learning—a review with new perspectives. *Mathematics*, 11(8):1777, 2023.

Cody Coleman, Edward Yeh, Stephen Mussmann, Baharan Mirzasoleiman, Peter Bailis, Percy Liang, Jure Leskovec, and Matei Zaharia. Selection via proxy: Efficient data selection for deep learning. In *ICLR*, 2020.

Anne-Sophie Collin and Christophe De Vleeschouwer. Improved anomaly detection by training an autoencoder with skip connections on images corrupted with stain-shaped noise. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 7915–7922. IEEE, 2021.

Steffen Czolbe, Oswin Krause, Ingemar Cox, and Christian Igel. A loss function for generative neural networks based on watson's perceptual model. In *Advances in Neural Information Processing Systems*, volume 33, pp. 2051–2061, 2020.

Burak Ercan, Onur Eker, Canberk Saglam, Aykut Erdem, and Erkut Erdem. Hypere2vid: Improving event-based video reconstruction via hypernetworks. *IEEE Transactions on Image Processing*, 33: 1826–1837, 2024.

Linus Ericsson, Henry Gouk, Chen Change Loy, and Timothy M. Hospedales. Self-supervised representation learning: Introduction, advances, and challenges. *Institute of Electrical and Electronics Engineers*, 39(3):42–62, 05 2022. doi: 10.1109/msp.2021.3134634. URL https://doi.org/10.1109/msp.2021.3134634.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pp. 1126–1135. PMLR, 2017.

Castañeda Alejandro Garcia, Jan van Gemert, Daan Brinks, and Nergis Tömen. Learning physics from video: Unsupervised physical parameter estimation for continuous dynamical systems. *CVPR*, 2025.

Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard Zemel, Wieland Brendel, Matthias Bethge, and Felix A Wichmann. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2(11):665–673, 2020.

Zhangxuan Gu, Li Niu, Haohua Zhao, and Liqing Zhang. Hard pixel mining for depth privileged semantic segmentation. *arXiv preprint arXiv:1906.11437*, 2020. URL https://arxiv.org/abs/1906.11437.

Guangsha Guo, Yang Feng, Hengyi Lv, Yuchen Zhao, Hailong Liu, and Guoling Bi. Event-guided image super-resolution reconstruction. *Sensors*, 23(4):1234–1250, 2023.

Kamal Gupta, Saurabh Singh, and Abhinav Shrivastava. Patch-vae: Learning local latent codes for recognition. in 2020 ieee. In *CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4745–4754, 2020.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16000–16009, 2022.

Jun-Ting Hsieh, Bingbin Liu, De-An Huang, Li F Fei-Fei, and Juan Carlos Niebles. Learning to decompose and disentangle representations for video prediction. *Advances in neural information processing systems*, 31, 2018.

Miguel Jaques, Michael Burke, and Timothy Hospedales. Physics-as-inverse-graphics: Unsupervised physical parameter estimation from video. *ICLR*, 2020.

Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. Focal frequency loss for image reconstruction and synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 13919–13929, 2021.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *ICLR*, 01 2014.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

Chris Lintott, Kevin Schawinski, Steven Bamford, Anže Slosar, Kate Land, Daniel Thomas, Edd Edmondson, Karen Masters, Robert C Nichol, M Jordan Raddick, et al. Galaxy zoo 1: data release of morphological classifications for nearly 900 000 galaxies. *Monthly Notices of the Royal Astronomical Society*, 410(1):166–178, 2011.

Ilya Loshchilov and Frank Hutter. Online batch selection for faster training of neural networks. *arXiv preprint arXiv:1511.06343*, 2015.

Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial autoencoders. In *arXiv preprint arXiv:1511.05644*, 2015.

Aamir Mustafa, Aliaksei Mikhailiuk, Dan Andrei Iliescu, Varun Babbar, and Rafał K Mantiuk. Training a task-specific image reconstruction loss. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 2319–2328, 2022.

Luke Oakden-Rayner. Exploring large-scale public medical image datasets. *Academic radiology*, 27 (1):106–112, 2020.

Silvia Pintea, Yancong Lin, Jouke Dijkstra, and Jan van Gemert. A step towards understanding why classification helps regression. *ICCV*, 2023. URL https://arxiv.org/abs/2308.10603.

Esther Puyol-Antón, Bram Ruijsink, Stefan K Piechnik, Stefan Neubauer, Steffen E Petersen, Reza Razavi, and Andrew P King. Fairness in cardiac mr image analysis: an investigation of bias due to data imbalance in deep learning based segmentation. In *Medical Image Computing and Computer Assisted Intervention(MICCAI)*, pp. 413–423. Springer, 2021.

Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. Prioritized experience replay. In *International Conference on Learning Representations (ICLR)*, 2016. URL https://arxiv.org/abs/1511.05952.

Abhinav Shrivastava, Abhinav Gupta, and Ross Girshick. Training region-based object detectors with online hard example mining. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

Nitish Srivastava, Elman Mansimov, and Ruslan Salakhudinov. Unsupervised learning of video representations using lstms. In *International conference on machine learning*, pp. 843–852. PMLR, 2015.

Chen Sun, Fabien Baradel, Kevin Murphy, and Cordelia Schmid. Learning video representations using contrastive bidirectional transformer. *arxiv-Cornell University*, 01 2019. doi: 10.48550/arXiv.1906. URL https://arxiv.org/abs/1906.05743.

Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pp. 1096–1103, 2008.

Haochen Wang, Kaiyou Song, Junsong Fan, Yuxi Wang, Jin Xie, and Zhaoxiang Zhang. Hard patches mining for masked image modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004.

Manuel Watter, Jost Tobias Springenberg, Joschka Boedecker, and Martin Riedmiller. Embed to control: A locally linear latent dynamics model for control from raw images. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 2746–2754, 2015.

Zifeng Wu, Chunhua Shen, and Anton van den Hengel. Bridging category-level and instance-level semantic image segmentation. *arXiv preprint arXiv:1605.06885*, 2016. URL https://arxiv.org/abs/1605.06885. Introduces online bootstrapping (hard pixel mining) for segmentation.

Boyu Yang, Yusen Liao, and Jun Ke. Reconstruction of moving small targets through scattering media: A hierarchical network approach integrating event information. *Optics and Lasers in Engineering*, 189:108944, 2025.

Lu Yang, He Jiang, Qing Song, and Jun Guo. A survey on long-tailed visual recognition. *Springer Science+Business Media*, 130(7):1837–1872, 05 2022. doi: 10.1007/s11263-022-01622-8. URL https://doi.org/10.1007/s11263-022-01622-8.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 586–595, 2018a.

Xiaofeng Zhang, Rong Wang, and Jianwei Ding. Abnormal event detection by learning spatiotemporal features in videos, 01 2018b. URL https://doi.org/10.1007/978-981-13-1702-6_42.

Yifan Zhang, Bingyi Kang, Bryan Hooi, Shuicheng Yan, and Jiashi Feng. Deep long-tailed learning: A survey. *IEEE Computer Society*, 45(9):10795–10816, 04 2023. doi: 10.1109/tpami.2023.3268118. URL https://doi.org/10.1109/tpami.2023.3268118.

Peng Zheng, Aleksandr Y Aravkin, Karthikeyan Natesan Ramamurthy, and Jayaraman Jayaraman Thiagarajan. Learning robust representations for computer vision. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 1784–1791, 2017.

A Appendix

A.1 Spatial imbalance, statistical rarity, and semantic rarity

In this section, we clarify how spatial data imbalance differs from other common forms of imbalance, and what we mean by "rare" in our method.

Sample-level and class-level imbalance. In supervised learning, data imbalance is usually defined at the level of labels or samples. For example, in a classification dataset, one class may contain many fewer examples than another class. This is a *class-level* imbalance: the minority class is rare because its label occurs infrequently. Similarly, in sample-level imbalance, some types of images or videos may appear less often than others. These settings are commonly addressed using class reweighting, resampling, or label-guided augmentation.

Spatial data imbalance is different. It can occur even when every sample contains the same object type or belongs to the same class. The issue is not necessarily that a semantic category is rare, but that fixed spatial locations in the image receive unequal exposure to the range of pixel values or image content that the model is expected to reconstruct. For example, in a damped pendulum sequence, the pendulum may be present in every frame, but some pixel locations are activated by the pendulum only rarely. Thus, the object itself is not rare at the dataset level, but its appearance at particular spatial locations is rare.

Foreground-background imbalance. Spatial imbalance is related to, but not identical to, foreground-background imbalance. Foreground-background imbalance describes the dominance of background pixels over foreground structures, which is common in medical, biological, astronomical, and physical imaging. However, foreground-background imbalance does not specify *where* image content occurs across the dataset. Spatial imbalance instead concerns the empirical distribution at fixed image coordinates: some locations repeatedly contain background or common appearances, while others only rarely contain informative pixel patterns. For example, in Galaxy Zoo images Lintott et al. (2011), faint stars or nearby sources may be treated as background for morphology classification, yet they can still be useful for other tasks such as artifact analysis, source detection, or studying imaging conditions. Thus, what is considered "background" is task-dependent, and rare location-wise pixel patterns should not be dismissed solely because they are not part of the main foreground object. This location-wise perspective is important for reconstruction losses, because each pixel's gradient contribution depends on how often that location exhibits different pixel values during training.

Statistical rarity in our method. In this paper, "rare" is used in a statistical sense. For each fixed spatial location ℓ, we estimate the empirical batch-level distribution of pixel intensities. A pixel-intensity value $x_{b,\ell}$ is statistically rare if it has low probability under this location-conditioned distribution. Our self-information score,

$$s_{b,\ell} = -\log(\widehat{p}_\ell(x_{b,\ell}) + \varepsilon),$$

therefore assigns larger values to low-probability pixel intensities at the same spatial location. This is a measure of *location-conditioned intensity rarity*, not a semantic measure of object importance.

Semantic rarity. Semantic rarity refers to rare objects, events, classes, or structures, such as a rare disease pattern, a rare galaxy morphology, or a rare cellular event. Semantic rarity usually requires labels, annotations, domain knowledge, or a representation that captures object-level meaning. Our method does not directly detect semantic rarity. Instead, it emphasizes pixel-intensity values that are statistically uncommon at fixed spatial locations. This distinction is important: a semantically rare structure may not have rare pixel intensities, and a statistically rare pixel value may not correspond to a semantically meaningful structure. Therefore, our method should be understood as targeting spatially conditioned pixel-statistical imbalance, rather than semantic rarity.

Why this distinction matters. The proposed self-information loss is most appropriate when rare image content is reflected in uncommon pixel-intensity values at specific spatial locations, as in settings with sparse foreground structures, repeated backgrounds, or non-uniform spatial sampling. It may be less effective when rare semantic structures have intensities similar to common structures, or when the data are not aligned in a shared coordinate system. Sample Propagation partially

complements this limitation by replaying hard-to-reconstruct samples, but it also operates through reconstruction difficulty rather than semantic labels. Accordingly, we interpret our method as an architecture-agnostic way to mitigate statistical spatial imbalance in unsupervised reconstruction, not as a general detector of rare semantic content.

A.2 Mathematical details

Probability calculation.

Pixel intensities are continuous on $[0, 1]$. Thus, we define a histogram over a continuous variable, which estimates a *density* (units: inverse intensity), not a discrete probability mass. Normalizing counts by the batch size, c/B, yields a *frequency* for a bin of width δ; converting to a density requires dividing by δ:

$$\widehat{p} \approx \frac{c/B}{\delta} = \frac{J}{B}\, c.$$

Dividing by B makes the estimate invariant to batch size; multiplying by $J = 1/\delta$ removes the dependence on the chosen bin width. Using a PMF-like c/B alone would shrink as bins get narrower (larger J), making $-\log(\cdot)$ incomparable across choices of J. The PDF-based surprise $s = -\log \widehat{p}$ is therefore bin-width–invariant and consistently highlights statistically rare (high-surprise) pixels.

A.3 Training details

All experiments were run on an NVIDIA A40 GPU. Unless otherwise stated, models were trained for 1000 epochs with batch size $B = 64$ using the Adam optimizer, learning rate 10^{-3}, and weight decay 10^{-5}. We did not use early stopping or learning-rate scheduling. Within each experiment, all methods were evaluated using the same backbone architecture, data split, preprocessing, batch size, number of epochs, and optimizer settings. Shared training hyperparameters were selected using the same validation protocol for all methods. Method-specific hyperparameters, such as the self-information trade-off coefficient λ, histogram bin count J, smoothing floor α, and SPP memory size M, were tuned separately when applicable. The selected values and search ranges are reported in the corresponding experiment sections below.

A.4 Dataset details

Table 3 summarizes the datasets used in the real-data experiments, including image resolution, sample counts, splits, and the source of spatial or event-frequency imbalance. All images are normalized to $[0, 1]$. Unless otherwise stated, models are trained on the training split only, hyperparameters are selected on the validation split, and final metrics are reported on the test split.

Pendulum. We use a real-world pendulum video Garcia et al. (2025) resized to $3 \times 72 \times 128$ pixels. The sequence contains a mostly static background and recurring ball positions. Due to damping, the pendulum spends more time near equilibrium than at extreme angles, producing unequal location-wise exposure across the image grid. This makes the dataset suitable for evaluating reconstruction under spatial imbalance.

Galaxy Zoo. We use galaxy images from Galaxy Zoo Lintott et al. (2011), resized to $3 \times 128 \times 128$ pixels. Images contain a mostly dark background and galaxies with diverse morphologies. Most objects are approximately centered, while off-center structures, faint arms, nearby sources, and other fine details occur less frequently. This dataset tests whether location-conditioned reconstruction weighting remains useful in a less controlled astronomical setting.

Mitosis. We use a time-lapse microscopy sequence from the MOTChallenge CTMC-v1 dataset Anjum & Gurari (2020), resized to $3 \times 128 \times 128$ pixels. The sequence contains common interphase cell appearances under relatively stable imaging conditions, while mitotic events occur briefly and are therefore underrepresented in time. Compared with Pendulum, this dataset is better described as containing event-frequency imbalance, rather than purely fixed-coordinate spatial imbalance.

Table 3: Dataset statistics and imbalance sources for the real-data experiments. Spatial imbalance refers to unequal exposure of fixed spatial locations to the range of appearances observed in the training set. Event-frequency imbalance refers to rare temporal events that occur only briefly within an image sequence.

Dataset	Resolution	Samples	Imbalance source
Synthetic Pendulum	$3 \times 64 \times 64$	500	Damping induces frequent near-equilibrium positions and rare extreme positions
Pendulum	$3 \times 72 \times 128$	644	Damping induces frequent near-equilibrium positions and rare extreme positions
Galaxy Zoo	$3 \times 128 \times 128$	1000	Mostly centered galaxies and dark background; off-center structures and fine details are less frequent
Mitosis	$3 \times 128 \times 128$	1200	Common interphase cell appearances; mitotic events are temporally rare

Table 4: **MLP Autoencoder architecture.** The input image of size (c, h, w) is flattened to dimension $c{\cdot}h{\cdot}w$, where c is the number of channels and h, w are the spatial dimensions. The encoder compresses the input into a latent vector of dimension l. The reduction factors d_i denote successive proportions used to shrink the layer dimensionality relative to the input size. The decoder mirrors this process to reconstruct.

Stage	**Layer**	**In**	**Out**	**Output Shape**
		Encoder		
1	Linear	$c\,h\,w$	$c\,h\,w/d_1$	$(\frac{c\,h\,w}{d_1})$
	ReLU			$(\frac{c\,h\,w}{d_1})$
2	Linear	$c\,h\,w/d_1$	$c\,h\,w/d_2$	$(\frac{c\,h\,w}{d_2})$
	ReLU			$(\frac{c\,h\,w}{d_2})$
3	Linear	$c\,h\,w/d_2$	l	(l)
		Decoder		
1	Linear	l	$c\,h\,w/d_2$	$(\frac{c\,h\,w}{d_2})$
	ReLU			$(\frac{c\,h\,w}{d_2})$
2	Linear	$c\,h\,w/d_2$	$c\,h\,w/d_1$	$(\frac{c\,h\,w}{d_1})$
	ReLU			$(\frac{c\,h\,w}{d_1})$
3	Linear	$c\,h\,w/d_1$	$c\,h\,w$	(c, h, w)
	Sigmoid			(c, h, w)

Preprocessing. For all datasets, pixel intensities are scaled to $[0, 1]$. Images are resized to the resolutions reported in Table 3. No label information is used during autoencoder training. For datasets derived from temporal sequences, train/validation/test splits are constructed without using future test frames during training.

A.5 Model Architectures

Across all experiments, we relied on the architectures detailed in Table 4, 5, 6, and 7; which define the AE, VAE, SAE, and MAE networks respectively. While the overall structure remained consistent, the layer dimensionality and size were adapted to the specific dataset or method.

Specification of Individual Experiments As discussed in the main paper, certain architectures have sufficient capacity to overfit the training set, thereby obscuring meaningful differences between reconstruction loss functions. To mitigate this effect and to more clearly expose the advantages of our approach, we systematically adjust dataset parameters in each experiment. The detailed configurations are reported here to ensure reproducibility and to provide a transparent account of how

Table 5: **Variational Autoencoder (VAE) architecture.** The input (c, h, w) is encoded through three convolutional *stages*, where each stage corresponds to a resolution-changing convolution (stride 2) followed by non-linearities. After flattening, fully connected layers reduce the representation to a latent distribution of dimension l, parameterized by mean (μ) and variance (σ). Channel sizes n_1, n_2, n_3 are the number of feature maps at each stage, and (h_i, w_i) are the spatial dimensions after stage i. The decoder mirrors this process with three transposed convolution *stages*, reconstructing the image back to (c, h, w). Intermediate fully connected layers are scaled by proportional factors d_1 and d_2 relative to n_3.

Stage	Layer	In	Out	Kernel / Stride / Pad	Output Shape
Encoder					
1	Conv2d	c	n_1	$3 \times 3,\ s = 2,\ p = 1$	(n_1, h_1, w_1)
	ReLU				(n_1, h_1, w_1)
2	Conv2d	n_1	n_2	$3 \times 3,\ s = 2,\ p = 1$	(n_2, h_2, w_2)
	ReLU				(n_2, h_2, w_2)
3	Conv2d	n_2	n_3	$3 \times 3,\ s = 2,\ p = 1$	(n_3, h_3, w_3)
	ReLU				(n_3, h_3, w_3)
	Flatten				$(n_3 h_3 w_3)$
MLP Encoder					
	Linear	$n_3 h_3 w_3$	$d_1 l$	–	$(d_1 l)$
	ReLU				$(d_1 l)$
	Linear	$d_1 l$	$d_2 l$	–	$(d_2 l)$
	ReLU				$(d_2 l)$
	Linear(μ)	$d_2 l$	l	–	(l)
	Linear(σ)	$d_2 l$	l	–	(l)
MLP Decoder					
	Linear	l	$d_2 l$	–	$(d_2 l)$
	ReLU				$(d_2 l)$
	Linear	$d_2 l$	$n_3 h_3 w_3$	–	$(n_3 h_3 w_3)$
	ReLU			$(n_3 h_3 w_3)$	
	Reshape			–	(n_3, h_3, w_3)
Decoder					
1	ConvTrans	n_3	n_2	$3 \times 3,\ s = 2,\ p = 1,\ op = 1$	(n_2, h_2, w_2)
	ReLU				(n_2, h_2, w_2)
2	ConvTrans	n_2	n_1	$3 \times 3,\ s = 2,\ p = 1,\ op = 1$	(n_1, h_1, w_1)
	ReLU				(n_1, h_1, w_1)
3	ConvTrans	n_1	c	$3 \times 3,\ s = 2,\ p = 1,\ op = 1$	(c, h, w)
	ReLU				(c, h, w)
	Conv2d	c	c	$3 \times 3,\ s = 1,\ p = 1$	(c, h, w)
	ReLU				(c, h, w)
	Conv2d	c	c	$3 \times 3,\ s = 1,\ p = 1$	(c, h, w)
	Sigmoid				(c, h, w)

the evaluation setup emphasizes the strengths and limitations of different methods. The parameter selection can be seen in table 8.

Table 6: **Sparse Autoencoder (SAE) architecture.** The encoder reduces the input image (c, h, w) through three convolutional layers, producing feature maps (n_3, h_0, w_0) with $h_0 = h/8$ and $w_0 = w/8$. This representation is flattened and mapped by an MLP into a latent vector of dimension l, with proportional factors d_1 and d_2 controlling the intermediate hidden sizes. A sparsity regularization is applied on l to encourage compact codes. The decoder mirrors this process, expanding the latent code through an MLP, reshaping into convolutional feature maps, and progressively upsampling with transposed convolutions to reconstruct (c, h, w).

Stage	**Layer (Parameters)**	**Output Shape**
Encoder		
Conv2d	$(c \to n_1),\ 3\times 3,\ s=2,\ p=1$	$(n_1,\ h/2,\ w/2)$
ReLU	–	–
Conv2d	$(n_1 \to n_2),\ 3\times 3,\ s=2,\ p=1$	$(n_2,\ h/4,\ w/4)$
ReLU	–	–
Conv2d	$(n_2 \to n_3),\ 3\times 3,\ s=2,\ p=1$	$(n_3,\ h/8,\ w/8)$
ReLU	–	–
Flatten	–	$(n_3\, h_0\, w_0)$
MLP Encoder		
Linear	$(n_3 h_0 w_0 \to d_1 l)$	$(d_1 l)$
ReLU	–	–
Linear	$(d_1 l \to d_2 l)$	$(d_2 l)$
ReLU	–	–
Linear	$(d_2 l \to l)$	(l)
MLP Decoder		
Linear	$(l \to d_2 l)$	$(d_2 l)$
ReLU	–	–
Linear	$(d_2 l \to d_1 l)$	$(d_1 l)$
ReLU	–	–
Linear	$(d_1 l \to n_3 h_0 w_0)$	$(n_3 h_0 w_0)$
Decoder		
ConvTranspose2d	$(n_3 \to n_2),\ 3\times 3,\ s=2,\ p=1,\ o_p=1$	$(n_2,\ 2h_0,\ 2w_0)$
ReLU	–	–
ConvTranspose2d	$(n_2 \to n_1),\ 3\times 3,\ s=2,\ p=1,\ o_p=1$	$(n_1,\ 4h_0,\ 4w_0)$
ReLU	–	–
ConvTranspose2d	$(n_1 \to c),\ 3\times 3,\ s=2,\ p=1,\ o_p=1$	$(c,\ 8h_0,\ 8w_0)$
ReLU	–	–
Conv2d	$(c \to c),\ 3\times 3,\ s=1,\ p=1$	$(c,\ 8h_0,\ 8w_0)$
ReLU	–	–
Conv2d	$(c \to c),\ 3\times 3,\ s=1,\ p=1$	$(c,\ 8h_0,\ 8w_0)$
Sigmoid	–	–

Table 7: **Masked Autoencoder (MAE) architecture.** The model first embeds the input image (N, C, H, W) into non-overlapping patches of size $P \times P$, producing $L = (HW/P^2)$ tokens of embedding dimension E, where C is the number of input channels. A learnable [CLS] token and fixed 2D sine–cosine positional embeddings are added. The encoder applies d Transformer blocks with h attention heads and MLP expansion ratio r, yielding latent representations of shape $(N, L+1, E)$. The decoder maps embeddings to dimension D_{dec} (e.g., 16), adds mask tokens and positional encodings, and processes them with d_{dec} Transformer blocks with h_{dec} heads. Finally, a projection stage normalizes and linearly predicts patch pixels of size P^2C, which are reassembled into the reconstructed image (N, C, H, W).

Stage	Module	Configuration	Output Shape
Embed	PatchEmbed	$P=16$, $D=E$ (e.g. 768/1024/1280), $C=3$	(N, L, E)
	Add Pos. Embed (+CLS)	fixed 2D sin-cos	$(N, L+1, E)$
Encoder	Transformer Blocks ($\times d$)	heads h, MLP-ratio r	$(N, L+1, E)$
	LayerNorm	$\mathrm{LayerNorm}(E)$	$(N, L+1, E)$
Decode	Linear Embed	$E \to D_{\text{dec}}$ (512)	$(N, L+1, D_{\text{dec}})$
	Mask Token & Pos. Embed	fixed 2D sin-cos	$(N, L+1, D_{\text{dec}})$
	Transformer Blocks ($\times d_{\text{dec}}$)	heads h_{dec}, MLP-ratio r	$(N, L+1, D_{\text{dec}})$
Proj	LayerNorm	$\mathrm{LayerNorm}(D_{\text{dec}})$	$(N, L+1, D_{\text{dec}})$
	Linear Predict	$D_{\text{dec}} \to P^2C$	$(N, L+1, P^2C)$
	Remove CLS and Unpatchify	—	(N, C, H, W)

Table 8: Experimental configurations across conditions, datasets, and models.

Condition	Dataset	Model	Parameters
Convergence	Pendulum	AE	$d_1 = 200,\ d_2 = 2000,\ l = 8,\ \lambda = 1.0$
		VAE	$n_1 = 8,\ n_2 = 16,\ n_3 = 32,\ d_1 = 16,\ d_2 = 4,\ l = 2,\ \lambda = 1.0$
		SAE	$n_1 = 8,\ n_2 = 16,\ n_3 = 32,\ d_1 = 16,\ d_2 = 4,\ l = 2,\ \lambda = 10.0$
		MAE	patch size = 16, depth= 16, heads = 16, $\lambda = 10.0$
	Mitosis	AE	$d_1 = 200,\ d_2 = 2000,\ l = 8,\ \lambda = 1.0$
		VAE	$n_1 = 8,\ n_2 = 16,\ n_3 = 32,\ d_1 = 16,\ d_2 = 4,\ l = 2,\ \lambda = 1.0$
		SAE	$n_1 = 8,\ n_2 = 16,\ n_3 = 32,\ d_1 = 16,\ d_2 = 4,\ l = 2,\ \lambda = 20.0$
		MAE	patch size = 16, depth= 16, heads= 16, $\lambda = 10.0$
	Galaxy	AE	$d_1 = 200,\ d_2 = 2000,\ l = 8$
		VAE	$n_1 = 8,\ n_2 = 16,\ n_3 = 32,\ d_1 = 16,\ d_2 = 4,\ l = 2$
		SAE	$n_1 = 8,\ n_2 = 16,\ n_3 = 32,\ d_1 = 16,\ d_2 = 4,\ l = 2$
		MAE	patch size = 16, depth= 16, heads = 16, $\lambda = 10.0$
Sample Unbalance	Pendulum	AE	$d_1 = 200,\ d_2 = 2000,\ l = 8,\ \lambda = 10.0$
		VAE	$n_1 = 8,\ n_2 = 16,\ n_3 = 32,\ d_1 = 16,\ d_2 = 4,\ l = 2,\ \lambda = 10.0$
		SAE	$n_1 = 8,\ n_2 = 16,\ n_3 = 32,\ d_1 = 16,\ d_2 = 4,\ l = 2,\ \lambda = 20.0$
		MAE	patch size = 16, depth = 16, heads = 16, $\lambda = 10.0$
	Mitosis	AE	$d_1 = 200,\ d_2 = 2000,\ l = 8,\ \lambda = 1.0$
		VAE	$n_1 = 8,\ n_2 = 16,\ n_3 = 32,\ d_1 = 16,\ d_2 = 4,\ l = 2,\ \lambda = 1.0$
		SAE	$n_1 = 8,\ n_2 = 16,\ n_3 = 32,\ d_1 = 16,\ d_2 = 4,\ l = 2,\ \lambda = 1.0$
		MAE	patch size = 16, depth = 16, heads = 16, $\lambda = 10.0$
	Galaxy	AE	$d_1 = 200,\ d_2 = 2000,\ l = 8,\ \lambda = 1.0$
		VAE	$n_1 = 8,\ n_2 = 16,\ n_3 = 32,\ d_1 = 16,\ d_2 = 4,\ l = 2,\ \lambda = 1.0$
		SAE	$n_1 = 8,\ n_2 = 16,\ n_3 = 32,\ d_1 = 16,\ d_2 = 4,\ l = 2,\ \lambda = 1.0$
		MAE	patch size = 16, depth = 16, heads = 16, $\lambda = 10.0$

A.6 COMPUTATIONAL OVERHEAD.

Table 9 reports the computational cost of the proposed components relative to standard AE training. The self-information loss increases memory and runtime because it computes batch-level location-wise histograms, while SPP approximately doubles the FLOPs because the candidate pool contains both the current batch and propagated samples. Importantly, SPP keeps the number of gradient-contributing samples fixed at B, so its overhead comes from evaluating the enlarged candidate pool rather than from increasing the optimization batch size.

Table 9: **Runtime and memory overhead of the proposed components.** We compare standard AE training, BEL, Sample Propagation (SPP), and the proposed self-information loss. Memory, FLOPs, and wall-clock time are reported per training step; lower values indicate lower computational cost. The self-information loss adds overhead due to batch-level histogram computation, while SPP increases FLOPs by evaluating an enlarged candidate pool before selecting the B hardest samples for the update.

Mode	Loss	Memory (GB)	FLOPs (GFLOPs)	Time (ms)
Standard	L2	**0.2134**	$\mathbf{3.86 \times 10^9}$	$\mathbf{5.51 \pm 0.18}$
Standard	SI	4.6117	3.88×10^9	24.33 ± 0.02
BEL	L2	0.4774	4.03×10^9	9.90 ± 0.25
SPP	L2	0.2207	7.71×10^9	18.00 ± 6.05
SPP	SI	4.6191	7.73×10^9	33.03 ± 3.32

We report computational cost for representative configurations rather than for every dataset–architecture–loss combination, since the full experimental grid is large and would add substantial redundancy. The comparison includes the standard L2 baseline, BEL as a strong reconstruction-loss baseline, and our proposed components, which capture the main sources of computational overhead. As expected, the self-information loss increases training memory and runtime because it computes location-wise batch statistics, and SPP increases training cost by evaluating an enlarged candidate pool. However, these costs apply only during training: at inference time, the trained autoencoder uses the same architecture and parameter count as the corresponding baseline, so its forward-pass cost is unchanged.

A.7 DETAILED EXPERIMENTAL RESULTS FOR REAL DATA

On Tables 10 to 12 we show the quantitative results for the real-world data previously summarized on Fig. 9a for the entropy experiments. Similarly, for the sample proagation experiments, the tables 13 to 15 show the statistical results corresponding to Fig. 9b.

High-resolution qualitative results. Figures 11 and 12 provide high-resolution versions of the qualitative comparisons summarized in the main paper. They are included to make differences in fine structures, such as pendulum traces, galaxy morphology, and cellular boundaries, easier to inspect.

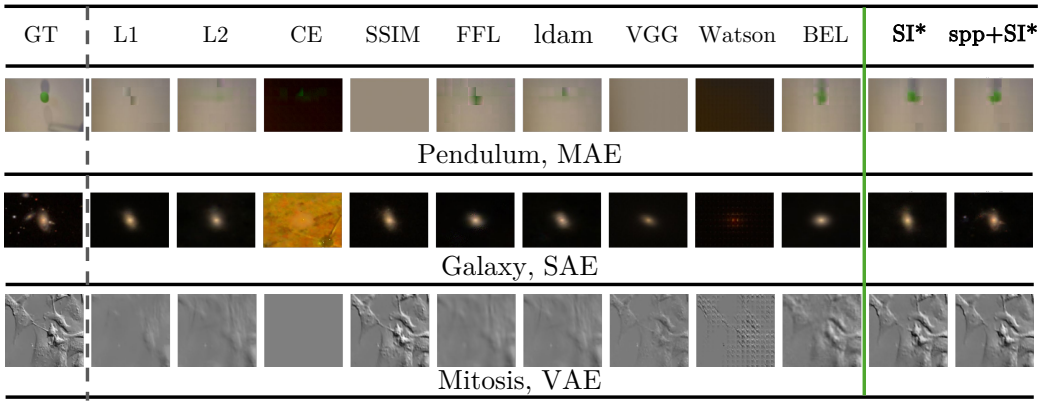


Figure 11: **High-resolution qualitative comparison of reconstruction losses.** This figure expands the qualitative comparison from Fig. 10a. Representative samples from Pendulum, Galaxy Zoo, and Mitosis are shown with ground truth and reconstructions from baseline losses and our SI-based objectives. The larger visualization highlights differences in fine structures that may be difficult to inspect in the compressed main-paper figure.

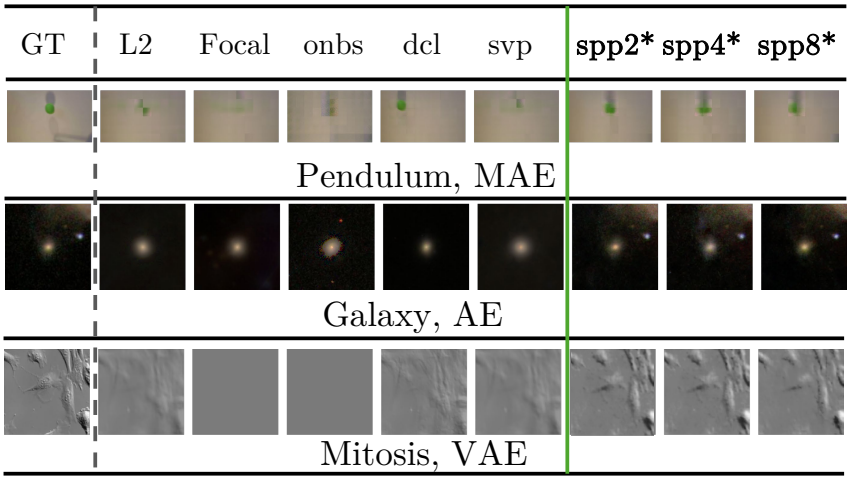


Figure 12: **High-resolution qualitative comparison of Sample Propagation.** This figure expands the Sample Propagation comparison from Fig. 10b. Representative samples from Pendulum, Galaxy Zoo, and Mitosis are shown with ground truth, sample-level baselines, and SPP variants. The larger visualization helps assess whether replaying hard-to-reconstruct samples improves preservation of rare or high-detail image content.

Table 10: **Pendulum reconstruction under repetitive dynamics.** We compare reconstruction losses across four autoencoder families (AE, VAE, SAE, MAE) on PENDULUM. Rows marked with * denote our SI-based objectives. Our methods achieve the strongest distortion metrics in most settings: for AE, SI* improves over the strongest baseline, BEL, by approximately $2\times$ lower MSE, $+3.3$ dB PSNR, and $+0.040$ SSIM; for VAE, SI* gives the best MSE/PSNR, while SSIM loss obtains the highest SSIM at worse distortion; for SAE, SPP + SI* gives the best MSE/PSNR and SI* the best SSIM; and for MAE, SI* is best on all metrics. These results indicate that location-conditioned self-information weighting reduces convergence toward frequent pendulum states and improves reconstruction of less frequent configurations. Bold values indicate the best result per metric and architecture.

Model	Loss	MSE	PSNR	SSIM
AE	L1	$1.126e^{-3} \pm 6.37e^{-4}$	$3.078e^{1} \pm 4.06e^{0}$	$8.777e^{-1} \pm 4.85e^{-2}$
	L2	$8.224e^{-4} \pm 3.35e^{-4}$	$3.127e^{1} \pm 2.01e^{0}$	$8.548e^{-1} \pm 3.29e^{-2}$
	CE	$5.145e^{-3} \pm 4.95e^{-4}$	$2.291e^{1} \pm 4.22e^{-1}$	$7.467e^{-1} \pm 1.33e^{-2}$
	SSIM	$1.588e^{-3} \pm 1.29e^{-4}$	$2.801e^{1} \pm 3.42e^{-1}$	$8.507e^{-1} \pm 1.04e^{-2}$
	FFL	$8.188e^{-4} \pm 3.45e^{-4}$	$3.128e^{1} \pm 1.98e^{0}$	$8.583e^{-1} \pm 3.38e^{-2}$
	ldam	$6.611e^{-4} \pm 3.57e^{-4}$	$3.229e^{1} \pm 2.00e^{0}$	$8.693e^{-1} \pm 2.17e^{-2}$
	VGG	$1.596e^{-3} \pm 1.14e^{-4}$	$2.798e^{1} \pm 3.03e^{-1}$	$8.442e^{-1} \pm 1.04e^{-2}$
	Watson	$2.296e^{-3} \pm 1.91e^{-4}$	$2.640e^{1} \pm 3.59e^{-1}$	$6.934e^{-1} \pm 7.87e^{-3}$
	BEL	$3.359e^{-4} \pm 1.67e^{-4}$	$3.511e^{1} \pm 1.69e^{0}$	$8.890e^{-1} \pm 2.21e^{-2}$
	SI*	$\mathbf{1.661e^{-4}} \pm 1.27e^{-4}$	$\mathbf{3.845e^{1}} \pm 2.13e^{0}$	$\mathbf{9.294e^{-1}} \pm 1.62e^{-2}$
	SPP + SI*	$9.833e^{-4} \pm 1.16e^{-4}$	$3.010e^{1} \pm 5.08e^{-1}$	$8.380e^{-1} \pm 1.87e^{-2}$
VAE	L1	$8.197e^{-4} \pm 9.51e^{-5}$	$3.089e^{1} \pm 4.77e^{-1}$	$8.846e^{-1} \pm 1.18e^{-2}$
	L2	$1.499e^{-3} \pm 1.44e^{-4}$	$2.826e^{1} \pm 4.13e^{-1}$	$7.985e^{-1} \pm 1.12e^{-2}$
	CE	$6.391e^{-3} \pm 5.84e^{-4}$	$2.196e^{1} \pm 4.02e^{-1}$	$7.675e^{-1} \pm 1.19e^{-2}$
	SSIM	$2.283e^{-4} \pm 4.36e^{-5}$	$3.648e^{1} \pm 7.01e^{-1}$	$\mathbf{9.644e^{-1}} \pm 4.44e^{-3}$
	FFL	$1.663e^{-3} \pm 9.29e^{-5}$	$2.780e^{1} \pm 2.40e^{-1}$	$6.879e^{-1} \pm 5.93e^{-3}$
	ldam	$2.725e^{-3} \pm 1.81e^{-4}$	$2.566e^{1} \pm 2.83e^{-1}$	$8.145e^{-1} \pm 1.01e^{-2}$
	VGG	$4.871e^{-4} \pm 1.72e^{-4}$	$3.329e^{1} \pm 1.07e^{0}$	$9.318e^{-1} \pm 1.57e^{-2}$
	Watson	$2.685e^{-3} \pm 3.11e^{-4}$	$2.574e^{1} \pm 4.95e^{-1}$	$8.183e^{-1} \pm 1.26e^{-2}$
	BEL	$1.483e^{-3} \pm 1.32e^{-4}$	$2.831e^{1} \pm 3.80e^{-1}$	$8.102e^{-1} \pm 1.08e^{-2}$
	SI*	$\mathbf{1.021e^{-4}} \pm 2.47e^{-5}$	$\mathbf{4.000e^{1}} \pm 7.94e^{-1}$	$9.419e^{-1} \pm 5.57e^{-3}$
	SPP + SI*	$1.064e^{-4} \pm 2.17e^{-5}$	$3.979e^{1} \pm 6.72e^{-1}$	$9.466e^{-1} \pm 5.72e^{-3}$
SAE	L1	$1.625e^{-3} \pm 1.08e^{-4}$	$2.790e^{1} \pm 2.97e^{-1}$	$8.470e^{-1} \pm 8.77e^{-3}$
	L2	$1.499e^{-3} \pm 1.36e^{-4}$	$2.826e^{1} \pm 3.92e^{-1}$	$8.029e^{-1} \pm 1.11e^{-2}$
	CE	$4.927e^{-3} \pm 4.10e^{-4}$	$2.309e^{1} \pm 3.63e^{-1}$	$7.851e^{-1} \pm 1.06e^{-2}$
	SSIM	$1.675e^{-3} \pm 1.19e^{-4}$	$2.777e^{1} \pm 3.03e^{-1}$	$8.439e^{-1} \pm 8.40e^{-3}$
	FFL	$1.659e^{-3} \pm 1.12e^{-4}$	$2.781e^{1} \pm 2.92e^{-1}$	$6.948e^{-1} \pm 6.48e^{-3}$
	ldam	$1.507e^{-3} \pm 1.46e^{-4}$	$2.824e^{1} \pm 4.16e^{-1}$	$7.932e^{-1} \pm 1.13e^{-2}$
	VGG	$1.639e^{-3} \pm 1.34e^{-4}$	$2.787e^{1} \pm 3.40e^{-1}$	$8.445e^{-1} \pm 9.75e^{-3}$
	Watson	$2.831e^{-3} \pm 1.72e^{-4}$	$2.549e^{1} \pm 2.61e^{-1}$	$8.347e^{-1} \pm 1.01e^{-2}$
	BEL	$1.649e^{-3} \pm 1.26e^{-4}$	$2.784e^{1} \pm 3.27e^{-1}$	$7.747e^{-1} \pm 1.04e^{-2}$
	SI*	$1.620e^{-4} \pm 2.74e^{-5}$	$3.795e^{1} \pm 5.93e^{-1}$	$\mathbf{9.479e^{-1}} \pm 4.83e^{-3}$
	SPP + SI*	$\mathbf{1.124e^{-4}} \pm 5.10e^{-5}$	$\mathbf{3.969e^{1}} \pm 1.11e^{0}$	$9.460e^{-1} \pm 6.58e^{-3}$
MAE	L1	$1.423e^{-3} \pm 2.63e^{-4}$	$2.856e^{1} \pm 9.13e^{-1}$	$8.423e^{-1} \pm 2.22e^{-2}$
	L2	$1.280e^{-3} \pm 2.41e^{-4}$	$2.902e^{1} \pm 9.39e^{-1}$	$7.989e^{-1} \pm 3.65e^{-2}$
	CE	$2.896e^{-1} \pm 4.87e^{-3}$	$5.383e^{0} \pm 7.35e^{-2}$	$1.673e^{-1} \pm 4.33e^{-2}$
	SSIM	$2.807e^{-3} \pm 1.88e^{-4}$	$2.553e^{1} \pm 2.86e^{-1}$	$8.231e^{-1} \pm 7.16e^{-3}$
	FFL	$1.392e^{-3} \pm 2.07e^{-4}$	$2.862e^{1} \pm 7.19e^{-1}$	$7.121e^{-1} \pm 6.71e^{-2}$
	ldam	$1.263e^{-3} \pm 2.52e^{-4}$	$2.909e^{1} \pm 1.00e^{0}$	$8.166e^{-1} \pm 3.30e^{-2}$
	VGG	$4.748e^{-3} \pm 3.79e^{-4}$	$2.325e^{1} \pm 3.46e^{-1}$	$7.622e^{-1} \pm 6.39e^{-3}$
	Watson	$1.730e^{-1} \pm 3.66e^{-3}$	$7.620e^{0} \pm 9.24e^{-2}$	$5.719e^{-1} \pm 4.70e^{-3}$
	BEL	$7.944e^{-4} \pm 3.79e^{-4}$	$3.142e^{1} \pm 1.86e^{0}$	$8.304e^{-1} \pm 3.44e^{-2}$
	SI*	$\mathbf{5.928e^{-4} \pm 3.86e^{-4}}$	$\mathbf{3.290e^{1} \pm 2.17e^{0}}$	$\mathbf{8.856e^{-1} \pm 2.27e^{-2}}$
	SPP + SI*	$6.137e^{-4} \pm 3.89e^{-4}$	$3.271e^{1} \pm 2.09e^{0}$	$8.847e^{-1} \pm 1.79e^{-2}$

Table 11: **Galaxy reconstruction with structured astronomical textures.** We compare reconstruction losses across four autoencoder families (AE, VAE, SAE, MAE) on GALAXY. Rows marked with * denote our SI-based objectives. Our methods achieve the best distortion metrics in all architectures while remaining competitive in SSIM, typically within 0.01–0.02 of the SSIM-trained baseline. For AE, SI* obtains the best MSE/PSNR (9.57×10^{-4}, 31.59 dB), improving over the strongest baseline by approximately 38% lower MSE and +0.69 dB. For VAE, SPP + SI* gives the best MSE/PSNR (8.78×10^{-4}, 31.79 dB), with a 9% MSE reduction and +0.73 dB. For SAE, SPP + SI* gives the best MSE/PSNR (1.30×10^{-3}, 30.38 dB), reducing MSE by approximately 32%. For MAE, SI* gives the highest PSNR (30.01 dB) and lower MSE (1.41×10^{-3}), improving MSE by approximately 24%. These results suggest that location-conditioned self-information weighting reduces oversmoothing on structured astronomical images and improves reconstruction fidelity across architectures. Bold values indicate the best result per metric and architecture.

Model	Loss	MSE	PSNR	SSIM
AE	L1	$1.301e^{-3} \pm 1.61e^{-3}$	$3.085e^{1} \pm 3.94e^{0}$	$8.359e^{-1} \pm 4.64e^{-2}$
	L2	$1.537e^{-3} \pm 1.45e^{-3}$	$2.954e^{1} \pm 3.42e^{0}$	$8.224e^{-1} \pm 4.92e^{-2}$
	CE	$1.707e^{-1} \pm 1.07e^{-2}$	$7.686e^{0} \pm 2.70e^{-1}$	$5.258e^{-1} \pm 3.48e^{-2}$
	SSIM	$1.268e^{-3} \pm 1.50e^{-3}$	$3.090e^{1} \pm 3.91e^{0}$	$\mathbf{8.527e^{-1}} \pm 4.53e^{-2}$
	FFL	$1.874e^{-3} \pm 2.21e^{-3}$	$2.900e^{1} \pm 3.62e^{0}$	$7.923e^{-1} \pm 6.07e^{-2}$
	ldam	$1.670e^{-3} \pm 1.82e^{-3}$	$2.936e^{1} \pm 3.53e^{0}$	$8.174e^{-1} \pm 5.03e^{-2}$
	VGG	$2.990e^{-3} \pm 5.33e^{-3}$	$2.717e^{1} \pm 3.57e^{0}$	$8.027e^{-1} \pm 4.76e^{-2}$
	Watson	$5.510e^{-3} \pm 9.70e^{-3}$	$2.388e^{1} \pm 2.75e^{0}$	$5.897e^{-1} \pm 8.58e^{-2}$
	BEL	$1.686e^{-3} \pm 1.89e^{-3}$	$2.935e^{1} \pm 3.55e^{0}$	$8.198e^{-1} \pm 5.00e^{-2}$
	SI*	$\mathbf{9.567e^{-4}} \pm 9.13e^{-4}$	$\mathbf{3.159e^{1}} \pm 3.38e^{0}$	$8.377e^{-1} \pm 4.34e^{-2}$
	SPP + SI*	$1.048e^{-3} \pm 6.15e^{-4}$	$3.026e^{1} \pm 1.96e^{0}$	$8.137e^{-1} \pm 4.75e^{-2}$
VAE	L1	$1.268e^{-3} \pm 2.92e^{-3}$	$3.045e^{1} \pm 3.12e^{0}$	$8.123e^{-1} \pm 5.24e^{-2}$
	L2	$9.686e^{-4} \pm 2.24e^{-3}$	$3.106e^{1} \pm 2.15e^{0}$	$7.954e^{-1} \pm 5.21e^{-2}$
	CE	$1.684e^{-1} \pm 1.31e^{-2}$	$7.748e^{0} \pm 3.35e^{-1}$	$6.202e^{-1} \pm 5.77e^{-2}$
	SSIM	$9.784e^{-4} \pm 2.85e^{-3}$	$3.150e^{1} \pm 2.82e^{0}$	$\mathbf{8.294e^{-1}} \pm 4.88e^{-2}$
	FFL	$1.283e^{-3} \pm 2.63e^{-3}$	$2.990e^{1} \pm 2.42e^{0}$	$7.656e^{-1} \pm 6.00e^{-2}$
	ldam	$1.004e^{-3} \pm 2.25e^{-3}$	$3.090e^{1} \pm 2.18e^{0}$	$7.891e^{-1} \pm 5.34e^{-2}$
	VGG	$2.412e^{-3} \pm 4.94e^{-3}$	$2.796e^{1} \pm 3.24e^{0}$	$8.022e^{-1} \pm 5.49e^{-2}$
	Watson	$1.886e^{-3} \pm 3.41e^{-3}$	$2.876e^{1} \pm 3.20e^{0}$	$6.720e^{-1} \pm 7.09e^{-2}$
	BEL	$3.127e^{-2} \pm 2.64e^{-3}$	$1.506e^{1} \pm 3.43e^{-1}$	$7.572e^{-1} \pm 6.00e^{-2}$
	SI*	$8.951e^{-4} \pm 2.69e^{-3}$	$3.163e^{1} \pm 2.47e^{0}$	$8.196e^{-1} \pm 5.30e^{-2}$
	SPP + SI*	$\mathbf{8.780e^{-4}} \pm 2.71e^{-3}$	$\mathbf{3.179e^{1}} \pm 2.42e^{0}$	$8.171e^{-1} \pm 5.21e^{-2}$
SAE	L1	$1.913e^{-3} \pm 4.06e^{-3}$	$2.952e^{1} \pm 3.88e^{0}$	$8.069e^{-1} \pm 5.59e^{-2}$
	L2	$2.224e^{-3} \pm 3.66e^{-3}$	$2.834e^{1} \pm 3.48e^{0}$	$7.757e^{-1} \pm 5.85e^{-2}$
	CE	$3.027e^{-1} \pm 1.55e^{-2}$	$5.198e^{0} \pm 3.24e^{-1}$	$6.933e^{-1} \pm 5.27e^{-2}$
	SSIM	$1.909e^{-3} \pm 4.45e^{-3}$	$2.980e^{1} \pm 4.04e^{0}$	$\mathbf{8.135e^{-1}} \pm 5.56e^{-2}$
	FFL	$2.320e^{-3} \pm 3.76e^{-3}$	$2.807e^{1} \pm 3.36e^{0}$	$7.583e^{-1} \pm 6.28e^{-2}$
	ldam	$2.228e^{-3} \pm 3.66e^{-3}$	$2.833e^{1} \pm 3.48e^{0}$	$7.795e^{-1} \pm 5.95e^{-2}$
	VGG	$3.534e^{-3} \pm 6.40e^{-3}$	$2.620e^{1} \pm 3.34e^{0}$	$7.730e^{-1} \pm 5.47e^{-2}$
	Watson	$1.476e^{-1} \pm 9.72e^{-3}$	$8.324e^{0} \pm 4.09e^{-1}$	$2.644e^{-1} \pm 1.04e^{-2}$
	BEL	$2.848e^{-3} \pm 6.37e^{-3}$	$2.737e^{1} \pm 3.36e^{0}$	$7.799e^{-1} \pm 5.22e^{-2}$
	SI*	$1.374e^{-3} \pm 2.86e^{-3}$	$3.025e^{1} \pm 3.34e^{0}$	$8.103e^{-1} \pm 5.48e^{-2}$
	SPP + SI*	$\mathbf{1.302e^{-3}} \pm 2.89e^{-3}$	$\mathbf{3.038e^{1}} \pm 3.20e^{0}$	$8.097e^{-1} \pm 5.40e^{-2}$
MAE	L1	$1.885e^{-3} \pm 3.40e^{-3}$	$2.911e^{1} \pm 3.54e^{0}$	$8.019e^{-1} \pm 5.63e^{-2}$
	L2	$1.907e^{-3} \pm 3.09e^{-3}$	$2.895e^{1} \pm 3.46e^{0}$	$7.969e^{-1} \pm 5.55e^{-2}$
	CE	$3.720e^{-2} \pm 8.35e^{-3}$	$1.438e^{1} \pm 8.25e^{-1}$	$1.883e^{-1} \pm 2.15e^{-2}$
	SSIM	$1.924e^{-3} \pm 3.51e^{-3}$	$2.903e^{1} \pm 3.54e^{0}$	$\mathbf{8.053e^{-1} \pm 5.47e^{-2}}$
	FFL	$2.070e^{-3} \pm 3.25e^{-3}$	$2.855e^{1} \pm 3.42e^{0}$	$7.869e^{-1} \pm 5.63e^{-2}$
	ldam	$1.850e^{-3} \pm 2.55e^{-3}$	$2.894e^{1} \pm 3.38e^{0}$	$7.950e^{-1} \pm 5.60e^{-2}$
	VGG	$6.195e^{-3} \pm 7.06e^{-3}$	$2.302e^{1} \pm 2.54e^{0}$	$1.120e^{-1} \pm 4.62e^{-3}$
	Watson	$7.818e^{-3} \pm 7.99e^{-3}$	$2.185e^{1} \pm 2.30e^{0}$	$5.033e^{-2} \pm 5.19e^{-3}$
	BEL	$1.851e^{-3} \pm 2.61e^{-3}$	$2.896e^{1} \pm 3.41e^{0}$	$7.976e^{-1} \pm 5.54e^{-2}$
	SI*	$1.409e^{-3} \pm 1.52e^{-3}$	$\mathbf{3.001e^{1} \pm 3.41e^{0}}$	$8.049e^{-1} \pm 5.38e^{-2}$
	SPP + SI*	$\mathbf{1.402e^{-3} \pm 1.51e^{-3}}$	$2.998e^{1} \pm 3.35e^{0}$	$8.053e^{-1} \pm 5.34e^{-2}$

Table 12: **Mitosis reconstruction with rare cellular events.** We compare reconstruction losses across four autoencoder families (AE, VAE, SAE, MAE) on MITOSIS. Rows marked with * denote our SI-based objectives. For AE, SI* achieves the best scores on all metrics (6.65×10^{-4} MSE, 32.10 dB PSNR, 0.918 SSIM), improving over the strongest baseline, LDAM, by approximately 38% lower MSE, +2.17 dB, and +0.047 SSIM. For VAE, **SPP + SI*** gives the best distortion metrics (1.05×10^{-3} MSE, 29.95 dB), with a 4.6% MSE reduction and +0.16 dB over the best baseline, while the SSIM-trained loss obtains the highest SSIM. For SAE, SI* is best on all metrics (3.01×10^{-3}, 25.49 dB, 0.740), reducing MSE by approximately 36% and improving PSNR by +1.85 dB. For MAE, **SPP + SI*** leads across metrics (3.88×10^{-3}, 24.30 dB, 0.630), reducing MSE by approximately 29%, improving PSNR by +1.56 dB, and increasing SSIM by +0.020. These results suggest that location-conditioned self-information weighting reduces oversmoothing and improves reconstruction of fine cellular structures. Bold values indicate the best result per metric and architecture.

Model	Loss	MSE	PSNR	SSIM
AE	L1	$1.546e^{-3} \pm 7.46e^{-4}$	$2.848e^{1} \pm 1.72e^{0}$	$8.588e^{-1} \pm 4.87e^{-2}$
	L2	$1.366e^{-3} \pm 4.27e^{-4}$	$2.883e^{1} \pm 1.25e^{0}$	$8.449e^{-1} \pm 2.80e^{-2}$
	CE	$6.967e^{-3} \pm 1.25e^{-3}$	$2.163e^{1} \pm 7.03e^{-1}$	$8.927e^{-2} \pm 1.25e^{-2}$
	SSIM	$2.744e^{-3} \pm 1.43e^{-3}$	$2.626e^{1} \pm 2.48e^{0}$	$7.933e^{-1} \pm 8.16e^{-2}$
	FFL	$1.572e^{-3} \pm 5.27e^{-4}$	$2.823e^{1} \pm 1.25e^{0}$	$8.128e^{-1} \pm 3.71e^{-2}$
	ldam	$1.073e^{-3} \pm 4.05e^{-4}$	$2.993e^{1} \pm 1.38e^{0}$	$8.711e^{-1} \pm 2.81e^{-2}$
	VGG	$2.410e^{-3} \pm 1.19e^{-3}$	$2.673e^{1} \pm 2.23e^{0}$	$8.502e^{-1} \pm 5.56e^{-2}$
	Watson	$6.681e^{-3} \pm 1.39e^{-3}$	$2.183e^{1} \pm 8.25e^{-1}$	$4.361e^{-1} \pm 6.97e^{-2}$
	BEL	$2.422e^{-3} \pm 7.46e^{-4}$	$2.636e^{1} \pm 1.33e^{0}$	$7.628e^{-1} \pm 4.73e^{-2}$
	SI*	$\mathbf{6.654e^{-4}} \pm 3.09e^{-4}$	$\mathbf{3.210e^{1}} \pm 1.60e^{0}$	$\mathbf{9.178e^{-1}} \pm 1.99e^{-2}$
	SPP + SI*	$1.478e^{-3} \pm 1.09e^{-4}$	$2.831e^{1} \pm 2.84e^{-1}$	$8.420e^{-1} \pm 2.65e^{-2}$
VAE	L1	$6.592e^{-3} \pm 1.63e^{-3}$	$2.192e^{1} \pm 9.59e^{-1}$	$5.223e^{-1} \pm 6.81e^{-2}$
	L2	$6.319e^{-3} \pm 1.67e^{-3}$	$2.212e^{1} \pm 1.03e^{0}$	$4.992e^{-1} \pm 4.92e^{-2}$
	CE	$7.555e^{-3} \pm 1.43e^{-3}$	$2.128e^{1} \pm 7.32e^{-1}$	$5.300e^{-1} \pm 6.07e^{-2}$
	SSIM	$1.096e^{-3} \pm 3.37e^{-4}$	$2.979e^{1} \pm 1.26e^{0}$	$\mathbf{9.094e^{-1}} \pm 1.59e^{-2}$
	FFL	$6.396e^{-3} \pm 1.59e^{-3}$	$2.206e^{1} \pm 9.67e^{-1}$	$4.542e^{-1} \pm 4.10e^{-2}$
	ldam	$6.320e^{-3} \pm 1.61e^{-3}$	$2.211e^{1} \pm 9.91e^{-1}$	$4.898e^{-1} \pm 4.42e^{-2}$
	VGG	$3.774e^{-3} \pm 1.46e^{-3}$	$2.456e^{1} \pm 1.69e^{0}$	$7.807e^{-1} \pm 4.72e^{-2}$
	Watson	$8.002e^{-3} \pm 1.72e^{-3}$	$2.105e^{1} \pm 8.30e^{-1}$	$4.519e^{-1} \pm 7.21e^{-2}$
	BEL	$3.611e^{-3} \pm 1.15e^{-3}$	$2.462e^{1} \pm 1.28e^{0}$	$6.562e^{-1} \pm 5.03e^{-2}$
	SI*	$1.054e^{-3} \pm 2.92e^{-4}$	$2.992e^{1} \pm 1.11e^{0}$	$8.798e^{-1} \pm 1.74e^{-2}$
	SPP + SI*	$\mathbf{1.046e^{-3}} \pm 2.92e^{-4}$	$\mathbf{2.995e^{1}} \pm 1.12e^{0}$	$8.736e^{-1} \pm 1.95e^{-2}$
SAE	L1	$7.381e^{-3} \pm 1.43e^{-3}$	$2.139e^{1} \pm 7.52e^{-1}$	$5.305e^{-1} \pm 6.05e^{-2}$
	L2	$6.312e^{-3} \pm 1.62e^{-3}$	$2.212e^{1} \pm 9.97e^{-1}$	$5.021e^{-1} \pm 4.73e^{-2}$
	CE	$7.345e^{-3} \pm 1.43e^{-3}$	$2.141e^{1} \pm 7.53e^{-1}$	$5.305e^{-1} \pm 6.05e^{-2}$
	SSIM	$4.709e^{-3} \pm 1.91e^{-3}$	$2.364e^{1} \pm 1.82e^{0}$	$6.965e^{-1} \pm 7.94e^{-2}$
	FFL	$6.385e^{-3} \pm 1.58e^{-3}$	$2.206e^{1} \pm 9.63e^{-1}$	$4.604e^{-1} \pm 4.16e^{-2}$
	ldam	$6.341e^{-3} \pm 1.58e^{-3}$	$2.209e^{1} \pm 9.64e^{-1}$	$4.682e^{-1} \pm 3.98e^{-2}$
	VGG	$7.207e^{-3} \pm 1.48e^{-3}$	$2.150e^{1} \pm 7.92e^{-1}$	$4.167e^{-1} \pm 4.68e^{-2}$
	Watson	$7.676e^{-3} \pm 1.42e^{-3}$	$2.121e^{1} \pm 7.16e^{-1}$	$3.781e^{-1} \pm 3.80e^{-2}$
	BEL	$6.295e^{-3} \pm 1.62e^{-3}$	$2.213e^{1} \pm 1.00e^{0}$	$5.029e^{-1} \pm 4.63e^{-2}$
	SI*	$\mathbf{3.014e^{-3} \pm 1.12e^{-3}}$	$\mathbf{2.549e^{1} \pm 1.54e^{0}}$	$\mathbf{7.403e^{-1} \pm 4.58e^{-2}}$
	SPP + SI*	$3.925e^{-3} \pm 1.43e^{-3}$	$2.435e^{1} \pm 1.59e^{0}$	$6.840e^{-1} \pm 6.81e^{-2}$
MAE	L1	$5.527e^{-3} \pm 1.21e^{-3}$	$2.268e^{1} \pm 9.35e^{-1}$	$5.553e^{-1} \pm 6.15e^{-2}$
	L2	$5.546e^{-3} \pm 1.25e^{-3}$	$2.266e^{1} \pm 9.22e^{-1}$	$5.165e^{-1} \pm 5.30e^{-2}$
	CE	$2.433e^{-1} \pm 2.11e^{-3}$	$6.139e^{0} \pm 3.74e^{-2}$	$7.364e^{-2} \pm 5.77e^{-3}$
	SSIM	$5.492e^{-3} \pm 1.40e^{-3}$	$2.274e^{1} \pm 1.08e^{0}$	$6.095e^{-1} \pm 6.05e^{-2}$
	FFL	$5.721e^{-3} \pm 1.31e^{-3}$	$2.253e^{1} \pm 9.30e^{-1}$	$5.097e^{-1} \pm 5.43e^{-2}$
	ldam	$5.492e^{-3} \pm 1.25e^{-3}$	$2.271e^{1} \pm 9.29e^{-1}$	$5.319e^{-1} \pm 5.13e^{-2}$
	VGG	$7.683e^{-3} \pm 1.44e^{-3}$	$2.121e^{1} \pm 7.30e^{-1}$	$6.523e^{-2} \pm 4.06e^{-3}$
	Watson	$7.942e^{-3} \pm 1.53e^{-3}$	$2.107e^{1} \pm 7.47e^{-1}$	$3.009e^{-1} \pm 5.31e^{-2}$
	BEL	$5.272e^{-3} \pm 1.24e^{-3}$	$2.290e^{1} \pm 1.02e^{0}$	$5.222e^{-1} \pm 6.36e^{-2}$
	SI*	$3.897e^{-3} \pm 1.07e^{-3}$	$2.425e^{1} \pm 1.18e^{0}$	$6.254e^{-1} \pm 5.43e^{-2}$
	SPP + SI*	$\mathbf{3.883e^{-3} \pm 1.15e^{-3}}$	$\mathbf{2.430e^{1} \pm 1.27e^{0}}$	$\mathbf{6.298e^{-1} \pm 6.06e^{-2}}$

Table 13: **Pendulum: effect of underrepresented-sample focusing on accuracy and robustness.** We compare losses across four architectures (AE, VAE, SAE, MAE). Rows with $*$ denote our *sppk* objective that prioritizes underrepresented samples via a memory of size k. Considering both mean and variability (mean±std), *spp* markedly improves average reconstruction on three architectures while keeping variance acceptable: **VAE**—*spp2* attains the best distortion and perceptual quality (MSE↓, PSNR↑, SSIM↑) with low–moderate spread; **SAE**—*spp2** achieves the strongest means across all metrics, with *spp8** slightly trading peak PSNR for reduced dispersion; **MAE**—*spp2** leads on all metrics and maintains tight standard deviations. For **AE**, focal loss offers the best means and small variance, while *spp4*/spp8** remain competitive in SSIM but do not surpass focal on averages. Across models, increasing memory (*spp2→spp4→spp8*) generally smooths variance at a modest cost in mean performance, indicating a controllable mean–robustness trade-off. Overall, underrepresented-sample focusing improves average reconstruction quality where data imbalance is most impactful (VAE/SAE/MAE) and provides a tunable handle on stability via the standard deviation.

Model	Baseline	MSE	PSNR	SSIM
AE	L2	$1.292e^{-3} \pm 3.23e^{-4}$	$2.905e^{1} \pm 1.27e^{0}$	$8.129e^{-1} \pm 3.71e^{-2}$
	Focal	$3.477e^{-4} \pm 1.56e^{-4}$	$3.489e^{1} \pm 1.54e^{0}$	$9.131e^{-1} \pm 1.61e^{-2}$
	onbs	$2.533e^{-3} \pm 7.90e^{-4}$	$2.636e^{1} \pm 2.22e^{0}$	$7.679e^{-1} \pm 3.89e^{-2}$
	dcl	$1.305e^{-3} \pm 3.71e^{-4}$	$2.904e^{1} \pm 1.34e^{0}$	$8.083e^{-1} \pm 3.63e^{-2}$
	svp	$1.009e^{-3} \pm 3.70e^{-4}$	$3.036e^{1} \pm 2.06e^{0}$	$8.315e^{-1} \pm 4.12e^{-2}$
	spp2*	$\mathbf{1.563e^{-3} \pm 1.49e^{-4}}$	$2.808e^{1} \pm 4.26e^{-1}$	$7.932e^{-1} \pm 1.77e^{-2}$
	spp4*	$1.041e^{-3} \pm 5.05e^{-4}$	$3.053e^{1} \pm 2.77e^{0}$	$\mathbf{8.368e^{-1} \pm 4.85e^{-2}}$
	spp8*	$1.147e^{-3} \pm 4.96e^{-4}$	$\mathbf{3.012e^{1} \pm 2.94e^{0}}$	$8.429e^{-1} \pm 4.67e^{-2}$
VAE	L2	$1.581e^{-3} \pm 9.09e^{-5}$	$2.802e^{1} \pm 2.48e^{-1}$	$7.832e^{-1} \pm 9.23e^{-3}$
	Focal	$1.524e^{-3} \pm 9.91e^{-5}$	$2.818e^{1} \pm 2.83e^{-1}$	$7.953e^{-1} \pm 1.08e^{-2}$
	onbs	$2.321e^{-3} \pm 3.76e^{-4}$	$2.642e^{1} \pm 8.52e^{-1}$	$\mathbf{6.651e^{-1} \pm 3.38e^{-2}}$
	dcl	$1.685e^{-3} \pm 2.03e^{-4}$	$2.776e^{1} \pm 5.08e^{-1}$	$7.191e^{-1} \pm 2.20e^{-2}$
	svp	$2.744e^{-3} \pm 1.83e^{-4}$	$2.562e^{1} \pm 2.84e^{-1}$	$8.143e^{-1} \pm 1.00e^{-2}$
	spp2*	$\mathbf{1.806e^{-4} \pm 8.21e^{-5}}$	$\mathbf{3.772e^{1} \pm 1.44e^{0}}$	$9.310e^{-1} \pm 9.63e^{-3}$
	spp4*	$2.754e^{-3} \pm 1.71e^{-4}$	$2.561e^{1} \pm 2.67e^{-1}$	$8.122e^{-1} \pm 1.02e^{-2}$
	spp8*	$6.442e^{-4} \pm 1.15e^{-4}$	$3.196e^{1} \pm 6.54e^{-1}$	$8.583e^{-1} \pm 1.13e^{-2}$
SAE	L2	$1.594e^{-3} \pm 9.58e^{-5}$	$2.798e^{1} \pm 2.58e^{-1}$	$7.757e^{-1} \pm 1.02e^{-2}$
	Focal	$\mathbf{1.587e^{-3} \pm 8.24e^{-5}}$	$2.800e^{1} \pm 2.24e^{-1}$	$7.902e^{-1} \pm 8.33e^{-3}$
	onbs	$2.540e^{-3} \pm 4.02e^{-4}$	$2.602e^{1} \pm 8.46e^{-1}$	$\mathbf{5.996e^{-1} \pm 2.71e^{-2}}$
	dcl	$1.691e^{-3} \pm 1.68e^{-4}$	$2.774e^{1} \pm 4.20e^{-1}$	$7.413e^{-1} \pm 1.94e^{-2}$
	svp	$1.529e^{-3} \pm 1.13e^{-4}$	$2.817e^{1} \pm 3.18e^{-1}$	$7.939e^{-1} \pm 1.17e^{-2}$
	spp2*	$1.537e^{-4} \pm 1.79e^{-4}$	$\mathbf{3.906e^{1} \pm 2.39e^{0}}$	$9.420e^{-1} \pm 1.44e^{-2}$
	spp4*	$7.563e^{-4} \pm 2.40e^{-4}$	$3.138e^{1} \pm 1.11e^{0}$	$8.656e^{-1} \pm 2.44e^{-2}$
	spp8*	$3.080e^{-4} \pm 1.49e^{-4}$	$3.545e^{1} \pm 1.57e^{0}$	$9.253e^{-1} \pm 1.49e^{-2}$
MAE	L2	$1.271e^{-3} \pm 2.46e^{-4}$	$2.906e^{1} \pm 9.62e^{-1}$	$8.181e^{-1} \pm 3.08e^{-2}$
	Focal	$1.283e^{-3} \pm 2.35e^{-4}$	$2.900e^{1} \pm 9.07e^{-1}$	$7.830e^{-1} \pm 3.85e^{-2}$
	onbs	$2.102e^{-3} \pm 3.21e^{-4}$	$2.683e^{1} \pm 7.05e^{-1}$	$4.907e^{-1} \pm 1.38e^{-2}$
	dcl	$1.331e^{-3} \pm 3.09e^{-4}$	$2.889e^{1} \pm 1.12e^{0}$	$7.887e^{-1} \pm 3.80e^{-2}$
	svp	$1.294e^{-3} \pm 2.32e^{-4}$	$2.896e^{1} \pm 8.92e^{-1}$	$7.892e^{-1} \pm 3.82e^{-2}$
	spp2*	$\mathbf{5.246e^{-4} \pm 3.28e^{-4}}$	$\mathbf{3.335e^{1} \pm 2.00e^{0}}$	$\mathbf{8.896e^{-1} \pm 1.53e^{-2}}$
	spp4*	$6.862e^{-4} \pm 4.47e^{-4}$	$3.227e^{1} \pm 2.18e^{0}$	$8.837e^{-1} \pm 1.85e^{-2}$
	spp8*	$5.593e^{-4} \pm 3.35e^{-4}$	$3.299e^{1} \pm 1.81e^{0}$	$8.823e^{-1} \pm 1.77e^{-2}$

Table 14: **Mitosis: underrepresented-sample focusing improves averages with controllable variance.** We compare losses across four architectures (AE, VAE, SAE, MAE); rows with $*$ denote our *sppk* objective (memory size k). Considering mean$\pm$std, **AE** benefits most from larger memory: *spp8* attains the best MSE/PSNR/SSIM means, with a small increase in dispersion—indicating sharper reconstructions at slightly higher variability. **VAE** shows large gains with *spp8*/spp4**, which substantially improve all three metrics while *reducing* MSE variability versus pixel/perceptual baselines, suggesting more stable training on rare cell patterns. For **SAE**, *spp4** offers the strongest averages across metrics but with higher SSIM spread; *spp8** trades a bit of mean performance for tighter variance, exposing a tunable accuracy–robustness frontier. **MAE** mirrors this behavior: *spp4** yields the best means with standard deviations comparable to L2/Focal. Overall, prioritizing underrepresented samples consistently improves average reconstruction (MSE↓, PSNR/SSIM↑) and provides a controllable handle on robustness via the standard deviation, especially effective for VAE/MAE and with memory size governing the mean–variance trade-off.

Model	Baseline	MSE	PSNR	SSIM
AE	L2	$6.270e^{-3} \pm 1.54e^{-3}$	$2.214e^{1} \pm 9.58e^{-1}$	$4.983e^{-1} \pm 3.89e^{-2}$
	Focal	$6.297e^{-3} \pm 1.52e^{-3}$	$2.212e^{1} \pm 9.51e^{-1}$	$4.966e^{-1} \pm 4.04e^{-2}$
	onbs	$7.454e^{-3} \pm 1.58e^{-3}$	$2.136e^{1} \pm 8.30e^{-1}$	$9.822e^{-2} \pm 3.45e^{-2}$
	dcl	$6.309e^{-3} \pm 1.46e^{-3}$	$2.210e^{1} \pm 8.94e^{-1}$	$4.901e^{-1} \pm 3.15e^{-2}$
	svp	$6.303e^{-3} \pm 1.45e^{-3}$	$2.210e^{1} \pm 8.91e^{-1}$	$4.912e^{-1} \pm 3.28e^{-2}$
	spp2*	$6.414e^{-3} \pm 1.56e^{-3}$	$2.204e^{1} \pm 9.48e^{-1}$	$5.136e^{-1} \pm 4.92e^{-2}$
	spp4*	$6.420e^{-3} \pm 1.51e^{-3}$	$2.203e^{1} \pm 9.19e^{-1}$	$5.119e^{-1} \pm 4.49e^{-2}$
	spp8*	$\mathbf{6.054e^{-3} \pm 1.59e^{-3}}$	$\mathbf{2.232e^{1} \pm 1.11e^{0}}$	$\mathbf{5.291e^{-1} \pm 5.71e^{-2}}$
ConvVAE	L2	$6.328e^{-3} \pm 1.50e^{-3}$	$2.209e^{1} \pm 9.34e^{-1}$	$4.783e^{-1} \pm 3.57e^{-2}$
	Focal	$6.363e^{-3} \pm 1.48e^{-3}$	$2.207e^{1} \pm 9.15e^{-1}$	$4.647e^{-1} \pm 3.43e^{-2}$
	onbs	$7.342e^{-3} \pm 1.47e^{-3}$	$2.142e^{1} \pm 7.75e^{-1}$	$3.648e^{-1} \pm 5.29e^{-2}$
	dcl	$6.350e^{-3} \pm 1.53e^{-3}$	$2.208e^{1} \pm 9.46e^{-1}$	$4.784e^{-1} \pm 3.76e^{-2}$
	svp	$6.413e^{-3} \pm 1.45e^{-3}$	$2.203e^{1} \pm 8.89e^{-1}$	$4.458e^{-1} \pm 2.99e^{-2}$
	spp2*	$2.085e^{-3} \pm 5.17e^{-4}$	$2.693e^{1} \pm 9.98e^{-1}$	$7.855e^{-1} \pm 2.38e^{-2}$
	spp4*	$\mathbf{1.762e^{-3} \pm 5.11e^{-4}}$	$\mathbf{2.771e^{1} \pm 1.20e^{0}}$	$8.045e^{-1} \pm 2.79e^{-2}$
	spp8*	$1.800e^{-3} \pm 4.83e^{-4}$	$2.759e^{1} \pm 1.07e^{0}$	$\mathbf{8.048e^{-1} \pm 2.28e^{-2}}$
SAE	L2	$6.397e^{-3} \pm 1.35e^{-3}$	$2.203e^{1} \pm 8.47e^{-1}$	$4.535e^{-1} \pm 2.89e^{-2}$
	Focal	$6.454e^{-3} \pm 1.30e^{-3}$	$2.198e^{1} \pm 8.19e^{-1}$	$4.302e^{-1} \pm 3.18e^{-2}$
	onbs	$7.276e^{-3} \pm 1.44e^{-3}$	$2.145e^{1} \pm 7.63e^{-1}$	$3.984e^{-1} \pm 4.46e^{-2}$
	dcl	$6.466e^{-3} \pm 1.40e^{-3}$	$2.199e^{1} \pm 8.72e^{-1}$	$4.381e^{-1} \pm 3.05e^{-2}$
	svp	$6.432e^{-3} \pm 1.31e^{-3}$	$2.200e^{1} \pm 8.18e^{-1}$	$4.268e^{-1} \pm 2.62e^{-2}$
	spp2*	$6.474e^{-3} \pm 1.35e^{-3}$	$2.197e^{1} \pm 8.28e^{-1}$	$4.901e^{-1} \pm 3.07e^{-2}$
	spp4*	$\mathbf{4.715e^{-3} \pm 1.37e^{-3}}$	$\mathbf{2.348e^{1} \pm 1.42e^{0}}$	$\mathbf{6.343e^{-1} \pm 7.53e^{-2}}$
	spp8*	$6.007e^{-3} \pm 1.05e^{-3}$	$2.228e^{1} \pm 7.26e^{-1}$	$5.137e^{-1} \pm 5.08e^{-2}$
MAE	L2	$5.508e^{-3} \pm 1.25e^{-3}$	$2.269e^{1} \pm 9.20e^{-1}$	$5.370e^{-1} \pm 5.20e^{-2}$
	Focal	$5.528e^{-3} \pm 1.23e^{-3}$	$2.267e^{1} \pm 9.21e^{-1}$	$5.325e^{-1} \pm 5.34e^{-2}$
	onbs	$7.342e^{-3} \pm 1.63e^{-3}$	$2.144e^{1} \pm 8.87e^{-1}$	$3.954e^{-1} \pm 5.36e^{-2}$
	dcl	$5.661e^{-3} \pm 1.37e^{-3}$	$2.259e^{1} \pm 9.83e^{-1}$	$5.087e^{-1} \pm 5.72e^{-2}$
	svp	$5.616e^{-3} \pm 1.28e^{-3}$	$2.261e^{1} \pm 9.26e^{-1}$	$5.145e^{-1} \pm 5.64e^{-2}$
	spp2*	$3.850e^{-3} \pm 1.05e^{-3}$	$2.430e^{1} \pm 1.16e^{0}$	$6.284e^{-1} \pm 5.22e^{-2}$
	spp4*	$\mathbf{3.765e^{-3} \pm 1.09e^{-3}}$	$\mathbf{2.442e^{1} \pm 1.23e^{0}}$	$\mathbf{6.355e^{-1} \pm 5.62e^{-2}}$
	spp8*	$3.799e^{-3} \pm 1.04e^{-3}$	$2.436e^{1} \pm 1.16e^{0}$	$6.327e^{-1} \pm 5.38e^{-2}$

Table 15: **Galaxy: underrepresented-sample focusing improves accuracy and stabilizes training.** We compare losses across four architectures (AE, VAE, SAE, MAE); rows with $*$ denote our *sppk* objective (memory size k). Considering mean$\pm$std, **AE** sees consistent gains with markedly lower variability: *spp2* attains the best MSE/PSNR/SSIM among AEs (9.19×10^{-4}, 31.71,dB, 0.829) with substantially smaller standard deviations than pixel/perceptual baselines. **VAE** benefits from larger memory: *spp8** yields the best MSE/PSNR (8.13×10^{-4}, 31.91,dB) and competitive SSIM (0.816), with variance comparable to or lower than L2/Focal. **SAE**: *spp8** provides the strongest triplet (1.22×10^{-3}, 30.70,dB, 0.818) while reducing spread relative to baselines. **MAE**: *spp8** achieves the top PSNR/SSIM (30.02,dB, 0.805), whereas *spp4** attains the lowest MSE (1.40×10^{-3}) with nearly identical variance—illustrating a controllable trade-off between mean accuracy and dispersion via memory size. Overall, prioritizing underrepresented samples improves average reconstruction (MSE↓, PSNR/SSIM↑) while keeping or lowering standard deviations, indicating better robustness across galaxy textures.

Model	Baseline	MSE	PSNR	SSIM
AE	L2	$1.465e^{-3} \pm 1.23e^{-4}$	$2.836e^{1} \pm 3.62e^{-1}$	$8.033e^{-1} \pm 1.38e^{-2}$
	Focal	$1.449e^{-3} \pm 1.03e^{-4}$	$2.840e^{1} \pm 3.06e^{-1}$	$8.087e^{-1} \pm 1.11e^{-2}$
	onbs	$2.476e^{-3} \pm 7.77e^{-4}$	$2.645e^{1} \pm 2.16e^{0}$	$7.642e^{-1} \pm 3.97e^{-2}$
	dcl	$1.470e^{-3} \pm 1.50e^{-4}$	$2.835e^{1} \pm 4.46e^{-1}$	$7.841e^{-1} \pm 1.52e^{-2}$
	svp	$1.460e^{-3} \pm 1.30e^{-4}$	$2.837e^{1} \pm 3.85e^{-1}$	$8.055e^{-1} \pm 1.42e^{-2}$
	spp2*	$1.464e^{-3} \pm 3.01e^{-4}$	$2.851e^{1} \pm 1.38e^{0}$	$8.267e^{-1} \pm 2.24e^{-2}$
	spp4*	$\mathbf{2.620e^{-4} \pm 2.81e^{-4}}$	$\mathbf{3.706e^{1} \pm 2.90e^{0}}$	$\mathbf{9.240e^{-1} \pm 3.00e^{-2}}$
	spp8*	$6.298e^{-4} \pm 7.51e^{-4}$	$3.455e^{1} \pm 4.52e^{0}$	$9.110e^{-1} \pm 4.51e^{-2}$
VAE	L2	$1.224e^{-3} \pm 2.12e^{-3}$	$3.019e^{1} \pm 2.70e^{0}$	$7.931e^{-1} \pm 5.30e^{-2}$
	Focal	$1.468e^{-3} \pm 2.13e^{-3}$	$2.928e^{1} \pm 2.58e^{0}$	$7.664e^{-1} \pm 5.47e^{-2}$
	onbs	$3.638e^{-3} \pm 7.04e^{-3}$	$2.598e^{1} \pm 3.01e^{0}$	$7.068e^{-1} \pm 4.80e^{-2}$
	dcl	$2.238e^{-3} \pm 4.07e^{-3}$	$2.851e^{1} \pm 3.61e^{0}$	$7.886e^{-1} \pm 5.62e^{-2}$
	svp	$1.165e^{-3} \pm 2.06e^{-3}$	$3.025e^{1} \pm 2.41e^{0}$	$7.840e^{-1} \pm 5.44e^{-2}$
	spp2*	$9.785e^{-4} \pm 2.34e^{-3}$	$3.116e^{1} \pm 2.57e^{0}$	$8.114e^{-1} \pm 5.14e^{-2}$
	spp4*	$8.682e^{-4} \pm 2.35e^{-3}$	$3.166e^{1} \pm 2.44e^{0}$	$\mathbf{8.163e^{-1} \pm 4.89e^{-2}}$
	spp8*	$\mathbf{8.126e^{-4} \pm 1.86e^{-3}}$	$\mathbf{3.191e^{1} \pm 2.45e^{0}}$	$8.156e^{-1} \pm 4.93e^{-2}$
SAE	L2	$2.179e^{-3} \pm 3.65e^{-3}$	$2.853e^{1} \pm 3.58e^{0}$	$7.945e^{-1} \pm 5.60e^{-2}$
	Focal	$3.497e^{-3} \pm 5.38e^{-3}$	$2.557e^{1} \pm 2.31e^{0}$	$7.822e^{-1} \pm 5.13e^{-2}$
	onbs	$3.353e^{-3} \pm 6.55e^{-3}$	$2.637e^{1} \pm 3.03e^{0}$	$7.151e^{-1} \pm 4.49e^{-2}$
	dcl	$2.884e^{-3} \pm 6.40e^{-3}$	$2.751e^{1} \pm 3.57e^{0}$	$7.923e^{-1} \pm 5.26e^{-2}$
	svp	$2.199e^{-3} \pm 3.72e^{-3}$	$2.851e^{1} \pm 3.58e^{0}$	$7.935e^{-1} \pm 5.54e^{-2}$
	spp2*	$6.160e^{-3} \pm 6.95e^{-3}$	$2.302e^{1} \pm 2.50e^{0}$	$7.397e^{-1} \pm 5.73e^{-2}$
	spp4*	$1.416e^{-3} \pm 2.71e^{-3}$	$3.022e^{1} \pm 3.50e^{0}$	$8.151e^{-1} \pm 5.22e^{-2}$
	spp8*	$\mathbf{1.220e^{-3} \pm 2.65e^{-3}}$	$\mathbf{3.070e^{1} \pm 3.27e^{0}}$	$\mathbf{8.185e^{-1} \pm 5.11e^{-2}}$
MAE	L2	$1.868e^{-3} \pm 2.62e^{-3}$	$2.892e^{1} \pm 3.38e^{0}$	$7.954e^{-1} \pm 5.64e^{-2}$
	Focal	$2.401e^{-3} \pm 2.52e^{-3}$	$2.727e^{1} \pm 2.72e^{0}$	$7.503e^{-1} \pm 5.03e^{-2}$
	onbs	$3.060e^{-3} \pm 6.32e^{-3}$	$2.692e^{1} \pm 3.20e^{0}$	$7.618e^{-1} \pm 5.05e^{-2}$
	dcl	$2.123e^{-3} \pm 3.81e^{-3}$	$2.864e^{1} \pm 3.55e^{0}$	$7.887e^{-1} \pm 6.39e^{-2}$
	svp	$1.903e^{-3} \pm 3.03e^{-3}$	$2.894e^{1} \pm 3.43e^{0}$	$7.907e^{-1} \pm 5.46e^{-2}$
	spp2*	$1.408e^{-3} \pm 1.42e^{-3}$	$2.994e^{1} \pm 3.35e^{0}$	$8.050e^{-1} \pm 5.40e^{-2}$
	spp4*	$\mathbf{1.399e^{-3} \pm 1.39e^{-3}}$	$2.993e^{1} \pm 3.30e^{0}$	$8.048e^{-1} \pm 5.33e^{-2}$
	spp8*	$1.411e^{-3} \pm 1.55e^{-3}$	$\mathbf{3.002e^{1} \pm 3.42e^{0}}$	$\mathbf{8.052e^{-1} \pm 5.41e^{-2}}$