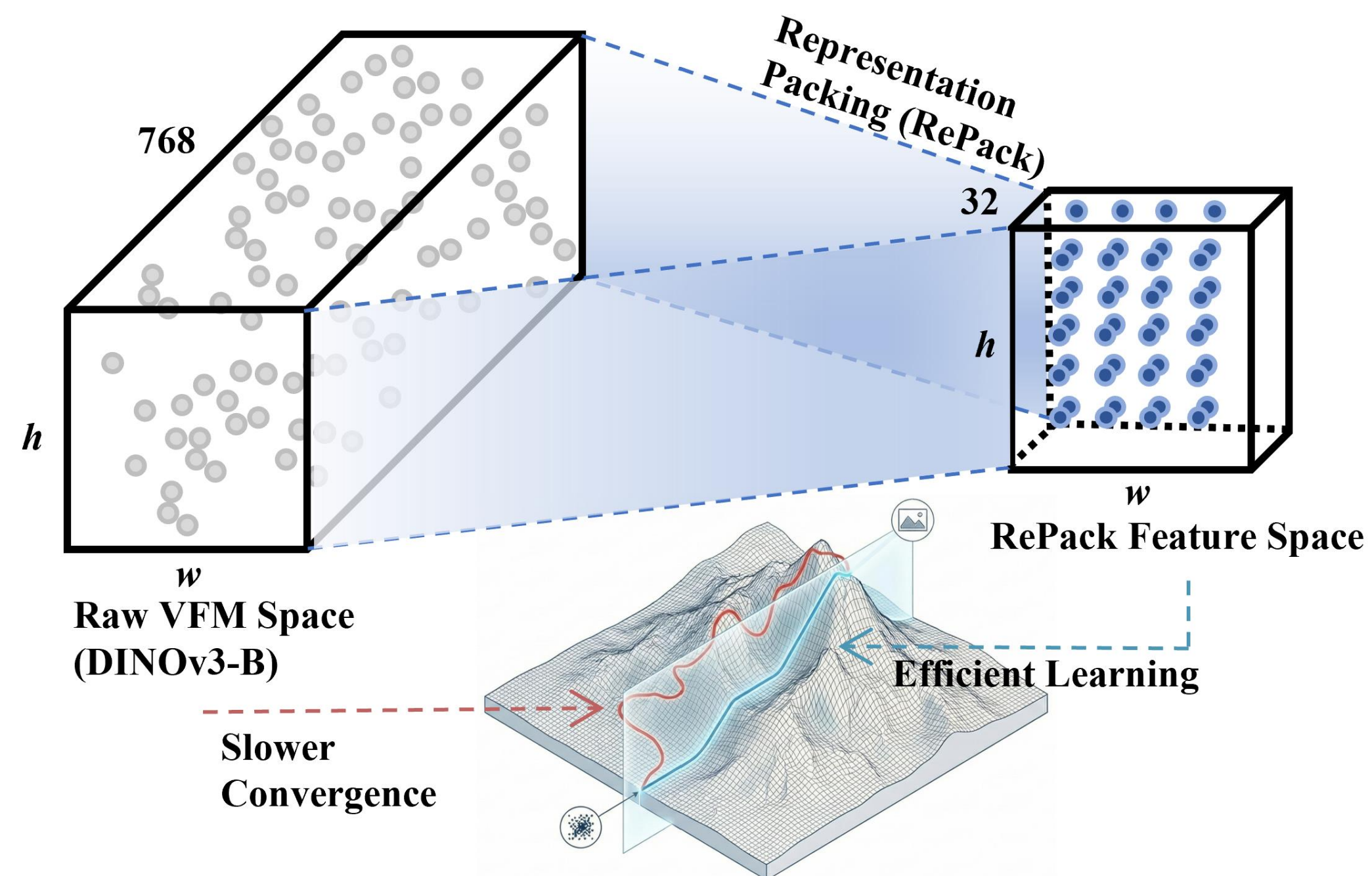


1. Motivation

Raw Vision Foundation Model (VFM) features from models like DINOv3 are semantically rich but high-dimensional and redundant (768 channels). Modeling such features directly with diffusion is inefficient and slows down convergence.

From Raw VFM Space to RePack Feature Space

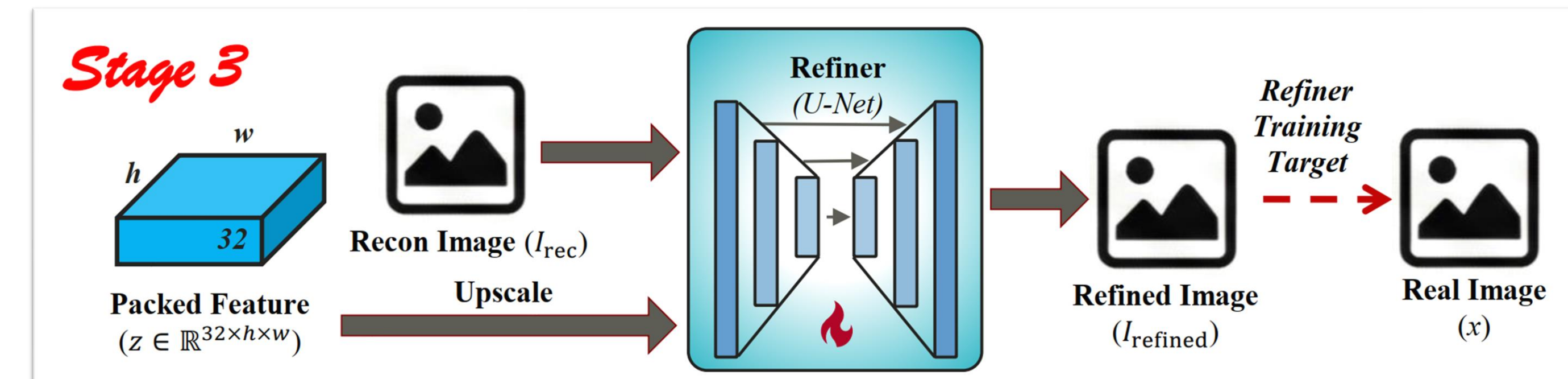
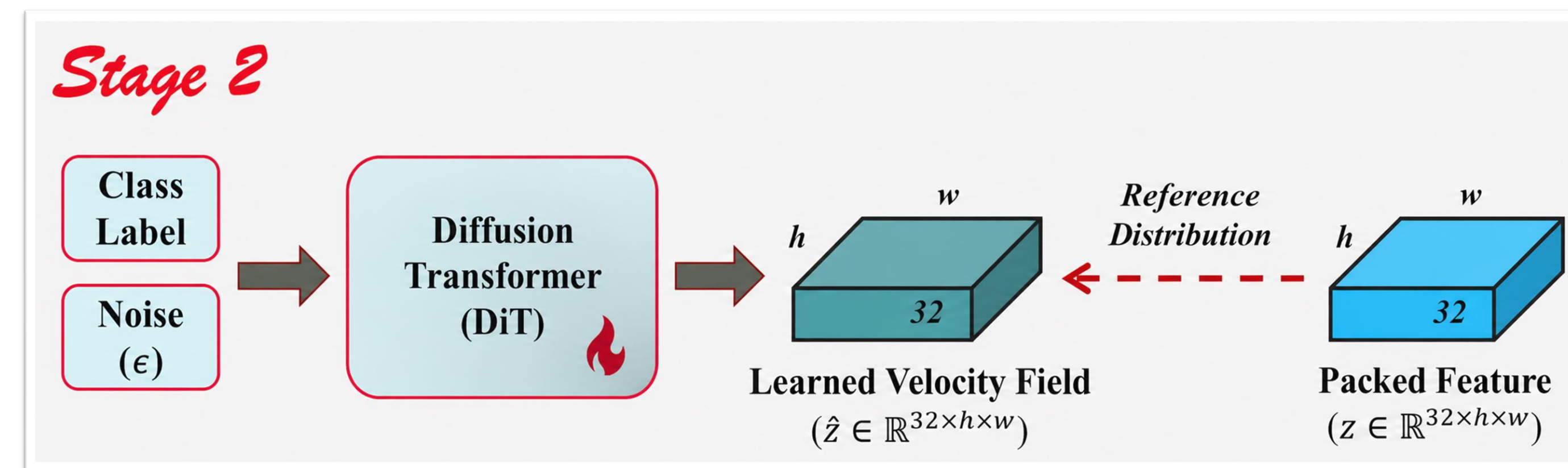
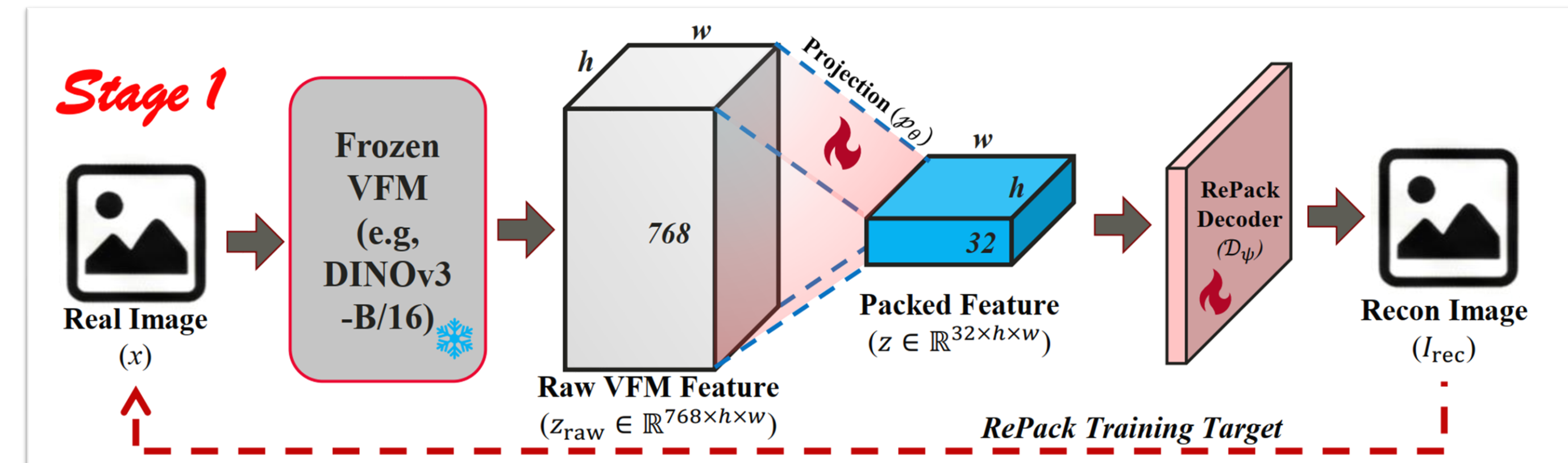


RePack compresses 768-channel raw VFM features into a compact 32-channel semantic manifold.

Key Contribution

- Propose **RePack**, a lightweight representation packing module for semantic compression.
- Train diffusion in a compact latent space for much faster convergence.
- Add a **Latent-Guided Refiner** to recover high-frequency details.
- Achieve strong ImageNet 256×256 performance with excellent training efficiency.

2. RePack then Refine Framework



Framework. RePack first compresses raw VFM features into a compact semantic latent:

$$z_{\text{raw}} = E_{\phi}(x) \in \mathbb{R}^{768 \times h \times w}, \quad z = P_{\theta}(z_{\text{raw}}) \in \mathbb{R}^{32 \times h \times w}.$$

The DiT then learns generation in the packed latent space:

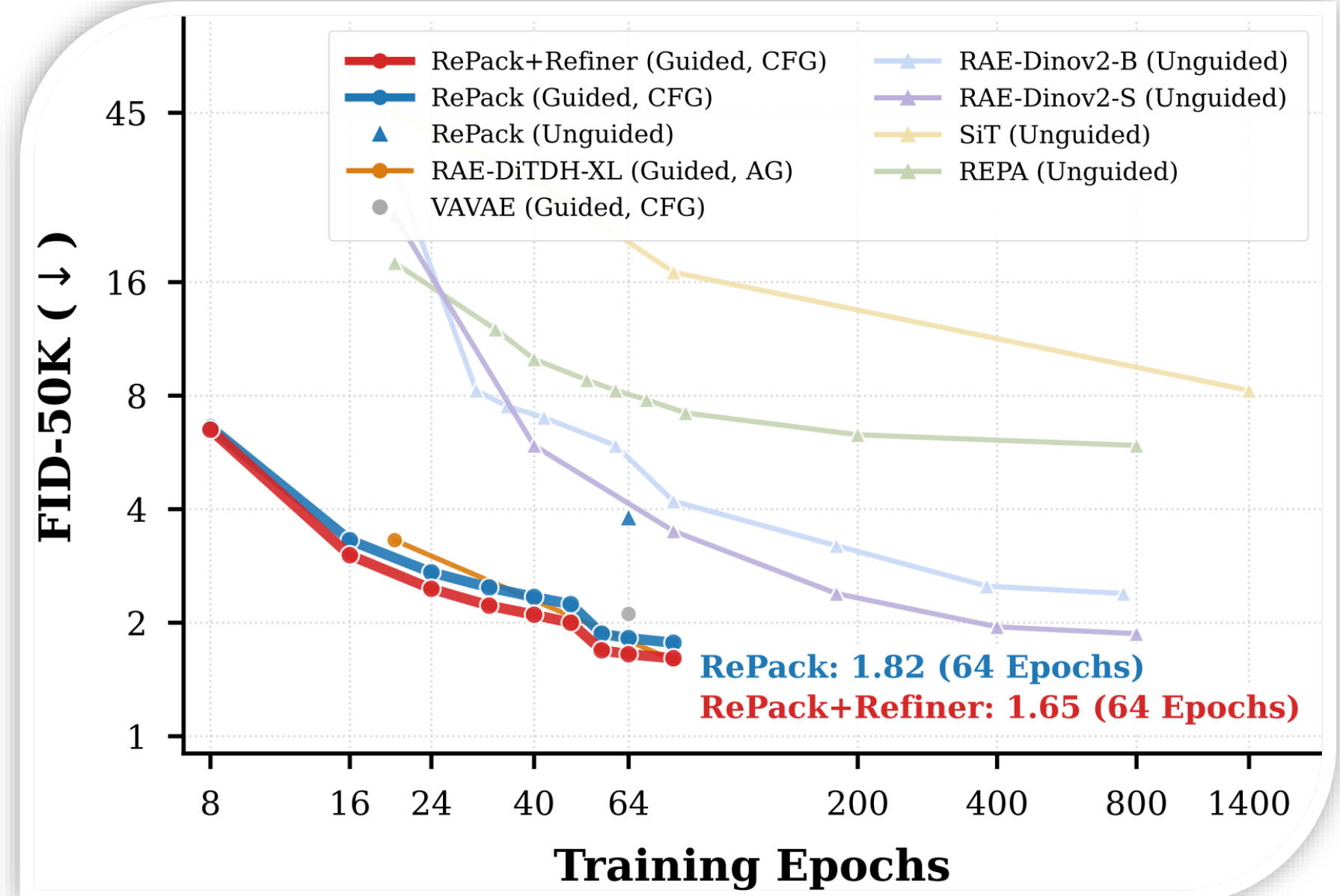
$$\hat{z} = \text{DiT}_{\omega}(\epsilon, y).$$

Finally, the decoder and Refiner recover image fidelity:

$$I_{\text{base}} = D_{\psi}(\hat{z}), \quad I_{\text{final}} = R_{\eta}(I_{\text{base}}, \text{Up}(\hat{z})).$$

Key idea: compress semantics first, model diffusion efficiently, then refine high-frequency details.

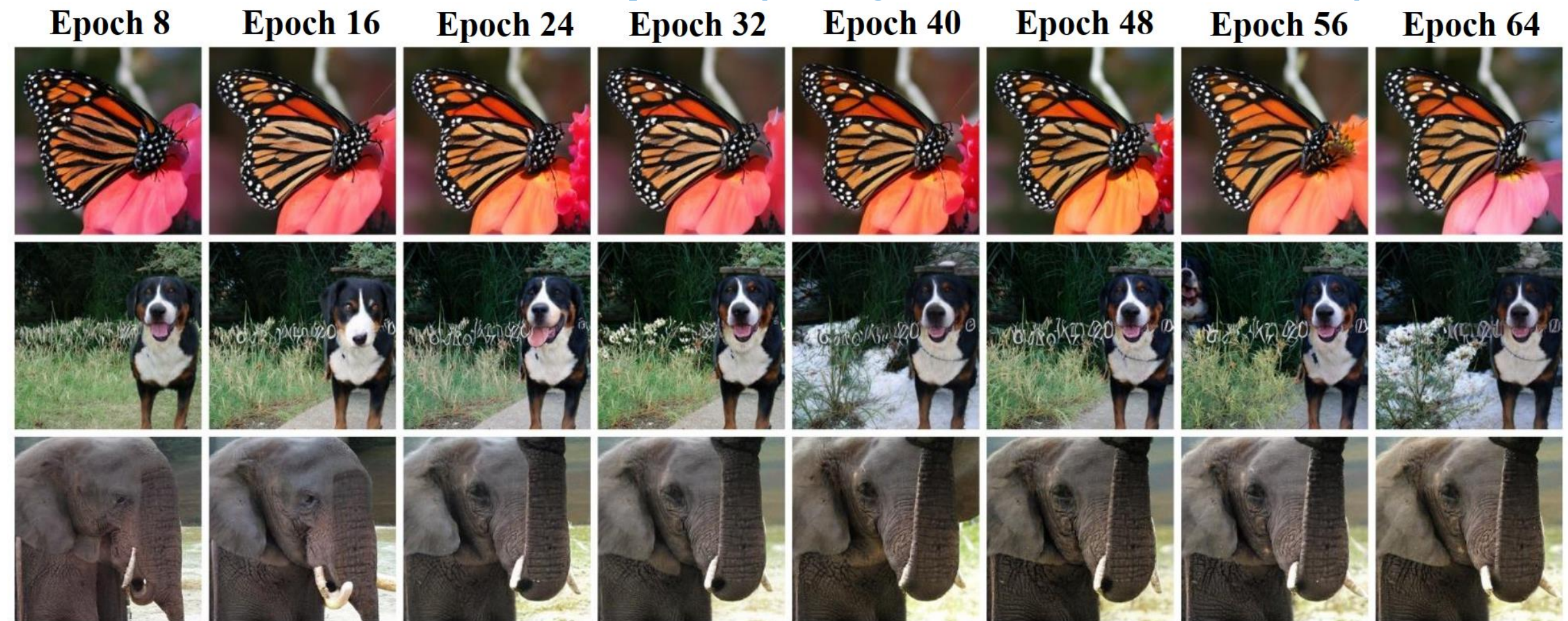
3. Experimental Results



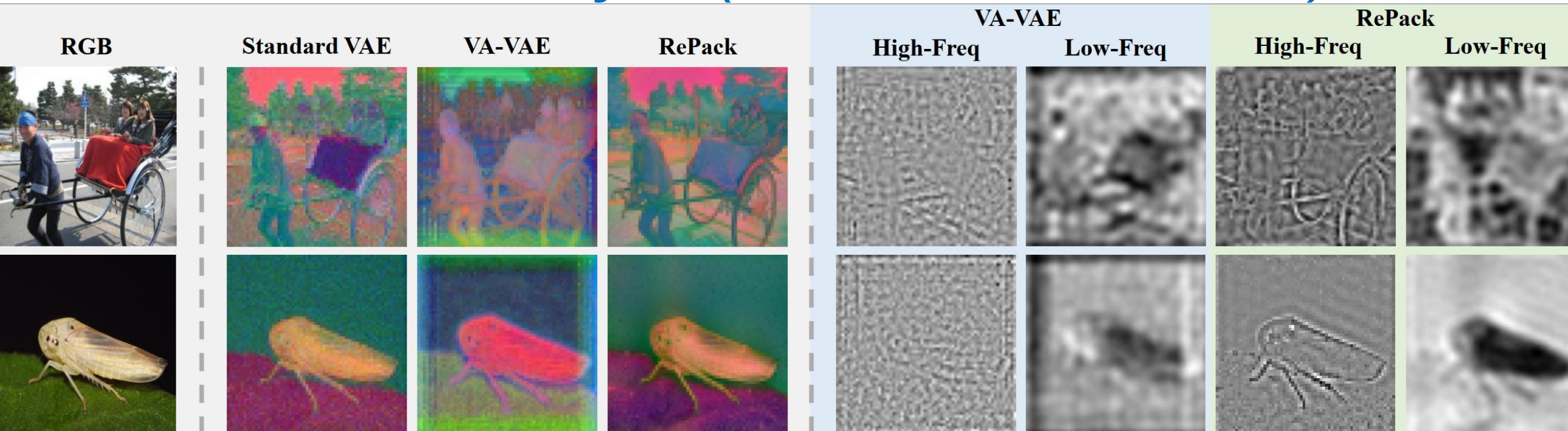
Headline Results

- RePack**
FD1.82 @ 64 epochs
- RePack + Refiner**
FD1.65 @ 64 epochs

Qualitative Samples (ImageNet-1K, 256×256)



Latent Analysis (DINOv3-B/16 Features)



RePack yields cleaner, more structured latent representations and suppresses aliasing noise while preserving semantic contours.

Takeaways



- Compressing VFM features makes diffusion learning easier and faster.
- RePack balances semantic purity and generation quality.
- The Refiner restores texture details and further improves fidelity.