# FINE: Factorizing Knowledge for Initialization of Variable-sized Diffusion Models

Supplementary Material

## A. Training Details

### A.1. Details of Knowledge Factorization

Algorithm 1 outlines the pseudo code for factorizing knowledge and encapsulating size-agnostic knowledge into shared singular vectors $U_\star$ and $V_\star$, referred to as learngenes.

**Algorithm 1** Knowledge Factorization for Encapsulating Size-agnostic Knowledge

**Input**: A DiT model $\varepsilon_\theta$ with $L$ layers, training dataset $\{(z^{(i)}, c^{(i)})\}_{i=1}^{m}$, number of epochs $N_{\text{ep}}$, batch size $B$, learning rate $\alpha$

**Output**: Shared singular vectors $\mathcal{U}$ and $\mathcal{V}$ (i.e., learngene) encapsulating size-agnostic knowledge

Random initialize the set of shared singular vectors $\mathcal{U}$ and $\mathcal{V}$, and layer-specific singular values $\mathcal{S}$
Construct weight matrices $\theta$ using $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{S}$ according to Eq. (7)
**for** $ep = 1$ to $N_{\text{ep}}$ **do**
  **for** each batch $\{(z^{(i)}, c^{(i)})\}_{i=1}^{B}$ **do**
    **for** each $U_\star$, $V_\star$ and $\Sigma_\star^{(l)}$ in $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{S}$ **do**
      Update $W_\star^{(l)}$ in $\theta$ based on the current values of $U_\star$, $V_\star$ and $\Sigma_\star^{(l)}$ under the rule of Eq. (6)
    **end for**
    Perform a forward propagate $\varepsilon_\theta(z_t, t, c)$ through the model for each noisy input $z_t$ at timestep $t$
    Calculate $\mathcal{L}_{\text{batch}} = \frac{1}{B}\sum_{i=1}^{B}\mathcal{L}(\varepsilon, \varepsilon_\theta(z_t^{(i)}, c^{(i)}, t))$ according to Eq. (1)
    Backward propagate the loss $\mathcal{L}_{\text{batch}}$ to compute the gradients with respect to $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{S}$: $\nabla_{\mathcal{U}}\mathcal{L}_{\text{batch}}$, $\nabla_{\mathcal{V}}\mathcal{L}_{\text{batch}}$ and $\nabla_{\mathcal{S}}\mathcal{L}_{\text{batch}}$
    Update $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{S}$
      $\mathcal{U} := \mathcal{U} - \alpha \cdot \nabla_{\mathcal{U}}\mathcal{L}_{\text{batch}}$
      $\mathcal{V} := \mathcal{V} - \alpha \cdot \nabla_{\mathcal{V}}\mathcal{L}_{\text{batch}}$
      $\mathcal{S} := \mathcal{S} - \alpha \cdot \nabla_{\mathcal{S}}\mathcal{L}_{\text{batch}}$
  **end for**
**end for**

### A.2. Details of Downstream Datasets

Table 6 provides an overview of six downstream datasets: CelebA-HQ [22], LSUN-Bedroom, LSUN-Church [45], Hubble, MRI and Pokemon. LSUN-Bedroom and LSUN-Church are subsets of the Large-Scale Scene Understanding (LSUN) dataset [45], containing scene images of bedrooms and churches, respectively, with a resolution of $256 \times 256$ pixels. CelebA-HQ is a high-quality variant of the CelebA dataset [31], featuring large-scale facial images of celebrities, resized to $256 \times 256$ pixels.

Table 6. Characteristics of downstream datasets.

| Dataset | Total | Resolution |
|---|---|---|
| CelebA | 30,000 | 256×256 |
| LSUN-Bedroom | 3,033,042 | 256×256 |
| LSUN-Church | 126,227 | 256×256 |
| Hubble | 2706 | 256×256 |
| MRI | 3753 | 256×256 |
| Pokemon | 833 | 256×256 |

Table 7. Hyper-parameters for FINE factorizing knowledge on ImageNet-1K.

| Training Settings | Configuration |
|---|---|
| optimizer | AdamW |
| learning rate | 1e-4 |
| weight decay | 0 |
| optimizer momentum | 0.9 |
| batch size | 64 |
| training steps | 300K |
| drop path | 0.5 |
| sigma share | B: 100 \| L:148 |
| class dropout | 0.1 |
| vae | stabilityai / sd-vae-ft-ema |

### A.3. Hyper-parameters

Table 7 and Table 8 present the basic settings, including batch size, training steps, optimizer and other settings for FINE encapsulating size-agnostic knowledge into shared singular vectors $U_\star$ and $V_\star$ and training the models initialized with learngenes on various datasets, respectively.

### A.4. Compared methods

**Direct Initialization.** Models are initialized using predefined rules (e.g., He-Init [5]) or observed patterns (e.g., Mimetic Init [43]).

**Conventional Knowledge Transfer.** These methods focus on transferring knowledge from pre-trained models to new ones. For example, LiGO [46] trains larger model by leveraging knowledge from a pre-trained smaller one, while Share init [28] reuses trained blocks across multiple layers to initialize models with variable depths. Laptop-Diff [56] and BK-SDM [25] adopt a prune-then-distill strategy, in which the pre-trained model is first layer-wise pruned to match the target architecture, followed by knowledge distillation to recover performance.

Table 8. Hyper-parameters for neural networks trained on downstream datasets.

| Dataset | Batch Size | Training Steps | Learning Rate | Drop Last | Droppath Rate | Optimizer |
|---|---|---|---|---|---|---|
| CelebA | 64 | 100K | 1e-4 | True | 0.1 | AdamW |
| Bedroom | 64 | 150K | 1e-4 | True | 0.1 | AdamW |
| Church | 64 | 150K | 1e-4 | True | 0.1 | AdamW |
| Hubble | 64 | 20K | 1e-4 | True | 0.1 | AdamW |
| MRI | 64 | 20K | 1e-4 | True | 0.1 | AdamW |
| Pokemon | 64 | 20K | 1e-4 | True | 0.1 | AdamW |

**Learngene-Based Methods.** We adapt existing learngene methods to diffusion models in this paper. Heur-LG [47] selects layers with minimal gradient changes during training as the learngenes, while Auto-LG [48] employs meta-learning to identify layers in the pre-trained model that share similar representations to those required by downstream tasks. TLEG [51] builds on the linear relationships observed between different layers of transformer architectures.

These methods provide diverse strategies for initializing diffusion models, each with varying reliance on prior knowledge and pre-learned patterns, advancing the state-of-the-art in model initialization.

## B. Additional Results

We provide additional images generated by the DiT-L/2 model, initialized with FINE at a resolution of 256×256, as illustrated in Figure 6-13.

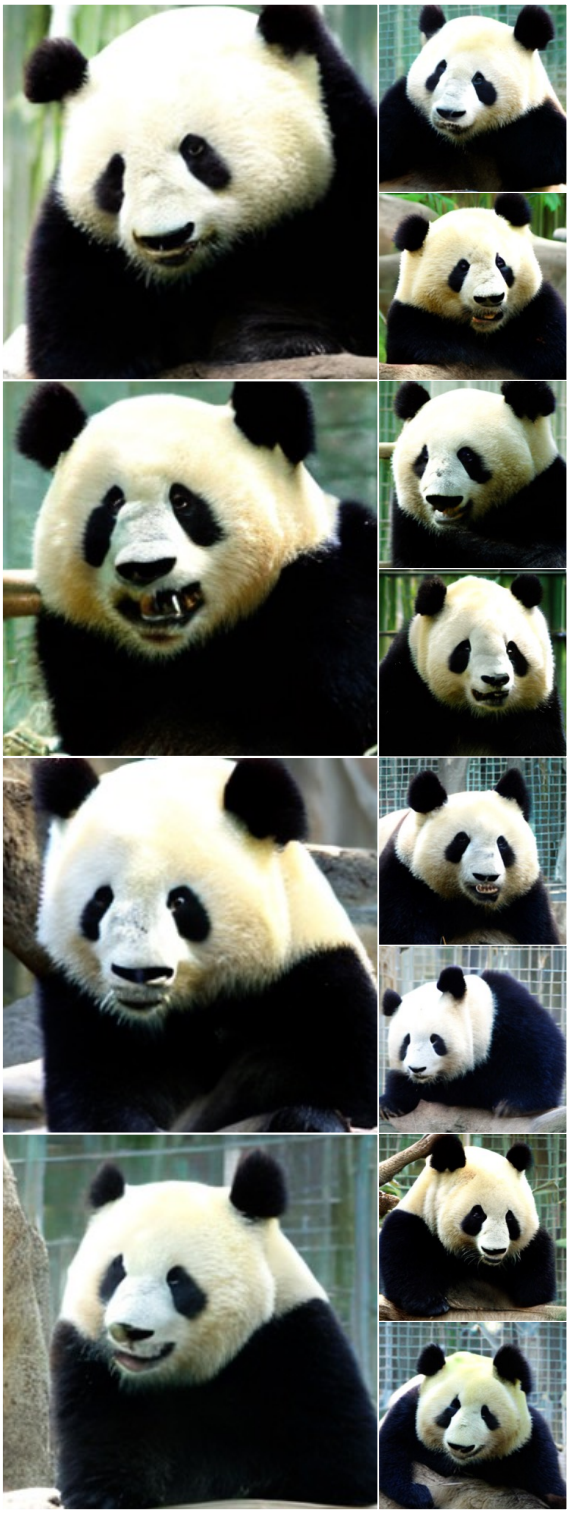

Figure 6. Images of n02510455 generated by FINE.

Figure 7. Images of n01514668 generated by FINE.

Figure 8. Images of n01534433 generated by FINE.

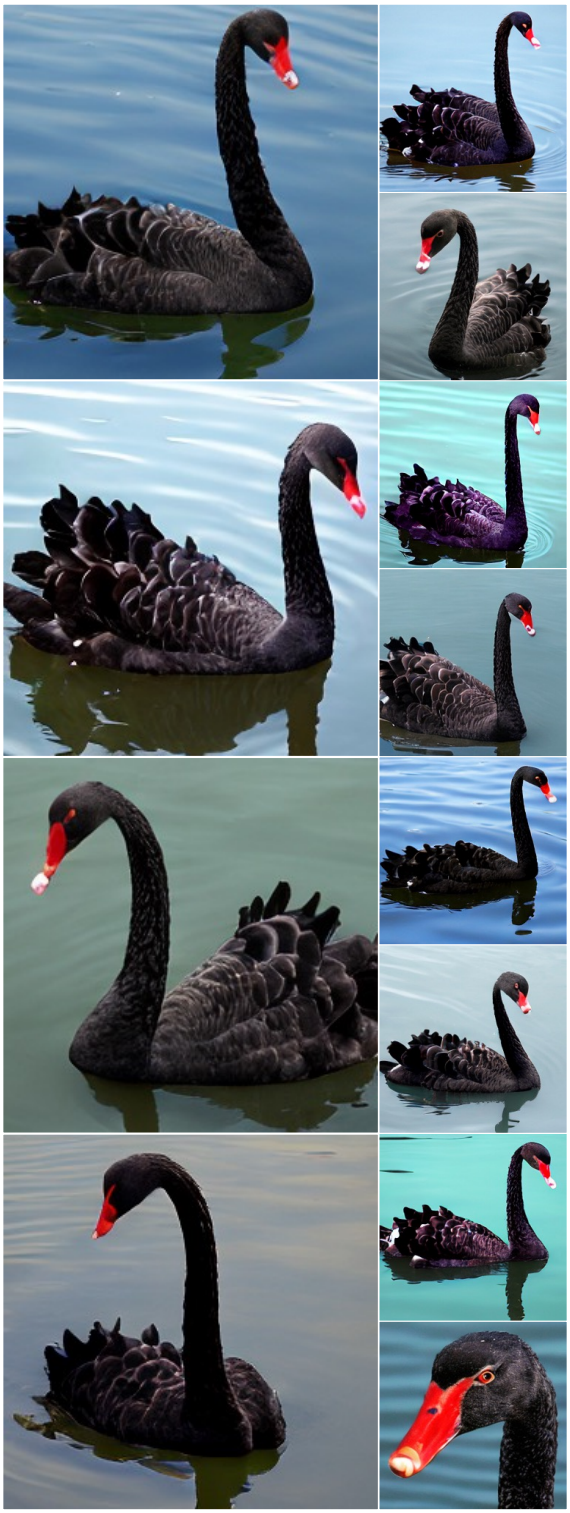

Figure 9. Images of n01860187 generated by FINE.

Figure 10. Images of n02002556 generated by FINE.

Figure 11. Images of n11939491 generated by FINE.

Figure 12. Images of n02782093 generated by FINE.

Figure 13. Images of n04335435 generated by FINE.