A. Formulation of (9)

Structural risk minimization principle minimizes the upper bound of the true risk on unseen data distribution, where the bound can be written as follows for a dataset containing n samples with the probability at least $1-\delta$:

$$E(J(x)) \leqslant \overline{E(J(x))} + 2R_n(\mathcal{F}) + \sqrt{\frac{\ln 1/\delta}{n}}, \tag{13}$$

where $E(J(x))$ and $\overline{E(J(x))}$ respectively illustrate the true expectation of the risk J for real data distribution x and the empirical expectation of that for sampled data from x, and $R_n(\mathcal{F})$ is the Rademacher complexity over the function class $\mathcal{F}$. The SRM principle requires the data to be sampled from i.i.d. distribution, while the latent images in selected timesteps should be more informative and representative. In order to extend the SRM principle in active timestep selection, we omitted x to reformulate the risk bound inequality:

$$E(J) \leqslant (E(J) - E_T(J)) + \overline{E_T(J)} + \mathcal{R}_0, \tag{14}$$

where $E(J)$ and $E_T(J)$ are the true risk of all latent images and the sampled data. $\mathcal{R}_0 = 2R_n(\mathcal{F}) + \sqrt{\frac{\ln 1/\delta}{n}}$ demonstrates the complexity of the diffusion model in the reverse process. In diffusion models, the data x consists of input samples z and target samples y, we can rewrite the first term of (14) as follows:

$$E(J) - E_T(J) = \int g(z)p(X)dz - \int g(z)p(X_s)dz, \tag{15}$$

where we rewrite $p(z|z \in X)$ and $p(z|z \in X_s)$ as $p(X)$ and $p(X_s)$ respectively for simplicity. X and X_s are the distribution of latent images generated in all timesteps and the selected ones respectively. As $g(z) = \int J \cdot p(y|z)dy$ is bounded and measurable, a bounded and continuous function $\hat{g}(z)$ can guarantee the boundness of (15):

$$\begin{aligned} E(J) - E_T(J) &\leqslant \sup_{\hat{g}(z)}[\int g(z)p(X)dz - \int g(z)p(X_s)dz] \\ &= MMD(p(X), p(X_s)), \end{aligned} \tag{16}$$

where $MMD(p(X), p(X_s))$ represents the maximum mean discrepancy between distribution $p(X)$ and $p(X_s)$. Finally, we rewrite the SRM principle in the following way:

$$E(J) \leqslant \overline{E_T(J)} + MMD(p(X), p(X_s)) + \mathcal{R}_0, \tag{17}$$

where we omit the data distribution x for simplicity. $\overline{E_T(J)}$ denotes the empirical risk of the latent images of selected timesteps for noise estimation.

B. Formulation about (11)

The definition of maximal mean discrepancy can be written as follows, where we denote $MMD(p(X), p(X_s))$ as M for simplicity:

$$\begin{aligned} \min_t M &= \sup ||\frac{1}{|U|}\sum_{\boldsymbol{x}_t \in U} \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t) - \frac{1}{|\mathcal{S}|}\sum_{\boldsymbol{x}_t \in \mathcal{S}} \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t)|| \\ &= \sup ||E_{\boldsymbol{x}_t \in U}(\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t)) - E_{\boldsymbol{x}_t \in S}(\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t))||, \end{aligned} \tag{18}$$

where U means the full set containing all original latent and timesteps for calibrating image selection, and $|\cdot|$ represents the number of elements in the set. $||\cdot||$ represent for L2 norm calculation. The upper bound of the first term of formula (18) can be written based on the upper confidence bound (UCB) principle as follows:

$$E_{\boldsymbol{x}_t \in U}(\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t)) = E_{\boldsymbol{x}_t \in S}(\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t)) + \varphi\sqrt{\frac{\ln N}{N_t + 1}}, \tag{19}$$

where N and N_t denote the number of sampling times for the t_{th} timestep and the total sampling times in calibration set construction respectively. φ is a constant in timestep sampling to achieve the exploitation-exploration trade-off in UCB principle. $\sqrt{\frac{\ln N}{N_t+1}}$ denotes for uncertainty between the distribution of the full set and the selected samples, which is reduced as the sampling times for the t_{th} timestep raise. N in formula (19) is designed to further explore the timesteps with more uncertainty, when the indeterminacy rises as t_{th} timestep is not selected. However, timestep t is large in diffusion models and the number of selected times N_t is always small for calculating the square root, which leads to large N and instability of uncertainty for the calibration set construction. Therefore, we simplify the design of uncertainty and obtain formula (11) in the paper as follows:

$$\min_t M = \frac{\varphi}{N_t + 1} \propto \frac{1}{N_t + 1}, \tag{20}$$

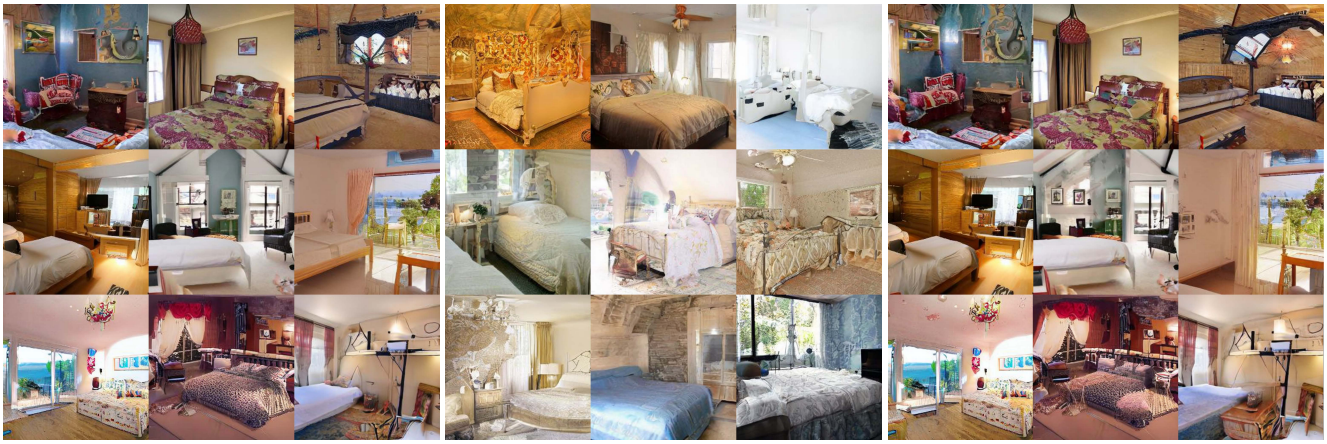
where we expect to select latent images in the timestep with few sampling times to further minimize the maximal mean discrepancy with high marginal benefits.

C. Samples

Additional samples: We show more samples generated by the 6-bit quantized LDM-4 diffusion model with different post-training quantization methods in Figure 5 (256×256 Church), Figure 6 (256×256 Bedroom), Figure 7 (256×256 CelebA-HQ), and Figure 8 (256×256 ImageNet). Compared with the conventional quantization method in diffusion models, our APQ-DM can still achieve high-quality details in plausible images for various datasets with weights and activations in low bitwidths, which are semblable to the full precision ones.

(a) Full Precision (b) PTQ4DM(6-bit) (c) ADP-DM(6-bit)

Figure 5. 256 × 256 LSUN-Church samples from 100 step LDMs in 6-bit with different post-training quantization methods.



(a) Full Precision (b) PTQ4DM(6-bit) (c) ADP-DM(6-bit)

Figure 6. 256 × 256 LSUN-Bedroom samples from 100 step LDMs in 6-bit with different post-training quantization methods.

(a) Full Precision (b) PTQ4DM(6-bit) (c) ADP-DM(6-bit)

Figure 7. 256 × 256 CelebA-HQ samples from 100 step LDMs in 6-bit with different post-training quantization methods.

Figure 8. 256×256 ImageNet samples from 100 step LDMs in 6-bit with APQ-DM.