

卒業研究 1 第 14 回プレゼンテーションレポート

情報理工学部 3 年 後藤 健一郎

2026-01-30

Table of contents

要旨	1
課題と目的	2
用いたデータ	2
用いた手法	2
Latent Diffusion Model (LDM)	2
DDPM 学習・サンプリング	3
実験設定	4
実験結果と考察	4
CelebA-HQ	4
まとめ	9
やり残した課題・工夫点・苦労した点	9
やり残した課題	9
工夫した点	9
苦労した点	9
参考文献	9

要旨

本レポートでは Latent Diffusion Models の論文に基づき、CelebA-HQ (256px) を対象とした再現実験を報告する。VAE (stabilityai/sd-vae-ft-mse) で画像を潜在空間へ圧縮し、潜在空間上で U-Net による拡散学習を実施した。拡散過程は cosine スケジュール、時間ステップ数 $T=1000$ を採用し、DDPM の逆拡散で生成した。wandb で学習曲線と生成例を記録し、学習の進行に伴う生成品質の変化を観察した。今後は評価指標の導入と、生成速度・品質の改善が課題である。

課題と目的

本課題の目的は、LDM の再現実験を通じて、論文の内容を踏まえた実装を手元で構築することにある。また DDPM では非常に長い時間のかかる学習を LDM の軽量化により再現する。さらに、CelebA-HQ (256px) における学習挙動と生成品質の変化を観察し、学習設定（スケジュール・最適化・混合精度）が安定性に与える影響を検証する。

用いたデータ

本実験では CelebA-HQ の顔画像データセット (256px) を用いた。

用いた手法

Latent Diffusion Model (LDM)

LDM は VAE により画像を潜在空間に圧縮し、潜在空間上で拡散モデルを学習する。低次元の latent space で学習を行うことによる学習を行うことにより、計算量とメモリ負荷を低減しつつ、生成品質を維持できる。アーキテクチャは元論文に準拠した。

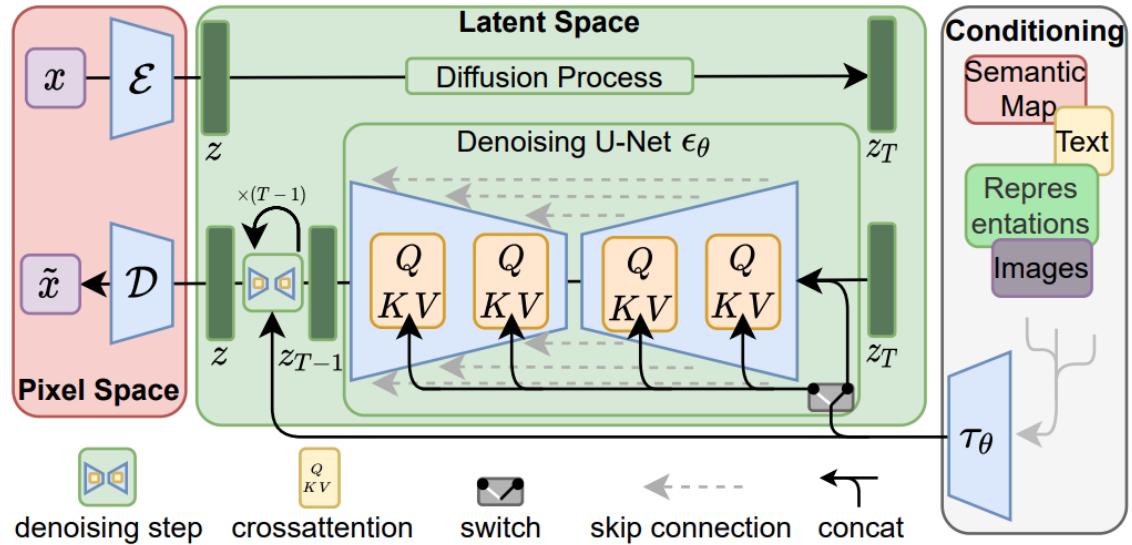


Figure1: 元論文に準拠した LDM の全体アーキテクチャ

潜在空間学習は VAE によるエンコード・デコードで構成し、拡散モデルは U-Net と Attention を組み合わせた。ノイズスケジュールには cosine を採用し、時間ステップ数は $T=1000$ とした。

VAE の downsample factor は自動算出され、本実験では $f=8$ となるため、256px 入力から 32×32 の潜在表現を得る。

また、学習開始前には VAE の latent space からの再構築性を確認し、入力と復元結果を比較した。

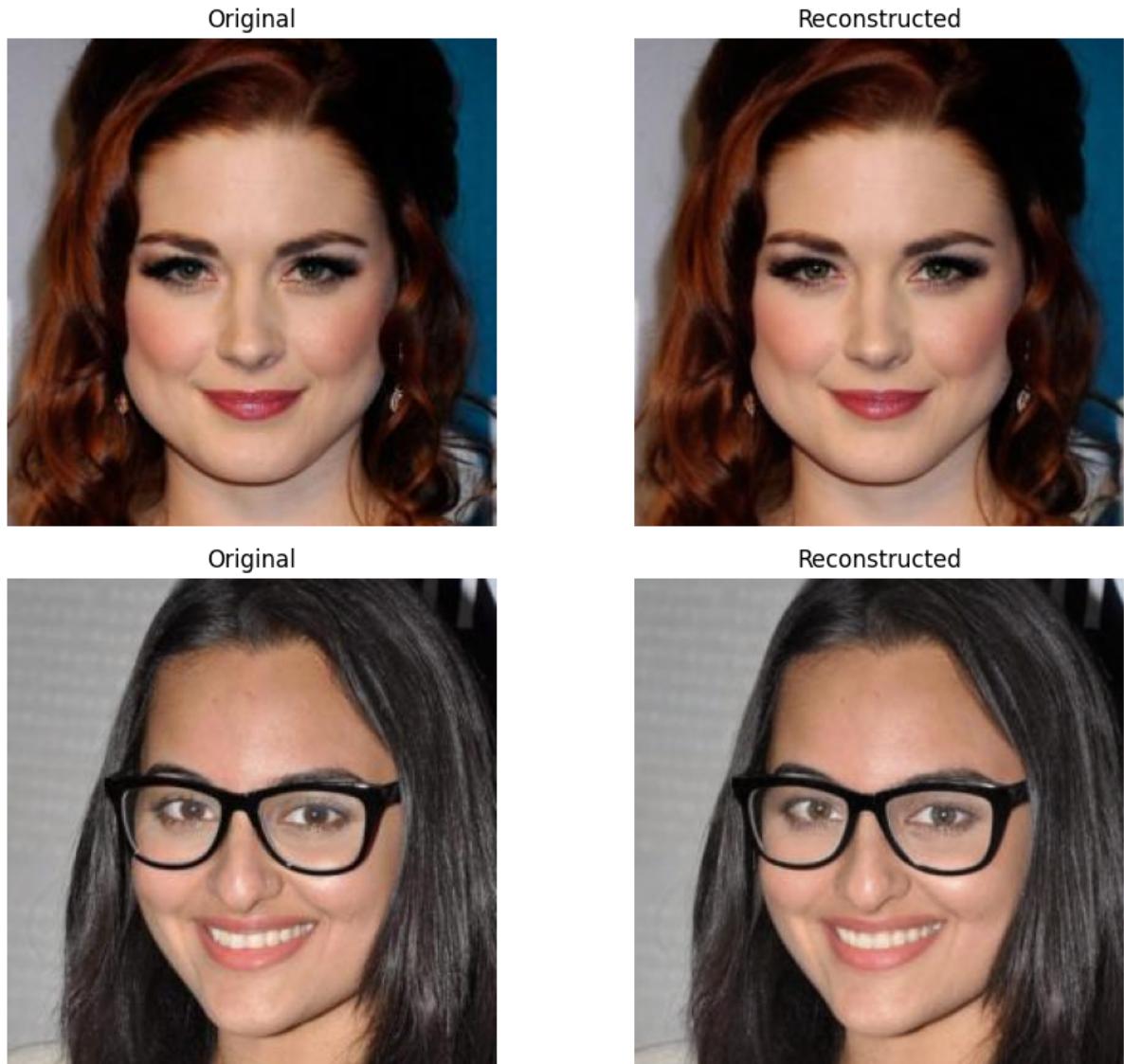


Figure2: VAE による入力画像と再構成画像の比較

DDPM 学習・サンプリング

拡散過程は DDPM の前向き過程でノイズ付加を行い、逆過程で生成を行う。学習は AdamW を用い、CosineAnnealingLR で学習率を調整した。

実験設定

表 1 に主要な実験設定を示す。

項目	設定値
計算環境	Google Colab (A100)
データセット	CelebA-HQ
解像度	256px
VAE	stabilityai/sd-vae-ft-mse (AutoencoderKL)
Downsample factor	8
Latent size	32×32
Batch	64
学習率	2e-4
Optimizer	AdamW
Scheduler	CosineAnnealingLR (T_max=100000)
Total steps	100000
Grad clip	1.0
Timesteps	1000
Beta schedule	cosine
Mixed precision	on (GPU 時)
Sampler	DDPM

実験結果と考察

CelebA-HQ

学習初期はノイズが強いが、学習の進行に伴って顔の輪郭や髪型といった大域構造が徐々に現れる傾向が観察された。一方で、生成品質の向上には長い学習時間と計算資源が必要であり、効率面が課題として残った。

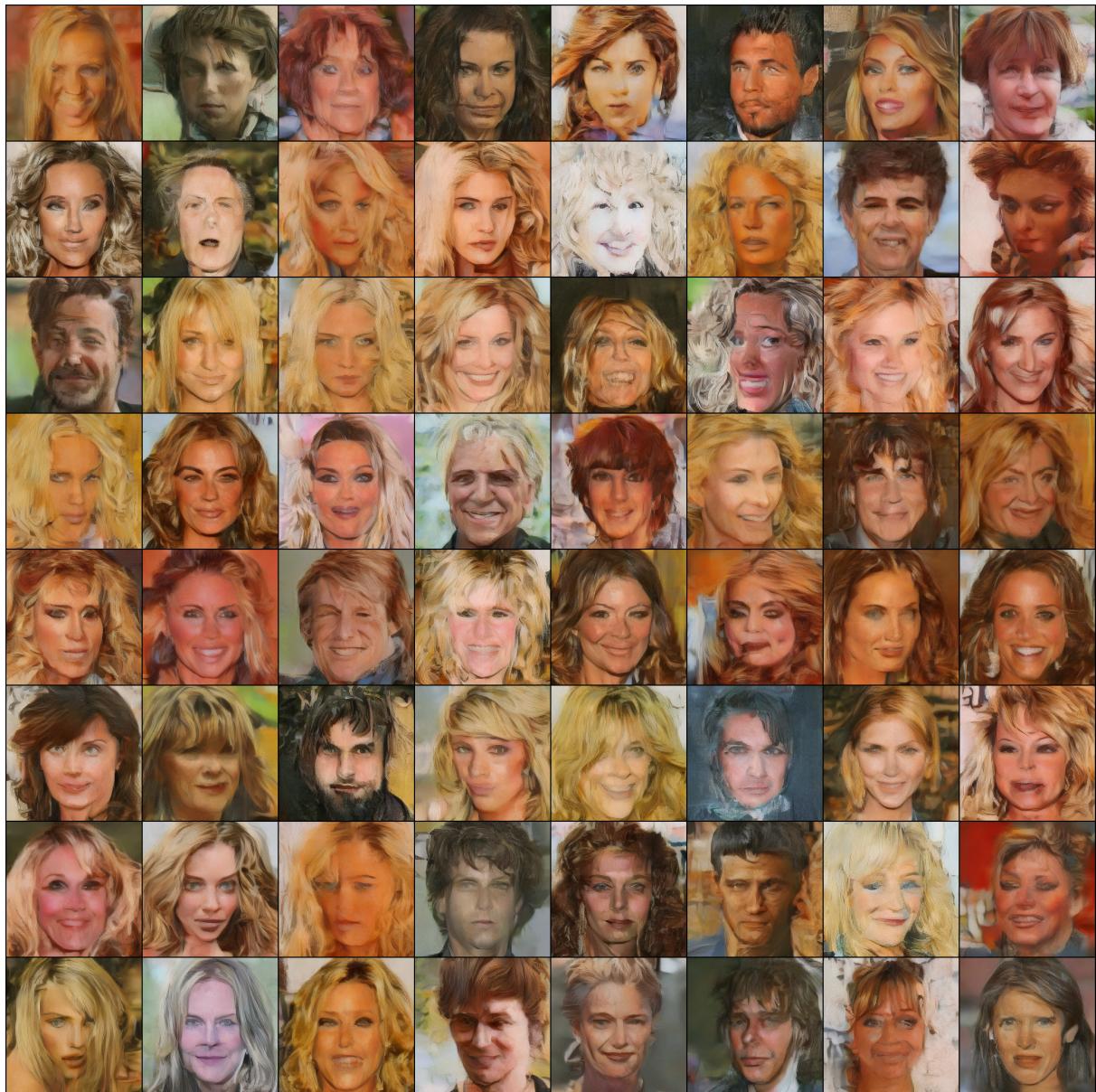


Figure3: 学習ステップ 10k における生成例

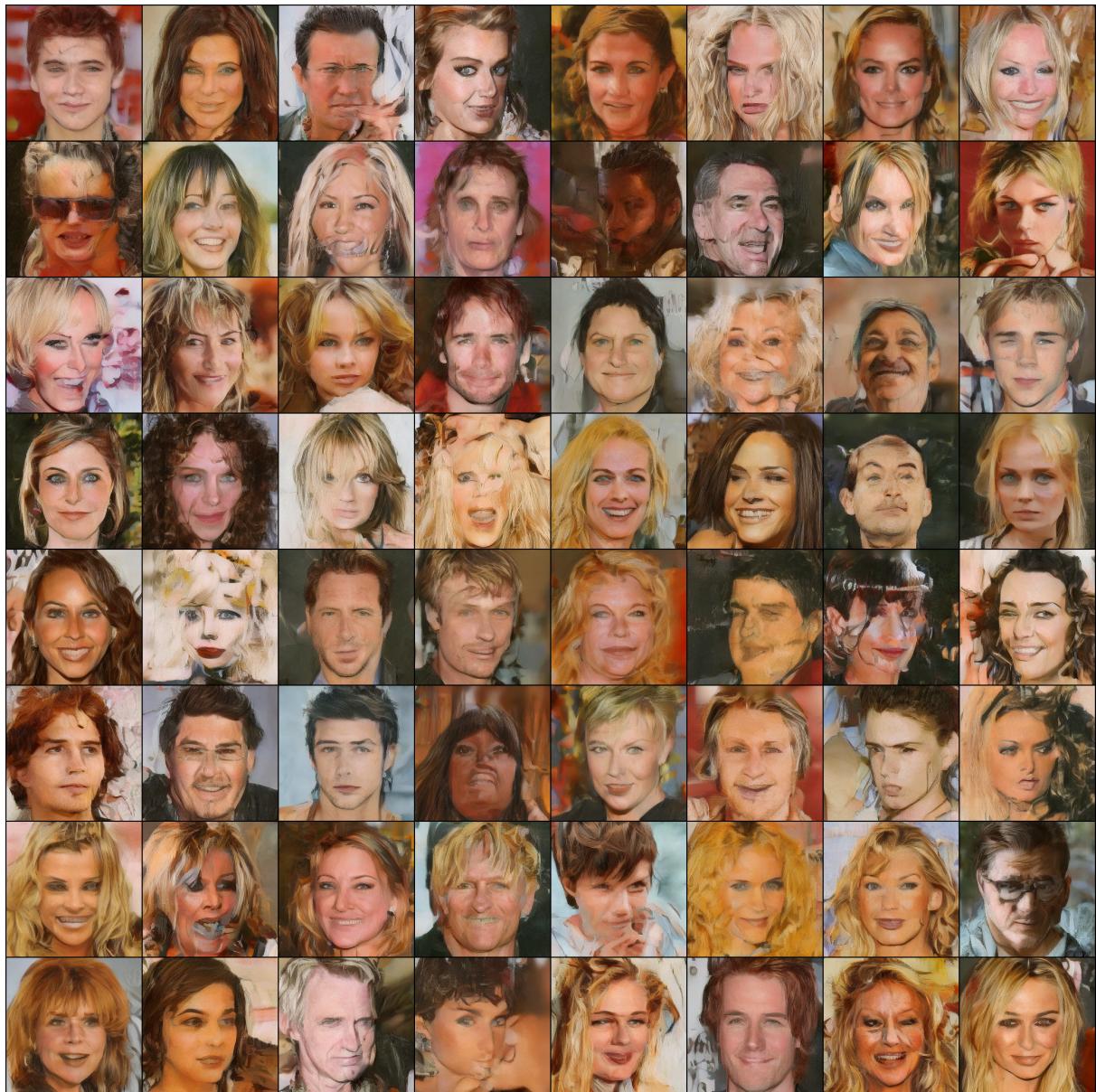


Figure4: 学習ステップ 50k における生成例

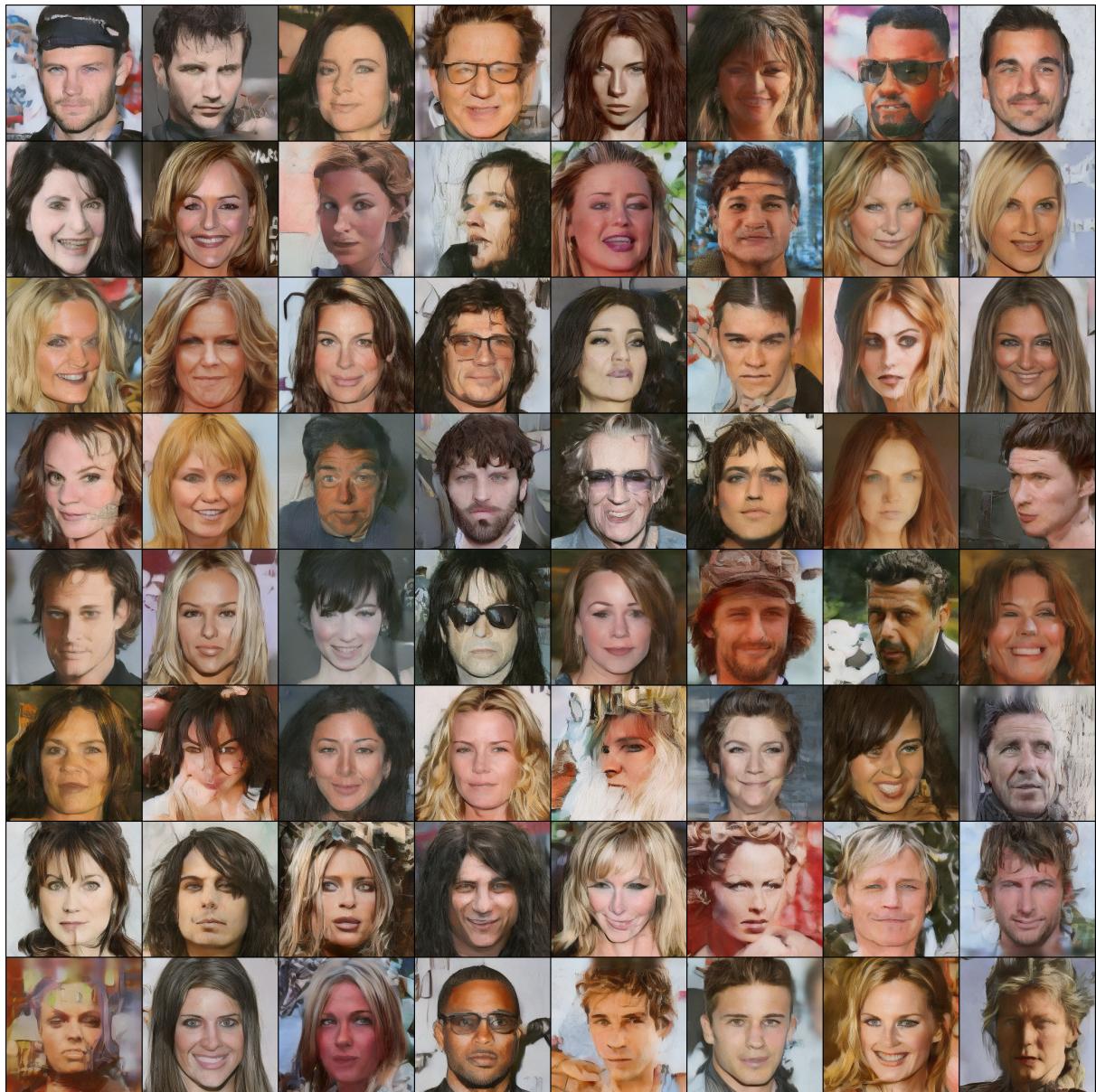


Figure5: 学習ステップ 100k における生成例

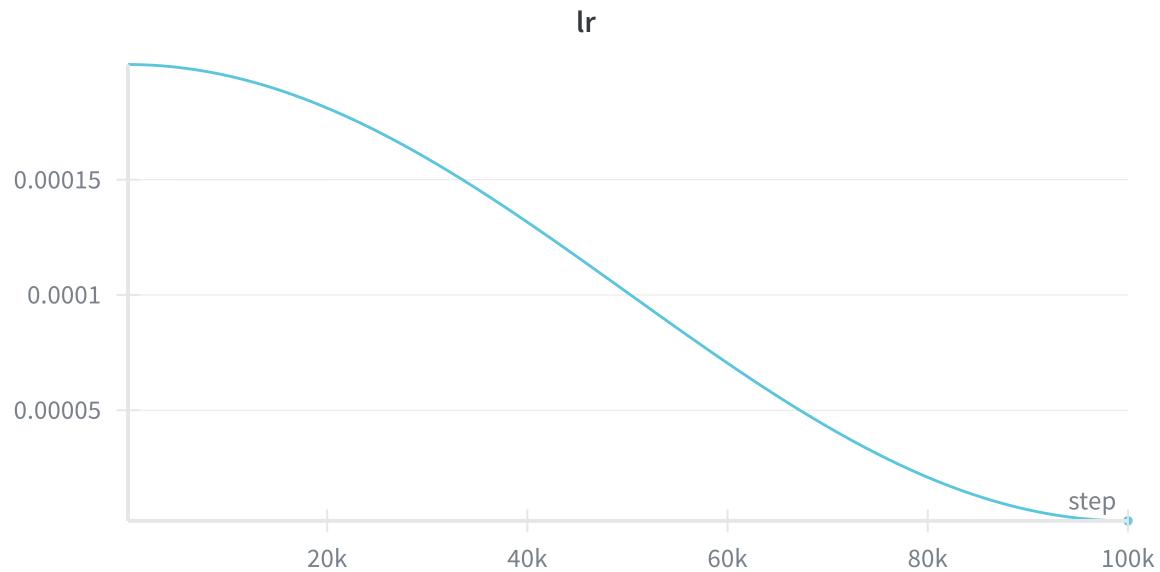


Figure6: 学習率の推移

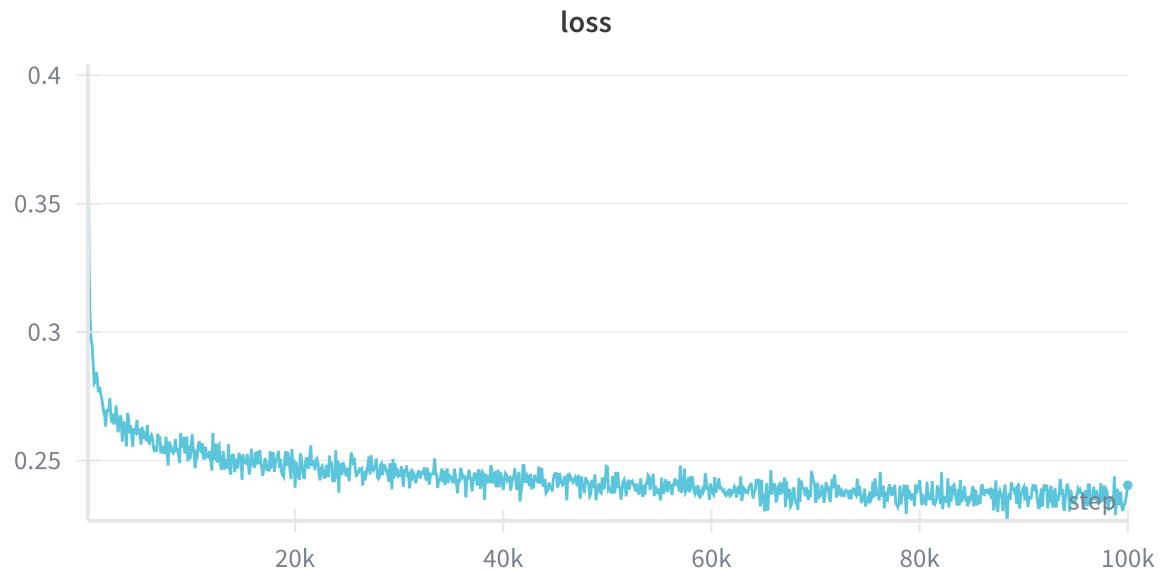


Figure7: Loss の推移

■考察

256px 入力に対して latent size を 32×32 に保てるため、潜在空間での拡散学習は一定の情報量を保持できる。さらに、cosine スケジュールと混合精度の組み合わせにより学習を安定化できた。一方で DDPM の逆拡

散は 1000 ステップを要するため生成が重く、今後は高速サンプラー（DDIM など）の導入やステップ削減が有効と考えられる。

まとめ

CelebA-HQ (256px) に対して LDM の再現実験を行い、潜在空間での学習のパイプラインを構築した。32×32 の潜在表現を用いることで大域構造の再現が可能であることを確認した。

やり残した課題・工夫点・苦労した点

やり残した課題

今後の課題として、FID などの定量評価指標の導入が必要である。生成速度改善のためには高速サンプラー（DDIM 等）の実装が有効と考えられる。また、学習ステップ数や学習率の最適化も未着手であり、特に学習終盤の学習率が小さすぎる点は改善余地がある。さらに、別データセットでの再現性確認も行いたい。

工夫した点

VAE による潜在表現の事前計算とキャッシュにより学習効率を向上させた。加えて、混合精度学習と勾配クリッピングを併用し、学習の安定性を確保した。

苦労した点

CelebA-HQ のデータ取得・前処理に時間を要した。また、高解像度設定により学習・生成が重く、試行回数を確保しづらかった。

参考文献

1. Ho, J., Jain, A., & Abbeel, P. (2020). *Denoising Diffusion Probabilistic Models*. NeurIPS.
2. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). *High-Resolution Image Synthesis with Latent Diffusion Models*. CVPR.
3. Kingma, D. P., & Welling, M. (2014). *Auto-Encoding Variational Bayes*. ICLR.