BMI / CS 771: Homework Assignment 4 Report

Qinxinghao Chen Handan Hu Chongwei Liu Bohan Wen

December 2025

Acknowledgment of AI Assistance

This assignment was completed with the assistance of OpenAI's ChatGPT (GPT-5). The tool was used throughout the process to brainstorm ideas, clarify relevant concepts, explain code logic, and check grammar. All final decisions regarding content, analysis, and conclusions were made by the author.

1 Team Contributions

- Qinxinghao Chen: Implemented the UNet decoder and ran FM experiments on MNIST.
- Handan Hu: Conducted AFHQ DDPM experiments, maintained the training pipeline, and contributed to bonus experiments.
- Chongwei Liu: Implemented latent diffusion, completed DDPM/FM core functions, and performed debugging and ablation studies.
- Bohan Wen: Wrote the report, prepared visualizations, and contributed to bonus experiments.

2 Implementation Summary

2.1 UNet

We filled in the decoder portion of the UNet by concatenating encoder skip connections with decoder features, applying a ResBlock (with time embedding), optional spatial transformer (with label conditioning), and then optional upsampling. Inputs are normalized to $[-1, 1]$ to stabilize training. The model takes both time and label embeddings: time uses sinusoidal encoding; labels are embedded and used inside transformer blocks.

2.2 DDPM

We implemented:

- `q_sample`: forward diffusion with analytic $\sqrt{\bar{\alpha}}$ coefficients.
- `p_sample`: DDPM reverse mean update.
- `generate`: reverse chain from noise.
- `compute_loss`: simplified noise-prediction loss.

2.3 Latent DDPM

AFHQ models use the provided tiny autoencoder. Images are encoded into latents, diffusion is done in latent space, and samples are decoded back to pixels.

2.4 Flow Matching (FM)

We implemented linearly interpolated paths between noise and data and trained the UNet to predict the constant velocity. Sampling uses Euler integration from $t = 0$ to 1 with no injected noise.

3 DDPM Experiments

3.1 MNIST DDPM

3.1.1 Training Curve

Figure 1 shows the loss curve. The loss decreases smoothly, with the fastest drop in the early steps.

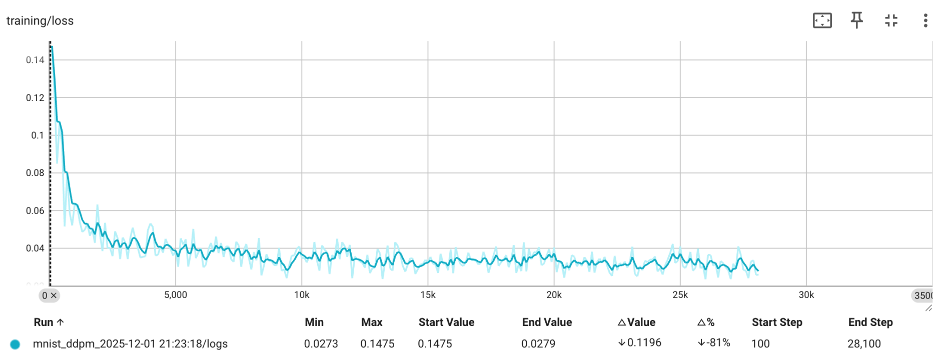

Figure 1: MNIST DDPM training loss.

3.1.2 Generated Samples (Early / Mid / Final)

Figure 2: MNIST DDPM samples at early (left), mid (middle), and final (right) epochs.

Short Discussion. Early samples show noisy digit-like blobs. Mid-stage samples have clearer shapes and mostly correct labels. Final samples look clean and consistent.

3.2 Latent DDPM on AFHQ

3.2.1 Training Curve

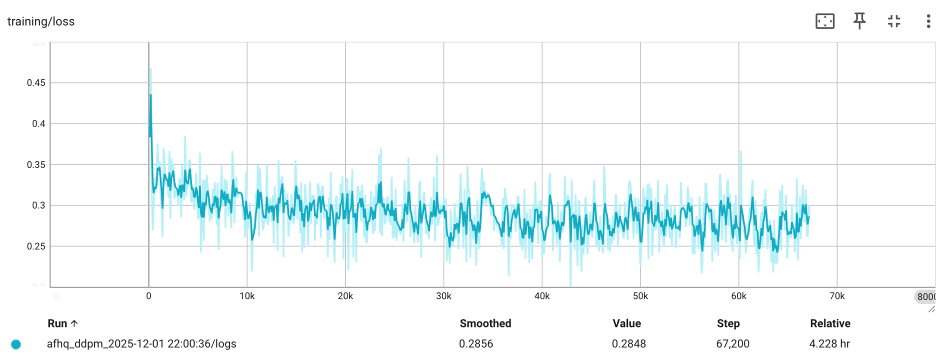

Figure 3: AFHQ DDPM loss curve.

3.2.2 Sample Progress (Early / Mid / Final)

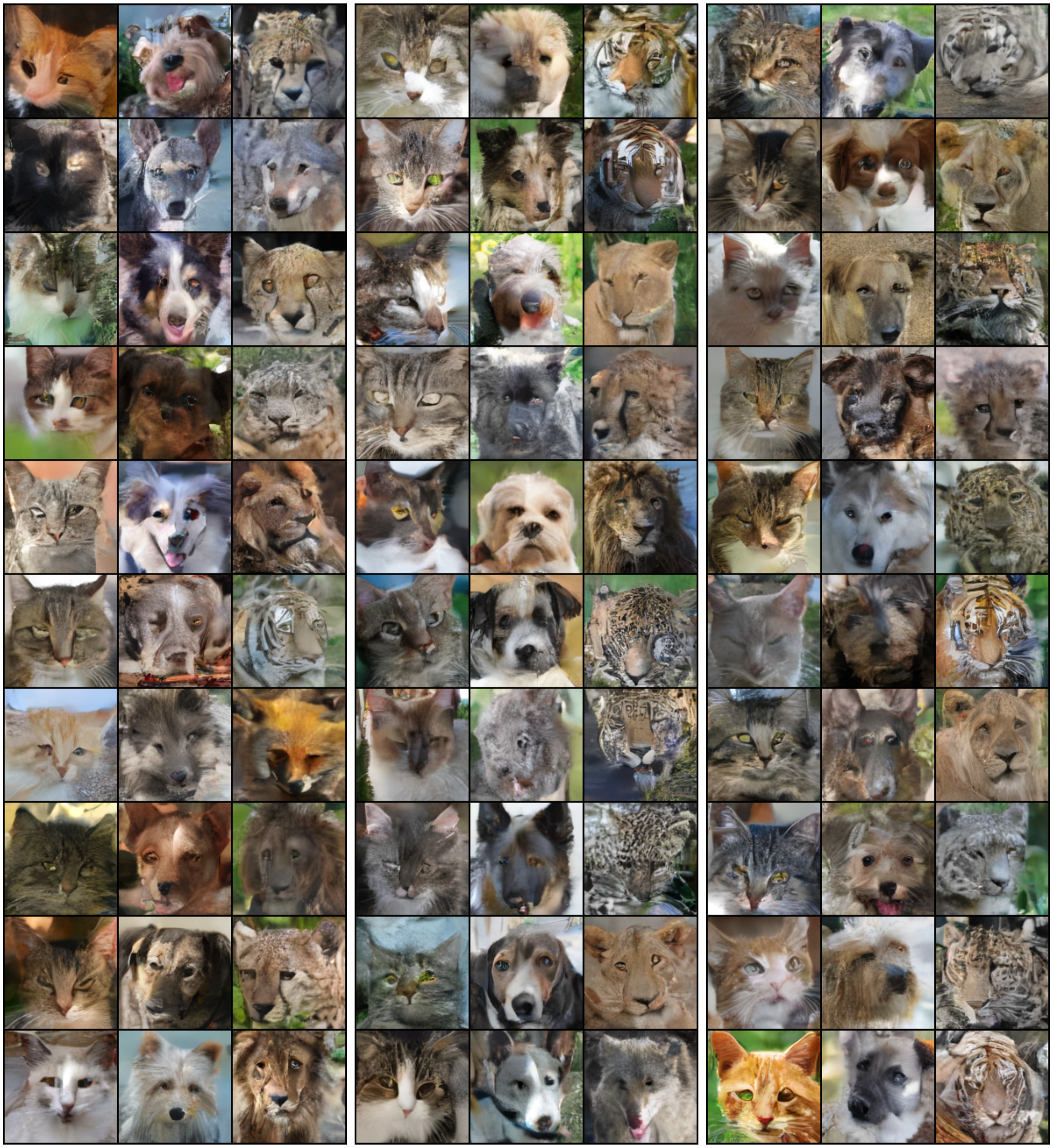

Figure 4: AFHQ DDPM samples (epoch 99 / 199 / 299).

Short Discussion. Early samples show coarse face structures but still blurry. Mid samples capture animal categories better (cats, dogs, big cats), although artifacts remain. Final samples are the clearest: identities and fur colors look much more consistent.

3.2.3 FID

The reported FID score for AFHQ DDPM is:

$$\textbf{FID} = \textbf{66.65}$$

4 Flow Matching (FM) Experiments

4.1 MNIST FM

4.1.1 Training Curve

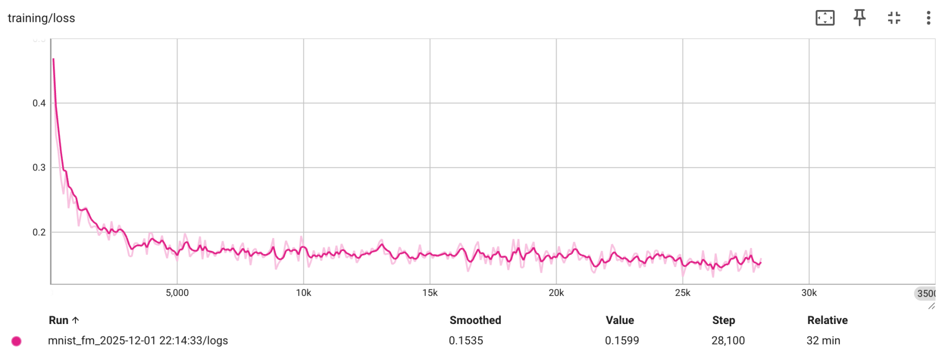

Figure 5: MNIST FM training loss.

4.1.2 Samples (Early / Mid / Final)

Figure 6: MNIST FM samples over training.

Short Discussion. Early outputs have digit outlines with missing parts. Mid-stage samples become more stable. Final samples are sharp and clearly show 0–9.

4.2 AFHQ FM

4.2.1 Training Curve

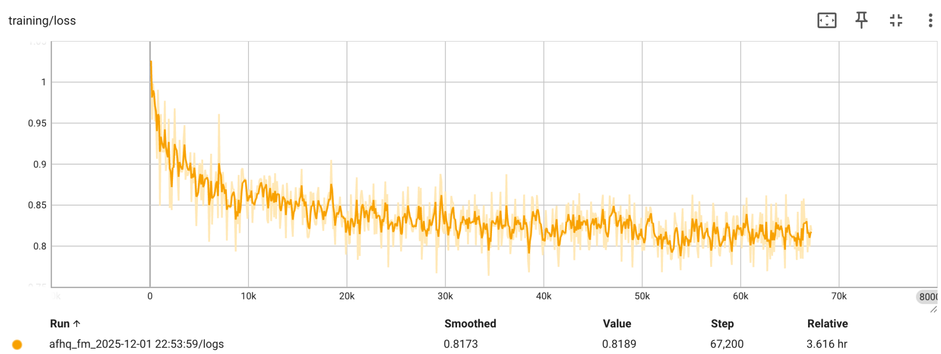

Figure 7: AFHQ FM training loss.

4.2.2 Sample Progress

Figure 8: AFHQ FM samples (epoch 99 / 199 / 299).

Short Discussion. FM improves steadily: early samples catch only rough shapes; mid samples show recognizable species; final samples are cleaner but tend to look slightly smoother than DDPM outputs.

4.2.3 FID

The reported FID score for AFHQ FM is:

$$\textbf{FID} = \textbf{66.80}$$

5 Conclusion

We implemented UNet, DDPM, latent DDPM, and Flow Matching models and trained them on MNIST and AFHQ. MNIST results were clean for both models. On AFHQ, latent DDPM slightly outperformed FM in FID (66.65 vs. 66.80), and DDPM samples tended to show a bit more texture. FM sampling was faster and simpler, and quality was competitive.