# Technical Report: MNIST Diffusion Migration to Purdue Gilbreth

PurdueHPC_Codex Workflow

February 24, 2026

## 1 Scope and Environment

This report captures migration of the MNIST DDPM training workflow from Gautschi to Gilbreth, including cluster policy adjustments and Slurm validation steps.

All work in this branch targets:

- Scratch workspace: `/scratch/gilbreth/rmaulik/codex_test`
- Source repository: `PurdueHPC_Codex` (branch `codex/gilbreth-diffusion`)
- Account: `rmaulik`
- Primary partition target: `a100-40gb`

## 2 Cluster Discovery and Policy Findings

Live Gilbreth inspection used:

- `slist`
- `sfeatures`
- `sinfo -s`
- `scontrol show partition`

Observed key finding for submission policy:

- Gilbreth requires explicit memory in batch submissions.

Initial submission error (before adding memory):

```
sbatch: error: Gilbreth requires you to explicitly request the amount
of memory you are requesting for your job.
sbatch: error: If you don't know your memory requirements,
a good default option in this partition is: --mem=240G per GPU
```

Branch scripts were updated accordingly with `#SBATCH -mem=240G`.

## 3 Diffusion Training Pipeline

The training stack and objective are unchanged from Gautschi:

- DDPM noise-prediction objective on MNIST
- Residual U-Net denoiser with sinusoidal time embeddings
- AdamW optimization + gradient clipping
- Run artifacts: samples, step/epoch loss plots, checkpoints, metrics JSON, and HTML dashboard

## 4 Gilbreth Slurm Configuration in This Branch

Primary short-run script:

- `submit_mnist_diffusion_gilbreth.slurm`
- `-p a100-40gb`
- `-gres=gpu:1`
- `-cpus-per-task=8`
- `-mem=240G`

Long-run script:

- `submit_mnist_diffusion_gilbreth_long.slurm`
- 80 epochs, 300 diffusion steps, same resource policy

## 5 Validation and Completed Run on Gilbreth (Feb 24, 2026)

| Job ID | Partition | State | Notes |
|---|---|---|---|
| 10330727 | a100-40gb | PENDING | Earlier queue attempt after adding `-mem` |
| 10330735 | a30 | PENDING | Alternate queue test; pending reason `AssocGrpGRES` |
| 10330760 | a100-40gb | COMPLETED | Final validated run on target partition |

Completed run details from `sacct` and `metrics.json`:

- Slurm elapsed: `00:03:41`
- Slurm state / exit code: `COMPLETED`, `0:0`
- Run tag: `gilbreth_10330760`
- Device: `cuda`
- Total training steps: `2345`
- Final epoch mean loss: `0.048376138451129896`
- Best epoch mean loss: `0.048376138451129896`

## 6 Dashboard and Artifacts

The completed run produced:

- `outputs/gilbreth_10330760/dashboard.html`
- `outputs/gilbreth_10330760/mnist_samples.png`
- `outputs/gilbreth_10330760/loss_curve_step.png`
- `outputs/gilbreth_10330760/loss_curve_epoch.png`
- `outputs/gilbreth_10330760/metrics.json`

The root live dashboard remained at:

- `outputs/dashboard.html`

and followed the latest run through `outputs/LATEST_RUN.txt = gilbreth_10330760`.

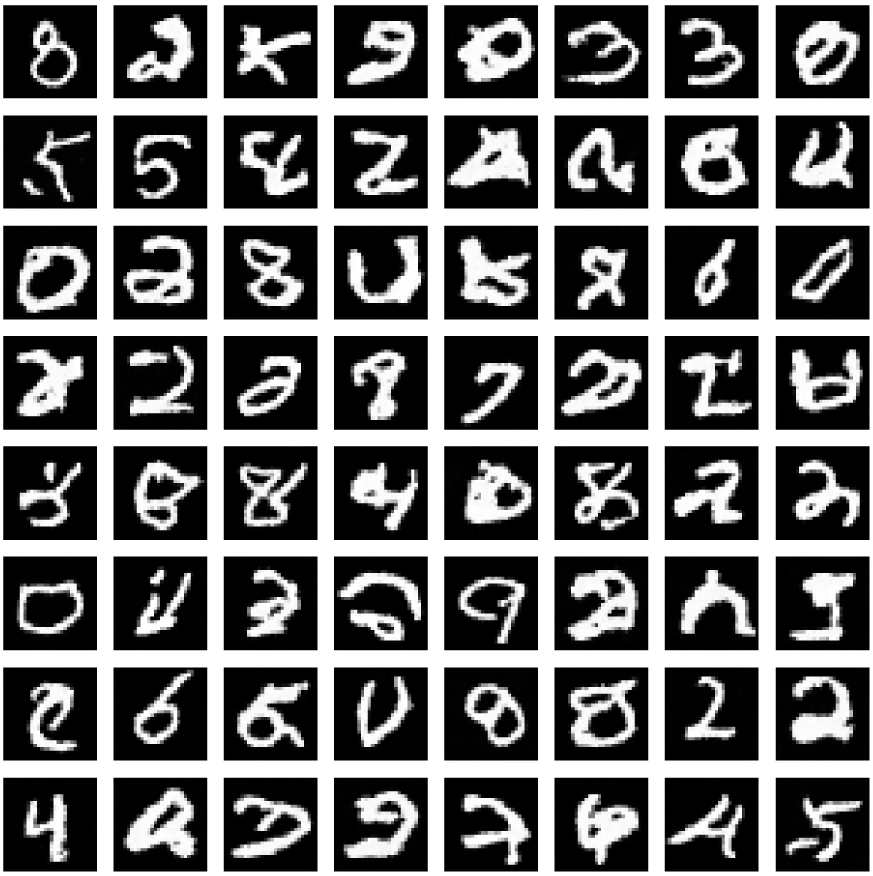

Figure 1: Generated MNIST samples from Gilbreth run `gilbreth_10330760`.

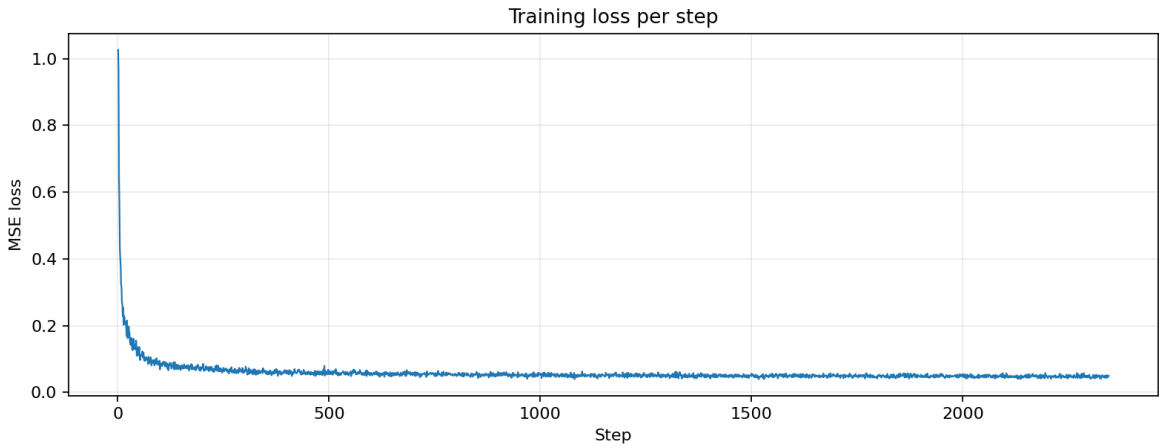


Figure 2: Per-step training loss on Gilbreth run `gilbreth_10330760`.

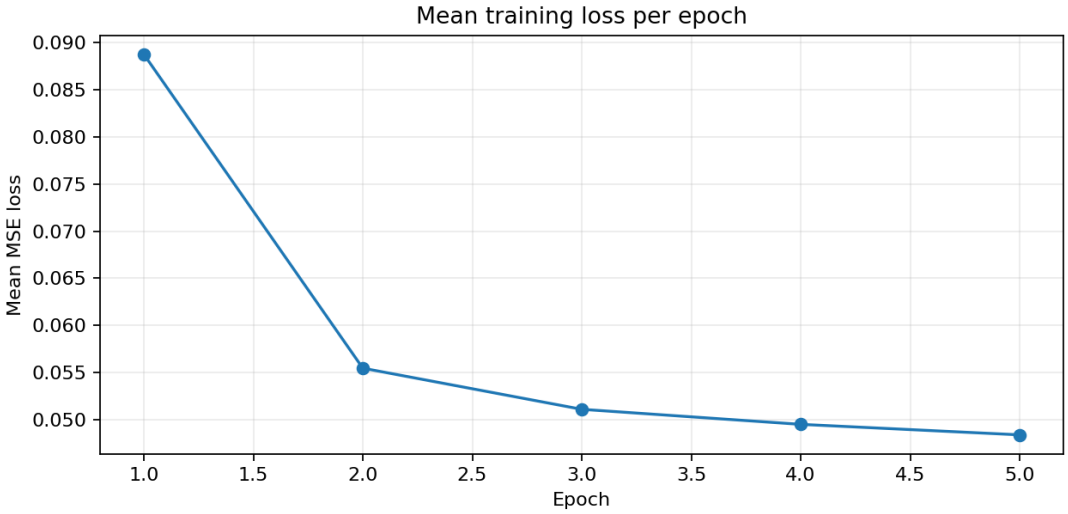


Figure 3: Per-epoch mean loss on Gilbreth run `gilbreth_10330760`.

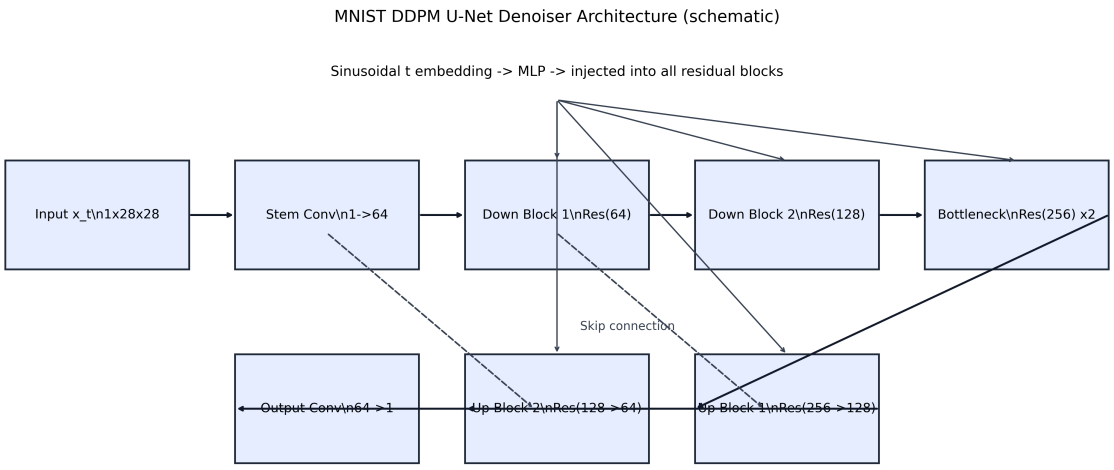


Figure 4: U-Net denoiser architecture schematic generated during the Gilbreth run.

# 7 Branch Artifacts Updated

- `README.md`
- `SKILL.md`
- `serve_dashboard.sh`
- `submit_mnist_diffusion_gilbreth.slurm`
- `submit_mnist_diffusion_gilbreth_long.slurm`
- `report/gilbreth_diffusion_report.tex`

# 8 Re-run Command

To repeat the validated short run:

```
cd /scratch/gilbreth/rmaulik/codex_test/PurdueHPC_Codex
sbatch submit_mnist_diffusion_gilbreth.slurm
```