Homework 4: Build your own model

CMU 10-799: Diffusion & Flow Matching

Spring 2026

Due: Friday, February 27, 2026 at 11:59 PM ET **Total Points:** 100

Late Due Date: Sunday, March 1, 2026 at 11:59 PM ET

Submission: Gradescope https://www.gradescope.com/courses/1207241

Starter Code: https://github.com/KellyYutongHe/cmu-10799-diffusion/

Introduction

Welcome to Homework 4! Can't believe it's already the final homework! In the last homework, you defined your problem, surveyed the field, and got a baseline working. Now it's time for the fun part: build your own model. In this homework, you'll implement your improvement ideas, run full experiments, and write up your findings. Think of this as completing your mini research paper: you're writing the Method, Experiments, and Conclusion sections. By the end, you'll have a complete research narrative: motivation → related work → baseline → your contribution → results → analysis.

What counts as a valid novelty?

We're not expecting NeurIPS-level novelty — this is a homework, not a research paper. But we do want to see you go beyond plugging in someone else's code and running it. The key question is: did you have to think? If your ideas required understanding the problem, making a design decision, and testing a hypothesis, then you've got what we're looking for. Here's a rough guide:

Good contributions (things that show thought and effort):

1. Combining ideas from two different papers in a new way
2. Adapting a method designed for a different setting to your specific problem
3. Proposing a new training trick, loss term, or architectural modification motivated by your HW3 observations
4. A thoughtful negative result — you tried something well-motivated and it didn't work, but you understand why
5. Anything that goes above and beyond these

Not sufficient on their own:

1. Only tuning hyperparameters (learning rate, batch size, number of steps)
2. Only training longer, on more data or a larger model
3. Implementing a second existing paper without any modification

Special note for HW 3 & 4:

1. **This homework is AI-friendly.** You may use any AI coding assistants, chatbots, or reference implementations. You may also use any other resources that you can find on the Internet. You may also use AI assistants for writing. However, you are responsible for every single word you submit. In other words, if something is wrong, unclear, or plagiarized, it's on you. AI is a tool, not an excuse. At the end of the homework, you'll document what resources you used.
2. **Build on your HW1&2 codebase.** Your implementation should extend from the codebase you've been developing. If you find useful code online (e.g., a reference implementation of your baseline method), you may incorporate it into your codebase directly. Just cite your sources clearly and explain what you used and modified.
3. **You must use the same dataset.** You may **not** switch to a different dataset (e.g., your own protein dataset) or use a different resolution for the last two homework. You may use additional training data or pretrained models if your method requires it, but all evaluation metrics (KID, FID, etc.) must be computed on the same CelebA 64×64 subset from HW1 and HW2. This ensures fair comparison across homeworks and across students.

Part I: Recap (10 points)

Before diving into your new contributions, briefly remind us of the setup from HW3. A reader needs to know what came before to understand what's new.

Q1. Where You Left Off **[10 pts]**

(a) [5 pts] **Problem statement:** Briefly restate your problem. What are your inputs, outputs, and assumptions? And how do you measure success? (This can be copy-pasted from HW3 if unchanged, or updated if your scope has shifted.)

(b) [5 pts] **Baseline summary:** What baseline method did you implement in HW3? Summarize the key idea in 1-2 paragraphs and report your baseline's quantitative results (primary metric).

Part II: Your Method (30 points)

This is the heart of HW4 and corresponds to the **Method** section of a research paper. Here you describe what you're doing differently from the baseline and why. A good method section should be clear enough that a classmate could reproduce your approach.

Q2. What's New? **[10 pts]**

(a) [3.5 pts] Describe the high-level idea and the key insights of your method. What is your improvement or novel contribution? What are you changing compared to your baseline?

(b) [3.5 pts] Why should this help? What's your hypothesis? Why do you expect this to improve over the baseline? Ground your reasoning in either theory, intuition from related work, or observations from your HW3 results.

(c) [3 pts] What makes this non-trivial? Explain why this goes beyond simple hyperparameter tuning. What required thought, creativity, or effort?

Q3. Method Details [20 pts]

(a) [15 pts] Describe your method in enough detail that a classmate could implement it. Include any new equations, algorithms (pseudocode or step-by-step), architectural changes, training modifications, or inference-time techniques. Diagrams are encouraged if they help.

(b) [5 pts] What are the important practical design choices specific to your method? (For example, (1) in order to efficiently compute the MeanFlow identity, the MeanFlow paper [Geng et al., 2025] decomposes their objective in the forms of Jacobian-vector products (2) in order to efficiently calculate the log likelihood, the flow matching paper [Lipman et al., 2023] uses the Hutchinson estimator to approximate the divergence.)

Part III: Experiments (40 points)

Now show us the results! This corresponds to the Experiments section of a research paper. Good experiments aren't just about showing your method works: they're about understanding why it works (or doesn't).

Q4. Experimental Setup [5 pts]

(a) [3 pts] Describe your experimental setup: model architecture and size, training details (batch size, learning rate, iterations, optimizer) or inference details (number of steps, guidance scale, solver), and any task-specific setup (e.g., degradation type for inverse problems, which attributes for conditional generation, distillation schedule for speed methods).

(b) [2 pts] How much compute did you use total across HW3 and HW4? (GPU type, total training/inference time)

Q5. Results [15 pts]

(a) [8 pts] **Quantitative comparison:** Report your primary metric for both your baseline (from HW3) and your improved method. Present this as a comparison table or plot. Did you beat your baseline? By how much?

(b) [7 pts] **Qualitative comparison:** Show a side-by-side comparison of outputs from your baseline and your method. Include a grid of samples (or equivalent for your track — e.g., before/after for editing, samples at different step counts for speed). What visual differences do you notice?

Note: If your task demonstrations require data that is difficult to show in a PDF, feel free to include a link to your demo here as well.

Q6. Ablation Study [10 pts]

(a) [5 pts] **Component ablation:** If your algorithm consists of multiple parts, show what happens when you remove each part individually. Present this as a table with rows for each ablation and columns for your metrics. What matters most?

(b) [5 pts] **Hyperparameter ablation:** Pick 1-2 key hyperparameters in your method and vary them. How sensitive is your method to these choices? Show the results as a table or plot.

Q7. Analysis [10 pts]

(a) [5 pts] What worked and what didn't? Be honest. If something you tried failed, tell us about it. Negative results are as valuable as positive results in this class. What surprised you?

(b) [5 pts] Show 2-3 examples where your method fails or produces poor results. Why do you think these failures occur?

Part IV: Discussion & Conclusion (20 points)

You're almost done! This corresponds to the **Discussion** and **Conclusion** sections of a research paper. Step back from the details and reflect on the bigger picture.

Q8. Discussion [10 pts]

(a) [5 pts] What are the main limitations of your approach? Be specific! Think about scalability, assumptions that might not hold, compute constraints, dataset-specific biases, etc.

(b) [5 pts] If you had another month (and unlimited compute), what would you try next? Propose 1-2 concrete next steps.

Q9. Conclusion [5 pts]

(a) [5 pts] Write a 1-paragraph conclusion summarizing your entire project arc: the problem you tackled, your approach, your key results, and the main takeaway. This should read like the conclusion of a research paper, i.e. concise and self-contained.

Q10. Reflection [5 pts]

(a) [2 pts] What was the most valuable thing you learned from this project (HW3 + HW4 combined)?

(a) [2 pts] If you were starting over, what would you do differently?

(a) [1 pts] List all the resources you used for this homework: AI tools, open source code, tutorials, papers, classmates, etc.

That's it! You have completed HW 4, and with it, your mini research project! Congrats on going from zero to a working, novel diffusion model in 7 weeks. Now go make some noise (or should we say, denoise?) in the real world!

References

Zhengyang Geng, Mingyang Deng, Xingjian Bai, J Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. *arXiv preprint arXiv:2505.13447*, 2025.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.