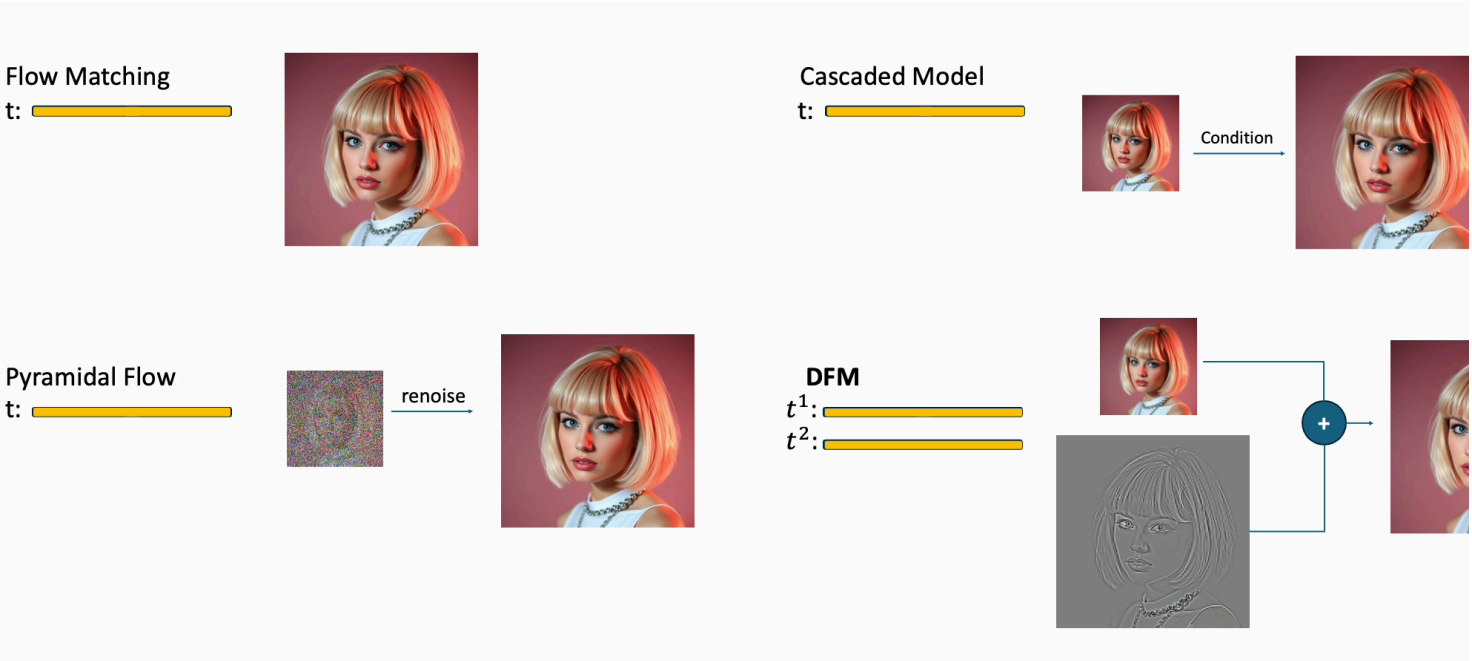
Improving Progressive Generation with Decomposable Flow Matching

Moayed Haji-Ali* Willi Menapace* Ivan Skorokhodov Arpit Sahni Sergey Tulyakov Vicente Ordonez Aliaksandr Siarohin

Snap Research RICE UNIVERSITY

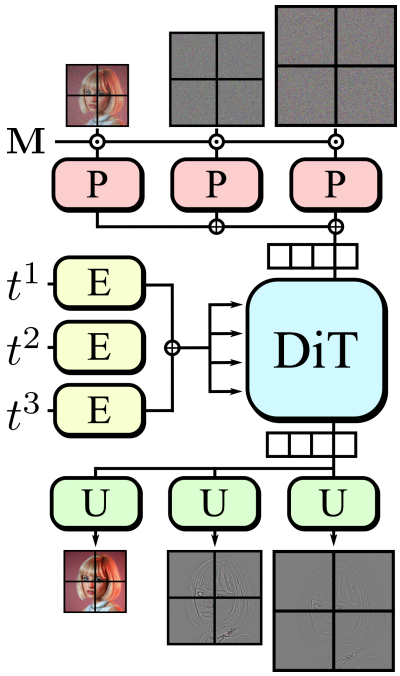
arXiv

TL;DR: Decomposable Flow Matching (DFM) is a simple framework to progressively generate visual modalities scale-by-scale, achieving up to 50% faster convergence compared to Flow Matching. Read the **paper** on arXiv for more details.



Method

Decomposable Flow Matching (DFM): A generative model combining multiscale decomposition with Flow Matching. DFM progres synthesizes different representation scales by generating coarse-structure scale first and incrementally refining it with finer scales.

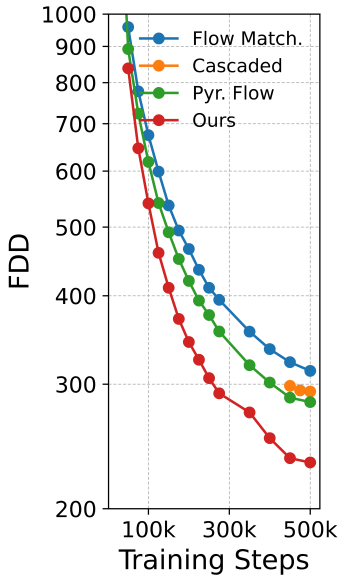


DFM Architecture: Our framework (DFM) progressively synthesizes images by combining multiscale decomposition with Flow Mat
modify the DiT architecture to use per-scale patchification and timestep-embedding layers while keeping the core DiT architecture u

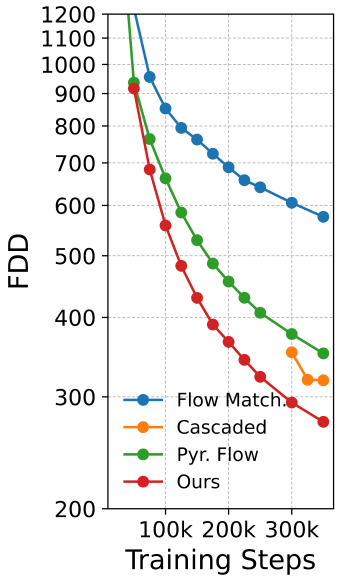
DFM Architecture | Training | Inference

Results

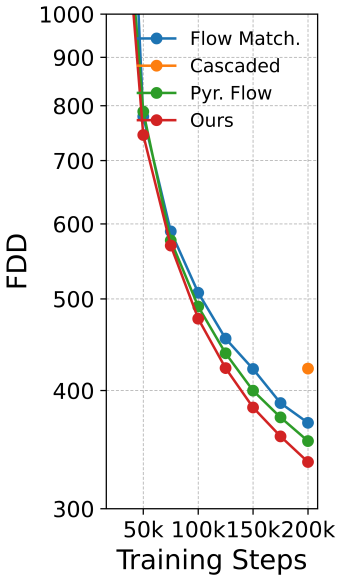
Across image and video generation, DFM outperforms the best-performing baselines, achieving the same Fréchet DINO Distance
Flow Matching baselines with up to 2x less training compute.



ImageNet-1k 512px



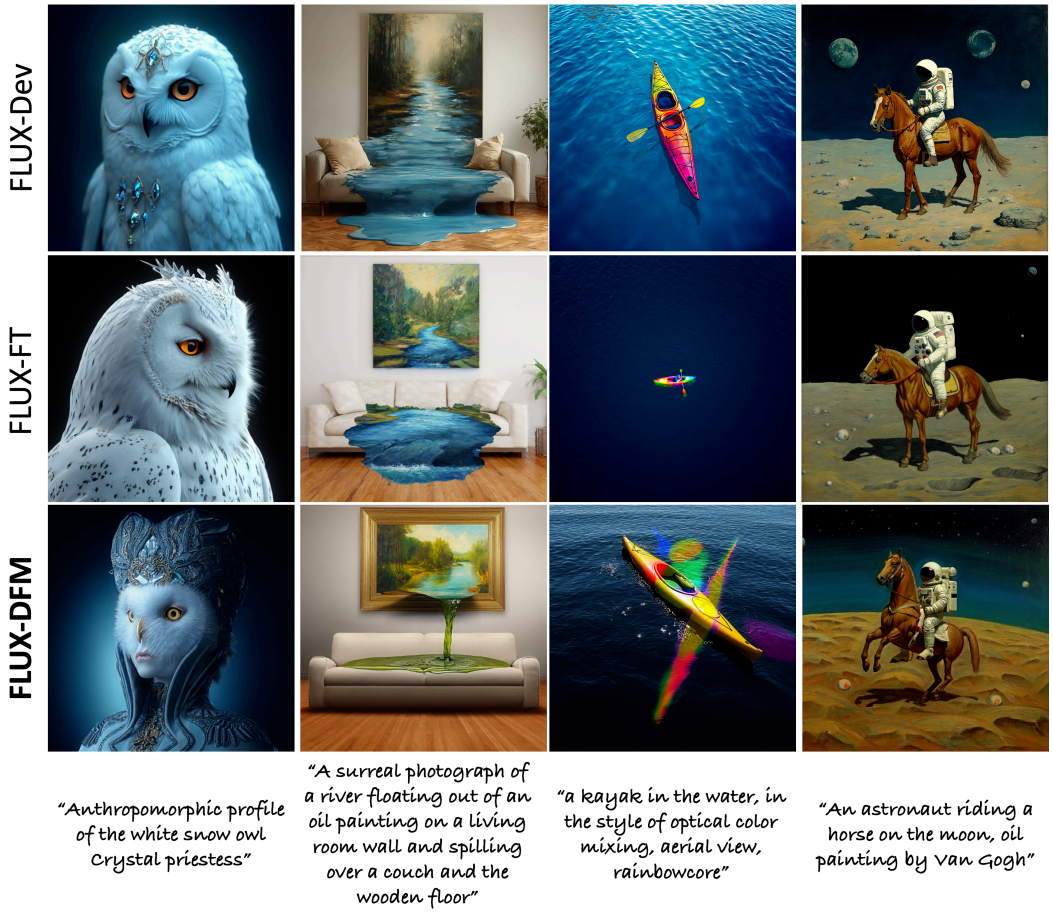
ImageNet-1k 1024px



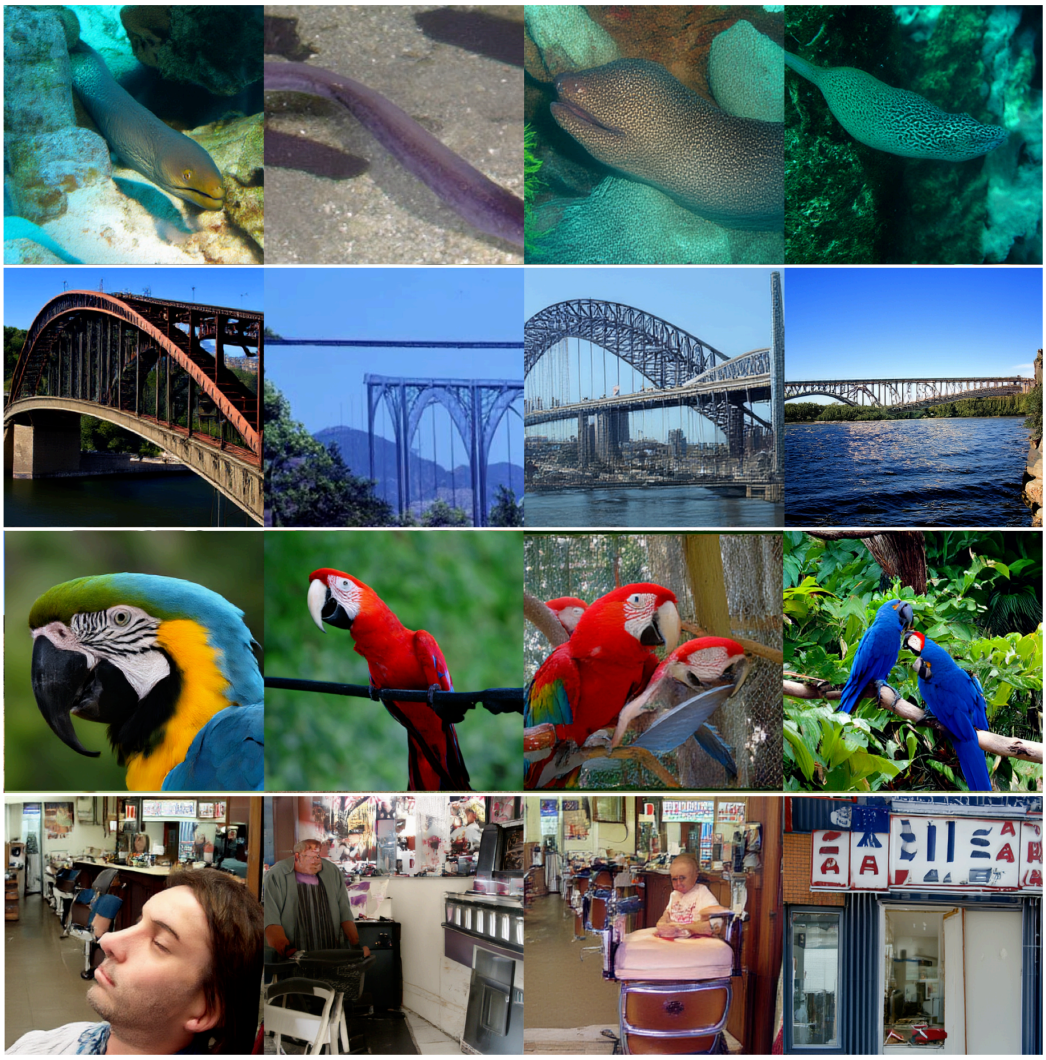
Kinetics-700 512px

Qualitative Results

Large-Scale Finetuning: Finetuning FLUX-dev with DFM (FLUX-DFM) achieves superior results than finetuning with standard full-finetuning (DFM-FT) for the same training compute.

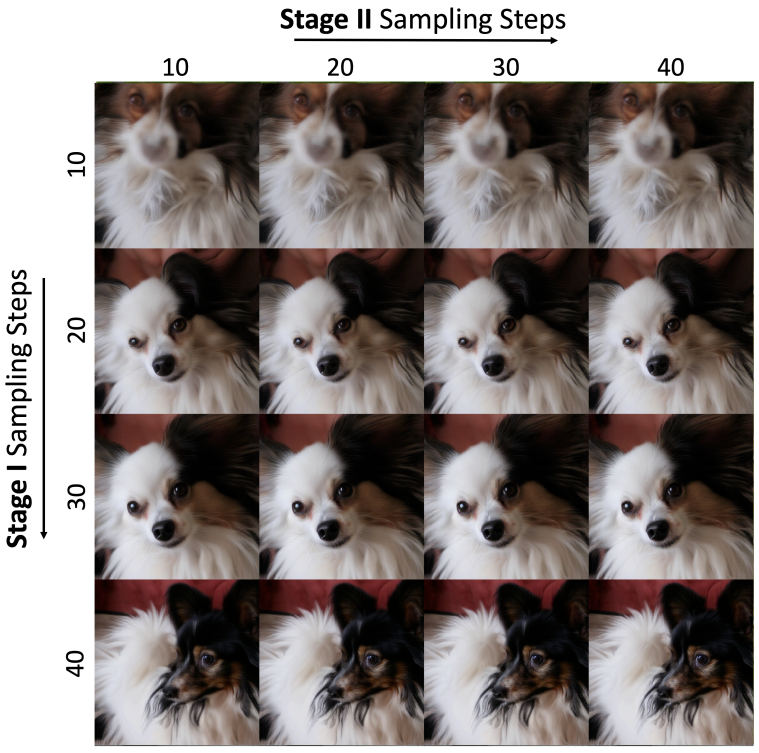


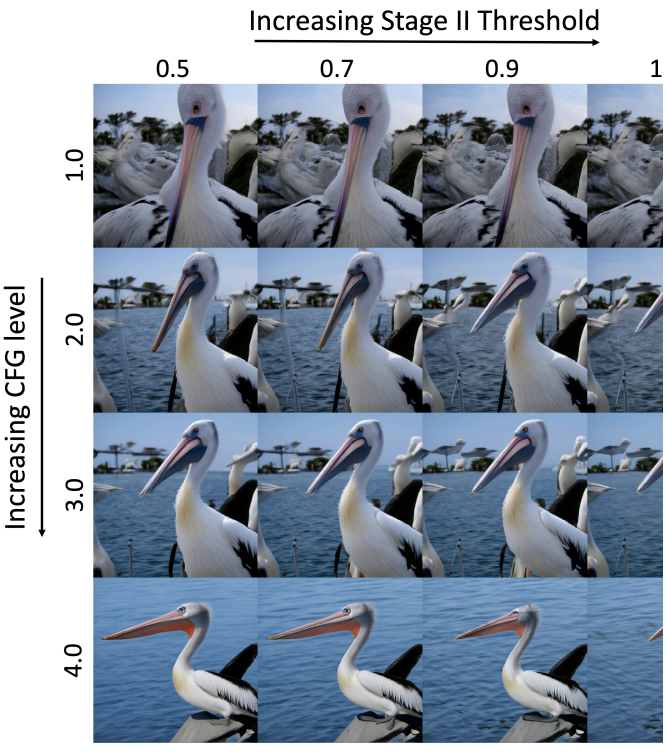
Training From Scratch for Image Generation: When trained from scratch on ImageNet-1k 512px, DFM achieves better quality tha baselines using the same training resources.



Training From Scratch for Video Generation: DFM is also suited for video generation, achieving better structural and visual qualit baselines when trained on the Kinetics-700 dataset with the same compute budget.

Ablations: We found that DFM benefits from more sampling steps in the coarse-structure stage and needs only a few in the high-fr stage, and it stays largely insensitive to the choice of sampling per-stage noise threshold, especially at high CFG values.





Citation

If you find this paper useful in your research, please consider citing our work:

```
@article{dfm,
title={Improving Progressive Generation with Decomposable Flow Matching},
author={Moayed Haji-Ali and Willi Menapace and Ivan Skorokhodov and Arpit Sahni and Sergey Tulyakov and
  Vicente Ordonez and Aliaksandr Siarohin},
journal={arXiv preprint arXiv:2506.19839}
year={2025}}
```