Accelerated Likelihood Maximization for Diffusion-based Versatile Content Generation

Hyunsoo Lee **Inwoo Hwang** **Young Min Kim**

Seoul National University

Project Page

ECCV 2026 | MALMÖ | SEPT 8–12

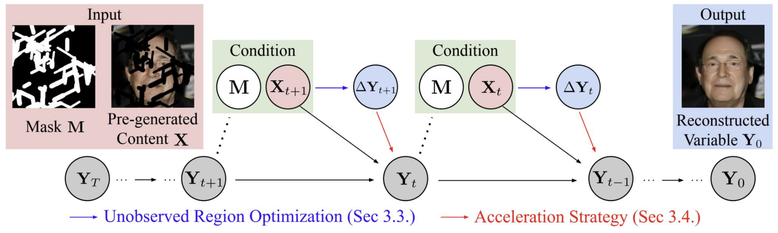
1 Introduction

- **Versatile Content Generation**
  - Versatility across various generative tasks conditioned on **partially observed or pre-generated inputs**
  - Support for inpainting, outpainting, and 3D mesh texture generation
- **Contributions**
  - A **training-free** and model-agnostic sampling scheme
  - An explicit score-based likelihood maximization strategy
  - An accelerated one-step optimization strategy
  - **Broad applicability** across diverse modalities

2 Background

- **Training-based Methods**
  - Strong task-specific performance at high training cost
  - Domain-limited models requiring retraining for other tasks
- **Diffusion Synchronization**
  - A representative training-free method for versatile content generation
  - Heuristic or indirect guidance
    - Reliance on extensive empirical search
    - Guidance restricted to the pre-generated region
    - **Lack of control over unobserved-region generation**

3 Framework



- **Optimization Objective**
  - Weighted sum of negative **log-likelihood** and negative **joint log-density**

$$-\lambda_1 \log p(\mathbf{X}_t, \mathbf{M} \mid \mathbf{Y}_t^i + \mathbf{M} \odot \Delta\mathbf{Y}_t^i, \mathbf{c}) - \lambda_2 \log p(\mathbf{X}_t, \mathbf{M}, \mathbf{Y}_t^i + \mathbf{M} \odot \Delta\mathbf{Y}_t^i \mid \mathbf{c})$$

- **Closed-Form Calculation**
  - 1st-order Taylor expansion with gradient descent

$$\Delta\mathbf{Y}_t^i = \mathbf{M} \odot (\lambda_1 \nabla_{\mathbf{Y}_t^i} \log p(\mathbf{X}_t, \mathbf{M} \mid \mathbf{Y}_t^i, \mathbf{c}) + \lambda_2 \nabla_{\mathbf{Y}_t^i} \log p(\mathbf{X}_t, \mathbf{M}, \mathbf{Y}_t^i \mid \mathbf{c}))$$

  - Score estimation using the pretrained diffusion model

$$\nabla_{\mathbf{Y}_t^i} \log p(\mathbf{Y}_t^i \mid \mathbf{c}) \simeq -\epsilon_\theta(\mathbf{Y}_t^i, t, \mathbf{c}) / \sqrt{1 - \alpha_t}$$

- **Update Equation**
  - **Acceleration strategy** for reducing *N*-step iterative minimum updates

$$\mathbf{Y}_t \leftarrow \mathbf{Y}_t + \mathbf{M} \odot (N\lambda_1(\epsilon_\theta(\mathbf{Y}_t, t, \mathbf{c}) - \epsilon_\theta(\mathbf{E}_t, t, \mathbf{c})) - N\lambda_2\epsilon_\theta(\mathbf{E}_t, t, \mathbf{c}))$$

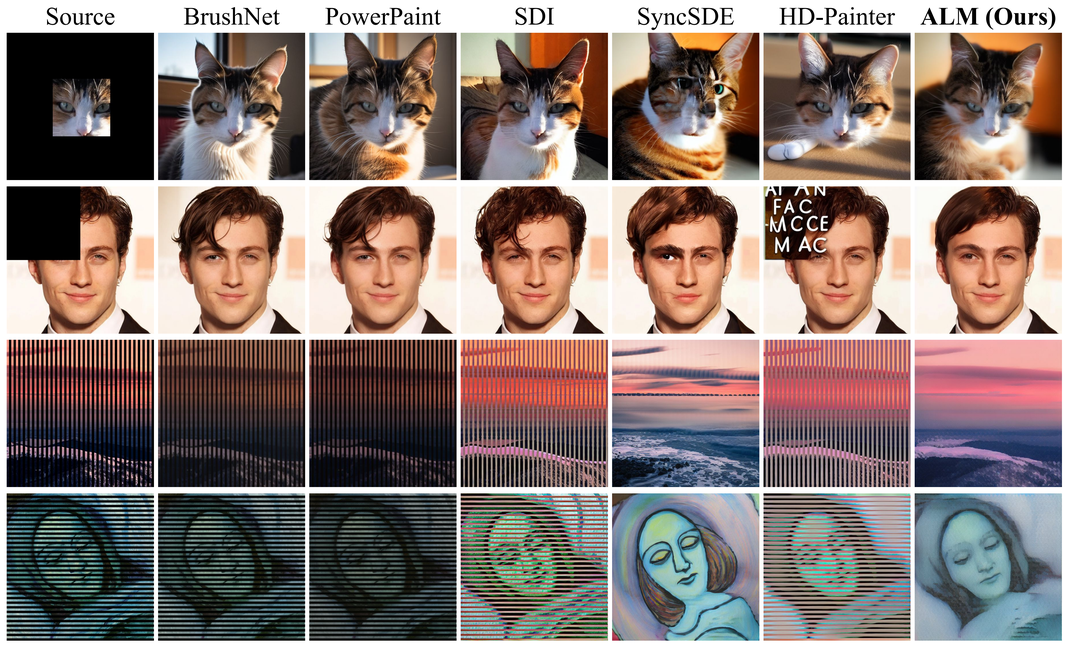
$$\text{with} \quad \mathbf{E}_t = \mathbf{X}_t \odot (\mathbf{1} - \mathbf{M}) + \mathbf{Y}_t \odot \mathbf{M}$$

  - Modified DDIM reverse process with optimized latent variable

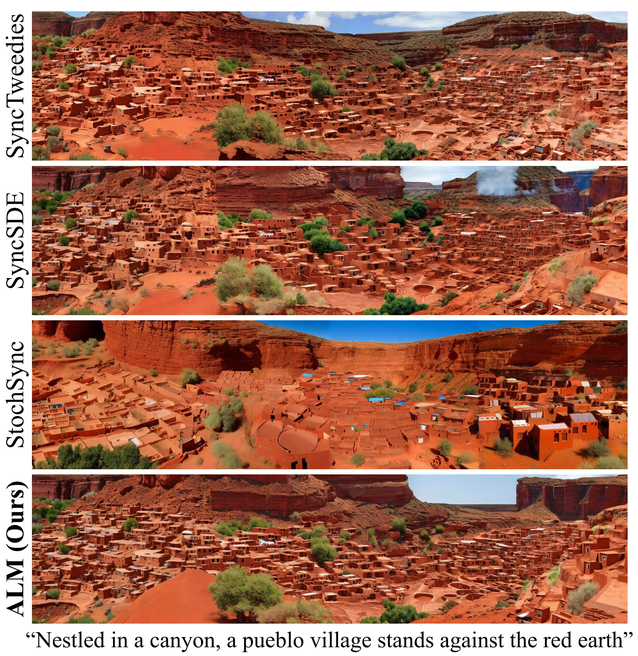
$$\nabla_{\mathbf{Y}_t} \log p(\mathbf{Y}_t \mid \mathbf{X}_t, \mathbf{c}) = \nabla_{\mathbf{Y}_t} \log p(\mathbf{Y}_t \mid \mathbf{c}) + \nabla_{\mathbf{Y}_t} \log p(\mathbf{X}_t \mid \mathbf{Y}_t, \mathbf{c})$$

4 Experiments

- **Image Inpainting**

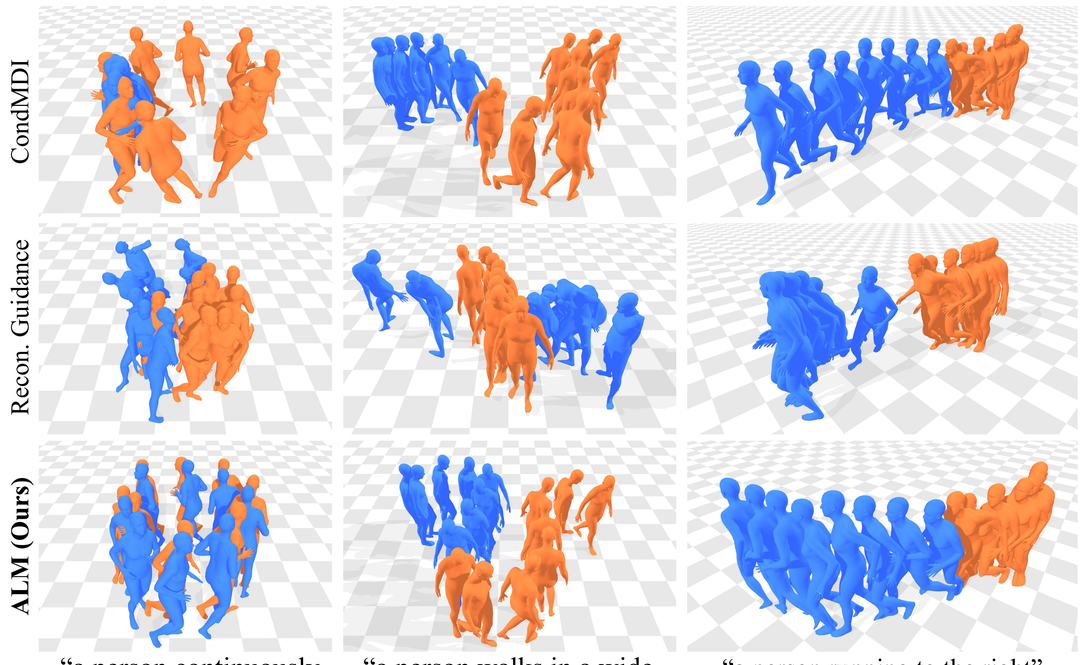


- **Wide Image Generation**



"Nestled in a canyon, a pueblo village stands against the red earth"

- **Human Motion Completion**
  - Given frames are highlighted in **orange**, while completed frames are shown in **blue**



"a person continuously jogs counter clockwise." "a person walks in a wide winding path, creating large semicircular loops." "a person running to the right"

- **3D Mesh Texturing**