Collaborative Diffusion

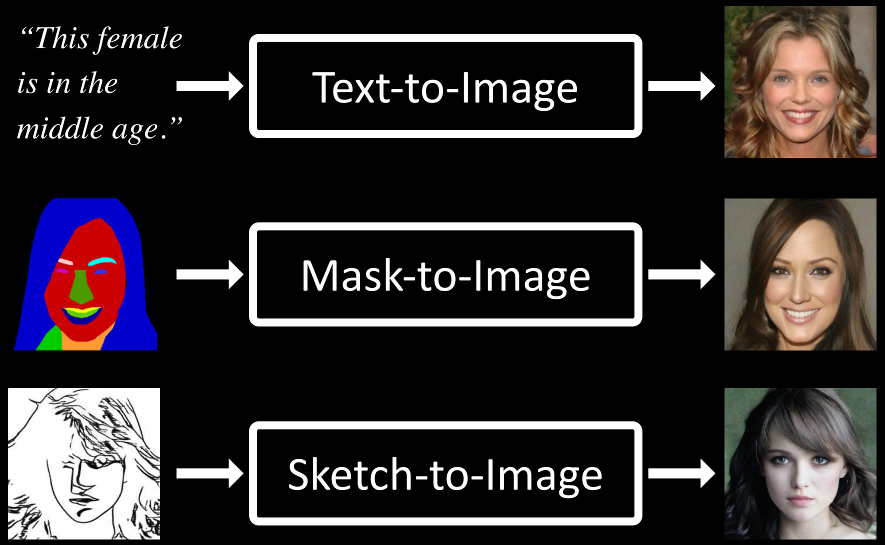
for Multi-Modal Face Generation and Editing

Ziqi Huang, Kelvin C.K. Chan, Yuming Jiang, Ziwei Liu

S-Lab, Nanyang Technological University

Motivation

Existing diffusion models are mainly uni-modal, that is., driven by only one modality of condition.



……

However, in real applications, users want multi-modal control. See examples next slide.

Task Highlight

(A) Multi-Modal Face Generation

given multi-modal controls

"This female is in the middle age."

……

synthesize high-quality image consistent with the controls

Task Highlight

(B) Multi-Modal Face Editing

given input image

and target multi-modal conditions

"This man has beard of medium length. He is in his thirties."

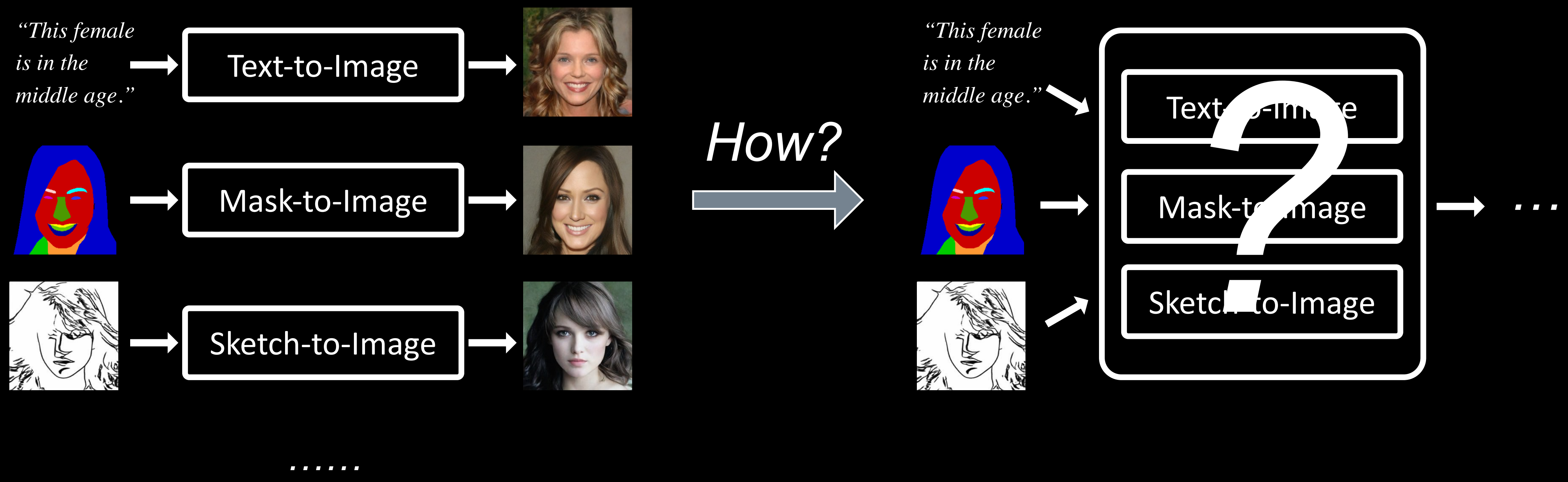
……

edit the image
to 1) satisfy the target conditions
while 2) preserving the facial identity

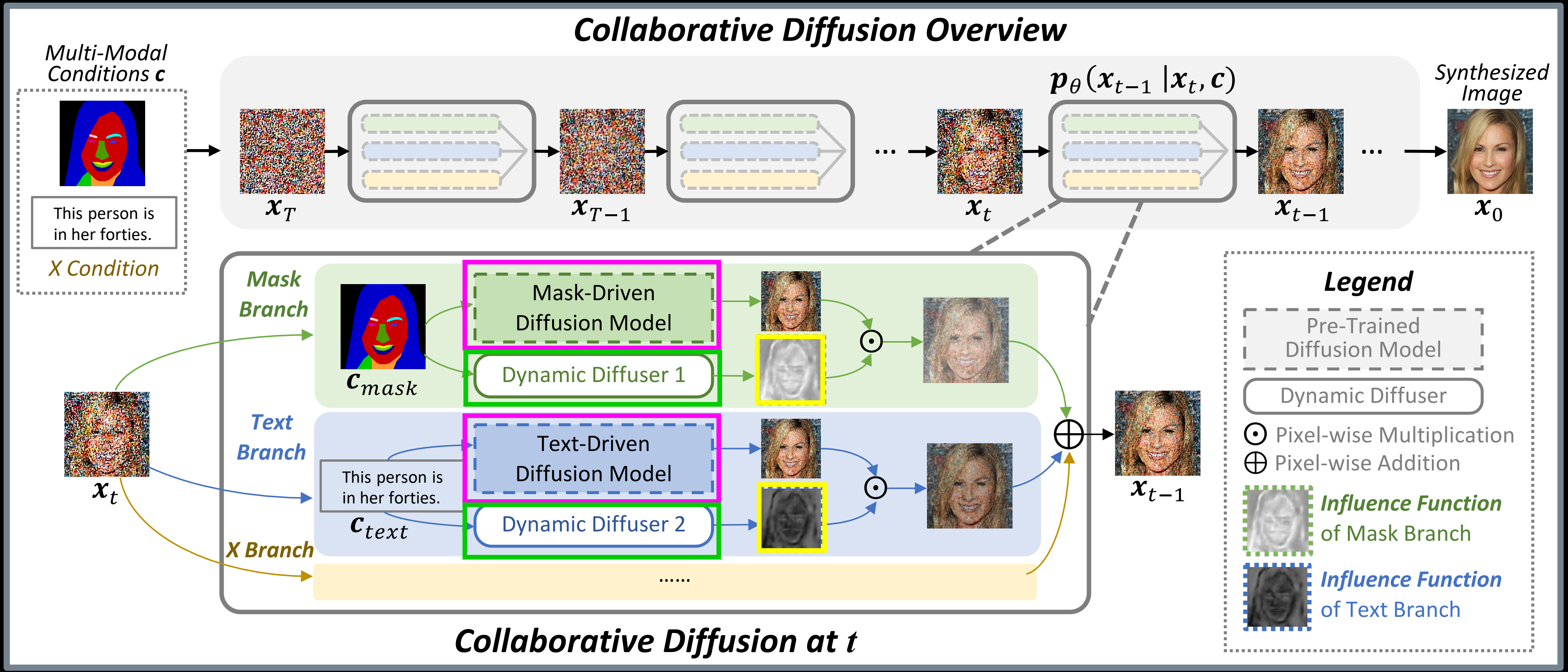
Motivation

Uni-Modal Diffusion Models

Multi-Modal Control



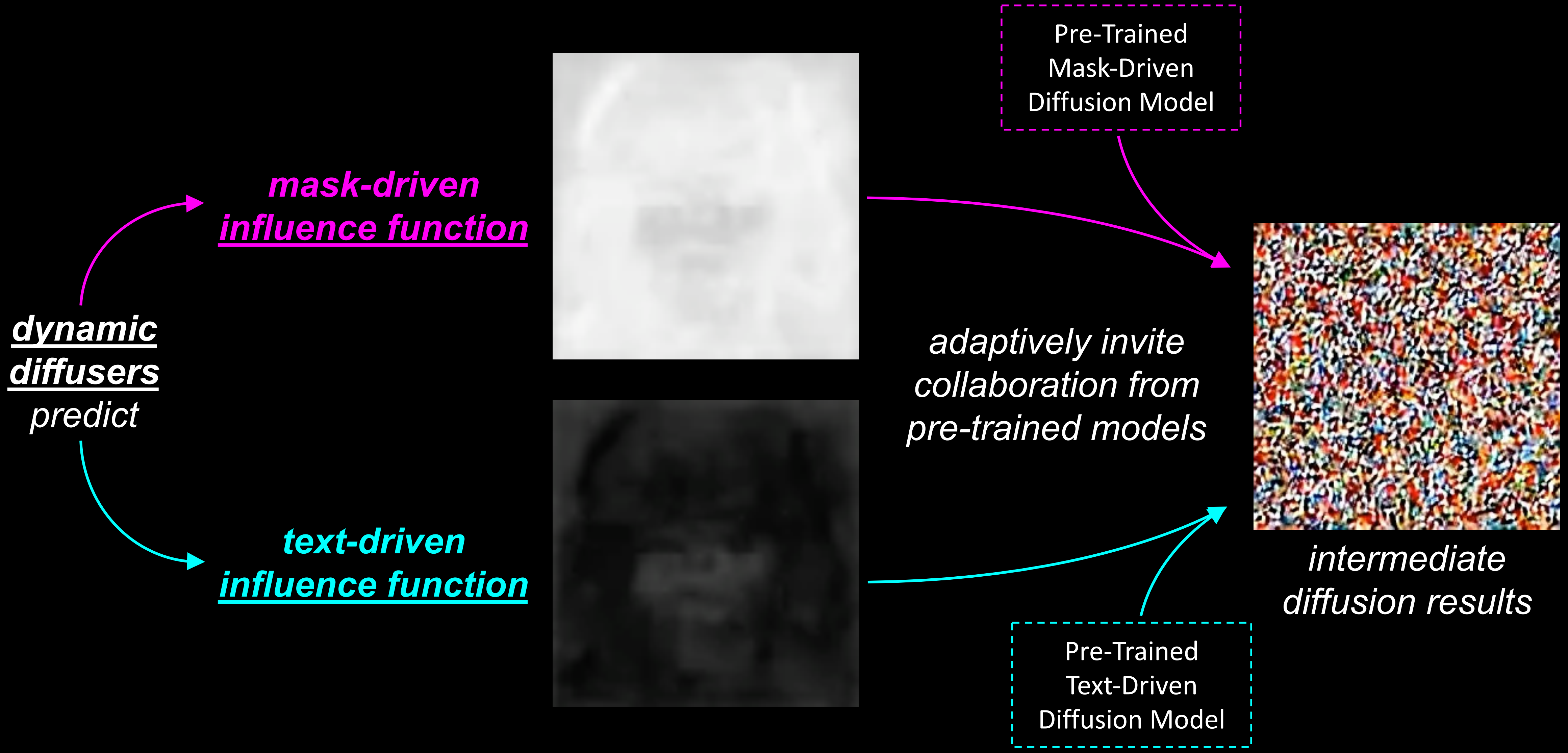
Collaborative Diffusion Framework



During the reverse process of diffusion models:

- *Pre-trained uni-modal diffusion models collaborate to achieve multi-modal control without being re-trained*
- ***Dynamic diffusers** predict spatial-temporal **influence functions** to enhance or suppress contributions from each pre-trained model*

Dynamic Diffusers *predict* ***Influence Functions***



Visual Results: Generation

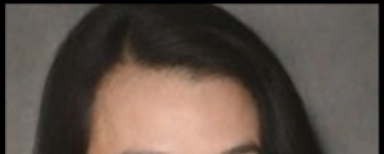
Visual Results: Editing

Input Image | *Target Mask* | *Target Text* | *Edited Image*


This woman looks like an elderly.


He is a young adult. He doesn't have any beard at all.





Summary

- We exploit pre-trained uni-modal diffusion models. They can collaborate to achieve multi-modal control without being re-trained.

- Collaborative Diffusion can be used to *extend arbitrary uni-modal approach* (e.g. face generation, face editing, motion generation, 3D generation) *to the multi-modal paradigm*.

Project Page

Code