## UniTransfer: Video Concept Transfer via Progressive Spatial and Timestep Decomposition

https://yu-shaonian.github.io/UniTransfer-Web/

Mid-Grain

A close-up shot of a woman

standing against a light-colored.

plain background. She wears a

pink t-shirt and has short bob

hairstyle.

Fine-Grain

A close-up shot of a woman

standing against a light-colored,

plain background. She wears a

pink t-shirt and has short bob

hairstyle. Her hands are raised near

her chest. She is looking at the

camera with a neutral expression.

Coarse-Grain

A close-up shot of a woman

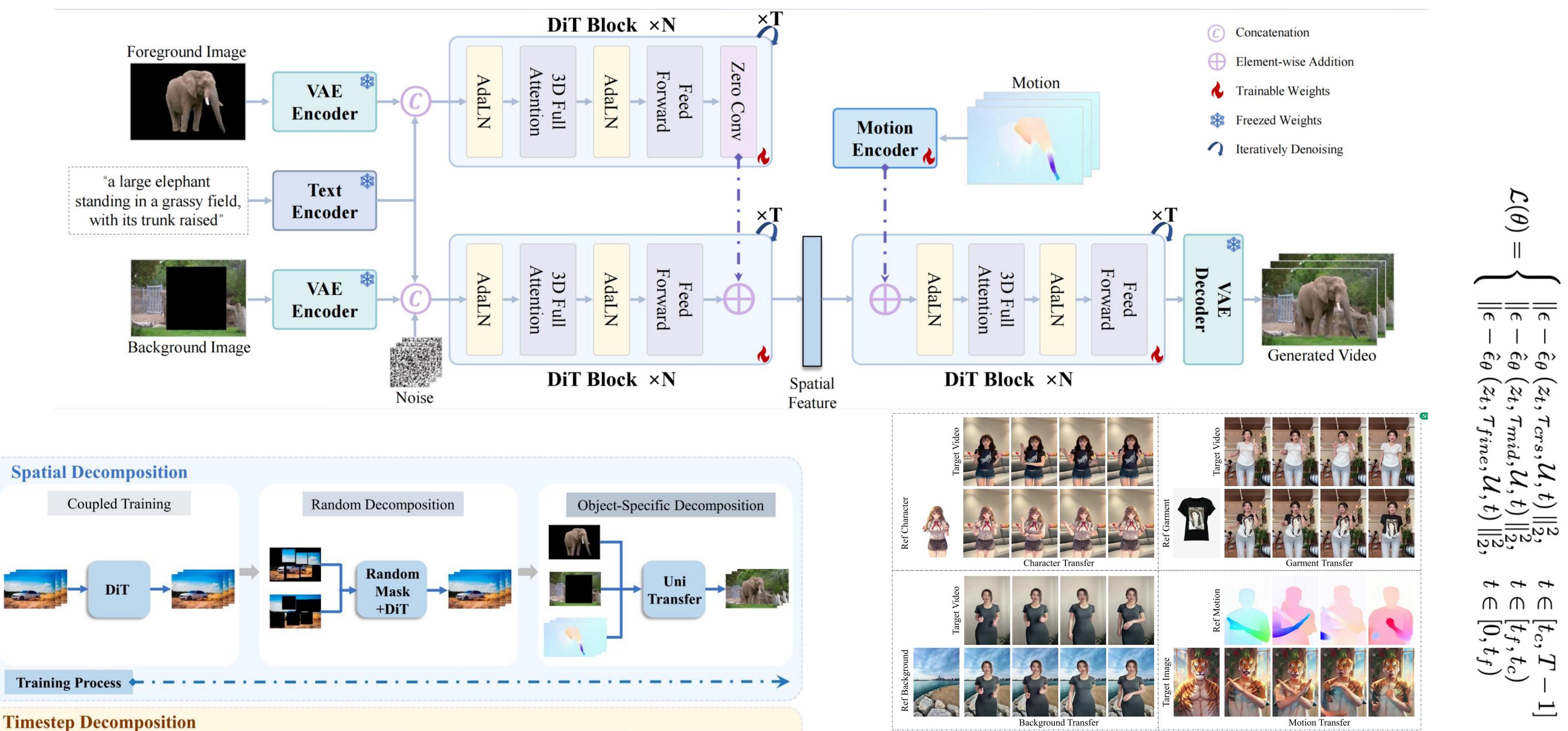
standing against a plain

background

**Denoising Process** 

Guojun Lei, Rong Zhang, Tianhang Liu, Chi Wang, Hong Li, Zhiyuan Ma, Weiwei Xu Zhejiang University Zhejiang Gongshang University





- We propose a DiT-based image-guided video concept transfer framework UniTransfer, which incorporates progressive spatial and timestep decomposition.
- We introduce a self-supervised pretraining strategy based on randomized masking to enhance the disentangled representation learning
- We further introduce an LLMs-guided chain-of-prompt mechanism to achieve the timestep decomposition. This progressive prompting strategy guides the generation process with stage-specific instructions, improving the VCT generation quality.