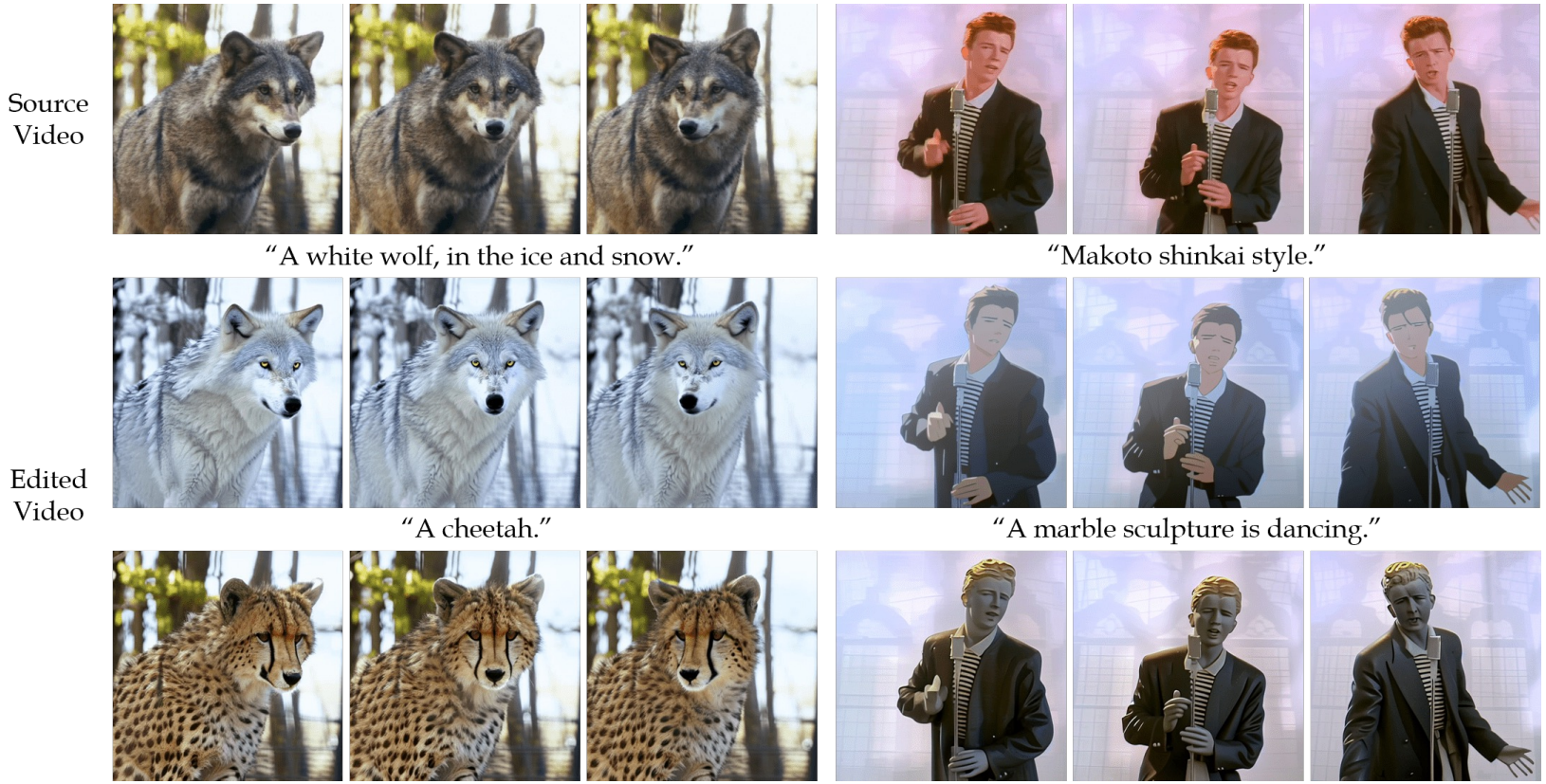
WAVE: Warping DDIM Inversion Features for Zero-shot Text-to-Video Editing

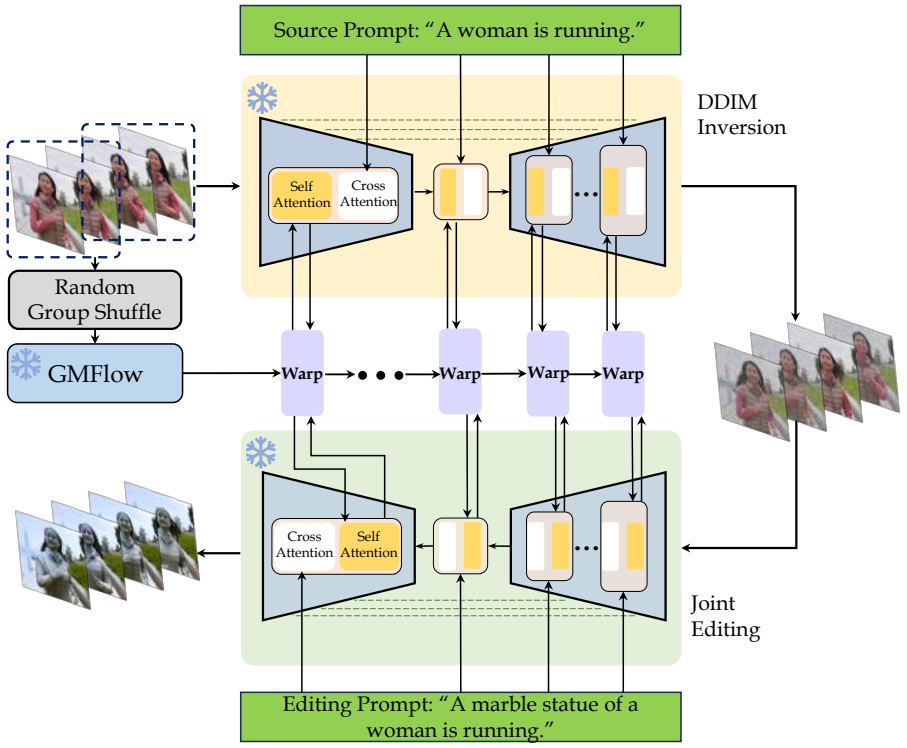
Yutang Feng*, Sicheng Gao*, Yuxiang Bao*, Xiaodi Wang, Shumin Han†,
Juan Zhang†, Baochang Zhang, and Angela Yao

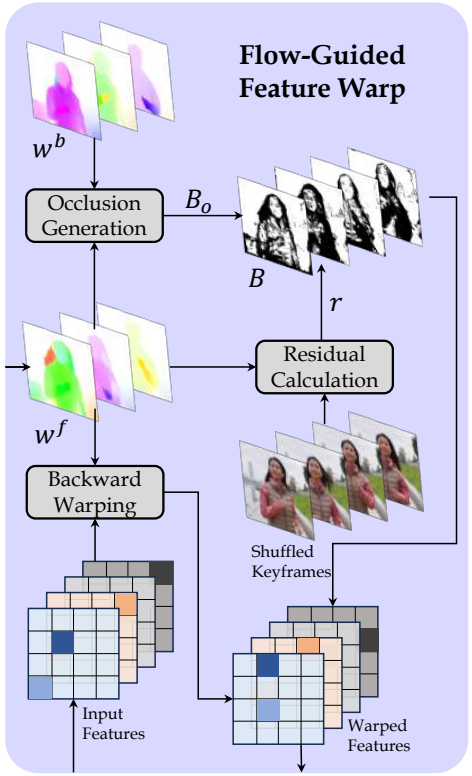
WAVE

Given a source video and a target textual prompt, WAVE employs a flow-guided feature warping strategy during both DDIM inversion and joint editing processes for powerful video editing.



Our framework





$$z_{t-1} = \sqrt{\alpha_{t-1}} \frac{z_t - \sqrt{1-\alpha_t}\varepsilon_\theta}{\sqrt{\alpha_t}} + \sqrt{1-\alpha_{t-1}}\varepsilon_\theta, \tag{4}$$

$$\hat{z}_t = \sqrt{\alpha_t} \frac{\hat{z}_{t-1} - \sqrt{1-\alpha_{t-1}}\varepsilon_\theta}{\sqrt{\alpha_{t-1}}} + \sqrt{1-\alpha_t}\varepsilon_\theta, \tag{5}$$

$$\hat{z}^i_{n+1} = B * w^b(\hat{z}^i_n) + (1-B)z^i_{n+1}, \tag{8}$$

Visualization

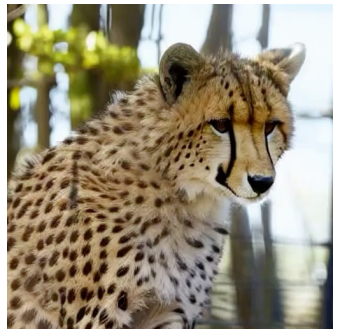
Input video

A marble statue of a woman is running

Bronze statue of a woman

Input video

A cheetah



A husky

Input video

A man wearing yellow trousers

Makoto Shinkai Style

Input video

Makoto Shinkai Style

Sunset ambiance

project page: https://ree1s.github.io/wave/

Thank you