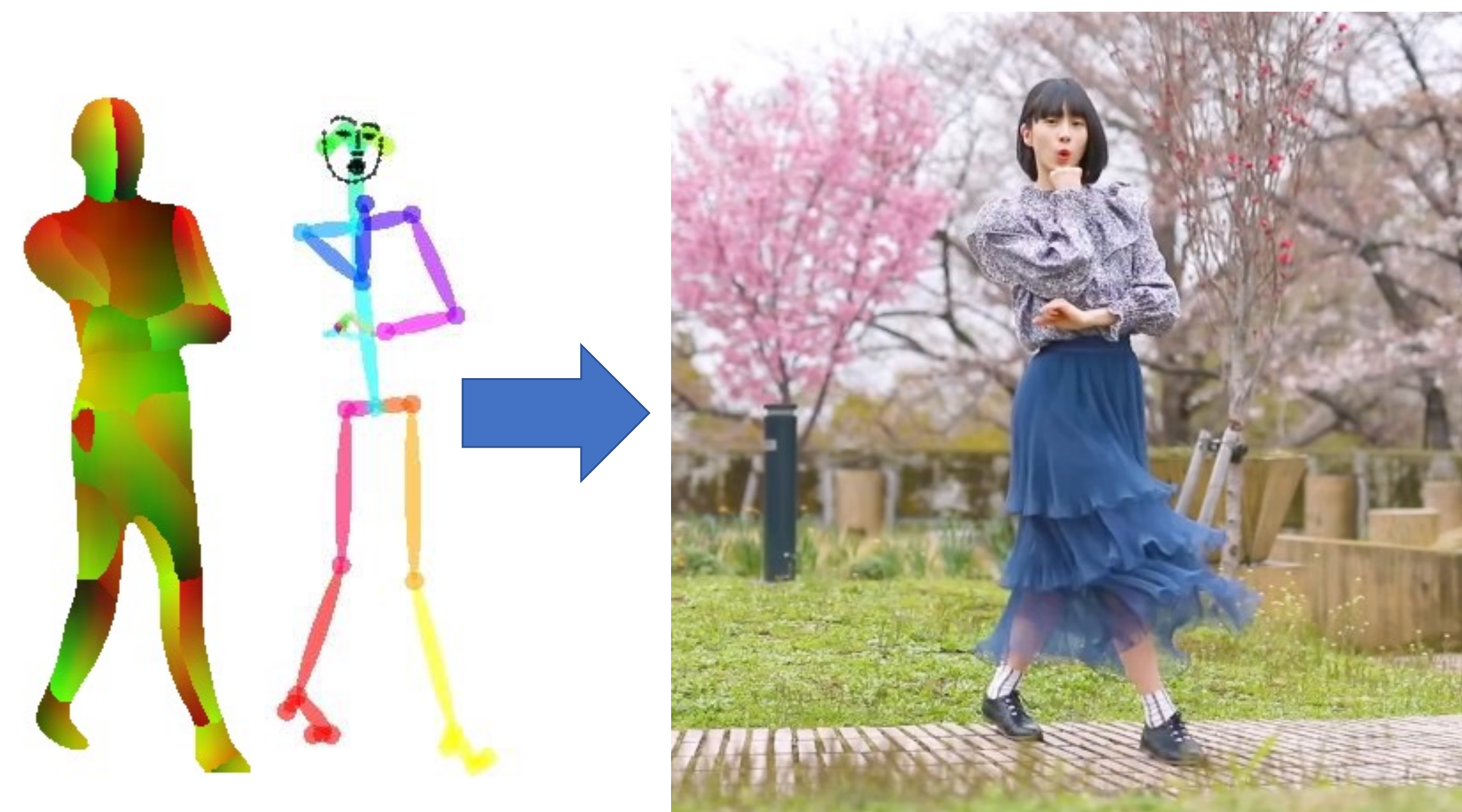# Dance In the Wild
Monocular Human Animation with Neural Dynamic Appearance Synthesis
Tuanfeng Wang, Duygu Ceylan, Krishna Kumar Singh, Niloy J. Mitra

Our Goal

Input:
- 2D pose/motion representation
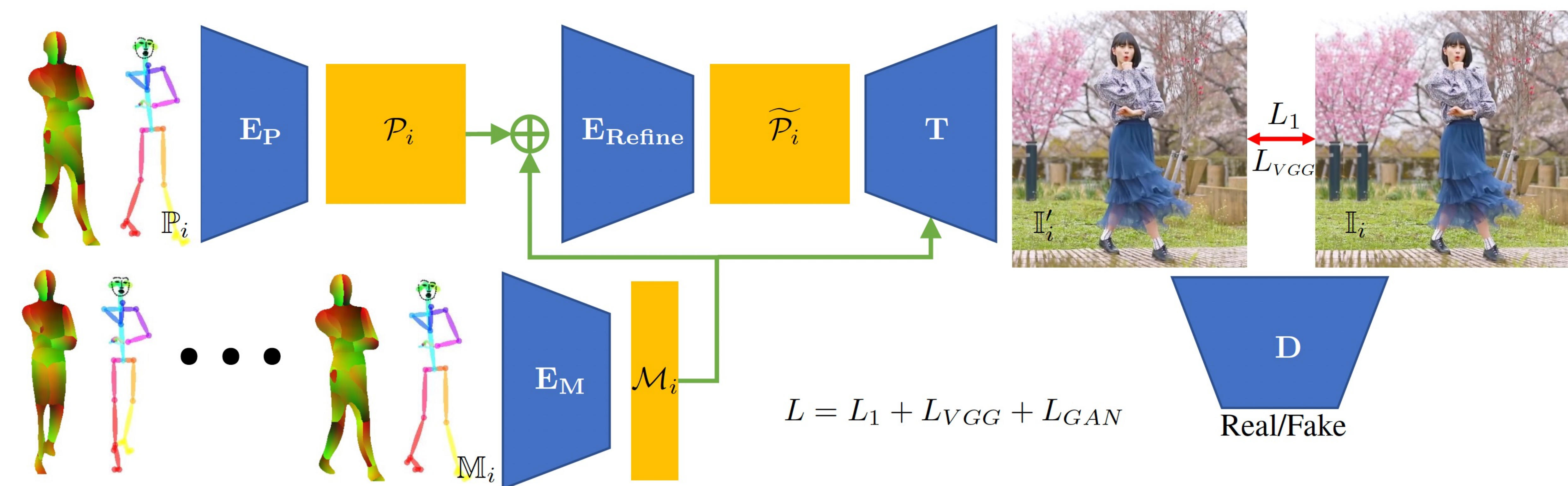
Output:
- 2D appearance synthesis

Tasks:
- Reconstruction
- Unseen motion synthesis (retargeting)

Reconstruction (SOTA)

Ground truth | Our | Pix2Pix | Vid2Vid | EDN | HF-NHMT

Application: retargeting

rest pose

target pose

Our framework

- Our network takes the 2D body dense UV and keypoints as input and learns a pose feature for each frame.
- By concatenating the pose inputs for the past few frames, we also learn motion features. The learned motion features are used to refine the pose features to improve the temporal coherency.
- We synthesize the final motion-aware dynamic appearance of the character using a StyleGAN based generator conditioned on the refined pose features and modulated by the motion features.

$$L = L_1 + L_{VGG} + L_{GAN}$$

$\mathbf{E_P}$ $\mathcal{P}_i$ $\mathbf{E_{Refine}}$ $\widetilde{\mathcal{P}_i}$ $\mathbf{T}$ $\mathbb{P}_i$ $\mathbf{E_M}$ $\mathcal{M}_i$ $\mathbb{M}_i$ $\mathbb{I}'_i$ L_1 L_{VGG} $\mathbb{I}_i$ $\mathbf{D}$ Real/Fake