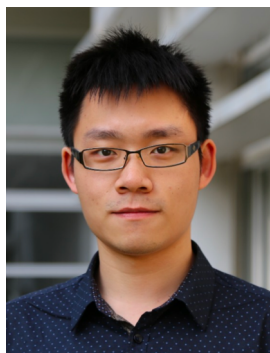


Disentangled Person Image Generation

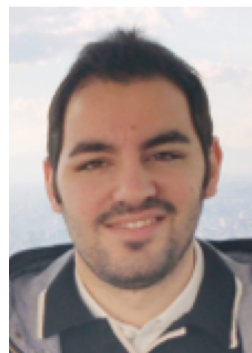
Paper ID: 1801 Project page: https://homes.esat.kuleuven.be/~liqianma/CVPR18_DPIG/



Liqian¹



Qianru²



Stamatis¹



Luc^{1,3}



Bernt²



Mario²

¹KU Leuven

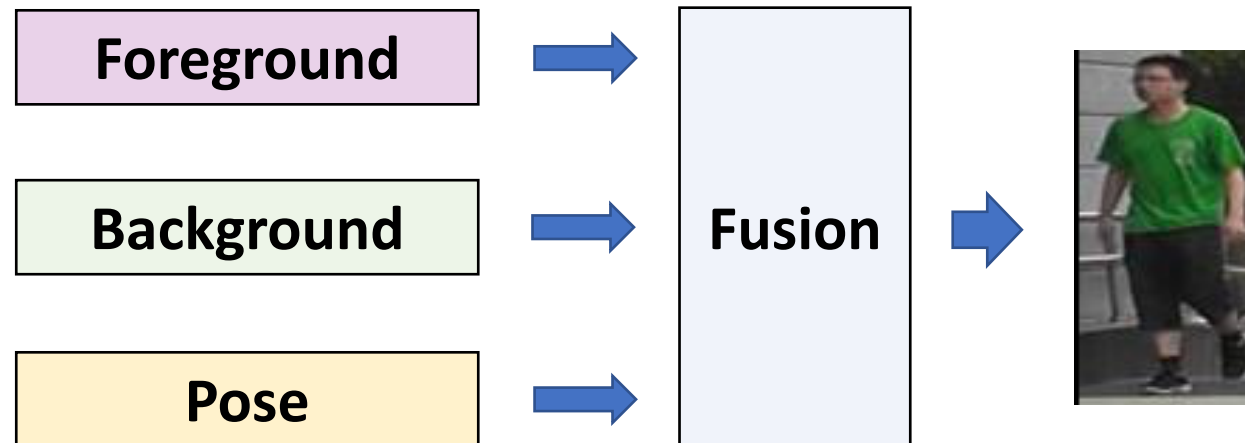
²MPI for Informatics

³ETH Zürich



max planck institut
informatik

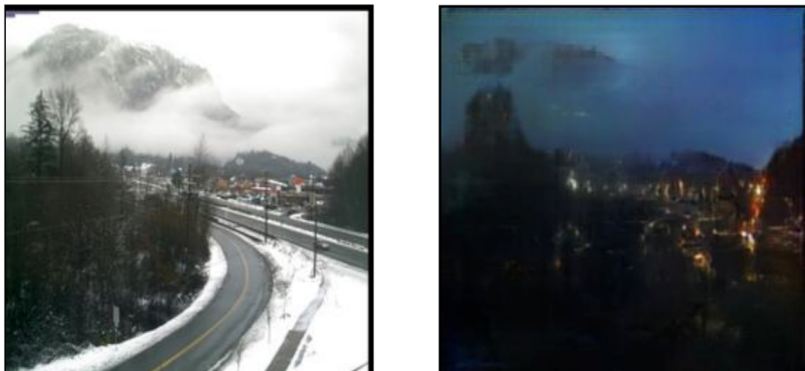
- **Motivation:** Learn image generation model for persons that explicitly represents foreground, background, and pose.
- **Task:** Synthesize person images, while independently controlling foreground, background, and pose, in a *self-supervised* way.
- **Key idea:** *Disentangle* person images into the aforementioned components, and then combine.



Related work

-- Disentangled Person Image Generation

Day → Night



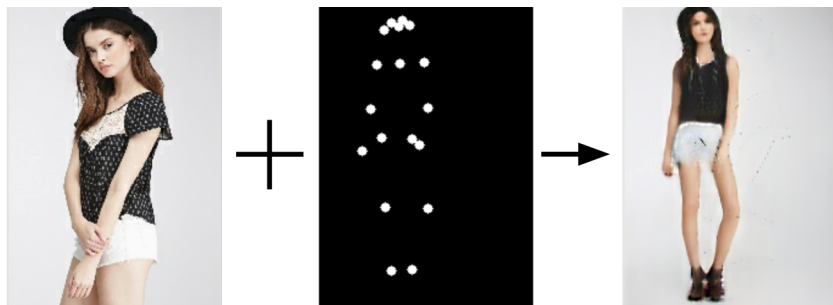
Pix2Pix [1]

Zebras ↔ Horses

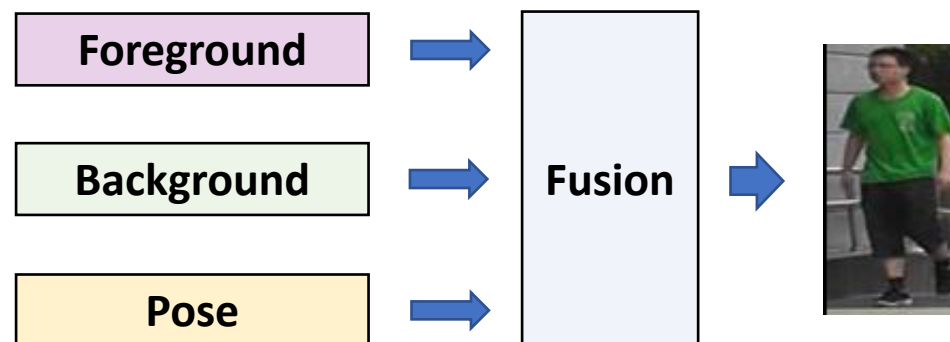


CycleGAN [2]

Condition image Target pose Generated image



PG² [3]



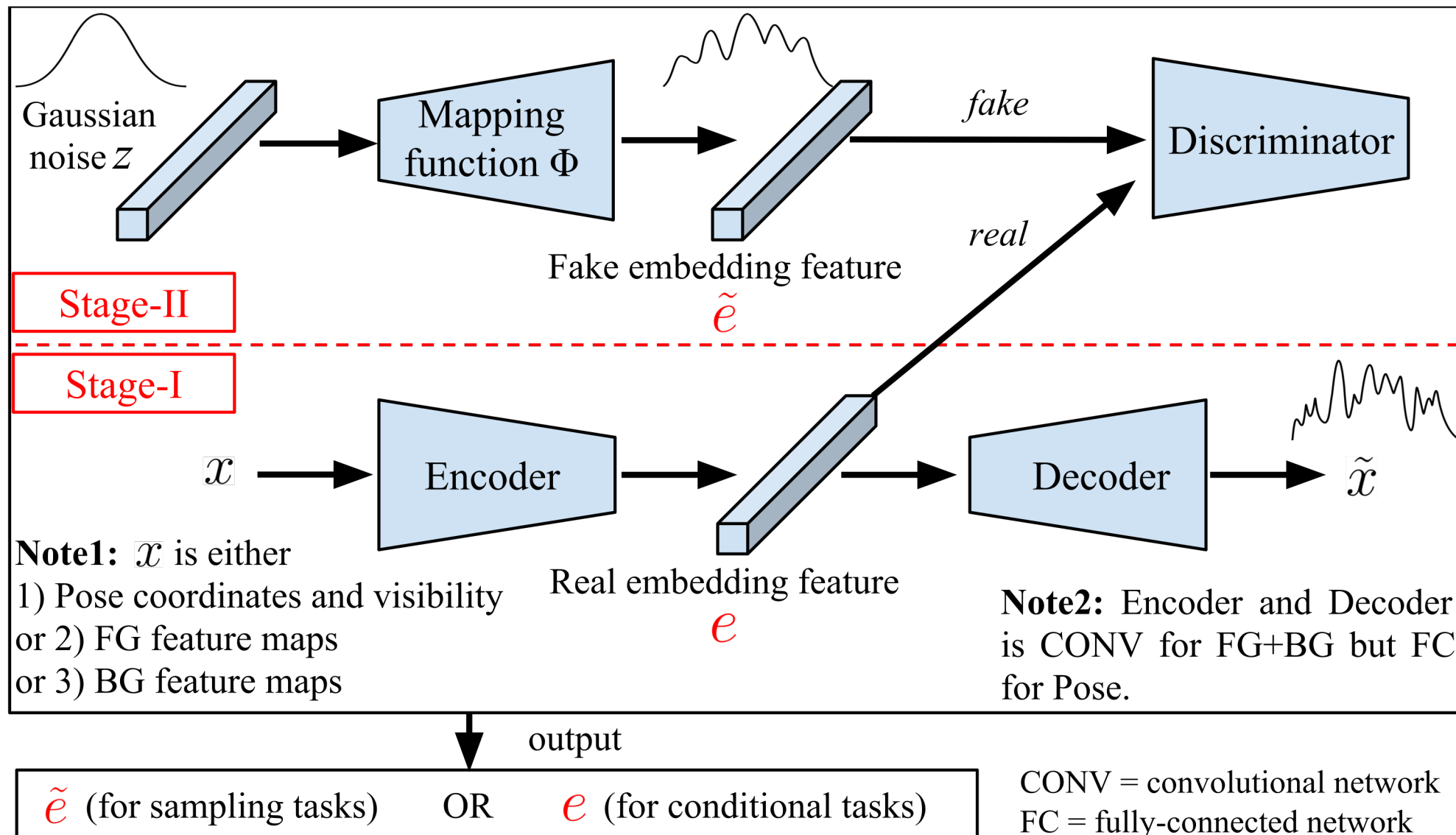
Ours

[1] P. Isola *et al.* Image-to-Image Translation with Conditional Adversarial Networks. In CVPR, 2017.

[2] J.-Y. Zhu *et al.* Unpaired imageto-image translation using cycle-consistent adversarial networks. In ICCV, 2017

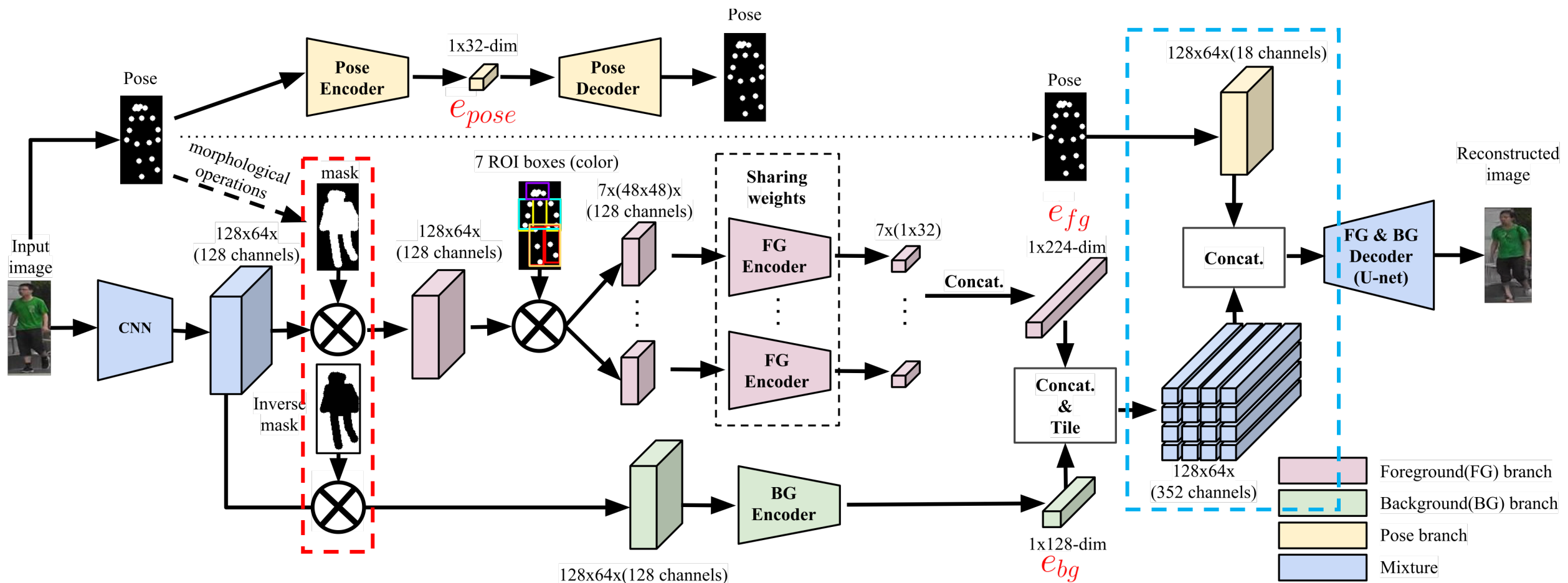
[3] L. Ma *et al.* Pose guided person image generation. In NIPS, 2017.

Framework



Framework Stage-I

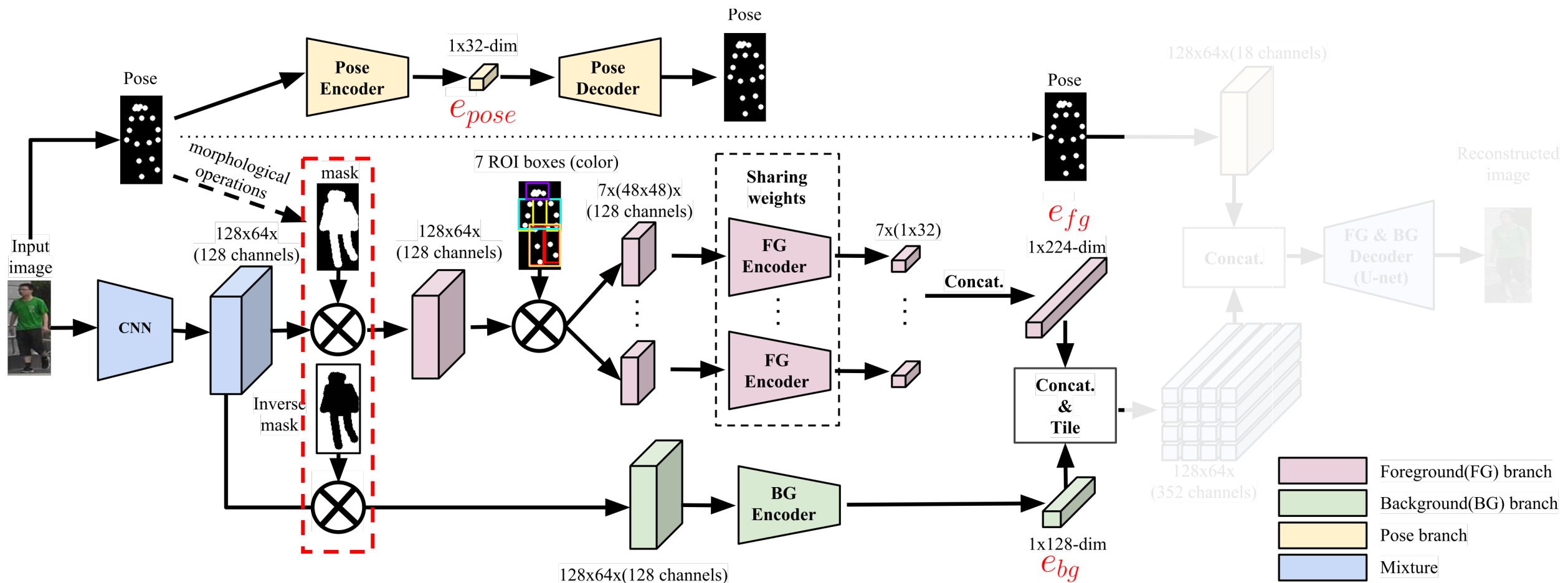
-- Disentangled Person Image Generation



Stage-I: Reconstruction pipeline. Three disentangled branches: foreground (FG), background (BG) and pose.

Framework Stage-I

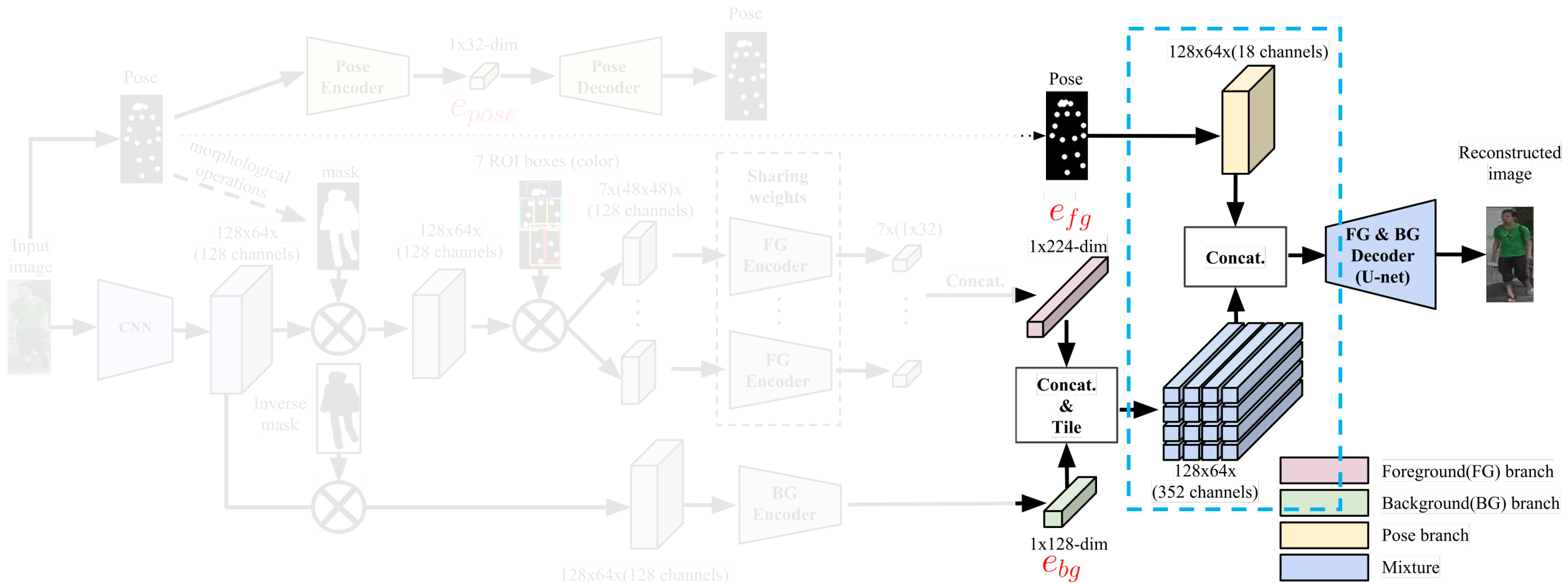
-- Disentangled Person Image Generation



Stage-I: Reconstruction pipeline. Three encoder branches for foreground (FG), background (BG) and pose.

Framework Stage-I

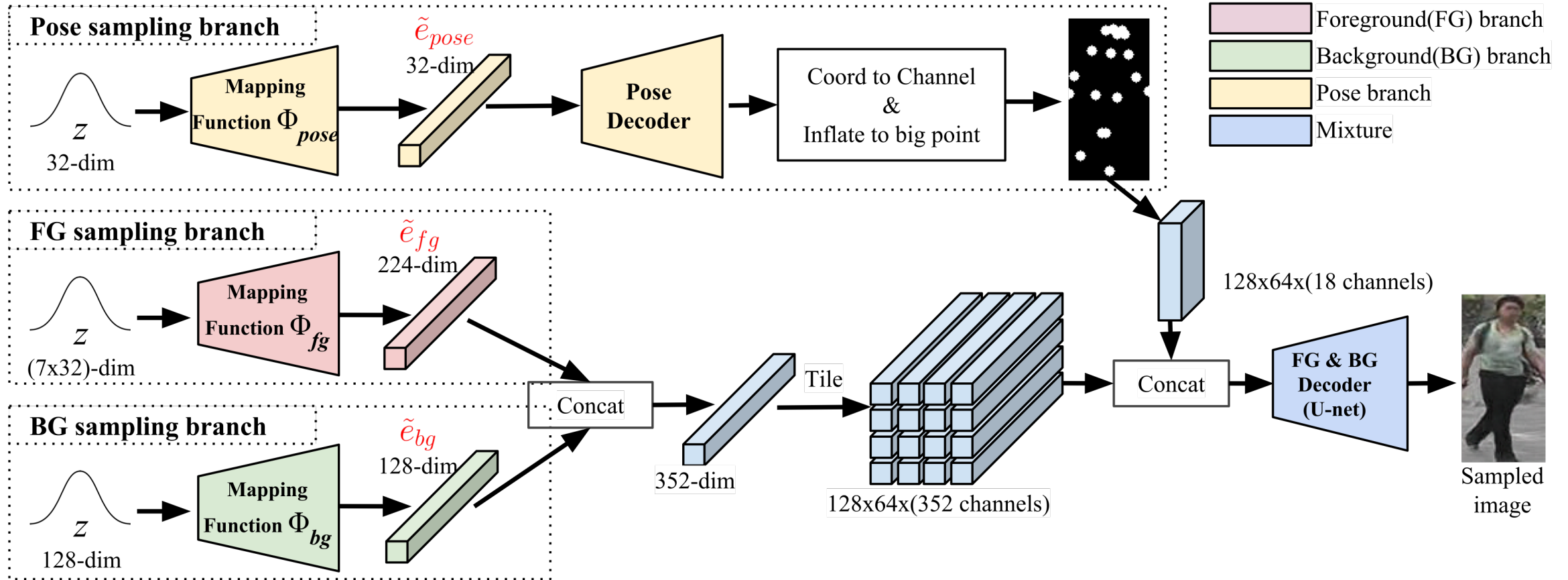
-- Disentangled Person Image Generation



Stage-I: Fuse three factors. Tile the appearance feature vector and then concatenate it with pose keypoints to encourage the Decoder to select appearance info. with pose keypoints.

Sampling phase (testing)

-- Disentangled Person Image Generation



Sampling phase: Sample foreground, background and pose from Gaussian noise to compose new person images.

Results

Experiments – sampling

-- Disentangled Person Image Generation



1) FG sampling (fixed BG and Pose)



3) Pose sampling (fixed FG and BG)



2) BG sampling (fixed FG and Pose)



4) FG, BG and Pose sampling

Market-1501: Sampling different factors



DeepFashion: Appearance and Pose sampling

Experiments – sampling

-- Disentangled Person Image Generation



VAE



DCGAN



Ours - Whole Body



Ours – BodyROI7



Ours - BodyROI7 with real pose



Real data

Sampling results comparison.

Our method can result in more realistic natural person images

Experiments – interpolation

-- Disentangled Person Image Generation

Same person



Same person



Different persons





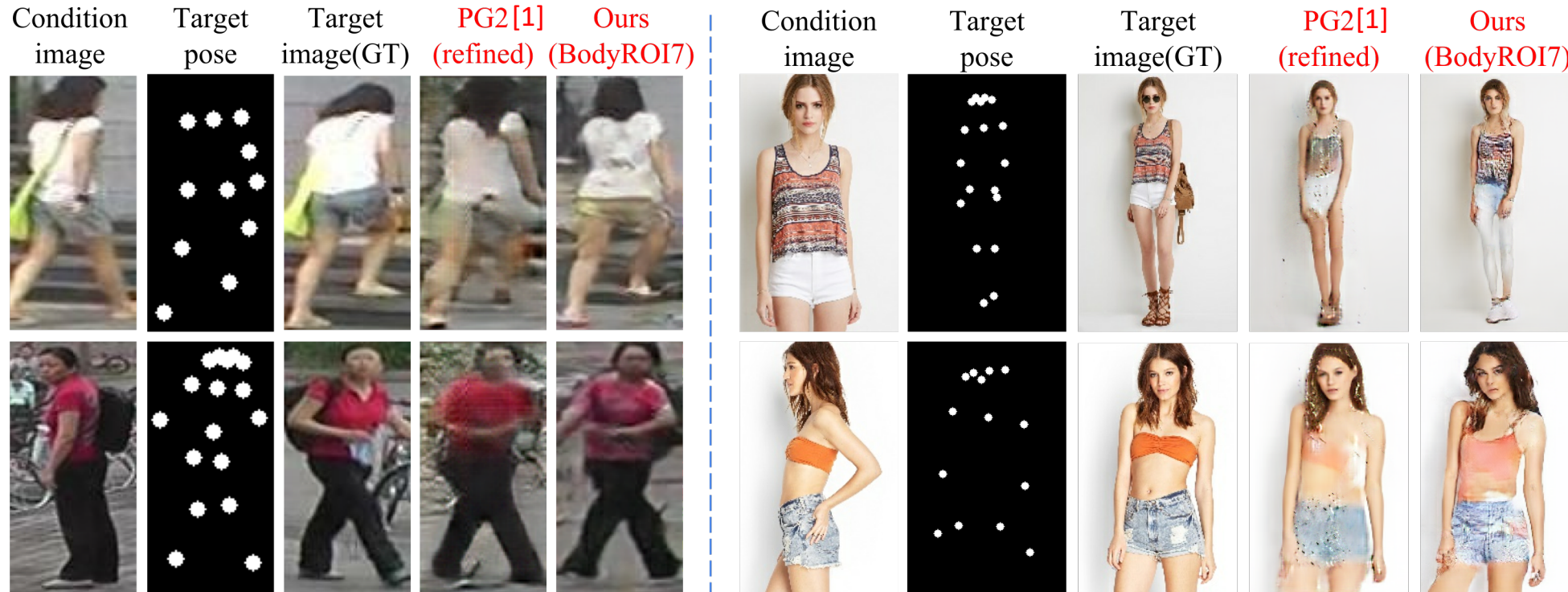
Our generated Virtual Market (VM) dataset

Model	Training data	Rank-1	mAP
Res50 [1]	CUHK03 (labeled)	0.300	0.115
Res50 [1]	Duke (labeled)	0.361	0.142
Res50	VM	0.338	0.134
Res50+PUL	VM+Market	0.369	0.156
Res50+PUL+KISSME	VM+Market	0.375	0.154

Our generated data can benefit unsupervised re-ID

Experiments – human pose transfer

-- Disentangled Person Image Generation

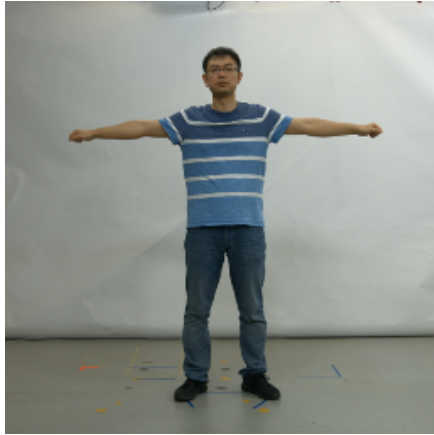


Model	DeepFashion		Market-1501			
	SSIM	IS	SSIM	IS	Mask-SSIM	Mask-IS
PG2[1]	0.762	3.090	0.253	3.460	0.792	3.435
Ours	0.614	3.228	0.099	3.483	0.614	3.491

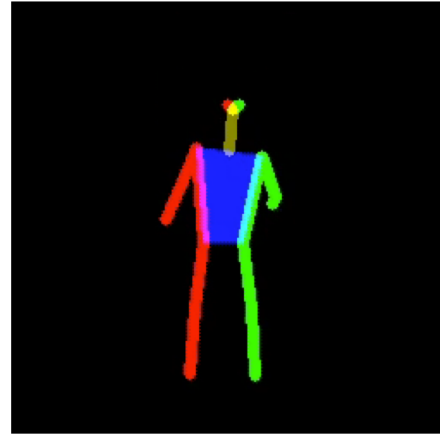
Experiments – video generation

-- Disentangled Person Image Generation

Person A



Person B's pose



+



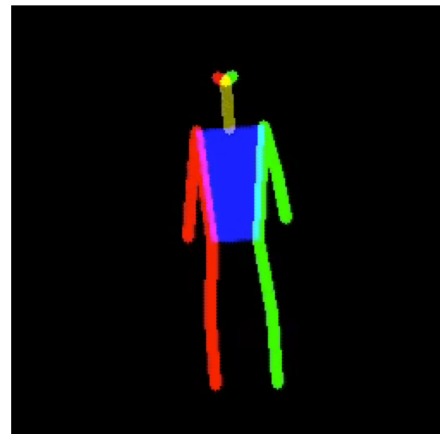
Result



Person B



Person A's pose



+



Result



Thank you !