Exploring Dual-task correlation for Pose Guided Person Image Generation

Pengze Zhang, Lingxiao Yang, Jianhuang Lai and Xiaohua Xie*

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Pose Guided Person Image Generation

(a) Source image (b) Target image (c) Vanilla CNN based method (d) Attention based method (e) Optical flow based method (f) Parsing map based method (g) Our method

Transform a person image from the source pose to a given target pose.

Analysis of Existing Methods

1. Solely focus on the **Source-to-Target Task**, which is an **ill-posed problem**, making it arduous to train a robust generator.

2. Cannot well capture the reasonable **texture mapping** between the source and target, especially when the person undergoes **large pose changes**.

Motivation

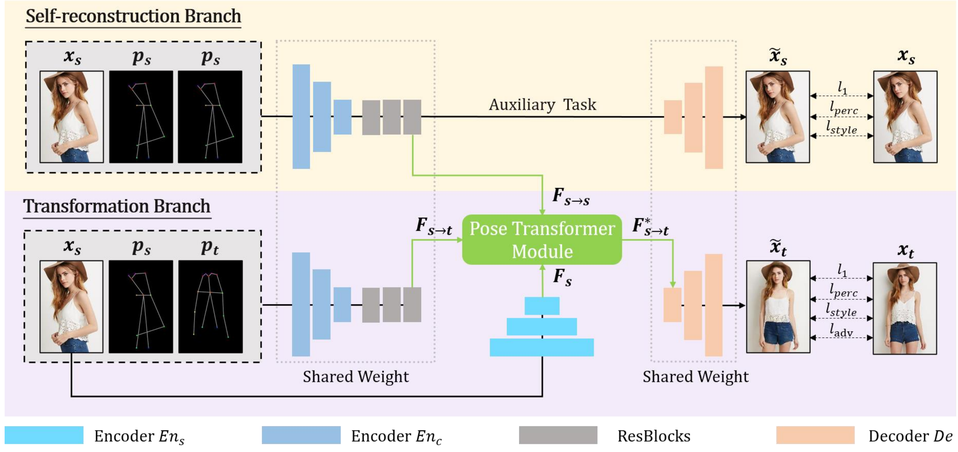
Task	Function	Difficulty	Feature
Source-to-Target Task	$G(\boldsymbol{x_s}, \boldsymbol{p_s}, \boldsymbol{p_t}) = \widetilde{\boldsymbol{x}}_t$	Hard	Aligned w $\boldsymbol{p_t}$
Source-to-Source Task	$\boldsymbol{G(x_s, p_s, p_s) = \widetilde{x}_s}$	**Easy**	**Aligned w $\boldsymbol{p_s}$**

1. Introduce an **auxiliary task**, i.e. **Source-to-Source Task**, by **Siamese structure**.

2. Design a **transformer-based module** to explore **texture correlation** between **dual-task** features.

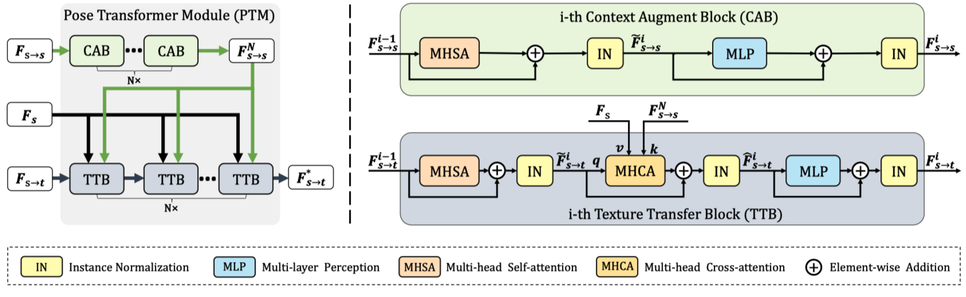
Code is available at: https://github.com/PangzeCheung/Dual-task-Pose-Transformer-Network

Dual-task Pose Transformer Network (DPTN)



Overview of our model. It contains a **self-reconstruction branch** for auxiliary **source-to-source task**, and **a transformation branch** for **source-to-target task**. These two branches share partial weights and are communicated by a **pose transformer module**.

Pose Transformer Module (PTM)



Structure of the **Pose Transformer Module (PTM)**. It contains two types of blocks: **Context Augment Block (CAB)** and **Texture Transfer Block (TTB)**. The CABs integrate the information of the feature $\boldsymbol{F_{s\to s}}$, while the TTBs transfer the real source image textures from $\boldsymbol{F_s}$ to optimize $\boldsymbol{F_{s\to t}}$ by capturing the correlation between features from the dual tasks.

Quantitative comparisons

Model	DeepFashion				Market1501				Number of Parameters ↓
	SSIM ↑	PSNR ↑	FID ↓	LPIPS ↓	SSIM ↑	PSNR ↑	FID ↓	LPIPS ↓	
PG2 [22] (NeurIPS'17)	0.7730	17.5324	49.5674	0.2928	0.2704	14.1749	86.0288	0.3619	437.09 M
VU-net [4] (CVPR'18)	0.7639	17.6582	15.5747	0.2415	0.2665	14.4220	44.2743	0.3285	139.36 M
DSC [27] (CVPR'18)	0.7682	18.0990	21.2686	0.2440	0.3054	14.3081	27.0118	0.3029	82.08 M
PATN [45] (CVPR'19)	0.7717	18.2543	20.7500	0.2536	0.2818	14.2622	22.6814	0.3194	41.36 M
DIAF [18] (CVPR'19)	0.7738	16.9004	14.8825	0.2388	0.3052	14.2011	32.8787	0.3059	49.58 M
DIST [24] (CVPR'20)	0.7677	18.5737	**10.8429**	0.2258	0.2808	14.3368	19.7403	0.2815	14.04 M
XingGAN [30] (ECCV'20)	0.7706	17.9226	39.3194	0.2928	0.3044	14.4458	22.5198	0.3058	42.77 M
PISE* [39] (CVPR'21)	0.7682	18.5208	11.5144	0.2080	—	—	—	—	64.01 M
SPIG* [21] (CVPR'21)	0.7758	18.5867	12.7027	0.2102	**0.3139**	14.4894	23.0573	0.2777	117.13 M
Ours	**0.7782**	**19.1492**	11.4664	**0.1957**	0.2854	**14.5207**	**18.9946**	**0.2711**	**9.79** M

Qualitative comparison

Visualization of PTM