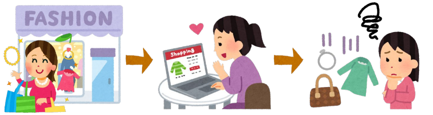
Pose aware Outfit Transfer between Unpaired in-the-wild images

Donnaphat Trakulwaranont, Marc A. Kastner, Shin'ichi Satoh

✉ eiam@nii.ac.jp

Introduction

- Fashion is one of highest revenue industries in the world.
- Online fashion shopping has some obstacles: Hard to judge a product’s look on oneself



Motivation

- However, outfit transfer method has various existing problems:
  - ○ Usually trained on studio photography (No backgrounds)
  - ○ Need for paired data
- We want to handle noisy backgrounds such as in street photography

Studio photography

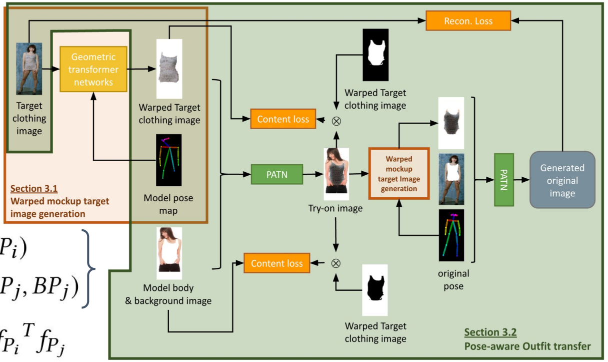
Paired data

Street photography (Noisy background)

Proposed method

Our method has 2 main modules

1. Warp mockup target clothing image generation
   - ○ It is used for generate target clothing image close to target model body shape.


Target clothing image
Geometric transformer networks
Warped Target clothing image
Model pose map
Section 3.1
Warped mockup target image generation
Model body & background image
Recon. Loss
Content loss
PATN
Try-on image
Warped Target clothing image
Warped mockup target Image generation
Content loss
Warped Target clothing image
original pose
PATN
Generated original image
Section 3.2
Pose-aware Outfit transfer

Feature extractor: $\left\{ \begin{array}{l} f_{P_i} = F_{e_A}(P_i) \\ f_{P_j} = F_{e_B}(P_j, BP_j) \end{array} \right\}$

Correlation: $c(i,j) = {f_{P_i}}^T f_{P_j}$

Transformation parameter: $\theta(i,j) = Regression(c(i,j))$, Transform image: $W^j_{P_i} = \mathcal{T}_\theta(P_i)$

2. Pose-aware Outfit transfer
   - ○ It is used to generate image of target person wear specific clothing, and the reconstruction part is introduced to train this module as we have real pair data.

Outfit Transfer Image: $O^f_j = PATN\left(B_j, W^j_{P_i}, BP_j\right)$, Reconstructed original Image: $O^f_i = PATN\left(B_i, W^i_{P_j}, BP_i\right)$

Training

- For warp mockup target clothing image generation
  - ○ use L1-Loss to train

$$\mathcal{L}(\theta, \theta_{GT}) = \frac{1}{N}\sum_{i=1}^{N} \|G'_i - G''_i\|$$

- For pose-aware outfit transfer
  - ○ use GAN-Loss, combination L1-loss and reconstruction loss to train

$$\mathcal{L} = \mathcal{L}_{GAN} + \mathcal{L}_{combL1} + \mathcal{L}_{recon}$$

$$\mathcal{L}_{recon} = \left\|\Psi_k(O_i) - \Psi_k(O^f_i)\right\|_1$$

$$\mathcal{L}_1 = \left\|W^j_{P_i} - (O^f_j \otimes PM_j)\right\|_1 + \left\|B_j - (O^f_j \otimes BM_j)\right\|_1$$

$$\mathcal{L}_{perL1} = \left\|\Psi_k(W^j_{P_i}) - \Psi_k(O^f_j \otimes PM_j)\right\|_1 + \left\|\Psi_k(B_j) - \Psi_k(O^f_j \otimes BM_j)\right\|_1$$

$$\mathcal{L}_{combL1} = \lambda_1 \mathcal{L}_1 + \lambda_2 \mathcal{L}_{perL1}$$

$$\mathcal{L}_{GAN} = \mathbb{E}_{BP_j \in \mathcal{P}_P, (O_i, O_j) \in \mathcal{X}} \left\{ \log \left[D_A(O_i, O_j) \cdot D_S(BP_j, O_j) \right] \right\} + \mathbb{E}_{BP_j \in \mathcal{P}_P, O_j \in \mathcal{X}, O^f_j \in \hat{\mathcal{X}}} \left\{ \log \left[\left(1 - D_A(O_j, O^f_j)\right) \cdot \left(1 - D_S(BP_j, O^f_j)\right) \right] \right\}$$

Evaluation

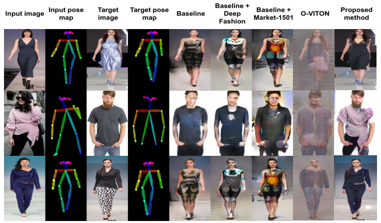
Quantitative results

- Using Structural similarity (SSIM) and Inception score (IS)
- Promising results: For SSIM and mask-SSIM, around 54% and 45% improvement, while for IS around 25% improvement

Model	SSIM	IS	Mask-SSIM	Mask-IS
Baseline	0.302	4.073	0.591	4.444
Baseline + DeepFashion	0.282	4.073	0.580	4.562
Baseline + Market-1501	0.256	3.912	0.565	4.064
O-VTON	0.253	2.648	0.320	3.292
Proposed method	0.467	5.096	0.860	3.994

Qualitative results

- Confirms a perception much closer to the expectation
- We can accurately preserve background information while still being able to correctly transfer the outfit



Real scenerio

- Test on myself
  - ○ Outfit images from fashion shopping website
  - ○ My image from webcam or real street photograph