DGPose: Deep Generative Models for Human Body Analysis

-IJCV 2020

-University of Oxford

Sep 18, 2020

Chuan Guo

Contributions

Preliminaries

Method

Datasets

Experiments

A versatile deep generative model for multiple purpose in human body analysis:

- Human image reconstruction
- Human image generation
- Pose transfer
- Pose estimation

in a semi-supervision manner.

Contributions

Preliminaries

Method

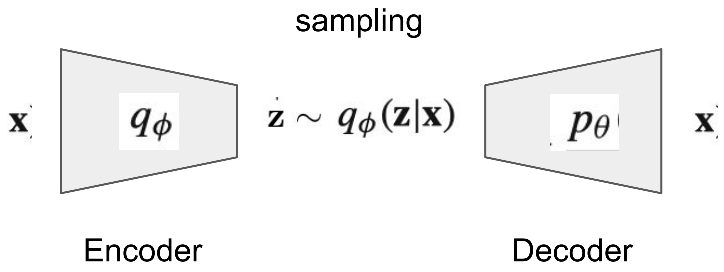
Datasets

Experiments

Preliminaries

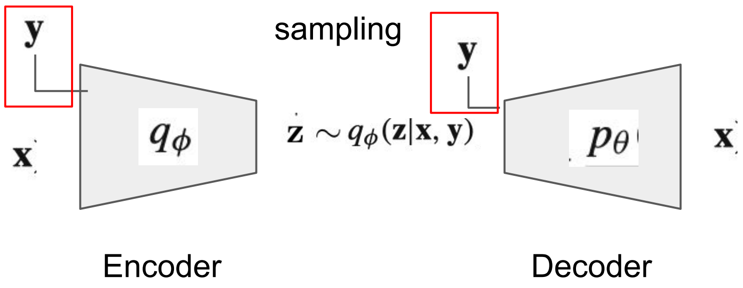
- Variational AutoEncoder (VAE)
- Conditional VAE
- Semi-supervised CVAE
- VAEGAN

Variational AutoEncoder



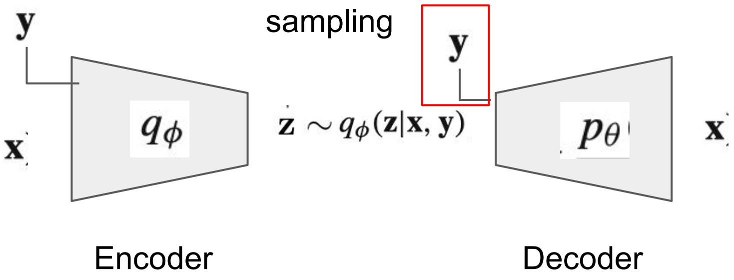
$$\begin{aligned}\log\ p_\theta(\mathbf{x}) &\geq \mathcal{L}_{\text{VAE}}(\phi, \theta; \mathbf{x}) \\ &= \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x})}\right].\end{aligned}$$

Conditional VAE



$$\log\ p_\theta(\mathbf{x}|\mathbf{y}) \geq \mathcal{L}_{\mathrm{CVAE}}(\phi, \theta; \mathbf{x}|\mathbf{y})$$
$$= \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x},\mathbf{y})}\left[\log \frac{p_\theta(\mathbf{x}, \mathbf{z}|\mathbf{y})}{q_\phi(\mathbf{z}|\mathbf{x}, \mathbf{y})}\right].$$

Semi-supervised CVAE



$$\mathcal{L}_{\mathrm{SS}}(\theta, \phi; \mathcal{D}) = \sum_{\mathbf{x}_u \in \mathcal{D}_u} \mathcal{L}_u(\theta, \phi; \mathbf{x}_u)$$

Unlabeled data

$$+ \gamma \sum_{(\mathbf{x}_s, \mathbf{y}_s) \in \mathcal{D}_s} \mathcal{L}_s(\theta, \phi; \mathbf{x}_s, \mathbf{y}_s)$$

labeled data

where $\mathcal{L}_u$ and $\mathcal{L}_s$ are defined as

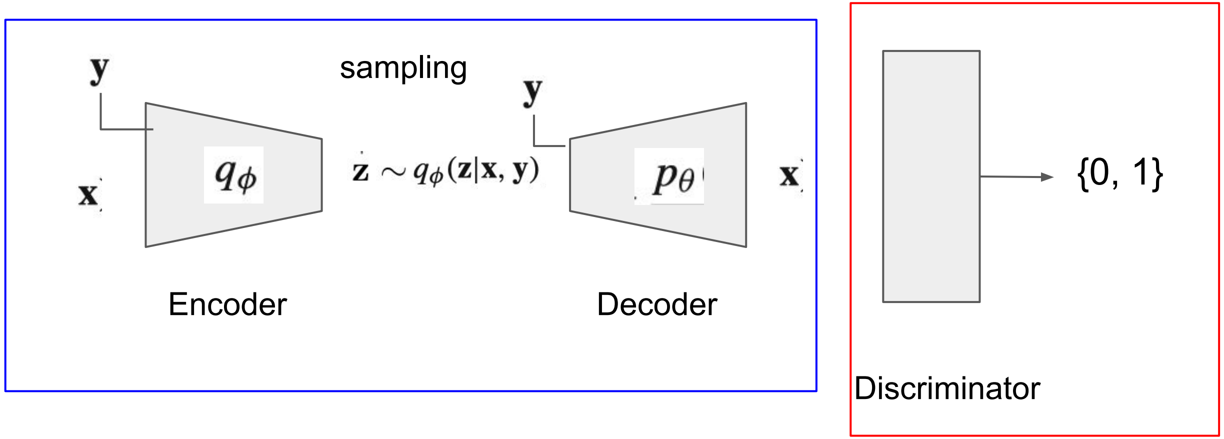
Standard VAE

$$\mathcal{L}_u(\theta, \phi; \mathbf{x}_u) = \mathcal{L}_{\mathrm{VAE}}(\theta, \phi; \mathbf{x}_u), \text{ and}$$

$$\mathcal{L}_s(\theta, \phi; \mathbf{x}_s, \mathbf{y}_s) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x}_s, \mathbf{y}_s)} \left[\log \frac{p_\theta(\mathbf{x}_s, \mathbf{z}|\mathbf{y}_s)}{q_\phi(\mathbf{z}|\mathbf{x}_s, \mathbf{y}_s)} \right] + \alpha \log q_\phi(\mathbf{y}_s|\mathbf{x}_s),$$

CVAE
+ Recognition Loss

CVAE-GAN



$$\mathcal{L} = \mathcal{L}_{\text{VAE}} + \mathcal{L}_{\text{GAN}}.$$

Contributions

Preliminaries

Methods

Datasets

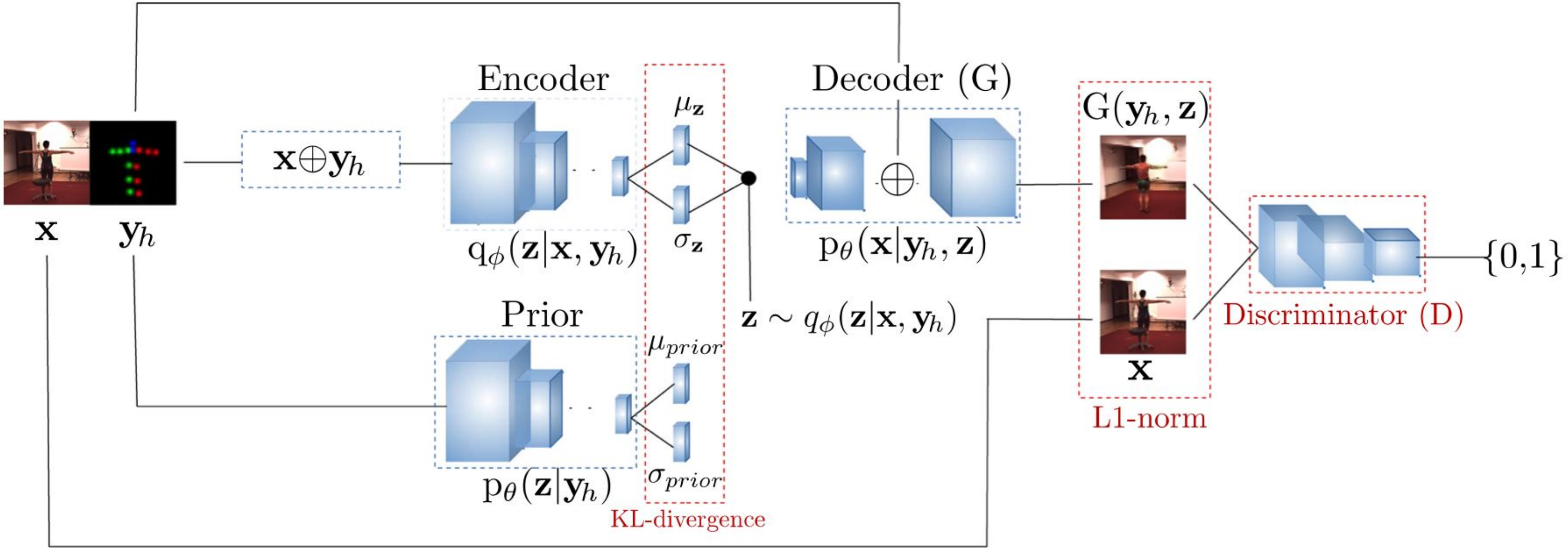
Experiments

Methods

- Conditional-DGPose
  - Full supervision

- Semi-DGPose
  - Semi-supervision

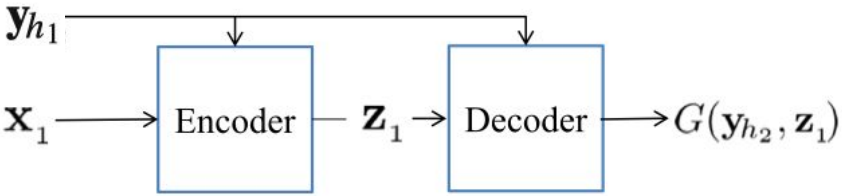
Architecture of Conditional-DGPose



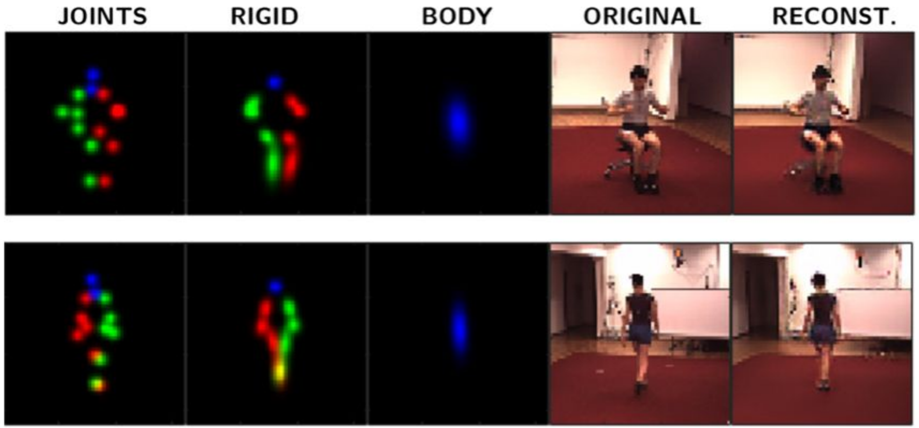
Loss = KL-divergence + L1-Norm + Adversarial

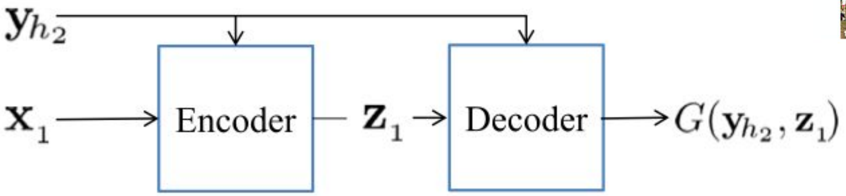
Applications of Conditional-DGPose

1.Reconstruction
2.Pose transfer
3. Conditional generation

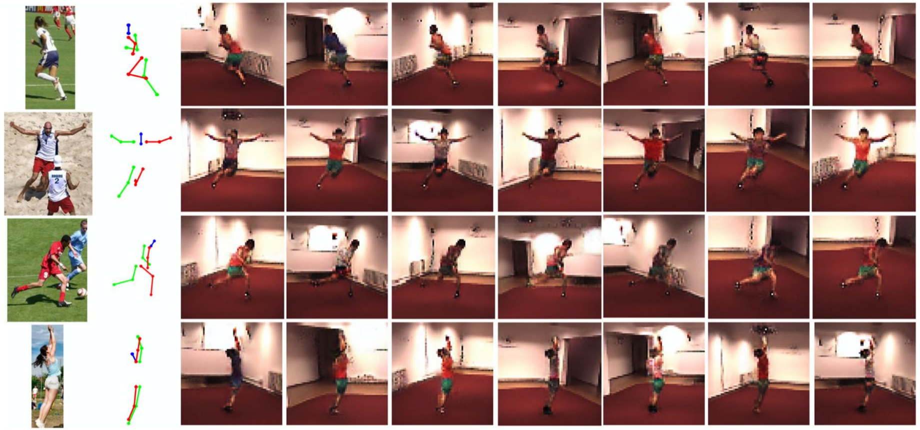


(a) Reconstruction
(same pose)

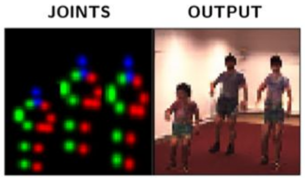




(b) Pose Transfer/Manipulation
(different pose)

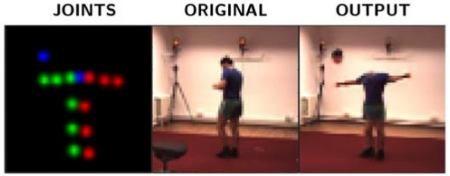


Cross-domain pose transfer

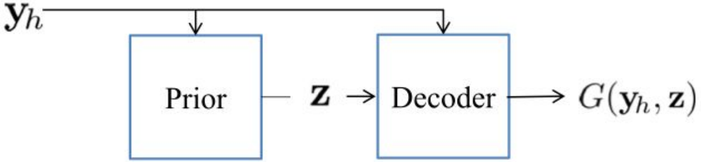


(a)

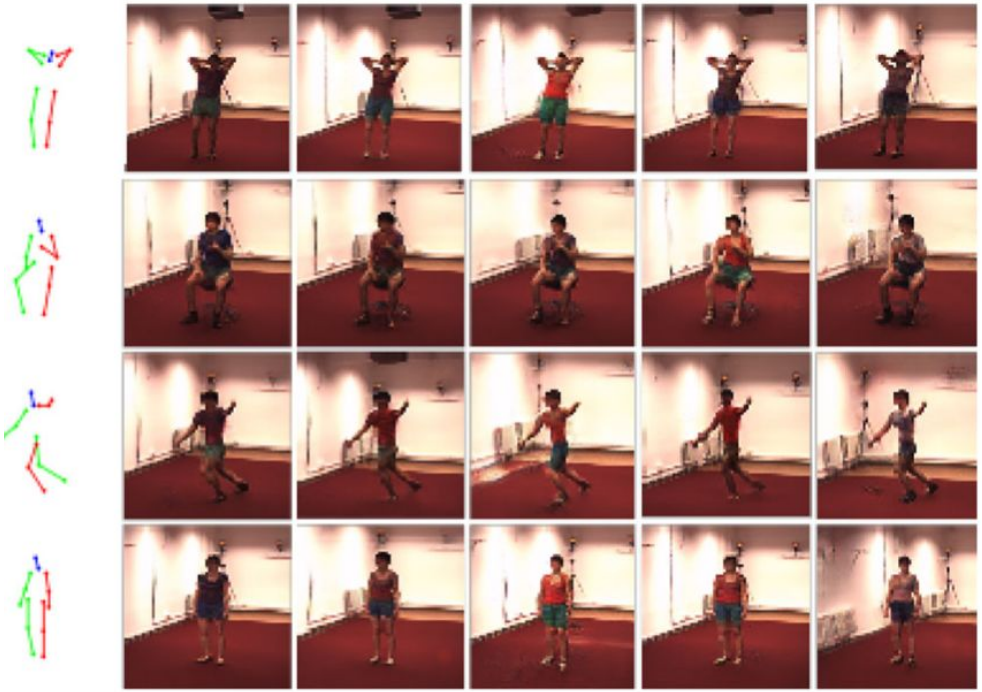
(b)



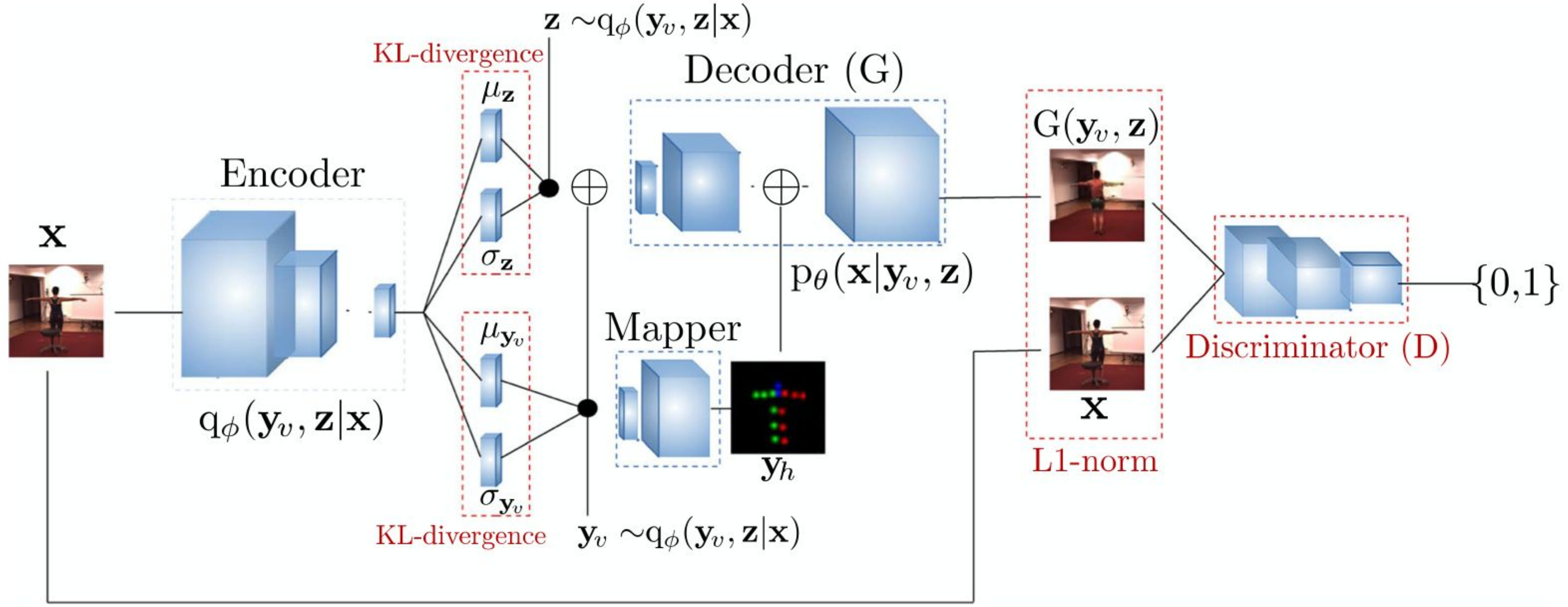
(c)



(c) Conditional image generation.



Architecture of Semi-DGPose

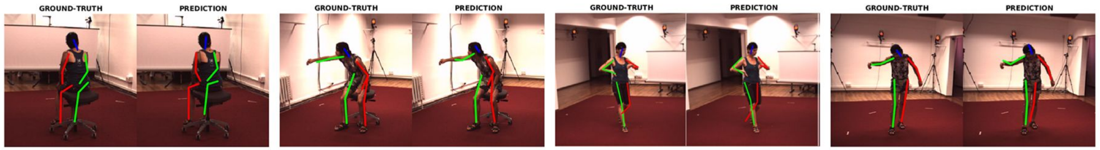


Mapper: an offline-learned neural unit which maps pose vector to pose heatmap.

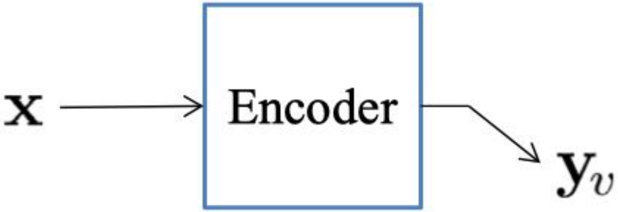
Loss = Loss_unlabel + Loss_label
Loss_unlabel = KL (top) + KL (bottom) + L1-norm + Adversarial
Loss_label = KL (top) + Pose regression loss + L1_norm + Adversarial

Applications of Semi-DGPose

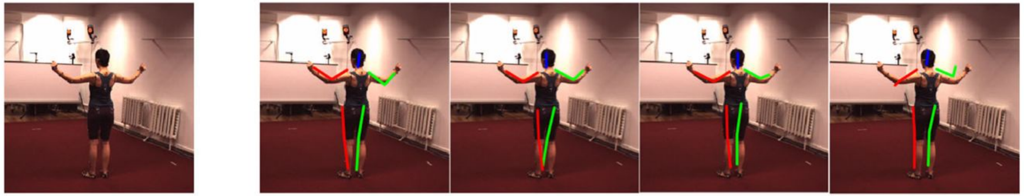
1. Pose estimation
2. Reconstruction
3. Indirect Pose transfer
4. Conditional generation



(a) PCK=92.9% (b) PCK=100.0% (c) PCK=96.4% (d) PCK=100.0%

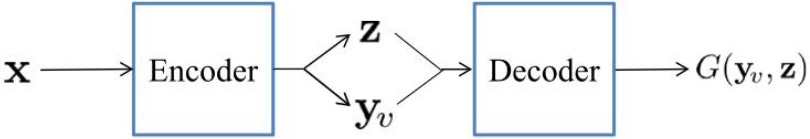


(a) Pose estimation.

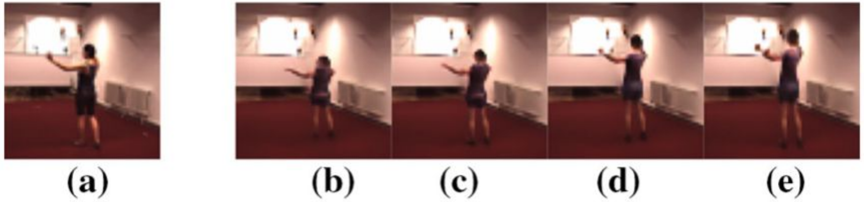


(a) (b) (c) (d) (e)

With 25%, 50%, 75%, 100% of supervision.



(b) Reconstruction.



Direct manipulation by change person's height.

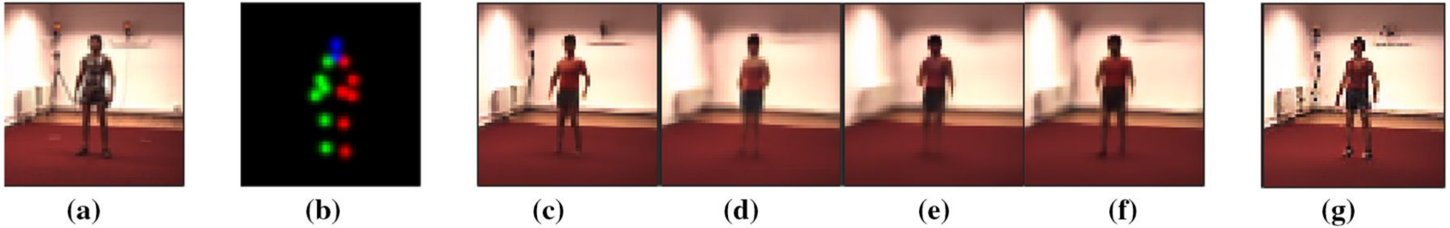
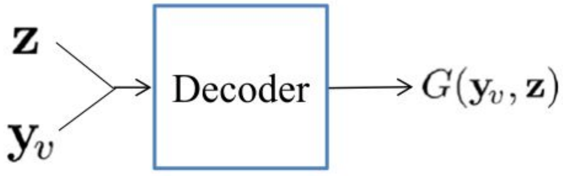
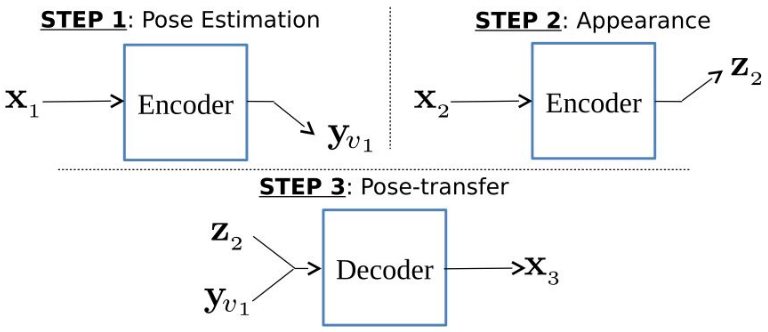


Image reconstruction with 100%, 75%, 50%, 25% of supervision, and Conditional-DGPose.

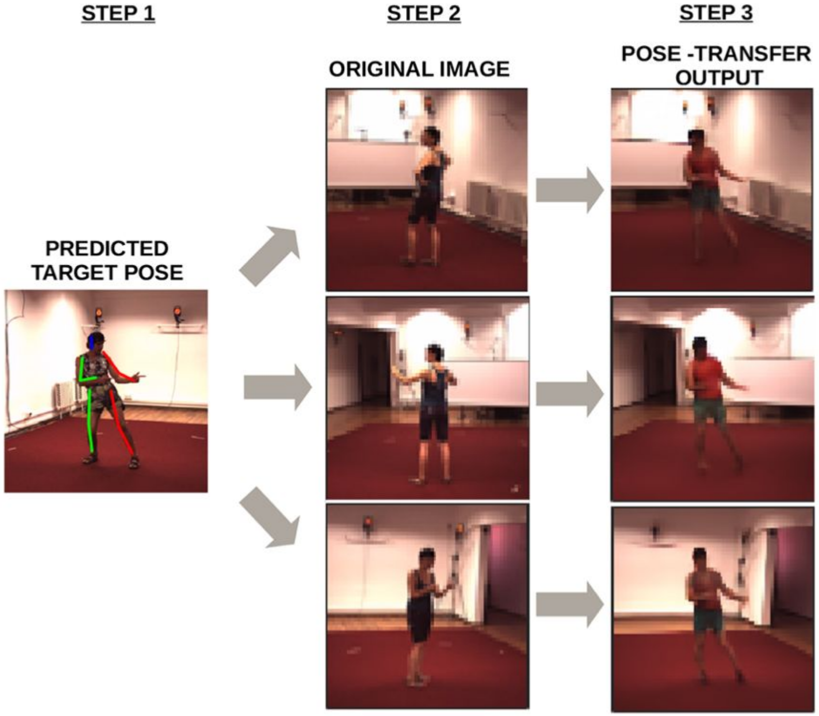


Not given.

(c) Image generation.



(d) Indirect Pose transfer.



Contributions

Preliminaries

Methods

Datasets

Experiments

- Human3.6 M
    - 317,989 and 1280 images for training and testing
    - Resolution of 1000 x 1000
- ChictopicalPlus
    - 23,011 and 2873 images for training and testing
    - Resolution of 286 x 286
- DeepFashion
    - 44,950 and 6560 images for training and testing
    - Resolution of 256 x 256

Contributions

Preliminaries

Methods

Datasets

Experiments

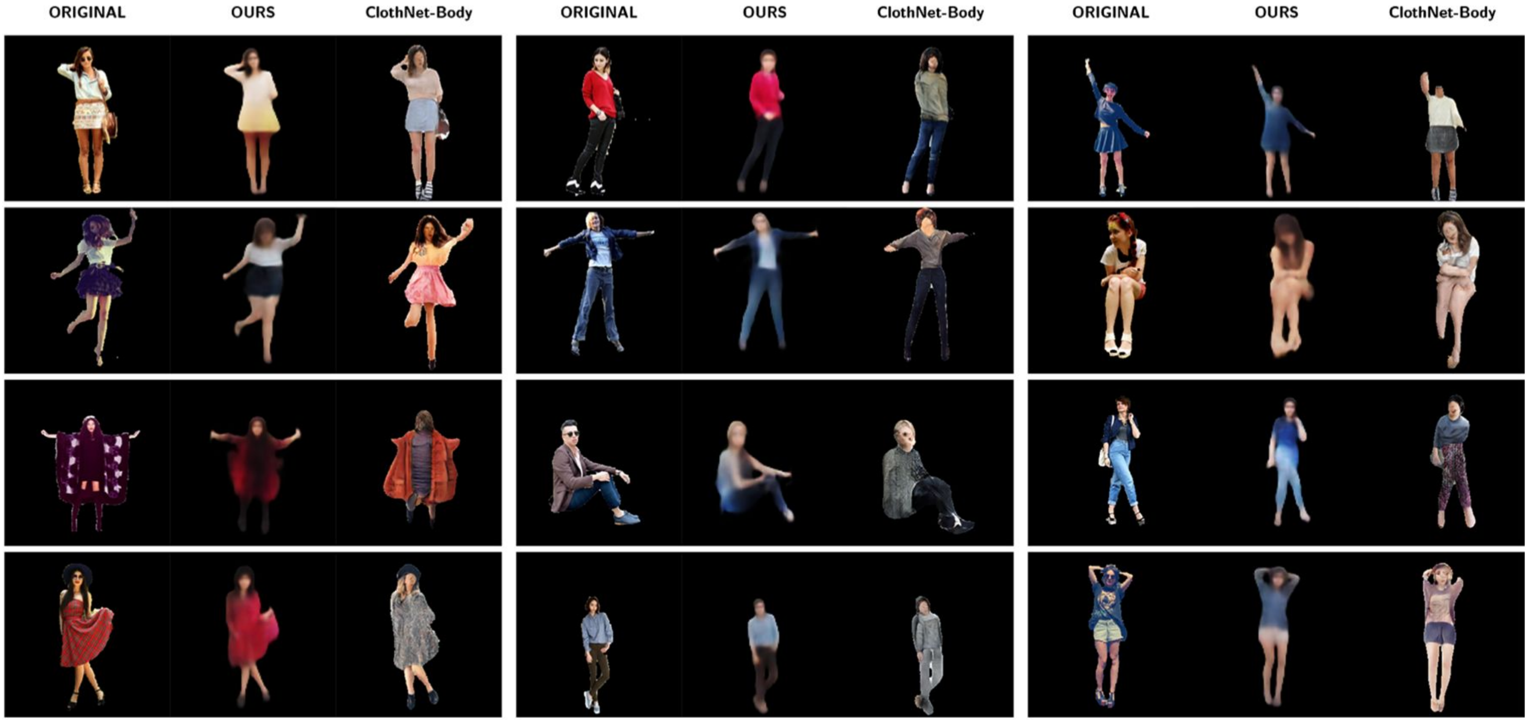
Metrics

- Image quality of Reconstructions
  - PSNR and SSIM
  - The higher, the better

- Accuracy of reconstructed poses
  - Extract pose from reconstructed image, and compare it to the ground truth pose
  - PCK. The higher, the better.

- Accuracy of pose estimation (Semi-DGPose)
  - PCK

Results of Conditional-DGPose



Results of Conditional-DGPose

Table 2 Image quality on ChictopiaPlus

	PSNR	SSIM
Conditional-DGPose	**21.33**	**0.88**
ClothNet-body (Lassner et al. 2017)	16.89	0.82

Best result is shown in bold

Quantitative evaluation w.r.t. image quality, showing that our method outperforms (Lassner et al. 2017) considering both metrics, the PSNR and the SSIM

Image Quality

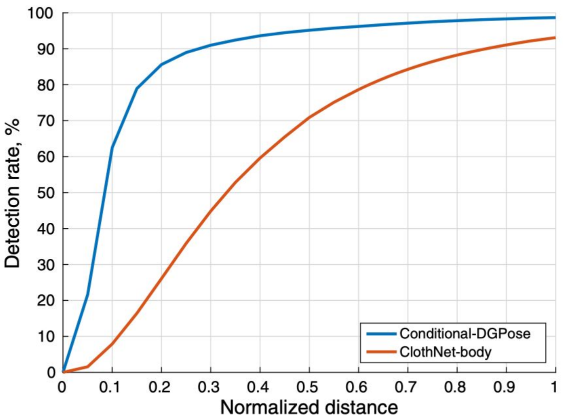


Fig. 20 Accuracy of Poses on ChictopiaPlus. The PCK scores over reconstructed images of our Conditional-DGPose (blue) significantly outperforms the ClothNet-body (Lassner et al. 2017) (red). Detection rate represents the percentage of joints correctly relocated in the reconstructions (Color figure online)

Accuracy of reconstructed pose

Results of Conditional-DGPose

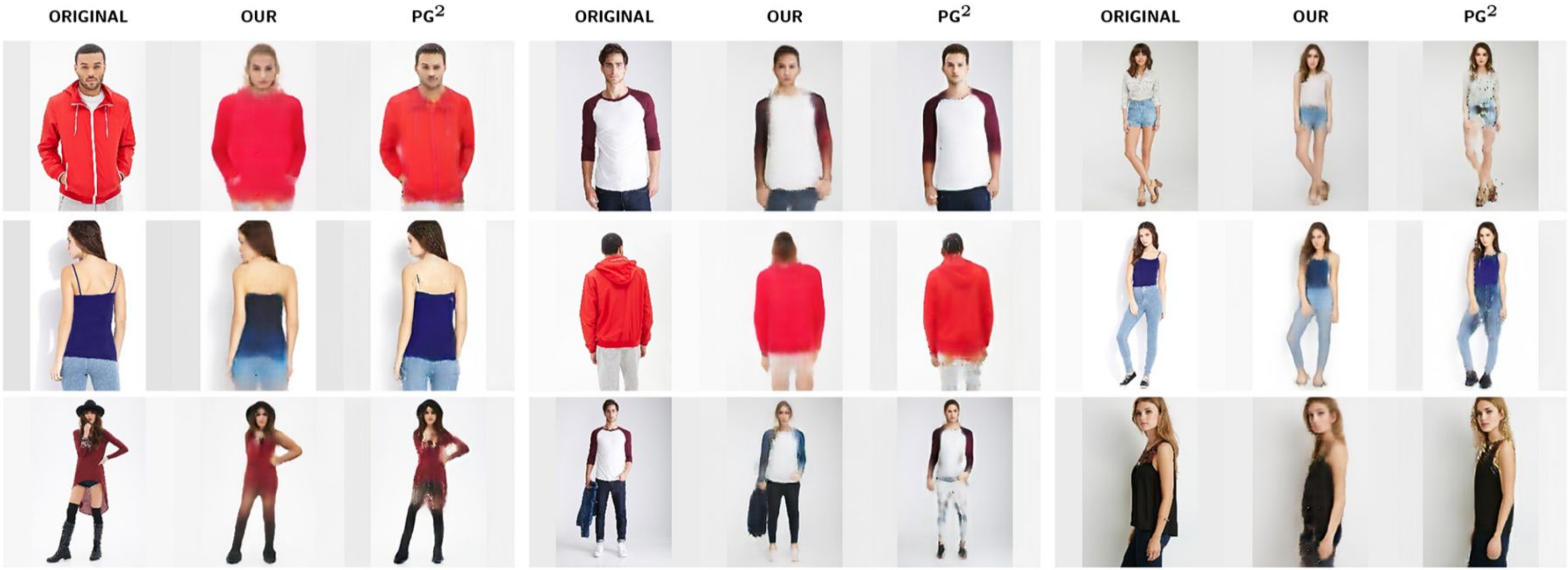


Image reconstruction

Results of Conditional-DGPose

Table 3 Image quality on DeepFashion

	PSNR	SSIM
Conditional-DGPose	18.38	0.79
PG^2 (Ma et al. 2017)	**18.96**	**0.83**

Best result is shown in bold
Quantitative evaluation w.r.t. image quality, showing that our method presents a performance only slightly below the baseline (Ma et al. 2017), considering both metrics, the PSNR and the SSIM, despite the fact it tackles a significantly more complex task than image-to-image translation

Image Quality

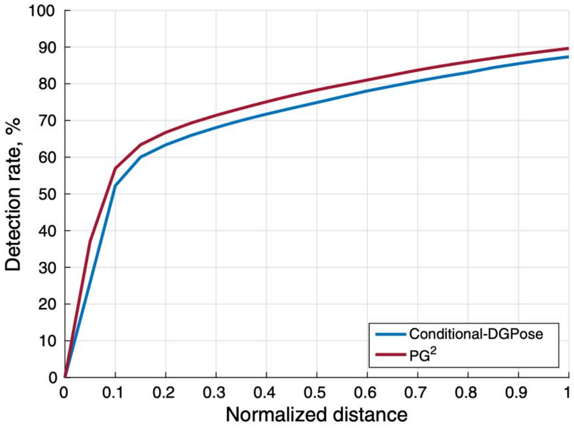


Fig. 21 Accuracy of Poses on DeepFashion. The PCK scores over

Accuracy of reconstructed pose

Results of Semi-DGPose

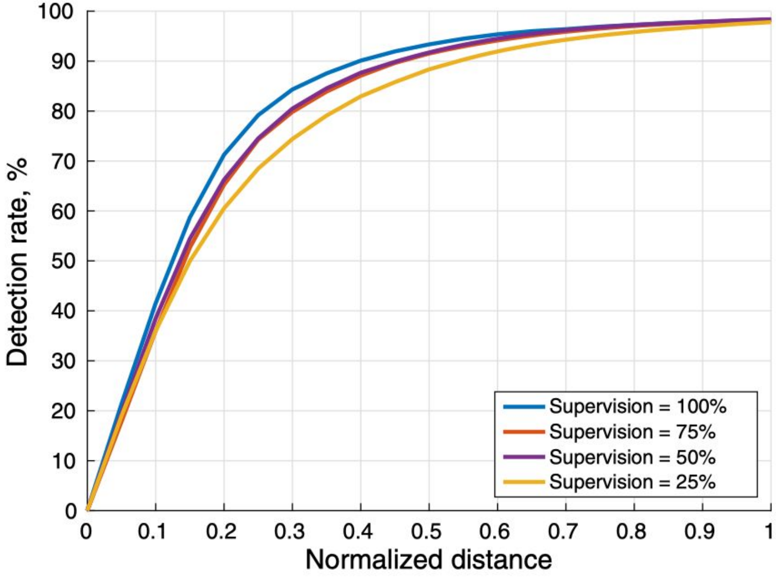
Different level of supervision

Table 4 Image quality on Human3.6M

Level of supervision	PSNR	SSIM
100%	22.27	0.89
75%	21.49	0.87
50%	21.36	0.86
25%	20.06	0.83

Quantitative evaluations of the Semi-DGPose with different levels of supervision using the PSNR and SSIM metrics

Image Quality



Pose estimation accuracy

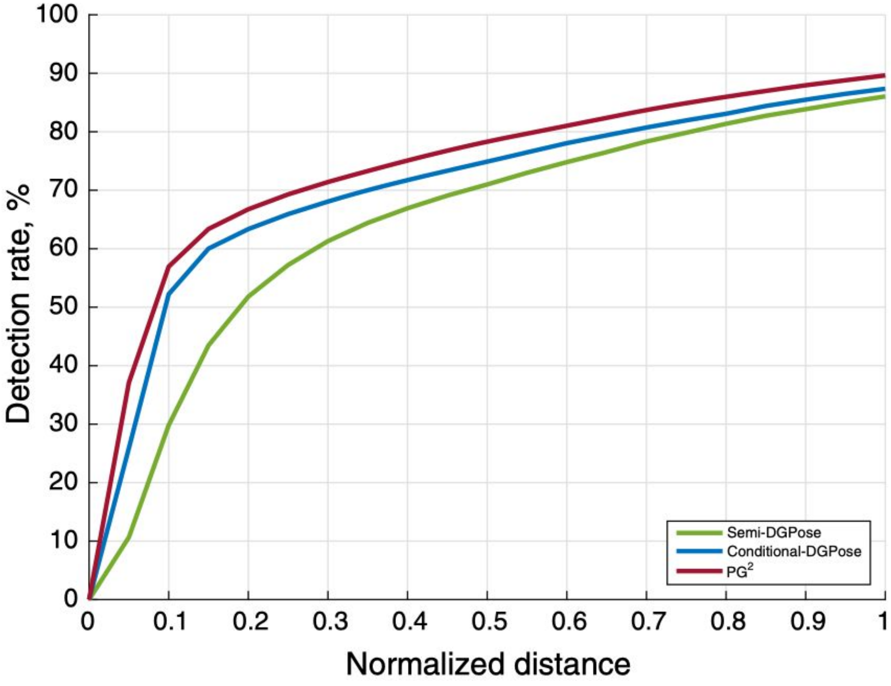
Different methods

Table 5 Image quality on DeepFashion

	PSNR	SSIM
Semi-DGPose	16.84	0.76
Conditional-DGPose	18.38	0.79
PG^2 (Ma et al. 2017)	**18.96**	**0.83**

Best result is shown in bold

Image Quality



Accuracy of pose reconstruction

Thanks!