Learning from Synthetic Humans

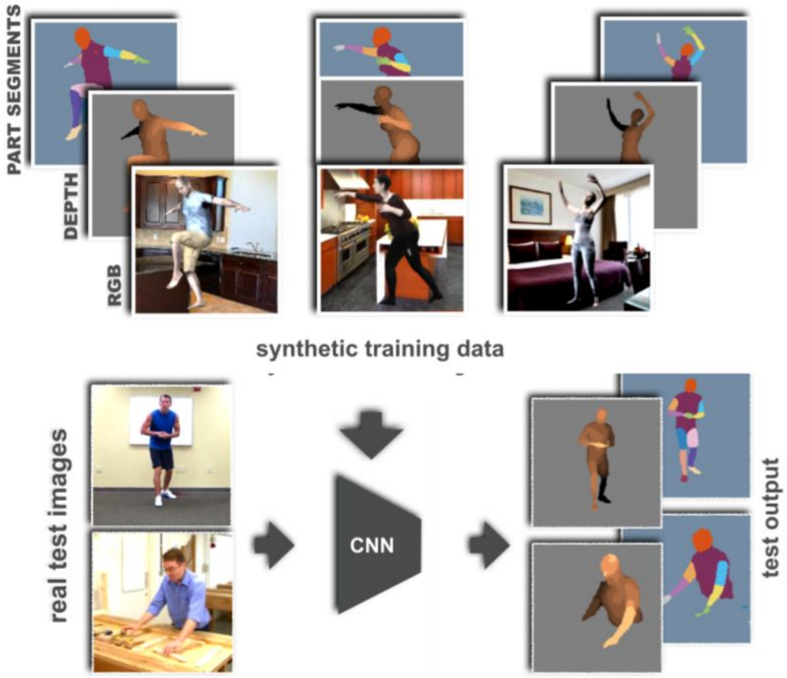
[1]

Gül Varol, Javier Romero, Xavier Martin, Naureen Mahmood,
Michael J. Black, Ivan Laptev, Cordelia Schmid

Presented by Taylor Kessler Faulkner

Motivation

- CNNs can effectively learn 2D human poses
- Labeled real human data is expensive and difficult in large amounts
- Goal: create synthetic data that is not hand-annotated



[1]

Goals

- Create a realistic synthetic dataset (SURREAL)
- Test whether a CNN can learn from SURREAL
  - Depth
  - Human parts segmentation
- Large synthetic person dataset with depth, segmentation, and ground truth

[1]

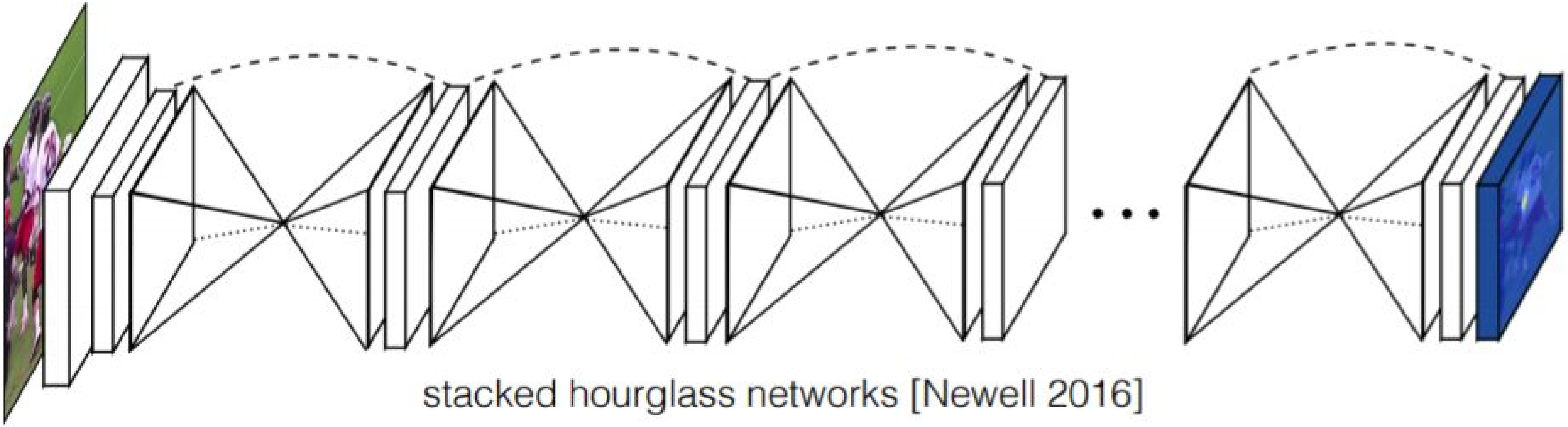
SURREAL Creation

- Body model: SMPL
- Body shape, texture: CAESAR
- Body pose: CMU MoCap marker data
- Background: LSUN
- Ground truth: Blender
- Random: 3D pose, shape, texture, viewpoint, lighting, background image

[1]

Network

- Adapted from 2D pose estimation



stacked hourglass networks [Newell 2016]

- Models spatial relations at different resolutions [1]
- Uses human body structure to obtain pixel-wise output [1]

Depth and Segmentation

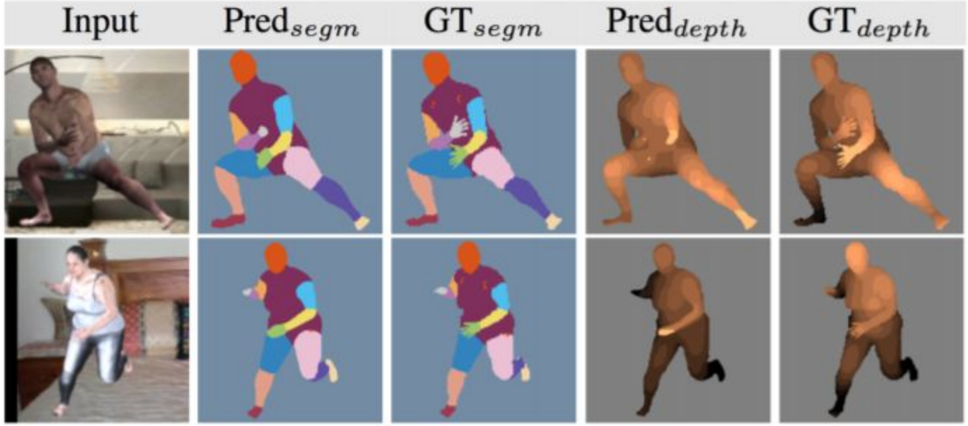
- Pixel-wise classification
- Segmentation: each pixel is classified
  - Head, torso, upper legs, lower legs, upper arms, lower arms, hands, feet, background
- Depth: Pelvis set as center
  - 9 depth levels in front, 9 levels behind

[1]

Experimental Evaluation

- Segmentation evaluation
  - Intersection over union (IOU)
  - Pixel accuracy measures
- Depth estimation evaluation
  - Classification problem, but continuous values
  - Root-mean-squared-error (RMSE) b/w predicted and ground truth depth

Experiments - SURREAL Dataset

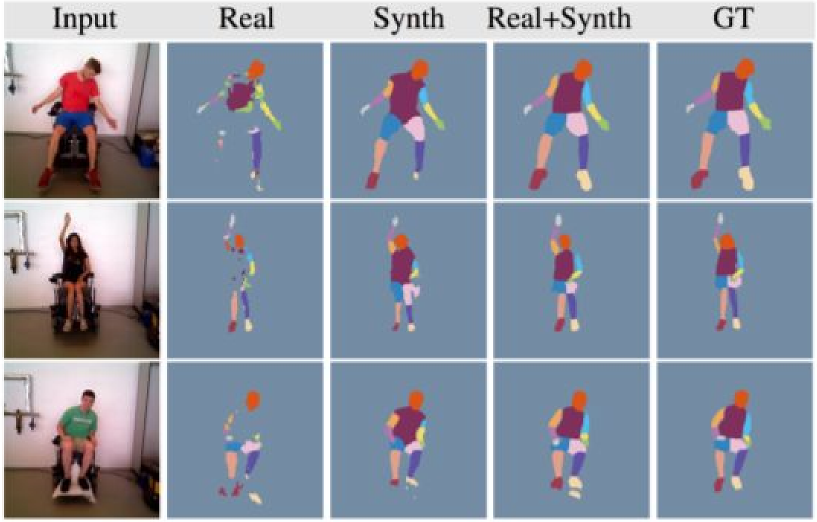


Segmentation	
IOU	69.13 %
Accuracy	80.61 %

Depth	
RMSE	72.9 mm
st-RMSE	56.3 mm

[2] Slide taken from authors' presentation

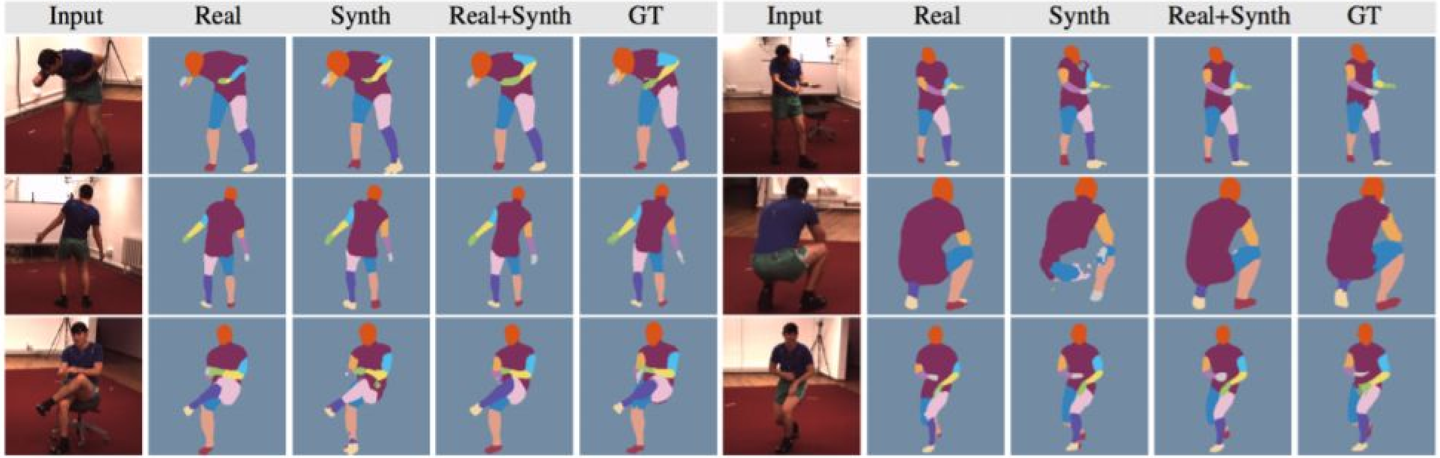
Experiments - Freiburg Sitting People Dataset



Training data	Head IOU	Torso IOU	$Legs_{up}$ IOU	mean IOU	mean Acc.
Real+Pascal[21]	-	-	-	64.10	81.78
Real	58.44	24.92	30.15	28.77	38.02
Synth	73.20	65.55	39.41	40.10	51.88
Synth+Real	72.88	80.76	65.41	59.58	78.14
Synth+Real+up	**85.09**	**87.91**	**77.00**	**68.84**	**83.37**

[2] Slide taken from authors' presentation

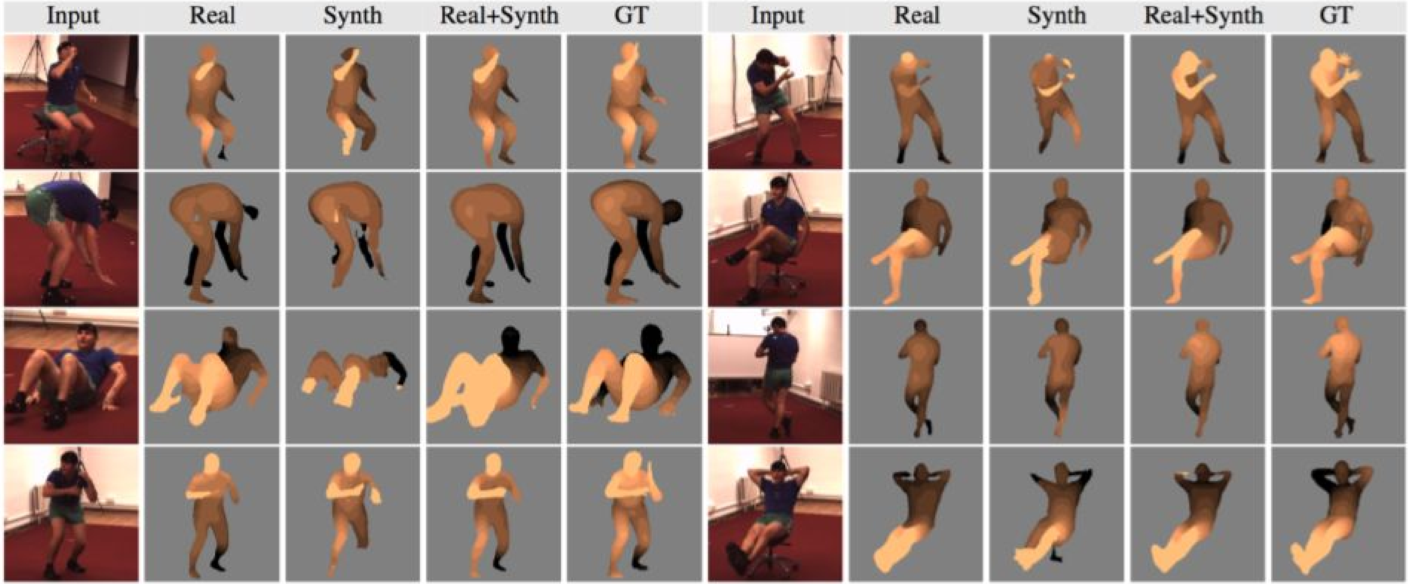
Experiments - Human3.6M Dataset



Training data	IOU fg+bg	IOU fg	Accuracy fg+bg	Accuracy fg
Real	49.61	46.32	58.54	55.69
Synthetic	46.35	42.91	56.51	53.55
Synthetic+Real	57.07	54.30	67.72	65.53

[2] Slide taken from authors' presentation

Experiments - Human3.6M Dataset



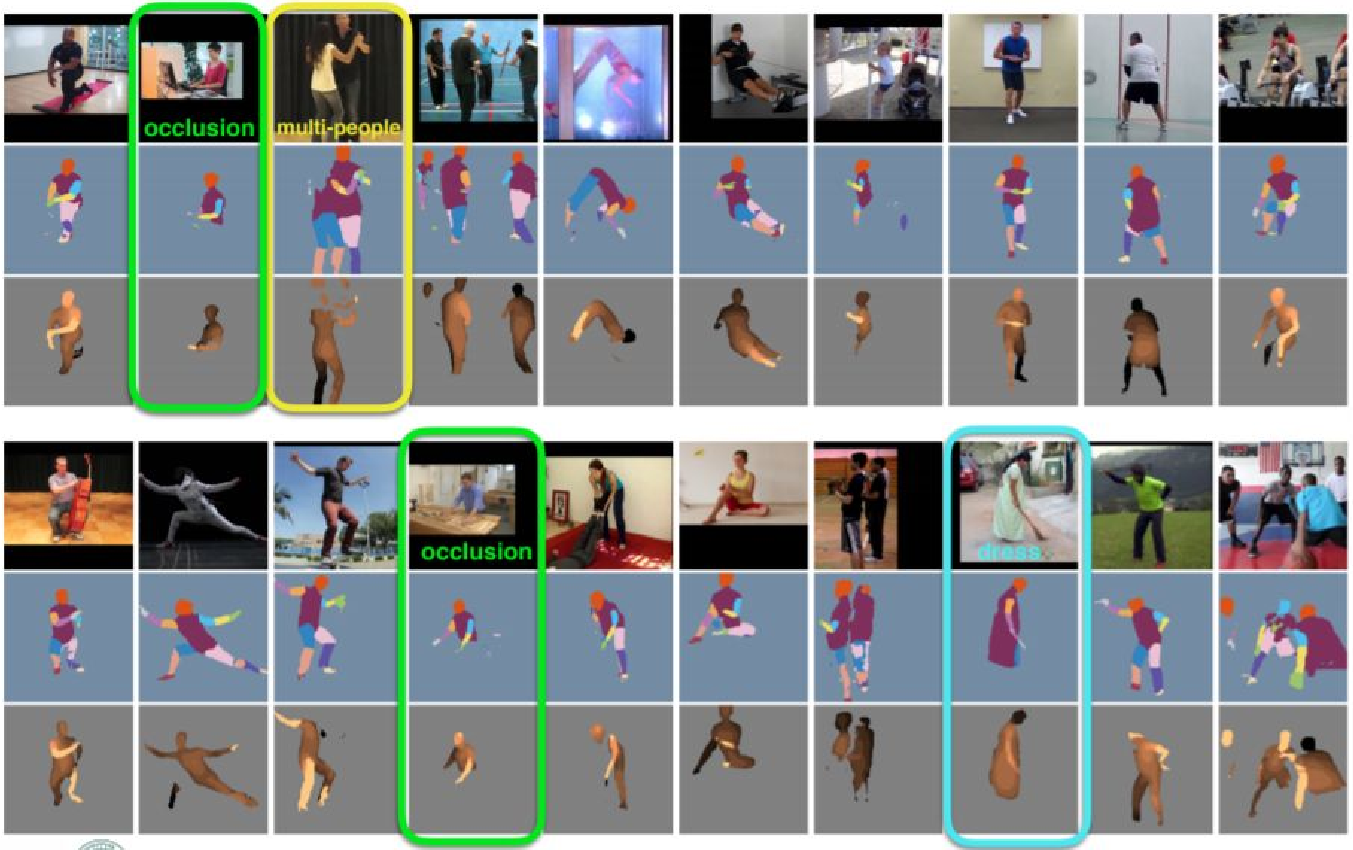
Training data	RMSE	st-RMSE	PoseRMSE	st-PoseRMSE
Real	96.3	75.2	122.6	94.5
Synthetic	111.6	98.1	152.5	131.5
Synthetic+Real	90.0	67.1	92.9	82.8

(mm)

Varol et al. Learning from Synthetic Humans, CVPR'17.

[2] Slide taken from authors' presentation

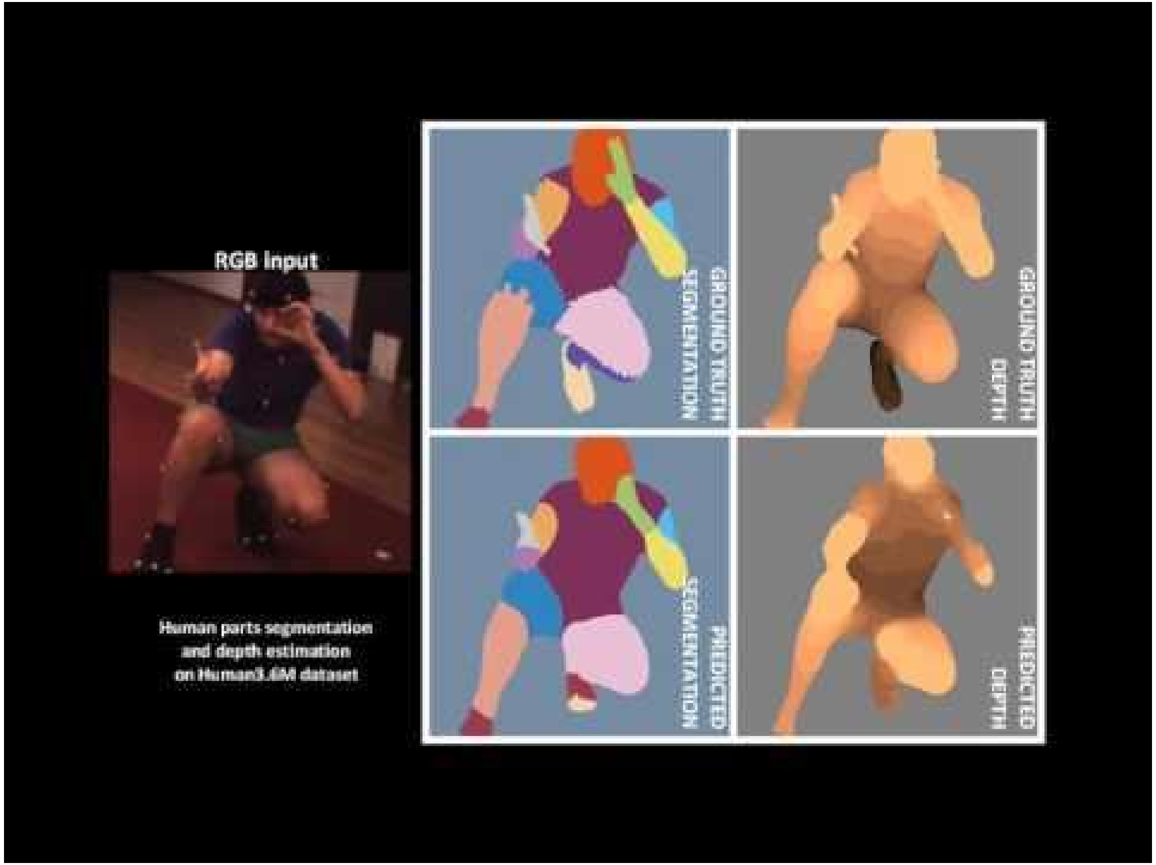
Experiments - MPII Human Pose Dataset



Varol et al. Learning from Synthetic Humans, CVPR'17.

[2] Slide taken from authors' presentation

Video


RGB Input
Human parts segmentation
and depth estimation
on Human3.6M dataset
GROUND TRUTH SEGMENTATION
GROUND TRUTH DEPTH
PREDICTED SEGMENTATION
PREDICTED DEPTH

[4]

Strengths and Weaknesses

- Easy to create realistic synthetic images
- Provides a good pre-training dataset for real data
- Backgrounds are unrealistic
  - No interaction with lighting
  - Human movement around objects in background is wrong
- Groups of people cause problems, so we can only test on single humans

Extensions

- Addition of occlusions and groups of people in dataset

- Better interactions with background image
    - Also provides occlusion data (objects in background)

Citations

[1] Learning from Synthetic Humans. G. Varol, J. Romero, X. Martin, N. Mahmood, M. Black, I. Laptev, C. Schmid. CVPR 2017.

[2] G. Varol, J. Romero, X. Martin, N. Mahmood, M. Black, I. Laptev and C. Schmid, "Learning from Synthetic Humans", 2017. http://www.di.ens.fr/willow/research/surreal/varol_cvpr17_presentation.pdf

[3] G. Varol, J. Romero, X. Martin, N. Mahmood, M. Black, I. Laptev and C. Schmid, *[CVPR'17] SURREAL dataset - Learning from Synthetic Humans*. 2017.

[4] G. Varol, J. Romero, X. Martin, N. Mahmood, M. Black, I. Laptev and C. Schmid, [CVPR'17] SURREAL synthetic training results on Human3.6M. 2017.

[5] G. Varol, J. Romero, X. Martin, N. Mahmood, M. Black, I. Laptev and C. Schmid, [CVPR'17] SURREAL synthetic training results on Youtube Pose


Output: Segmentation
Output: Depth
Output: Segmentation
Output: Depth
Output: Segmentation
Output: Depth

[5]