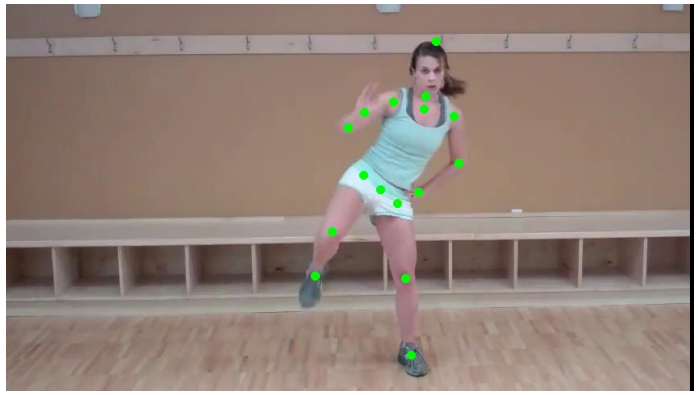
Single-Person 2D Joint Estimation

- Neural Network: Hourglass model
- Two different datasets: MPII and Unite the People
- Comparison of model performance on the two datasets
- GPU: GTX 1080 8GB
- Training time: 30 hours per Dataset
- Keras + Tensorflow backend

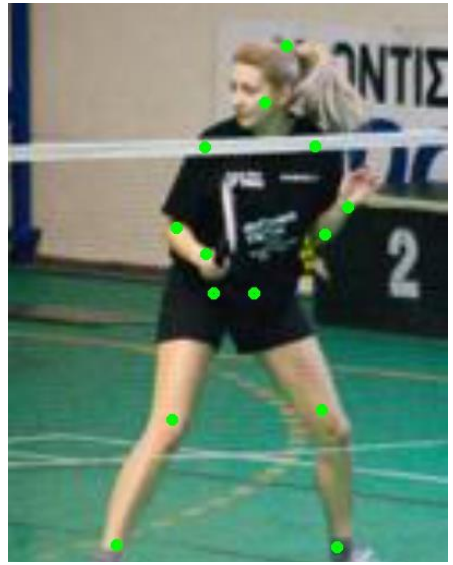
MPII Dataset

- Single- and multi-person
- 25k total photos containing over 40k individuals
- 55k training images and 3k test images
- 16 joints per individual (one heatmap per joint)
- If no available information on joint: (0,0)

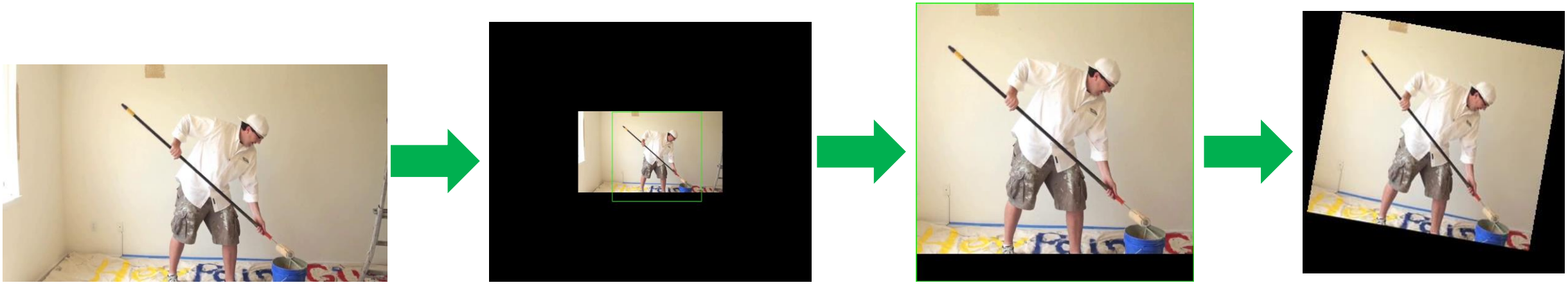


Unite the People Dataset

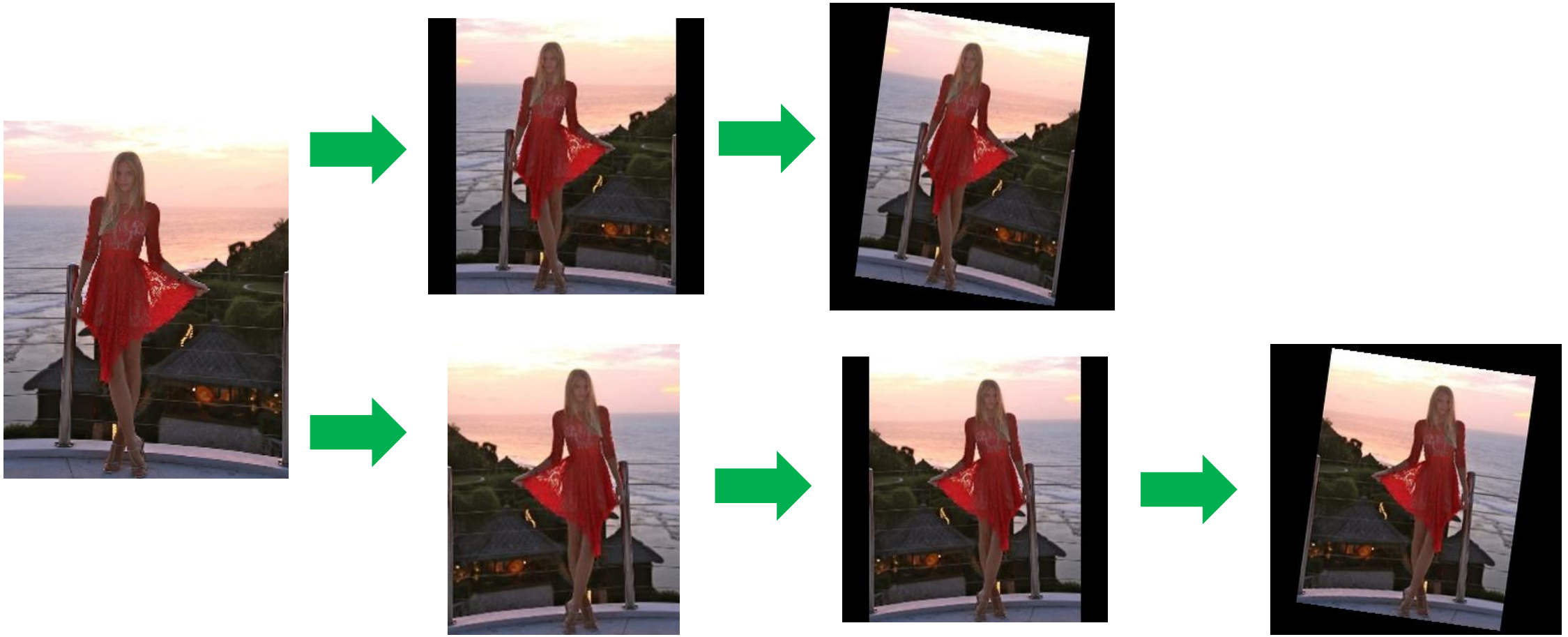
- Single- and multi-person dataset
- 8k+ photos
- 45k training images and 1.6k test images
- 14 joints per individual (one heatmap per joint)
- If no available information on joint: (0,0)



MPII Dataset: Data Augmentation

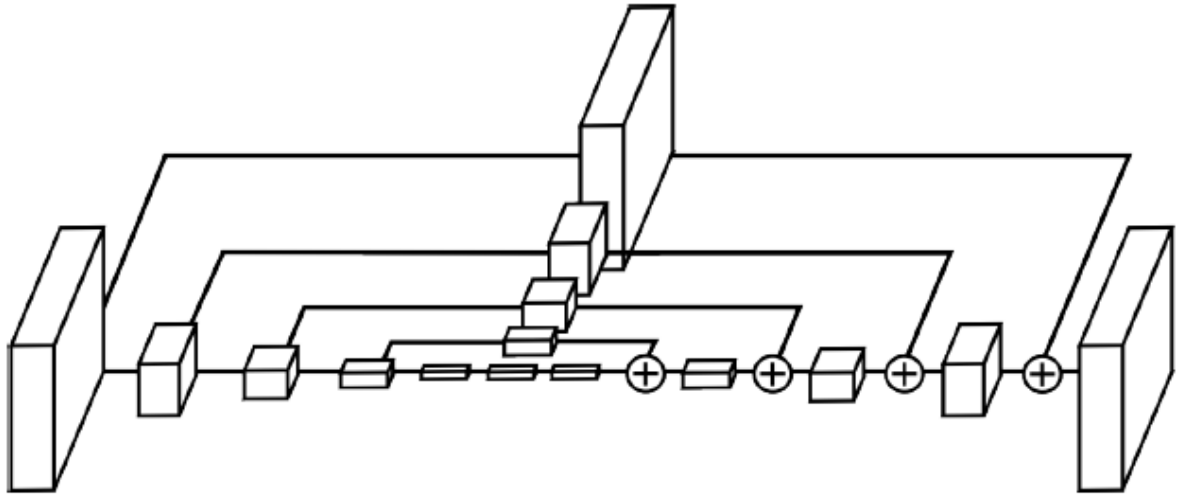


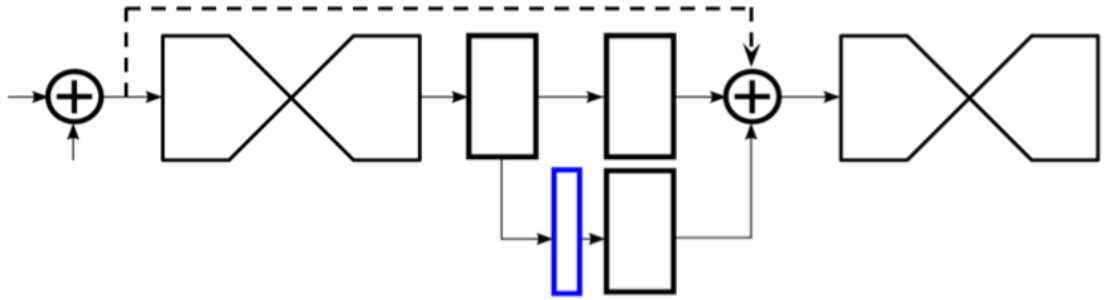
Unite the People Dataset: Data Augmentation



Hourglass Model

- The upper branches gradually extract deeper features first in their original size and then by halving the image size at each level (from top to bottom)
- The Stacked Hourglass structure is obtained by stacking multiple hourglasses end-to-end, feeding the output of one as input into the next

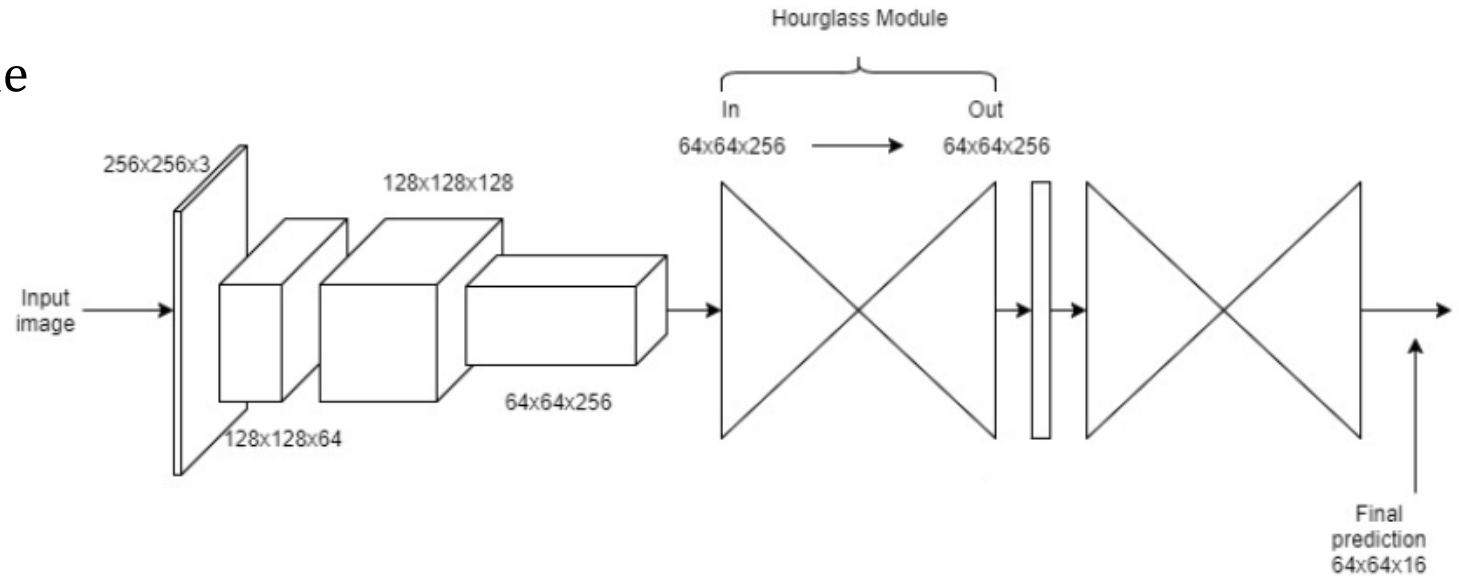




Stacked Hourglass Model

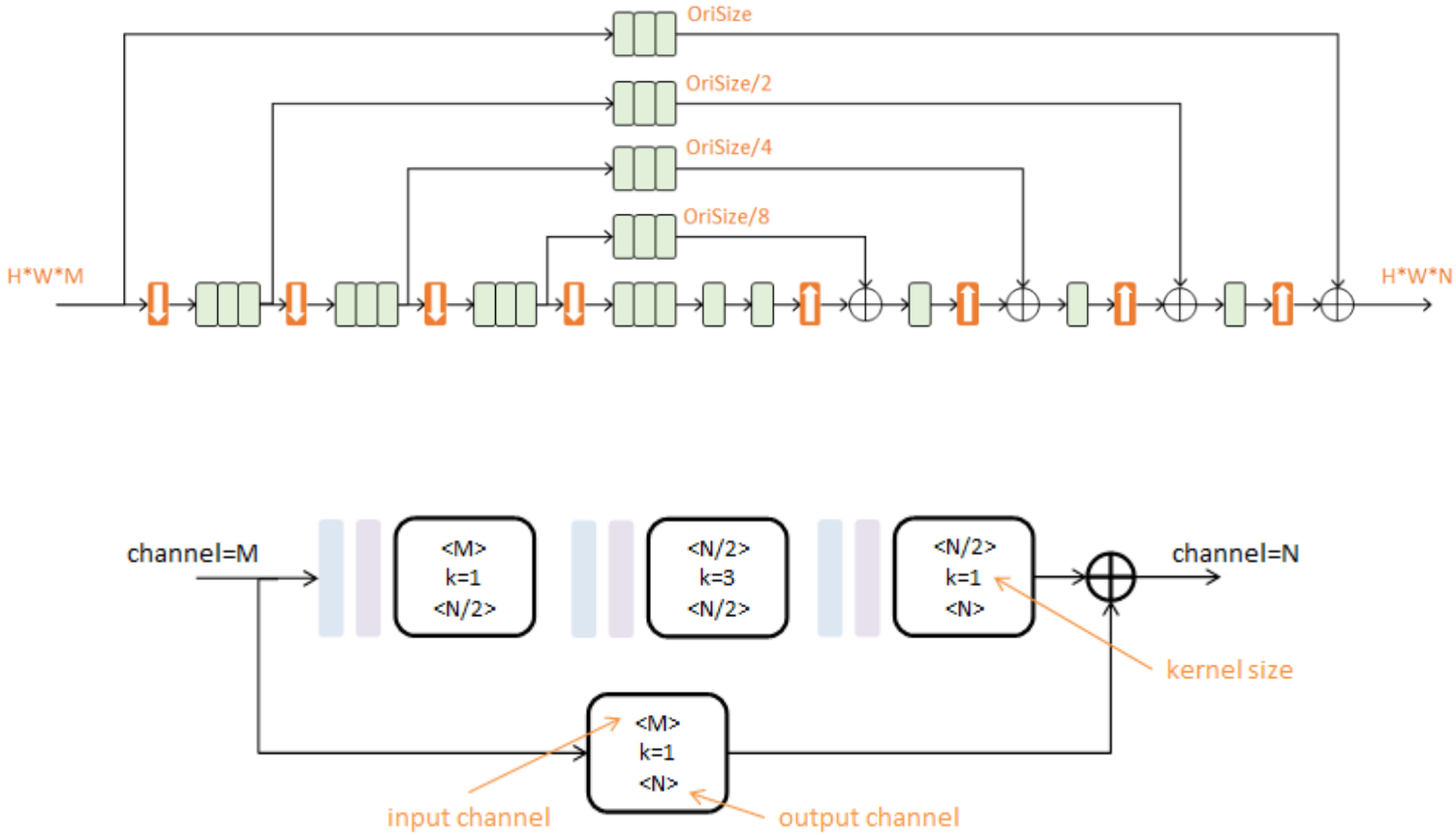
- Additional modules are present before the hourglass and allow the model to segment the pictures using convolution layers

- Ground truths and outputs are heatmap arrays (64 x 64) with:
  - One heatmap per joint
  - Ground truth heatmaps have small gaussian peaks around correct joint position



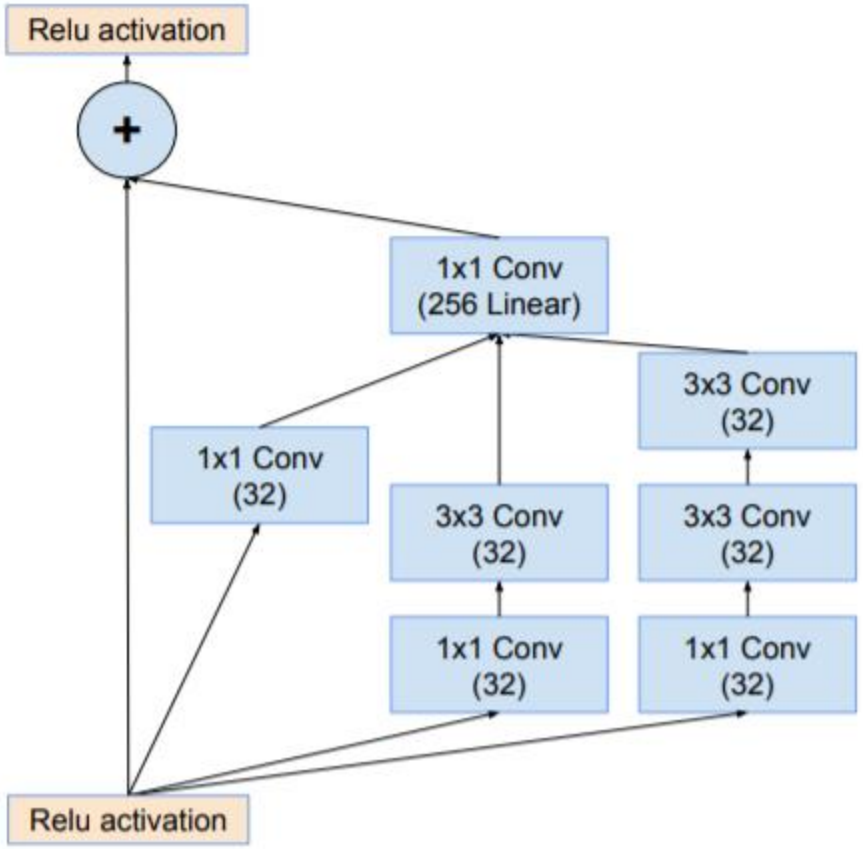
Residual Blocks

- The Hourglass model is comprised of residual blocks (green blocks)
- The blue rectangles are Batch Normalizations and the purple are ReLU activations

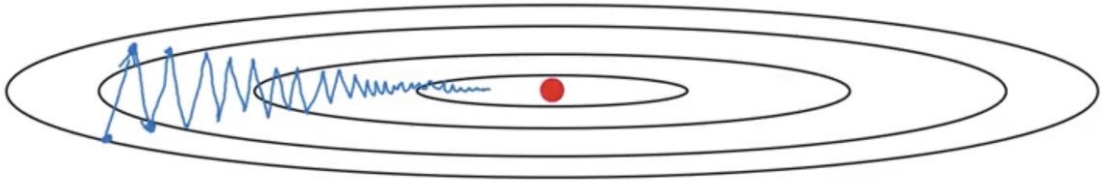


The model

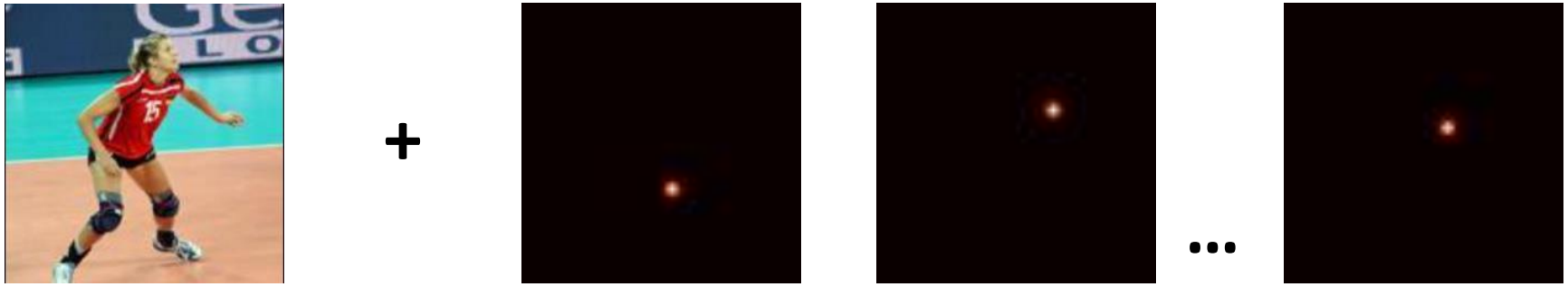
- Two Hourglasses at level 4
- Residual blocks: Inception-ResNet-A module of Inception-ResNet-v1 network



Training details (1)



- RMSProp optimisation algorithm
- Model Input:
  - Image (256x256) + 16 heatmaps (64x64)

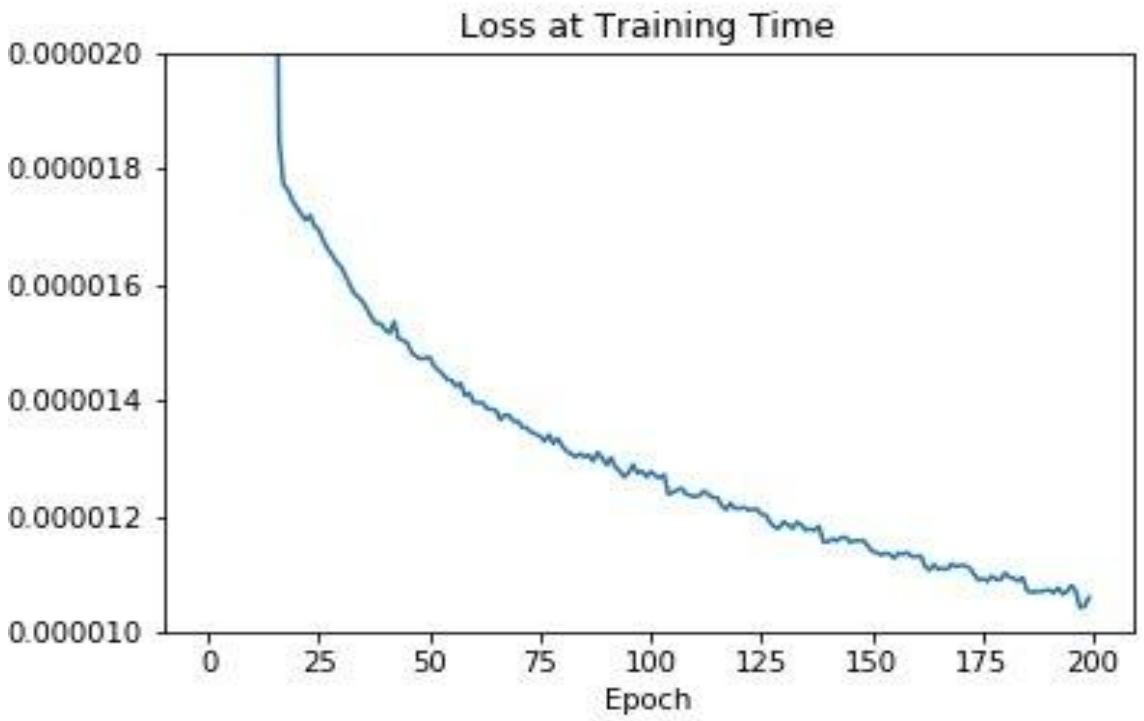


- Model Output:
  - 16 predictions (64x64)

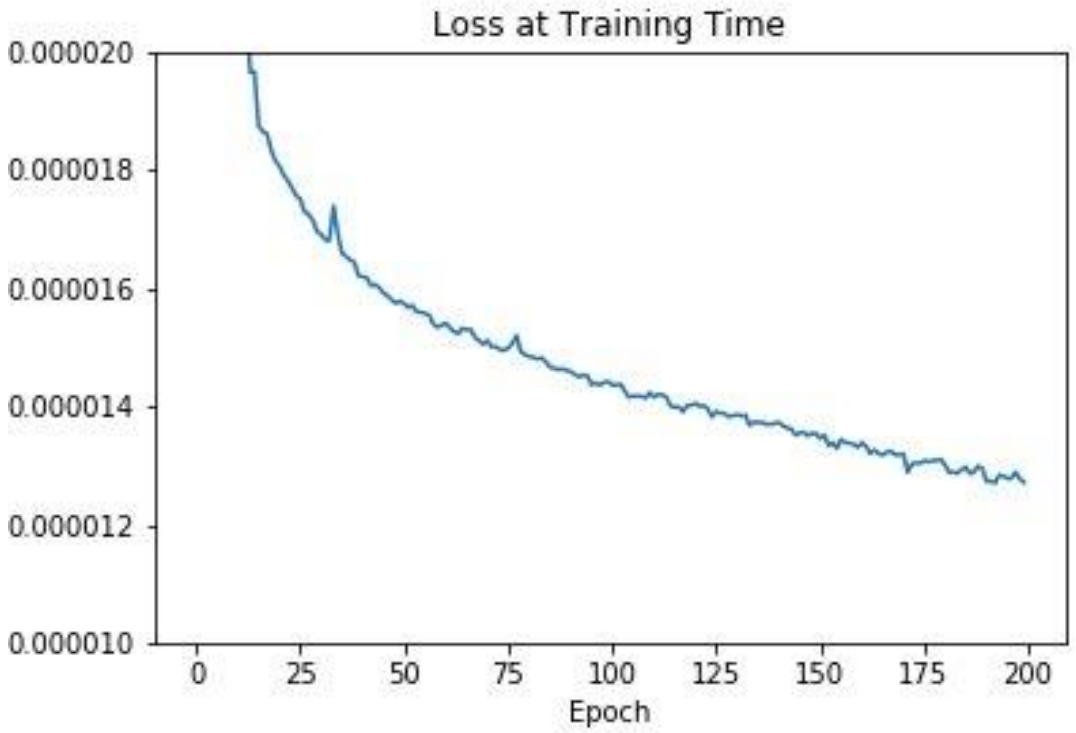
Training details (2)

- Epochs : 200
- Step size : 800

MPII Dataset



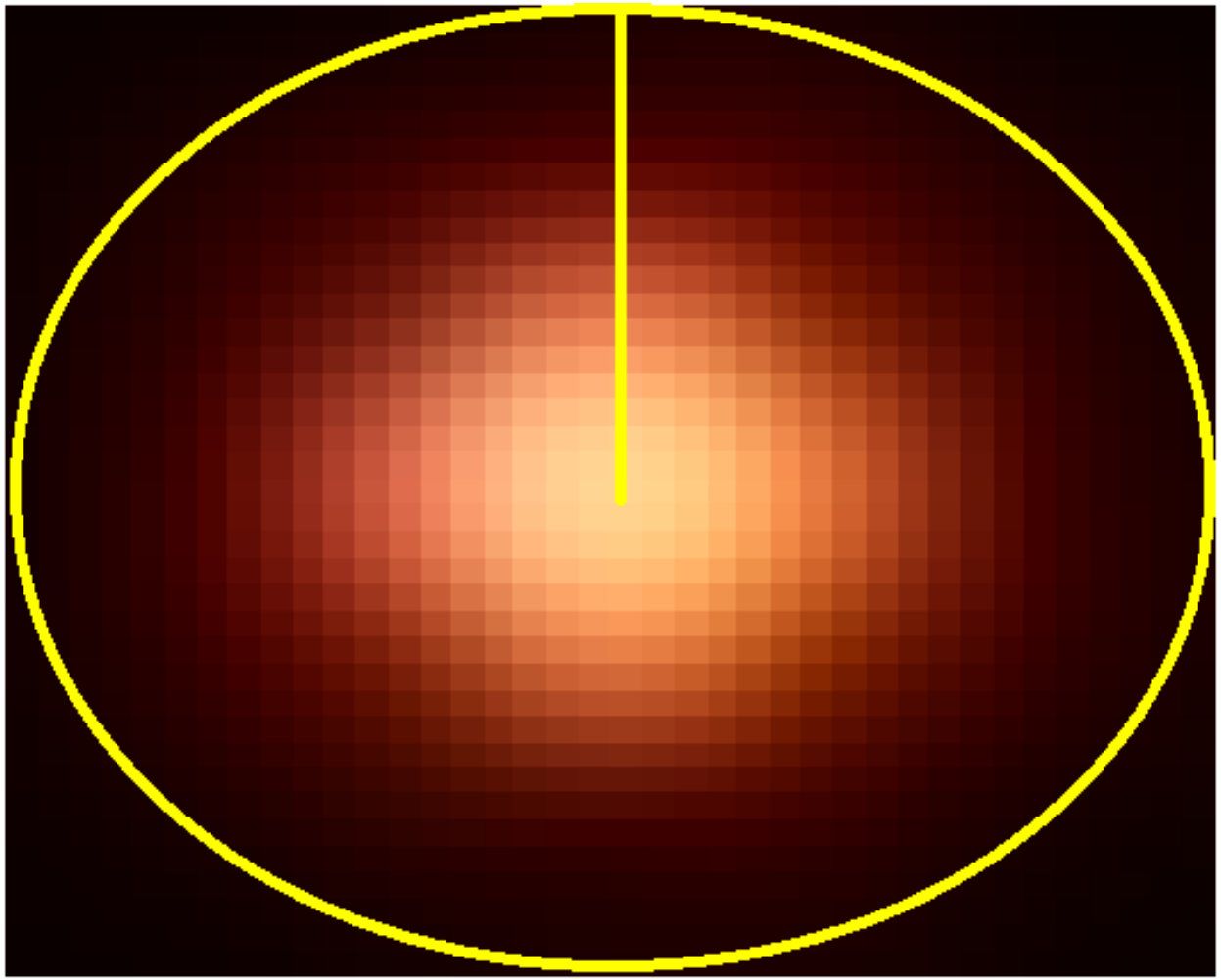
UP14



Evaluation (PCK)

Original PCK: Percentage of true predicted key points – calculated using as threshold half the size of the head in the ground truth

Custom PCK: Percentage of true predicted key points – calculated using as threshold a circle of radius 4 around the true joint in the ground truth



Results MPII-dataset (PCK)

Results on the Train Set

Right Ankle	Right Knee	Right Hip	Left Hip	Left Knee	Left Ankle	Pelvis	Thorax	Upper Neck	Head Top	Right Wrist	Right Elbow	Right Shoulder	Left Shoulder	Left Elbow	Left Wrist
56%	72%	76%	76%	72%	67%	80%	89%	88%	86%	61%	72%	84%	85%	73%	61%

Results on the Test Set

Right Ankle	Right Knee	Right Hip	Left Hip	Left Knee	Left Ankle	Pelvis	Thorax	Upper Neck	Head Top	Right Wrist	Right Elbow	Right Shoulder	Left Shoulder	Left Elbow	Left Wrist
54%	57%	65%	65%	66%	63%	72%	84%	84%	82%	56%	65%	77%	79%	68%	56%

Average PCK

Train	Test	Random Choice
78%	74%	~2%

Results UP14-dataset (PCK)

Results on the Train Set

Right Ankle	Right Knee	Right Hip	Left Hip	Left Knee	Left Ankle	Right Wrist	Right Elbow	Right Shoulder	Left Shoulder	Left Elbow	Left Wrist	Upper Neck	Head Top
69%	64%	88%	88%	66%	69%	68%	73%	79%	76%	73%	73%	95%	93%

Results on the Test Set

Right Ankle	Right Knee	Right Hip	Left Hip	Left Knee	Left Ankle	Right Wrist	Right Elbow	Right Shoulder	Left Shoulder	Left Elbow	Left Wrist	Upper Neck	Head Top
57%	59%	71%	76%	61%	65%	62%	65%	70%	68%	64%	69%	91%	89%

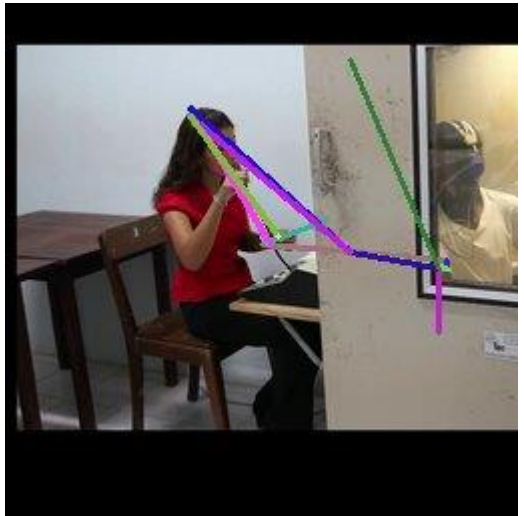
Average PCK

Train	Test	Random Choice
78%	71%	~2%

Results (Visualization)

MPII dataset



UP14 dataset




Limitations and Further Work

- Difficulties faced:
  - Joint estimation=point estimation (in a radius of 4 pixels) is a complex problem
  - Individuals in the dataset have several different (distorted) poses
  - Hardware limitations, both in terms of GPU power and memory
  - Making the model more complex with Inception-Residual module
- Room for improvement regarding:
  - Higher model hyper parameters → longer training time
  - More complex model → more hourglasses and layers
  - Trying other residual module types
  - Occlusion-robust model

https://youtu.be/pW6nZXeWlGM?t=36

Resources

- Alejandro Newell, Kaiyu Yang, and Jia Deng, Stacked Hourglass Networks for Human Pose Estimation, arXiv:1603.06937, 2016.
- http://publications.lib.chalmers.se/records/fulltext/253624/253624.pdf
- http://human-pose.mpi-inf.mpg.de/
- http://files.is.tuebingen.mpg.de/classner/up/
- https://github.com/yuanyuanli85/Stacked_Hourglass_Network_Keras
- https://github.com/wbenbihi/hourglasstensorlfow
- https://keras.io/