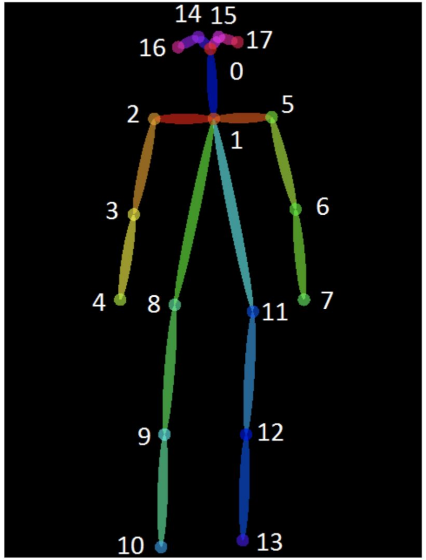
Body Pose estimation

Michael Maldonado

Understanding Body Pose Estimation

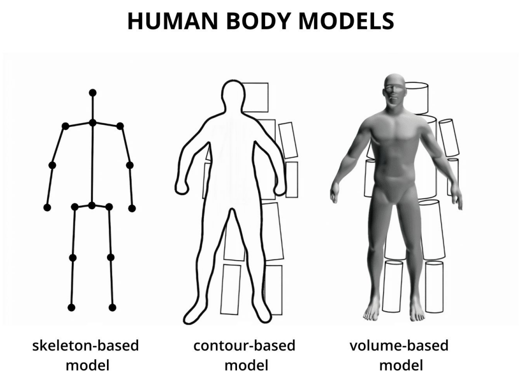
What is Human Body Pose Estimation?

- Key Problem in Computer Vision.
- Goal is to detect position and orientation of body based on specific landmarks.
- Difficult and still largely unsolved



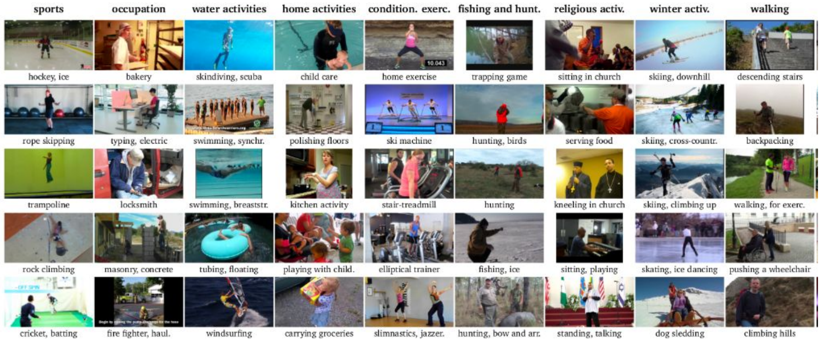
Types and Approaches

- 2D vs 3D pose estimation
- Models of representing body
  - Skeleton, Contour, and Volume based
- Bottom up vs Top Down approach



Datasets Used for Pose Estimation

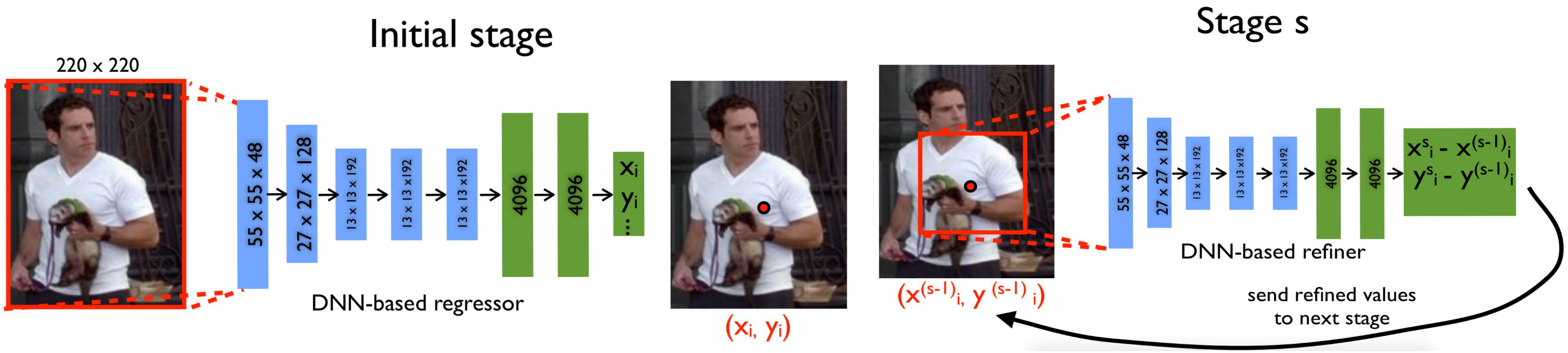
- Leeds Sports Pose Dataset
  - 1k testing/training images
  - Mainly poses from various sports
- MPII Human Pose Dataset
  - 18k training 7k testing images
  - 410 different human activities



Research Works in Body pose estimation

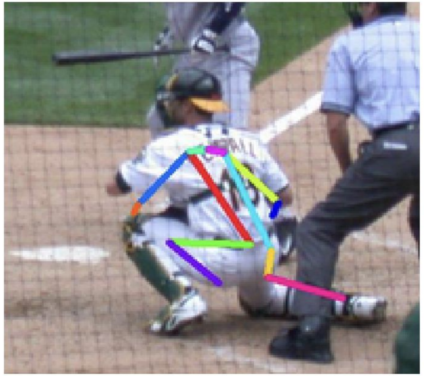
DeepPose: Human Pose Estimation via Deep Neural Networks

- One of the fundamental papers about Human Body Pose Estimation
- First major paper that applied DNNs to Pose estimation.
- Estimation is a joint regression problem.



Motivations for *DeepPose*

- Applying DNNs for object poses had not been done at that time.
- DNNs may provide holistic reasoning.
  - Variability in poses, occluded joints, etc.
- No need to implement complex detectors which limit expressiveness.




Results of *DeepPose*

- Achieved state-of-art or better results on several challenging datasets.

Method	Arm		Leg		Ave.
	Upper	Lower	Upper	Lower	
DeepPose	0.8	0.75	0.71	0.5	0.69
Pishchulin [17]	0.80	0.70	0.59	037	0.62
Johnson et al. [13]	0.75	0.67	0.67	0.46	0.64
Yang et al. [26]	0.69	0.64	0.55	0.35	0.56

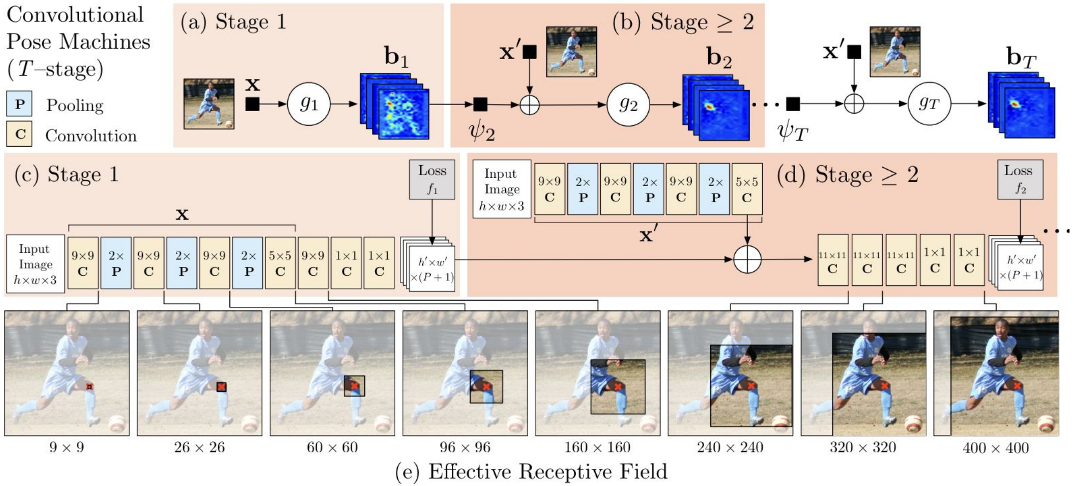
Convolutional Pose Machines

- Combines benefits from DNNs and Traditional pose machine architectures.

Input Image (a) Stage 1 (b) Stage 2 (c) Stage 3

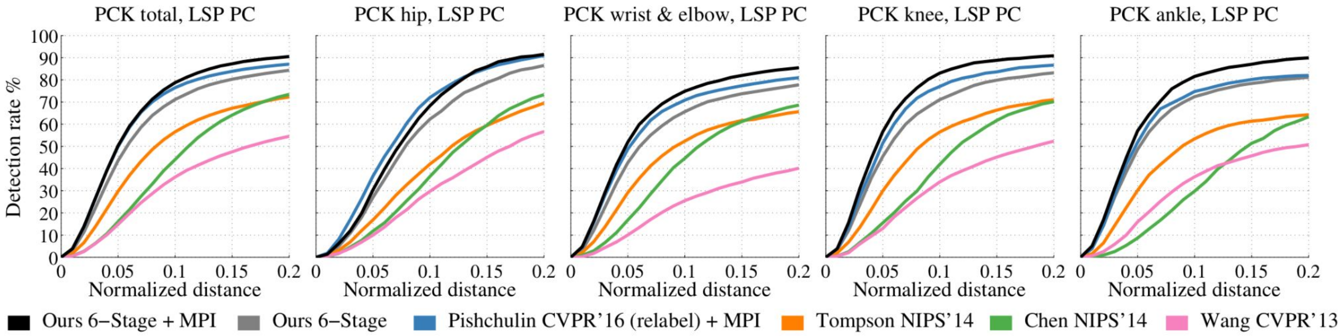
Motivations for *Convolutional Pose Machines*

- Main motivation is to learn long range spatial relationships.
- Address the vanishing gradient problem



Results of *Convolutional Pose Machines*

- Accuracy is significantly higher than other methods in challenging non frontal view of person
- Model does not need another module dedicated to location refinement for high-precision accuracy.



Stacked Hourglass Networks for Human Pose Estimation

- Introduced a new model architecture that pools and upscales input at multiple stages

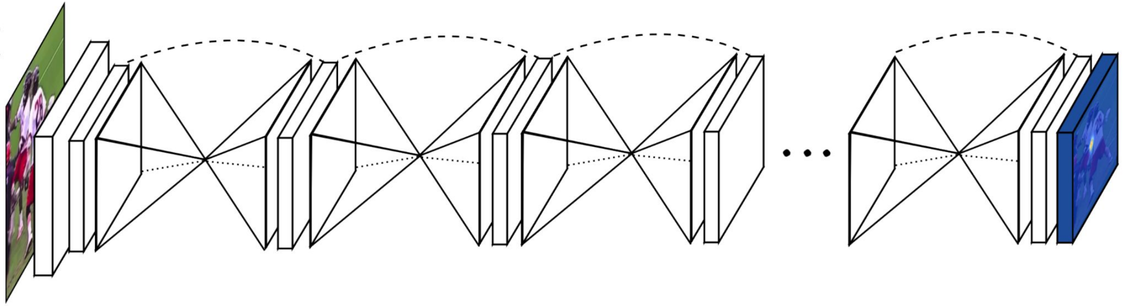
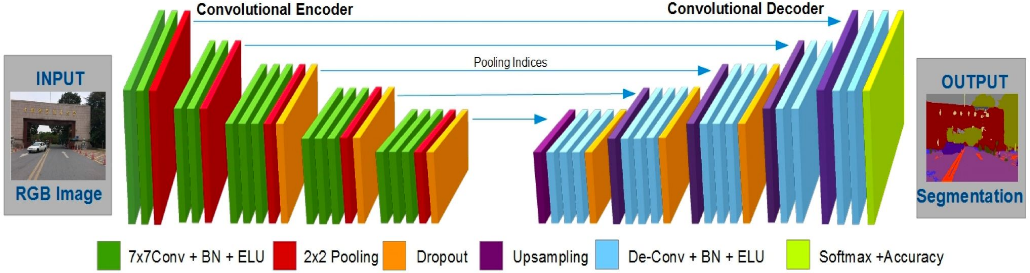
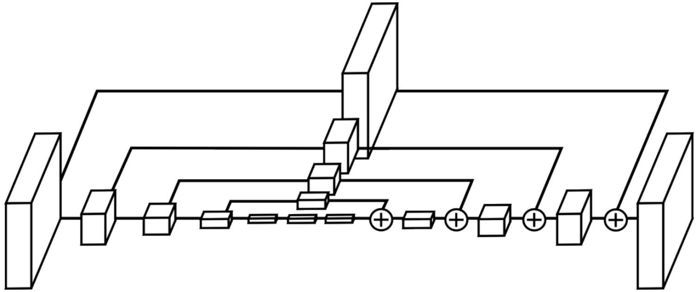


Fig. 1. Our network for pose estimation consists of multiple stacked hourglass modules which allow for repeated bottom-up, top-down inference.

Motivations for *Stacked Hourglass Networks for Human Pose Estimation*

- To be able to capture information at every scale of network.
- Local evidence is key for individual joints, final pose requires global overview





An illustration of a single "hourglass" module. Each box in the figure corre- to a residual module as seen in Figure 4. The number of features is consistent the whole hourglass.

Results for *Stacked Hourglass Networks for Human Pose Estimation*

- Achieves state-of-the-art results on the MPII dataset.

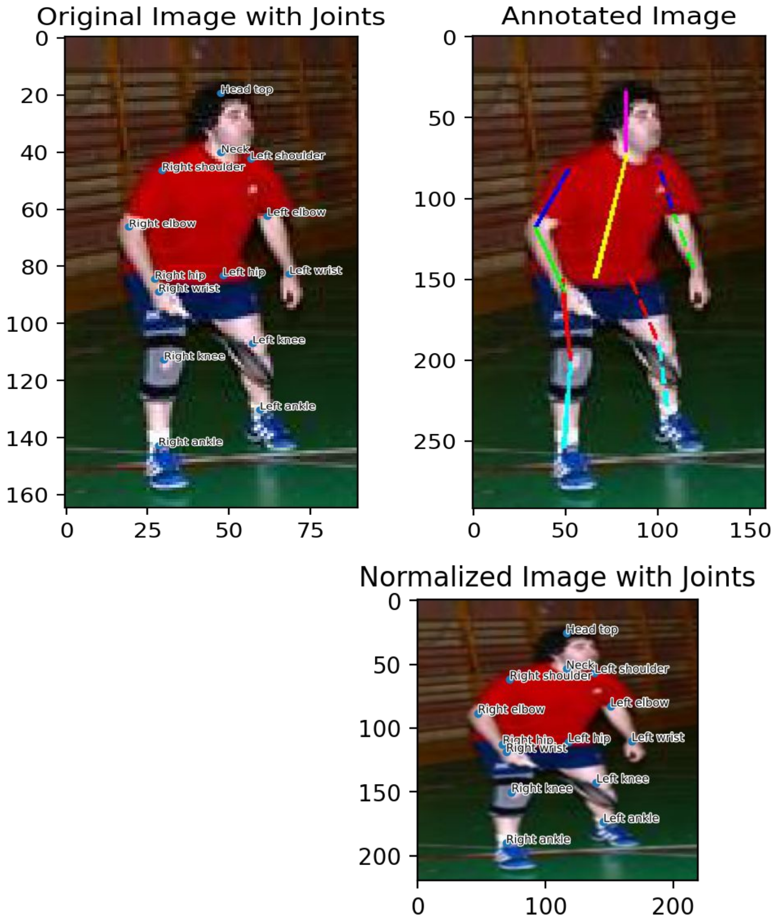
Fig. 7. PCKh comparison on MPII

	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Total
Tompson et al. [16], CVPR'15	96.1	91.9	83.9	77.8	80.9	72.3	64.8	82.0
Carreira et al. [19], CVPR'16	95.7	91.7	81.7	72.4	82.8	73.2	66.4	81.3
Pishchulin et al. [17], CVPR'16	94.1	90.2	83.4	77.3	82.6	75.7	68.6	82.4
Hu et al. [27], CVPR'16	95.0	91.6	83.0	76.6	81.9	74.5	69.5	82.4
Wei et al. [18], CVPR'16	97.8	95.0	88.7	84.0	88.4	82.8	79.4	88.5
Our model	**98.2**	**96.3**	**91.2**	**87.1**	**90.1**	**87.4**	**83.6**	**90.9**

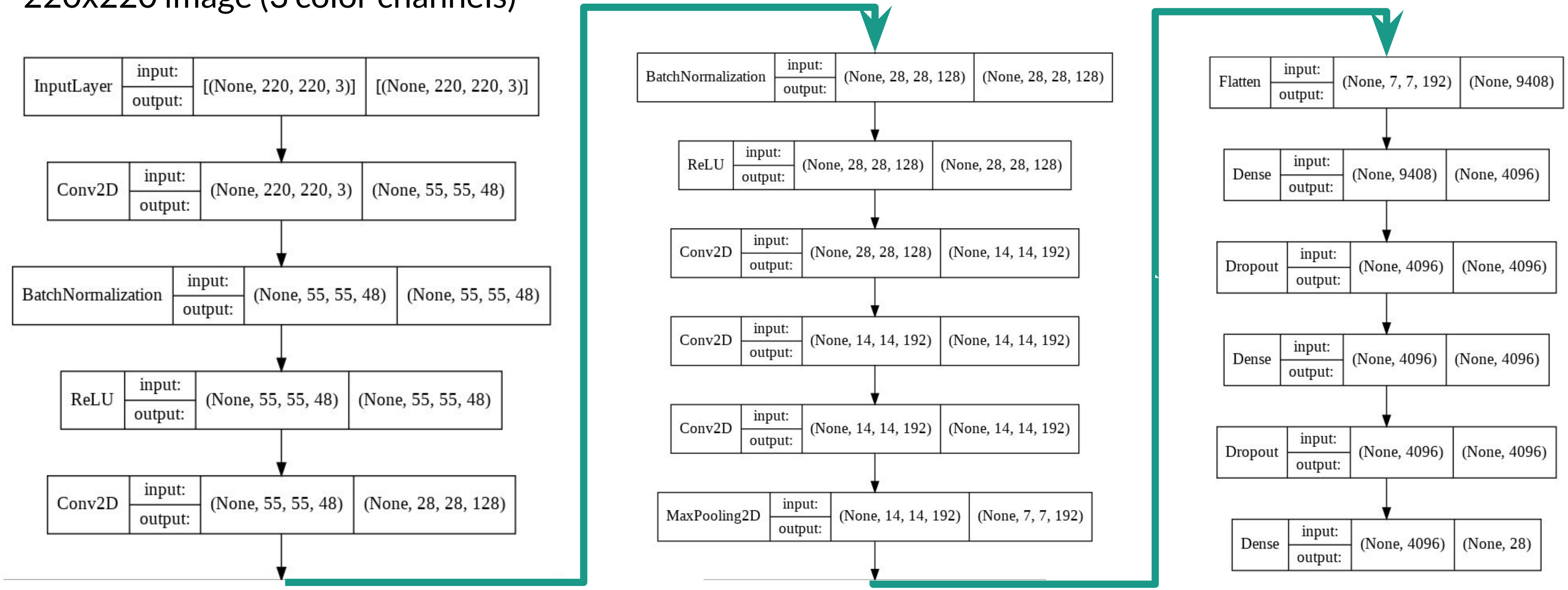
My Implementation

Model

- Based on First stage layed out in the *DeepPose* paper.
- Trained on Leeds Sports Pose Dataset
    - 220x220 pixel image
    - 14 joint coordinates per image



220x220 image (3 color channels)

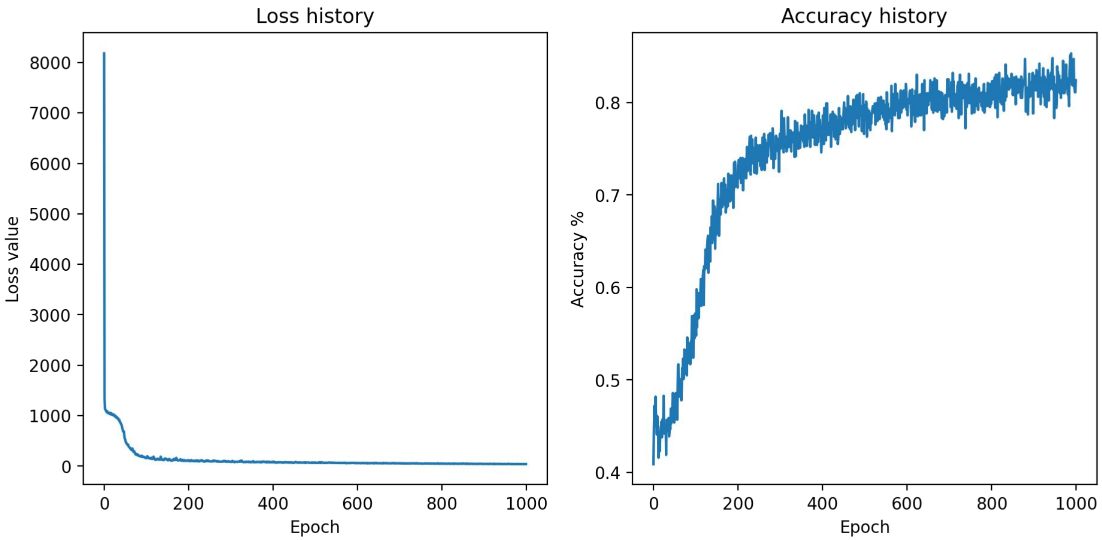


28 individual positions, later
transformed to 14 (x, y) pairs

Training Plots

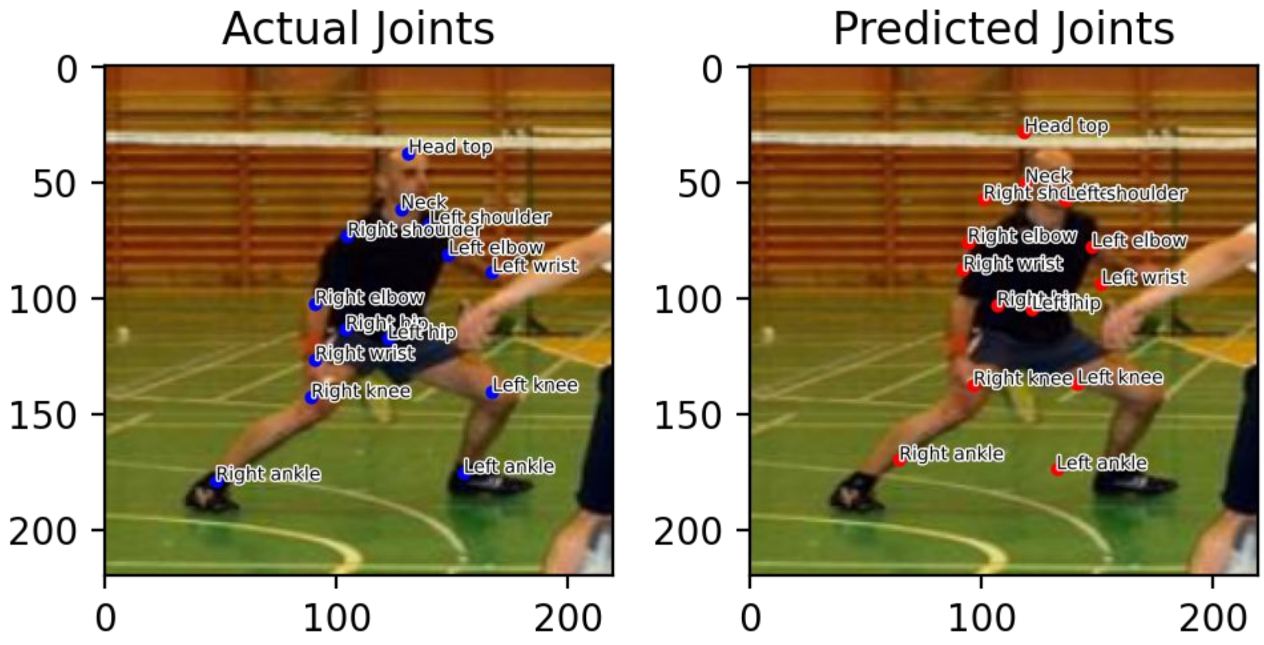
Results

- Trained for 1000 epochs
- Significantly lower testing accuracy compared to training



	Loss	Accuracy
Train	40.4414	0.8240
Test	868.5184	0.4639

Comparing Predicted pose vs Actual



Citations

- A. Toshev and C. Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1653-1660, doi: 10.1109/CVPR.2014.214.
    - https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/42237.pdf
- Wei, S. E., Ramakrishna, V., Kanade, T., & Sheikh, Y. (2016). Convolutional pose machines. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (pp. 4724-4732).
    - https://arxiv.org/pdf/1602.00134.pdf
- Newell, A., Yang, K., & Deng, J. (2016, October). Stacked hourglass networks for human pose estimation. In European conference on computer vision (pp. 483-499). Springer, Cham
    - https://arxiv.org/pdf/1603.06937.pdf

Thank You!