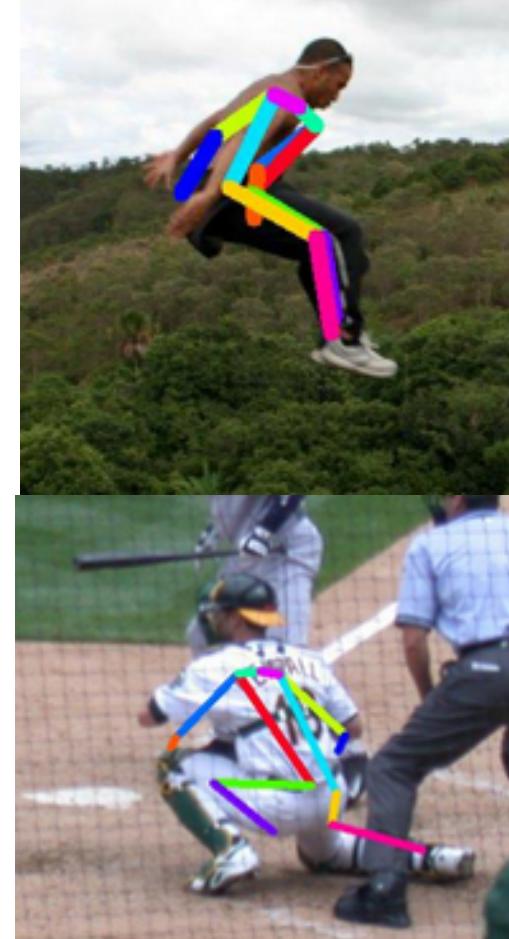


# Convolutional Pose Machines

# Pose Estimation

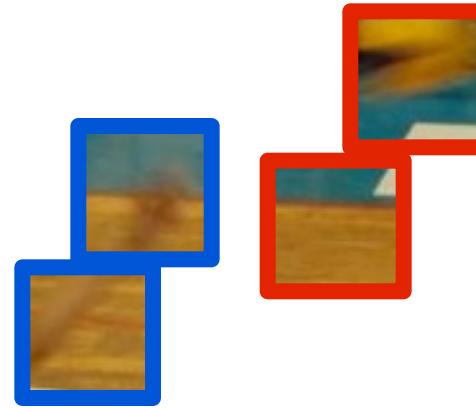
- Key-points
  - 14+ Punkte
  - Hintergrund
  - Skelett Rekonstruktion
- Schwierigkeiten
  - Winkel
  - Unvollständige Informationen
  - Wenig geometrische Zusammenhänge  
=> Extremitäten sind besonders schwer



Which patch corresponds to a body part?



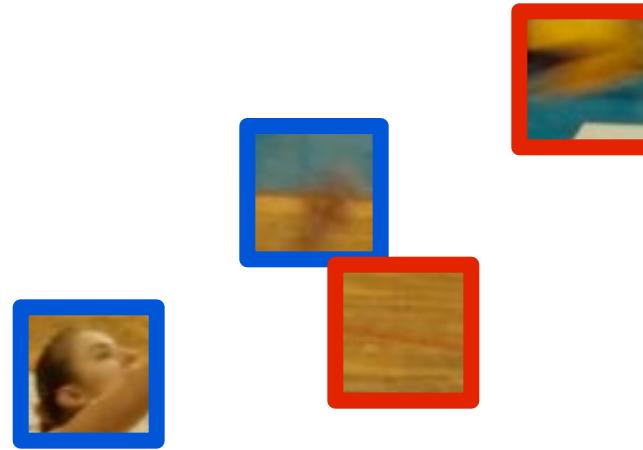
Which patch corresponds to a body part?



Which patch corresponds to a body part?



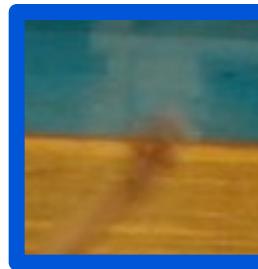
Which patch corresponds to a body part?



Which patch corresponds to a body part?



Which patch corresponds to a body part?



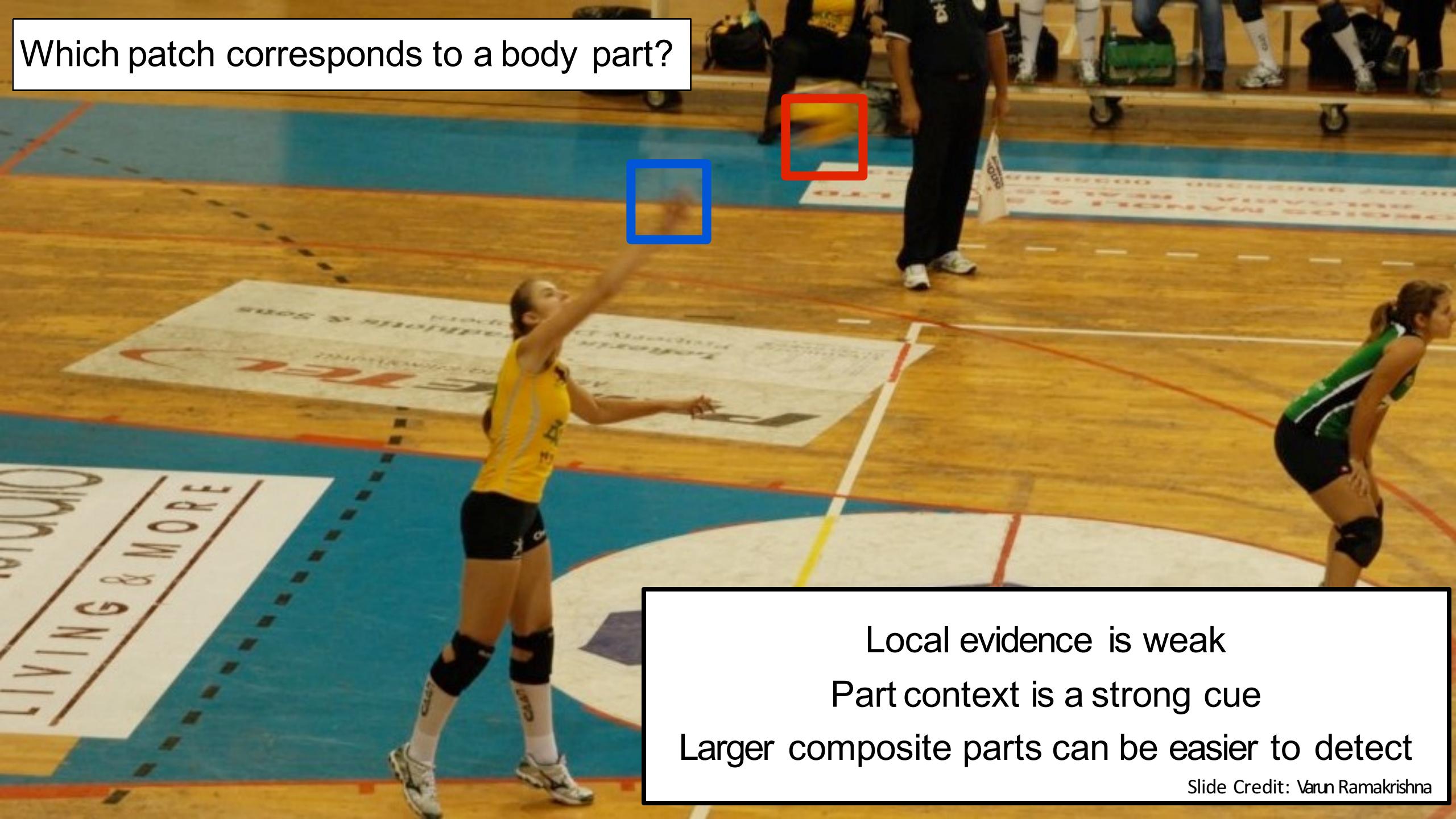
Which patch corresponds to a body part?



Which patch corresponds to a body part?



Which patch corresponds to a body part?



Local evidence is weak  
Part context is a strong cue  
Larger composite parts can be easier to detect

Slide Credit: Varun Ramakrishna

# Related Work

## Pictorial structures

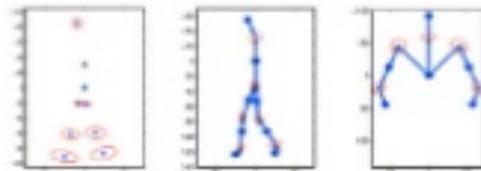
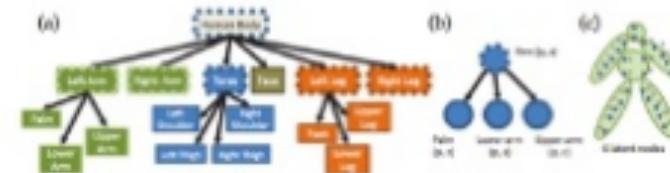


Figure 3. Priors on the part configurations (left to right): Pedestrian detection (star vs. tree model) and upper body detection.

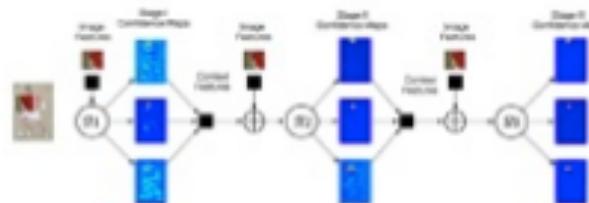
[Mykhaylo et al., CVPR'2009]

## Hierarchical models



[Tian et al., ICCV'2011]

## Sequential prediction



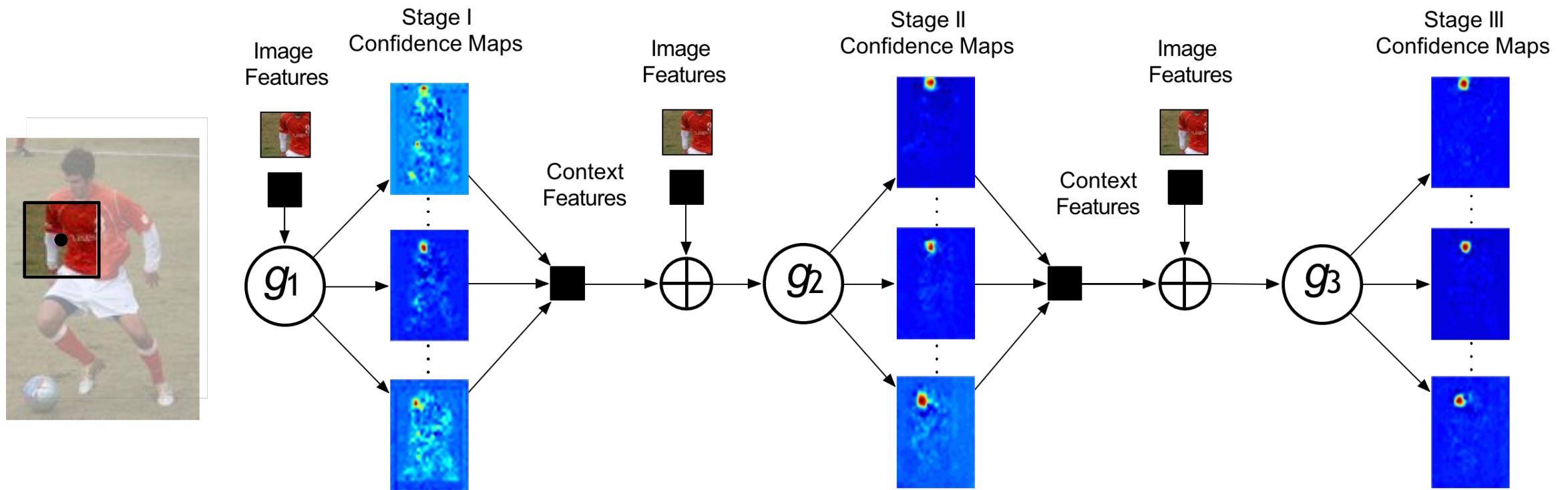
[Ramakrishna et al., 2014]

## Convolutional architectures



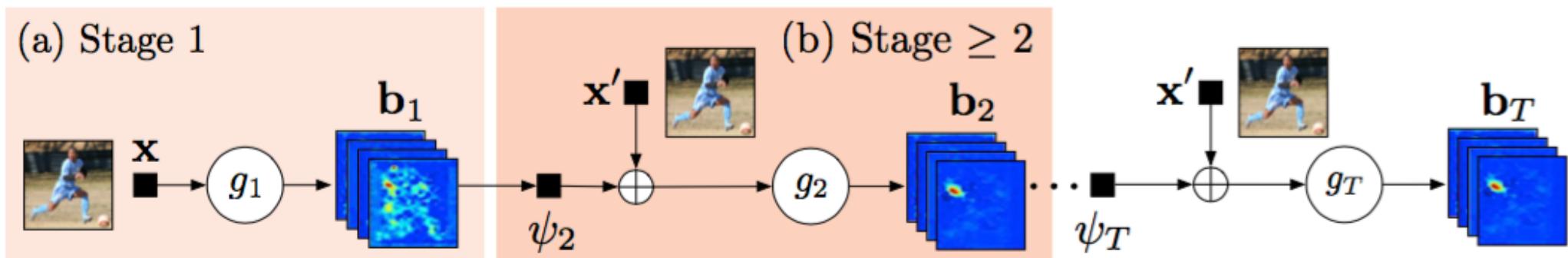
[A. Toshev and C. Szegedy, CVPR'2013]

# Pose Machines

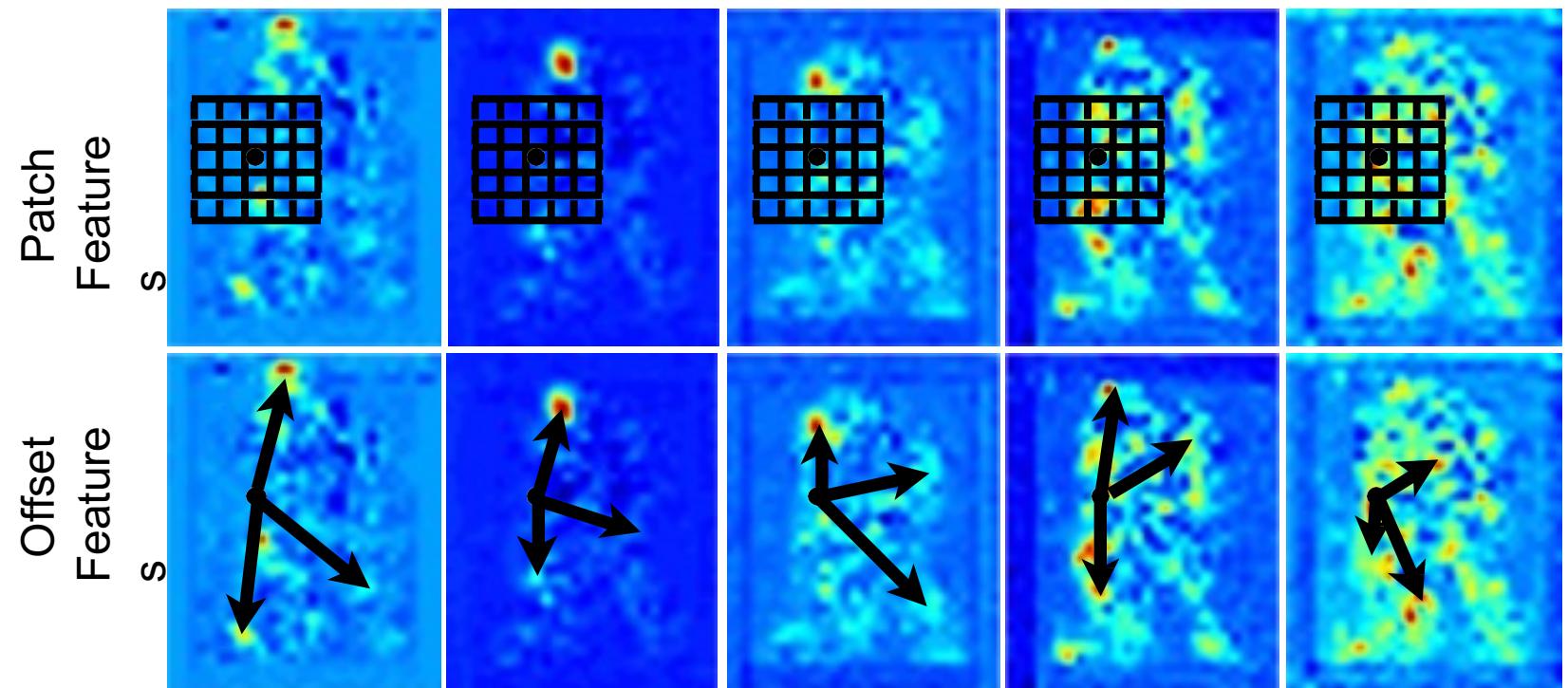


# Pose Machines

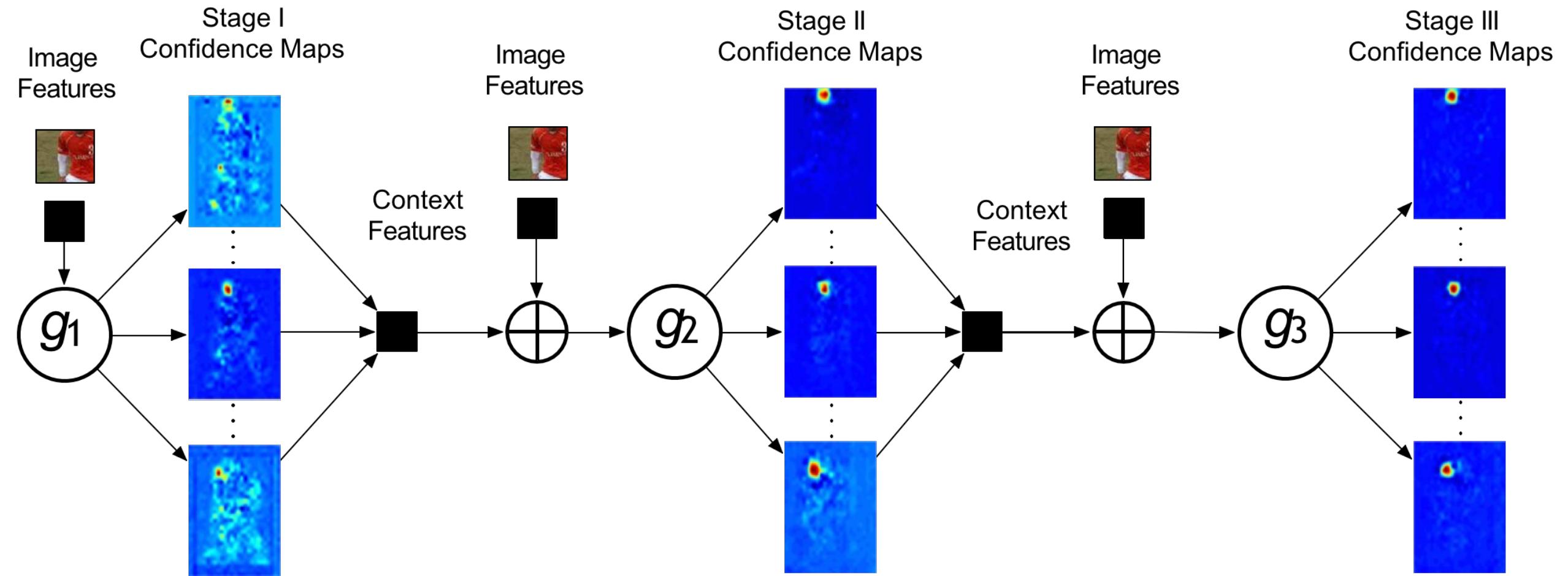
- Feature extraction
  - Histogram of oriented gradients (HoG)
- Predictor
  - Boosted random forest



# Pose Machines – Kontext erfassen

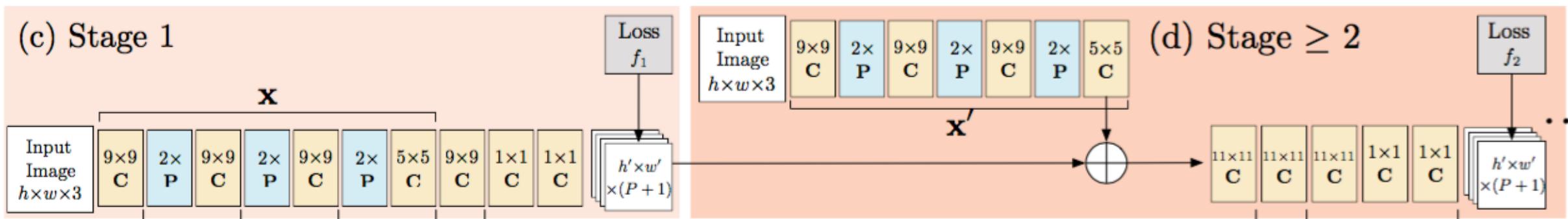


# Pose Machines

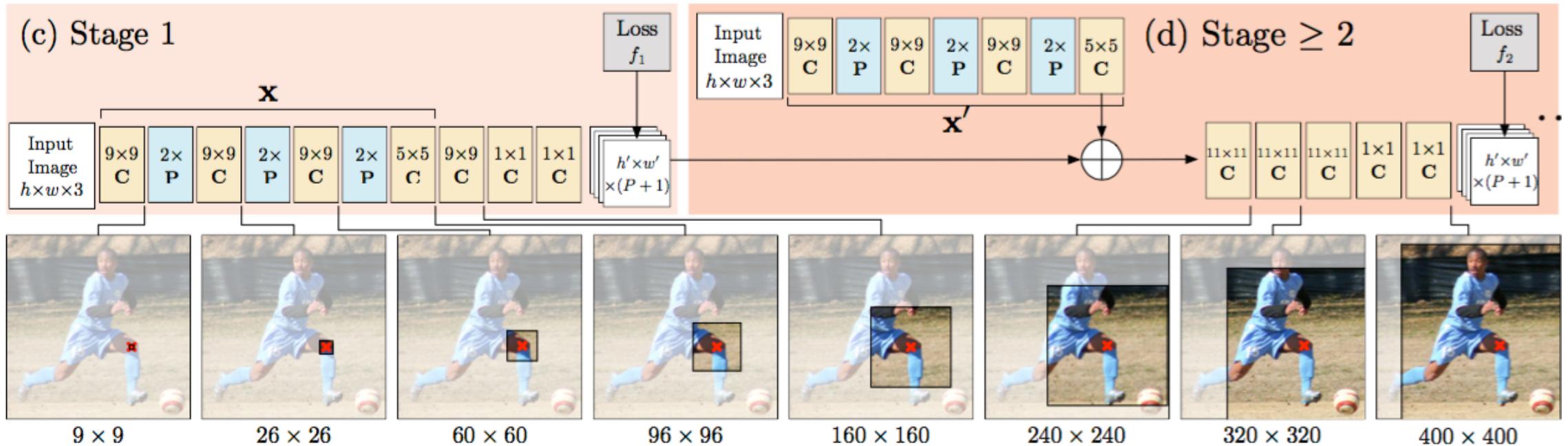


# Convolutional Pose Machines

- Fully convolutional
- End-to-End training
- Implicit spatial relationships



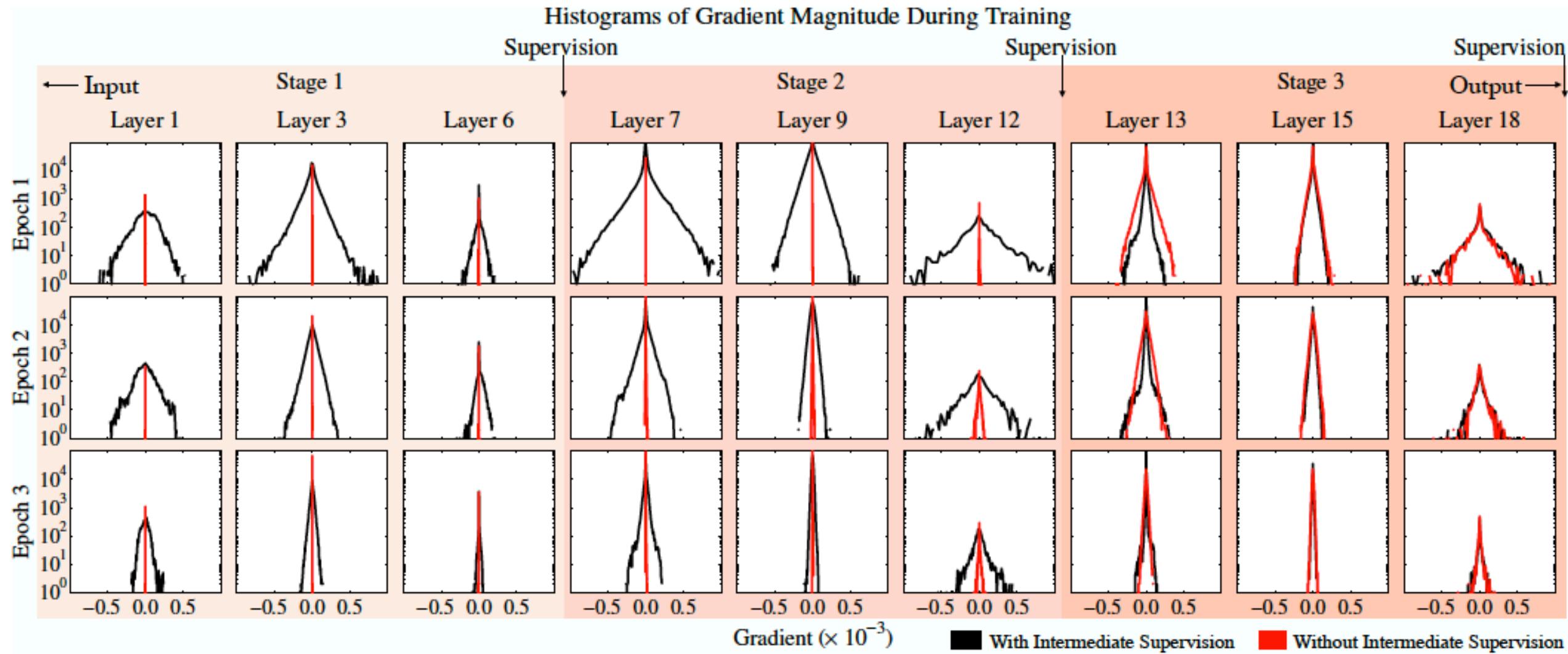
# Receptive field



$$r_{out} = r_{in} + (\text{kernel} - 1) * d_{in}$$

$$d_{out} = d_{in} * \text{stride}$$

# Vanishing Gradients



# Effect of spatial realtioships



Input Image

(a) Stage 1

(b) Stage 2

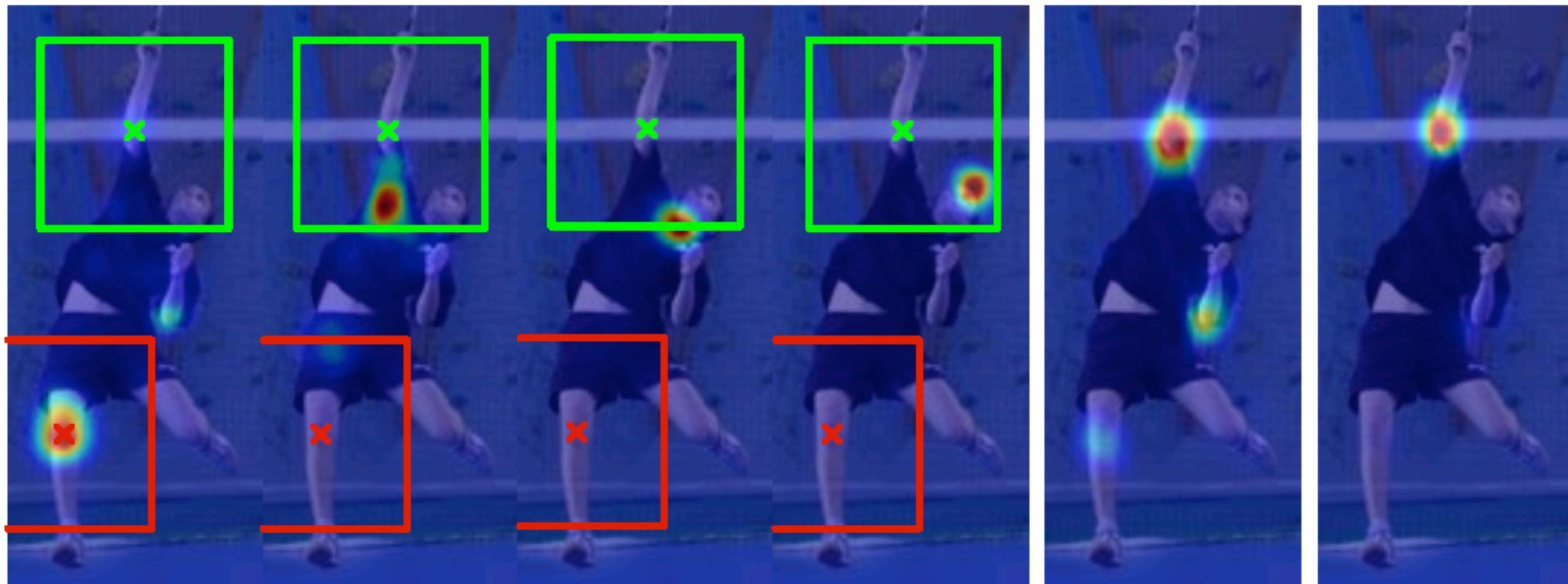
(c) Stage 3

# Spatial relationships and receptive field

stage 1

stage 2

stage 3



R. Elbow

R. Shoulder

Neck

Head

R. Elbow

R. Elbow

# Evaluation

- Eingabebilder 368x368
- Bearbeitet
  - Skaliert
  - Zugeschnitten
  - Gepaddet
- MPII Human Pose Dataset
  - <28000 Bilder
- Augmentierung
  - Rotation [-40°,40°]
  - Skalieren [0.7,1.3]
  - Horizontal spiegeln

# Resultate

- PCKh-0.5 Metrik
- Convolutional Pose Machines (2016) 88.0%
- Convolutional Pose Machines+ (2016) 88.5%
- Deeply Learned Compositional Models for Human Pose Estimation (2018)  
92.3%



# Danke!

- <https://www.youtube.com/watch?v=3bF6PRwFADY>