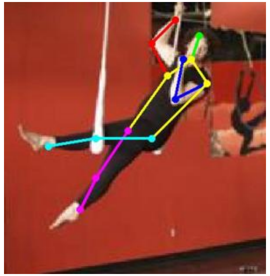
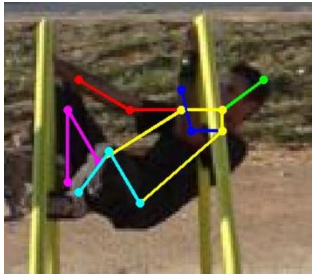
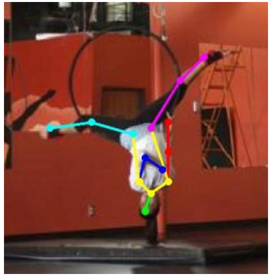
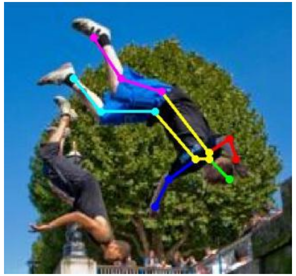


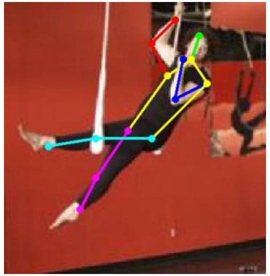
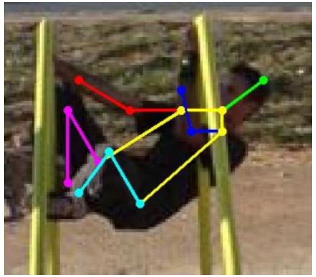
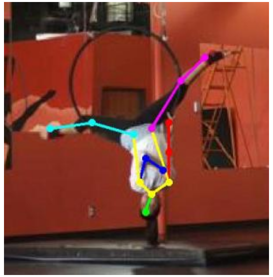
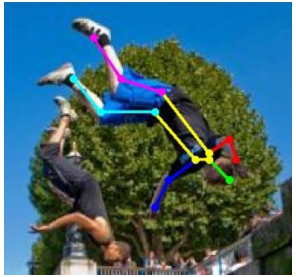
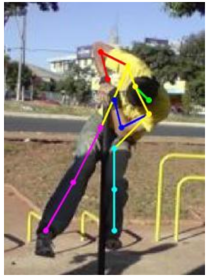





人体姿态识别年度总结

欧阳万里

香港中文大学

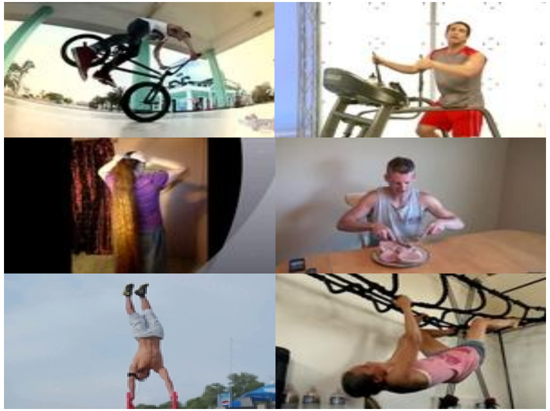
What is Human Pose Estimation?

Results are generated by our proposed methods (without temporal constraints).

Video Credit: Peter Jasko solo - M-idzomer 2013

Applications



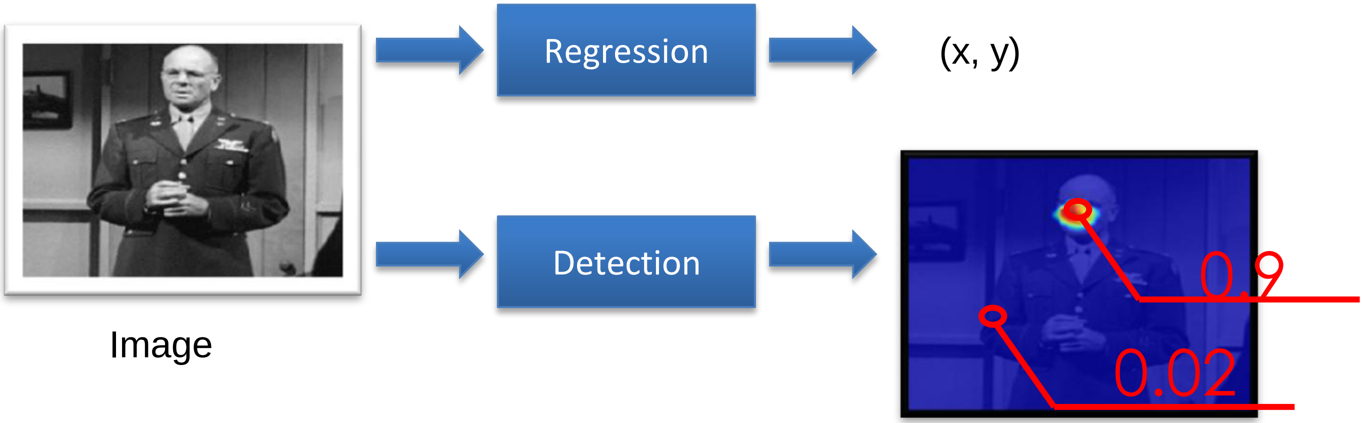
Action Recognition

HCI, Game and Animation

Clothing Parsing
[Yamaguchi et al. CVPR’14]

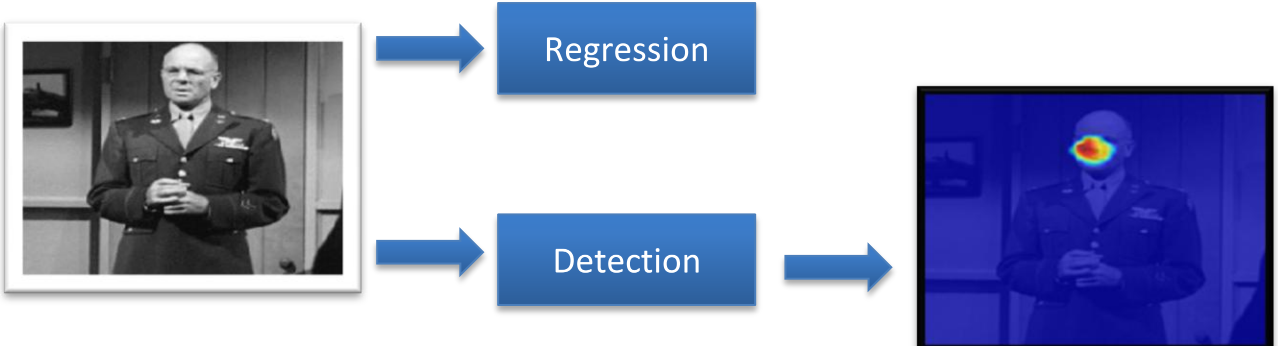
Regression or detection?

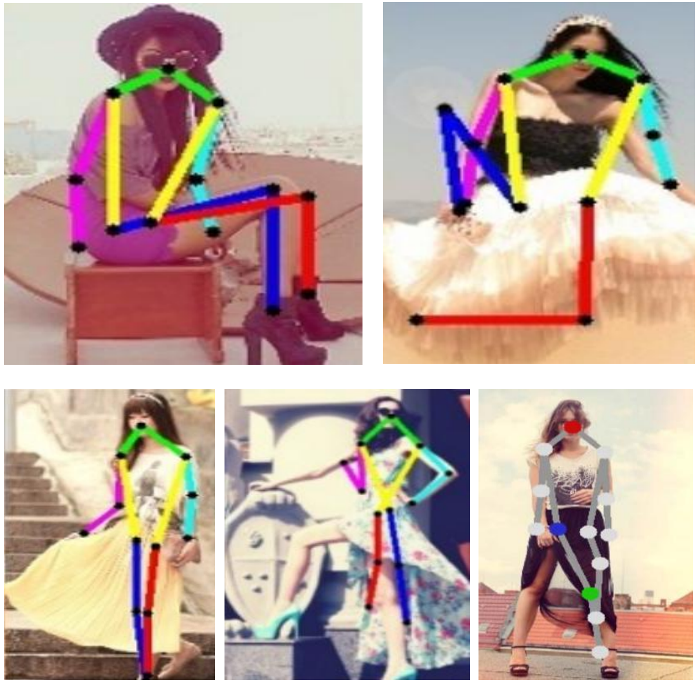
- Output coordinates
  - To regress the body locations
- Output heatmaps
  - To detect the body locations



Limitations

- Regression
  - Low accuracy in high precision region caused by flexible body movement
  - Hard to extend when pose estimation is used for unknown number of persons
- Detection





Dantone et al. CVPR 2013

Challenge

- Body movement
- Foreshortening
- Clothing

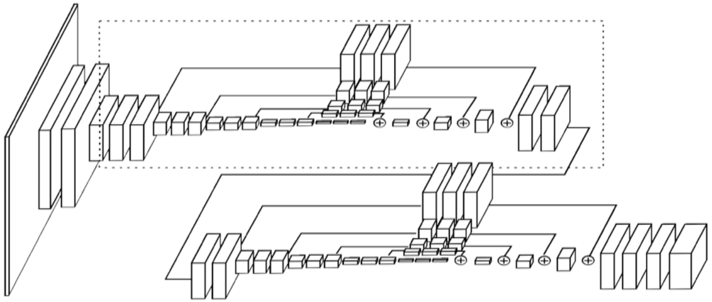
Can you tell which part is from an image patch?

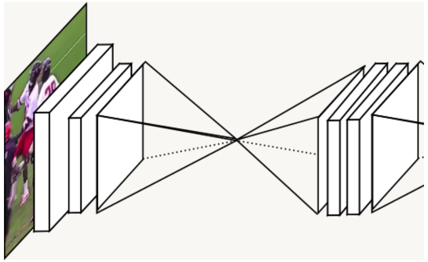
What about this patch?

- Local appearance is insufficient
- Global appearance is helpful

Stacked hourglass network

- Hourglass
  - Subsampling: lower resolution for less computation and larger receptive field
  - => upsampling: higher resolution for more accurate localization

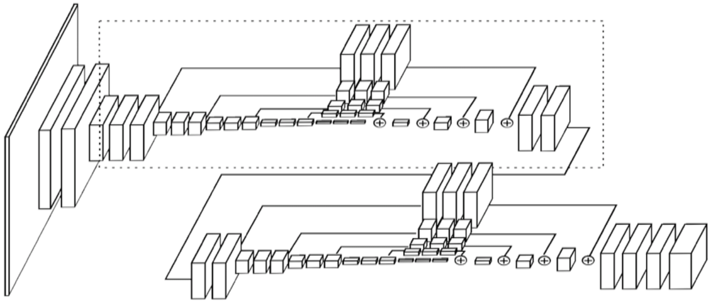


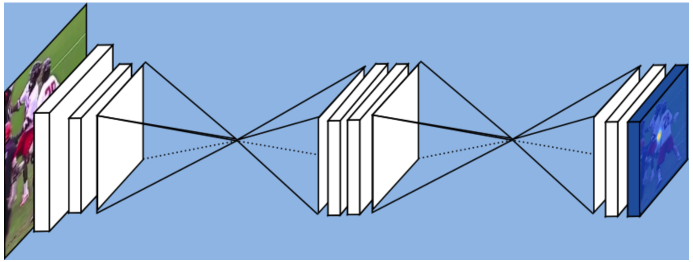


Newell, Alejandro, Kaiyu Yang, and Jia Deng. "Stacked hourglass networks for human pose estimation." *ECCV*, 2016. [@ University of Michigan]

Stacked hourglass network

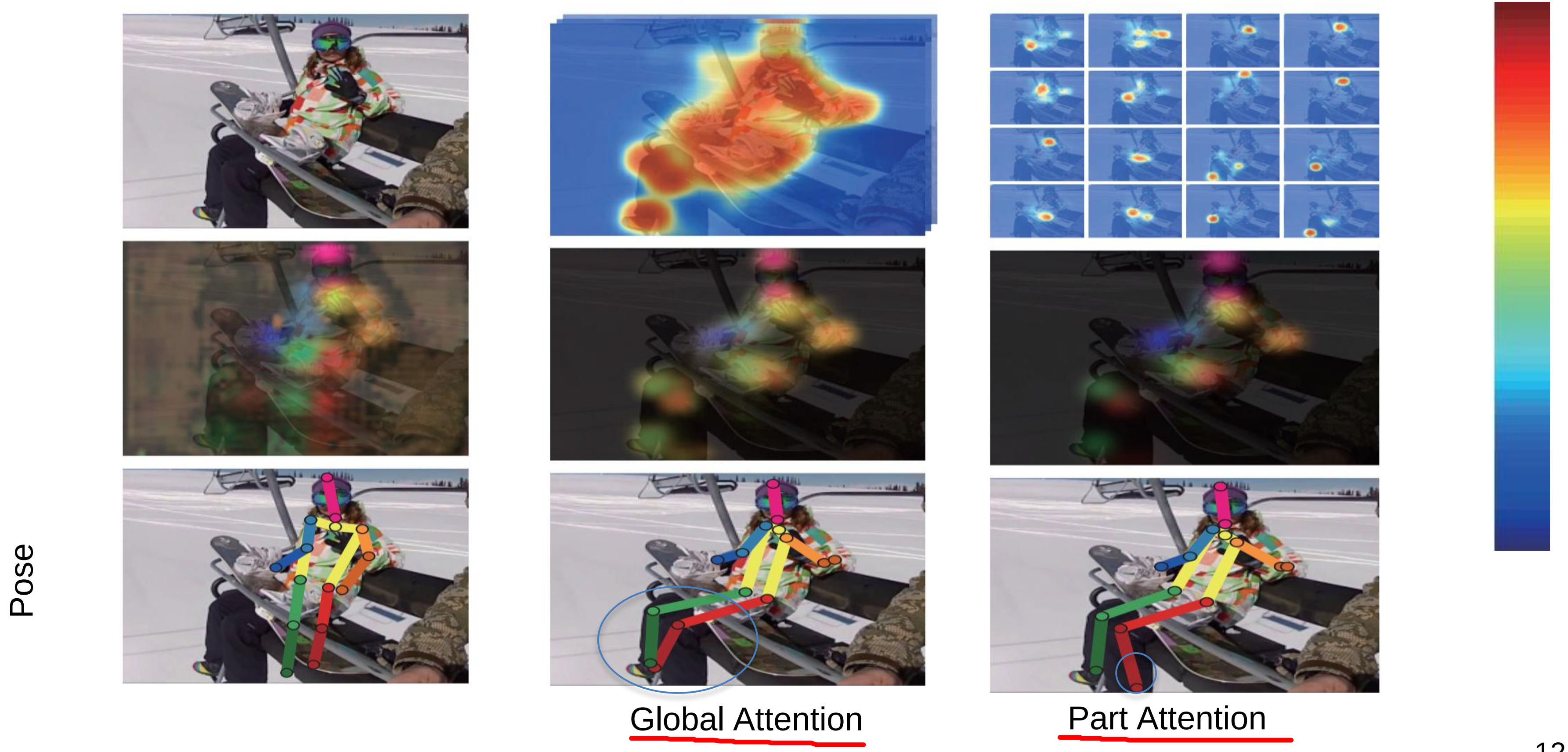
- ❖ Hourglass, subsampling => upsampling
- ❖ Stack multiple hourglass structures





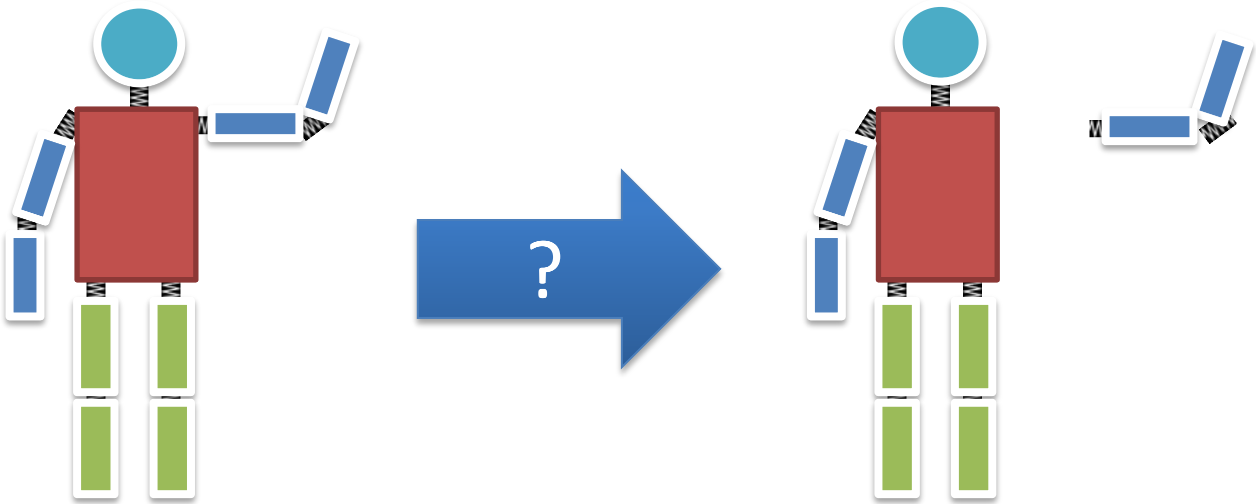
Newell, Alejandro, Kaiyu Yang, and Jia Deng. "Stacked hourglass networks for human pose estimation." *ECCV*, 2016. [@ University of Michigan]

How attention helps human pose estimation?

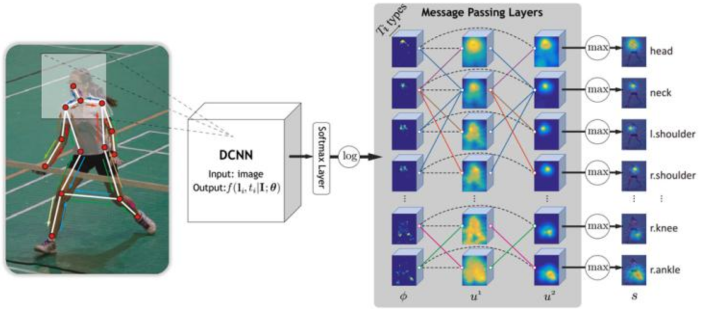


X. Chu, W. Yang, W. Ouyang , X. Wang, A. Yuille. "Multi-Context Attention for Human Pose Estimation", *Proc. CVPR* , 2017.

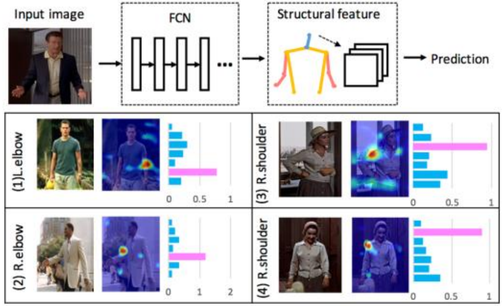
Structure also matters...



Structure also matters...



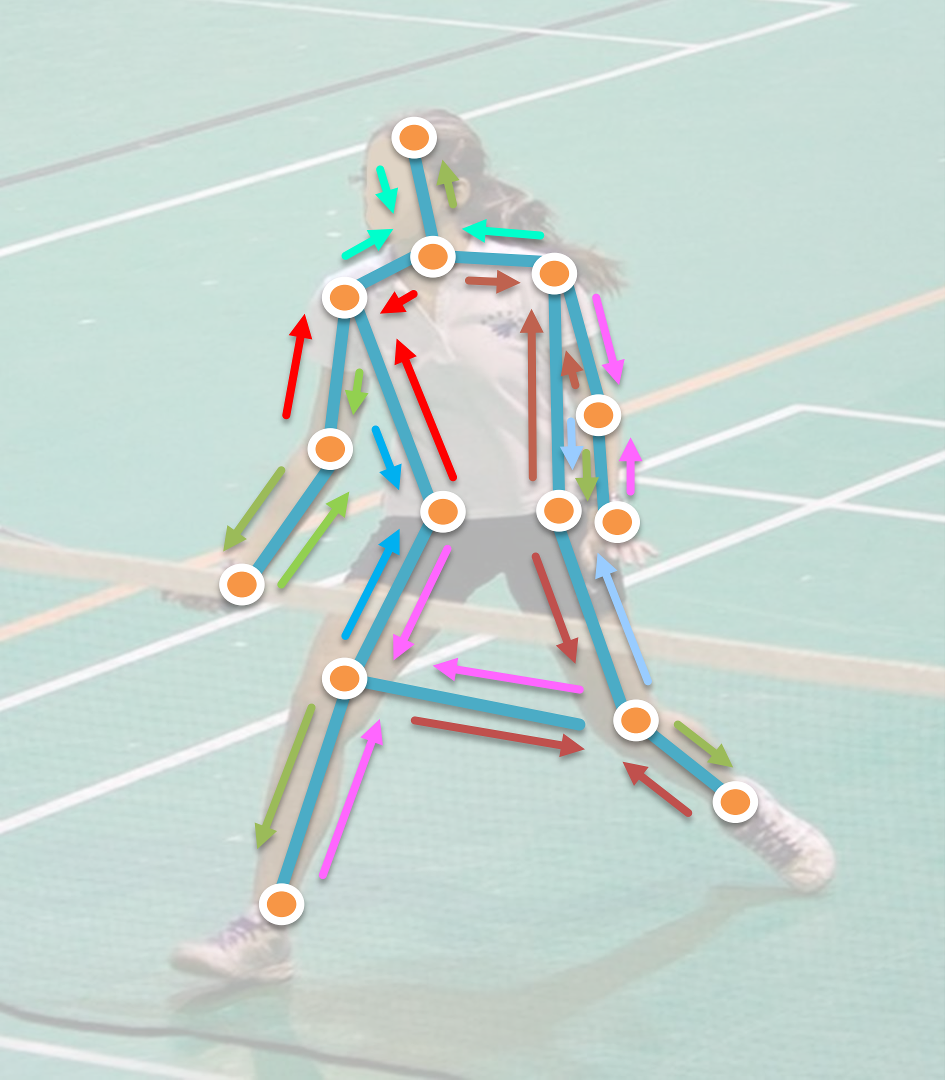
Deep Mixture of Parts



Structured Feature Learning

(source code provided)

Wei Yang, Wanli Ouyang, Hongsheng Li and Xiaogang Wang "End-to-End Learning of Deformable Mixture of Parts and Deep Convolutional Neural Networks for Human Pose Estimation", In *Proc. CVPR* 2016 (Oral).
X. Chu, Wanli Ouyang , H. Li, and X. Wang. "Structured feature learning for pose estimation", In *Proc. CVPR* 2016.

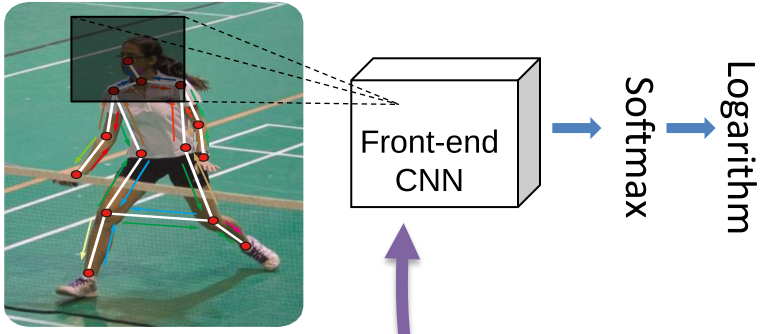


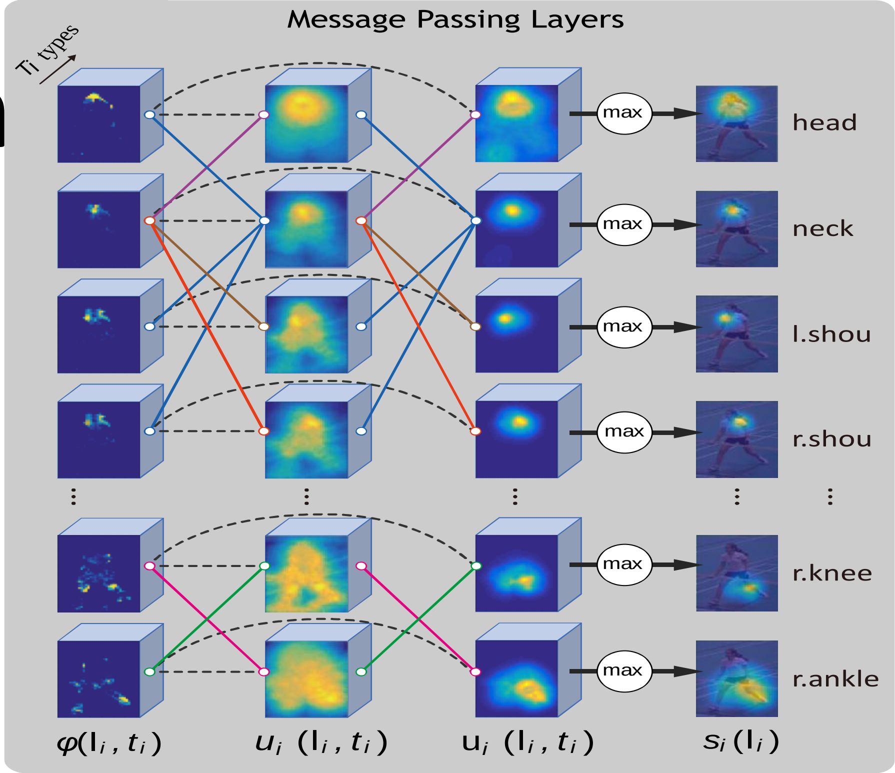
Graph model

$G = (V, E)$

- **Vertices**
  - Locations and mixture types of body parts
  - Modeled by a front-end CNN
- **Edges** message passing
  - Pairwise spatial relationships between body parts
  - Modeled by message passing layers

Fram

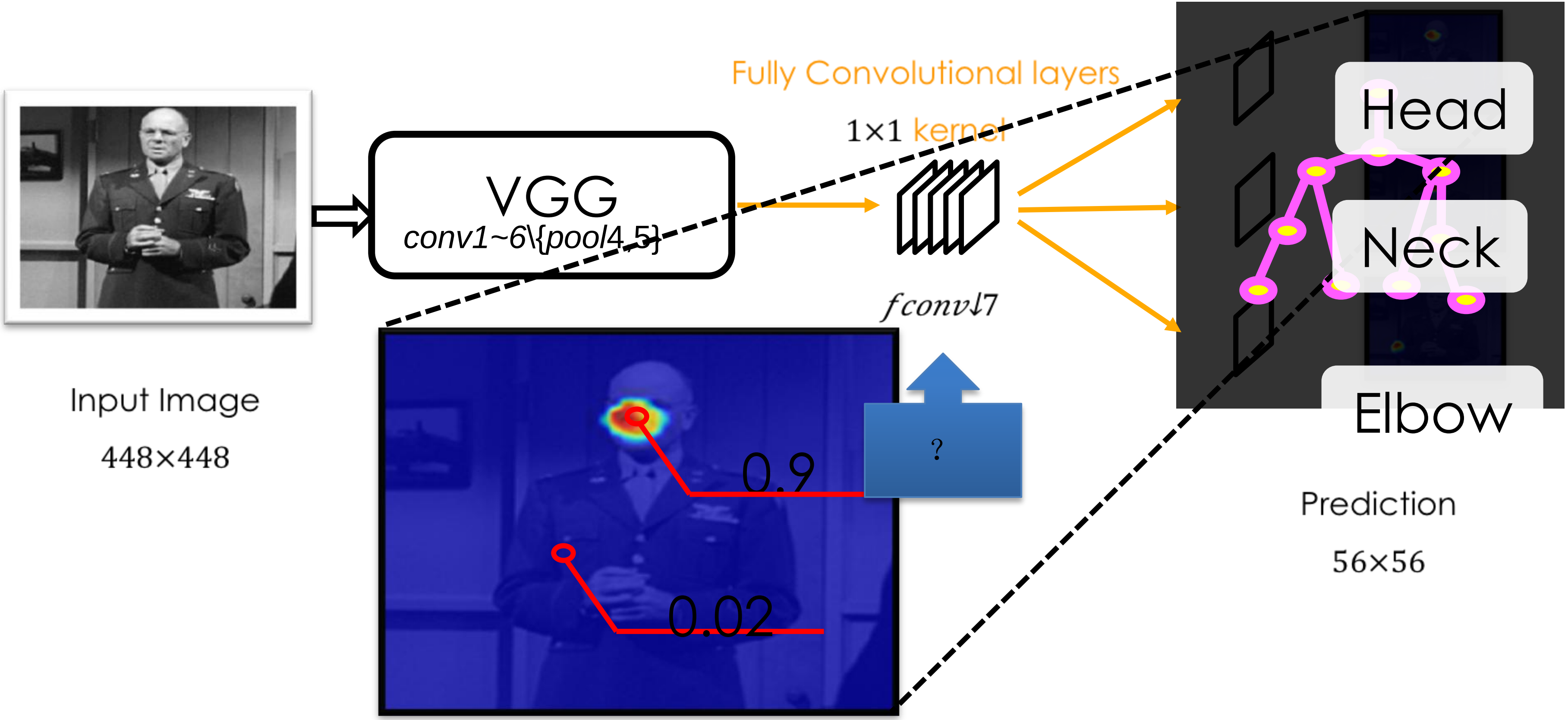




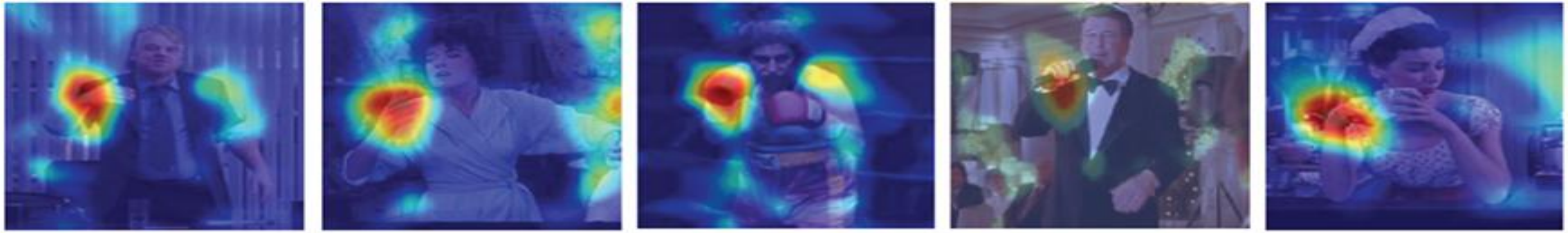
$$F(\boldsymbol{l}, \mathbf{t}|I;\boldsymbol{\theta},\boldsymbol{\omega}) = \sum_{i\in V} \phi(l_i, t_i|I;\theta) + \sum_{(i,j)\in E} \psi(l_i, l_j, t_i, t_j|I;\omega_{i,j}^{t_i,t_j})$$

Wei Yang, Wanli Ouyang, Hongsheng Li and Xiaogang Wang "End-to-End Learning of Deformable Mixture of Parts and Deep Convolutional Neural Networks for Human Pose Estimation", In *Proc. CVPR* 2016 (Oral).

Fully convolutional net for Human pose estimation

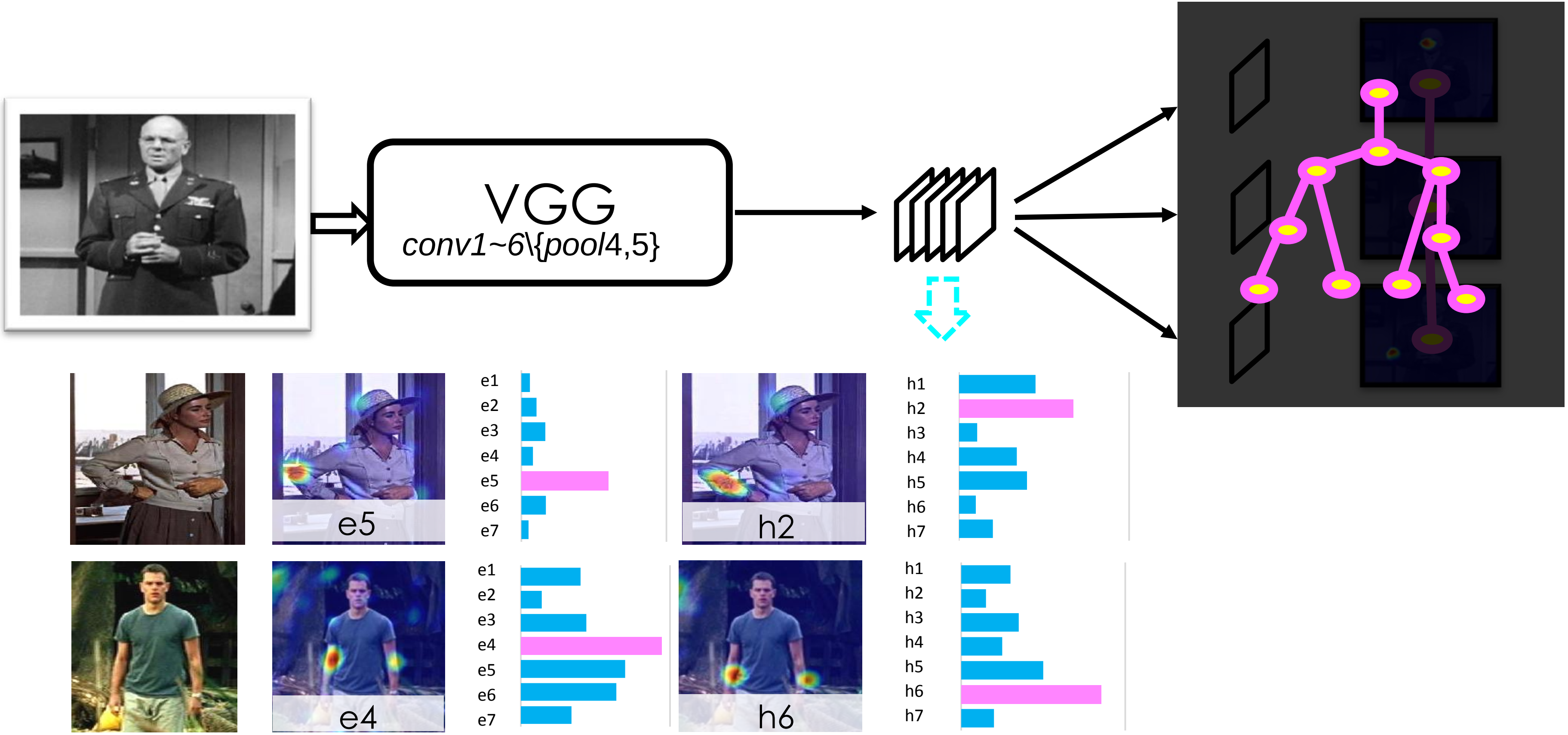


High responding images for channel 1 for neck

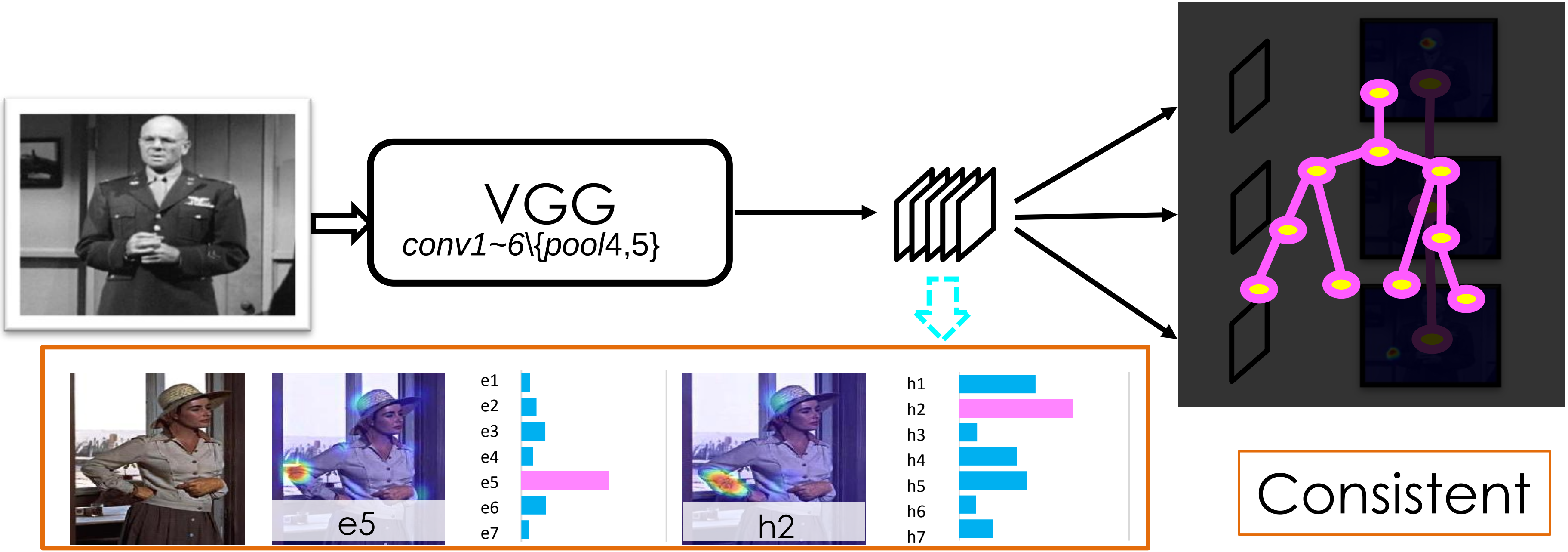


High responding images for channel 2 for left shoulder

Fully convolutional net for Human pose estimation

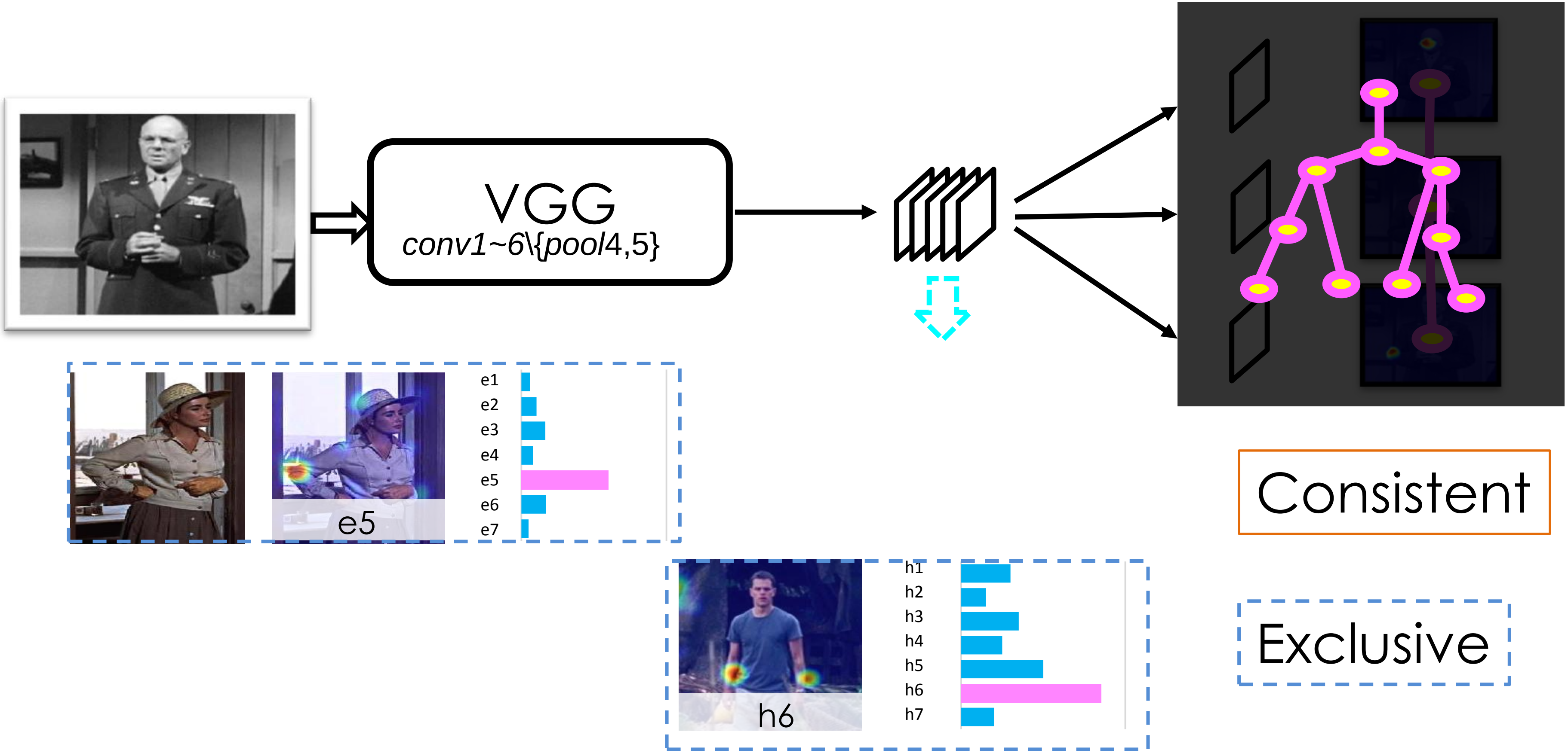


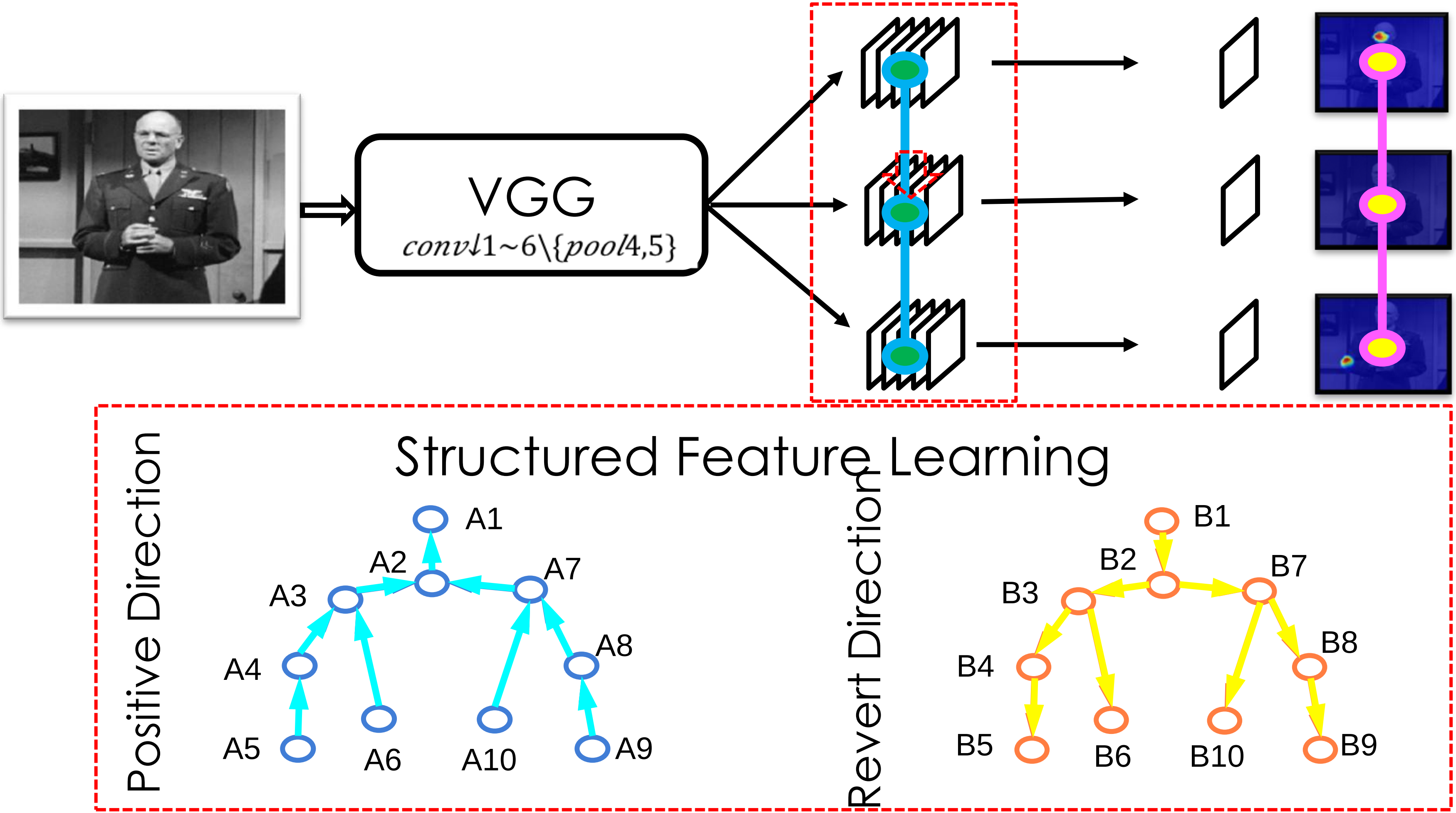
Fully convolutional net for Human pose estimation



Consistent

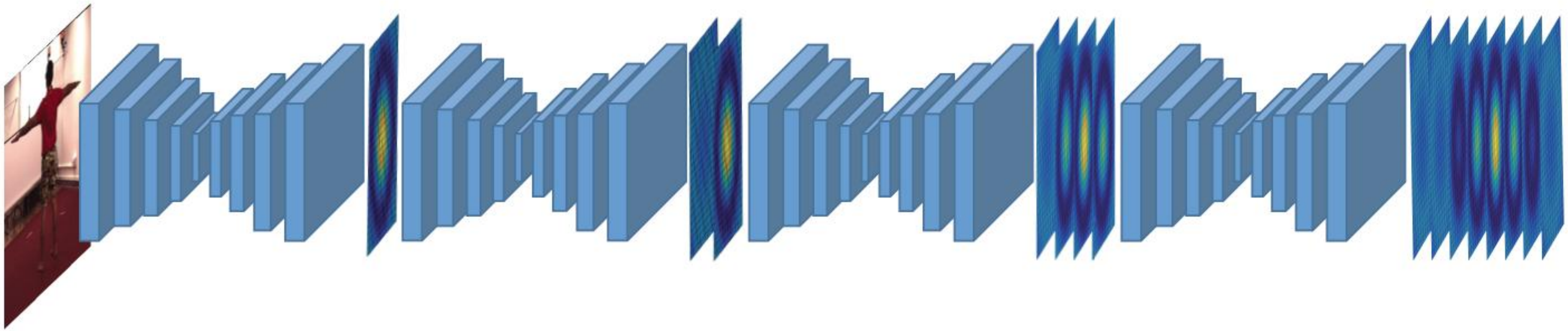
Fully convolutional net for Human pose estimation





X. Chu, W. Ouyang , H. Li, and X. Wang. "Structured feature learning for pose estimation", In *Proc. CVPR* 2016.

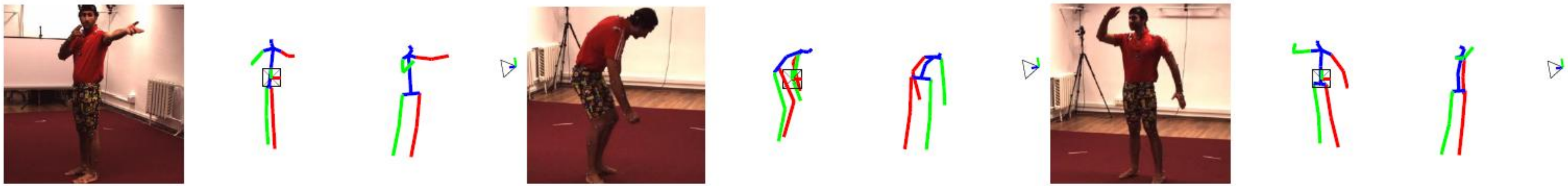
2D pose to 3D pose (from image)



2D pose

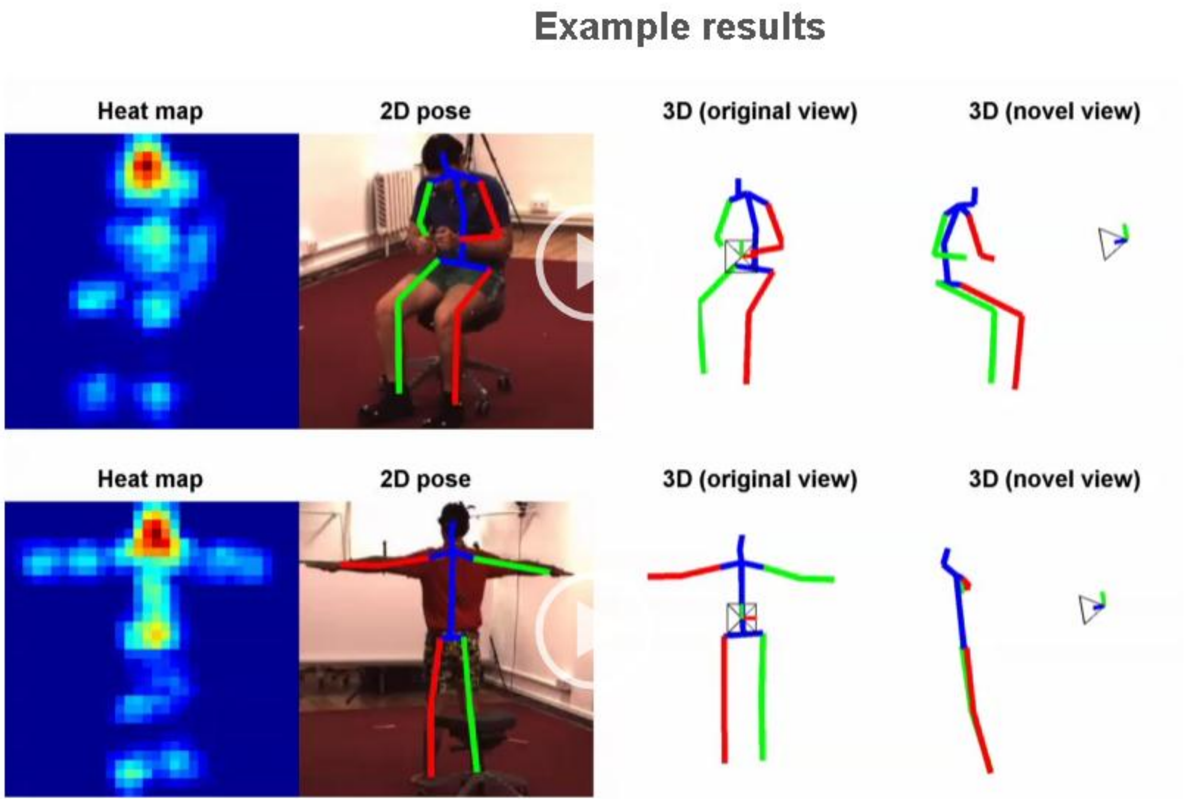
Coarse 3D pose

Fine 3D pose



Pavlakos, Georgios, et al. "Coarse-to-Fine Volumetric Prediction for Single-Image 3D Human Pose." *CVPR,* 2017.

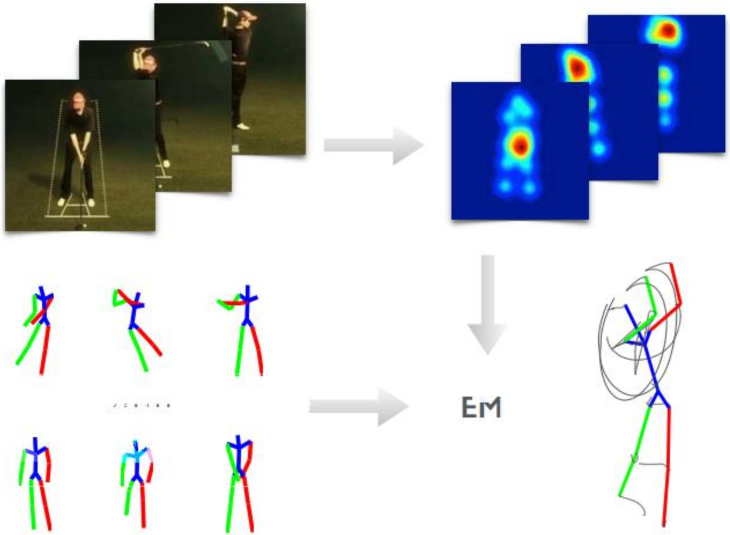
2D pose to 3D pose (from video)



Zhou, Xiaowei, et al. "Sparseness meets deepness: 3D human pose estimation from monocular video." *CVPR*, 2016.

2D pose to 3D pose (from video)

- Sparsity-driven 3D geometric prior
  - 3D pose can be represented as a linear combination of predefined basis poses
- Temporal smoothness
  - Poses are similar in adjacent frames



Zhou, Xiaowei, et al. "Sparseness meets deepness: 3D human pose estimation from monocular video." *CVPR*, 2016.

Other interesting papers

- Toshev, Alexander, and Christian Szegedy. "Deeppose: Human pose estimation via deep neural networks." CVPR 2014.
- Tompson, Jonathan J., et al. "Joint training of a convolutional network and a graphical model for human pose estimation." NIPS, 2014.
- Chen, Xianjie, and Alan L. Yuille. "Articulated pose estimation by a graphical model with image dependent pairwise relations." NIPS, 2014.
- Jain, Arjun, et al. "Learning human pose estimation features with convolutional networks." ICLR, 2014.
- Jain, Arjun, et al. "Modeep: A deep learning framework using motion features for human pose estimation." ACCV, 2014.
- Tompson, Jonathan, et al. "Efficient object localization using convolutional networks." CVPR. 2015.
- Fan, Xiaochuan, et al. "Combining local appearance and holistic view: Dual-source deep neural networks for human pose estimation." CVPR, 2015.
- Pfister, Tomas, James Charles, and Andrew Zisserman. "Flowing convnets for human pose estimation in videos." ICCV, 2015.
- Chu, Xiao, et al. "Structured feature learning for pose estimation." CVPR, 2016.
- Yang, Wei, et al. "End-to-end learning of deformable mixture of parts and deep convolutional neural networks for human pose estimation." CVPR, 2016.
- Xiao Chu, Wei Yang, W. Ouyang , Xiaogang Wang, Alan Yuille. ”Multi-Context Attention for Human Pose Estimation”, *Proc. CVPR* , 2017.
- Gkioxari, Georgia, Alexander Toshev, and Navdeep Jaitly. "Chained Predictions Using Convolutional Neural Networks." ECCV,2016.
- J. Charles, et al, Personalizing Human Video Pose Estimation, CVPR16
- Carreira, Joao, et al. "Human pose estimation with iterative error feedback.” CVPR 2016.
- Insafutdinov E, Pishchulin L, Andres B, et al. DeeperCut: A Deeper, Stronger, and Faster Multi-person Pose Estimation Model. ECCV 2016.
- Pishchulin, L., Insafutdinov, E., Tang, S., Andres, B., Andriluka, M., Gehler, P., & Schiele, B. DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation, CVPR 2016


рахмат
danke
謝謝
ngiyabonga
Баярлалаа
spas
teşekkür ederim
mersi
kia ora
barka
welalin
tack
mahalo
tapadh leat
спасибо
faafetai lava
vinaka
спасибі
blagodaram
dank je
misaotra
matondo
paldies
grazzi
kiitos
dankie
nanni
nandri
dhanyavad
thank you
gracias
хвала
asante
manana
bedankt
hvala
mauruuru
köszönöm
akun
dankon
ačiū
obrigada
tenki
enkosi
bayarlalaa
gràcie
djiere dieuf
tau
mochchakkeram
murakoze
dziękuję
дякую
mamnun
chokrane
chnorakaloutioun
gratias ago
gràcies
sulpáy
go raibh maith agat
sobodi
děkuji
sağolun
sukriya
kop khun krap
taiku
obrigado
arigatô
takk
dakujem
trugarez
mèsi
ありがとう
tanemirt
grazie
didi madloba
najis tuke
rahmet
diolch
dhanyavadagalu
shukriya
mercé
мерси
kam sah hamnida
terima kasih
ευχαριστώ
merci
rahmat
xièxie
감사합니다