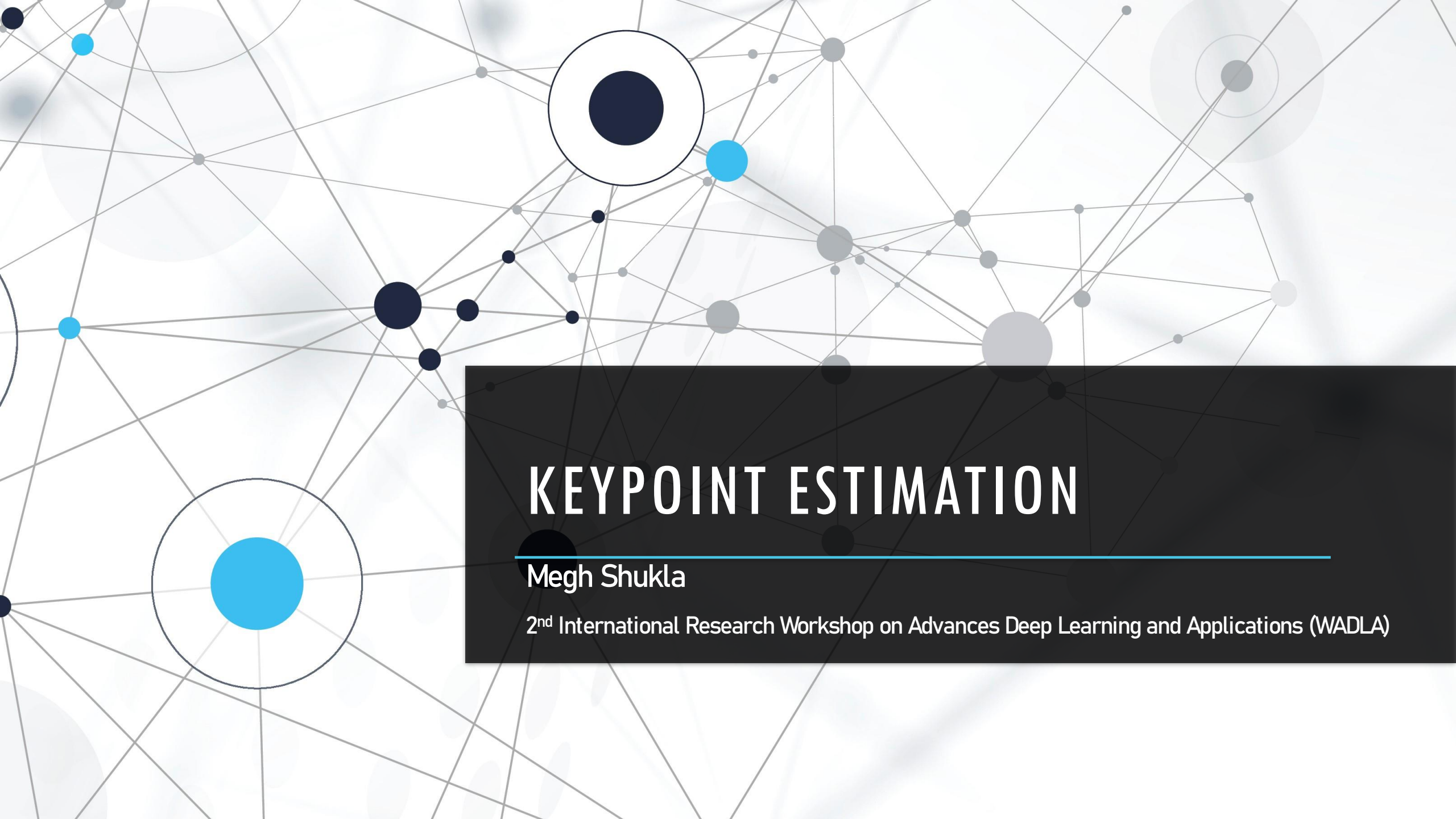
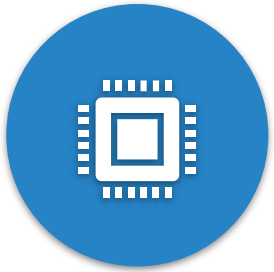


KEYPOINT ESTIMATION

Megh Shukla

2nd International Research Workshop on Advances Deep Learning and Applications (WADLA)

KEYPOINT ESTIMATION



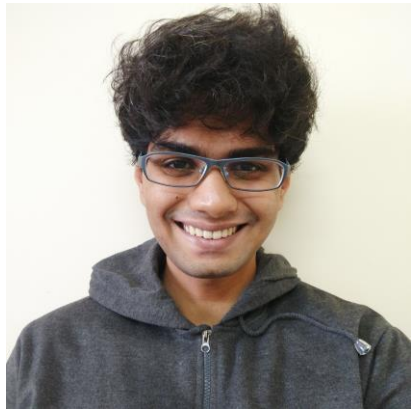
Modelling

Data

Demo

Modelling: How does keypoint estimation work?
Data: How do we collect data to make keypoint estimation work?
Demo: Talk is cheap, show me the code!

BIO



That's me, pre-pandemic!

Research Engineer

Mercedes-Benz Research and Development India

Active Learning for Human Pose Estimation [1, 2, 3]

Source

Master of Technology

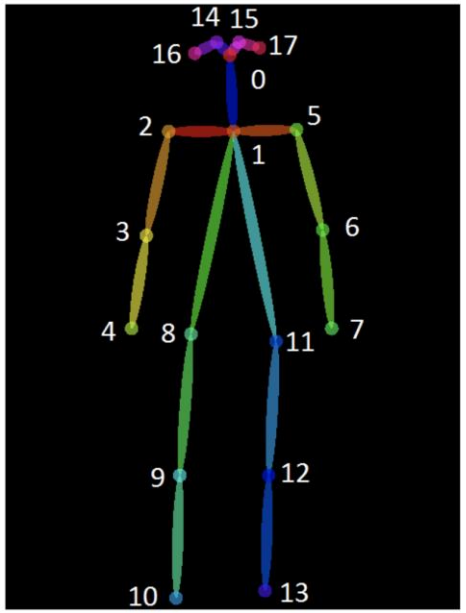
Indian Institute of Technology Bombay

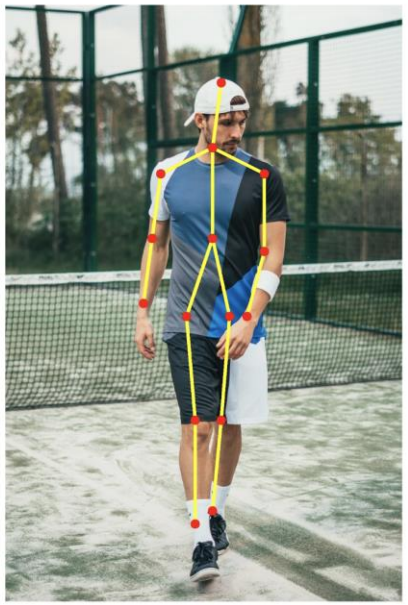
Bachelor of Engineering

University of Mumbai

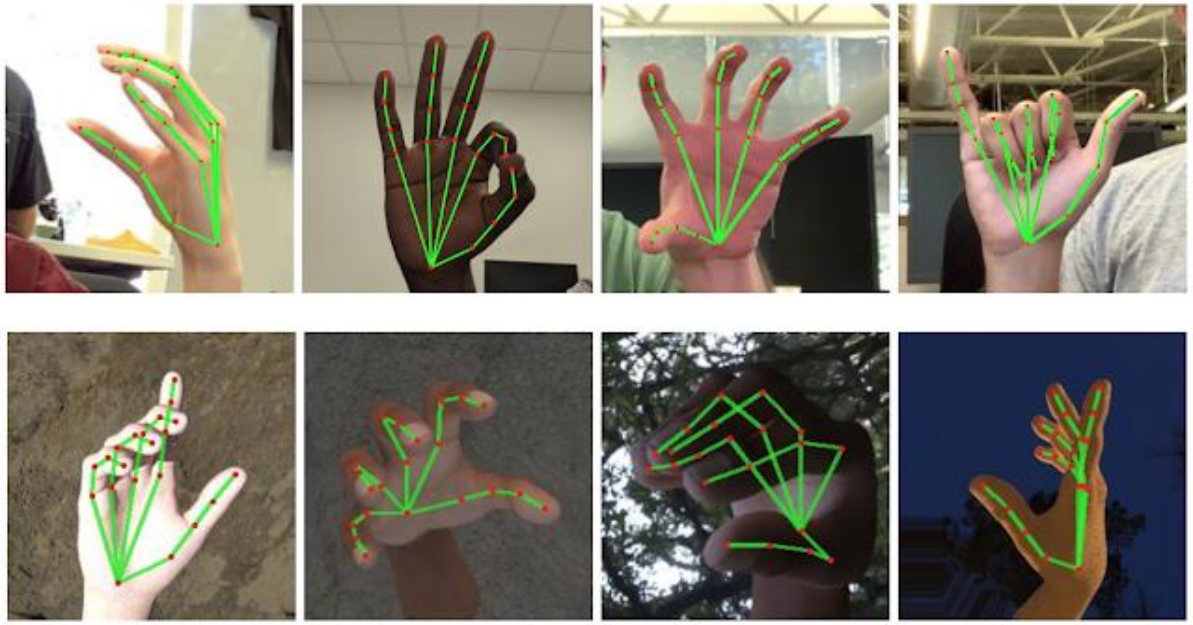
[1] Megh Shukla, "Bayesian Uncertainty and Expected Gradient Length - Regression: Two Sides Of The Same Coin?" In WACV 2022
[2] Megh Shukla and Shuaib Ahmed, "A Mathematical Analysis Of Learning Loss For Active Learning In Regression" In CVPRW 2021
[3] Megh Shukla and Shuaib Ahmed, "A Method For Annotating One Or More Images Of A User" In Mercedes-Benz AG Patent (Filed)

WHAT IS KEYPOINT ESTIMATION?

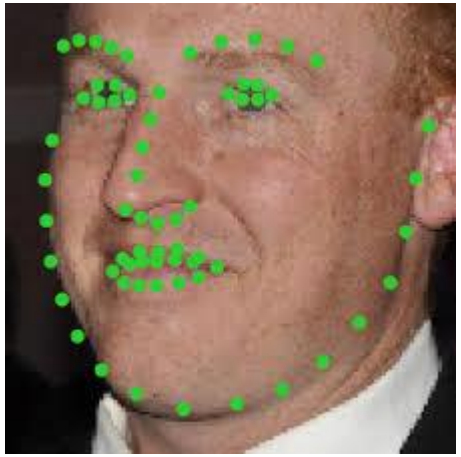




Source

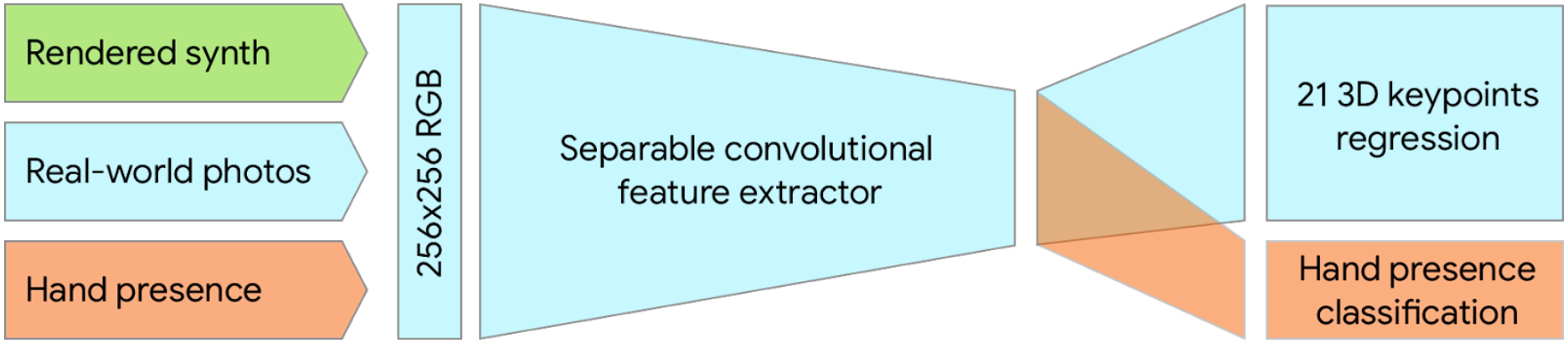


Source

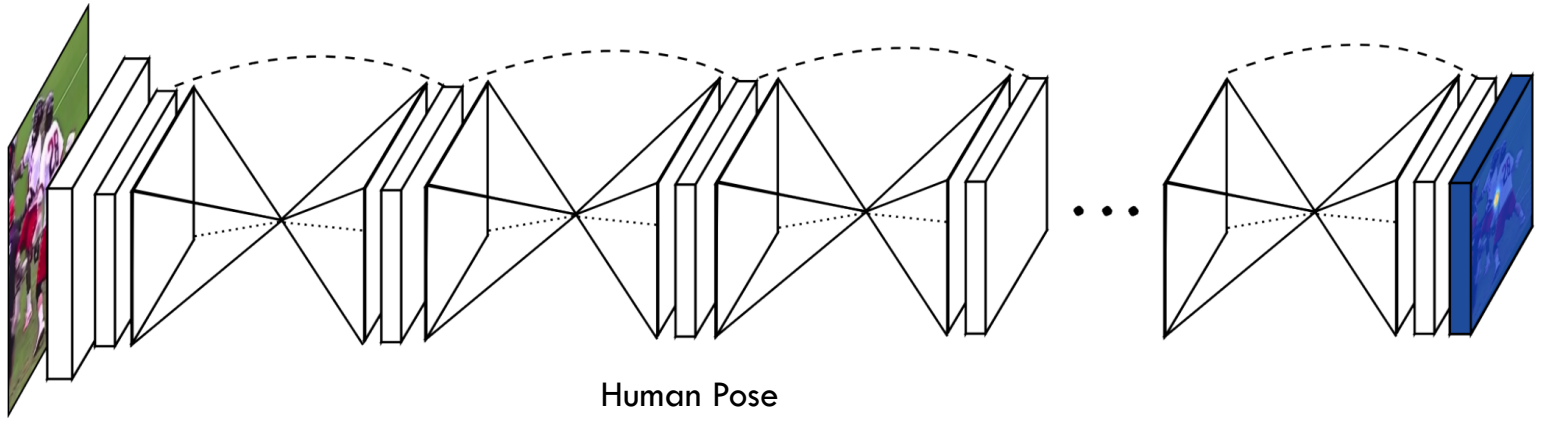


Source

EXAMPLES: KEYPOINT ESTIMATION



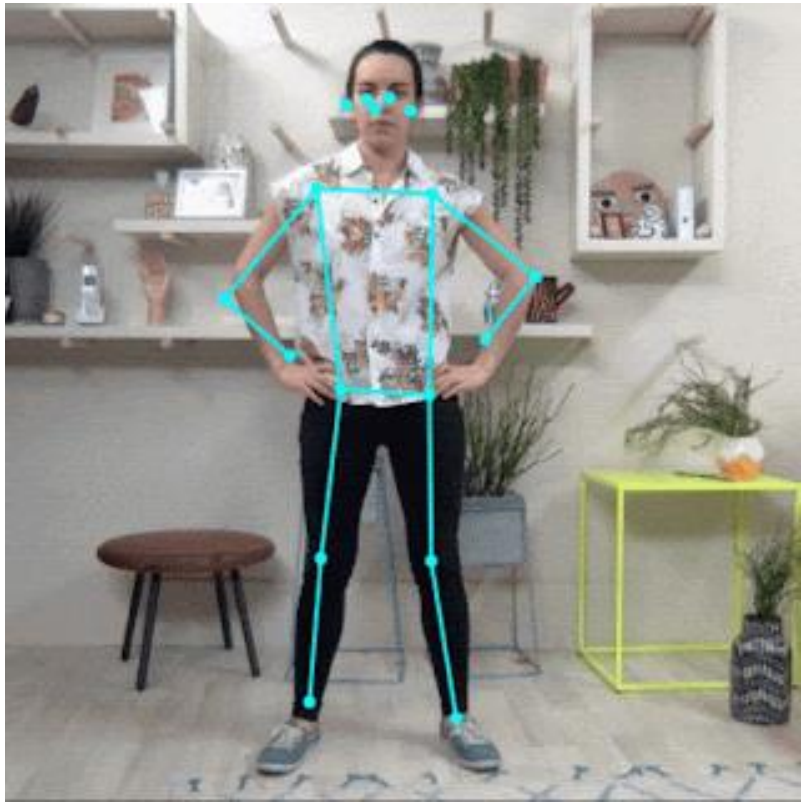
Hand Pose



Human Pose

PART 1: MODEL

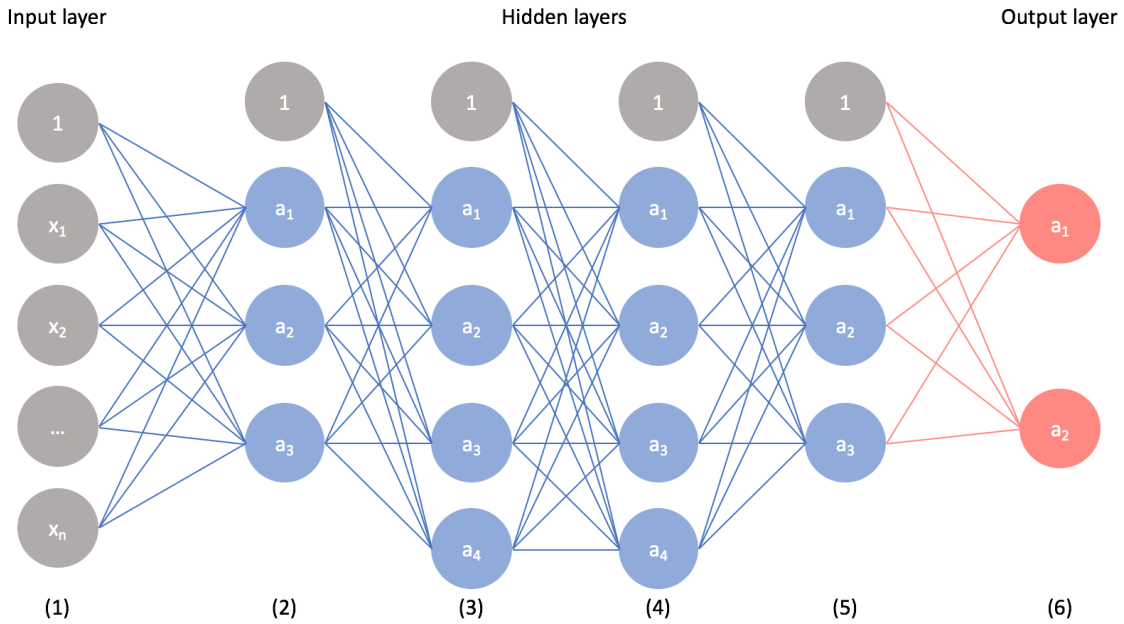
HUMAN POSE ESTIMATION

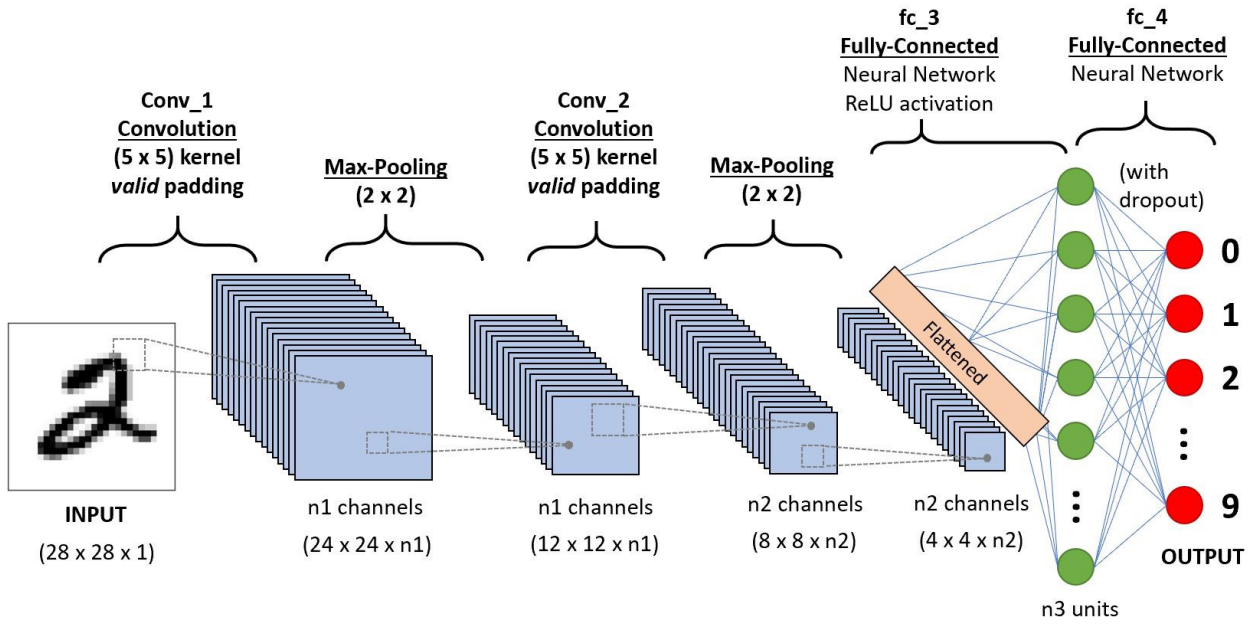


Source

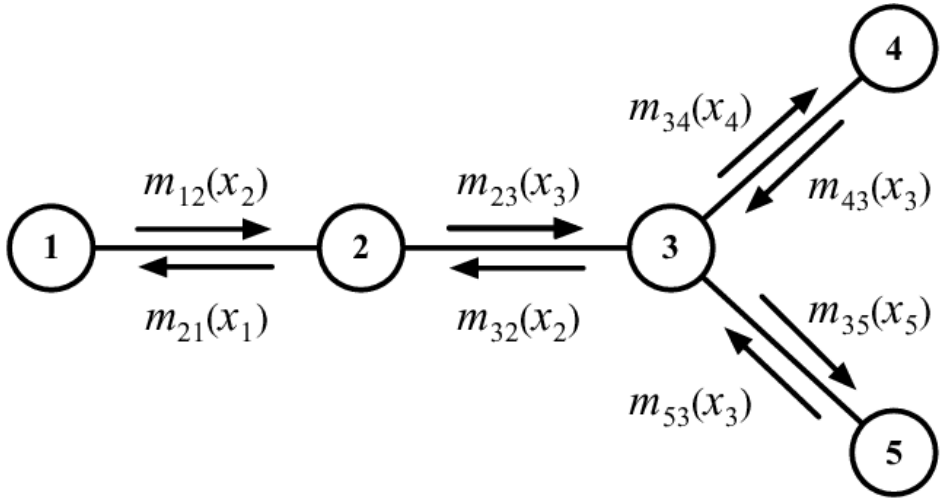
Eg: Action Recogniton

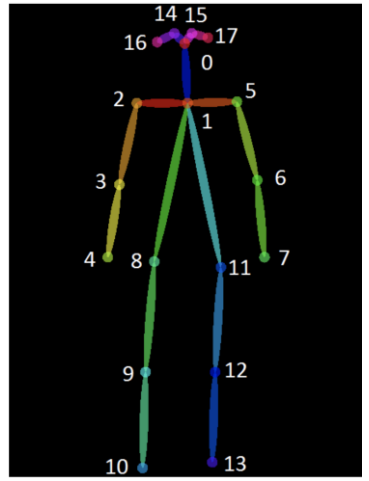
BACKGROUND: DEEP LEARNING

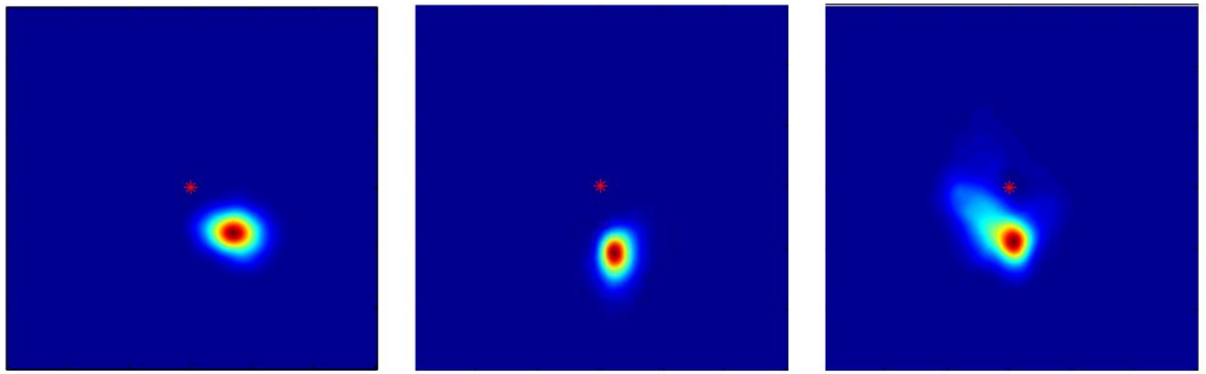


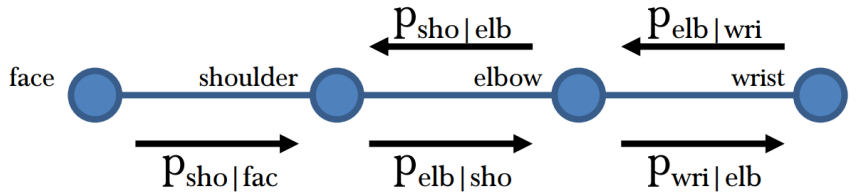


BELIEF PROPAGATION





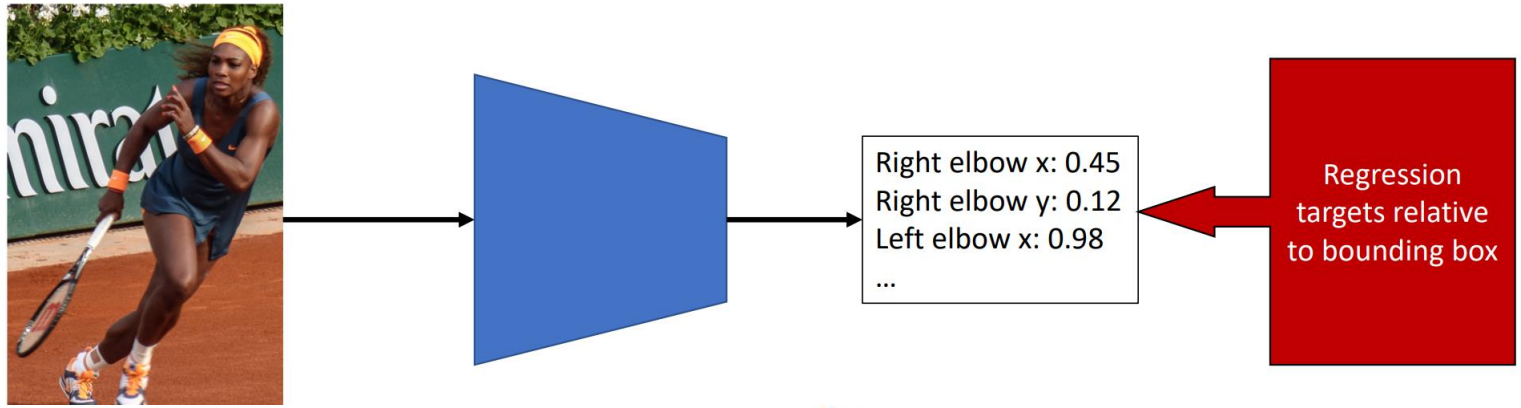


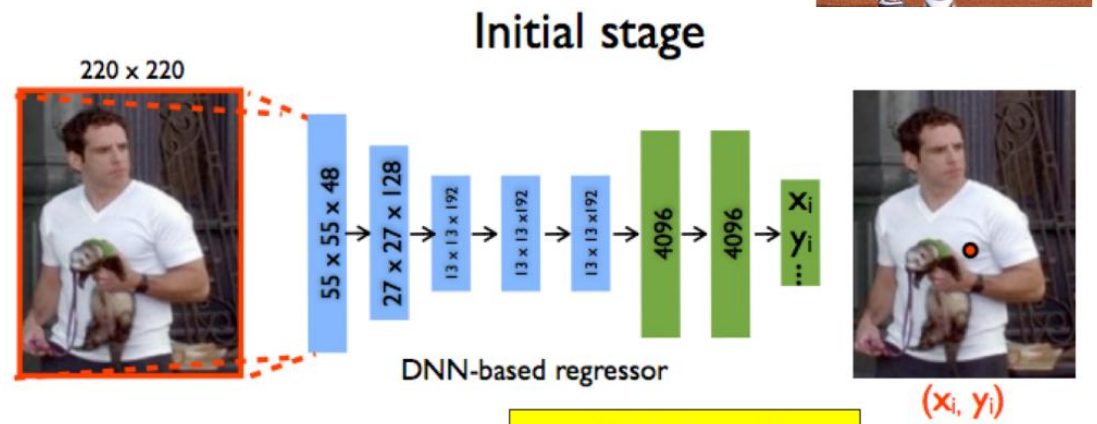


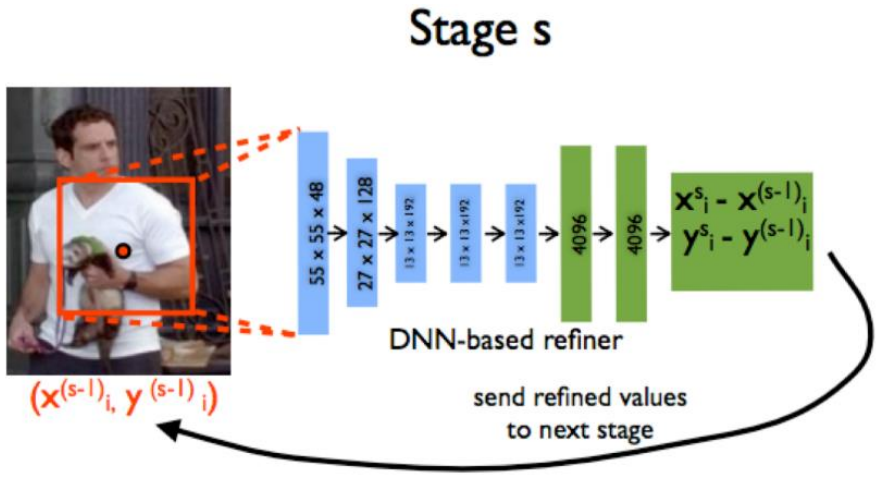
More on belief propagation: [belief_propagation.pdf (emtiyaz.github.io)]

Jain et al., "Learning Human Pose Estimation Features with Convolutional Networks", ICLR 2014
Also see: Tompson et al., "Joint Training of a Convolutional Network and a Graphical Model for Human Pose Estimation", NeurIPS 2014

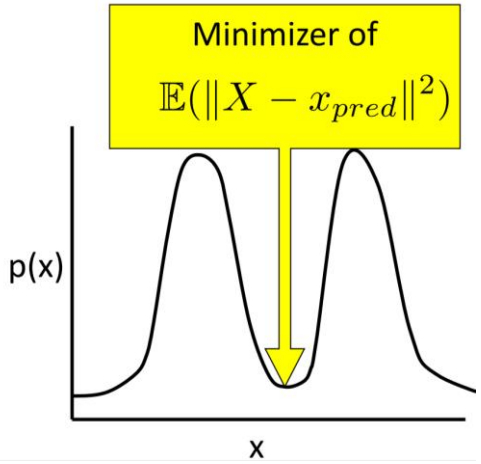
DEEP POSE







Multimodal distributions?

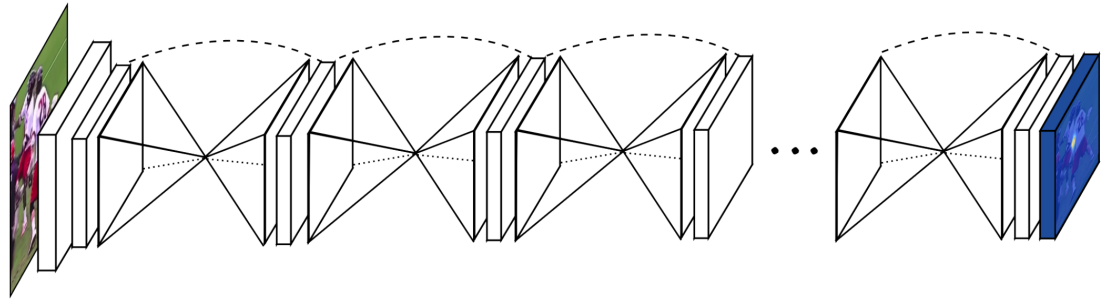


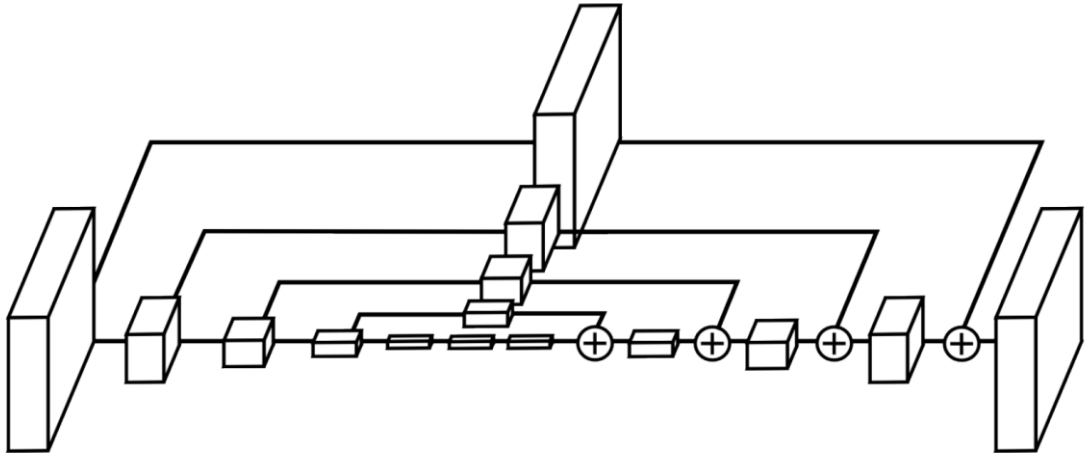
Additional issue…
Do scalar (vector) quantities reflect keypoints accurately?

Source

Toshev and Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks", CVPR 2014

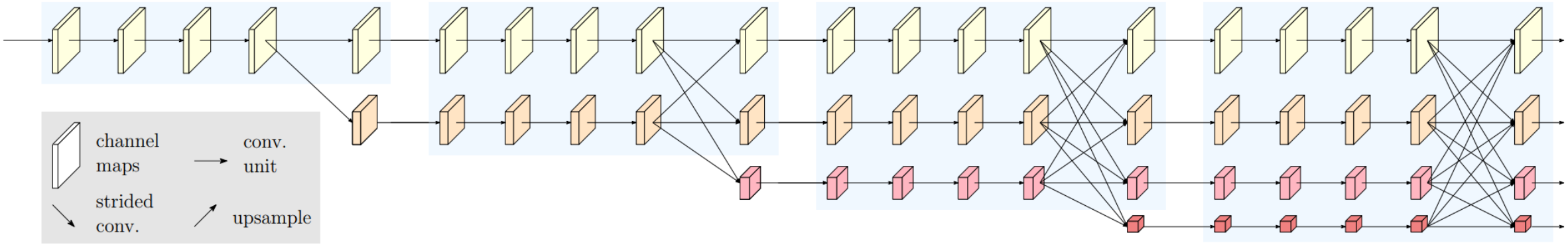
STACKED HOURGLASS

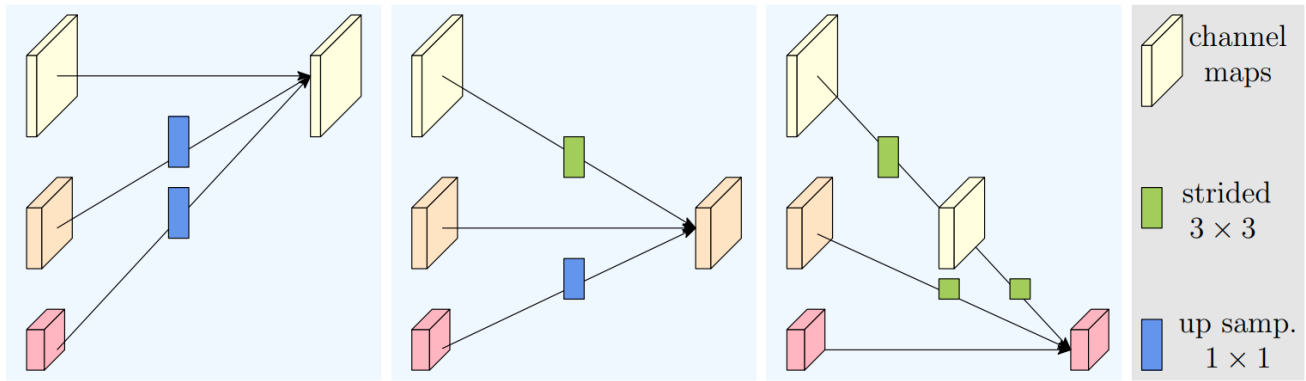




Newell et al., “Stacked Hourglass Networks for Human Pose Estimation”, ECCV 2016

HIGH RESOLUTION NETWORK





Wang et al., "Deep High-Resolution Representation Learning for Visual Recognition", T-PAMI 2020

OPEN POSE

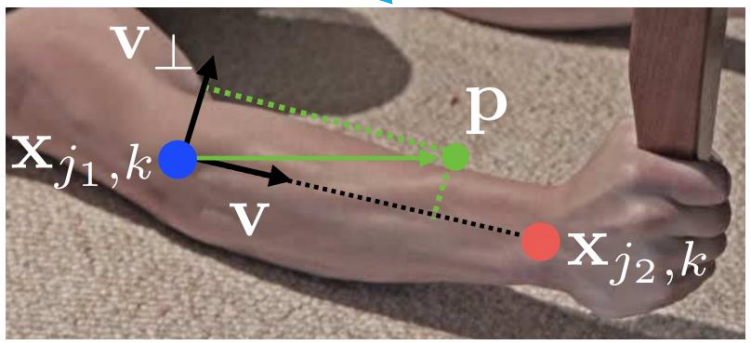
(a) Input Image

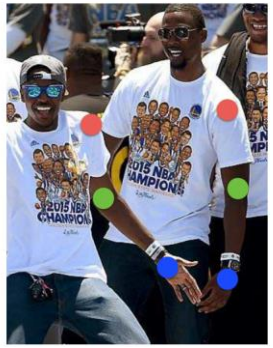
(b) Part Confidence Maps

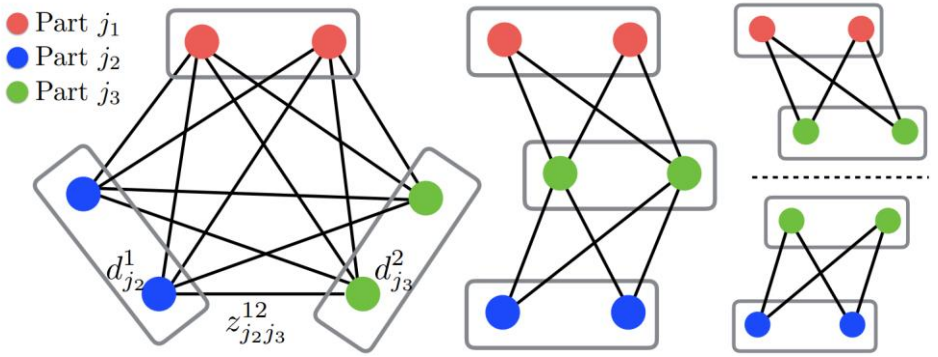
(c) Part Affinity Fields

(d) Bipartite Matching

(e) Parsing Results





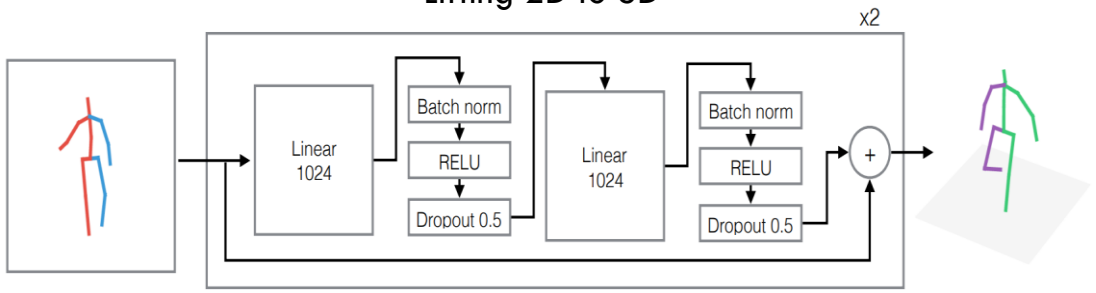


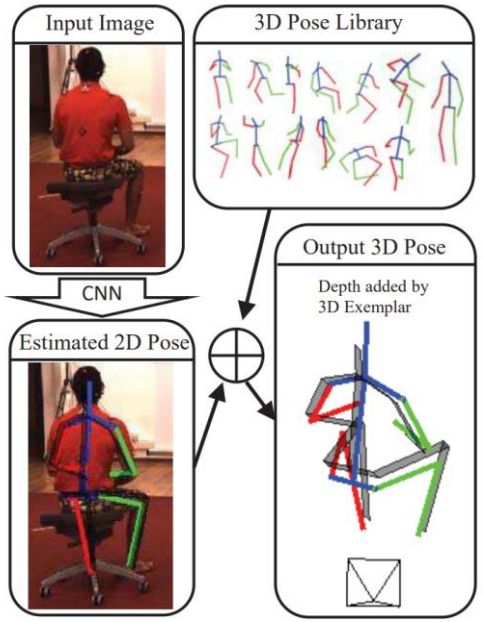
Cao et al., "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", T-PAMI

Also see: Kreiss et al., "OpenPifPaf: Composite Fields for Semantic Keypoint Detection and Spatio-Temporal Association", IEEE Transactions – ITS 2021

3D HUMAN POSE ESTIMATION

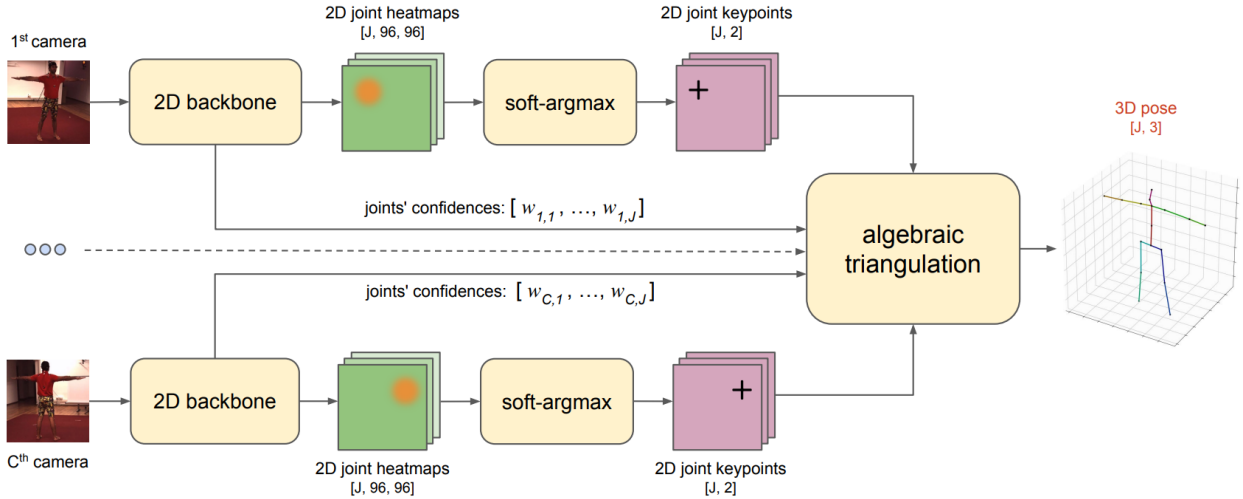
Lifting 2D to 3D





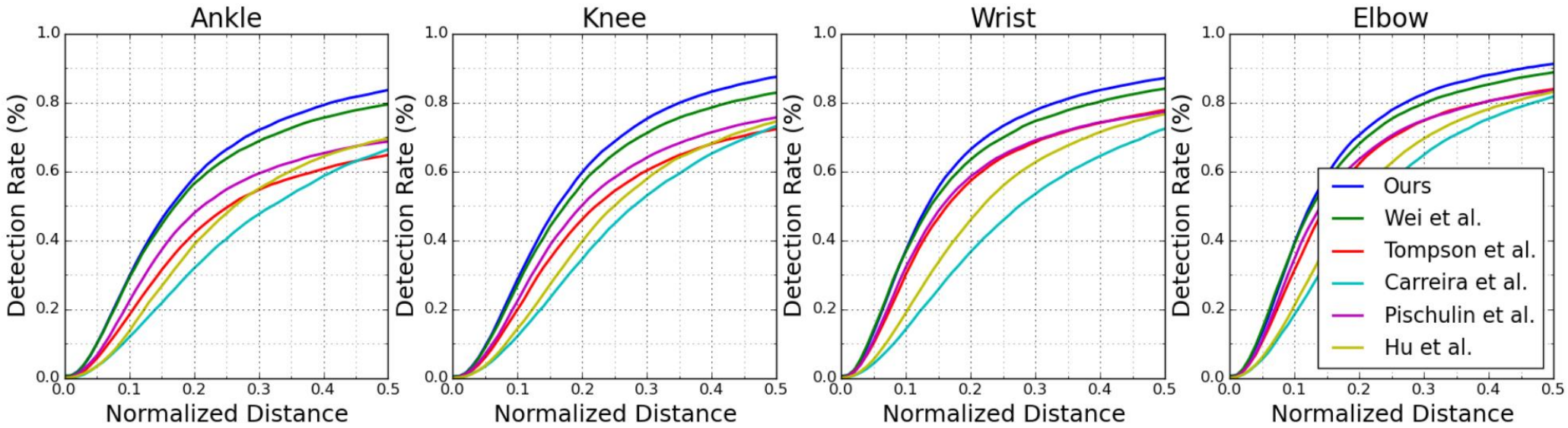
Martinez et al., "A simple yet effective baseline for 3d human pose estimation", ICCV 2017
Chen and Ramanan, "3D Human Pose Estimation = 2D Pose Estimation + Matching", CVPR 2017

Triangulation using multiple views



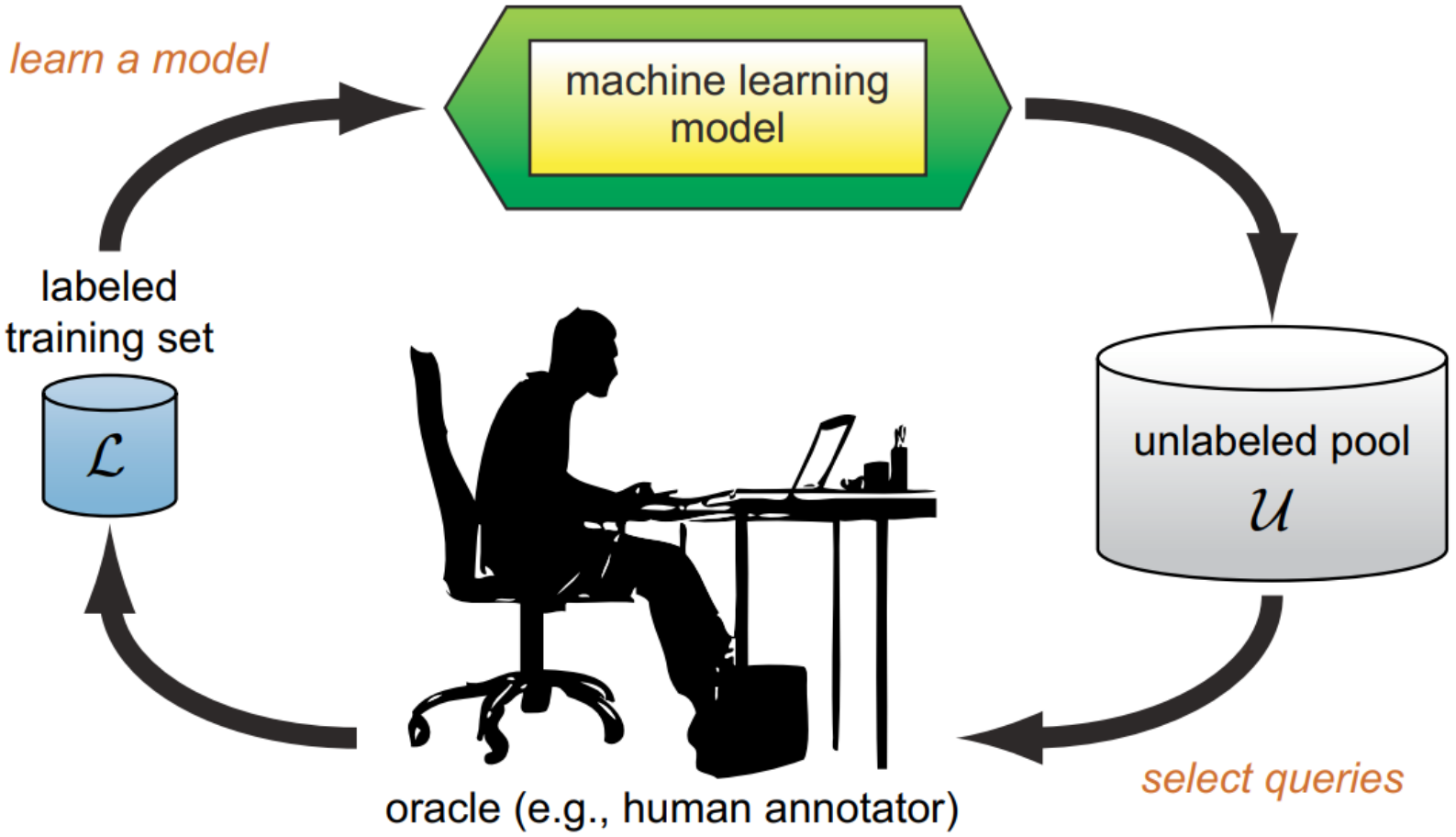
Iskakov et al., "Learnable Triangulation of Human Pose", ICCV 2019

EVALUATION METRIC: PERCENTAGE CORRECT KEYPOINTS



PART 2: DATA

WE HAVE THE MODEL. GREAT!
BUT HOW DO WE CURATE OUR DATASET?



http://burrsettles.com/pub/settles.activelearning.pdf

WHY ACTIVE LEARNING?

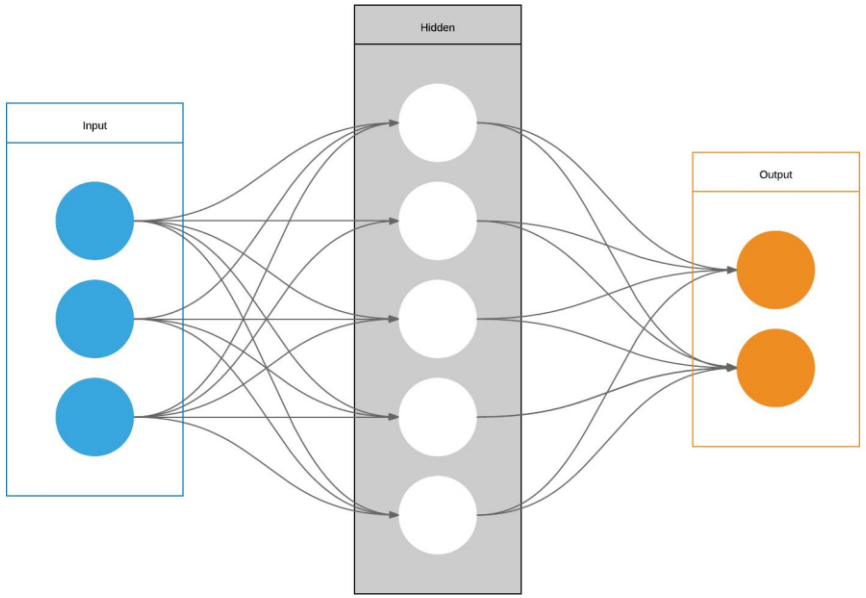
An *intelligent* way of curating datasets

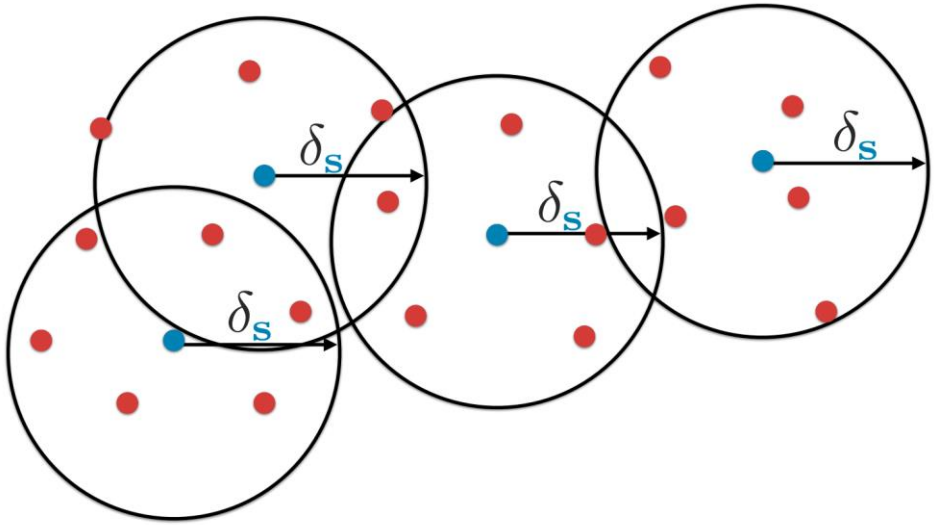
Cost Savings
Smaller datasets
Lower annotation costs

Faster Deployment
Faster annotation
Faster training

Bias Elimination
Reduces overlap / oversampling

CORE-SET

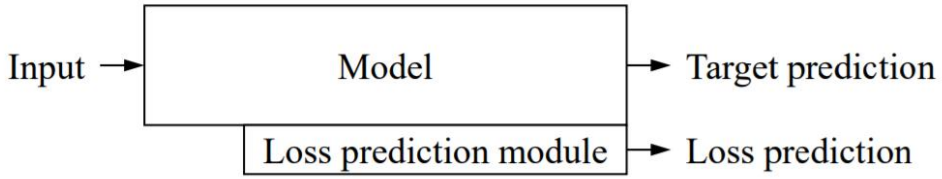




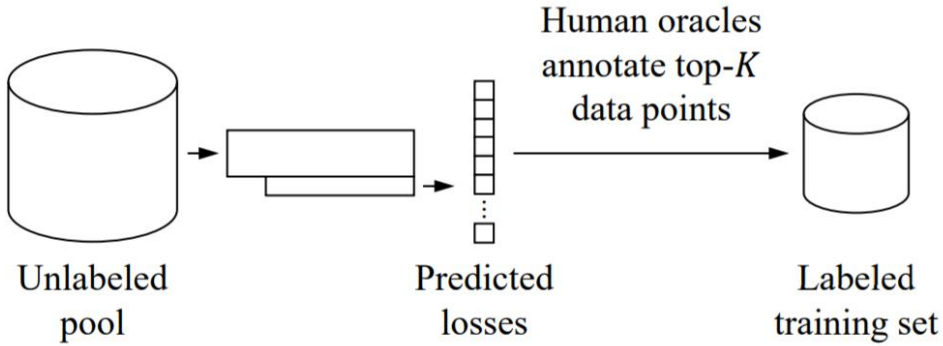
$$\left| \frac{1}{n} \sum_{i \in [n]} l(\mathbf{x}_i, y_i, A_{\mathbf{S}}) - \frac{1}{|\mathbf{S}|} \sum_{j \in \mathbf{S}} l(\mathbf{x}_j, y_j; A_{\mathbf{S}}) \right| \leq \mathcal{O}(\delta_{\mathbf{S}}) + \mathcal{O}\left(\sqrt{\frac{1}{n}}\right)$$

Sener and Savarese, “Active Learning for Convolutional Neural Networks: A Core-Set Approach”, ICLR 2018

LEARNING LOSS



(a) A model with a loss prediction module

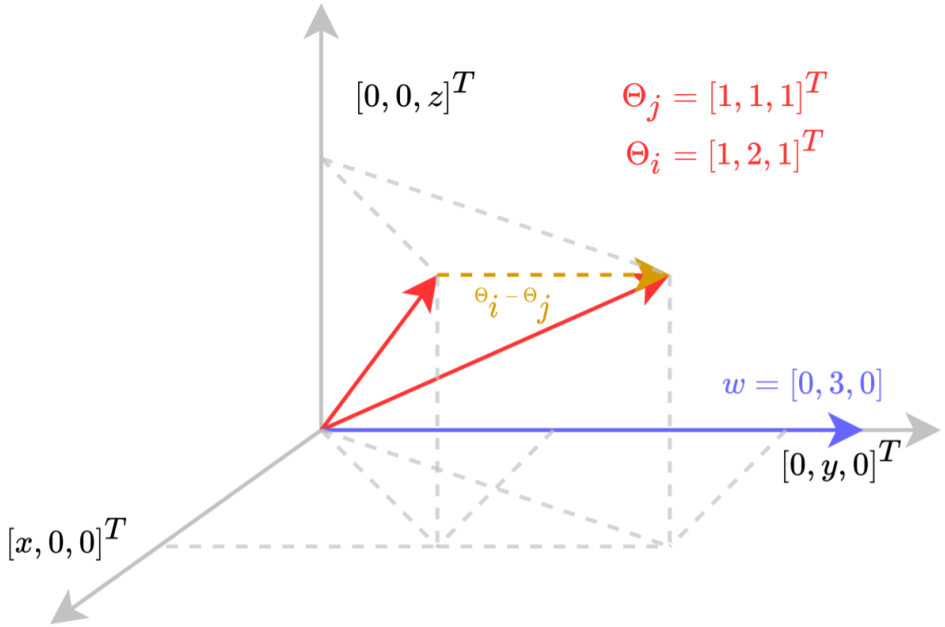


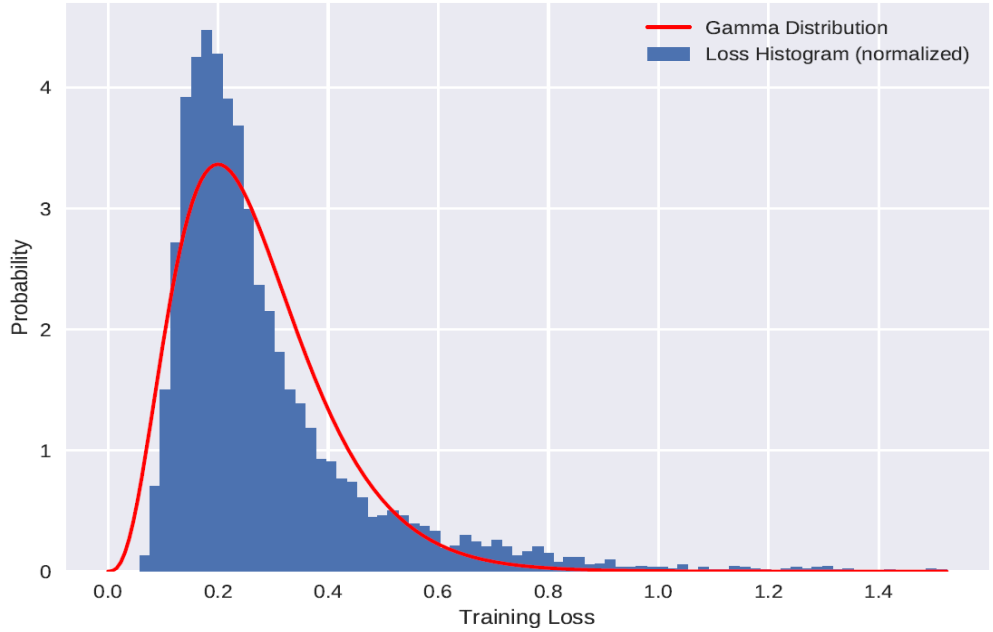
$$\mathbb{L}_{loss}(w, \theta_i, \theta_j) = \mathrm{KL}(p||q) = p_i \log \frac{p_i}{q_i} + p_j \log \frac{p_j}{q_j}$$

Yoo and Kweon, "Learning Loss For Active Learning", CVPR 2019
Shukla and Ahmed, "A Mathematical Analysis of Learning Loss for Active Learning in Regression", CVPRW 2021

LEARNING LOSS

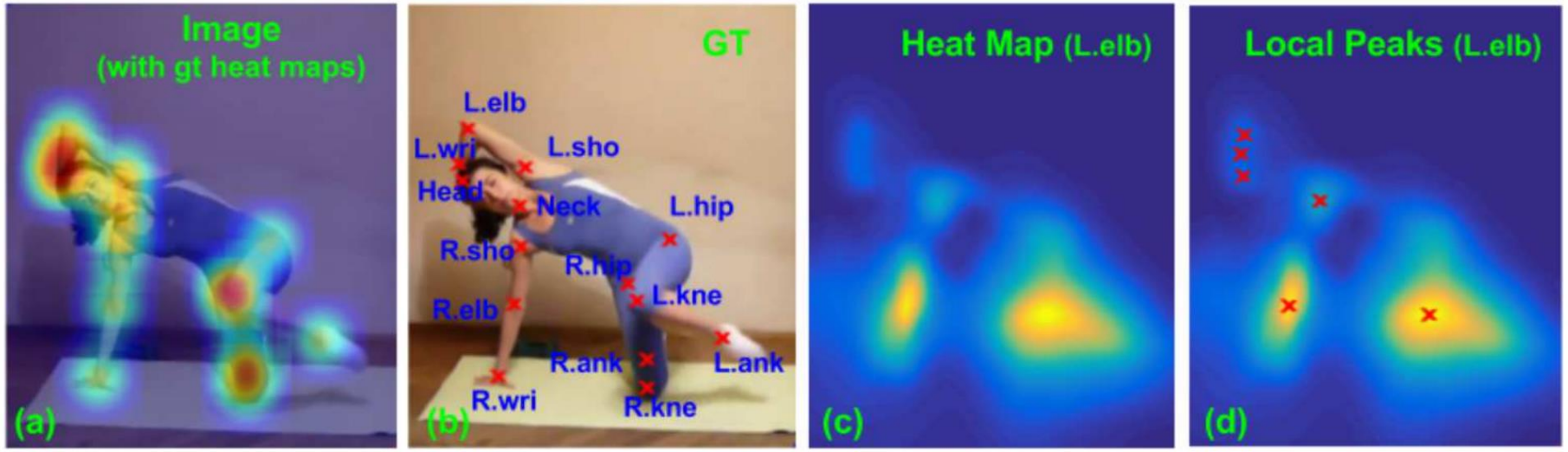
$$\nabla_w \mathbb{L}_{loss}(w, \theta_i, \theta_j) = (q_i - p_i)(\theta_i - \theta_j)$$
$$\nabla_\theta \mathbb{L}_{loss}(w, \theta_i, \theta_j) = (q_i - p_i)w$$





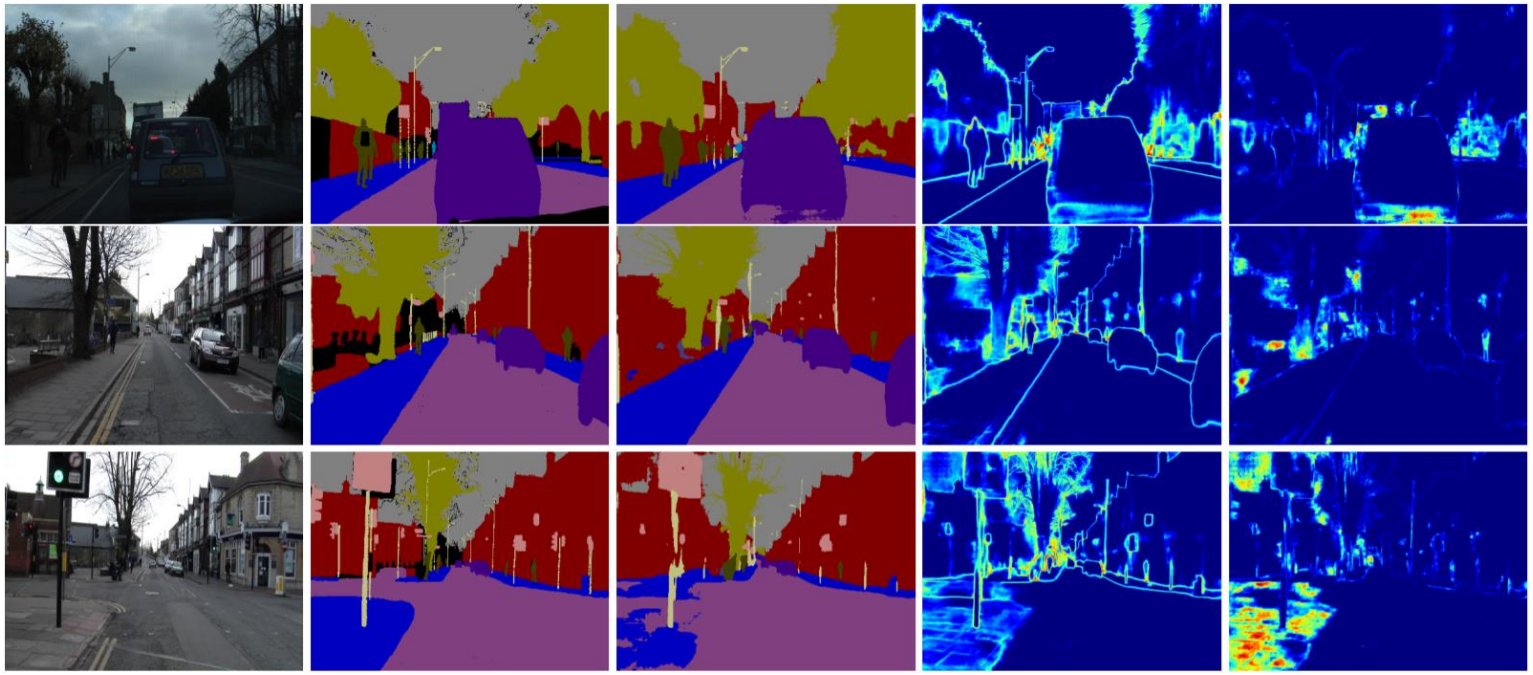
Yoo and Kweon, "Learning Loss For Active Learning", CVPR 2019
Shukla and Ahmed, "A Mathematical Analysis of Learning Loss for Active Learning in Regression", CVPRW 2021

MULTI-PEAK ENTROPY



Liu and Ferrari, "Active Learning for Human Pose Estimation", ICCV 2017

BAYESIAN UNCERTAINTY



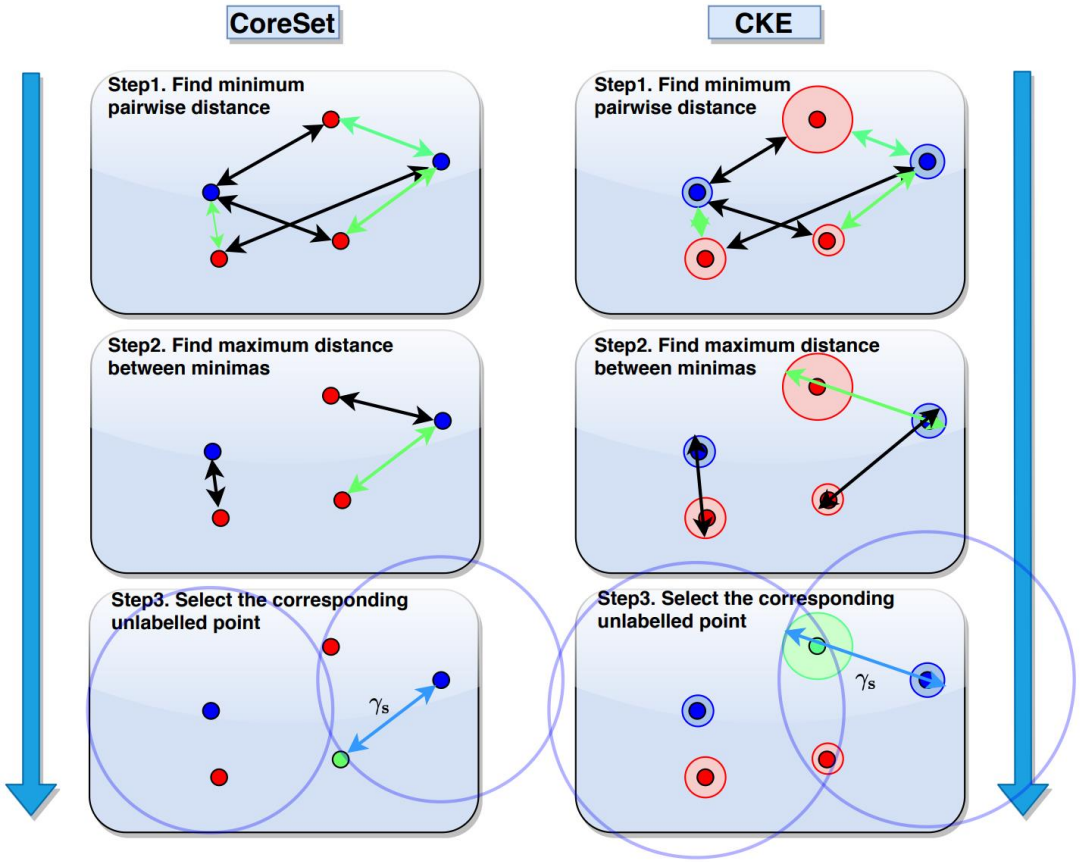
$$Var(y) \approx \frac{1}{T}\sum_{t=1}^{T}\hat{y}_t^2 - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{y}_t\right)^2 + \frac{1}{T}\sum_{t=1}^{T}\hat{\sigma}_t^2$$

Epistemic

Aleatoric

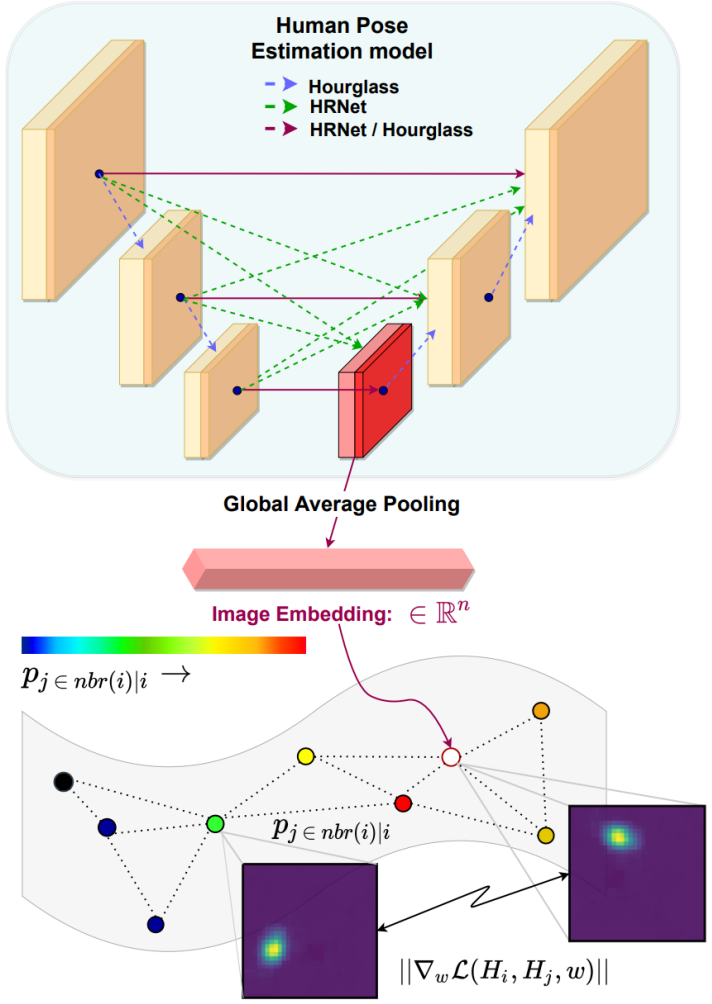
Kendall and Gal, "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?", NeurIPS 2017

BAYESIAN UNCERTAINTY + CORE-SET

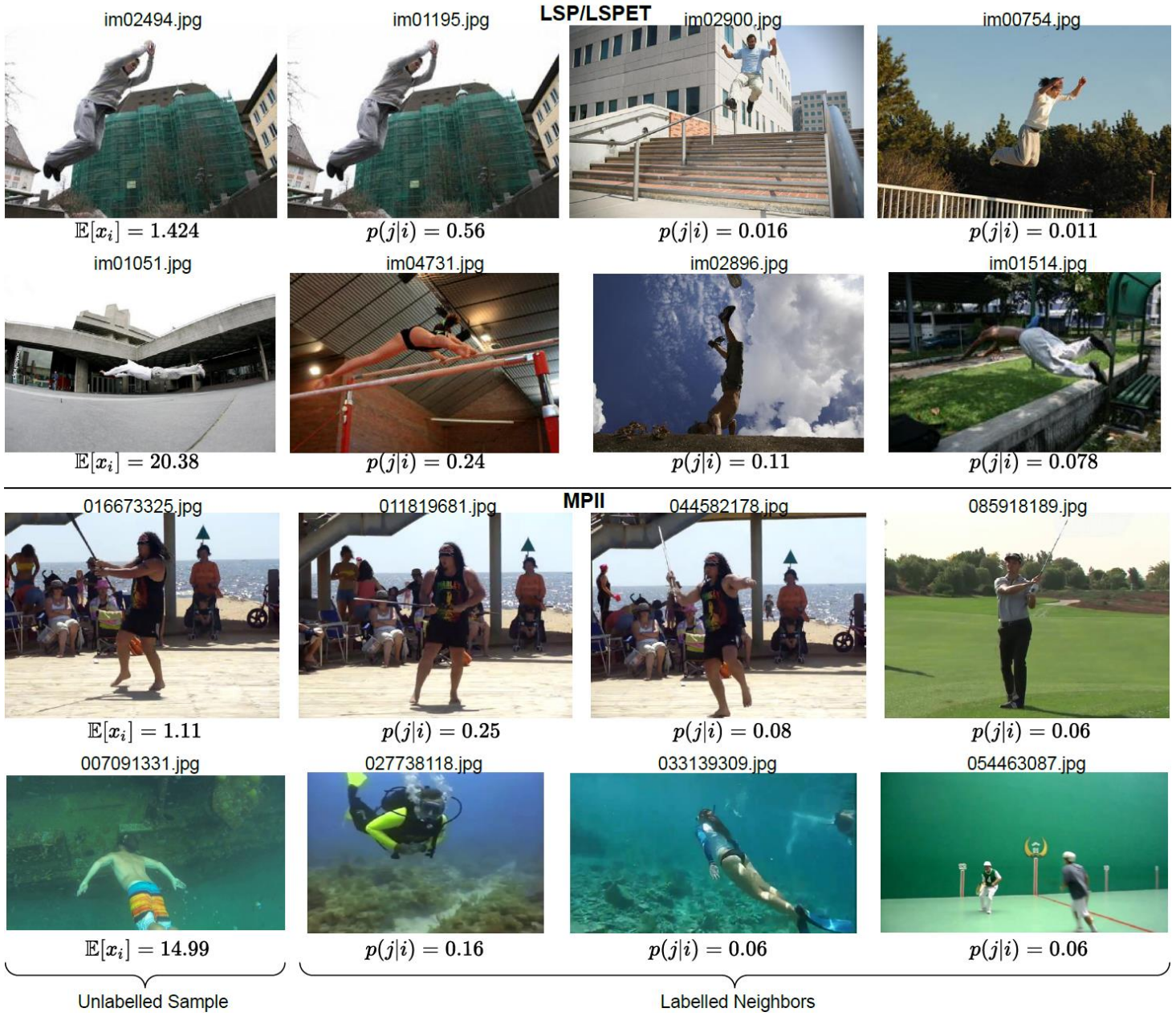


Caramalau et al., "Active Learning for Bayesian 3D Hand Pose Estimation", WACV 2021

EXPECTED GRADIENT LENGTH



Megh Shukla, "Bayesian Uncertainty and Expected Gradient Length – Regression: Two Sides of the Same Coin?", WACV 2022



Megh Shukla, "Bayesian Uncertainty and Expected Gradient Length – Regression: Two Sides of the Same Coin?", WACV 2022

PART 3: DEMO

jupyter

PREREQUISITES – NVIDIA GPU