G33:Single Person 2-D Pose Estimation from Images using CNNs

Ishan Bhatt (ivbhatt) | Aishwarya Seth (aseth) | Akhil Bommadevara (abommad)

Motivation:

Human Pose-Estimation(HPE) has a wide range of applications:

- Commercial(Person tracking, Activity Recognition)
- Entertainment (Animation, VR)
- Medical (Physical Therapy, Assisted Living)

Problem Definition:

Attempted to build an HPE system that:

- Takes RGB images as input
- Assumes images to have a clear subject person without significant occlusions
- Detects the key-points(KPs) of the subject person

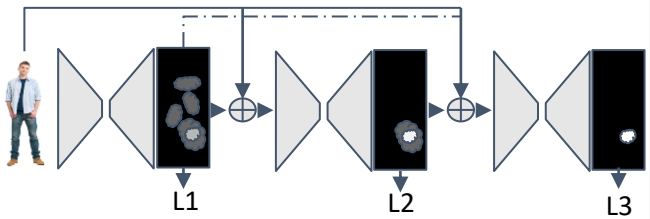
Data:

- MPII Human Pose dataset (~25K images) of ~40K people with annotated body joints **(left/right) (ankle, knee, hip, shoulder, elbow, wrist)** + head + thorax(upper-neck) + pelvis
- ~4K people (bounding-boxes) with all 12 key-points visible
- ~2.5K high-quality images (subject person is clear)

Augmentation:

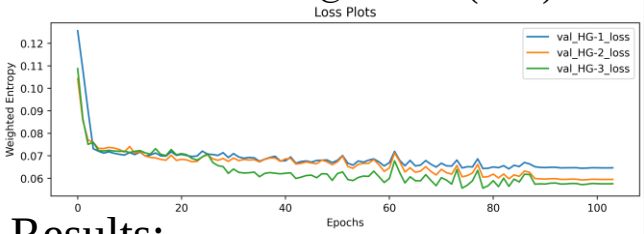
- 50% FlipLR
- Upto: 10% scale; 5% translate; 5° rotate
- Random-order
- 8x the data!

Model:



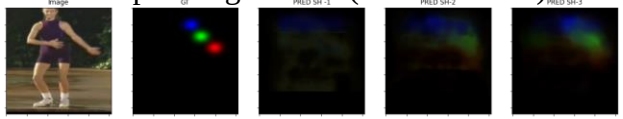
Training:

- ~12K train images (128x128); 10% val
- ~2 hrs. (~100 epochs) on Nvidia Tesla T4
- ~250K parameters
- Our PCKh@0.5 = 81.5;
- State-of-the-Art PCKh@0.5=92.5 (2019)



Results:

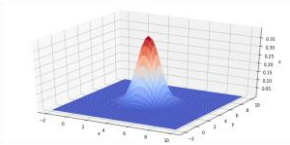
- Heat-maps being refined (left-hand KPs)



- Predictions on unseen images

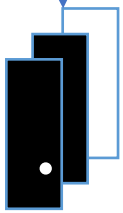
Pipeline:

Images with multiple people; occlusions

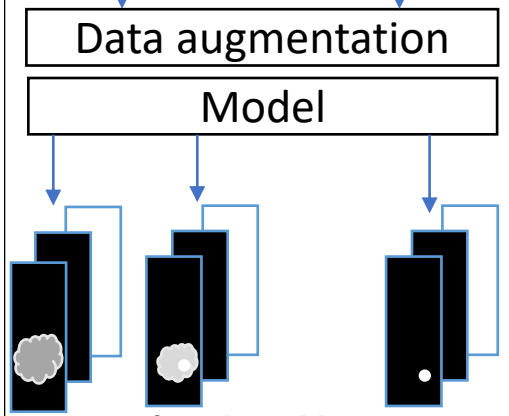


Gaussian Peaks!
μ = KP coordinate;
σ = 15 px

Person-images fully-visible
12 Key-Points (KP)



13-channel heatmaps



K set of 13-channel heatmaps;
K = number of modules