Learning 3D Human Pose Estimation

Monocular 3D Human Pose Estimation In The Wild Using Improved CNN Supervision

Seminar on Current Works in Computer Vision

Presenter: Ikrima Bin Saeed

July 26, 2017

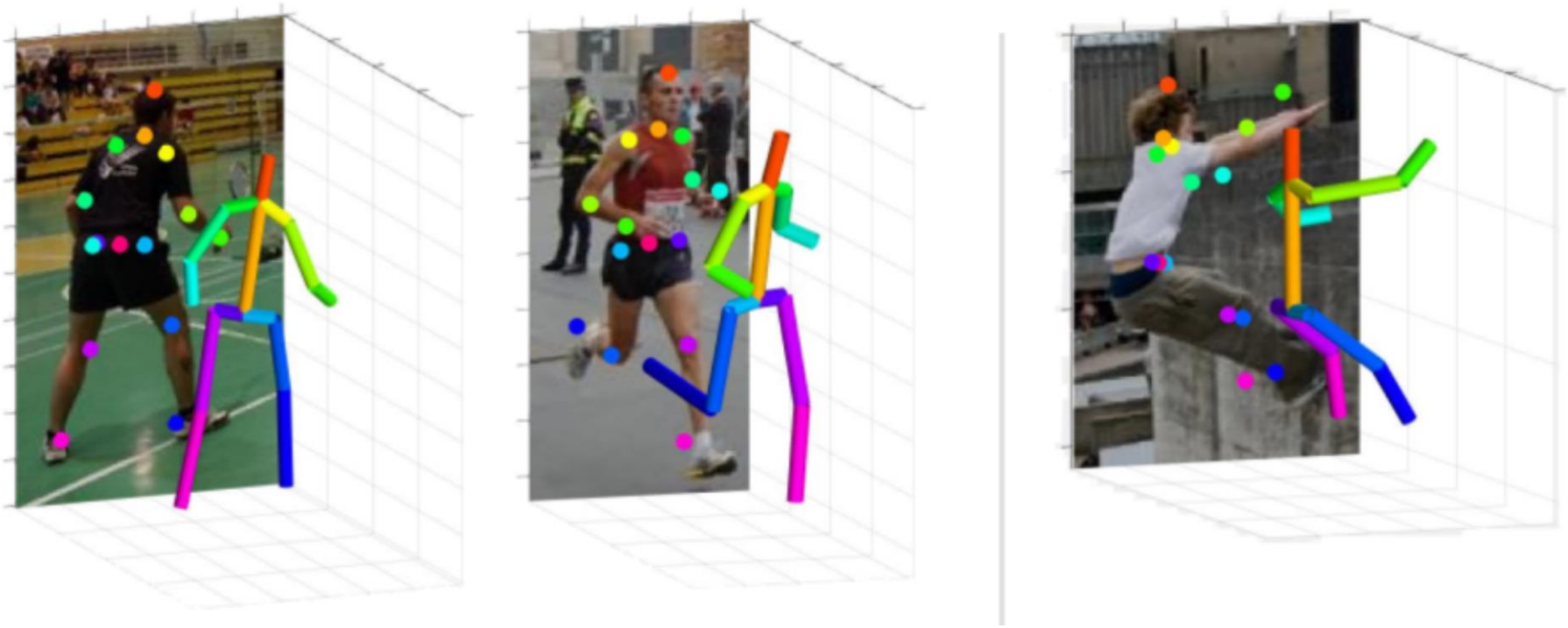
The Focus of My Presentation

- Introduction
- Common Approaches
- Current Approach
  - Bounding Box
  - 3D Pose Prediction
  - Global 3D Pose Computation
- Data Set Contribution
- Evaluation
  - Perspective Correction
  - Multi-level Corrective Skip Connections
  - New Data Set
  - Transfer Learning
  - Multi-modal Prediction
  - State of the Art
- Summary

The Focus of My Presentation

- Introduction
- Common Approaches
- Current Approach
  - Bounding Box
  - 3D Pose Prediction
  - Global 3D Pose Computation
- Data Set Contribution
- Evaluation
  - Perspective Correction
  - Multi-level Corrective Skip Connections
  - New Data Set
  - Transfer Learning
  - Multi-modal Prediction
  - State of the Art
- Summary

3D Human Pose Estimation Using a Single Image



The Focus of My Presentation

- Introduction
- Common Approaches
- Current Approach
  - Bounding Box
  - 3D Pose Prediction
  - Global 3D Pose Computation
- Data Set Contribution
- Evaluation
  - Perspective Correction
  - Multi-level Corrective Skip Connections
  - New Data Set
  - Transfer Learning
  - Multi-modal Prediction
  - State of the Art
- Summary

Common Approaches

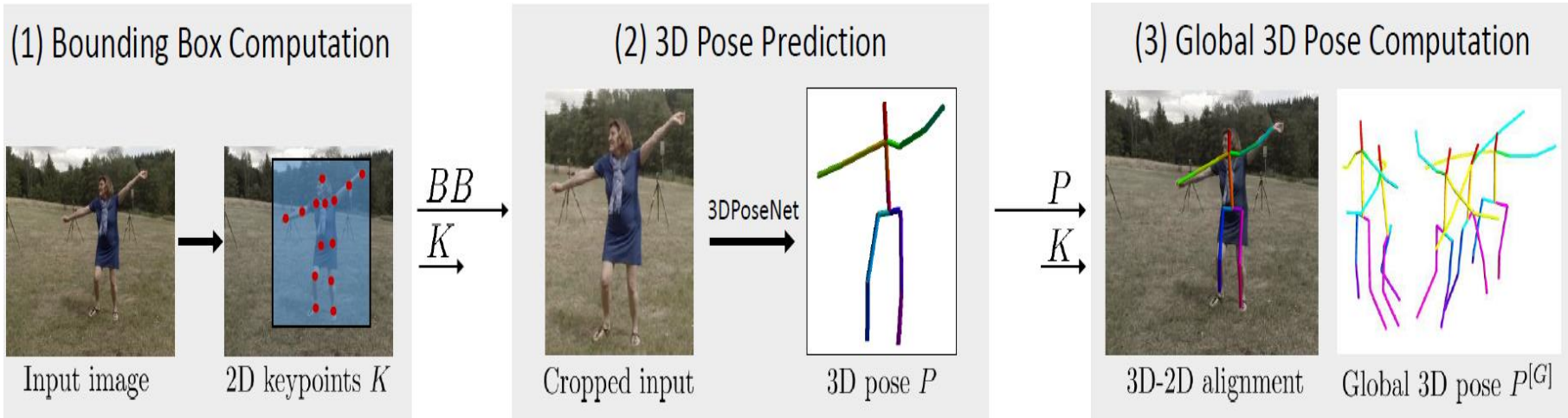
- Fit 3D models to 2D key points
  - Computationally expensive
  - Unstable under depth ambiguities

- Direct CNN-based 3D regression
  - Performance far below than that of CNN based methods for 2D human pose

The Focus of My Presentation

- Introduction
- Common Approaches
- Current Approach
  - Bounding Box
  - 3D Pose Prediction
  - Global 3D Pose Computation
- Data Set Contribution
- Evaluation
  - Perspective Correction
  - Multi-level Corrective Skip Connections
  - New Data Set
  - Transfer Learning
  - Multi-modal Prediction
  - State of the Art
- Summary

Approach Overview

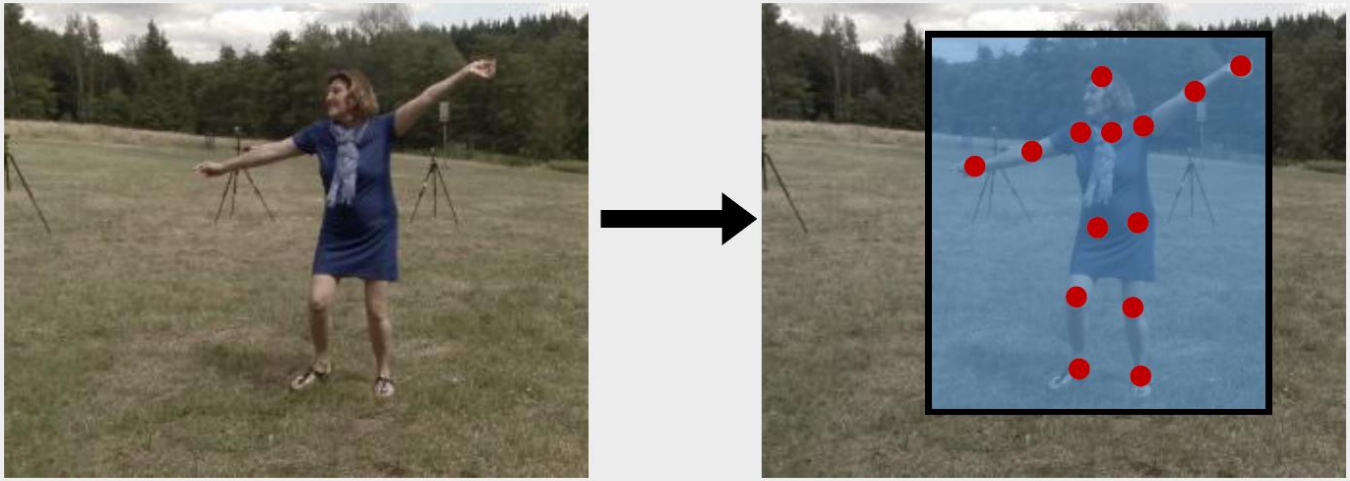


- 3 main steps from 2D input image to global 3D human pose

The Focus of My Presentation

- Introduction
- Common Approaches
- Current Approach
  - Bounding Box
  - 3D Pose Prediction
  - Global 3D Pose Computation
- Data Set Contribution
- Evaluation
  - Perspective Correction
  - Multi-level Corrective Skip Connections
  - New Data Set
  - Transfer Learning
  - Multi-modal Prediction
  - State of the Art
- Summary

Bounding Box Computation



Bounding Box Computation

Input Image

I

Input image source: ESPN Cricinfo

Bounding Box Computation

Input Image

I

Heat Map

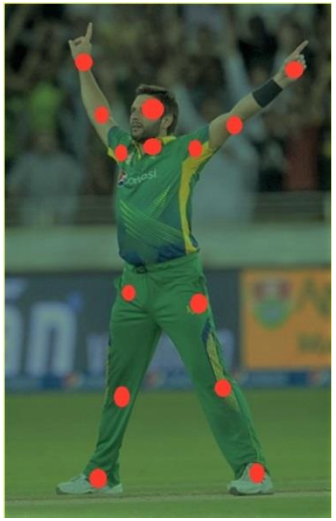
H

Input image source: ESPN Cricinfo

Bounding Box Computation

Input Image
I

Heat Map
H



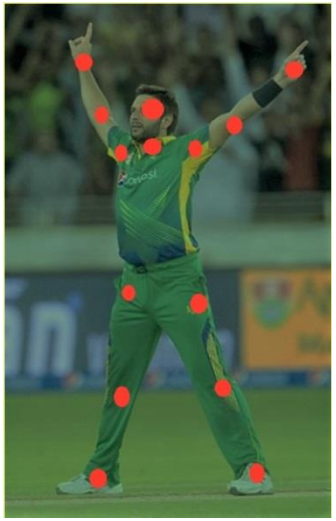
2D Joint Locations K

Input image source: ESPN Cricinfo

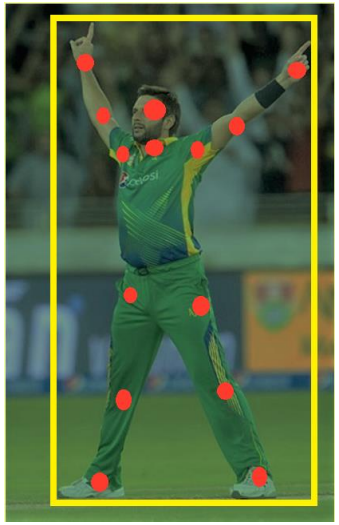
Bounding Box Computation

Input Image
I

Heat Map
H



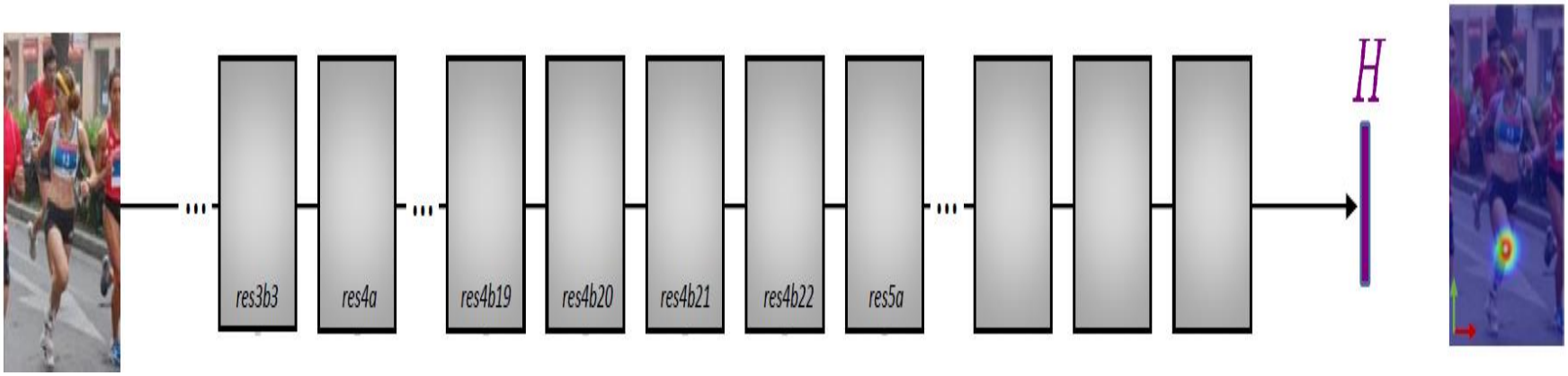
2D Joint Locations K



Bounding Box BB

Input image source: ESPN Cricinfo

2DPoseNet Architecture

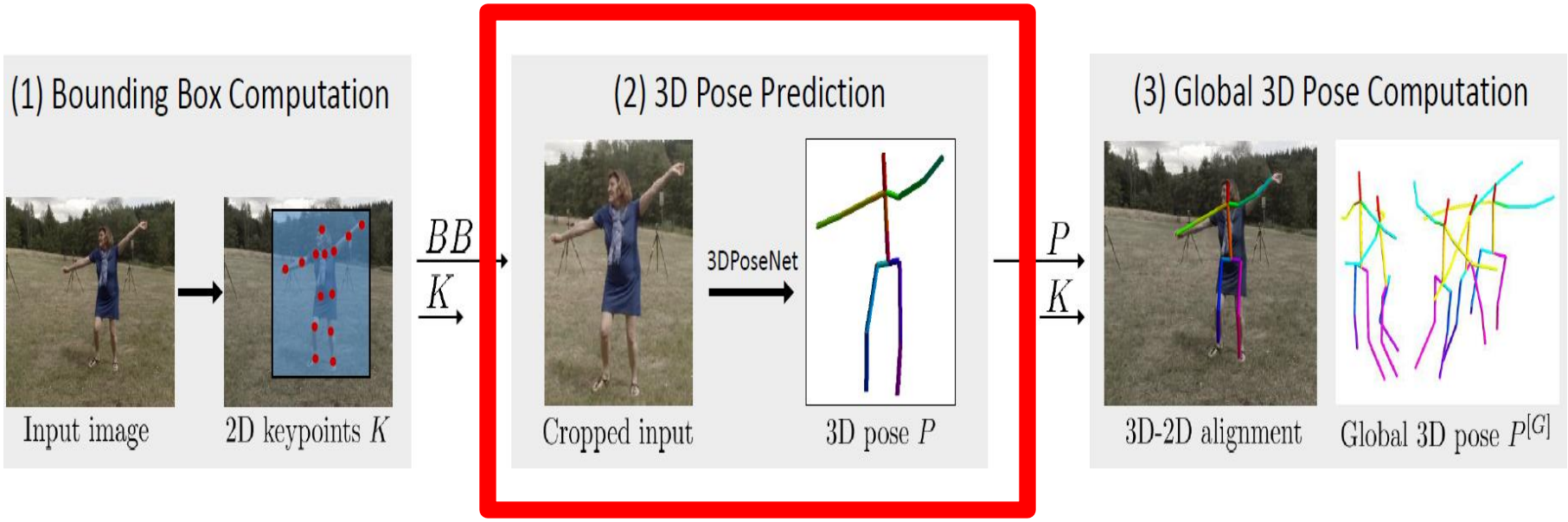


- Training data: MPII + LSP
- Architecture based on Resnet-101

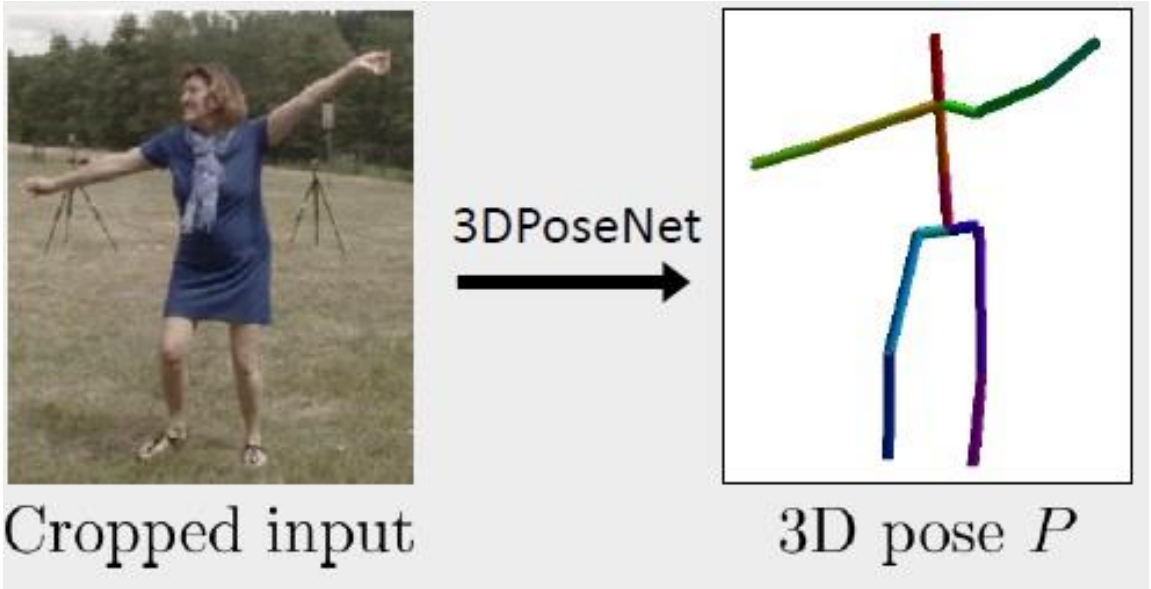
The Focus of My Presentation

- Introduction
- Common Approaches
- Current Approach
  - Bounding Box
  - 3D Pose Prediction
  - Global 3D Pose Computation
- Data Set Contribution
- Evaluation
  - Perspective Correction
  - Multi-level Corrective Skip Connections
  - New Data Set
  - Transfer Learning
  - Multi-modal Prediction
  - State of the Art
- Summary

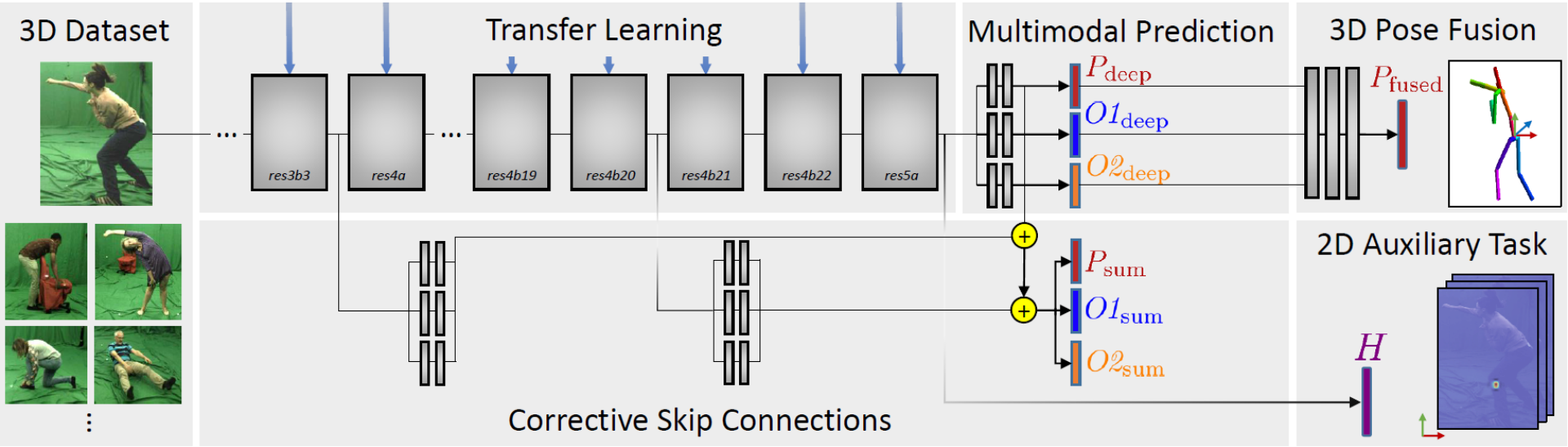
Approach Overview



3D Pose Regression



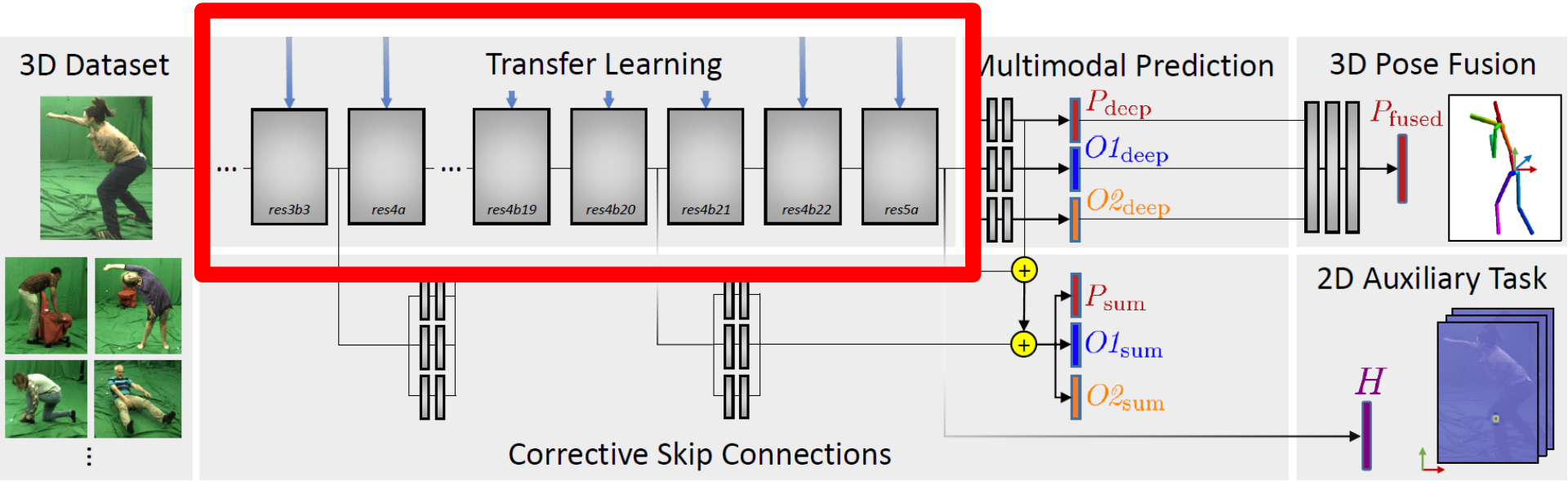
3DPoseNet Architecture



Main Components:

- CNN based on ResNet-101 with Transfer Learning
- Multi-level Corrective Skip Connections
- Multi-modal Prediction
- 3D Pose Fusion
- 2D Pose Estimation

3DPoseNet Architecture

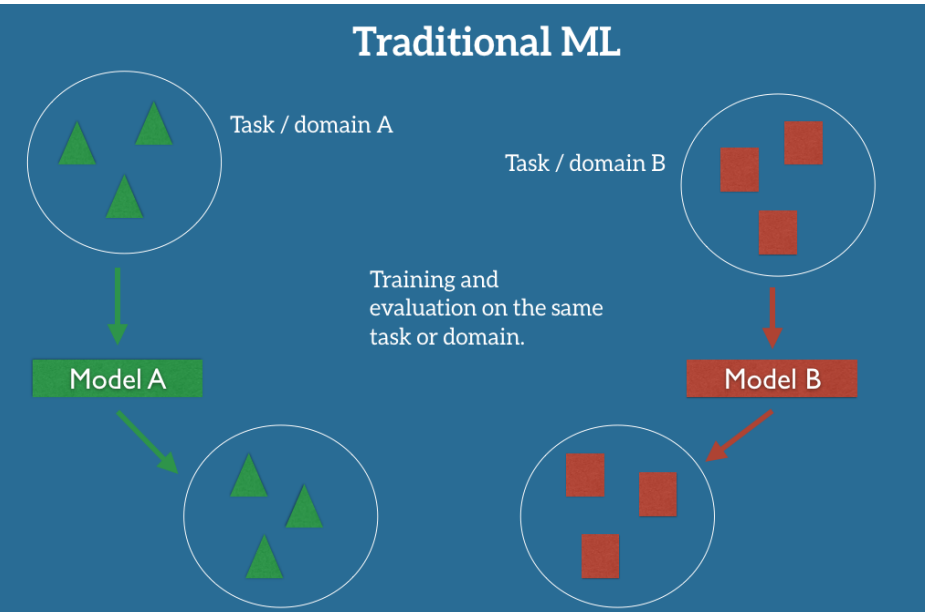


Main Components:

- CNN based on ResNet-101 with Transfer Learning
- Multi-level Corrective Skip Connections
- Multi-modal Prediction
- 3D Pose Fusion
- 2D Pose Estimation

Transfer Learning

Transfer the knowledge acquired from one task to help solve a new task



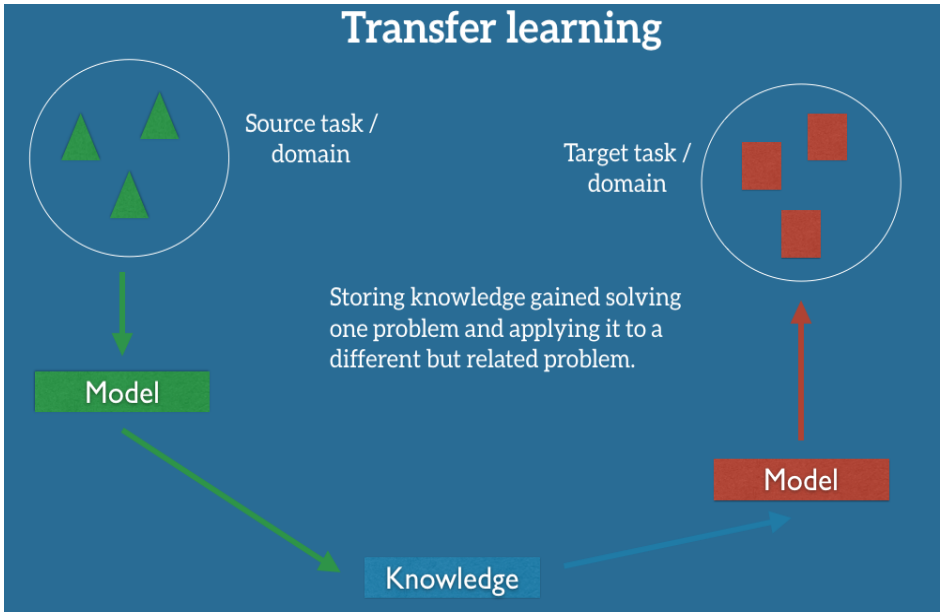
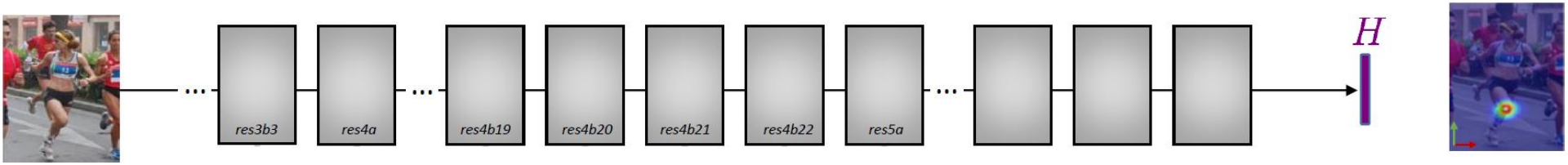


Image Source: http://ruder.io/transfer-learning/

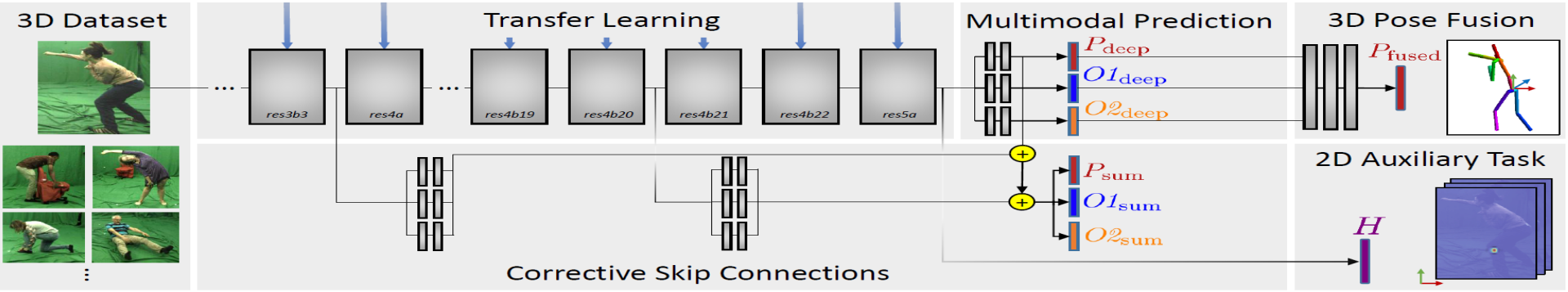
Transfer Learning in Our Case

ResNet-101

2DPoseNet

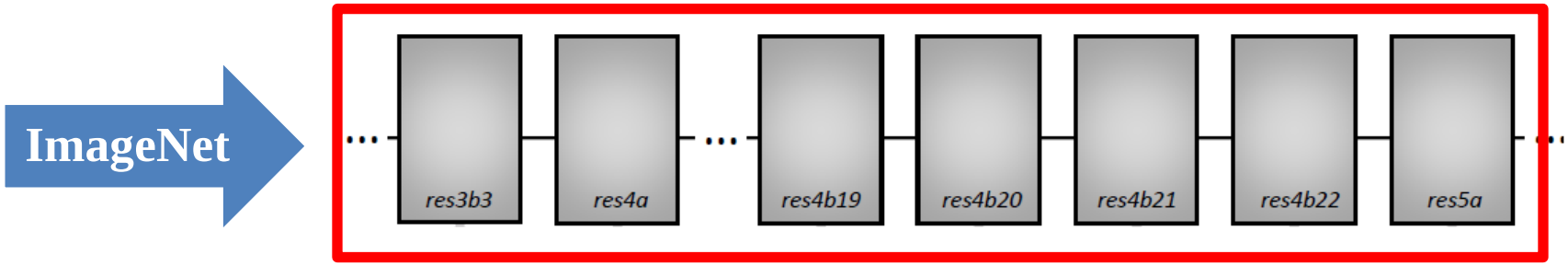


3DPoseNet

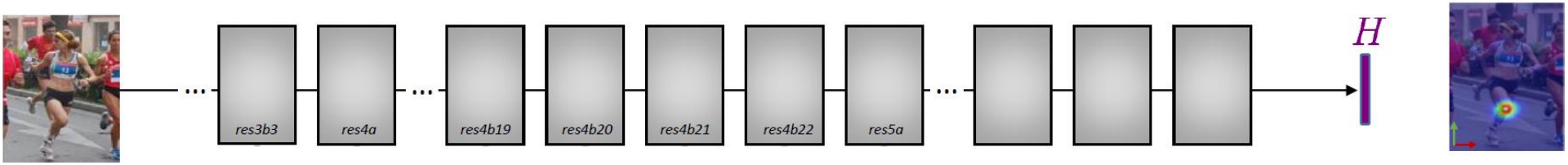


Transfer Learning: Train ResNet-101

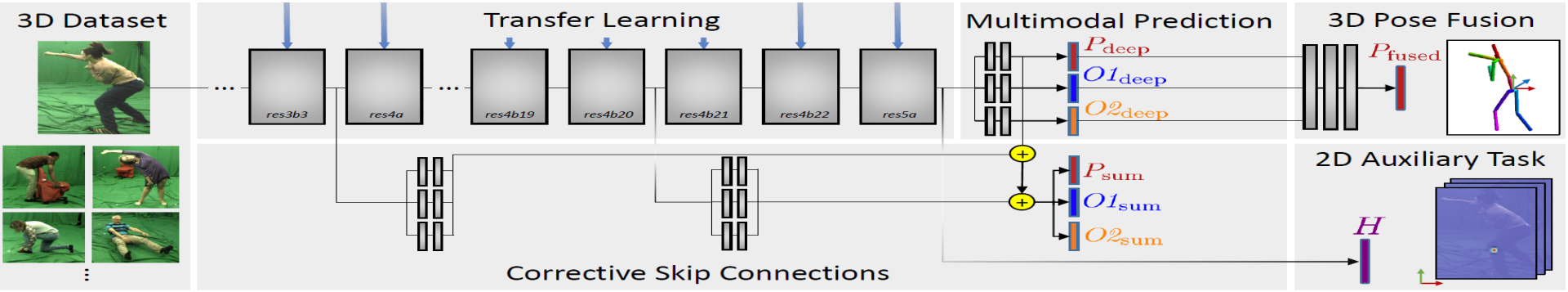
ResNet-101



2DPoseNet

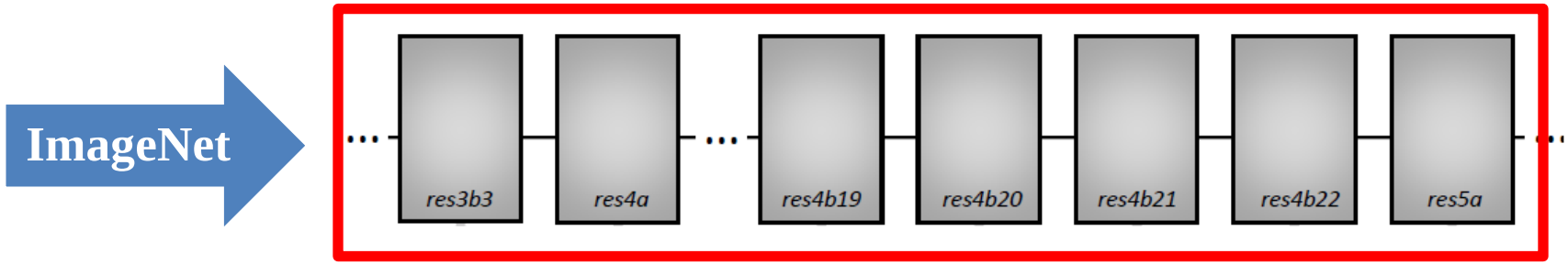


3DPoseNet

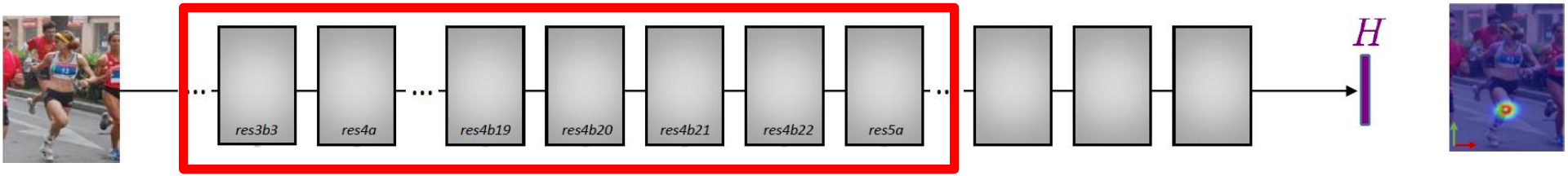


Transfer Learning: Identify Similar Structure

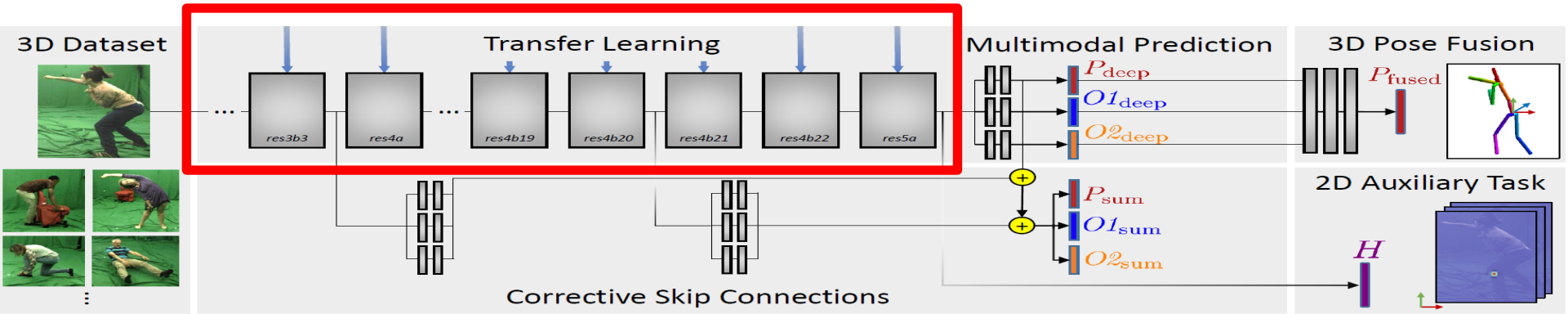
ResNet-101



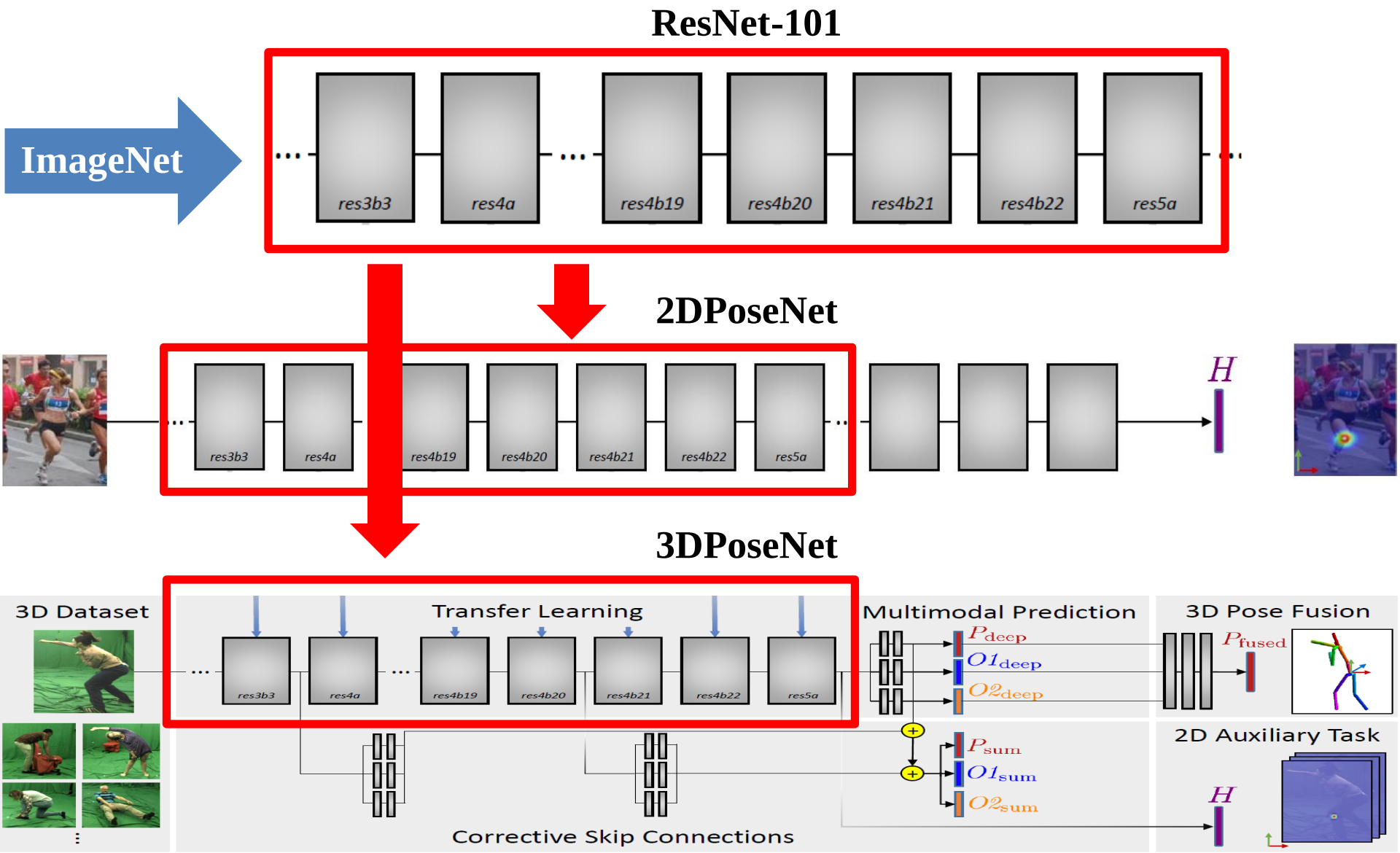
2DPoseNet



3DPoseNet



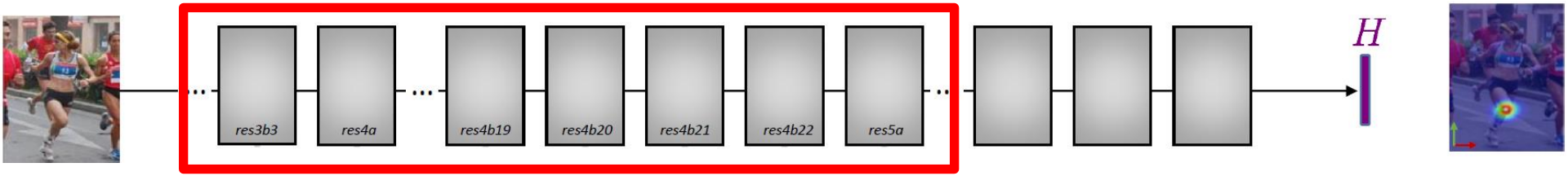
Transfer Learning: Weight Initialization Step 1



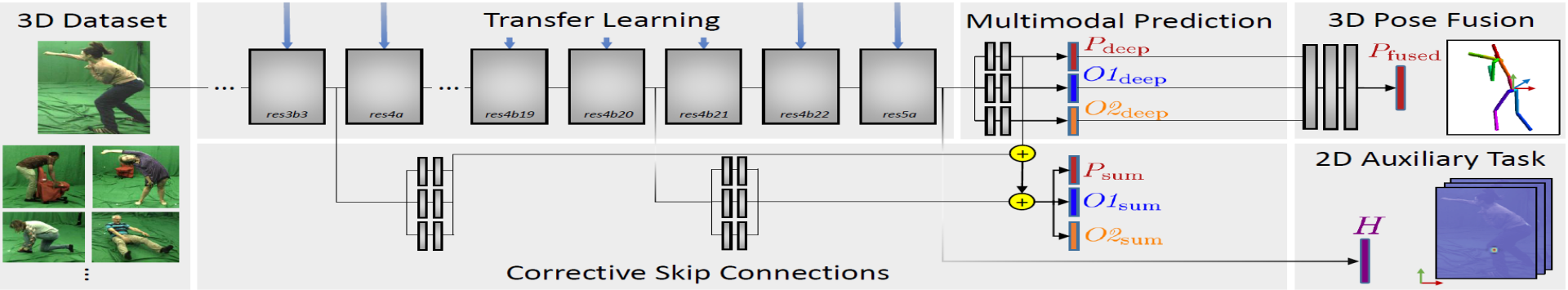
Transfer Learning: Train 2DPoseNet

ResNet-101

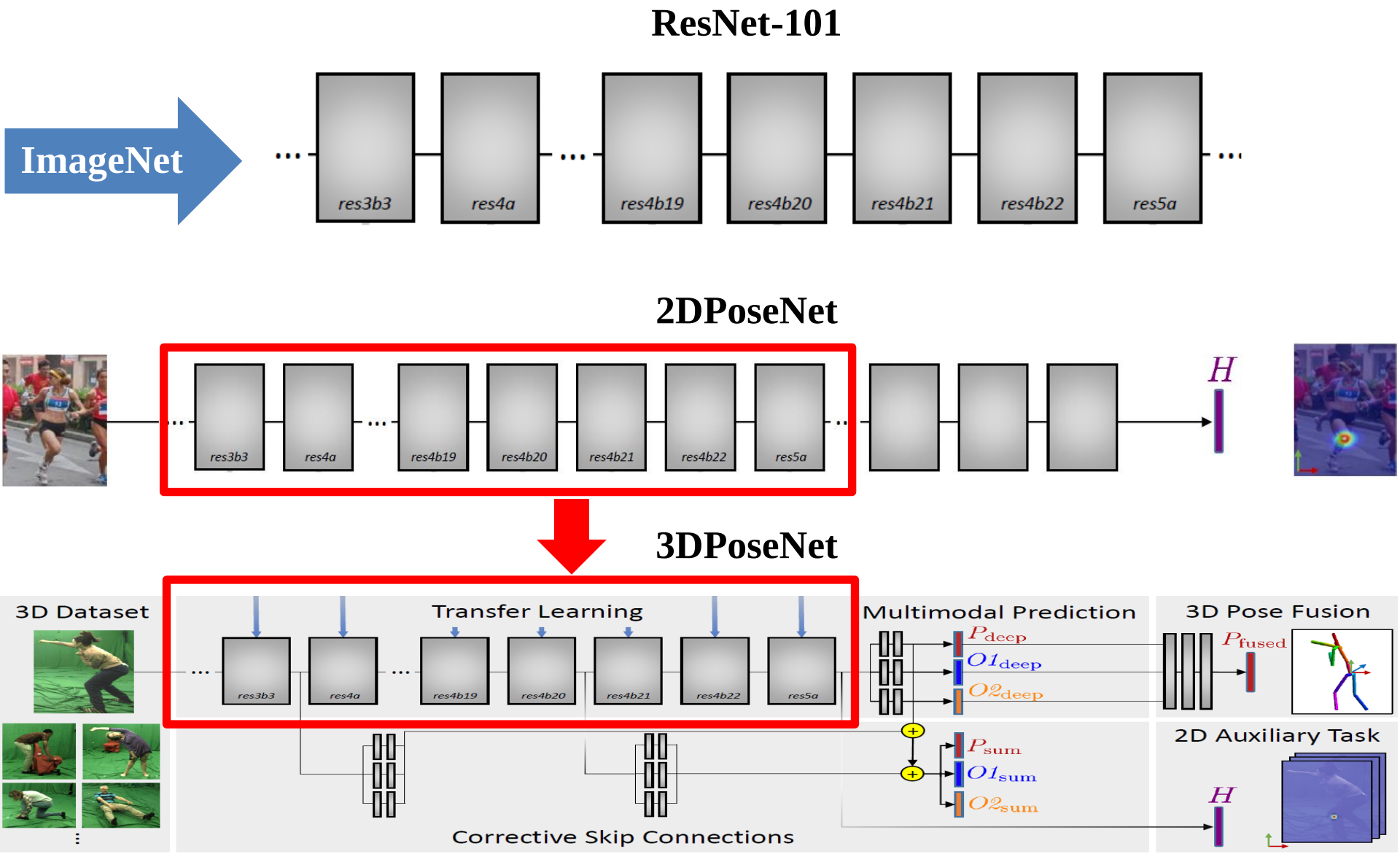
2DPoseNet



3DPoseNet



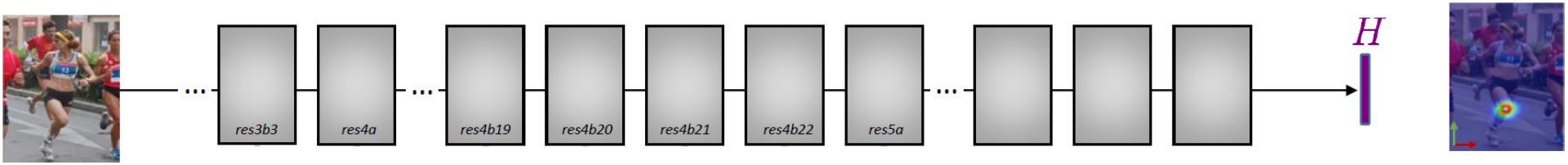
Transfer Learning: Weight Initialization Step 2



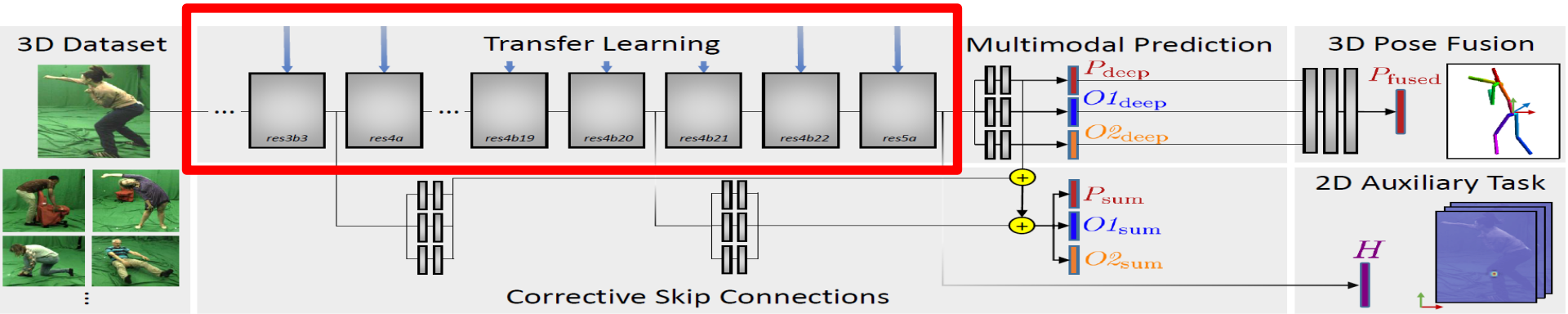
Transfer Learning: Train 3DPoseNet

ResNet-101

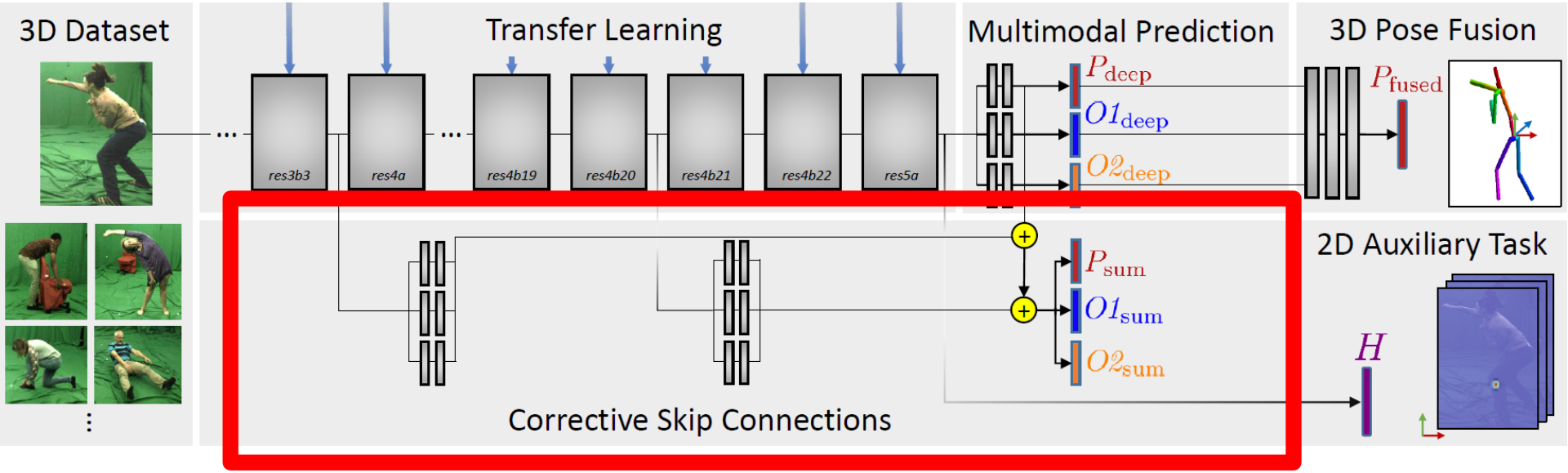
2DPoseNet



3DPoseNet



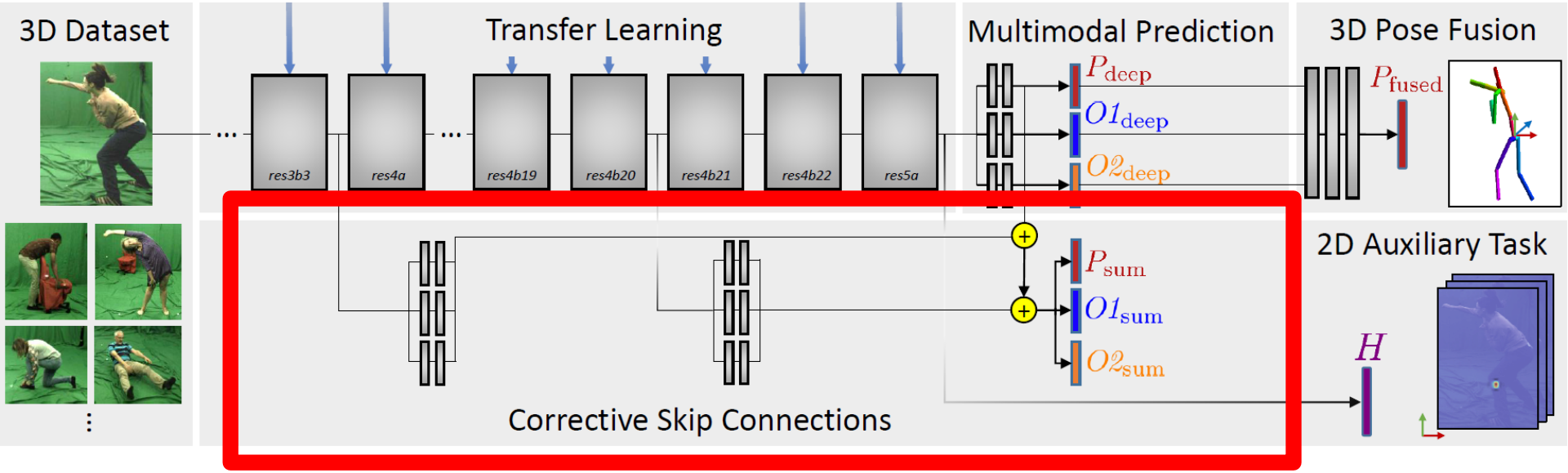
3DPoseNet Architecture



Main Components:

- CNN based on ResNet-101 with Transfer Learning
- Multi-level Corrective Skip Connections
- Multi-modal Prediction
- 3D Pose Fusion
- 2D Pose Estimation

Multi-level Corrective Skip Connections

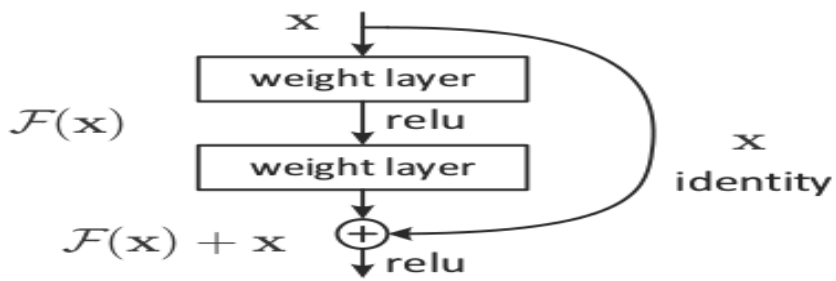


- Used only during training
- Key insights
  - Information flow from different levels help in regression problems
  - Deepest level should contribute more than shallower ones

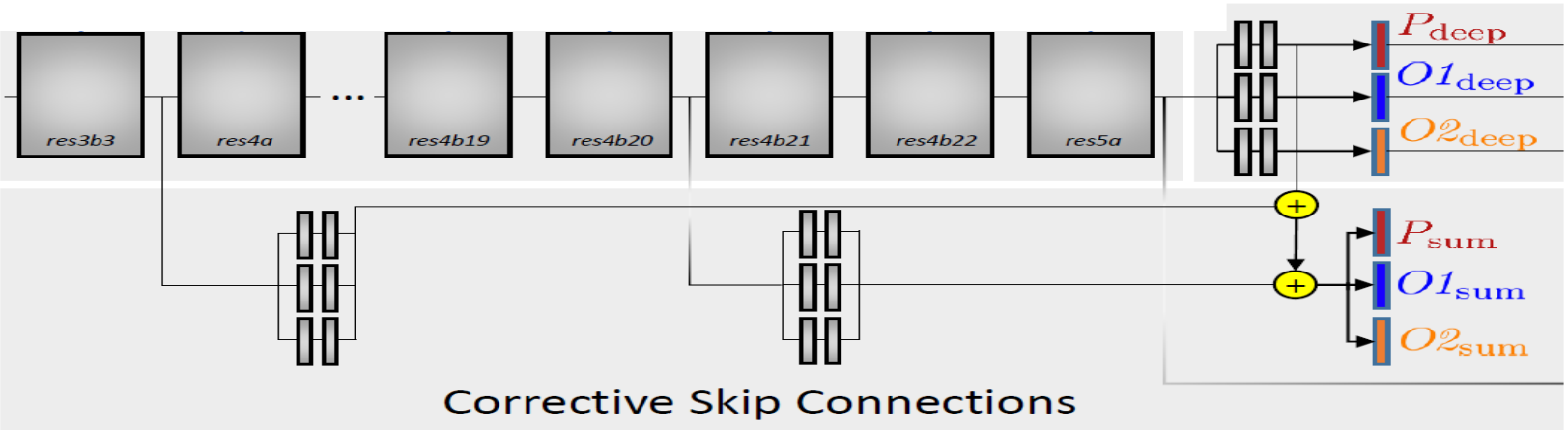
Multi-level Corrective Skip Connections

- The insight "**Deepest level should contribute more than shallower ones**" is not satisfied if we use regular skip connections

Image Source: *Kaiming He et al.* 2016



Normal Skip Connection



Multi-level Skip Connections

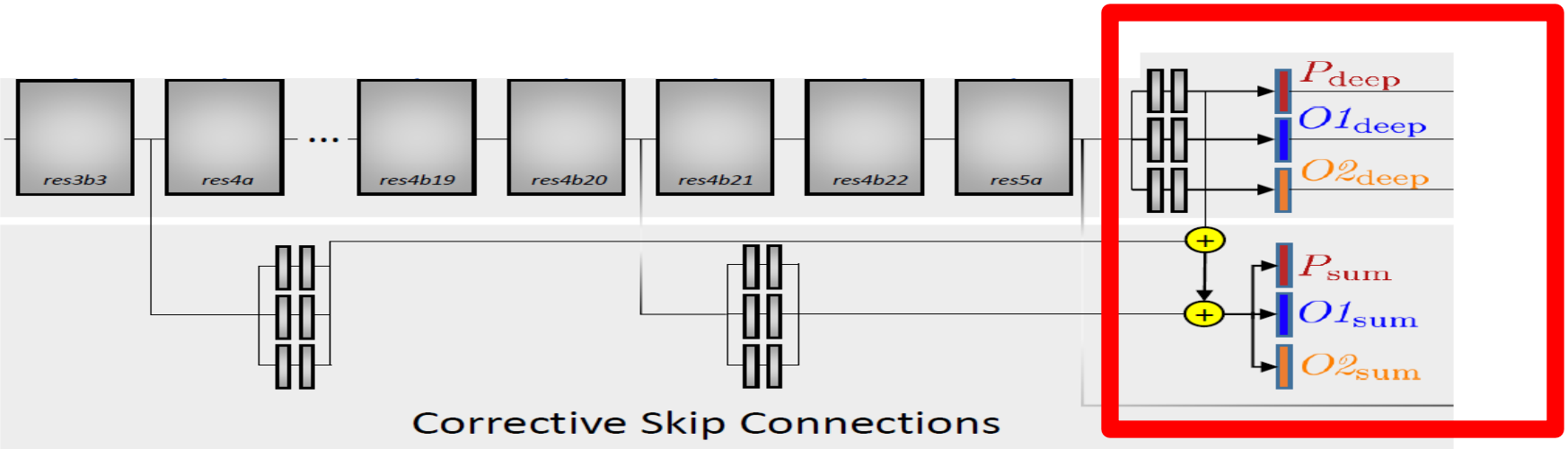
Multi-level Corrective Skip Connections

- The insight "**Deepest level should contribute more than shallower ones**" is not satisfied if we use regular skip connections

- By using multi-level skip connections, the loss is given as:

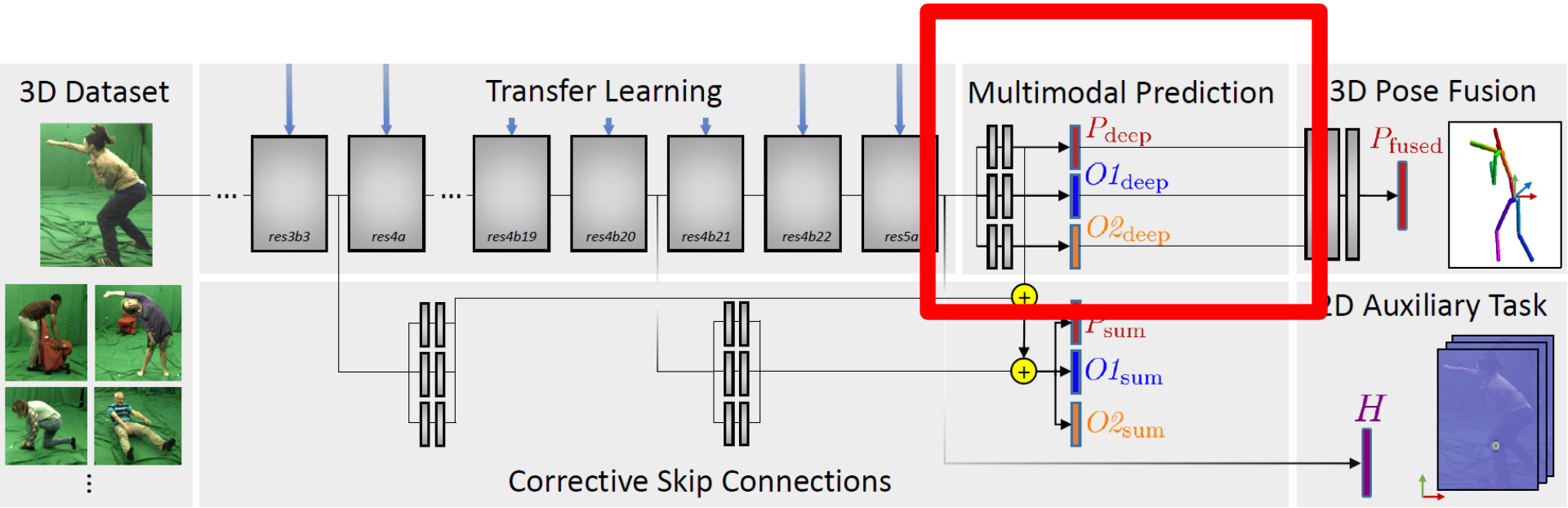
$$P_{sum} = P_{deep} + loss\ contribution\ from\ shallower\ terms$$

- This structure enforces the above insight



Multi-level Skip Connections

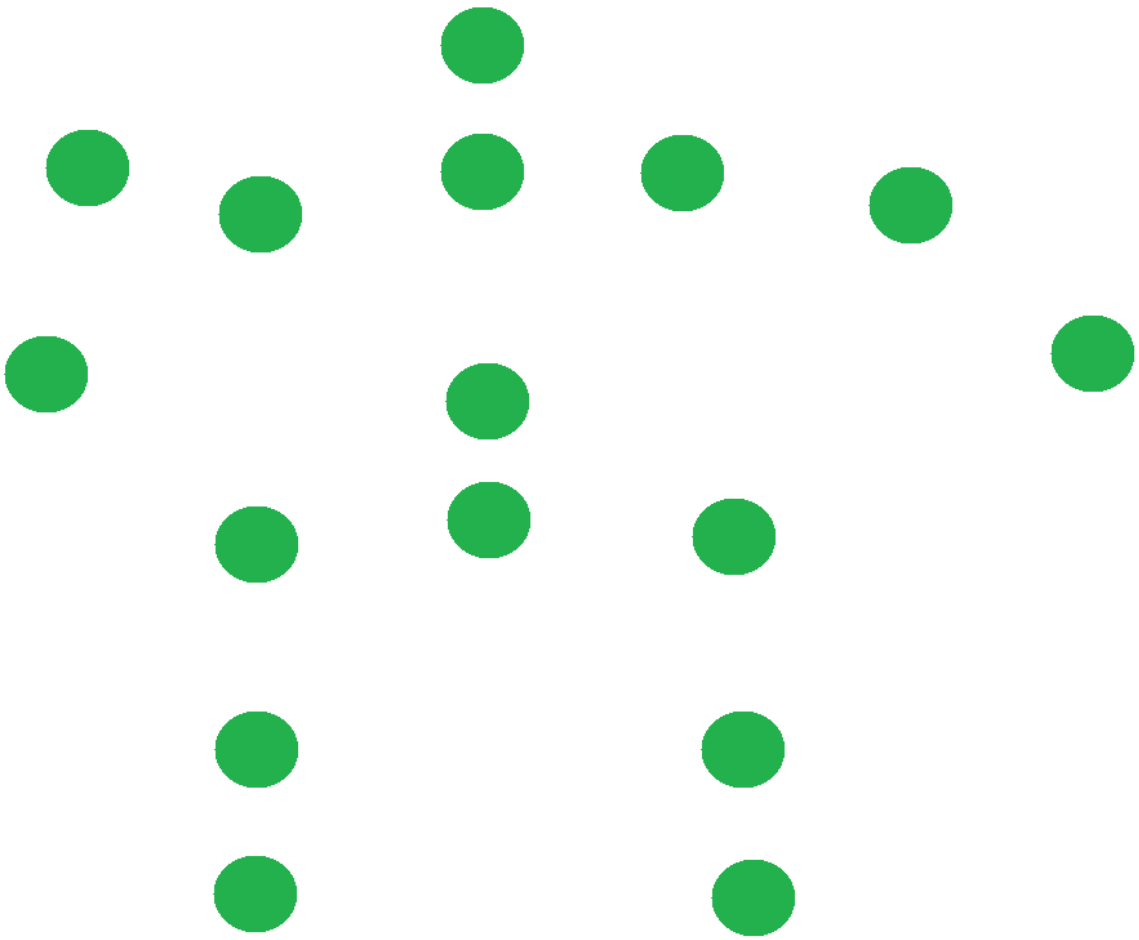
3DPoseNet Architecture



Main Components:

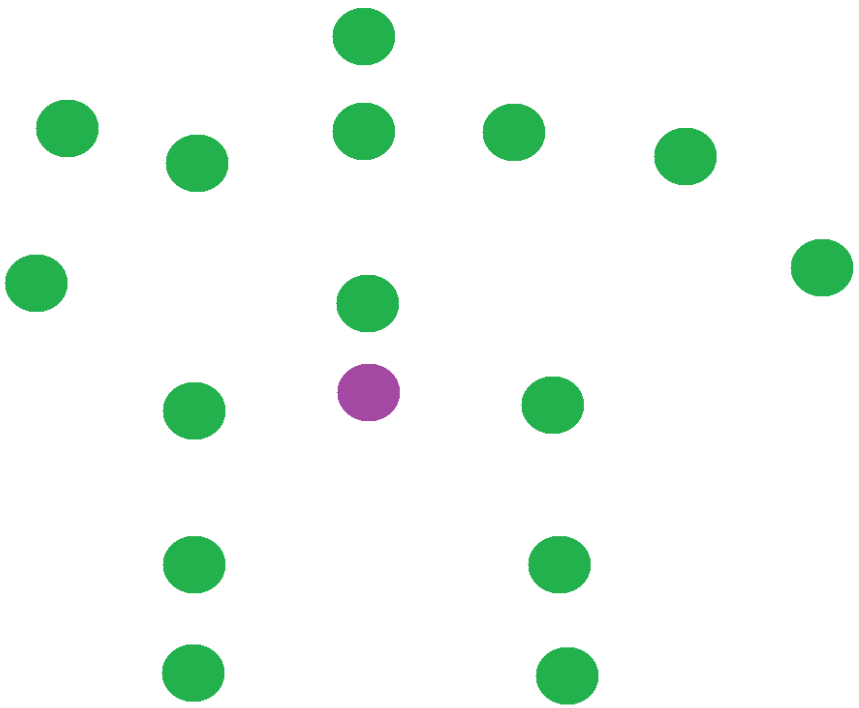
- CNN based on ResNet-101 with Transfer Learning
- Multi-level Corrective Skip Connections
- Multi-modal Prediction
- 3D Pose Fusion
- 2D Pose Estimation

Representation of Joints



Representation of a Pose

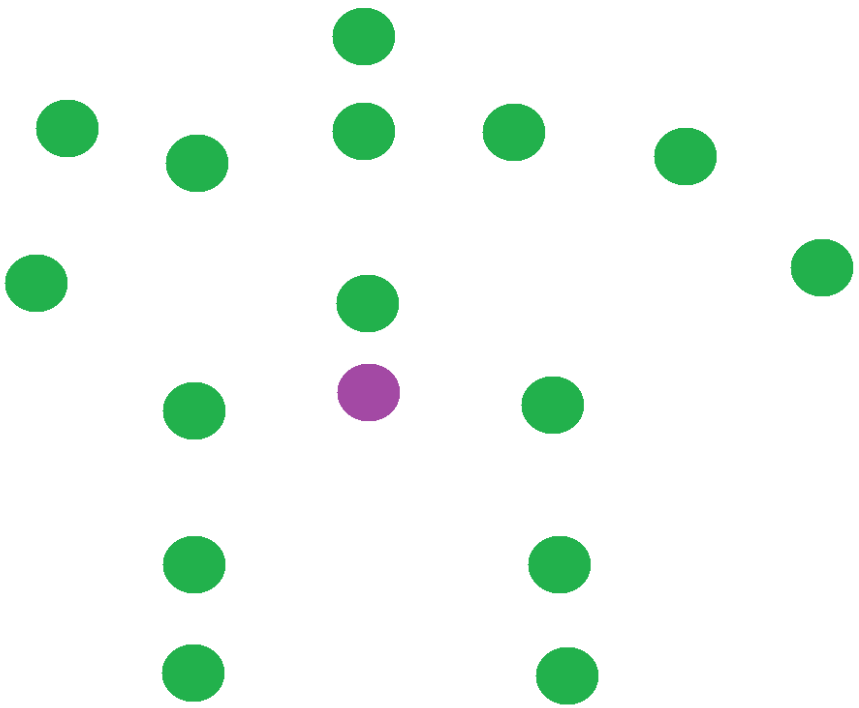
- Select a joint
- Represent other joints relative to this one



Representation of a Pose

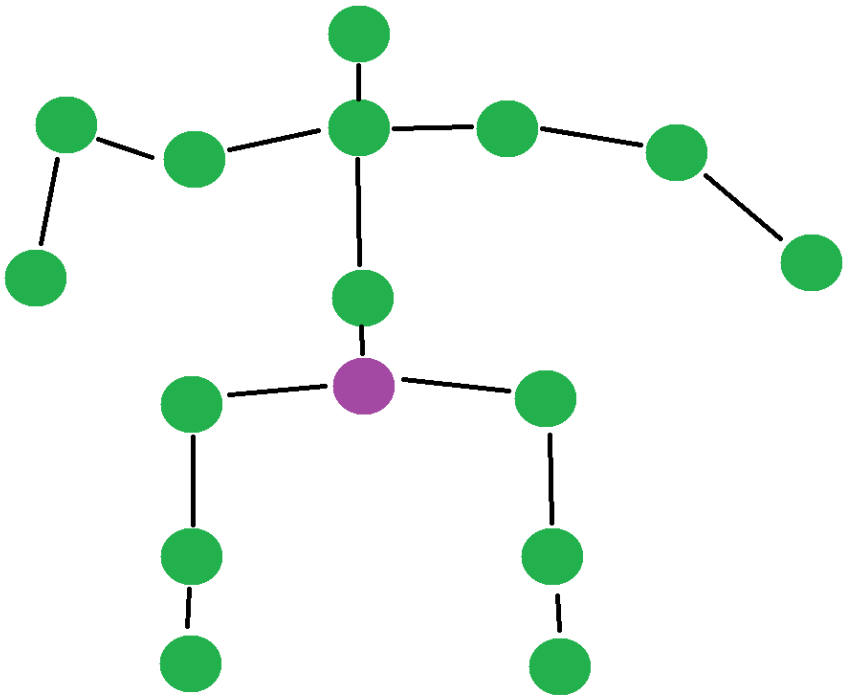
- Select a joint
- Represent other joints relative to this one

Problem! This representation is not always optimal



Representation Using First Order Kinematics

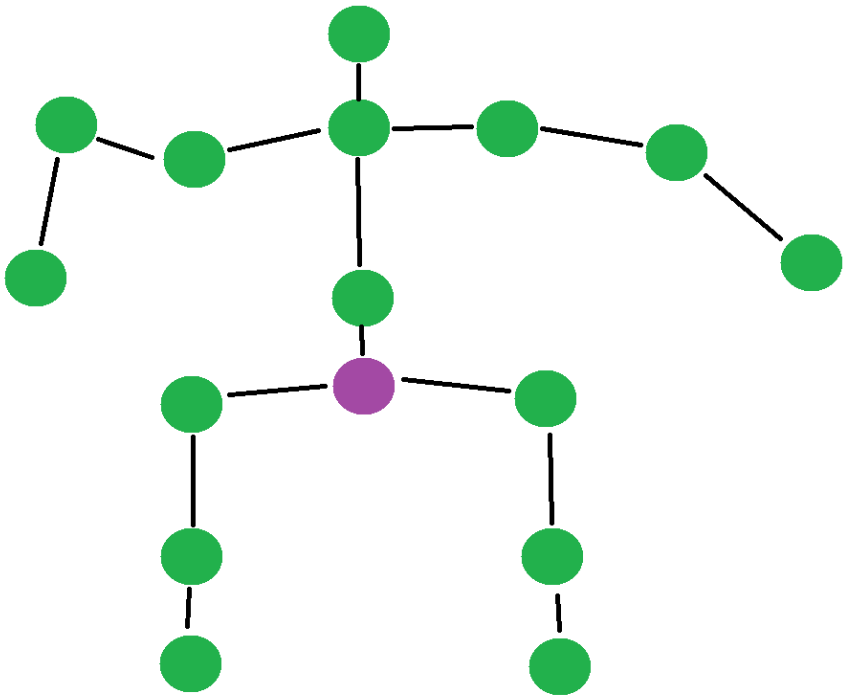
- S. Li et al. suggested representation using first order parents in kinematic structure improves performance



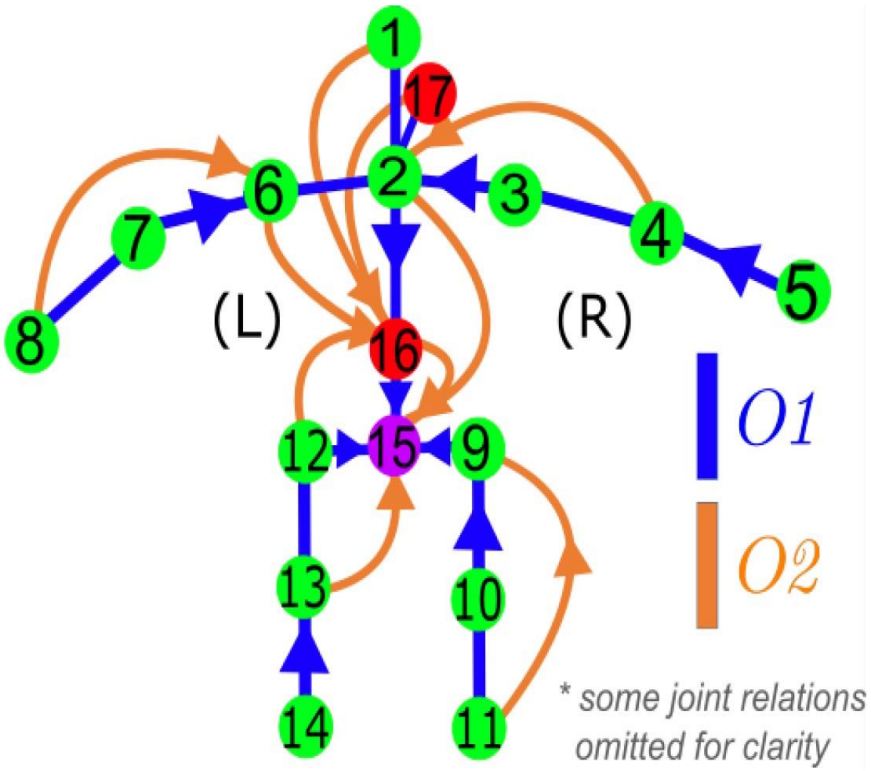
Representation Using First Order Kinematics

- S. Li et al. suggested representation using first order parents in kinematic structure improves performance

- Authors of this paper found out that this is not universally true. Optimal order relationship in kinematic structure depends on pose and visibility.

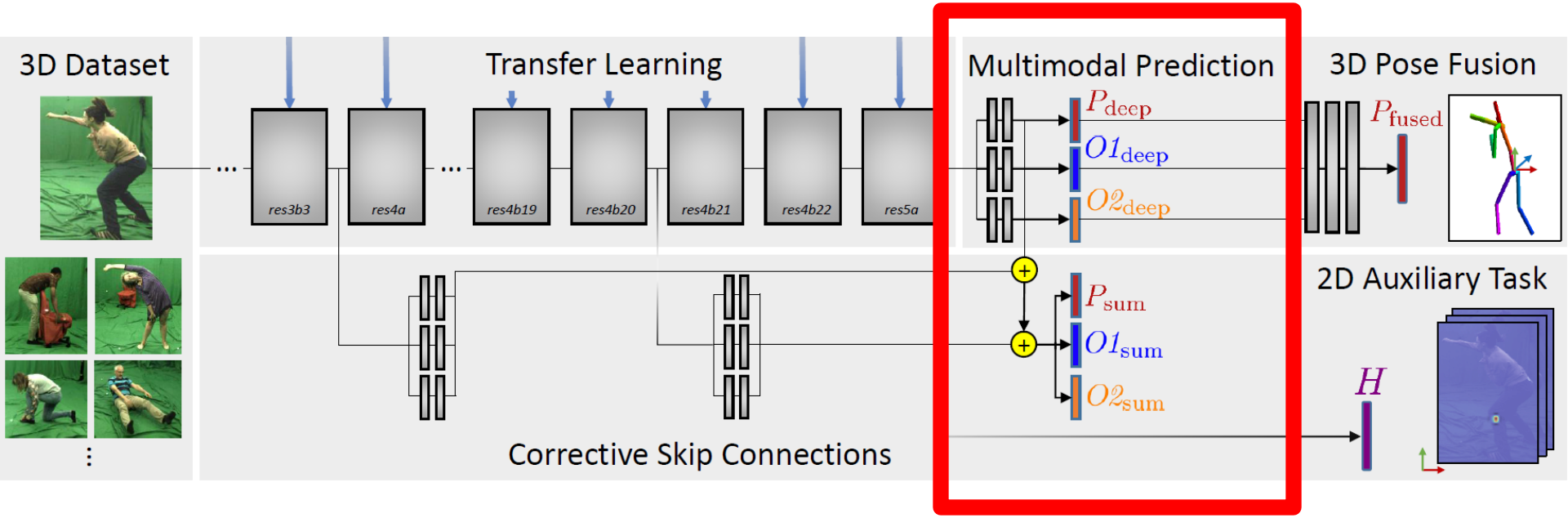


Different Representations of a Pose

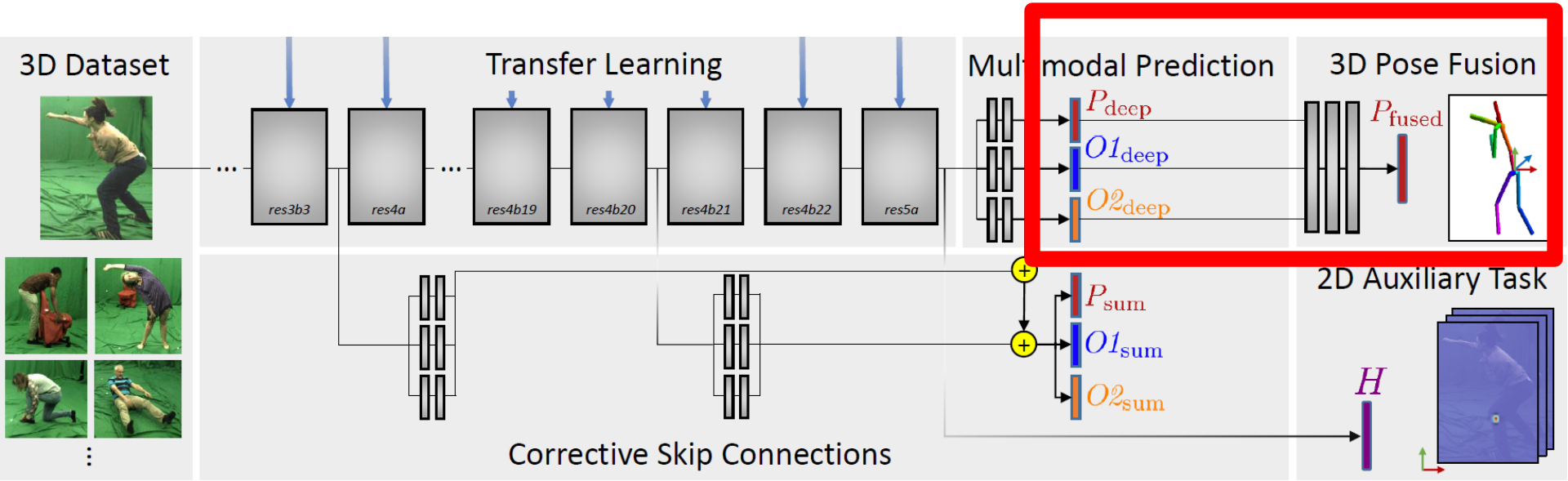


P: joints relative to root (pelvis here)
O1: joints relative to first order kinematics
O2: joints relative to second order kinematics

Multi-modal Prediction



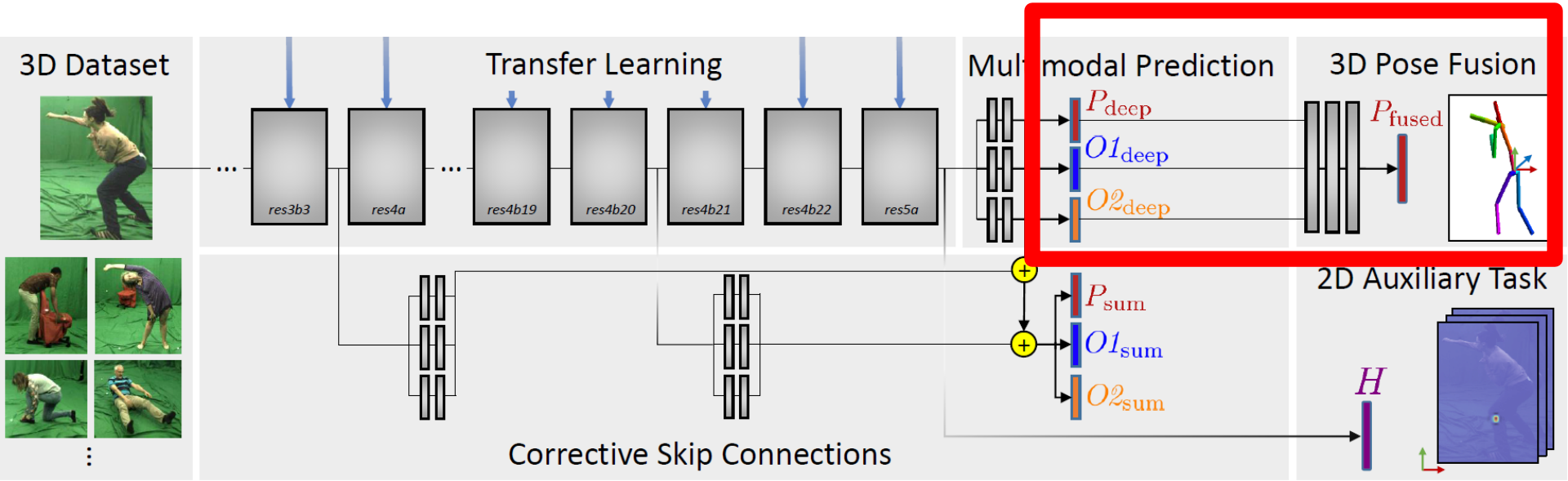
3DPoseNet Architecture



Main Components:

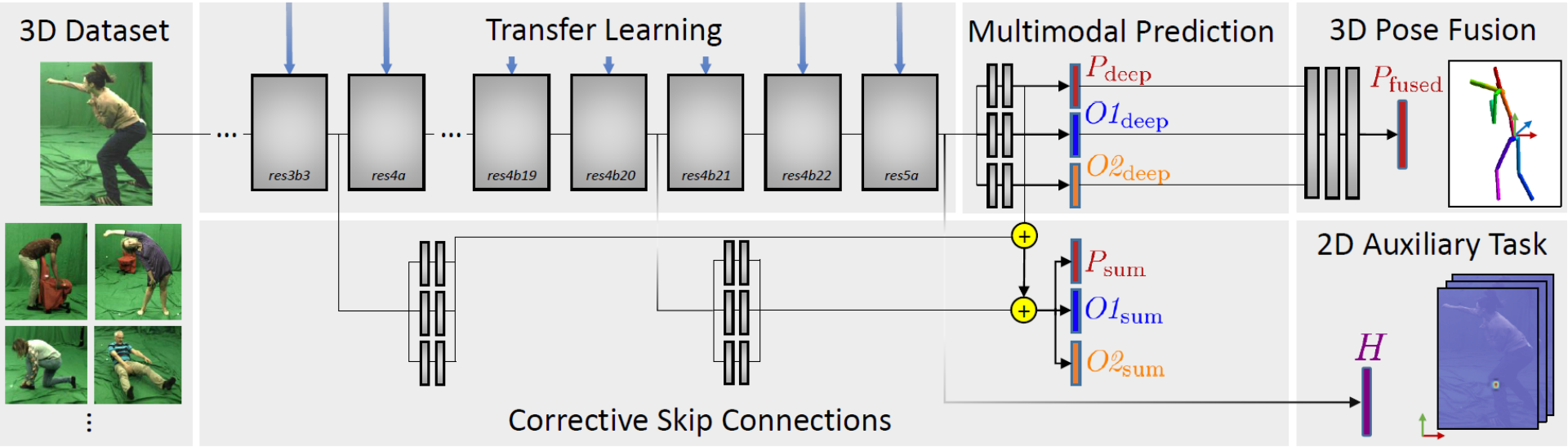
- CNN based on ResNet-101 with Transfer Learning
- Multi-level Corrective Skip Connections
- Multi-modal Prediction
- 3D Pose Fusion
- 2D Pose Estimation

3D Pose Fusion



- Now better constraints are available per joint
- Network automatically finds the optimal combination of different pose representations

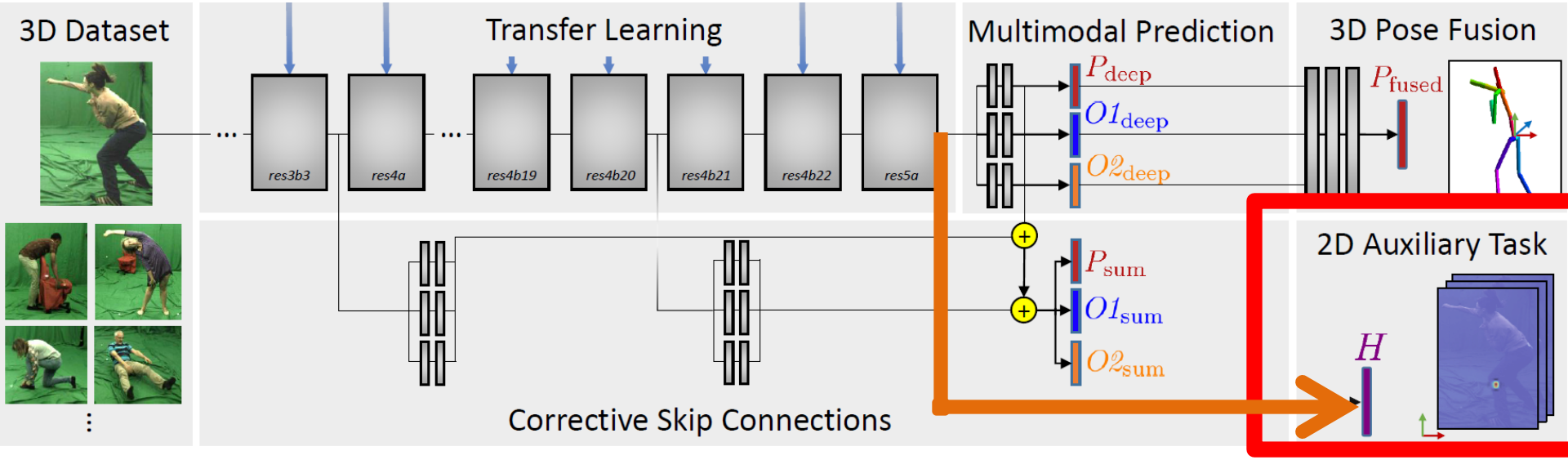
3DPoseNet Architecture



Main Components:

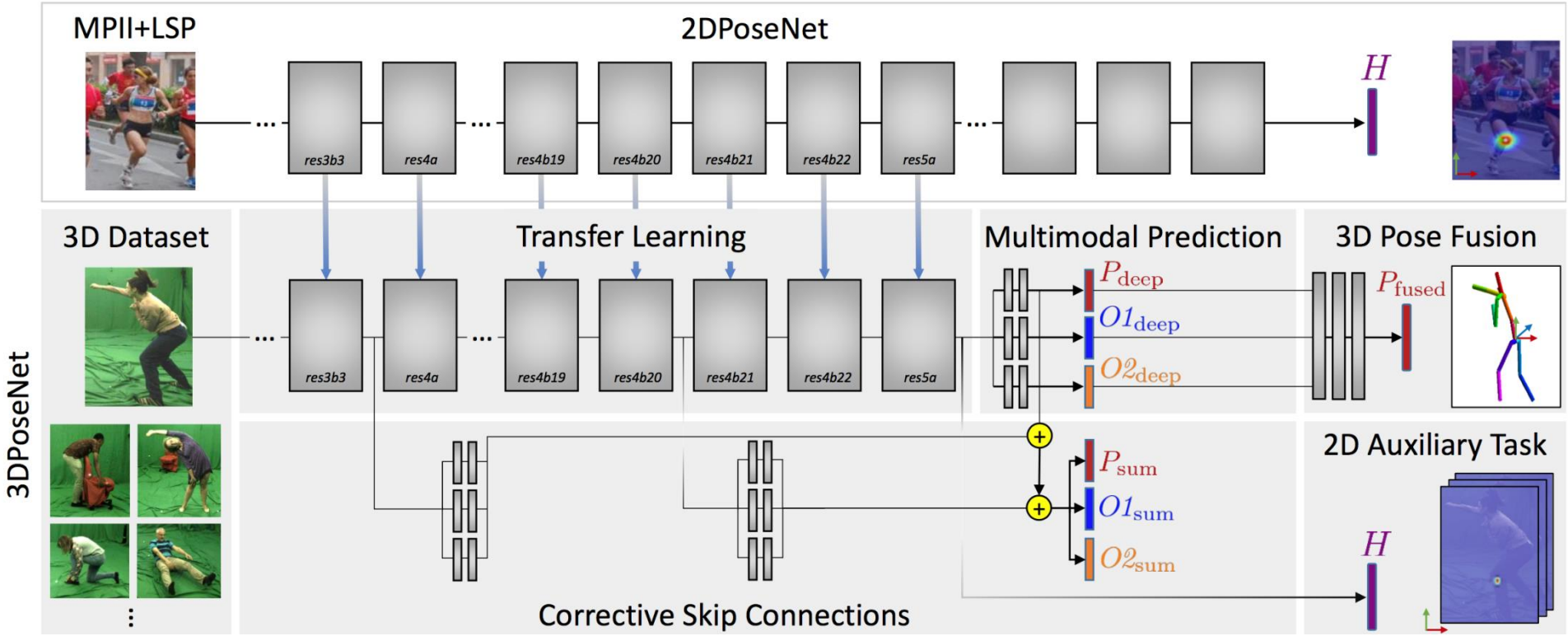
- CNN based on ResNet-101 with Transfer Learning
- Multi-level Corrective Skip Connections
- Multi-modal Prediction
- 3D Pose Fusion
- 2D Pose Estimation

2D Pose Estimation



Accomplishing an additional task without any significant overheads

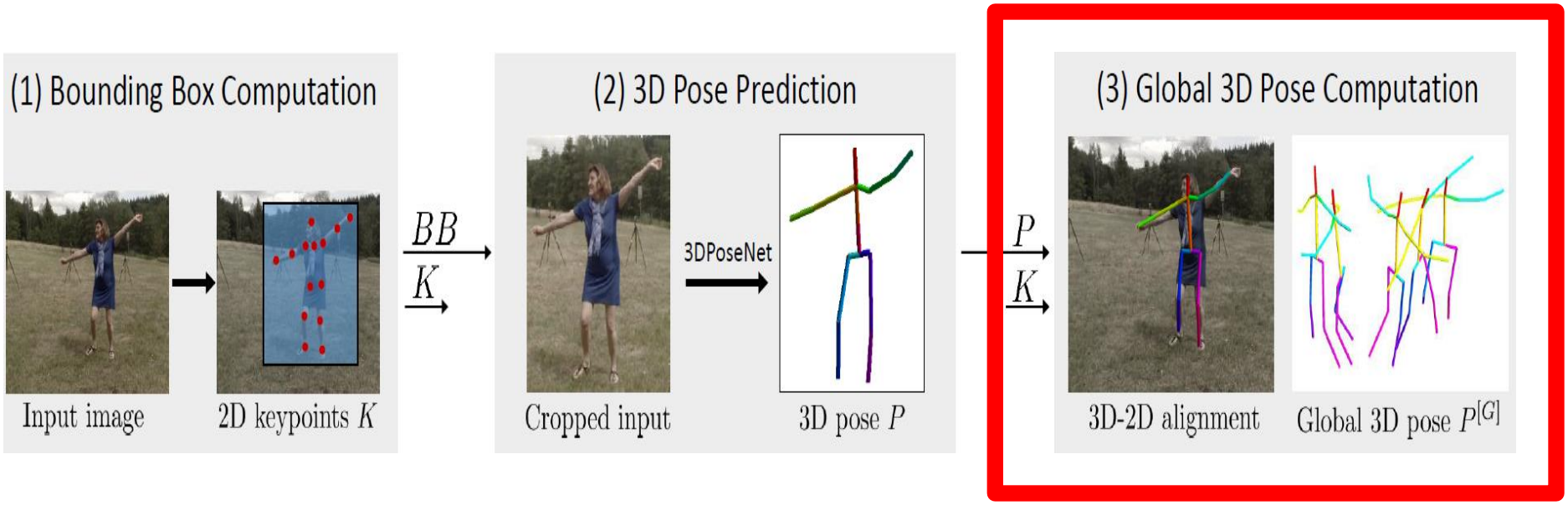
Network Architecture



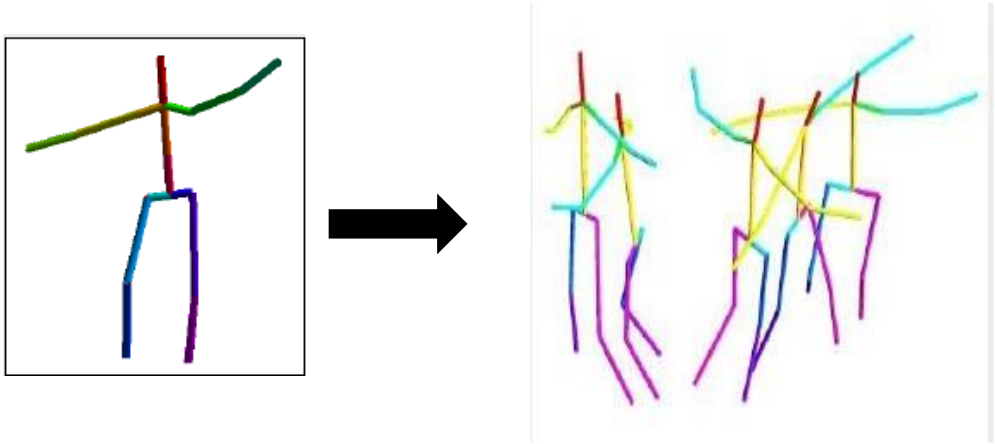
The Focus of My Presentation

- Introduction
- Common Approaches
- Current Approach
  - Bounding Box
  - 3D Pose Prediction
  - Global 3D Pose Computation
- Data Set Contribution
- Evaluation
  - Perspective Correction
  - Multi-level Corrective Skip Connections
  - New Data Set
  - Transfer Learning
  - Multi-modal Prediction
  - State of the Art
- Summary

Approach Overview



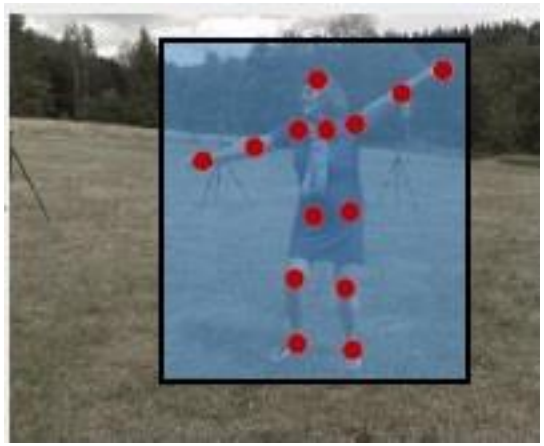
Global 3D Pose Computation



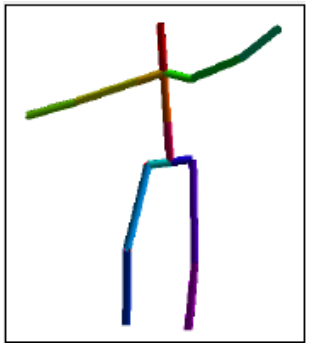
P_{fused} **Global 3D Pose** $P^{[G]}$

- From pelvis-centered pose P_{fused} , we need to get a global3D pose $P^{[G]}$

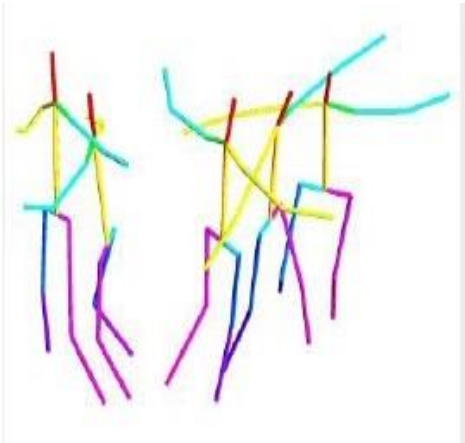
Global 3D Pose Computation



K



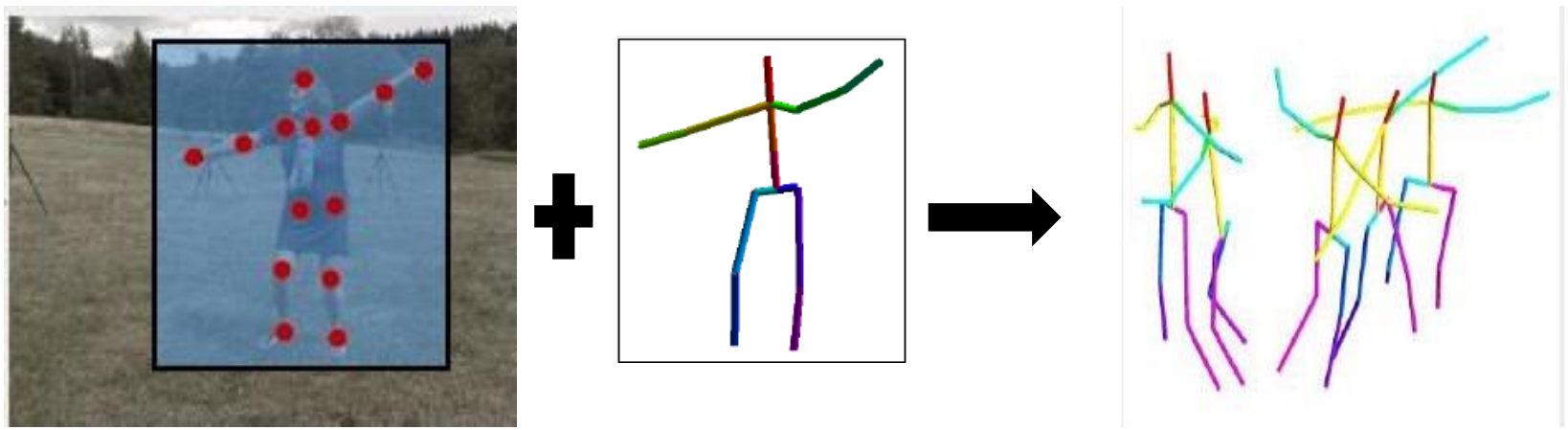
P_{fused}



Global 3D Pose $P^{[G]}$

- From pelvis-centered pose P_{fused} , we need to get a global3D pose $P^{[G]}$
- P_{fused} is obtained from the normalized image that we get from the bounding box **BB**

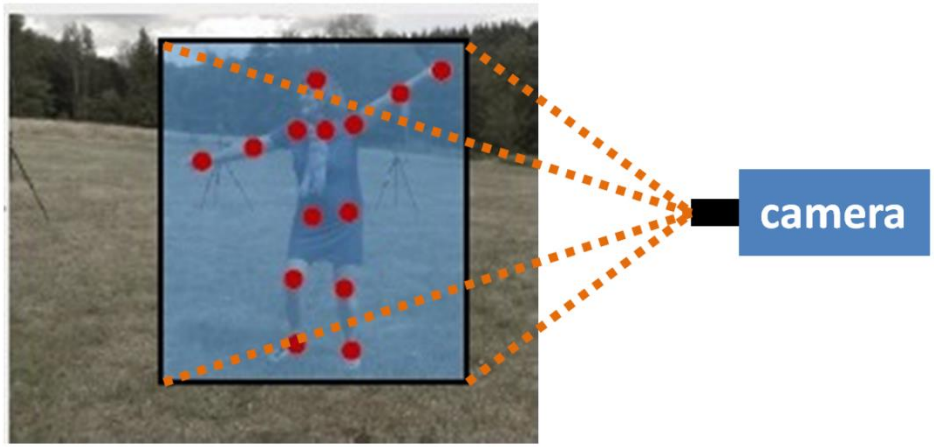
Global 3D Pose Computation

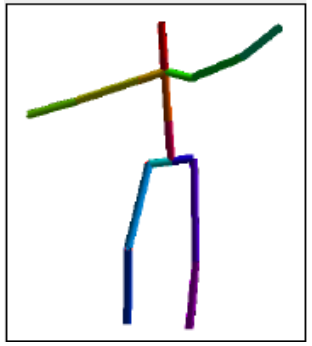


K $\boldsymbol{P_{fused}}$ **Global 3D Pose** $\boldsymbol{P^{[G]}}$

- From pelvis-centered pose $\boldsymbol{P_{fused}}$, we need to get a global3D pose $\boldsymbol{P^{[G]}}$
- $\boldsymbol{P_{fused}}$ is obtained from the normalized image that we get from the bounding box **BB**
- If we know the joint locations **K** in the original image, we can know the **BB** and thus know the normalization parameters used for $\boldsymbol{P_{fused}}$

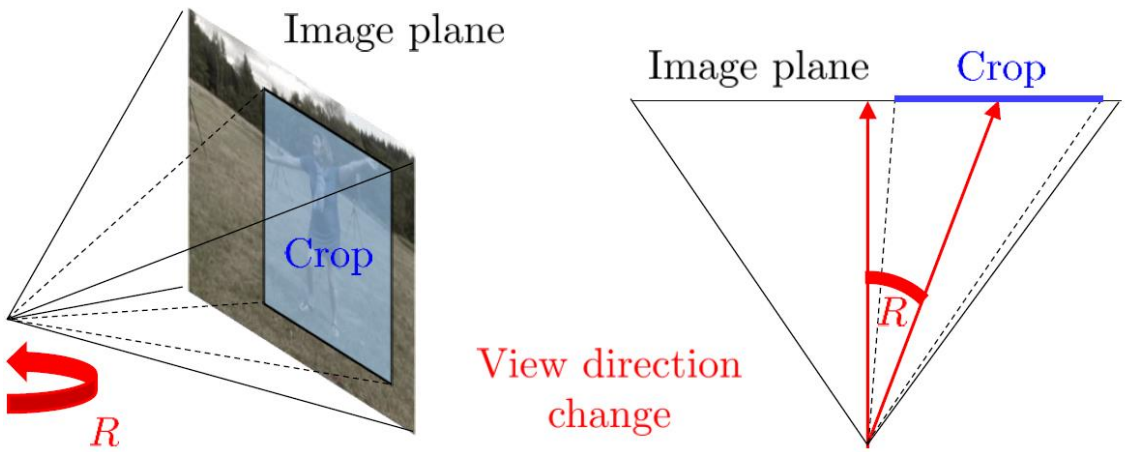
Perspective Correction





- If we imagine a virtual camera that only sees the cropped part of the image, then $\boldsymbol{P}_{fused}$ can be viewed as the pose inside the camera frame

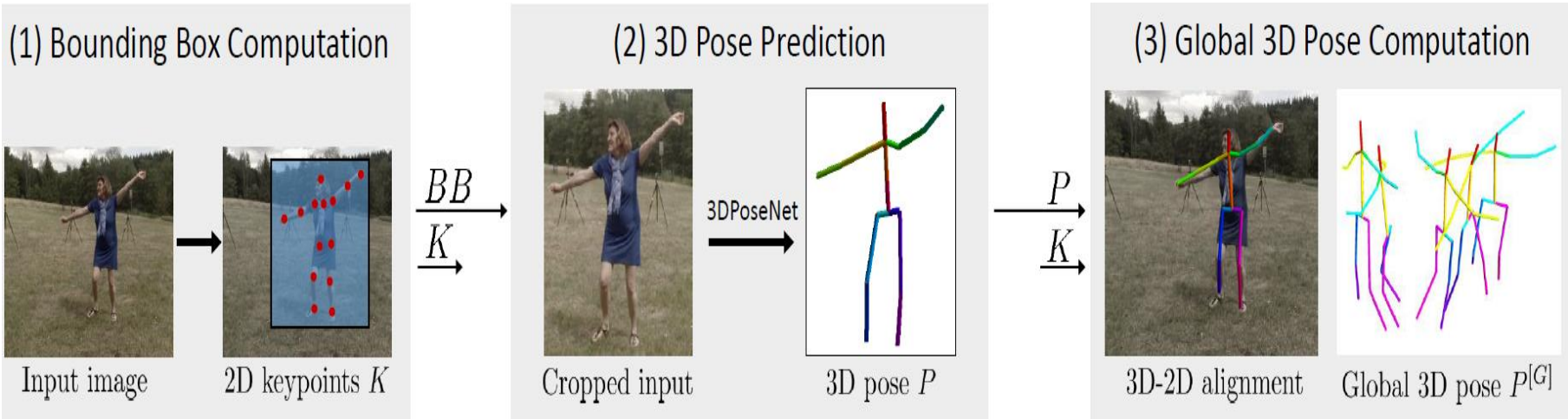
Perspective Correction



- Now, we need to find **R** and **T** such the view frame of this virtual camera equal the cropped image, then we can get the global 3D pose $\boldsymbol{P}^{[G]}$

$$\boldsymbol{P}^{[G]} = (\,\mathrm{R}\,|\,\mathrm{T}\,)\,\boldsymbol{P}_{fused}$$

Process Completed



The Focus of My Presentation

- Introduction
- Common Approaches
- Current Approach
  - Bounding Box
  - 3D Pose Prediction
  - Global 3D Pose Computation
- Data Set Contribution
- Evaluation
  - Perspective Correction
  - Multi-level Corrective Skip Connections
  - New Data Set
  - Transfer Learning
  - Multi-modal Prediction
  - State of the Art
- Summary

Another Contribution of this Paper: New Dataset

Existing 3D Pose Datasets

Marker-based Motion Capture	Marker-less Motion Capture
• Relatively cheaper with fewer cameras • Requires skin-tight clothing to capture 3D data	• Hundreds of cameras are required • Requires very controlled environment • Diverse clothing available

A. Baak et al. showed that motion tracking techniques do not work in general scenes

MPI-INF-3DHP Dataset

Salient Features

- Multi-camera studio
- Similar to the marker-less motion capture
- Works on everyday apparel in contrast to skin tight clothing
- 8 actors (4m+4f)
- Covered more pose classes than Human3.6M
- Different viewpoints and different elevations
- > 1.3M frames
- Augmentation
- Diverse test set

MPI-INF-3DHP Dataset

Salient Features

- Multi-camera studio
- Similar to the marker-less motion capture
- Works on everyday apparel in contrast to skin tight clothing
- 8 actors (4m+4f)
- Covered more pose classes than Human3.6M
- Different viewpoints and different elevations
- > 1.3M frames
- Augmentation
- Diverse test set

Data Augmentation

MPI-INF-3DHP Dataset

Salient Features

- Multi-camera studio
- Similar to the marker-less motion capture
- Works on everyday apparel in contrast to skin tight clothing
- 8 actors (4m+4f)
- Covered more pose classes than Human3.6M
- Different viewpoints and different elevations
- > 1.3M frames
- Augmentation
- Diverse test set

Diverse Test Set

- Test sets of **Human3.6M** and **HumanEva** are recorded indoors
- Test set of **MPI-INF-3DHP** is more general due to
  - More diverse motions
  - Greater variation in camera view-point
  - Outdoor scenes

The Focus of My Presentation

- Introduction
- Common Approaches
- Current Approach
  - Bounding Box
  - 3D Pose Prediction
  - Global 3D Pose Computation
- Data Set Contribution
- Evaluation
  - Perspective Correction
  - Multi-level Corrective Skip Connections
  - New Data Set
  - Transfer Learning
  - Multi-modal Prediction
  - State of the Art
- Summary

Experiments and Evaluation

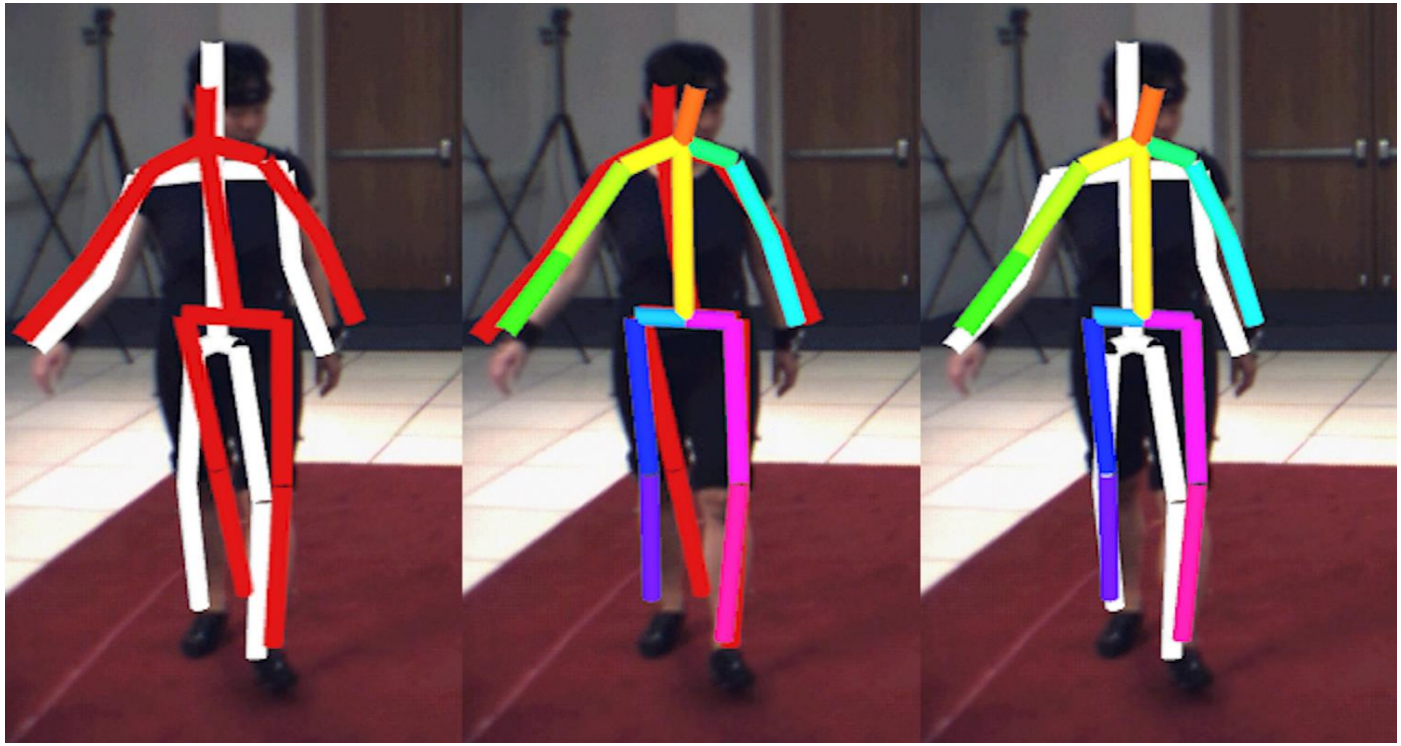
Metrics for Evaluation

- Mean Per Joint Position Error (MPJPE)
- Percentage of Correct Keypoints (PCK)
- Area Under the Curve (AUC)

The Focus of My Presentation

- Introduction
- Common Approaches
- Current Approach
  - Bounding Box
  - 3D Pose Prediction
  - Global 3D Pose Computation
- Data Set Contribution
- Evaluation
  - Perspective Correction
  - Multi-level Corrective Skip Connections
  - New Data Set
  - Transfer Learning
  - Multi-modal Prediction
  - State of the Art
- Summary

Evaluating Perspective Correction



- On MPI-INF-3DHP, performance improves by **3** PCK
- On HumanEva, error is decreases by **3 mm** MPJPE

The Focus of My Presentation

- Introduction
- Common Approaches
- Current Approach
  - Bounding Box
  - 3D Pose Prediction
  - Global 3D Pose Computation
- Data Set Contribution
- Evaluation
  - Perspective Correction
  - Multi-level Corrective Skip Connections
  - New Data Set
  - Transfer Learning
  - Multi-modal Prediction
  - State of the Art
- Summary

Evaluating Multi-level Corrective Skip Connections

Mean Per Joint Position Error (MPJPE) in mm on Human3.6M Test Data

	Direct	Discuss	Eating	Greet	Phone	Posing	Purch.	Sitting	Sit Down	Smoke	Take Photo	Wait	Walk	Walk Dog	Walk Pair	Total
Initialized with Resnet-101 (ImageNet) weights																
Base	98.98	100.14	86.07	101.83	101.34	96.74	94.89	125.28	158.31	100.21	112.49	99.57	83.39	109.61	95.79	104.32
Regular Skip	113.34	112.26	97.40	110.50	108.63	112.09	105.67	125.97	173.41	109.34	120.87	107.75	97.30	126.05	117.45	115.29
Corr. Skip	**92.57**	99.08	85.46	**95.43**	96.93	**89.56**	95.67	123.54	160.98	97.13	**107.56**	**93.86**	**76.99**	110.93	**88.73**	101.09
+ Fusion	93.80	99.17	84.73	95.60	94.48	89.40	93.15	119.94	154.61	95.94	106.09	94.13	77.25	108.82	87.38	99.79
Transfer Learning from 2DPoseNet weights																
Corr. + Fusion	59.69	69.74	60.55	68.77	76.36	59.05	75.04	96.19	122.92	70.82	85.42	68.45	54.41	82.03	59.79	74.14

Key Observations:

- Error increases if we use regular skip connections
- Error decreases if we use multi level skip connections

Evaluating Multi-level Corrective Skip Connections

Percentage of Correct Keypoints (PCK) and Area Under the Curve (AUC) on MPI-INF-3DHP Test Data. GS = green screen background

3D Data	Network Arch.	Studio GS	Studio no GS	Outdoor	All	
		3DPCK	3DPCK	3DPCK	3DPCK	AUC
Human 3.6m	Base	21.1	32.5	10.8	22.6	8.8
	Corr Skip	22.2	33.9	18.5	25.1	8.7
	+Fusion	22.3	34.2	20.0	26.0	9.5
Ours Unaug.	Base	73.6	42.9	19.5	49.0	23.3
	Corr Skip	66.9	38.2	27.9	46.8	20.9
	+Fusion	67.6	39.6	28.5	47.8	21.8
Ours Aug.	Base	77.2	59.5	48.7	63.7	31.1
	Corr. Skip	71.1	51.7	36.1	55.4	26.0
	+Fusion	73.5	53.1	37.9	57.3	28.0

Key Observations:

- Nearly 8% 3DPCK improvement for the outdoor scenes
- Multi-level skip connections improve generality
- Performance decreases for augmented MPI-INF-3DHP training data

The Focus of My Presentation

- Introduction
- Common Approaches
- Current Approach
  - Bounding Box
  - 3D Pose Prediction
  - Global 3D Pose Computation
- Data Set Contribution
- Evaluation
  - Perspective Correction
  - Multi-level Corrective Skip Connections
  - New Data Set
  - Transfer Learning
  - Multi-modal Prediction
  - State of the Art
- Summary

Evaluating New Data Set

Percentage of Correct Keypoints (PCK) and Area Under the Curve (AUC) on MPI-INF-3DHP Test Data. GS = green screen background

3D Data	Network Arch.	Studio GS	Studio no GS	Outdoor	All	
		3DPCK	3DPCK	3DPCK	3DPCK	AUC
Human 3.6m	Base	21.1	32.5	10.8	22.6	8.8
	Corr Skip	22.2	33.9	18.5	25.1	8.7
	+Fusion	22.3	34.2	20.0	26.0	9.5
Ours Unaug.	Base	73.6	42.9	19.5	49.0	23.3
	Corr Skip	66.9	38.2	27.9	46.8	20.9
	+Fusion	67.6	39.6	28.5	47.8	21.8
Ours Aug.	Base	77.2	59.5	48.7	63.7	31.1
	Corr. Skip	71.1	51.7	36.1	55.4	26.0
	+Fusion	73.5	53.1	37.9	57.3	28.0

Evaluating New Data Set

Percentage of Correct Keypoints (PCK) and Area Under the Curve (AUC) on MPI-INF-3DHP Test Data. GS = green screen background

3D Dataset	Method	Studio GS	Studio no GS	Outdoor	All	
		3DPCK	3DPCK	3DPCK	3DPCK	AUC
Human3.6m	Domain adapt.	44.1	42.6	35.2	41.4	17.7
Human3.6m	Corr. + Fusion	70.8	62.3	58.5	64.7	31.7
Ours Unaug.	Corr + Fusion	84.1	68.9	59.6	72.5	36.9
Ours.	Base + Fusion	82.8	68.0	62.3	72.4	36.9
Aug.	Corr. + Fusion	82.6	66.7	62.0	71.7	36.4
Ours Aug.	Corr. + Fusion	84.6	72.4	69.7	76.5	40.8
+	no persp. corr.	81.9	68.6	67.4	73.5	37.6
Human3.6m	with pred. BB	80.4	71.2	69.8	74.4	39.6

The Focus of My Presentation

- Introduction
- Common Approaches
- Current Approach
  - Bounding Box
  - 3D Pose Prediction
  - Global 3D Pose Computation
- Data Set Contribution
- Evaluation
  - Perspective Correction
  - Multi-level Corrective Skip Connections
  - New Data Set
  - Transfer Learning
  - Multi-modal Prediction
  - State of the Art
- Summary

Evaluating Transfer Learning from 2DPoseNet to 3DPoseNet

Mean Per Joint Position Error (MPJPE) in mm on Human3.6M Test Data

	Direct	Discuss	Eating	Greet	Phone	Posing	Purch.	Sitting	Sit Down	Smoke	Take Photo	Wait	Walk	Walk Dog	Walk Pair	Total
Initialized with Resnet-101 (ImageNet) weights																
Base	98.98	100.14	86.07	101.83	101.34	96.74	94.89	125.28	158.31	100.21	112.49	99.57	83.39	109.61	95.79	104.32
Regular Skip	113.34	112.26	97.40	110.50	108.63	112.09	105.67	125.97	173.41	109.34	120.87	107.75	97.30	126.05	117.45	115.29
Corr. Skip	**92.57**	99.08	85.46	**95.43**	96.93	**89.56**	95.67	123.54	160.98	97.13	**107.56**	**93.86**	**76.99**	110.93	**88.73**	101.09
+ Fusion	93.80	99.17	84.73	95.60	94.48	89.40	93.15	119.94	154.61	95.94	106.09	94.13	77.25	108.82	87.38	99.79
Transfer Learning from 2DPoseNet weights																
Corr. + Fusion	59.69	69.74	60.55	68.77	76.36	59.05	75.04	96.19	122.92	70.82	85.42	68.45	54.41	82.03	59.79	74.14

The Focus of My Presentation

- Introduction
- Common Approaches
- Current Approach
  - Bounding Box
  - 3D Pose Prediction
  - Global 3D Pose Computation
- Data Set Contribution
- Evaluation
  - Perspective Correction
  - Multi-level Corrective Skip Connections
  - New Data Set
  - Transfer Learning
  - Multi-modal Prediction
  - State of the Art
- Summary

Evaluating Multi-modal Prediction and 3D Pose Fusion

Mean Per Joint Position Error (MPJPE) in mm on Human3.6M Test Data

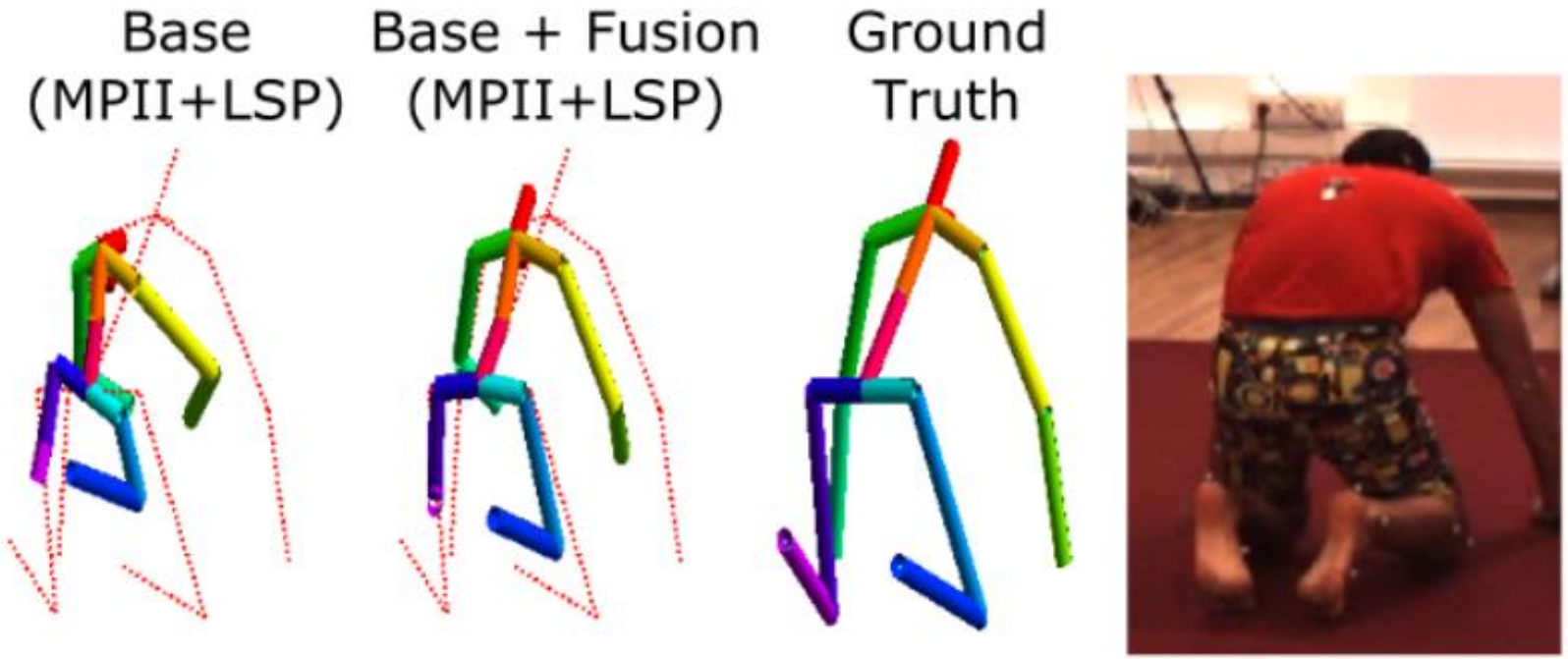
	Direct	Discuss	Eating	Greet	Phone	Posing	Purch.	Sitting	Sit Down	Smoke	Take Photo	Wait	Walk	Walk Dog	Walk Pair	Total
Initialized with Resnet-101 (ImageNet) weights																
Base	98.98	100.14	86.07	101.83	101.34	96.74	94.89	125.28	158.31	100.21	112.49	99.57	83.39	109.61	95.79	104.32
Regular Skip	113.34	112.26	97.40	110.50	108.63	112.09	105.67	125.97	173.41	109.34	120.87	107.75	97.30	126.05	117.45	115.29
Corr. Skip	**92.57**	99.08	85.46	**95.43**	96.93	**89.56**	95.67	123.54	160.98	97.13	**107.56**	**93.86**	**76.99**	110.93	**88.73**	101.09
+ Fusion	93.80	99.17	84.73	95.60	94.48	89.40	93.15	119.94	154.61	95.94	106.09	94.13	77.25	108.82	87.38	99.79
Transfer Learning from 2DPoseNet weights																
Corr. + Fusion	59.69	69.74	60.55	68.77	76.36	59.05	75.04	96.19	122.92	70.82	85.42	68.45	54.41	82.03	59.79	74.14

Evaluating Multi-modal Prediction and 3D Pose Fusion

Percentage of Correct Keypoints (PCK) and Area Under the Curve (AUC) on MPI-INF-3DHP Test Data. GS = green screen background

3D Data	Network Arch.	Studio GS	Studio no GS	Outdoor	All	
		3DPCK	3DPCK	3DPCK	3DPCK	AUC
Human 3.6m	Base	21.1	32.5	10.8	22.6	8.8
	Corr Skip	22.2	33.9	18.5	25.1	8.7
	+Fusion	22.3	34.2	20.0	26.0	9.5
Ours Unaug.	Base	73.6	42.9	19.5	49.0	23.3
	Corr Skip	66.9	38.2	27.9	46.8	20.9
	+Fusion	67.6	39.6	28.5	47.8	21.8
Ours Aug.	Base	77.2	59.5	48.7	63.7	31.1
	Corr. Skip	71.1	51.7	36.1	55.4	26.0
	+Fusion	73.5	53.1	37.9	57.3	28.0

Evaluating Multi-modal Prediction and 3D Pose Fusion



Large Self Occlusion

The Focus of My Presentation

- Introduction
- Common Approaches
- Current Approach
  - Bounding Box
  - 3D Pose Prediction
  - Global 3D Pose Computation
- Data Set Contribution
- Evaluation
  - Perspective Correction
  - Multi-level Corrective Skip Connections
  - New Data Set
  - Transfer Learning
  - Multi-modal Prediction
  - State of the Art
- Summary

Comparison With the State of the Art

- Base structure of **Deep Kinematic Pose** consists of ResNet that is pre-trained on ImageNet
- **SMPLify** uses a different technique. It first predicts 2D body joint locations . Then tries to fit 3D shape to these 2D joints
- **25%** improvement over the state of the art regression based model

Table 4. Comparison of results on Human3.6m [28] with the state of the art. Human3.6m, Subjects 1,5,6,7,8 used for training, and 9,11 used for testing. ^{S} = Scaled to test subject specific skeleton, computed from T-pose. ^{T}= Uses Temporal Information, $^{J14/J17}$ = Joint set evaluated, ^{A} = Uses Best Alignment To GT per frame, ^{Act} = Activitywise Training, $^{1/10/64}$ = Test Set Frame Sampling

Method	Total MPJPE (mm)
Deep Kinematic Pose[78]$^{\mathbf{J17},B}$	107.26
Sparse. Deep. [80]$^{T,\mathbf{J17},B,10,Act}$	113.01
Motion Comp. Seq. [65]$^{T,\mathbf{J17},B}$	124.97
LinKDE [28]$^{\mathbf{J17},B,Act}$	162.14
Du et al. [75]$^{T,\mathbf{J17},B}$	126.47
Rogez et al. [51]$^{(\mathbf{J13}),B,64}$	121.20
SMPLify [10]$^{\mathbf{J14},B,A,(\text{First cam.})}$	82.3
Ours (with 2DPoseNet Transfer)	
Corr. Skip + Fusion $^{\mathbf{J17},B}$	74.11
Corr. Skip + Fusion $^{\mathbf{J17},B,\mathbf{S}}$	68.61
Corr. Skip + Fusion $^{\mathbf{J14},B,\mathbf{A}}$	54.59

Comparison With the State of the Art

- Training Data Set: Human3.6M
- Evaluation Data Set: MPI-INF-3DHP

- **Deep Kinematic Pose** attains **13.8 %** 3DPCK
- This method attains
  - **26%** 3DPCK without Transfer Learning
  - **64.7%** 3DPCK with Transfer Learning

Some Thoughts

Percentage of Correct Keypoints (PCK) and Area Under the Curve (AUC) on MPI-INF-3DHP Test Data. GS = green screen background

3D Data	Network Arch.	Studio GS	Studio no GS	Outdoor	All	
		3DPCK	3DPCK	3DPCK	3DPCK	AUC
Human 3.6m	Base	21.1	32.5	10.8	22.6	8.8
	Corr Skip	22.2	33.9	18.5	25.1	8.7
	+Fusion	22.3	34.2	20.0	26.0	9.5
Ours Unaug.	Base	73.6	42.9	19.5	49.0	23.3
	Corr Skip	66.9	38.2	27.9	46.8	20.9
	+Fusion	67.6	39.6	28.5	47.8	21.8
Ours Aug.	Base	77.2	59.5	48.7	63.7	31.1
	Corr. Skip	71.1	51.7	36.1	55.4	26.0
	+Fusion	73.5	53.1	37.9	57.3	28.0

The Focus of My Presentation

- Introduction
- Common Approaches
- Current Approach
  - Bounding Box
  - 3D Pose Prediction
  - Global 3D Pose Computation
- Data Set Contribution
- Evaluation
  - Perspective Correction
  - Multi-level Corrective Skip Connections
  - New Data Set
  - Transfer Learning
  - Multi-modal Prediction
  - State of the Art
- Summary

Summary

- Direct3D Human Pose Estimation Using a Single Image
- Contributions of this paper
  - Transfer Learning
  - Multi-Level Corrective Skip Connections
  - 3D Pose Fusion
  - New Data Set
  - Perspective Correction
- Achieves better performance than the state of the art
- Further work needed for real-time estimation

References

- S. Li and A. B. Chan. 3d human pose estimation from monocular images with deep convolutional neural network. In Asian Conference on Computer Vision (ACCV), pages 332–347, 2014.
- *Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun*; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778
- Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler and Bernt Schiele. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. IEEE CVPR'14
- Sam Johnson and Mark Everingham "Learning Effective Human Pose Estimation from Inaccurate Annotation" In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR2011)*
- A. Baak, M.M¨uller, G. Bharaj, H.-P. Seidel, and C. Theobalt. A data-driven approach for real-time full body pose reconstruction from a depth camera. In IEEE International Conference on Computer Vision (ICCV), 2011.
- L. Sigal, A. O. Balan, and M. J. Black. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. International Journal of Computer Vision (IJCV), 87(1-2):4–27, 2010.
- C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 36(7):1325–1339, 2014.
- X. Zhou, X. Sun, W. Zhang, S. Liang, and Y. Wei. Deep kinematic pose regression. In ECCV Workshop on Geometry Meets Deep Learning, 2016
- X. Zhou, M. Zhu, S. Leonardos, K. Derpanis, and K. Daniilidis. Sparseness Meets Deepness: 3D Human Pose Estimation from Monocular Video. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

References

- B. Tekin, A. Rozantsev, V. Lepetit, and P. Fua. Direct Prediction of 3D Body Poses from Motion Compensated Sequences. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- Y. Yu, F. Yonghao, Z. Yilin, and W. Mohan. Marker-less 3D Human Motion Capture with Monocular Image Sequence and Height-Maps. In European Conference on Computer Vision (ECCV), 2016.
- G. Rogez and C. Schmid. Mocap-guided data augmentation for 3d pose estimation in the wild. In Advances in Neural Information Processing Systems, pages 3108–3116, 2016.
- F. Bogo, A. Kanazawa, C. Lassner, P. Gehler, J. Romero, and M. J. Black. Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image. In European Conference on Computer Vision (ECCV), 2016.

Monocular 3D Human Pose Estimation

Ikrima Bin Saeed