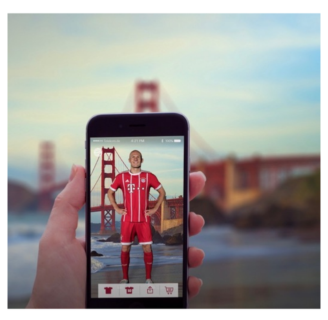
Novel Self-Supervised Deep Neural Networks for 3D Human Shape and Motion Reconstruction From a Monocular Video

Introduction

Real-world applications of 3D Human Motion Reconstruction:

3D broadcasting
Vizrt at IBC2019: The Big AR Sports Show

Virtual Reality
Ryanking999, iStock

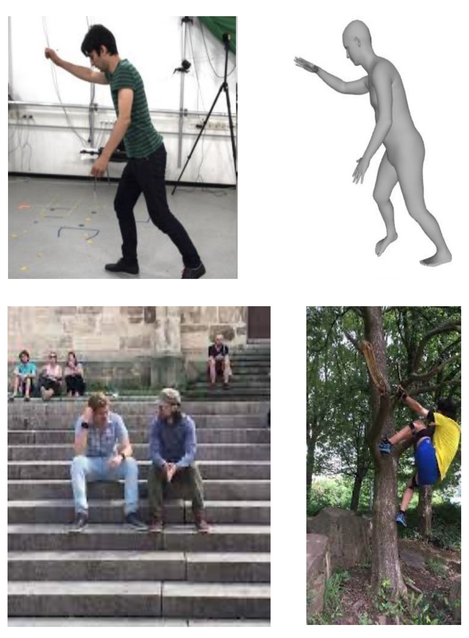


Augmented Reality
FC Bayern, 2017

Telepresence
Microsoft Research, 2016

Challenges:

- 3D reconstruction is a **missing information recovery** problem due to the absence of depth information from images/videos;
- Many **hard-to-obtain training pairs** consisting of human images/videos and their corresponding **3D models** are needed;
- Performance degradation occurs due to **poor domain adaption** between controlled settings and in-the-wild environments.



Research Problem & Objectives

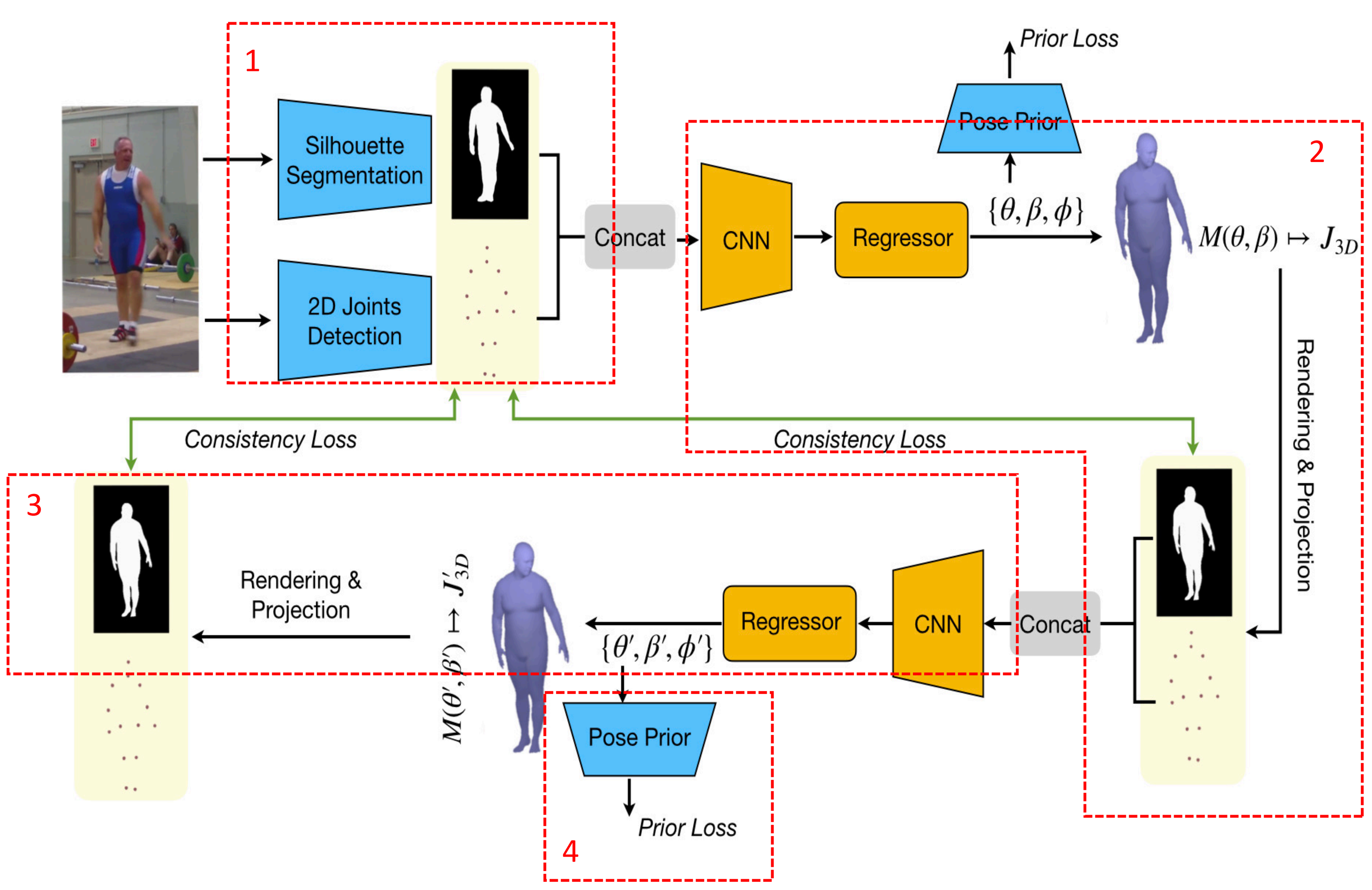
Limitations of Existing Methods:

- 3D sensors are costly & not readily available in the real world
- Annotated 2D-3D training pairs are required for supervision
- Not generalizable to versatile human motion due to heavy reliance on 3D supervision

Objectives:

1. Design a deep neural network to reconstruct a 3D model of human motion with **self-supervision** instead of relying on annotated 2D-3D training pairs;
2. Employ 2D joint locations and silhouettes to form a **geometric representation** to combat the negative impacts of appearance-based representations.

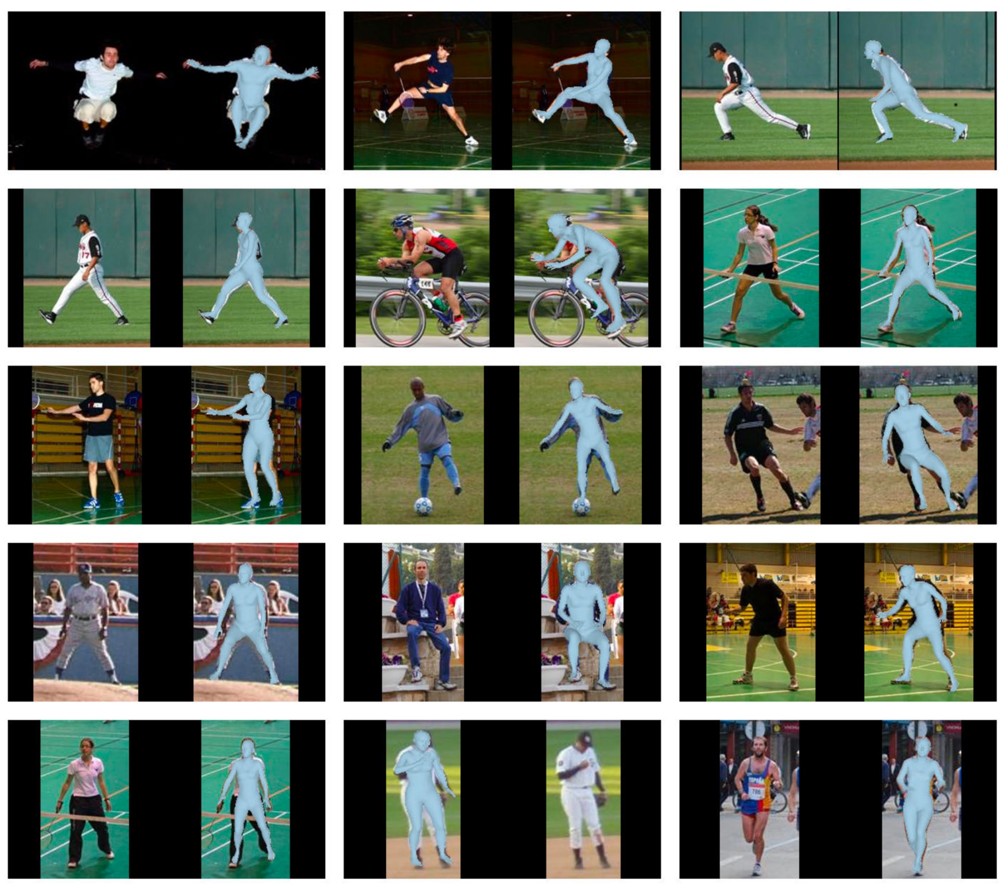
Geometric Consistency-based Self-Supervised Neural Network (GC-SSN) Architecture



1. **Geometric Representation**: silhouette and 2D joints extracted from input image
2. **3D Human Model Generator**: convolutional neural network (CNN) and multilayer perceptron nonlinearly regresses features to obtain parameters for reconstructing the 3D model under self-supervision
3. **Cycle-Consistency**: the rendered 2D representation of the reconstructed 3D model is fed through the 3D Model Generator again
4. **Pose Prior**: the reconstructed 3D model is compared with the distribution of all possible human poses to penalize unnatural reconstructions

Experiments & Results

- GC-SSN trained and tested on public benchmark datasets
- Accurate 3D human motion models reconstructed from low-resolution input images
- Outperforms state-of-the-art

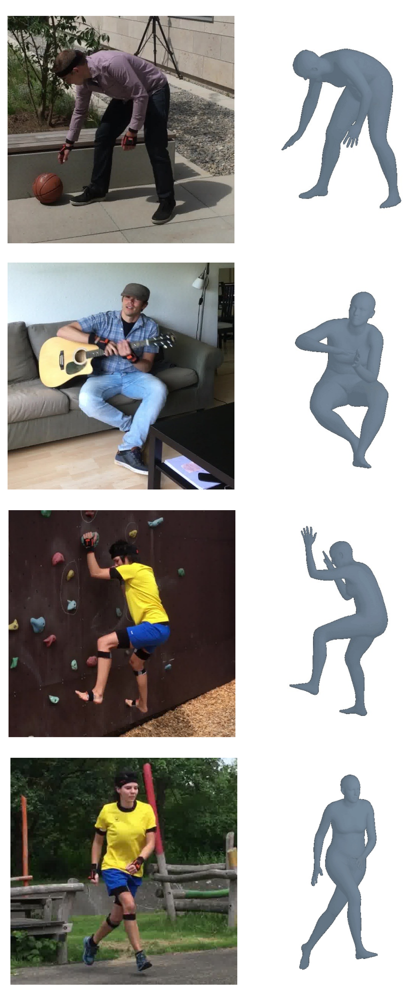


Frame-based Methods	Human3.6M MPJPE ↓	Human3.6M PA-MPJPE ↓	3DPW MPJPE ↓	3DPW PA-MPJPE ↓	3DPW MPVPE ↓
SMPLify [Bogo et al., 2016]	-	82.3	-	-	-
HMR [Kanazawa et al., 2018]	88.0	56.8	-	81.3	-
GraphCMR [Kolotouros et al., 2019b]	-	50.1	-	70.2	-
SPIN [Kolotouros et al., 2019a]	-	**41.1**	-	59.2	116.4
Pose2Mesh [Choi et al., 2020]	64.9	46.3	88.9	58.3	106.3
GC-N (2D+3D GT) (Mine)	**62.3**	44.2	**85.3**	**56.5**	**102.1**

*All photos, graphs, and images are created by the researcher unless indicated otherwise.

Conclusions

- Novel **GC-SSN** proposed to reconstruct 3D human motion
- **Geometric representation** and **cycle-consistency** overcome appearance domain gap
- GC-SSN is **self-supervised**, avoiding all manual annotations and 3D GT data acquisitions
- GC-SSN **outperforms state-of-the-art** approaches
- GC-SSN accurately handles 3D human shape and **motion reconstruction from 2D videos**



GC-SSN integrated into a HoloLens-enabled augmented reality-based remote coaching application