A

Seminar Presentation on

“ 3D Human Pose Machines with Self-Supervised Learning”

Prepared by:

Varpe Pratiksha Maruti

Roll No:-3261

Guided by:

Mr.R.S.Gaikwad Sir

Academic Year- 2020-21

Department of Computer Engineering

Amrutvahini College of Engineering, Sangamner

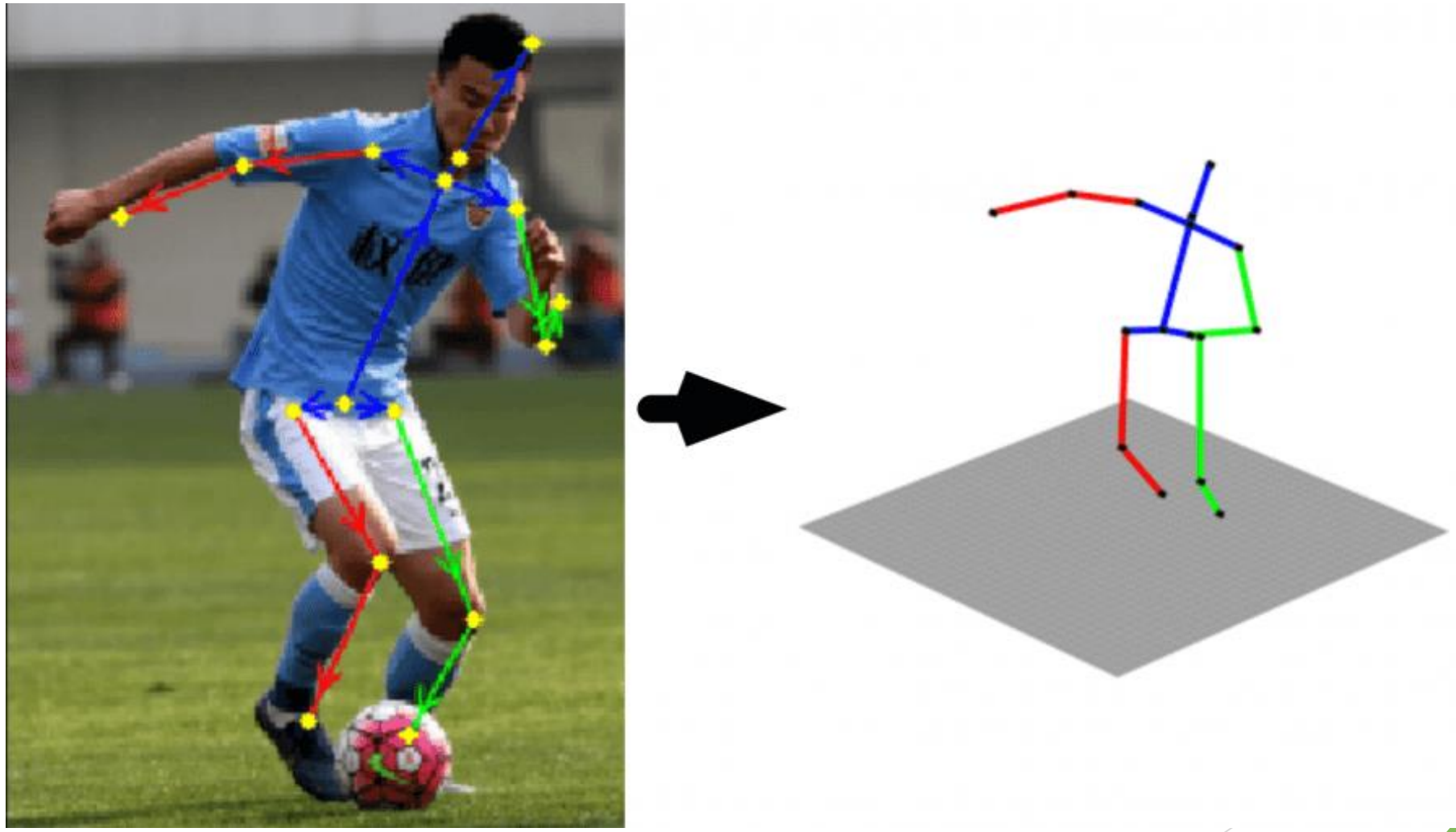
Content

Introduction

The estimation of 3D human pose from a single image is a longstanding problem with many applications .In fact, estimating human pose from images is quite challenging with respect to large variances in human appearances, arbitrary view-points, invisibilities of body parts.

To address this issue they propose an effective yet efficient 3D human pose estimation framework, which implicitly learns to integrate the 2D spatial relationship, temporal coherency and 3D geometry knowledge by utilizing the advantages afforded by Convolutional Neural Networks(CNNs)

Monocular Image to 3D Human Pose



Motivation

- Appearance/size/shape of people can vary dramatically
- The bones and joints are observable indirectly
- Occlusions
- High Dimensionally of the state space
- Lose of depth information in 2D image projections

Objectives

- To estimate the 3D human pose for images
- To detect the XYZ coordinates of a specific number of joints(keypoints) on human body by using an image contain person
- To train a model capable of interfering 3D keypoints directly from the provided image
- To detect 2D points keypoints and then transform them into 3D

LITERATURE SURVEY

Sr No.	Paper	Technology	Limitations
1	C. Wang, Y. Wang, Z. Lin, A. L. Yuille, and W. Gao, "Robust estimation of 3D human poses from a single image," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2014, pp. 2369–2376.	Robust .3D pose estimation	Frequently get large errors in the position of some 2D points
2	H. Yasin, U. Iqbal, B. Kruger, A. Weber, and J. Gall, "A dual-source € approach for 3D pose estimation from a single image," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 4948–4956.	Dual source 3D pose estimation	Time efficiency . Processing an image requires more than 20 sec
3	M. Sanzari, V. Ntouskos, and F. Pirri, "Bayesian image based 3D pose estimation," in Proc. Eur. Conf. Comput. Vis., 2016, pp. 566–582.	Hierarchical Bayesian non-parametric model	Overlapping objects
4	S. Li and A. B. Chan, "3D human pose estimation from monocular images with deep convolutional neural network," in Proc. Asian Conf. Comput. Vis., 2014, pp. 332–347.	Convolutional neural network	The prediction problem
5	L. Sigal, M. Isard, H. Haussecker, and M. J. Black : Estimating 3D human pose " Int. J. Comput. Vis., vol. 98, no. 1, pp. 15–48, 2012.	Convolutional neural network	the inability to recover from tracking failture

Dataset Description

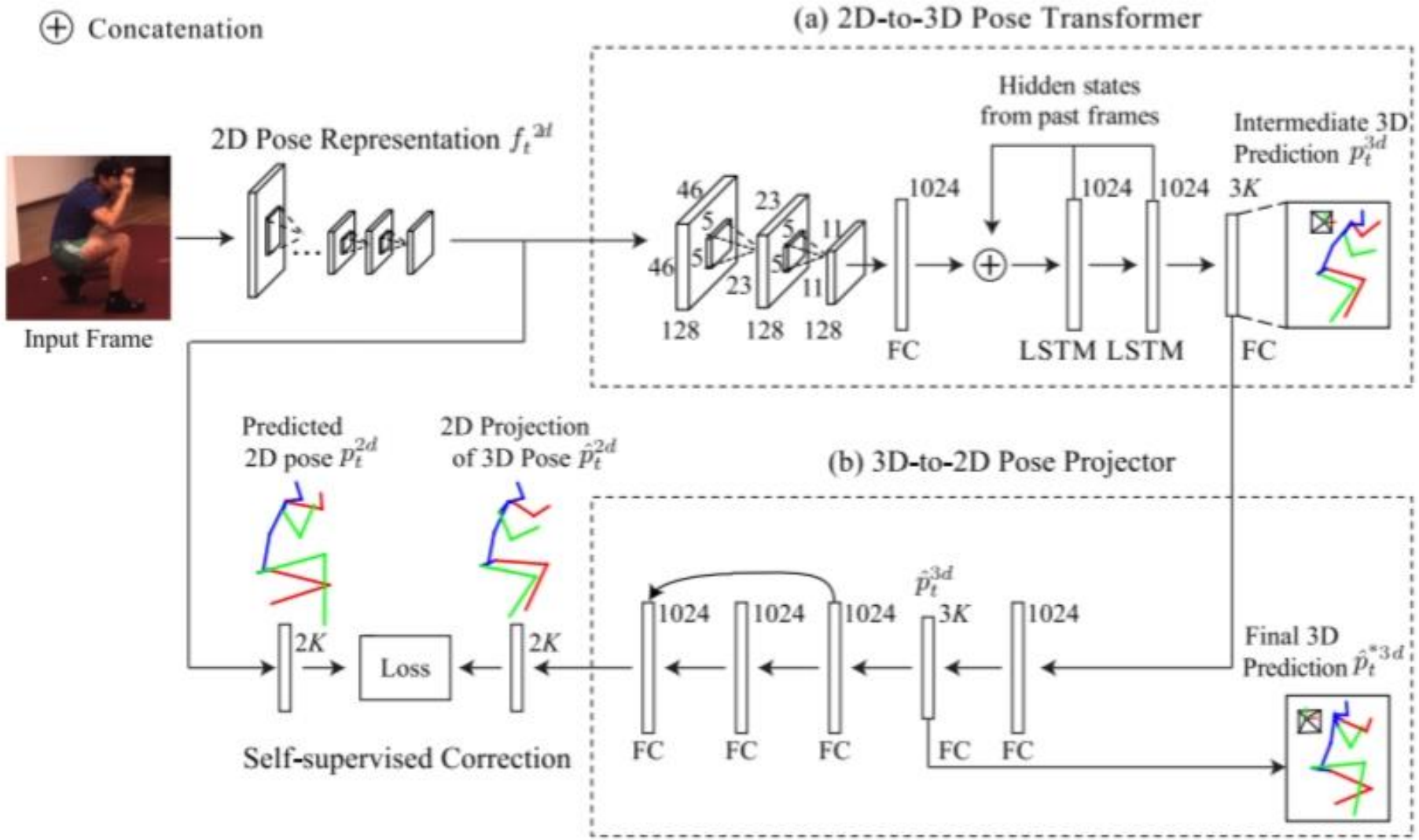
- **Human3.6M Dataset**:

The human3.6M dataset that provides 3.6 millions 3D human pose images and corresponding annotations from the controlled laboratory environment.This dataset capture 11 professional actor performing in 15 scenarios under 4 different viewpoint

- **HumanEva-1 Dataset**

The humaneva-1 dataset contain video sequences of four subjects performing six common actions and its also provide 3D pose annotation for each frame in video sequence

System Architecture



Algorithm

- Given the input image sequence with N frames
- The 2D-pose sub-network is first employed to extract the 2D pose-aware features and predict the 2D pose for the tth frame of the input sequence.
- Then, the extracted 2D poseaware features are further fed into the 2D-to-3D pose transformer module to obtain the intermediate 3D pose , where 2D-to-3D pose transformer module is composed of the hidden states learned from the past frames.
- Finally, the predicted 2D poses and intermediate 3D pose are fed into the 3D-to-2D projector module with two functions, to obtain the final 3D poses

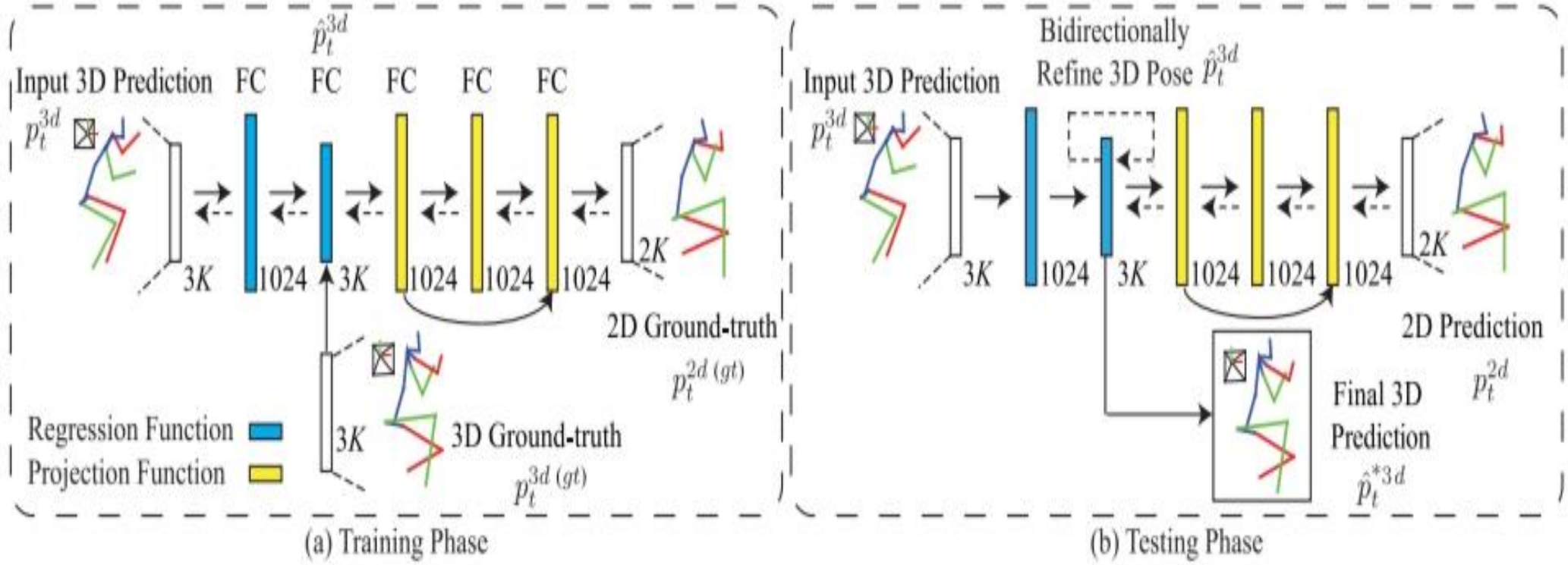
2D Pose Sub-Network

- The objective of the 2D pose sub-network is to encode each frame in a given monocular sequence with a compact representation of the pose information, e.g., the body shape of the human. The shallow convolution layers often extract the common low-level information, which is a very basic representation of the human image. the 2D pose sub-network takes the 368*368 image as input, and it outputs the 2D pose-aware feature maps with a size of 128*46*46 and the predicted 2D pose vectors with 2K entries being the argmax positions of these feature maps.

2D-to-3D Pose Transformer Module

- Based on the features extracted by the 2D pose sub-network, the 3D pose transformer module is employed to adapt the 2D pose-aware features in an adapted feature space for the later 3D pose prediction. two convolutional layers and one fully connected layer are leveraged. Each convolutional layer contains 128 different kernels with a size of 5*5 and a stride of 2, and a max pooling layer with a2*2kernel size and a stride of 2 is appended on the convolutional layers. Finally, the convolution features are fed to a fully connected layer with 1,024 units to produce the adapted feature vector. In this way, the 2D pose-aware features are transformed into the 1,024-dimensional adapted feature vector.

3D-to-2D Projector Module



Advantages and Disadvantages

- **Advantages**

High accuracy without compromise on execution performance

Simple to train as per the paper

Minimal domain knowledge

Built to avoid errors in existing models such as incorrect localization or recognition

- **Disadvantages**

High computational cost

Application

- Motion Capture and Augmented Reality:

  An interesting application of human pose estimation is for CGI applications. Graphics styles, fancy enhancements , equipments and artwork can be superimposed on the person if their human pose can be estimated. By tracking the variations of this human pose, the rendered graphics can “naturally fit” the person as they move

- Activity Recognition

Tracking the variations in the pose of a person over a period of time can also be used for activity, gesture and gait recognition. There are several use cases for the same, including:

- *Applications to detect if a person has fallen down or is sick.*
- *Applications that can autonomously teach proper work out regimes, sport techniques and dance activities.*
- *Applications that can understand full-body sign language. (Ex: Airport runway signals, traffic policemen signals, etc.).*
- *Applications that can enhance security and surveillance.*

Conclusion

Great strides have been made in the field of human pose estimation, which enables us to better serve the myriad applications that are possible with it. Moreover, research in related fields such as Pose Tracking can greatly enhance its productive utilization in several fields. This model presented a 3D human pose machine that can learn to integrate rich spatio-temporal long-range dependencies and 3D geometry knowledge in an implicit and comprehensive manner

They further enhanced their model by developing a novel self-supervised correction mechanism, which involves two dual learning tasks, i.e., 2D-to-3D pose transformation and 3D-to-2D pose projection, under a self-supervised proposed self-supervised correction mechanism can bridge the domain gap between 3D and 2D human poses to leverage the external 2D human pose data without requiring additional 3D annotations mechanism

REFERENCES

- [1] Kaze Wang, Liang Lin, Chenhan Jiang, Chen Qian,and Pengxu Wei,"3D Human pose estimation"in IEEE trans.on pattern analysis and machine intelligence VOL.42 NO.5,MAY 2020
- [2] H. Yasin, U. Iqbal, B. Kruger, A. Weber, and J. Gall, "A dual-source € approach for 3D pose estimation from a single image," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 4948–4956.
- [3S. Li and A. B. Chan, "3D human pose estimation from monocular images with deep convolutional neural network," in Proc. Asian Conf. Comput. Vis., 2014, pp. 332–347.
- [4] M. Sanzari, V. Ntouskos, and F. Pirri, "Bayesian image based 3D pose estimation," in Proc. Eur. Conf. Comput. Vis., 2016, pp. 566–582.
- [5] C. Wang, Y. Wang, Z. Lin, A. L. Yuille, and W. Gao, "Robust estimation of 3D human poses from a single image," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2014, pp. 2369–2376.
- [6] C. Held, J. Krumm, P. Markel, and R. P. Schenke, "Intelligent video surveillance," Comput., vol. 3, no. 45, pp. 83–84, 2012
- [7] A. Toshev and C. Szegedy, "DeepPose: Human pose estimation via deep neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2014, pp. 1653–1660.

THANK YOU