

# VOccl3D: A Video Benchmark Dataset for 3D Human Pose and Shape Estimation under real Occlusions

Yash Garg, Saketh Bachu, Arindam Dutta, Rohit Lal, Sarosij  
Bose, Calvin-Khang Ta, M Salman Asif, Amit Roy-Chowdhury



Vision and Learning Group

Computational Sensing and Image Processing lab



# Introduction/Motivation

## Dataset ???

- In the current literature, we still lack a realistic human pose dataset with significant occlusions.

## Purpose of Dataset ???

- We need to train the existing pose estimation models with occluded-human dataset.
- Need a better benchmark dataset to report the performance of methods that claim to work under occlusion.



a) 3DPW-AdvOcc@80



b) Occluded Human3.6M



c) Syn. Occ. Human3.6M



d) OCMotion



e) 3DPW



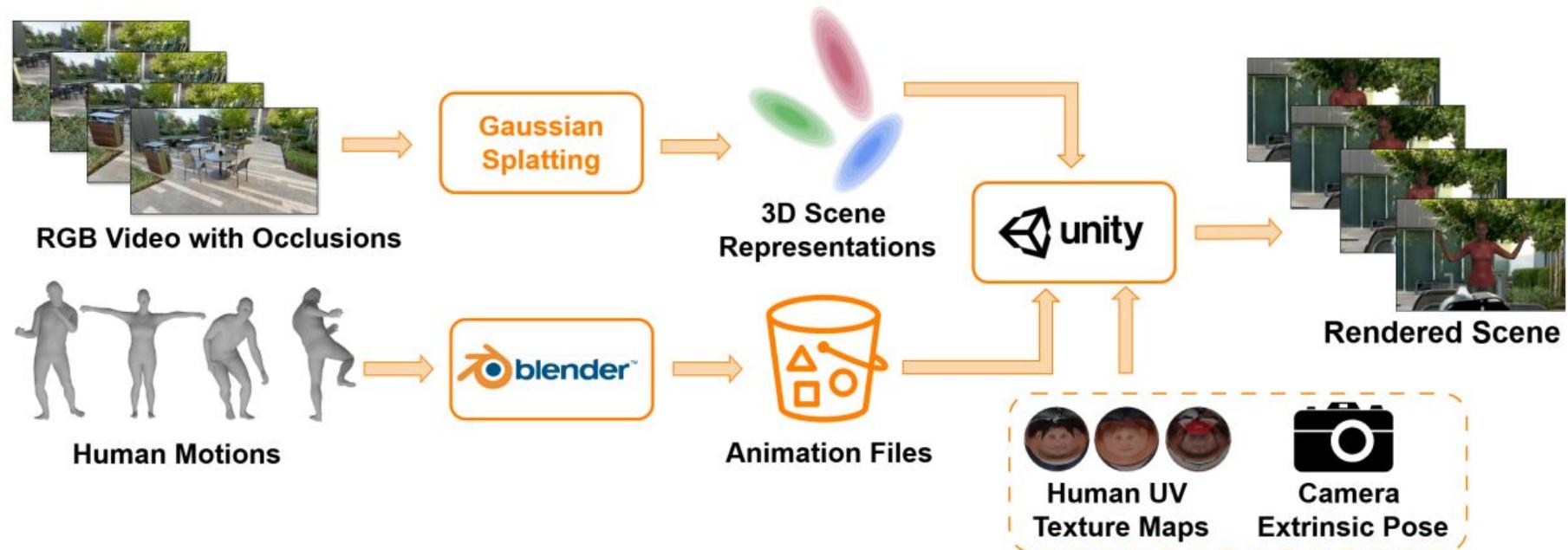
f) Ours

# Dataset

- We propose **VOccl3D**: a Video-based human **Occlusion** dataset with **3D** body pose and shape annotations



# Dataset



# Dataset

- **Annotations of VOccl3D:**
  - Camera Intrinsic
  - Camera Extrinsic
  - SMPLX pose
  - SMPLX shape
  - Gender
  - 2D-keypoints
  - Bounding Box
- **Attributes of VOccl3D:**
  - 40 background scenes
  - ~400 videos
  - ~240,000 RGB images
  - ~200 clothing texture



Figure: Different ground-truth modalities

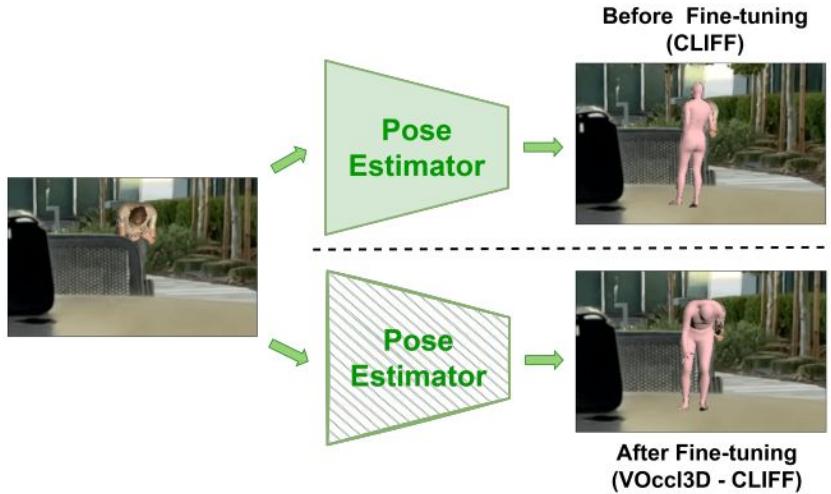
# Dataset Samples - Image



# Dataset Samples - Video



# Experimental Results



# Experimental Results

Occlusion category:

- **Hard-Occlusion:** 4-9 visible keypoints (22 total)
- **Medium-Occlusion:** 10-15 visible keypoints (22 total)
- **Low-Occlusion:** 16-20 visible keypoints (22 total)

# Experimental Results

Occlusion category:

- **Hard-Occlusion:** 4-9 visible keypoints (22 total)
- **Medium-Occlusion:** 10-15 visible keypoints (22 total)
- **Low-Occlusion:** 16-20 visible keypoints (22 total)

Method	Hard-Occlusion			Medium-Occlusion			Low-Occlusion		
	MPJPE	PA-MPJPE	PVE	MPJPE	PA-MPJPE	PVE	MPJPE	PA-MPJPE	PVE
CLIFF [28]	192.22	114.35	247.41	121.70	78.56	158.46	98.82	67.64	126.92
BEDLAM-HMR [4]	167.35	102.92	214.39	102.55	68.13	134.59	86.55	54.48	110.15
BEDLAM-CLIFF [4]	154.86	99.53	199.95	90.97	65.03	119.63	74.95	52.65	96.60
HMR2.0 [13]	169.71	100.49	215.17	113.88	71.78	145.62	88.53	59.08	114.39
STRIDE with BEDLAM-CLIFF [27]	155.64	100.44	-	91.14	65.38	-	75.02	53.21	-
WHAM [48]	152.15	102.14	177.07	110.97	76.81	127.45	93.90	66.68	106.51
VOccl3D-B-CLIFF	<b>136.34</b>	<b>89.94</b>	<b>175.92</b>	82.48	<b>58.78</b>	<b>106.84</b>	<b>69.46</b>	<b>46.32</b>	<b>88.19</b>
STRIDE with VOccl3D-B-CLIFF	136.43	90.28	-	<b>82.37</b>	58.98	-	69.65	46.86	-

# Experimental Results



# Experimental Results



**Top Row: OcclType1-3DPW**  
**Bottom Row: OcclType2-3DPW**

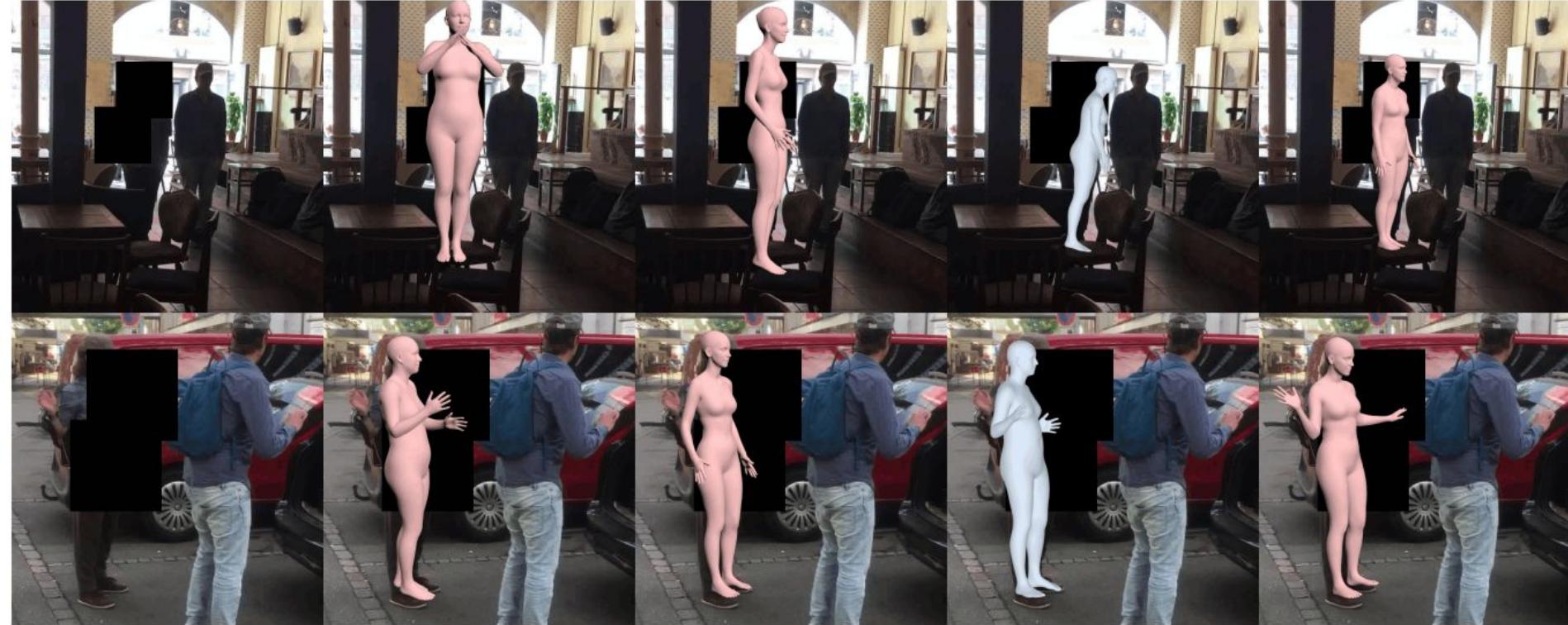
# Experimental Results



Top Row: OcclType1-3DPW  
Bottom Row: OcclType2-3DPW

Method	3DPW			OcclType1-3DPW			OcclType2-3DPW		
	MPJPE	PA-MPJPE	PVE	MPJPE	PA-MPJPE	PVE	MPJPE	PA-MPJPE	PVE
CLIFF [28]	73.9	46.4	87.6	98.15	62.27	118.47	99.49	62.16	119.82
BEDLAM-HMR [4]	79.0	47.6	93.1	108.62	66.53	128.05	106.19	64.05	125.66
BEDLAM-CLIFF [4]	72.0	46.6	85.0	98.71	64.26	117.41	96.80	61.32	115.33
HMR2.0 [13]	81.2	54.3	143.7	103.40	69.66	164.55	99.01	66.17	158.79
VOccl3D-B-CLIFF	72.0	47.3	84.5	95.89	63.43	<b>114.28</b>	94.36	60.44	<b>112.01</b>
VOccl3D-CLIFF	<b>71.10</b>	<b>45.98</b>	<b>84.25</b>	<b>95.17</b>	<b>61.83</b>	114.95	<b>93.74</b>	<b>59.66</b>	112.59

# Experimental Results



Image

CLIFF

BEDLAM-CLIFF

HMR2.0

VOccl3D-B-CLIFF

# Impact of Human Detector on HPS Estimation

- Any human pose and shape estimation method, first pass RGB image to human tracker/human detector.

# Impact of Human Detector on HPS Estimation

- Any human pose and shape estimation method, first pass RGB image to human tracker/human detector.
- The tracked/detected cropped human image is given as input to pose estimation model.

# Impact of Human Detector on HPS Estimation

- Any human pose and shape estimation method, first pass RGB image to human tracker/human detector.
- The tracked/detected cropped human image is given as input to pose estimation model.

## No Occlusion

- Tracker output 
- Pose and shape estimation 

## Occlusion

- Tracker output 
- Pose and shape estimation 

# Impact of Human Detector on HPS Estimation

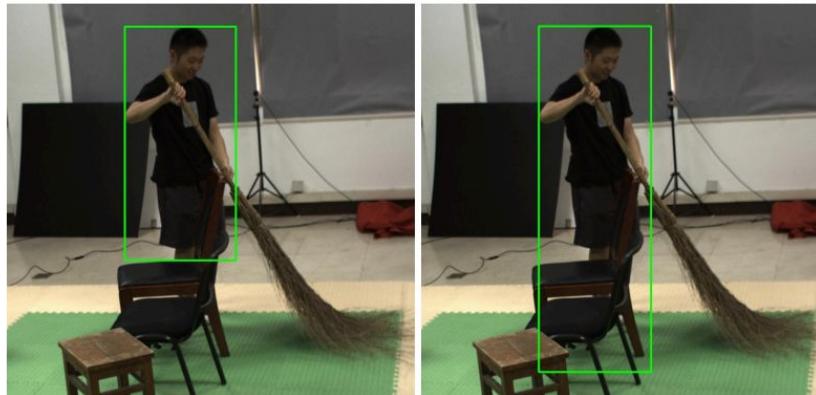
- Fine-tuned YOLO11 on VOccl3D to improve the performance of tracker under occlusion.

Method	3DPW		OCMotion	
	mAP50	mAP75	mAP50	mAP75
YOLO11	58.99	47.14	98.84	91.80
VOccl3D-YOLO11	<b>59.89</b>	<b>48.26</b>	<b>99.10</b>	<b>91.95</b>

# Impact of Human Detector on HPS Estimation

- Fine-tuned YOLO11 on VOccl3D to improve the performance of tracker under occlusion.

Method	3DPW		OCMotion	
	mAP50	mAP75	mAP50	mAP75
YOLO11	58.99	47.14	98.84	91.80
VOccl3D-YOLO11	<b>59.89</b>	<b>48.26</b>	<b>99.10</b>	<b>91.95</b>



YOLO11

VOccl3D-YOLO11

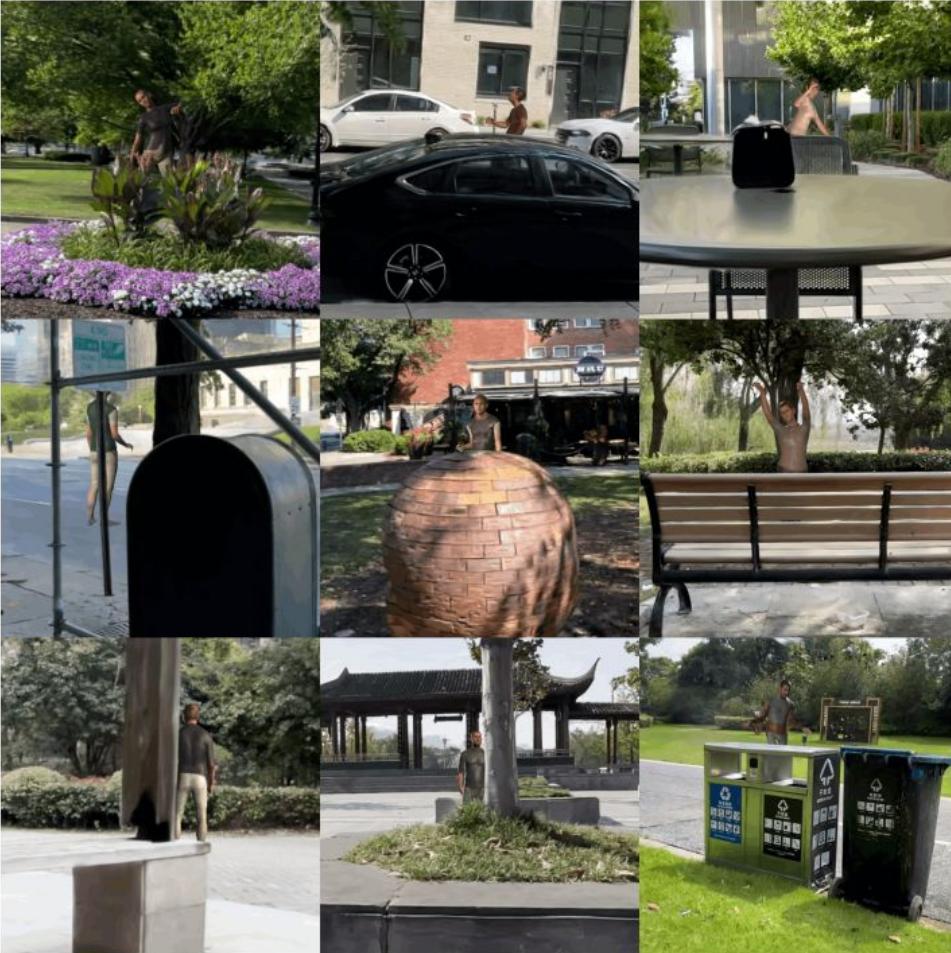
# Impact of Human Detector on HPS Estimation

- Fine-tuned YOLO11 on VOccl3D to improve the performance of tracker under occlusion.

Method	3DPW		OCMotion	
	mAP50	mAP75	mAP50	mAP75
YOLO11	58.99	47.14	98.84	91.80
VOccl3D-YOLO11	<b>59.89</b>	<b>48.26</b>	<b>99.10</b>	<b>91.95</b>

Method	3DPW		OCMotion		
	MPJPE	PA-MPJPE	PVE	MPJPE	PA-MPJPE
VOccl3D-CLIFF w/GT	71.10	45.98	84.25	64.29	39.64
VOccl3D-CLIFF w/YOLO11	116.52	63.35	139.74	67.16	41.30
VOccl3D-CLIFF w/VOccl3D-YOLO11	114.85	62.66	137.19	66.65	41.15
					82.40

# Additional samples of VOccl3D



# Thank You !

For additional info, please  
read our paper

SCAN ME

