2023 여름 세미나

2023.07.07

Sogang University
Vision & Display Systems Lab, Dept. of Electronic Engineering

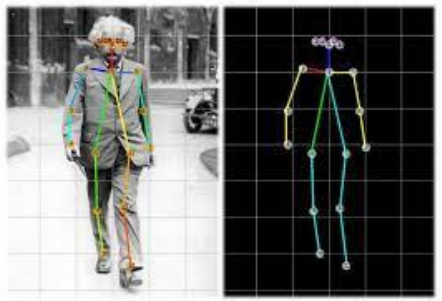
Presented By
양창희

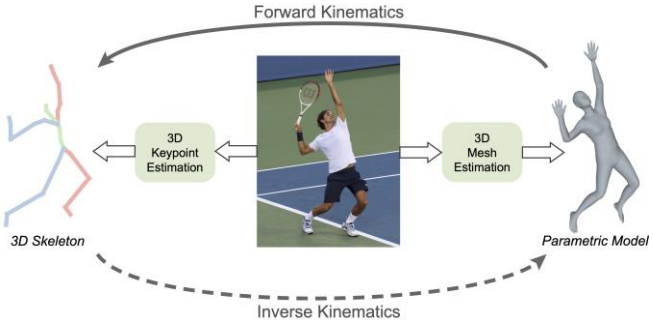
Outline

- Intro
  - What is human pose estimation
  - How to do 2d pose estimation
  - How to do 3d pose estimation
  - How to do 3d mesh estimation
- 2023 CVPR Human paper
  - Implicit 3D Human Mesh Recovery using Consistency with pose and shape from Unseen-view
  - PoseExaminer: Automated Testing of Out-of-distribution Robusteness in Human Pose and Shape estimation
  - A Characteristic Function-based Method for Bottom-up Human Pose Estimation
  - Human Pose Estimation in Extremely Low-Light Conditions
  - Scene-aware Egocentric 3D Human Pose Estimation
- Conclusion

Intro

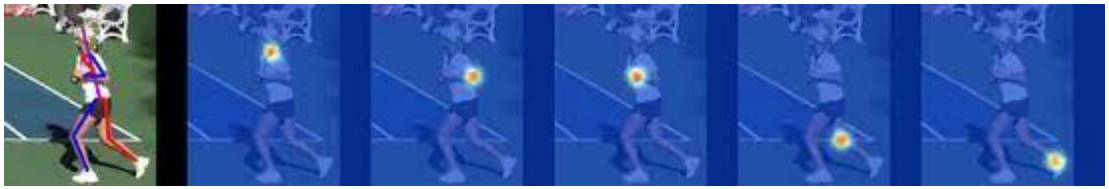
- What is human pose estimation
  - 2D pose estimation
    - Image안에 존재하는 사람의 keypoint를 추정하는 task
      - Keypoint (Ex) head, neck, ankle, …)
    - Top-down, bottom-up approach 존재
  - 3D pose estimation
    - Image안에 존재하는 사람의 world coordinate를 구하는 task
    - Top-down, bottom-up approach, Temporal, 2d to 3d 존재
  - 3D mesh estimation
    - Image안에 존재하는 사람의 world coordinate mesh를 구하는 Task
    - Top-down, bottom-up, 2d to mesh, temporal, optimization 등 존재





Intro

- How to do 2d pose estimation
  - Regression, Heatmap approach
    - Regression 방법은 말 그대로 즉시 pose를 찾는 기술을 보여줌
      - 대부분 resnet50 backbone을 거쳐서 바로 회귀하는 방법을 이용함
        - ✓정확도 측면에서 성능이 떨어져서 사용하지 않는 기술임
    - Heatmap은 가우시안 분포를 이용해 만들어 inference할 때 softargmax를 이용해 joint의 위치를 구함



      - 대부분의 2d pose에서 사용하는 기술로 regression보다 높은 성능을 보여줌
  - 최근에 들어서 메인 conference 마다 하나씩 개제가됨
    - 그만큼 Accept되기 힘든 분야이며 성능을 이끌어내기 힘듦

Intro

- How to do 3d pose estimation
  - Regression, Heatmap, Temporal, 2d to 3d approach
    - Regression: 2d pose와 마찬가지로 잘 사용되지 않는 기술임
      - 속도 측면으로 가끔 사용함
    - Heatmap: 대부분의 3D pose 논문에서 사용함
      - 3D heatmap등을 사용하면서 성능을 이끌어냄
    - Temporal: 비디오 데이터에서 사용하는 방법으로 시간 데이터를 이용해 3D joint를 얻어냄
      - 3D Heatmap과 성능 비교를 했을 때 엄청난 성능을 이끌어내지는 못함
      - Model 크기 제한도 존재함
    - 2D to 3D: Temporal의 단점을 보완하기 위해서 나온 프레임워크 모델이라고 생각 가능
      - 2D pose를 구하고 이를 기반으로 3D Pose를 이끌어내는 방법임
      - 대표적인 논문
        - ✓3D human pose estimation in video with temporal convolutions and semi-supervised training 가 있음

Intro

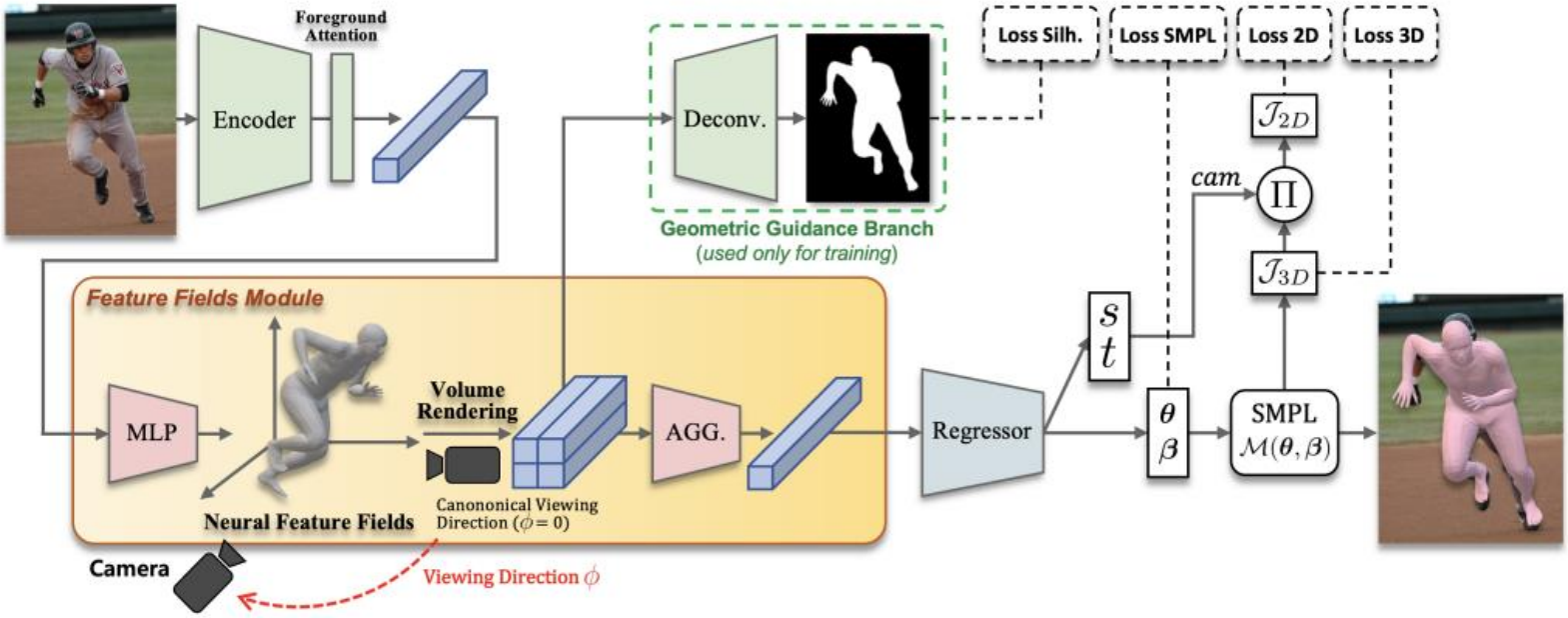
- How to do 3d mesh estimation
  - Regression, Heatmap, Temporal, 2d to mesh, optimization
    - Regression(model-based): 대부분 사용하는 방식
      - SMPL parameter를 추정해야 되다보니 heatmap보다 더욱 활용됨
    - Heatmap(model-free): I2L-Mesh Net
      - I2L-Mesh Lixel heatmap 방식을 이용함
    - Temporal(resnet50 backbone): resnet50의 latent space를 이용해 구함
      - 시간 효율상 resnet50 backbon을 이용해서 구함
    - 2D to mesh: pose2mesh
      - 2D pose를 구한후 3d joint를 구한 다음 mesh를 구하는 방식임
    - Optimization (EFT, SPIN 등)
      - 대부분의 pseudo gt로 사용하기 위해서 적용하는 방식이면서 성능 개선이 이뤄져 사용함

2023 CVPR Human papaer

- Why choose this paper
  - CVPR 2023 논문 총 5개 간단하게 revie를 목표로함
    - 해당 논문들을 고른 이유는 흥미로운 concept의 모델이 많아서 고름
    - Implicit 3D Human Mesh Recovery using Consistency with pose and shape from Unseen-view
      - Viewing point를 바꿔서 optimization을 진행
    - PoseExaminer: Automated Testing of Out-of-distribution Robusteness in Human Pose and Shape estimation
      - 강화학습을 이용해 새로운 evaluation metric을 제안
    - A Characteristic Function-based Method for Bottom-up Human Pose Estimation
      - Heatmap approach의 문제점을 제시하면서 해결 방안을 보임
    - Human Pose Estimation in Extremely Low-Light Conditions
      - 극도로 어두운 이미지에서의 인간 자세 추정을 목표로함
    - Scene-aware Egocentric 3D Human Pose Estimation
      - Egocentric dataset을 만듦과 동시에 ego pose estimation를 진행함

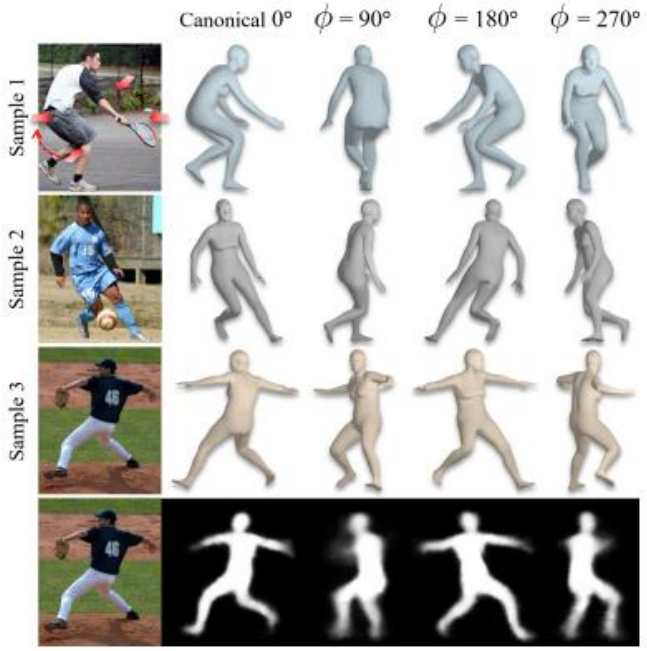
2023 CVPR Human papaer

- Implicit 3D Human Mesh Recovery using Consistency with pose and shape from Unseen-view
  - 문제점 제시
    - 기존의 human Mesh recovery 논문들은 촬영된 방향만을 고려한다.
    - 이를 개선하면서 optimization을 할 수 있도록 만든다.



2023 CVPR Human papaer

- PoseExaminer: Automated Testing of Out-of-distribution Robusteness in Human Pose and Shape estimation
  - 결과



	Method	3PDW MPJPE ↓	3PDW PA-MPJPE ↓	3PDW PVE ↓
Temporal	HMMR [20]	116.5	72.6	139.3
	DSD [50]	-	69.5	-
	Arnab *et al.* [2]	-	72.2	-
	Doersch *et al.* [11]	-	74.7	-
	VIBE [23]	93.5	56.5	113.4
	TCMR [8]	95.0	55.8	111.3
	MPS-Net [55]	91.6	54.0	109.6
Frame-based	HMR [19]	130.0	76.7	-
	GraphCMR [26]	-	70.2	-
	SPIN [25]	96.9	59.2	116.4
	PyMAF [62]	92.8	58.9	110.1
	I2L-MeshNet [39]	100.0	60.0	-
	ROMP [49]	89.3	53.5	105.6
	HMR-EFT [17]	-	54.2	-
	PARE [24]	82.9	52.3	99.7
	ImpHMR (Ours)	**81.8**	**49.8**	**96.4**
	ImpHMR (Ours) w. 3DPW	74.3	45.4	87.1

2023 CVPR Human papaer

- Implicit 3D Human Mesh Recovery using Consistency with pose and shape from Unseen-view
  - 문제점 제시
    - Human pose and shape estimation 모델들이 훈련 데이터 시나리오에서 너무 피팅 되어 있어 문제를 야기
    - Evaluation metric이 평균화 되어 있어서 실패 case를 제대로 못찾음

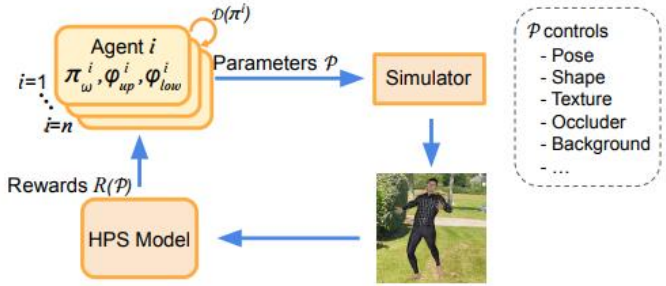
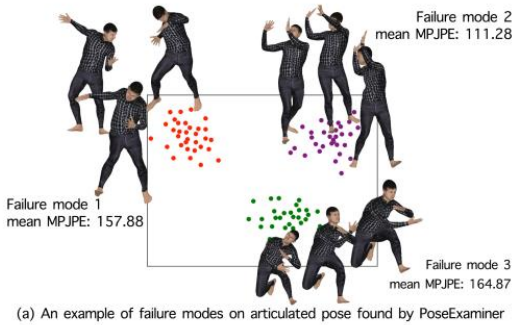
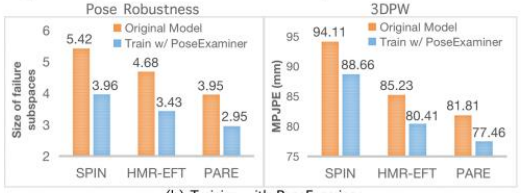


Figure 2. **PoseExaminer model pipeline.** Multiple RL agents collaborate to search for worst cases generated by a policy π_ω and find subspaces boundaries defined by ϕ_{up} and ϕ_{low}. The human simulator is conditioned on parameters generated by the agents. m images are then generated for each agent and are used to test a given HPS method. The prediction error of the HPS model serves as the reward signal to update the policy parameters of agents.



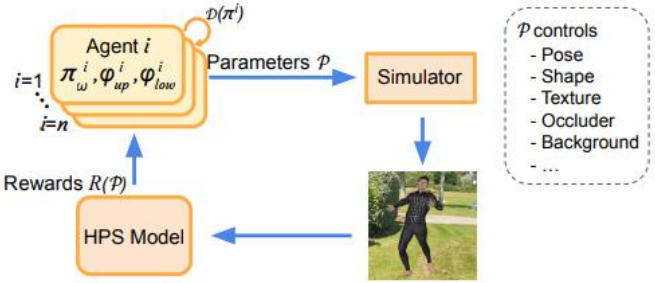
(a) An example of failure modes on articulated pose found by PoseExaminer



(b) Training with PoseExaminer

2023 CVPR Human papaer

- Implicit 3D Human Mesh Recovery using Consistency with pose and shape from Unseen-view
  - Evaluation 방법
    - P controls을 통해서 새로운 이미지를 만들고 human pose and shape model에 성능을 평가해 어떤 pose, shape이 잘 안나오는지 판단이 가능함
    - 또한 이를 이용해 augmentation 느낌으로 훈련도 가능함
      - 대부분의 모델이 성능이 개선된것을 볼 수 있음
    - Shape, pose를 만들 때 latent space를 강화학습을 통해 훈련을 진행함

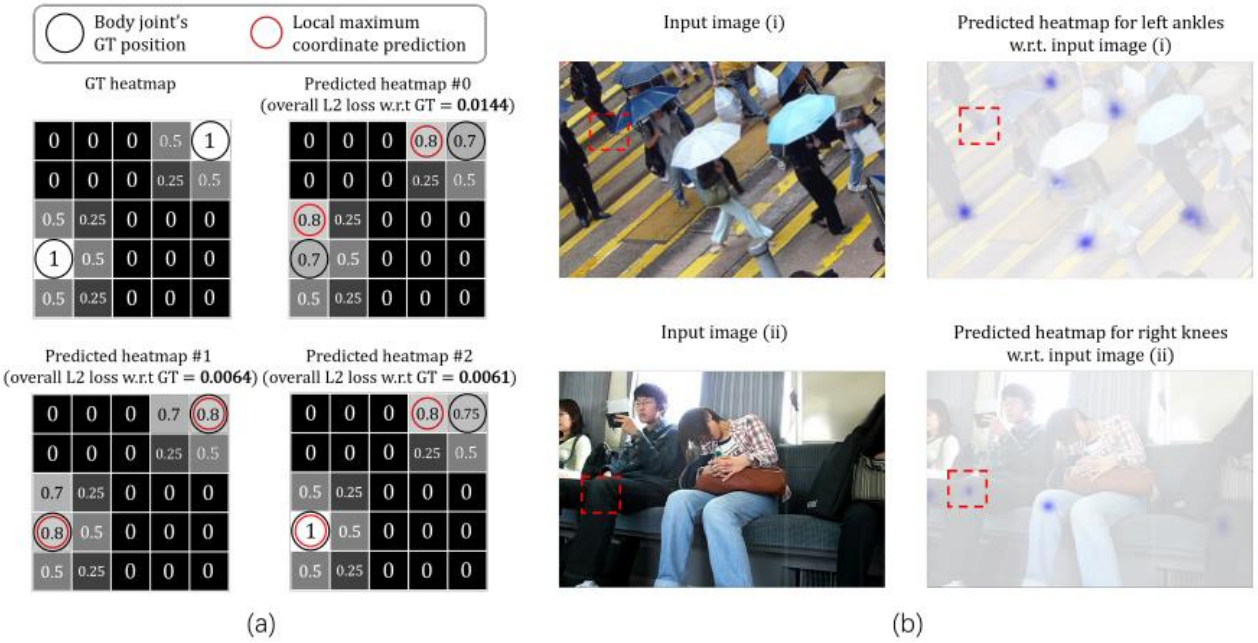


	Robustness (PoseExaminer)		ID dataset (3DPW)			OOD dataset (cAIST)		Extreme poses (cAIST-EXT)	
	Succ. Rate↓	Region Size↓	MPJPE↓	PA-MPJPE↓	PVE↓	MPJPE↓	PA-MPJPE↓	MPJPE↓	PA-MPJPE↓
SPIN [20]	95.0%	5.417	94.11	57.54	111.12	108.09	68.59	133.65	81.03
+ PE (Ours)	76.6% (-18.4)	3.964 (-1.435)	88.66 (-5.45)	54.34 (-3.20)	103.94 (-7.18)	98.88 (-9.21)	65.28 (-3.31)	120.98 (-12.67)	76.82 (-4.21)
HMR-EFT [15]	84.1%	4.675	85.23	51.88	107.88	98.55	66.01	122.19	78.39
+ PE (Ours)	63.6% (-20.5)	3.429 (-1.246)	80.41 (-4.82)	49.15 (-2.73)	101.43 (-6.45)	88.53 (-10.02)	64.00 (-2.01)	112.78 (-9.41)	75.15 (-3.24)
PARE [18]	75.3%	3.948	81.81	50.78	102.27	99.15	62.43	117.45	72.68
+ PE (Ours)	48.4% (-26.9)	2.953 (-0.995)	77.46 (-4.35)	48.01 (-2.77)	94.86 (-7.41)	87.44 (-11.71)	59.80 (-2.63)	109.82 (-7.63)	70.73 (-1.95)

Table 4. **Fine-tuning with PoseExaminer.** 'PE' is short for PoseExaminer. Despite being synthetic, fine-tuning on the failure modes discovered by PoseExaminer can significantly improve the robustness and real-world performance of all methods on all benchmarks.

2023 CVPR Human papaer

- A Characteristic Function-based Method for Bottom-up Human Pose Estimation
  - 문제점 제기
    - Heatmap Approach의 문제점을 제시함



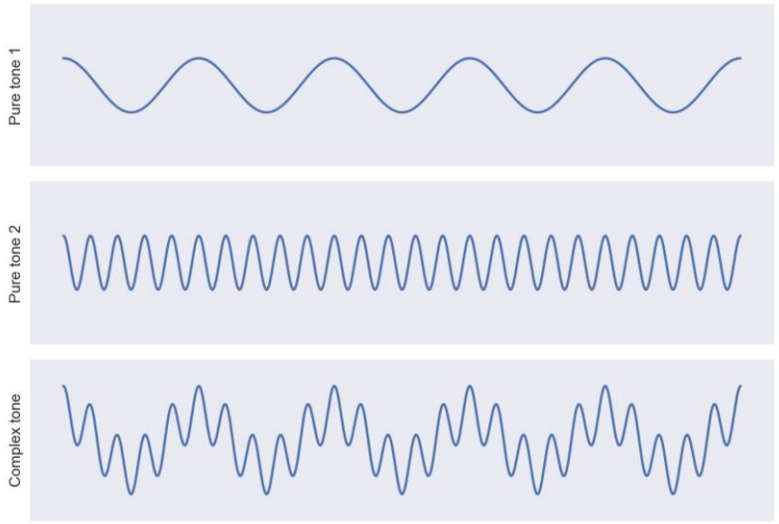
- 아래 그림처럼 loss가 줄어드는 방향이 적절한 pose 위치를 찾지 못한다고 언급하고 있음

2023 CVPR Human papaer

- A Characteristic Function-based Method for Bottom-up Human Pose Estimation
  - 해결방안
    - 아래 수식과 같은 Heatmap의 characteristic function을 만듦

$$\varphi_D(\mathbf{t}) = E_{\mathbf{x}\sim D}[e^{i\langle \mathbf{t},\mathbf{x}\rangle}] = \int_{\mathbb{R}^N} e^{i\langle \mathbf{t},\mathbf{x}\rangle}\, dD$$

    - 이 수식은 푸리에 변환을 이용해서 만든 수식임
      - 왜 푸리에 변환을 사용했나?



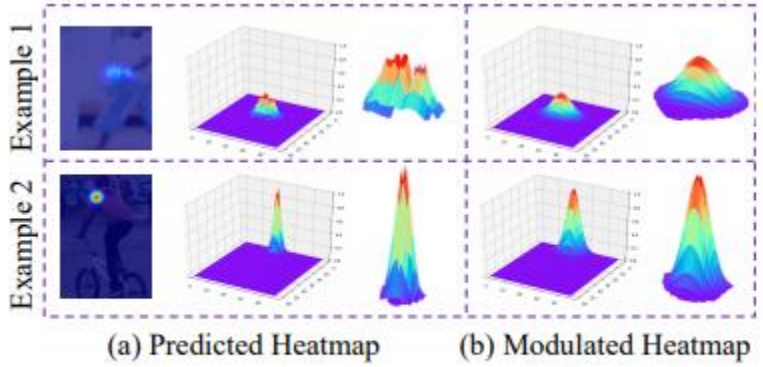


Figure 3: Illustration of heatmap distribution modulation. (a) Predicted heatmap; (b) Modulated heatmap distribution.

2023 CVPR Human papaer

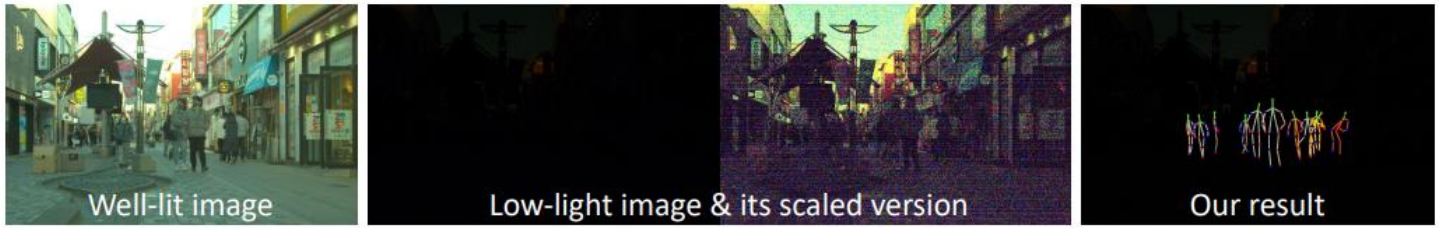
- A Characteristic Function-based Method for Bottom-up Human Pose Estimation
  - 결과

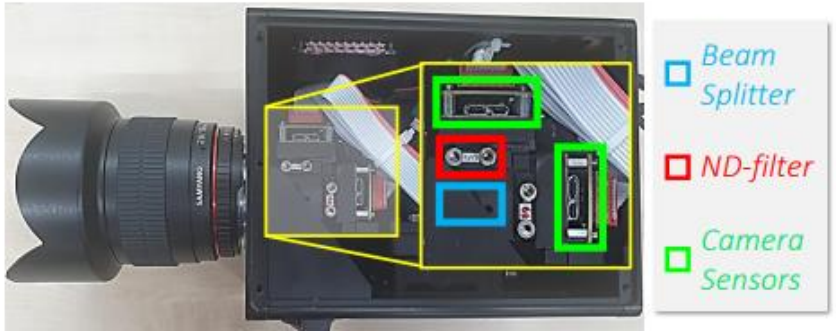
	(a)	(b)	(c)	(d)
Body joint name:	Right shoulders	Right eyes	Right elbows	Left shoulders
Baseline (HrHRNet-W32) result:				
Ours result:				

Method	Venue	Backbone	Input size	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}
HGG [16]	ECCV 2020	Hourglass	512	68.3	86.7	75.8	-	-
Point-Set Anchors [38]	ECCV 2020	HRNet-W48	640	69.8	88.8	76.3	-	-
DEKR [11]	CVPR 2021	HRNet-W48	640	72.3	88.3	78.6	68.6	78.6
SWAHR [22]	CVPR 2021	HrHRNet-W32	512	71.4	88.9	77.8	66.3	78.9
SWAHR [22]	CVPR 2021	HrHRNet-W48	640	73.2	89.8	79.1	69.1	79.3
PoseTrans [15]	ECCV 2022	HrHRNet-W32	512	71.2	88.2	77.2	66.5	78.0
HrHRNet [6]	CVPR 2020	HrHRNet-W32	512	69.9	87.1	76.0	65.3	77.0
+ Ours		HrHRNet-W32	512	**71.8(↑1.9)**	88.9	78.1	67.3	78.4
HrHRNet [6]	CVPR 2020	HrHRNet-W48	640	72.1	88.4	78.2	67.8	78.3
+ Ours		HrHRNet-W48	640	**73.7(↑1.6)**	89.9	79.6	69.6	79.5

2023 CVPR Human papaer

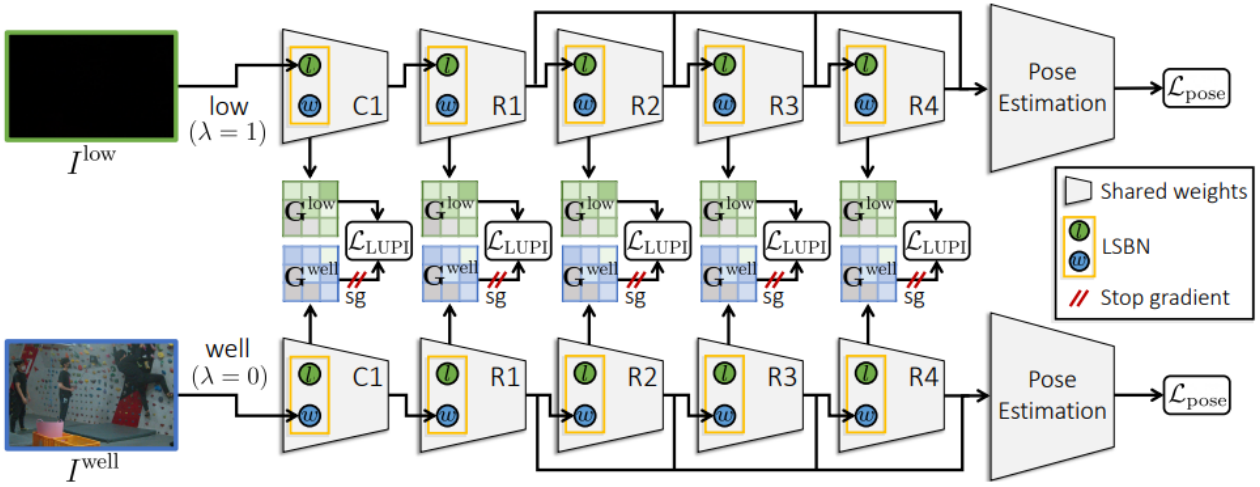
- Human Pose Estimation in Extremely Low-Light Conditions
  - 문제제기
    - 극도로 어두운 이미지에서의 인간 자세 추정의 연구가 어려움
    - 데이터 수집의 어려움
    - 어두운 이미지에서의 자세 추정 및 annotation이 어려움
  - 해결 방안
    - 카메라에서 어두운 이미지 + 밝은 이미지를 만들고 밝은 이미지에서만 annotation을 진행함





2023 CVPR Human papaer

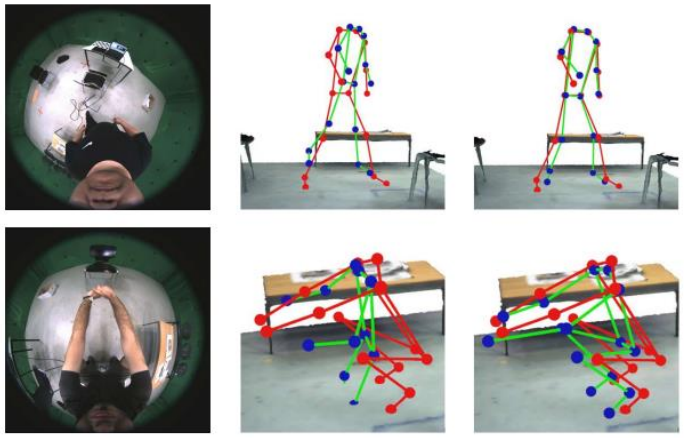
- Human Pose Estimation in Extremely Low-Light Conditions
  - 해결 방안
    - Teacher & Student Model을 설계함



	Training data			AP@0.5:0.95					Param. (M)	Latency (sec)
	LL	WL	Enhanced-LL	LL-N	LL-H	LL-E	LL-A	WL		
Baseline-low	✓			32.6	25.1	13.8	24.6	1.6	27.37	1.07
Baseline-well		✓		23.5	7.5	1.1	11.5	**68.8**	27.37	1.07
Baseline-all	✓	✓		33.8	25.4	14.3	25.4	57.9	27.37	1.07
LLFlow + Baseline-all		✓	✓	35.2	20.1	8.3	22.1	65.1	66.23	3.34
LIME + Baseline-all		✓	✓	38.3	25.6	12.5	26.6	63.0	27.37	1.65
DANN	✓	✓		34.9	24.9	13.3	25.4	58.6	27.37	1.07
AdvEnt	✓	✓		35.6	23.5	8.8	23.8	62.4	27.37	1.07
Ours	✓	✓		**42.3**	**34.0**	**18.6**	**32.7**	68.5	27.53	1.07

2023 CVPR Human papaer

- Scene-aware Egocentric 3D Human Pose Estimation
  - 문제점 제기
    - 대부분의 human pose estimation의 경우 egocentric에서 진행하지 않고 일반적인 이미지에서 만 진행함
      - 이때의 문제점은 XR, VR, AR과 같은 기기에서 pose estimation이 적절히 되기 어렵다는 점이 있음
    - 또한 지금까지의 egocentric dataset들은 depth에 대한 정보가 없기 때문에 정확한 pose 추정이 어렵다고 언급함
      - Dataset으로 depth에 대한 정보를 포함한 Dataset을 구축함



2023 CVPR Human papaer

- Scene-aware Egocentric 3D Human Pose Estimation
  - 해결방안
    - Depth 정보를 포함하는 Dataset을 만듦과 동시에 scene에대한 depth estimation model을 만들어
    - 3D pose의 대략적인 위치를 찾을 수 있는 모델을 만듦

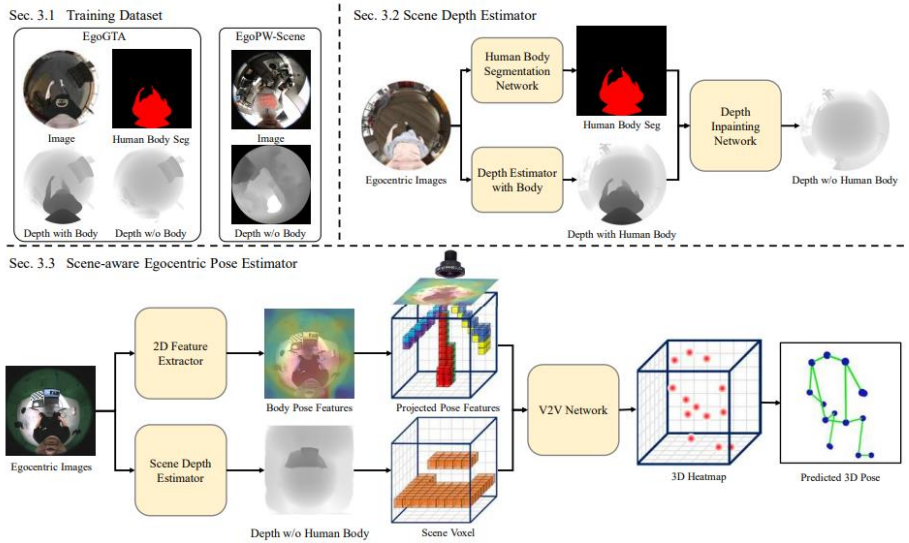


Figure 2. Overview of our method. We first render synthetic training dataset EgoGTA and in-the-wild training dataset EgoPW-Scene. Both datasets contain egocentric depth maps for subsequent training process (Sec. 3.1). Next, we train an egocentric scene depth estimator that predicts a depth map without the human body and a depth inpainting network (Sec. 3.2). Finally, we combine the 2D body pose features and scene depth map into a common voxel space. The 3D body pose heatmaps are regressed from the voxel space with a V2V network and the final pose prediction is obtained with soft-argmax (Sec. 3.3).

Method	MPJPE	PA-MPJPE
Our test dataset		
Mo^{2}Cap^2 [39]	200.3	121.2
xR-egopose [32]	241.3	133.9
EgoPW [35]	189.6	105.3
Ours	**118.5**	**92.75**
Method	PA-MPJPE	BA-MPJPE
Wang *et al.*'s dataset [36]		
Mo^{2}Cap^2 [39]	102.3	74.46
xR-egopose [32]	112.0	87.20
EgoPW [35]	81.71	64.87
Ours	**76.50**	**61.92**
Mo^{2}Cap^2 test dataset [39]		
Mo^{2}Cap^2 [39]	91.16	70.75
xR-egopose [32]	86.85	66.54
EgoPW [35]	83.17	64.33
Ours	**79.65**	**62.82**

2023 CVPR Human papaer

- Conclusion
  - 최근에 들어서 다양한 문제제기를 하는 논문이 증가함
    - 이 말의 즉슨 2D, 3D, Mesh등 대부분의 model들의 성능이 어느정도 적절히 달성했다는 것을 의미함.
      - 어느정도 성능이 달성했음으로 다양한 문제제기를 통해 연구를 진행함
  - 발전 가능성이 높은 연구 분야 [내 생각]
    - Egocentric pose estimation
      - Apple의 vr기기가 새로 생김과 동시에 해당 분야 발전 가능성이 높고 VR기기에 들어갈 카메라 모듈이 대부분 Egocentric이기 때문에 충분히 고려해볼 만한 분야임
    - 극도로 어려운 환경에서의 pose estimation 개선
      - 이번에 처음 문제제기가 됨과 동시에 Dataset이 만들어짐 VR등 다양한 환경에서 고려해야하기 때문에 연구하기 좋은 분야
        - ✓하지만 기존의 모델 성능을 개선 시키면서 진행해야 하기 때문에 어려운 task가 될 수 있음
  - Complex pose dataset!
    - 최근 들어서 요가와 같은 어려운 포즈 데이터 셋이 생기기 시작했음 이쪽을 target으로 해도 좋음!
      - 3D Human Pose Estimation via Intuitive Physics [2023 CVPR accept]