한국컴퓨터그래픽스학회
Korea Computer Graphics Society Vol. 32, No. 2, P. 33~45

Pose-Guided Reconstruction of Partially Visible Pedestrians under Vehicle Occlusion

Hyunjun Jang, Jae-Chern Yoo*

Dept. of Electrical and Computer Engineering, Sungkyunkwan University

10hour@skku.edu, *yoojc@skku.edu

차량 가림 환경에서 부분적으로 가시적인 보행자의 포즈 유도 재구성

장현준, 유재천*

성균관대학교 전자전기컴퓨터공학과

Abstract

This paper presents a pose-guided reconstruction framework that recovers full-body pose and silhouette from minimal lower-limb cues visible through the gap beneath an occluding vehicle. The implemented and evaluated pipeline consists of Occluded Lower-Limb Pose Detection (OLPD), Full-Body Pose Inference (FBPI), and Pose-Guided Silhouette Reconstruction (PGSR), which detect visible lower-limb keypoints, infer full-body keypoints, and generate a silhouette probability mask. We also introduce Minimum Required Visibility (MRV) to quantify the minimum visible body ratio required to satisfy a predefined reconstruction criterion. Under the adopted IoU $\geq$ 0.80 and Dice/F1 $\geq$ 0.90 criterion, the lowest visible ratio among successful test samples was 12.87%, and the measured latency of the FBPI + PGSR reconstruction path was 8.676 ms per sample, excluding OpenPose BODY_25 inference, vehicle-mask prediction, keypoint filtering, and HSV. OLPD results should be interpreted as a mask-given synthetic upper-bound because the current implementation assumes access to the synthetic vehicle-layer mask produced by the compositing pipeline. The reported 8.676 ms should be interpreted as a local, hardware-unspecified measurement of the FBPI+PGSR reconstruction path only. It excludes OpenPose BODY_25 inference, vehicle-mask prediction, keypoint filtering, and HSV, and therefore does not establish end-to-end real-time performance. Human Silhouette Verification (HSV) is not implemented or evaluated in the current paper and is discussed only as a prospective future extension.

요 약

본 논문은 차량에 의해 부분적으로 가려진 보행자를 대상으로, 차량 하부 틈새를 통해 관측되는 최소한의 하지 단서만으로 전신 자세와 실루엣을 복원하는 pose-guided reconstruction framework를 제안한다. 구현 및 평가된 파이프라인은 OLPD, FBPI, PGSR의 3단계로 구성되며, 가시적인 하지 keypoint를 검출한 뒤 전신 keypoint를 추론하고 최종적으로 silhouette probability mask를 생성한다. 또한 사전 정의된 복원 기준을 만족하는 데 필요한 최소 가시 비율을 정량화하기 위해 MRV를 도입한다. 통제된 synthetic paired dataset 실험에서, 채택한 IoU $\geq$ 0.80 및 Dice/F1 $\geq$ 0.90 기준하에 성공한 테스트 샘플들 중 가장 낮은 가시 비율은 12.87%였으며, OpenPose BODY_25 추론, vehicle-mask prediction, keypoint filtering, HSV를 제외한 측정된 FBPI+PGSR 복원 경로 기준 샘플당 8.676 ms의 처리 시간을 보였다. 현재 구현에서 OLPD는 compositing pipeline이 생성한 synthetic vehicle-layer mask에 접근한다고 가정하므로, OLPD 결과는 mask-given synthetic upper-bound로 해석되어야 한다. 보고된 8.676 ms는 하드웨어 정보가 별도로 기록되지 않은 로컬 환경에서 측정된 FBPI+PGSR 복원 경로 전용 수치로 해석되어야 하며, 전체 end-to-end 실시간 성능을 입증하지 않는다. HSV는 현재 구현 또는 평가에 포함되지 않았으며, 향후 확장 모듈로 별도 논의한다.

Keywords: Occluded Pedestrian Reconstruction, Human Pose Estimation, Full-Body Pose Inference, Pose-Guided Silhouette Reconstruction, Traffic Safety Perception.

키워드: 차량 가림 보행자 재구성, 인체 자세 추정, 전신 자세 추론, 포즈 유도 실루엣 재구성, 교통 안전 인식

*corresponding author: Jae-Chern Yoo / Dept. of Electrical and Computer Engineering, Sungkyunkwan University (yoojc@skku.edu)

Received : 2026.04.01./ Review completed : 1st 2026.04.16. / Accepted : 2026.04.30.
DOI : 10.15701/kcgs.2026.32.2.33
ISSN : 1975-7883(Print)/2383-529X(Online)

1. Introduction

Roadside pedestrian accidents often occur when a pedestrian suddenly emerges from behind a parked vehicle before sufficient visual evidence becomes available to onboard perception modules. This risk is particularly high in dense roadside-parking environments, where only a narrow under-vehicle gap may reveal limited lower-limb cues.

Existing pose estimators and occlusion-recovery methods do not directly address full-body silhouette reconstruction from extremely sparse under-vehicle observations [1, 2]. In parallel, ghost-probe approaches primarily focus on hazard-zone prediction rather than explicit shape-level recovery from partial visual evidence. This leaves a practical gap between visible-cue detection and reconstruction-oriented reasoning for severely occluded pedestrians.

We propose a pose-guided reconstruction framework whose implemented and evaluated pipeline comprises OLPD, FBPI, and PGSR for recovering full-body structure and silhouette from sparse lower-limb cues beneath parked vehicles. A separate HSV stage is treated only as a prospective future extension and is not implemented or evaluated in this paper.

The main contributions are as follows. First, we formulate partial under-vehicle pedestrian visibility as a pose-guided reconstruction problem. Second, we introduce an implemented three-stage cascade for recovering full-body pose and silhouette from sparse lower-limb cues. Third, we introduce MRV to quantify occlusion tolerance. Fourth, we construct a synthetic paired dataset for controlled evaluation.

2. Related Work

2.1 Occlusion-Aware Human Pose and Mesh Recovery

OpenPose (Cao et al., 2017) and CrowdPose (Li et al., 2019) provide the visible-keypoint estimation foundation of our framework [1, 2]. More recent work has pushed pose and mesh reasoning under occlusion by explicitly modeling part visibility, spatial context, diffusion priors, or richer whole-body structure. For example, PARE uses body-part-guided attention to improve 3D human body estimation under partial occlusion [3]; visibility-aware transformer reasoning improves 2D pose estimation under occlusion by suppressing unreliable occluder features [4]; Occluded Human Mesh Recovery (OCHMR) augments top-down mesh recovery with spatial context for person-person occlusion [5]; AiOS moves toward an all-in-one-stage expressive pose-and-shape pipeline without an external detector [6]; DPMesh exploits diffusion priors for occluded human mesh recovery [7]; and ScoreHMR uses score-guided diffusion for 3D human recovery across several inverse-problem settings [8]. These studies demonstrate strong recovery capability when broader human extent and richer contextual evidence are available. However, they are not direct one-to-one solutions for the present setting, because our target problem is under-vehicle occlusion with only minimal lower-limb cues and a lightweight staged reconstruction output of full-body keypoints and silhouette probability masks rather than full 3D meshes.

In the KCGS literature, K-SMPL, silhouette-driven human-shape estimation, sensor-based motion reconstruction, and low-cost interactive motion-capture systems provide useful representation- and motion-level context [9, 10, 11, 12], but they are not designed for under-vehicle occlusion or minimal lower-limb cue recovery.

2.2 Human De-Occlusion and Amodal Completion

Human de-occlusion and amodal completion address the inference of invisible structure behind occluders. Earlier human-specific work includes object-occluded human shape and pose recovery [13], human de-occlusion with invisible-mask and appearance recovery [14], and 2D pose-guided complete silhouette estimation for occluded human bodies [15]. Subsequent work has expanded both realism and scope: OccNeRF studies human rendering from object-occluded monocular videos using geometry and visibility priors [16]; Amodal Completion via Progressive Mixed Context Diffusion uses context-guided diffusion for completion [17]; pix2gestalt synthesizes wholes for zero-shot amodal segmentation [18]; and Amodal Ground Truth and Completion in the Wild broadens amodal benchmarking and completion in real images [19]. Complementary visible-mask extraction backbones such as Mask2Former and SAM strengthen general-purpose segmentation quality [20, 21]. These methods are highly relevant because they infer invisible masks, shapes, or appearance from partial observations or strengthen the visible mask extraction on which downstream completion may depend. Nevertheless, most of them target generic amodal or appearance completion, broader segmentation, or settings with richer visible context than ours, and are not explicitly formulated as lightweight staged reconstruction from minimal lower-limb cues under under-vehicle occlusion.

pair were always assigned to the same split; they were never separated across training, validation, and test partitions. Because the current synthetic construction used child-pedestrian foreground assets, the reported results should be interpreted as a controlled feasibility study under a child-pedestrian setting rather than as a universal benchmark for all pedestrian appearances.

4.2 Quantitative Results and Visual Comparison

Because HSV was not implemented, the quantitative results reported in this section correspond only to the implemented OLPD, FBPI, and PGSR components. Table 2 separates the measured latency scope. Because the original experiment log did not preserve the CPU/GPU model, inference runtime framework, batch size, or warm-up/repetition counts, a fully specified timing protocol cannot be reported retrospectively. The reported 8.676 ms should therefore be interpreted as a local, hardware-unspecified measurement of the FBPI + PGSR reconstruction path only. It excludes OpenPose BODY_25 inference, vehicle-mask prediction, keypoint filtering, and HSV. The 0.077 ms OLPD-related timing reported here does not measure OpenPose inference; it measures only keypoint filtering time after OpenPose outputs are available, using the known synthetic vehicle mask. When this post-OpenPose keypoint-filtering step is added to FBPI + PGSR, the combined measured latency becomes 8.753 ms, but this value still excludes OpenPose BODY_25 inference, vehicle-mask prediction, and HSV. No manual keypoint annotation was used for FBPI evaluation in this section. Accordingly, FBPI metrics should be interpreted as consistency with OpenPose-derived reference poses rather than as actual human pose accuracy. Moreover, because the dataset is constructed from only 10 background assets, 10 vehicle assets, and 10 child pedestrian assets, the reported numbers should be interpreted as a controlled feasibility study under synthetic conditions rather than as evidence of broad real-world recovery. In this section, oracle-mask OLPD results should be interpreted strictly as mask-given synthetic upper-bound results obtained with the synthetic vehicle-layer mask from the compositing pipeline; they do not measure automatic vehicle segmentation or deployment-time perception.

4.2.1 Oracle-Mask vs Predicted-Mask Comparison

Before comparing against external generic completion baselines, we first separate the oracle-mask setting from the predicted-mask setting. Fig. 9 compares these two regimes. In the oracle-mask setting, the synthetic vehicle-layer mask is provided directly by the compositing pipeline, so the resulting OLPD behavior should be interpreted only as a mask-given synthetic upper-bound rather than as automatic vehicle segmentation or deployment-time perception. By contrast, the predicted-mask panel in Fig. 9 is included only as an illustrative external segmentation example in which an externally estimated vehicle mask is substituted for the oracle mask and then passed to the same downstream reconstruction path. The exact external mask-prediction procedure used to prepare that illustrative panel is not documented in sufficient detail in the current manuscript assets and is not part of the implemented benchmarked pipeline reported in this paper. Accordingly, the predicted-mask setting in Fig. 9 should not be read as a reproducible model specification or formal method comparison. Moreover, the current manuscript does not provide quantitative evaluation for predicted vehicle-mask IoU, OLPD lower-limb F1 under the predicted mask, FBPI OKS under the predicted mask, PGSR IoU/Dice under the predicted mask, or separated mask-prediction latency. Therefore, the predicted-mask results in Fig. 9 should be interpreted as qualitative evidence only rather than as quantitative support for practical deployment.

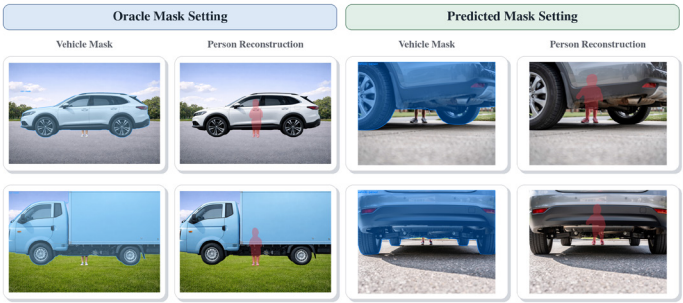

Figure 9: Qualitative comparison between the oracle-mask and predicted-mask settings. Left: the synthetic vehicle mask is provided directly by the compositing process and used as an oracle occlusion mask; this oracle-mask condition should be interpreted as a mask-given synthetic upper-bound. Right: an externally estimated vehicle mask is used in place of the oracle mask and the same downstream reconstruction path is applied. The predicted-mask panel is presented only as an illustrative external segmentation example; the exact mask-prediction procedure is not specified as part of the implemented benchmarked pipeline in the current manuscript. Because no predicted-mask quantitative evaluation is reported, this panel should be interpreted as qualitative evidence only.

4.2.2 Stage-Wise Sensitivity to Mask Errors

To analyze how vehicle-mask errors affect the OLPD–FBPI–PGSR pipeline, we applied four perturbation types: undersegmentation, oversegmentation, boundary jitter, and partial missing. Fig. 10 reports vehicle-mask IoU, OLPD lower-limb F1, FBPI OKS with respect to OpenPose-derived reference poses on occluded samples, and PGSR mask IoU across mild-to-extreme severity levels.

pletion can produce strong upward lift, but the proposed staged pipeline remains more lightweight and structurally controlled for the under-vehicle setting. These observations should be interpreted as illustrative qualitative tendencies rather than as a quantitative benchmark.

4.2.4 Pipeline-Level Summary

For the implemented system, we therefore summarize the results at the pipeline level through visibility tolerance and component-wise latency scope rather than through stage-wise diagnostic tables. Any OLPD-related interpretation in this summary remains limited to the mask-given synthetic upper-bound under the oracle synthetic vehicle-mask setting; it does not evaluate automatic vehicle segmentation or deployment-time perception.

To make the MRV metric operational, let A_{vis} denote the number of visible pedestrian pixels in the occluded sample and A_{full} denote the number of pedestrian pixels in the paired non-occluded ground-truth mask. The visible ratio is defined as $r_{\mathrm{vis}} = A_{\mathrm{vis}}/A_{\mathrm{full}}$. We define Minimum Required Visibility (MRV) as the minimum visible ratio among test samples whose reconstruction satisfies both $\mathrm{IoU} \geq 0.80$ and $\mathrm{Dice\ (F1)} \geq 0.90$. Under this criterion, the observed MRV on the test set was 12.87%.

Because the original experiment log did not preserve the CPU/GPU model, inference runtime framework, batch size, or warm-up/repetition counts, Table 2 should be read as a local, hardware-unspecified latency summary of the measured reconstruction subpaths only. These values do not establish end-to-end real-time performance.

Table 2: Latency scope of the implemented reconstruction path.

Path	Included operation	Excluded operations	Mean latency
Keypoint filtering	post-OpenPose vehicle-mask-based joint filtering	OpenPose, vehicle-mask prediction, FBPI, PGSR, HSV	0.077 ms
FBPI forward pass	PartialToFullKeypointNet OpenPose, vehicle-mask prediction, keypoint filtering, PGSR, HSV		1.464 ms
PGSR forward pass plus resize/post-process	KeypointToMaskNet OpenPose, vehicle-mask prediction, keypoint filtering, FBPI, HSV		7.212 ms
FBPI+PGSR	FBPI inference + PGSR inference	OpenPose, vehicle-mask prediction, keypoint filtering, HSV	8.676 ms
Post-OpenPose path FBPI + PGSR	keypoint filtering + OpenPose, vehicle-mask prediction, HSV		8.753 ms

At the same time, the interpretation range of these results is narrow. The 10-background, 10-vehicle, and 10-child-pedestrian asset composition limits scene diversity, vehicle geometry variation, and especially the pose, scale, and silhouette distributions observed during training and evaluation. Because only child assets are used, the reported performance does not establish equivalent behavior for adult pedestrians or for broader variation in body shape, clothing, and self-occlusion. In addition, multi-person scenes and fully invisible cases are outside the evaluated distribution. Accordingly, the present numbers should be read as feasibility under controlled synthetic conditions rather than as evidence of robust real-world recovery.

5. Conclusions

5.1 Key Findings and Implications

This paper presented an implemented three-stage pose-guided reconstruction framework for partially visible pedestrians under vehicle occlusion. The current work focuses on recovering full-body structure and silhouette from sparse lower-limb evidence observable through the under-vehicle gap, rather than implementing a complete warning decision system. On the adopted synthetic dataset, the measured FBPI + PGSR reconstruction path required 8.676 ms per sample, excluding OpenPose BODY_25 inference, vehicle-mask prediction, keypoint filtering, and HSV, and adding only the post-OpenPose keypoint-filtering step yielded 8.753 ms. These values should be interpreted as local, hardware-unspecified measurements of the partial reconstruction path only, and they do not establish end-to-end real-time performance. Successful reconstruction was maintained down to a minimum visible ratio of 12.87% under the adopted criterion. The FBPI stage should be interpreted as full-body keypoint inference consistent with paired OpenPose reference poses rather than as measured human pose accuracy against manual keypoint annotation. In addition, PGSR in the current implementation is trained on paired OpenPose reference full-body poses from non-occluded images but is evaluated downstream with residual-fused FBPI outputs on occluded samples. Therefore, the reported PGSR behavior should be interpreted as a pose-conditioned upper-bound with respect to pose-input quality rather than as performance of a model trained on the exact deployment-time pose input. The additional oracle-versus-predicted mask comparison provides qualitative evidence only through an illustrative external segmentation example; it is not a reproducible benchmarked mask-prediction method specification in the current manuscript. At the same time, any OLPD-related interpretation in the current manuscript remains limited to the mask-given synthetic upper-bound under the oracle-mask setting and does not measure automatic vehicle segmentation or deployment-time perception. Taken together, these findings support feasibility under controlled synthetic conditions rather than robust real-world recovery, while HSV and downstream alert logic remain future work.

5.2 Limitations and Future Work

A notable limitation of this study is that the proposed Human Silhouette Verification (HSV) stage was not implemented or quantitatively evaluated in the current system. Although HSV was conceptually designed as a final quality-control module for rejecting implausible reconstructed masks, the present work reports no HSV-specific quantitative results and evaluates only the first three stages, namely lower-limb pose extraction, full-body pose inference, and pose-guided silhouette reconstruction.

A further limitation is that OLPD currently relies on the synthetic vehicle-layer mask available from the compositing pipeline. Accordingly, all reported OLPD results should be interpreted as mask-given synthetic upper-bound results for lower-limb keypoint filtering under a known occlusion mask. They do not evaluate automatic vehicle segmentation or deployment-time perception.

In addition, several aspects of the current experimental design directly limit how the reported results should be interpreted. The validation is restricted to a fully synthetic paired dataset constructed from 10 background images, 10 vehicle images, and 10 Pedestrian (Kid) assets. This small asset pool narrows scene diversity and constrains the pose, scale, and silhouette distributions seen during training and testing. Because the pedestrian assets are child-only, the reported numbers do not establish equivalent performance for adult pedestrians or for broader variation in body shape, clothing, and self-occlusion. Each sample also contains a single pedestrian instance, fully invisible cases were excluded because reliable lower-body keypoints could not be extracted by OpenPose, and the full-body keypoint supervision is derived from the paired OpenPose reference pose on the non-occluded image rather than from manual keypoint annotation. No manually annotated keypoint subset was used to validate actual human pose accuracy; therefore, FBPI metrics should be interpreted as consistency with OpenPose-derived reference poses. The same paired non-occluded reference pose is also used as the PGSR training input, whereas inference uses the residual-fused FBPI output on the occluded sample. Consequently, the present manuscript includes a pose-input train–test distribution gap at Stage 3, and PGSR results should be interpreted as pose-conditioned upper-bound evidence rather than as a fully matched train/test evaluation under predicted pose input. In addition, the surviving manuscript assets do not preserve the exact internal projection from the 25×3 PGSR pose tensor to the first spatial encoder feature map, so Fig. 6 should be interpreted as a conceptual schematic rather than a fully reproducible layer-by-layer architecture. Moreover, Fig. 9 uses an illustrative externally estimated mask example whose exact mask-prediction procedure is not specified as part of the benchmarked pipeline, so that panel should be interpreted as qualitative evidence only. Accordingly, the present quantitative results should be interpreted as a controlled feasibility study under restricted synthetic conditions rather than as evidence of broad road-scene generalization.

Future work should therefore include real-image validation, stricter cross-asset evaluation, adult pedestrians, broader body-shape and clothing variation, multi-person scenes, fully invisible cases, quantitative benchmarking with predicted masks including predicted vehicle-mask IoU, OLPD lower-limb F1 under predicted masks, FBPI OKS under predicted masks, PGSR IoU/Dice under predicted masks, separated mask-prediction latency, and implementation of the HSV stage.

References

[1] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7291–7299, 2017.

[2] J. Li, C. Wang, H. Zhu, Y. Mao, H.-S. Fang, and C. Lu, "CrowdPose: Efficient Crowded Scenes Pose Estimation and a New Benchmark," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10863–10872, 2019.

[3] M. Kocabas, C.-H. P. Huang, O. Hilliges, and M. J. Black, "PARE: Part Attention Regressor for 3D Human Body Estimation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 11127–11137, 2021.

[4] P. Sun, K. Gu, Y. Wang, L. Yang, and A. Yao, "Rethinking Visibility in Human Pose Estimation: Occluded Pose Reasoning via Transformers," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 5903–5912, 2024.

[5] R. Khirodkar, S. Tripathi, and K. Kitani, "Occluded Human Mesh Recovery," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1715–1725, 2022.

[6] Q. Sun, Y. Wang, A. Zeng, W. Yin, C. Wei, W. Wang, H. Mei, C.-S. Leung, Z. Liu, L. Yang, and Z. Cai, "AiOS: All-in-One-Stage Expressive Human Pose and Shape Estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1834–1843, 2024.

[7] Y. Zhu, A. Li, Y. Tang, W. Zhao, J. Zhou, and J. Lu, "DPMesh: Exploiting Diffusion Prior for Occluded Human

Mesh Recovery," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1101–1110, 2024.

[8] A. Stathopoulos, L. Han, and D. Metaxas, "Score-Guided Diffusion for 3D Human Recovery," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 906–915, 2024.

[9] B. Choi and S.-H. Lee, "K-SMPL: Korean Body Measurement Data Based Parametric Human Model," *Journal of the Korea Computer Graphics Society*, 28(4):1–11, 2022.

[10] D. Ahn and S. I. Park, "3D Human Shape Estimation from a Silhouette Image by using Statistical Human Shape Spaces," *Journal of the Korea Computer Graphics Society*, 29(1):13–22, 2023.

[11] H. Kim, K. Kang, G. Park, and T. Kwon, "Deep Learning-Based Motion Reconstruction Using Tracker Sensors," *Journal of the Korea Computer Graphics Society*, 29(5):11–20, 2023.

[12] J. Kim, D. Kang, Y. Lee, and T. Kwon, "Real-time Interactive Animation System for Low-Priced Motion Capture Sensors," *Journal of the Korea Computer Graphics Society*, 28(2):29–41, 2022.

[13] T. Zhang, B. Huang, and Y. Wang, "Object-Occluded Human Shape and Pose Estimation From a Single Color Image," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7376–7385, 2020.

[14] Q. Zhou, S. Wang, Y. Wang, Z. Huang, and X. Wang, "Human De-Occlusion: Invisible Perception and Recovery for Humans," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3691–3701, 2021.

[15] S. Xu, X. Li, and X. Lin, "2D Pose-guided Complete Silhouette Estimation of Human Body in Occlusion," in *26th International Conference on Pattern Recognition (ICPR)*, pp. 3618–3624, 2022.

[16] T. Xiang, A. Sun, J. Wu, E. Adeli, and L. Fei-Fei, "Rendering Humans from Object-Occluded Monocular Videos," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3239–3250, 2023.

[17] K. Xu, L. Zhang, and J. Shi, "Amodal Completion via Progressive Mixed Context Diffusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9099–9109, 2024.

[18] E. Ozguroglu, R. Liu, D. Suris, D. Chen, A. Dave, P. Tokmakov, and C. Vondrick, "pix2gestalt: Amodal Segmentation by Synthesizing Wholes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3931–3940, 2024.

[19] G. Zhan, C. Zheng, W. Xie, and A. Zisserman, "Amodal Ground Truth and Completion in the Wild," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28003–28013, 2024.

[20] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, "Masked-Attention Mask Transformer for Universal Image Segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1290–1299, 2022.

[21] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment Anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4015–4026, 2023.

[22] H. Fu, Y. Gu, Y. Yan, Y. Shen, Y. Wu, and L. Sun, "SDR-GAIN: A High Real-Time Occluded Pedestrian Pose Completion Method for Autonomous Driving," *arXiv preprint arXiv:2306.03538*, 2023.

[23] P. Hardy and H. Kim, "LInKs "Lifting Independent Keypoints" - Partial Pose Lifting for Occlusion Handling With Improved Accuracy in 2D-3D Human Pose Estimation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 3426–3435, 2024.

[24] M. Armando, S. Galaaoui, F. Baradel, T. Lucas, V. Leroy, R. Bregier, P. Weinzaepfel, and G. Rogez, "Cross-view and Cross-pose Completion for 3D Human Understanding," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1512–1523, 2024.

[25] H. Zheng, H. Li, W. Dai, Z. Zheng, C. Li, J. Zou, and H. Xiong, "HiPART: Hierarchical Pose AutoRegressive Transformer for Occluded 3D Human Pose Estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16807–16817, 2025.

[26] W. Qu, J. Du, S. Yuan, J. Wang, Y. Sun, S. Liu, Y. Zhu, J. Rao, J. Yu, S. Cao, R. Xia, X. Tang, X. Wu, and D. Luo,

"DPGP: A Hybrid 2D-3D Dual Path Potential Ghost Probe Zone Prediction Framework for Safe Autonomous Driving," in *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 20341–20348, 2025.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," in *Advances in Neural Information Processing Systems 30*, pp. 5998–6008, 2017.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, 2015.

〈저자소개〉

유재천
- Current Position: Professor
- Affiliation: Department of Electrical and Computer Engineering, Sungkyunkwan University
- Education
- Ph.D. in Electrical and Electronic Engineering, POSTECH, 2001
- Career / Experience
- Senior Research Engineer, Samsung Electronics Co., Ltd., 1986 - 1996
- Professor, Samsung Institute of Technology, 1994
- Adjunct Professor, POSTECH Information Research Laboratories and Graduate School for Information Technology, 2002 - 2006
- Research Assistant Professor, POSTECH Graduate School for Information Technology, 2006 - 2008
- 관심분야: Design of deep learning architectures and their applications, AI-based medical image diagnosis, Design of AI-controlled robotic hands, AI-based digital healthcare platforms
- https://orcid.org/0000-0003-3721-0207

장현준
- Current Position: Master's Student
- Affiliation: Department of Electrical and Computer Engineering, Sungkyunkwan University
- Education
- B.S. in Electronic Engineering, Hankuk University of Foreign Studies, 2022
- M.S. Student in Department of Electrical and Computer Engineering, Sungkyunkwan University, 2025 - Present
- 관심분야: Image processing, Vision