DepthHMR: Leveraging Depth Around Humans for Multi-Human Mesh Generation

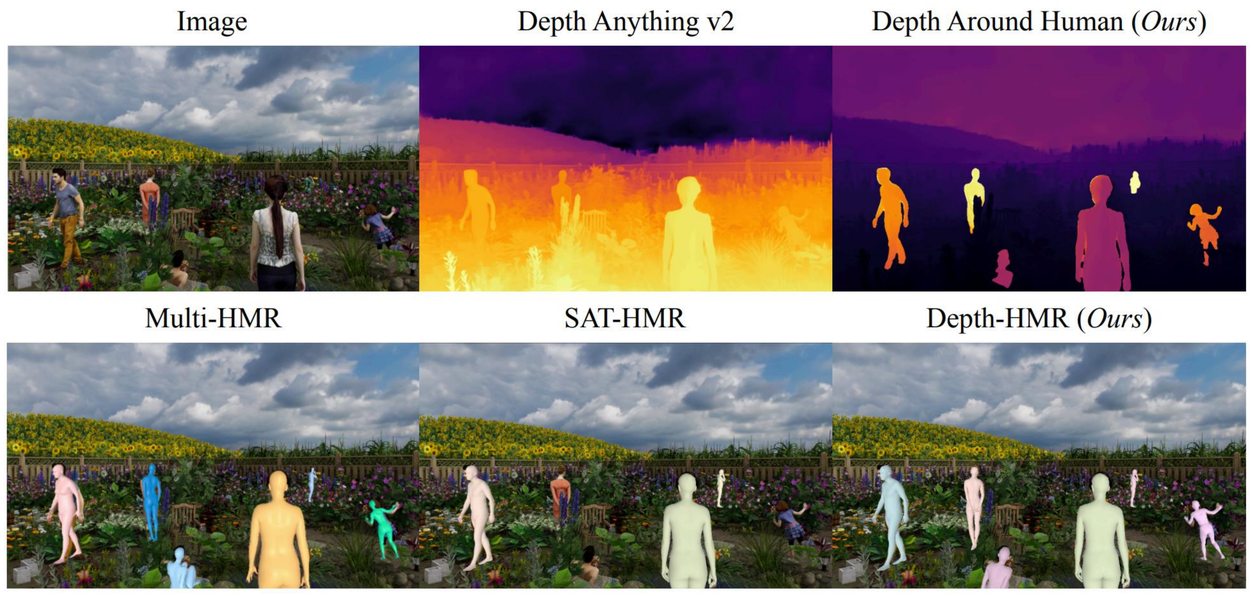
Nikhil Sharma, Jiachen Tao, Junyi Wu, Yan Yan

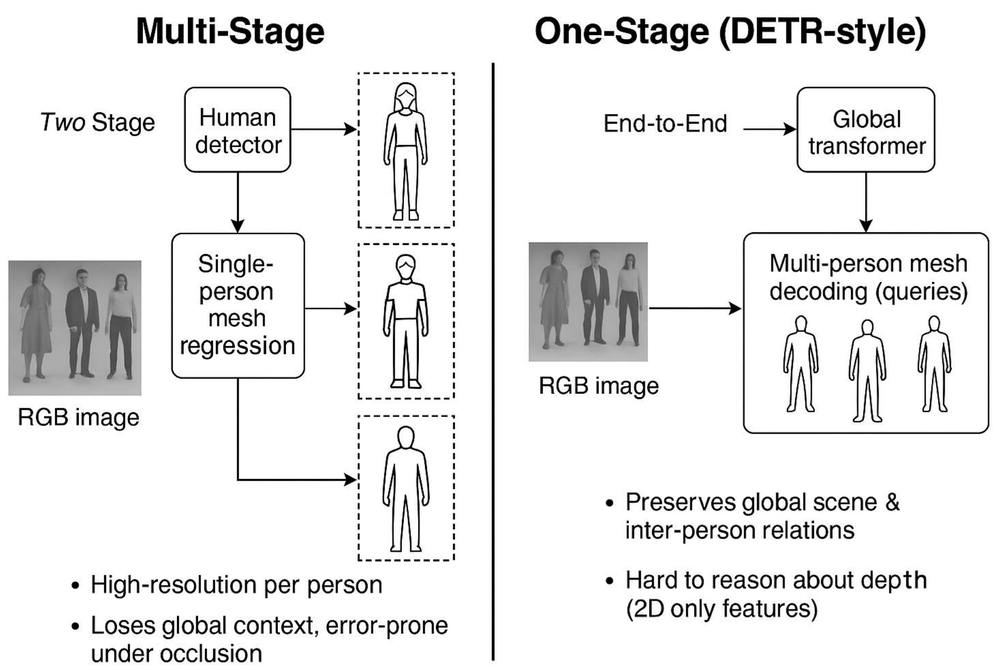
University of Illinois Chicago

BMVC 2025

Motivation

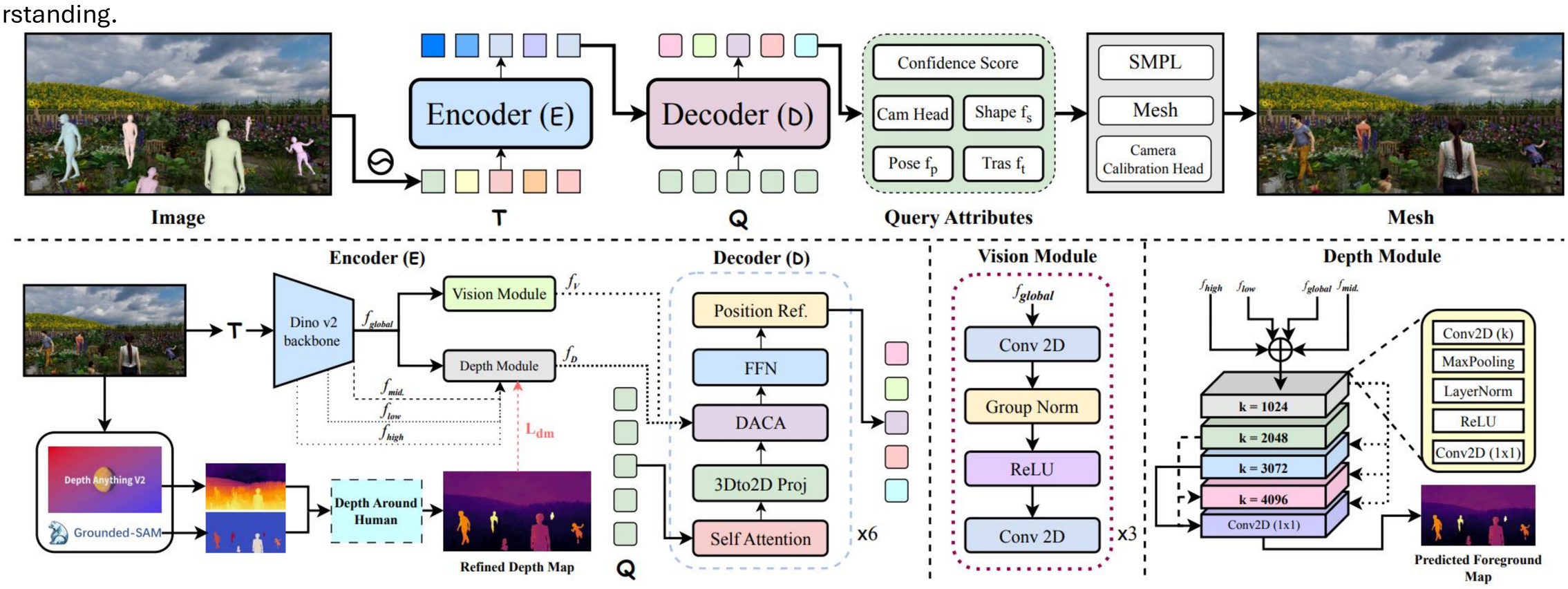
Recovering multi-human 3D meshes from a single RGB image still suffers from severe depth ambiguity, especially for occluded and distant individuals. Both Multi-Stage and One-Stage methods lack explicit 3D reasoning, making it difficult to handle occlusion and scale variations





Method

We integrate explicit depth reasoning into a DETR-style one-stage framework, leveraging the Depth Around Human (DAH) module to enhance 3D understanding.



Experimental Results

Table 1: Comparison with SOTA methods on AGORA test set. "Res" represents the input resolution of the image. All metrics except F1-Score are reported in "mm". The **best** and *second best* results are highlighted.

Method	Res	F1-Score ↑	MPJPE ↓	MVE ↓	NMJE ↓	NMVE ↓
ROMP [35]	512	0.91	108.1	103.4	118.8	113.6
BEV [36]	512	0.93	105.3	100.7	113.2	108.3
PSVT [28]	512	0.93	97.7	94.1	105.1	101.2
AiOS [34]	1333	0.94	*63.9*	*57.5*	*68.0*	**61.2**
Multi-HMR [1]	896	0.93	82.8	77.6	89.0	83.4
Multi-HMR [1]	1288	*0.95*	65.3	61.1	68.7	64.3
SAT-HMR [33]	644	*0.95*	67.9	63.3	71.5	66.6
DepthHMR *(Ours)*	644	**0.96**	**63.4**	**54.7**	**66.0**	*61.7*

Table 2: Comparison with SOTA methods on real-world 3DPW [37] and MuPoTS [21] test sets. "Res" represents the input resolution. "Ind" refers to cropping individuals for single-person mesh estimation.

Method	Res	3DPW		MuPoTS (PCK)	
		PA-MPJPE ↓	MVE ↓	All ↑	Matched ↑
3DCrowdNet [5]	Ind	51.5	98.3	72.7	73.3
ROMP [35]	512	47.3	93.4	69.9	72.2
BEV [36]	512	46.9	92.3	70.2	75.2
PSVT [28]	512	45.7	84.9	-	-
Multi-HMR [1]	896	41.7	75.9	85.0	89.3
SAT-HMR [33]	644	41.6	73.7	89.0	90.1
DepthHMR *(Ours)*	644	**41.6**	**71.3**	**84.3**	**89.7**

Figure 3: Qualitative comparison with SOTA methods on in-the-wild images from internet.

Figure 1: Qualitative comparison with SOTA methods on the samples from AGORA test set