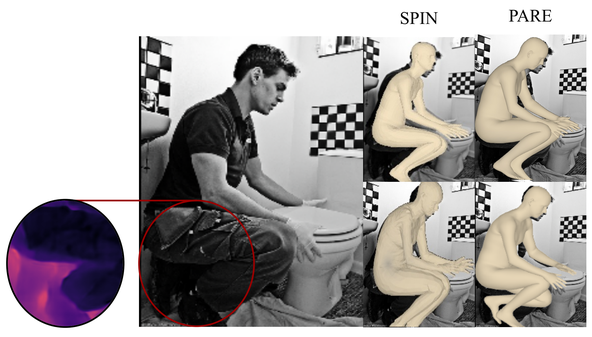
Distribution and Depth-Aware Transformers for 3D Human Mesh Recovery

Jerrin Bright, Bavesh Balaji, Harish Prakash, Yuhao Chen, David Clausi, John Zelek
University of Waterloo, Waterloo, Ontario, Canada

AI · CRV conference

Key Contributions

- A novel image-based HMR model named **D2A-HMR** that adeptly models the underlying distributions and integrates pseudo-depth priors for efficient and accurate mesh recovery.
- By leveraging a **residual log-likehood** approach, we refine the model by learning the disparity between the underlying predicted and ground truth distribution.
- Validation of the enhanced performance through the integration of pseudo-depth and distribution-aware modules in HMR, shows robustness in complex pose scenarios.



Loss Functions

Distribution regularizer [1]:

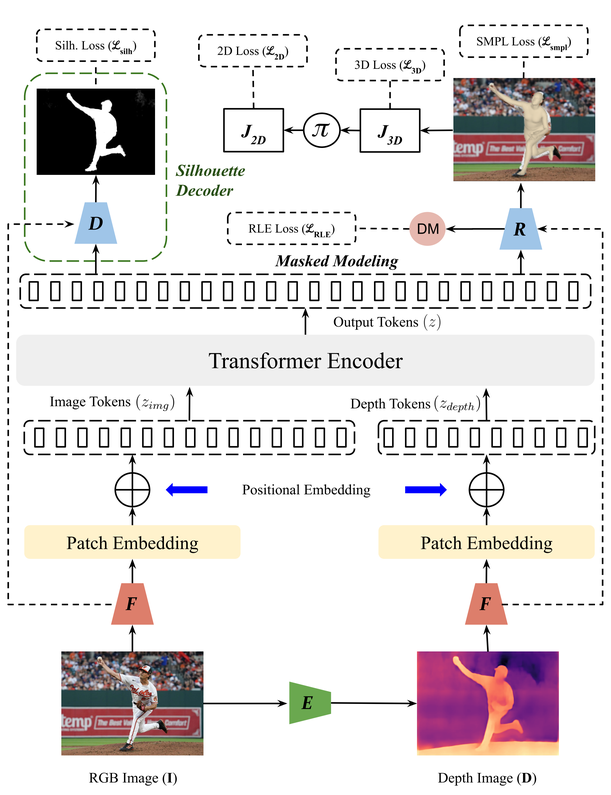
$$\mathcal{L}_{RLE} = -\log Q(\bar{\mu}_g) - \log G_\phi(\bar{\mu}_g) - \log c + log\ \sigma \quad (1)$$

Overall objective function:

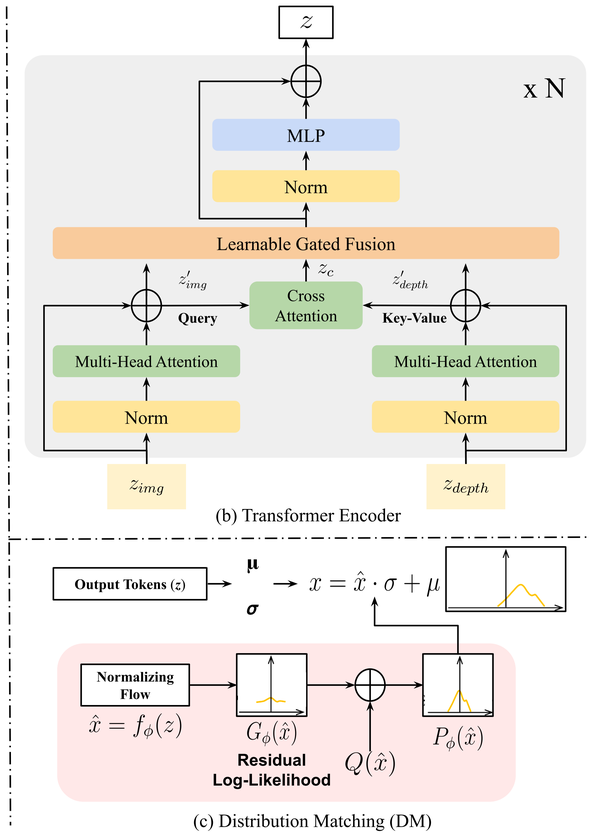
$$\mathcal{L} = \lambda_d \mathcal{L}_{RLE} + \lambda_v \mathcal{L}_v + \lambda_{3D}\mathcal{L}_{3D} + \lambda_{2D}\mathcal{L}_{2D} + \lambda_s \mathcal{L}_{silh} \quad (2)$$

Methodology

An end-to-end transformer architecture meticulously designed to minimize the disparity between distributions and incorporate scene-depth leveraging prior depth information.



(a) Proposed D2A-HMR framework



(b) Transformer Encoder

(c) Distribution Matching (DM)

Quantitative Results

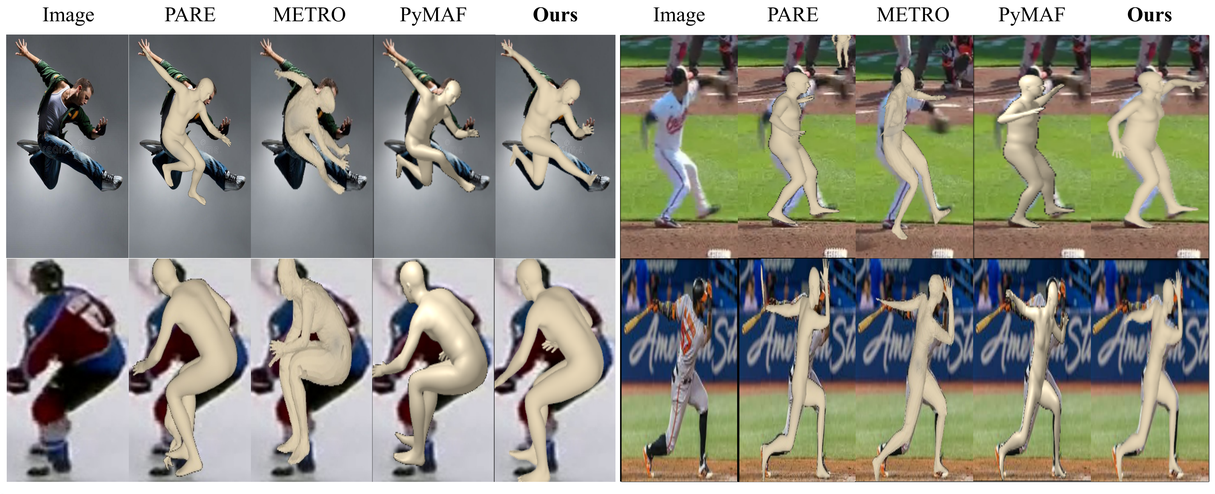
- Comparison on 3DPW and Human3.6M datasets.

Method	Human3.6M		3DPW	
	mPJPE	PA-mPJPE	mPJPE	PA-mPJPE
HMMR	-	58.1	116.5	72.6
TCMR	62.3	41.1	95.0	55.8
VIBE	65.6	41.4	93.5	56.5
HMR	88.0	56.8	130.0	81.3
SPIN	62.5	41.1	96.9	59.2
PyMAF	57.7	40.5	92.8	58.9
ROMP	-	-	105.6	53.5
HMR-EFT	63.2	43.8	85.1	52.2
PARE	76.8	50.6	82.0	50.9
ProHMR	-	41.2	95.1	59.5
Pose2Mesh	64.9	47.0	89.2	58.9
METRO	54.0	36.7	77.1	47.9
Ours	53.8	36.2	80.5	48.4

- Comparison on the MLBPitchDB dataset [2].

Method	Acc. ↑	mPJPE ↓
HMR	65.9	61.3
SPIN	84.7	32.1
ProHMR	76.1	48.2
ROMP	77.4	48.9
METRO	81.5	37.8
PARE	84.0	33.7
Ours	87.9	30.6

Qualitative Results



Ablation Study

- Pseudo-depth and distribution modeling evaluated on 3DPW dataset.

Depth	Distribution	mPJPE ↓	PA-mPJPE ↓
✓		92.7	61.8
	✓	90.0	56.9
✓	✓	**80.5**	**48.4**

- Silhouette decoder and masked modeling evaluated on 3DPW dataset.

Silhouette	Masked Modeling	mPJPE ↓	PA-mPJPE ↓
✓		89.5	62.2
	✓	84.7	51.4
✓	✓	**80.5**	**48.4**

Acknowledgement

D2A-HMR Preprint:

References

[1] Jiefeng Li, Siyuan Bian, Ailing Zeng, Can Wang, Bo Pang, Wentao Liu, and Cewu Lu. Human pose regression with residual log-likelihood estimation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 11025–11034, 2021.

[2] Jerrin Bright, Yuhao Chen, and John Zelek. Mitigating motion blur for robust 3d baseball player pose modeling for pitch analysis. In *Proceedings of the 6th International Workshop on Multimedia Content Analysis in Sports*, pages 63–71, 2023.