3D-LFM: Lifting Foundation Model

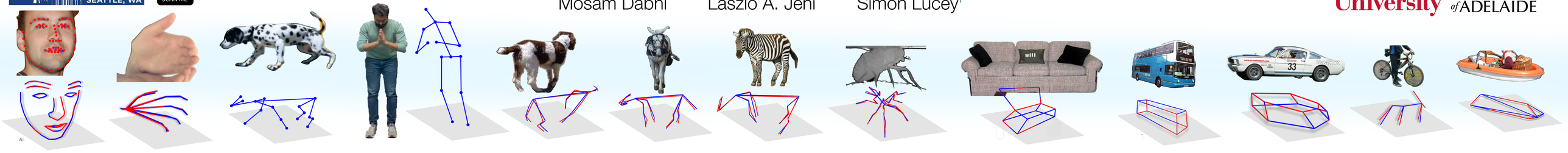
Mosam Dabhi* László A. Jeni* Simon Lucey†

Carnegie Mellon University

THE UNIVERSITY of ADELAIDE

CVPR JUNE 17-21, 2024 SEATTLE, WA

SCAN ME



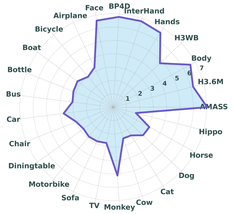
Universal model for 3D Lifting

Single model for diverse categories: Handles body parts, animal species, and everyday objects, eliminating the need for specialized models

Scalable and robust: Works without requiring correspondence or object-specific data, thanks to permutation equivariance, TPE, and Procrustean training

Transfers learnings: Combined training shares spatial features across categories, enabling transfer learning.

Generalizes to OOD: Handles unseen categories and configurations



Challenges tackled

C1: Handling correspondence for diverse categories

Permutation equivariant architecture with TPE-based graph transformer.

C2: Making the Transformer More Efficient

Procrustean transformer to focus on deformable aspects within a canonical frame.

C3: Differentiating among unique categories

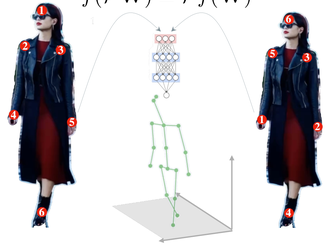
Graph-based transformer with TPE and joint connectivity information for precise category differentiation.

Core innovations

Permutation equivariance

Enables the model to process inputs in any order without affecting the output

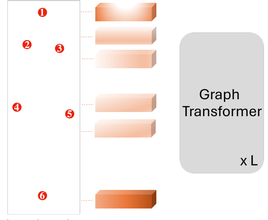
$$f(\mathcal{P}\mathbf{W}) = \mathcal{P}f(\mathbf{W})$$



TPE

Encodes spatial positions of landmarks into tokens for transformer processing.

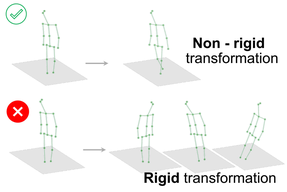
$$\sqrt{\frac{2}{D}}\left[\sin(\mathbf{W}_c\boldsymbol{\omega}+b);\cos(\mathbf{W}_c\boldsymbol{\omega}+b)\right]$$



Landmarks

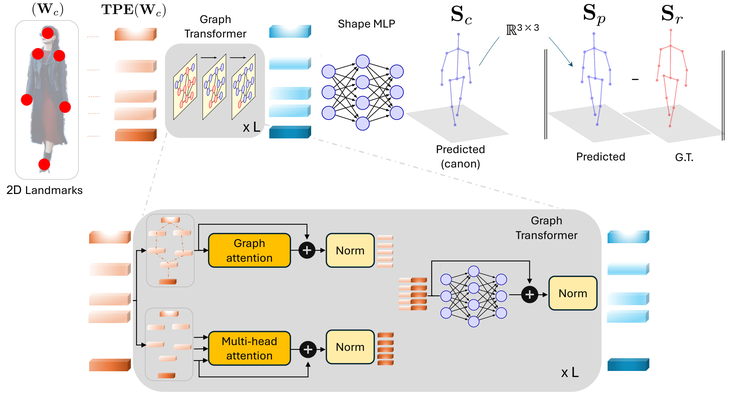
Procrustean transformer

Separates rigid transformations from non-rigid deformations in 3D data.



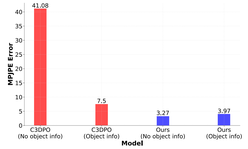
Aligns 3D predictions to a canonical frame, focusing learning non-rigid deformations only.

Architecture & Training



The training process involves encoding 2D landmarks into TPE tokens, which are then processed by a Graph Transformer through multiple layers. The Shape MLP predicts the canonical shape, and the Procrustean loss function minimizes the difference between predicted and ground truth (G.T.) structures.

Multi-category handling

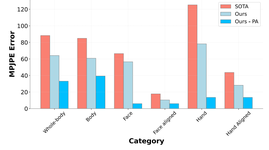


3D-LFM's ability to handle diverse categories without object-specific training

Evaluated on PASCAL3D+ dataset. Compared with C3DPO.

3D-LFM effectively handles multi-object 2D to 3D lifting, demonstrating its robustness and scalability.

Benchmark: WholeBody 3D SOTA



3D-LFM against leading specialized methods for human body, face, and hands categories.

Evaluated on H3WB dataset, compared against Jointformer (extra data), with results verified by H3WB team.

3D-LFM outperforms SOTA across all categories in 2D to 3D lifting without needing category-specific semantics.

Foundation Model for 3D Lifting

Validate 3D-LFM as a foundational model for diverse 2D-3D lifting tasks, handling multiple object types and data imbalances.

Transfer learning

Evaluated 3D-LFM trained on 30+ categories to improve performance on under-sampled categories through shared features, comparing isolated vs. combined dataset training.

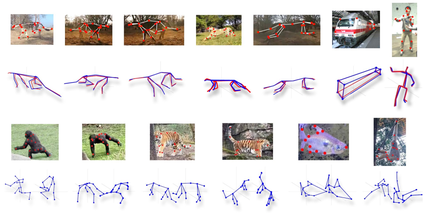
Dataset	Isolated	Combined
Animals3D	21.22	**12.5**
OpenMonkey	19.45	**9.59**
PASCAL3D+	4.31	**1.1**
AMASS	1.67	**1.66**

Combined training significantly reduces MPJPE (mm), showing better generalization and scalability for diverse 2D-3D lifting tasks.

Handling out of distribution data

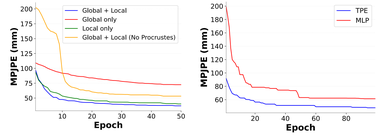
Evaluated on unseen categories from Acinoset, PASCAL3D+, and unseen rigs.

Successful 2D-3D lifting for unseen categories and in-the-wild data; despite joint variations. Quantitative results are in the paper.



Design ablations

Combining global and local attention with Procrustean alignment reduces MPJPE most effectively. TPE converges faster and more efficiently than MLP.



Takeaways

3D-LFM offers a scalable and adaptable solution for 2D-3D lifting, addressing data imbalances and generalizing to new categories. It achieves state-of-the-art results on benchmarks and excels in OOD scenarios with limited computational resources. This work lays a foundation for future advancements in 3D pose estimation and reconstruction, highlighting immense possibilities for diverse applications.