# 🐼PandaPose:3D Human Pose Lifting from a Single Image via Propagating 2D Pose Prior to 3D Anchor Space

Jinghong Zheng[1], Changlong Jiang[1], Yang Xiao[1, †], Jiaqi Li[1], Haohong Kuang[1], Hang Xu[1], Ran Wang[1], Zhiguo Cao[1], Min Du[2], Joey Tianyi Zhou[3]

[1]Huazhong University of Science and Technology [2]ByteDance Inc. [3]A*STAR, Singapore

## Introduction

- Estimating 3D human pose from a single RGB image is highly challenging due to (1) *2D pose noise inevitably propagating to 3D* and (2) *self-occlusion causing depth ambiguity*.
- Existing image-based methods rely on in-plane 2D features or direct 2D→3D regression, which suffer from:
  - ❌ Heavy reliance on accurate 2D joints
  - ❌ Lack of explicit depth modeling
  - ❌ Difficulty handling occluded joints
  - ❌ One-to-one joint regression with weak robustness

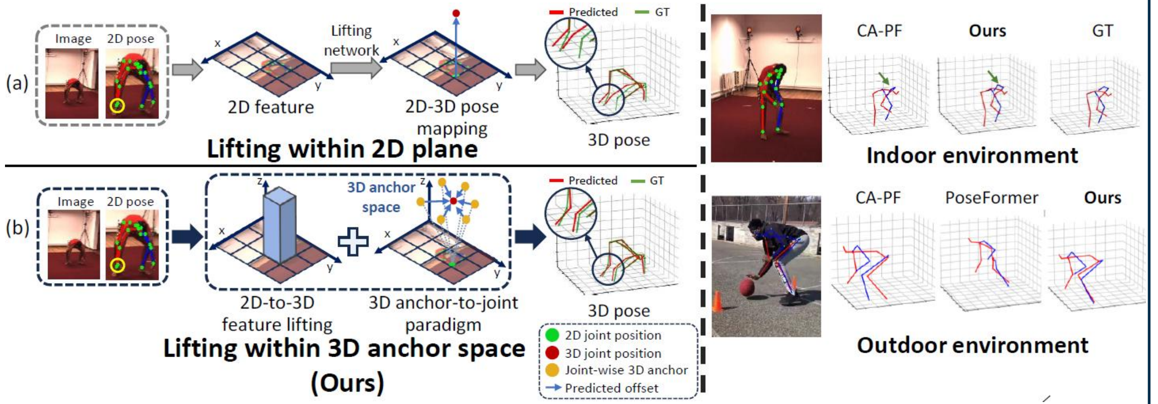



Fig1 Comparison between different 2D-to-3D human pose lifting manners

## Contribution

**3D Anchor Space as a Unified Intermediate Representation**

✓*Joint-wise adaptive 3D anchors — provide robust 3D priors to mitigate noisy 2D inputs*

✓*Joint-wise depth distribution — fine-grained depth estimation to resolve depth ambiguity and self-occlusion*

✓*3D Anchor-feature interaction decoder — fuse 3D anchors, depth cues and lifted 3D features*

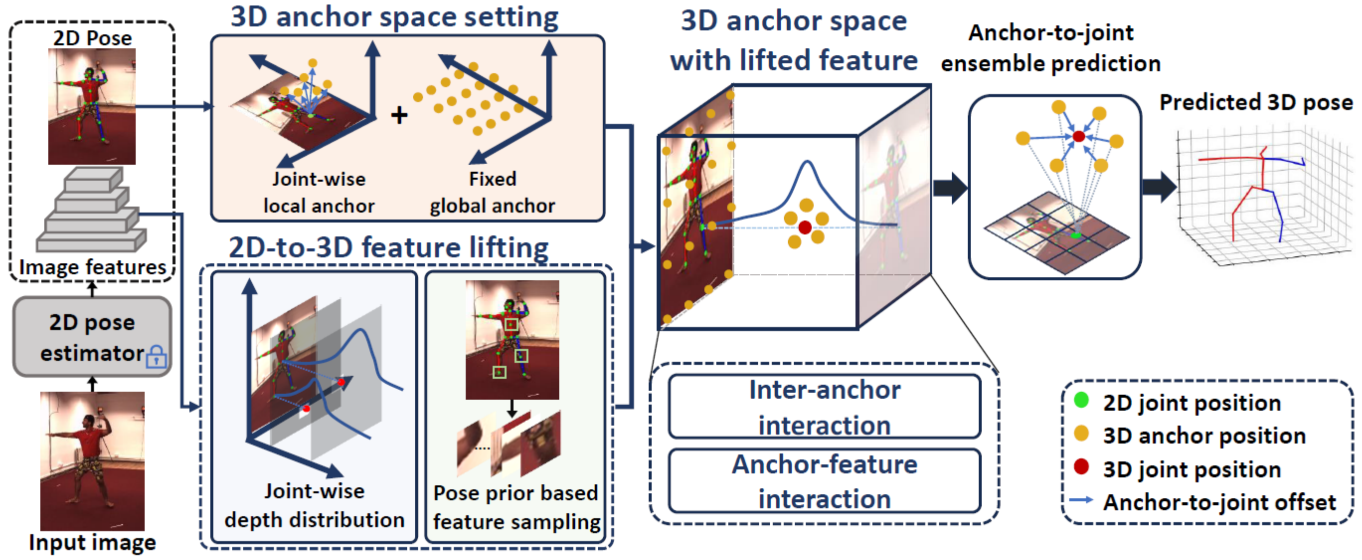
😀 *A robust, occlusion-resistant 3D pose lifting pipeline that surpasses prior methods in both normal and challenging settings.*

➡ *Paper, code are available at:* [https://github.com/DeepZheng/PandaPose](https://github.com/DeepZheng/PandaPose)

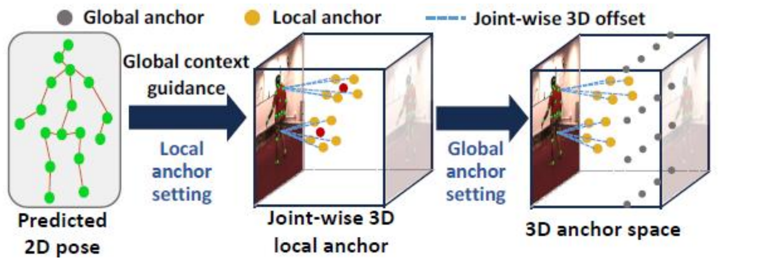
*Feel free to contact* deepzheng@hust.edu.cn / deepzheng@foxmail.com

## Method

### Overview of PandaPose

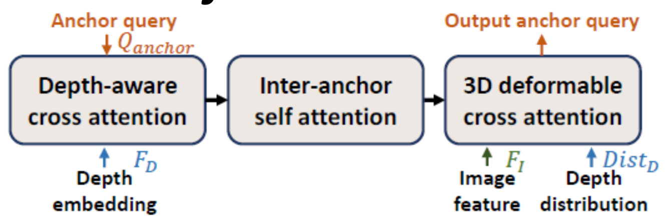



### 3D anchor setting




*1) Predicts learnable 3D local anchors per joint, conditioned on global 2D pose context*
*2) Combined with global fixed anchors to maintain global geometry cues*

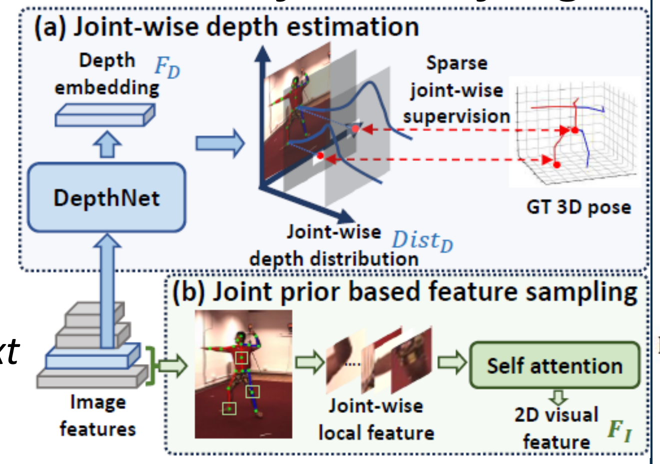
### Anchor-feature interaction




$$DCA(a) = \sum_{n \in N} W_n \phi(F_{3D}, P_a + \Delta S_n).$$

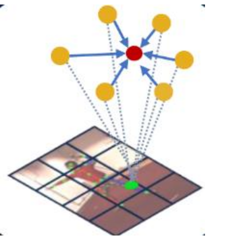
*Fusing adaptive 3D anchors with depth-aware and image features to produce unified anchor queries that enable accurate and occlusion-robust 3D pose prediction.*

### 2D-to-3D feature lifting




*Predict depth distribution per joint instead of one global depth map and uses 3D GT joint depth as sparse supervision*

### Anchor-to-joint ensemble prediction



$$P_j^{3D} = \sum_{a \in A} \tilde{W}_{a,j}(P_a + O_{a,j}),$$

*Achieves stable 3D reconstruction even with noisy 2D inputs*

## Results

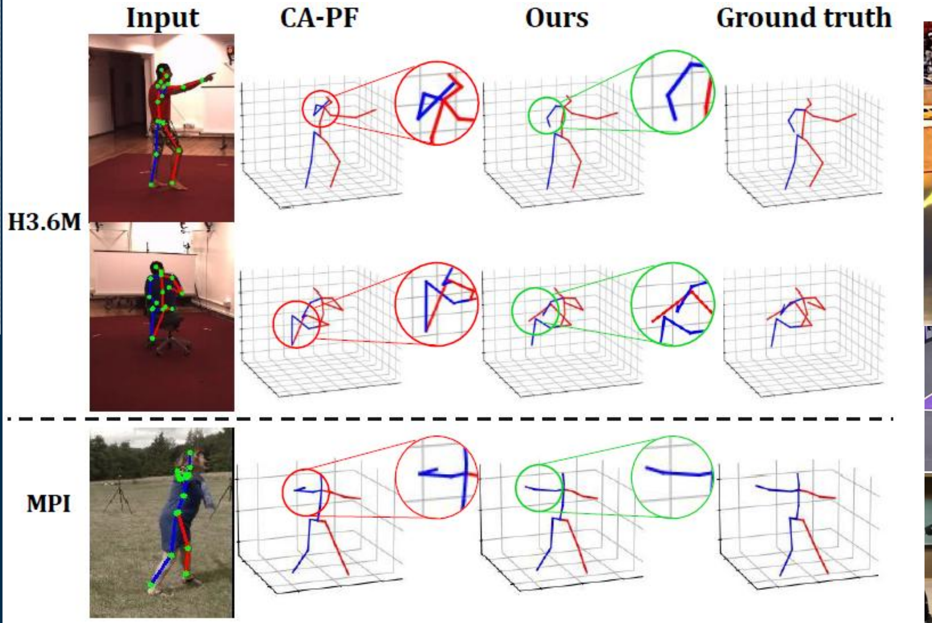
### Quantitative Comparison

| | Method | Venue | Frame | Parameters (M) for Lifting Module | MPJPE ↓ | PA-MPJPE ↓ |
|---|---|---|---|---|---|---|
| Sequence based | PoseFormer [57] | ICCV'21 | 81 | 9.5 | 44.3 | 34.6 |
| | MHFormer [20] | CVPR'22 | 351 | 24.8 | 43.0 | 34.4 |
| | MixSTE [50] | CVPR'22 | 243 | 33.6 | 40.9 | 32.6 |
| | P-STMO [31] | ECCV'22 | 243 | 4.6 | 43.0 | 34.4 |
| | STCFormer [35] | CVPR'23 | 243 | 18.9 | 41.0 | 32.0 |
| | KTPFormer [28] | CVPR'24 | 243 | 35.2 | 40.1 | 31.9 |
| Image based | ***Full test set*** | | | | | |
| | GraphSH [43] | CVPR'21 | 1 | 3.7 | 51.9 | - |
| | HCSF [48] | ICCV'21 | 1 | - | 47.9 | 39.0 |
| | GraFormer [56] | CVPR'22 | 1 | - | 51.8 | - |
| | Diffpose [10] | CVPR'23 | 1 | 1.9 | 49.7 | - |
| | Zhou *et al.*[59] | AAAI'24 | 1 | - | 46.4 | - |
| | HiPART [58] | CVPR'25 | 1 | 2.4 | 42.0 | - |
| | CA-PF [55] | NeurIPS'23 | 1 | 14.1 | 41.4 | 33.5 |
| | PandaPose (Ours) | | 1 | 15.2 | **39.8** (1.6↓) | **32.7** (0.8↓) |
| | ***Challenging subset*** | | | | | |
| | CA-PF [55] | NeurIPS'23 | 1 | 14.1 | 82.4 | 82.0 |
| | PandaPose (Ours) | | 1 | 15.2 | **73.1** (9.3↓) | **69.9** (12.1↓) |

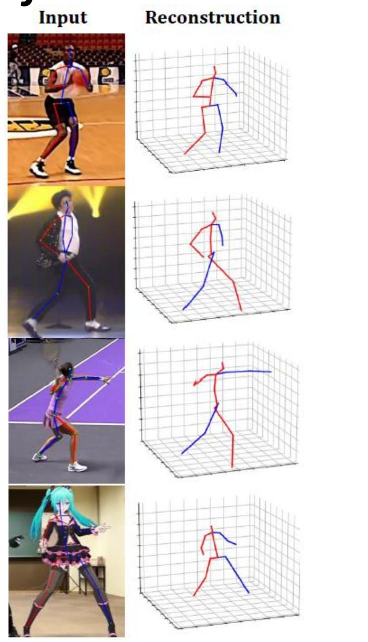
*Achieves **SOTA** on Human3.6M, with a significant improvement **(12.1mm)** in challenging subset*

### Qualitative Comparison

#### Challenging case in dataset




#### OOD sample from Internet




### Ablation Studies

| Global fixed anchor | Adaptive local anchor | MPJPE↓ (Full) | MPJPE ↓ (Challenging) |
|---|---|---|---|
| PandaPose w/o anchor | | 42.1 | 81.9 |
| ✓ | | 40.8 (1.3↓) | 76.2 (5.0↓) |
| | ✓ | 40.1 (2.1↓) | 74.0 (7.2↓) |
| ✓ | ✓ | 39.8 (2.3↓) | 73.1 (8.1↓) |

Table 4: Anchor setting strategy comparison.

| Anchor feature | Depth distribution | MPJPE↓ (Full) | MPJPE ↓ (Challenging) |
|---|---|---|---|
| 2D | - | 40.9 | 80.8 |
| 3D | Single | 40.3 (0.6↓) | 75.9 (4.9↓) |
| 3D | Joint-wise | 39.8 (1.1↓) | 73.1 (7.7↓) |

Table 5: Ablation study of joint-wise 3D feature lifting.

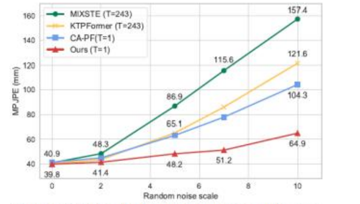



Figure 9: We add Gaussian noise with varying scales to the input 2D poses of different methods to test the robustness to noisy inputs.