# Animal Pose Refinement in 2D Images with 3D Constraints

Xiaowei Dai[1], Shuiwang Li[2], Qijun Zhao[1,3], Hongyu Yang[1,3]

[1]National Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University [2]College of Information Science and Engineering, Guilin University of Technology [3]College of Computer Science, Sichuan University

## Introduction

### Motivation

- Animal pose has many potential applications in various fields.
- It is impractical to bring wild animals into a laboratory environment for scanning in specific poses.
- However, uncontrollable illumination, complex backgrounds and random occlusions in in-the-wild animal images often lead to large errors in pose estimation.

### Objective

- To reduce the pose estimation errors, we propose a method for refining the initial animal pose with 3D prior constraints.
- We construct a 3D animal pose dataset using synthetic methods for 3D pose dictionary learning.
- We collect and manually annotate images to build a 2D animal pose dataset for algorithm evaluation.

## Dataset

### 3D Pose Dataset

- This dataset contains more than 400 samples.

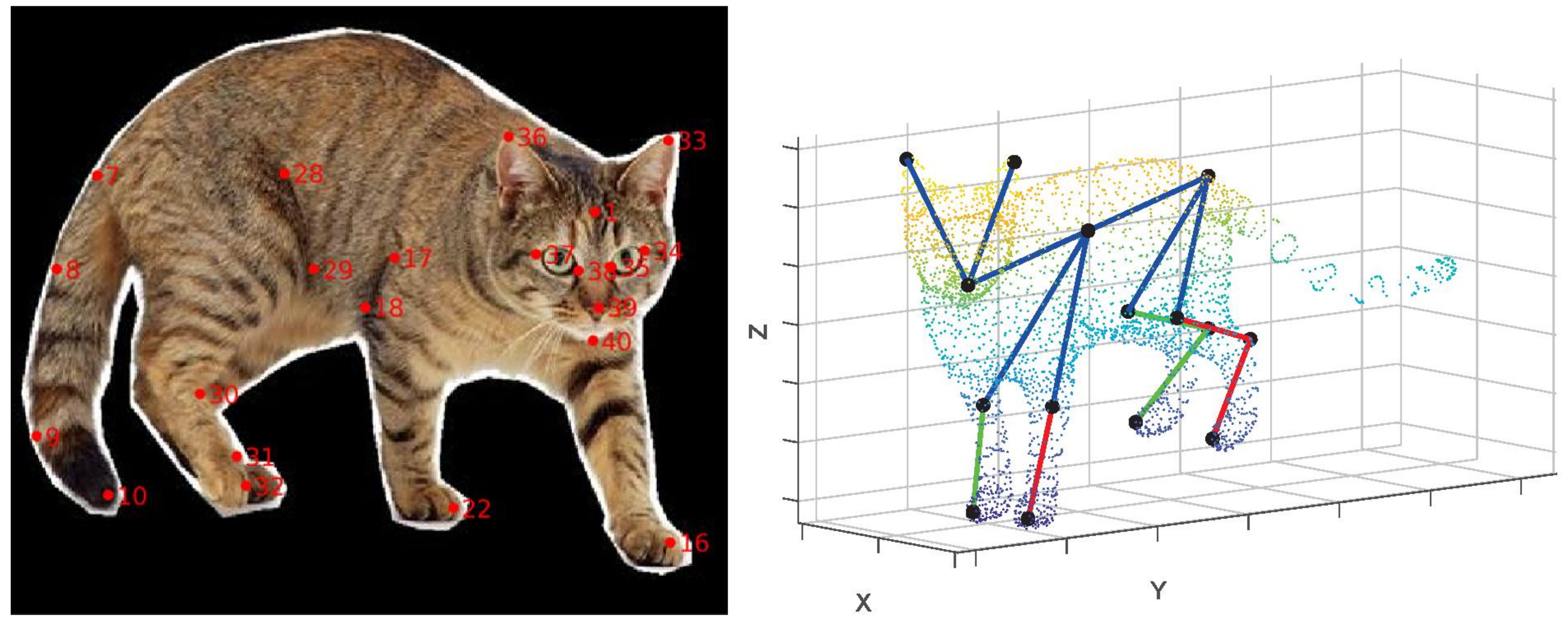


Figure 1: (a) are examples of keypoints in the dataset Cat. Red dots indicate the keypoints. (b) The joints defined in ATRW[1].

**References**. [1] Li S, Li J, Tang H, et al. Atrw: A benchmark for amur tiger reidentification in the wild. In Proc. Multimedia, 2020.
[2] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In Proc. CVPR, 2016.

## Method

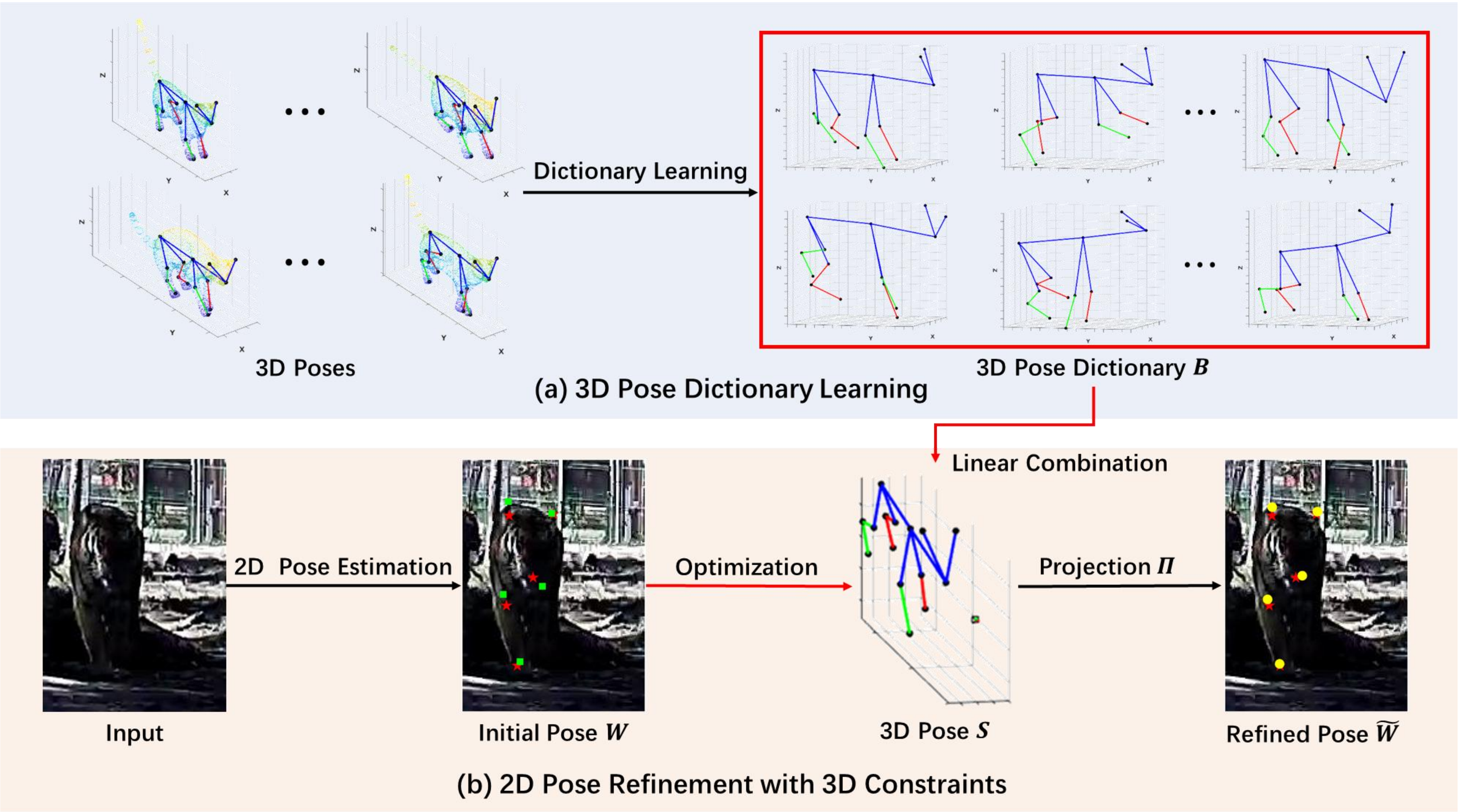


Figure 2: Overview.

### 3D Pose Dictionary Learning

- we use dictionary learning to find a good basis for 3D poses.

$$\min_{B,C} \sum_{n=1}^{N} \frac{1}{2} \left\| S_n - \sum_{k=1}^{K} C_{n,k} B_k \right\|_F^2 + \lambda \|C\|_1$$

$$s.t. \quad C_{n,k} \geq 0, \quad \|B_k\|_F \leq 1, \quad \forall_k \in [1, K], \quad n \in [1, N]$$

$S_n$ denotes a 3D pose; $B_k$ is the basis pose; $C_{n,k}$ represents the $k$th coefficient.

### 2D Pose Refinement with 3D Constraints

- we use the following objective function to estimate a latent 3D pose.

$$\min_{M_{i,\dots,}M_K} \frac{1}{2} \left\| W - \sum_{k=1}^{K} M_k B_k \right\|_F^2 + \alpha \sum_{k=1}^{K} \|M_k\|_2$$

- we can finally obtain our refined 2D pose $\widetilde{W}$.

$$\widetilde{W} = \Pi S = \Pi \sum_{k=1}^{K} c_k R_k B_k = \sum_{k=1}^{K} M_k B_k$$

## Experiments

### Comparison with State of the Art

| Method | (0, 45] | (45, 65] | (65, 100] |
|---|---|---|---|
| ResNet [2] | 35.0 | 56.8 | 84.3 |
| Ours | **37.2** | **57.4** | **84.3** |
| HRNet [3] | 35.9 | 57.0 | 84.7 |
| Ours | **38.8** | **57.8** | **84.7** |
| CC-SSL [4] | 32.1 | 55.4 | 75.4 |
| Ours | **35.1** | **55.4** | **75.4** |
| UDA [5] | 34.4 | 55.2 | 76.5 |
| Ours | **38.5** | **55.7** | **76.5** |

### Qualitative Results

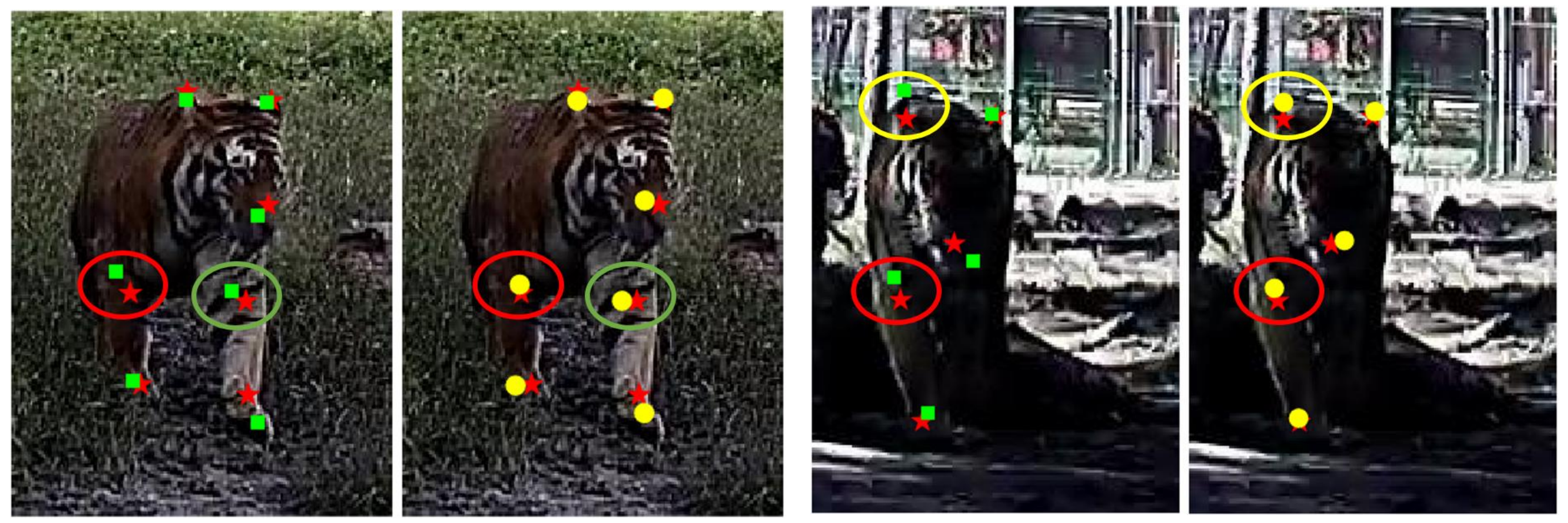

Figure 3: Examples. Ground truth, initial and refined poses are shown in red, green and yellow colors, respectively.

## Conclusion

- It is worth emphasizing that the proposed method works as a plug-in post-processing module and can be attached to existing animal pose estimation.
- We employ a dictionary to encode 3D prior constraints to refine the initial animal pose.
- To address the scarcity of 3D animal data, we use synthetic data for dictionary learning.

**References.** [3] Sun, Ke, et al. Deep high-resolution representation learning for human pose estimation. In Proc. CVPR, 2019.
[4] Mu, Jiteng, et al. Learning from synthetic animals. In Proc. CVPR, 2020.
[5] C. Li and G. H. Lee. From synthetic to real: Unsupervised domain adaptation for animal pose estimation. In Proc. CVPR, 2021.