Taran Lau

tlau44@gatech.edu

Week 5 Report

Lizard CV Movement Weekly Report

Time-Log

- Very brief sections of bullet points for the following:
    - What did you do this week?
        - Updated project website to include detailed description
        - Added tools list and installation guides for them to repo
        - Used deep cut lab to create new labeled data from videos
    - What are you going to do next week
        - Create additional marker points in labeled data for snout, neck, elbows, hands, feet.
    - Blockers, things you want to flag, problems, etc.
        - No attendance from any advisors at scheduled meeting.

Abstracts:

Paper:

X-Pose: Detecting Any Keypoints

Abstract:

This work aims to address an advanced keypoint detection problem: how to accurately detect any keypoints in complex real-world scenarios, which involves massive, messy, and open-ended objects as well as their associated keypoints definitions. Current high-performance keypoint detectors often fail to tackle this problem due to their two-stage schemes, under-explored prompt designs, and limited training data. To bridge the gap, we propose X-Pose, a novel end-to-end framework with multi-modal (i.e., visual, textual, or their combinations) prompts to detect multi-object keypoints for any articulated (e.g., human and animal), rigid, and soft objects within a given image. Moreover, we introduce a large-scale dataset called UniKPT, which unifies 13 keypoint detection datasets with 338 keypoints across 1,237 categories over 400K instances. Training with UniKPT, X-Pose effectively aligns text-to-keypoint and image-to-keypoint due to the mutual enhancement of multi-modal prompts based on cross-modality contrastive learning. Our experimental results demonstrate that X-Pose achieves notable improvements of 27.7 AP, 6.44 PCK, and 7.0 AP compared to state-of-the-art non-promptable, visual prompt-based, and

textual prompt-based methods in each respective fair setting. More importantly, the in-the-wild test demonstrates X-Pose's strong fine-grained keypoint localization and generalization abilities across image styles, object categories, and poses, paving a new path to multi-object keypoint detection in real applications. Our code and dataset are available at https://github.com/IDEA-Research/X-Pose.

Summary:

The paper titled "X-Pose: Detecting Any Keypoints" introduces an innovative end-to-end framework designed to accurately detect keypoints across a wide array of objects in complex real-world scenarios. Traditional keypoint detectors often struggle with diverse and unstructured environments due to their reliance on two-stage processes and limited prompt designs. To overcome these limitations, the authors propose X-Pose, which utilizes multi-modal prompts—visual, textual, or their combinations—to detect keypoints for various objects, including articulated (e.g., humans and animals), rigid, and soft objects within a single image. This approach enables the model to handle the vast diversity and complexity inherent in real-world images.

To support the training of X-Pose, the researchers developed UniKPT, a comprehensive dataset that consolidates 13 keypoint detection datasets, encompassing 338 keypoints across 1,237 categories and over 400,000 instances. Training with UniKPT allows X-Pose to effectively align textual and visual information through cross-modality contrastive learning. Experimental results demonstrate that X-Pose achieves significant improvements over state-of-the-art methods, with increases of 27.7 AP, 6.44 PCK, and 7.0 AP in respective evaluations. Furthermore, tests in uncontrolled environments highlight X-Pose's robust ability to generalize across various image styles, object categories, and poses, marking a significant advancement in multi-object keypoint detection for practical applications.

What did you do and prove it

Tell us in about a paragraph what you did. Additionally, provide relevant “proof” of it.

- Contribute to repo: https://github.gatech.edu/tlau44/Lizard-Movement-Spring2025
    - installation guide for tools
    - project objectives and useful links
    - sample data and labels