# A Large-Scale Video Dataset for Moving Camouflaged Animals Understanding

Tuan-Anh Vu[1,2] Ziqiang Zheng[1] Qing Guo[2] Ivor Tsang[2] Sai-Kit Yeung[1]

[1]The Hong Kong University of Science and Technology [2]Centre for Frontier AI Research (CFAR), A*STAR

CVPR SEATTLE, WA JUNE 17-21, 2024

## Motivations

- Training DL models is often **resource-intensive**, requiring **large amounts** of **accurately annotated** data.
- Enhancing data **diversity** and **scale** is essential to mitigate algorithmic biases.
- Only **a few** studies have specifically targeted **camouflaged animals**.

| Dataset | Venue | # videos / frames | # species | Frequency | Class. | B.Box | Mask | Motion | Coarse OF | Expres. |
|---|---|---|---|---|---|---|---|---|---|---|
| CAD [32] | ECCV'16 | 9 / 839 | 6 | every 5 frames | ✓ | | ✓ | | | |
| MoCA [18] | ACCV'20 | 141 / 37,250 | 67 | every 5 frames | ✓ | ✓ | | ✓ | | |
| MoCA-Mask [4] | CVPR'22 | 87 / 22,939 | 44 | every 5 frames | ✓ | | ✓ | | | |
| **CamoVid50K** | - | **149 / 44,270** | **70** | **every frames** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

**Table 1. Comparison with existing video camouflaged animal datasets.** Class.: Classification Label, B.Box: Bounding Box, Motion: Motion of Animal, Coarse OF: Coarse Optical Flow, Expres.: Expression.

## Contributions

- **CamoVid40k** - a **large-scale** video dataset dedicated to the understanding of camouflaged animals, comprises **149** videos with **44,270** finely annotated frames, covering **69** animal categories.
- We provide **a wider variety of annotation types** (animal categories, bounding box, annotated mask, coarse optical flow, expression), making it a more effective benchmark for CAU tasks.
- This dataset supports **a broad range of downstream tasks** as shown in Figure 1, including classification, detection, segmentation (semantic, referring, motion), and optical flow estimation, etc.

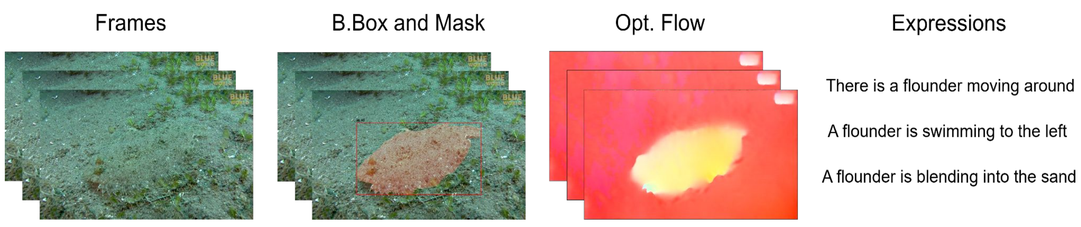


**Figure 1.** Example from our CamoVid40K dataset with bounding box, mask, coarse optical flow, and expression.

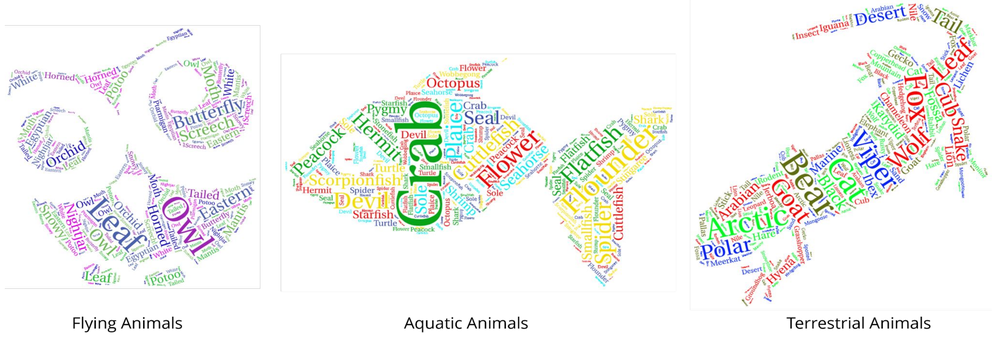


**Figure 2.** Word cloud of category distribution of camouflaged animals.

## Dataset Processing, Specification and Statistics

- **Data Curation:** we built our dataset from Camouflaged Animals Dataset (CAD), Moving Camouflaged Animals (MoCA), MoCA-Mask, Marine Video Kit (MVK) and crawled video from internet.
- **Preprocessing & Filtering:** we collected around 1,929 videos and manually checked and filtered blurry, irrelevant videos with obvious animals.
- **Box & Mask Annotation:** we utilized SAM model for mask initialization and XMem for mask propagation. Then, we adopt the perceptual camouflage score to quantify the effectiveness of animals' camouflage, i.e. how successfully an animal blends into its background.
- **Coarse Optical Flow Annotation:** we utilized RAFT method to compute all pairwise optical flow fields, then filtered the estimated optical flow using cycle and appearance consistency checks (using DINO features). Finally, we applied chain cycle consistent correspondences to create denser correspondences.
- **Motion Annotation:** we manually labeled our dataset by their types of motion (locomotion, deformation, still).
- **Expression Annotation:** we used GPT-4V to create short, accurate descriptions for each frame. For aquatic animals, we switched to for better accuracy. We then refined all annotations, choosing the best three captions per video sequence, and removed any objects that couldn't be identified.

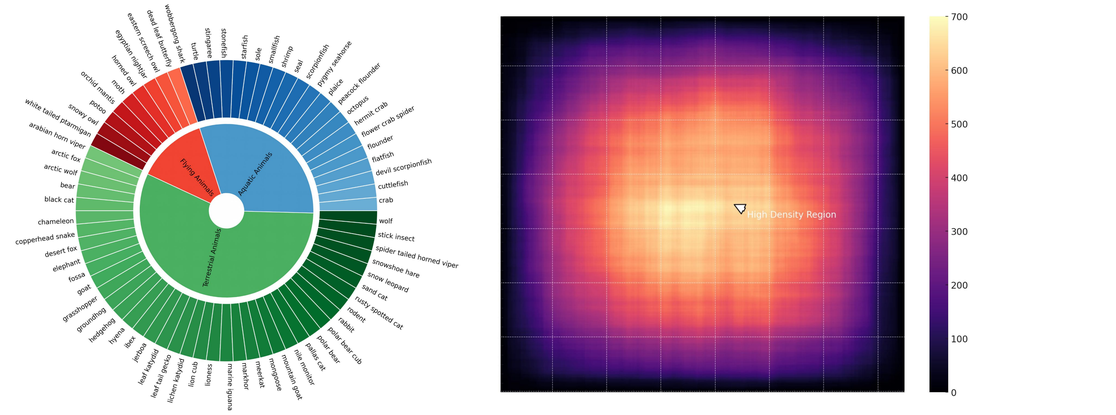


**Figure 3.** Left: Taxonomic structure of our dataset. Right: Spatial distribution of animals' position based on bounding box.

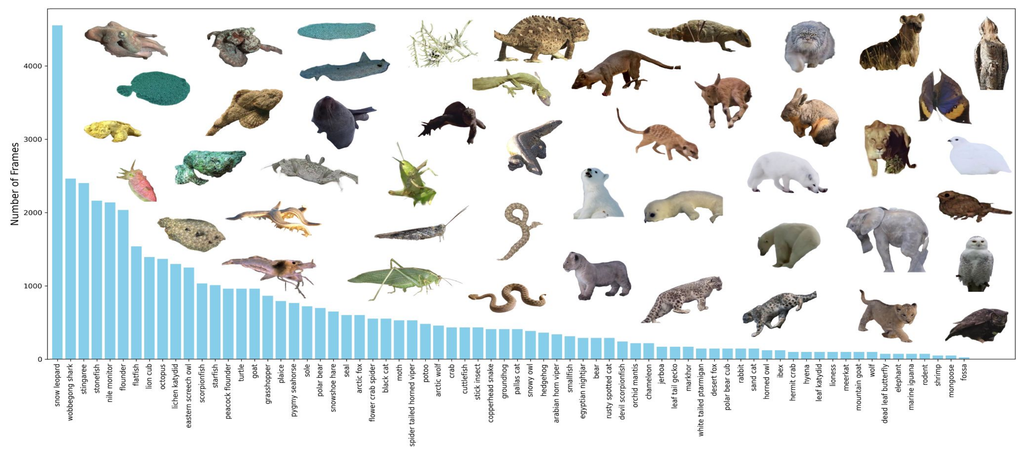


**Figure 4.** Category distribution (number of frames) and some visual examples (extracted animal masks) of our dataset.