## A Experiment on zero-shot classification

2 We conduct performance evaluations on OpenCLIP models of different sizes on Species196-L. All

3 models we tested are pretrained on LAION-2B (en) with  $224 \times 224$  image size. As shown in table 1,

4 we first compare different prompts settings and finally select {"{c}.", "photo of {c}, a type

5 of species.", "photo of {c}, a type of invasive species."} as our prompts.

Prompt number	Prompt	TOP-1 ACC.
(1)	{c}	10.7
(2)	a photo of $\{c\}$ , a type of species.	9.58
(3)	a photo of a {c}, a type invasive species.	9.58
(4)	photo of {c}, a type of species.	10.19
(5)	photo of {c}, a type of invasive species.	10.23
(6)	a photo of the {c}, a type of species.	10.06
(7)	a photo of the {c}, a type of invasive species.	9.8
(1) + (5)	-	10.87
(1) + (4) + (5)	-	10.94
(1) + (4) + (5) + (6)	-	10.91

Table 1: Comparison of different prompts and combination. The model used for comparison is VIT-B/32 and  $\{c\}$  denotes classnames.

6 We then examine the influence of using different kinds of classnames. In our analysis, we employ

7 scientific names, common names, appearance descriptions generated by ChatGPT, and Minigpt-4

8 as classnames. In some cases, the AI generated descriptions is irrelevant to the required response.

9 Table 2 presents examples of both relevant and irrelevant descriptions generated by the model.

The results (see Table 3) indicate that in our dataset, using common names as classnames yields the best zero-shot classification performance. For the generated descriptions used as classnames, ChatGPT outperforms the scientific names as classnames in most models, while MiniGPT-4 has lower inference accuracy due to the high proportion of irrelevant descriptions. Compared to other fine-grained datasets, the significantly lower accuracy of Species196-L suggests that it poses a new challenge in the field of zero-shot fine-grained classification.

	Species196-L						
Model	Scientific name	Common name	Description (ChatGPT)	Description (Minigpt-4)	Cars	FGVC Aircraft	Flowers102
ViT-B/32	10.94	16.71	10.25	5.76	86.05	24.551	71.62
ViT-B/16	10.40	16.90	10.81	6.59	88.50	26.97	71.34
ViT-L/14	12.07	18.43	12.69	6.29	92.64	36.75	75.83
ViT-H/14	15.00	22.57	15.95	6.23	93.36	42.60	80.13

Table 3: Experiment results of zero-shot classification on *Species196-L* as well as other fine-grained datasets.

### **B** Examples of annotations in *Species196-L* dataset

Figure 1 displays several examples of annotation boxes in our *Species196-L* dataset. Although each
image in our dataset contains only one class of label, the objective detection experimental results
shows that the dataset is challenging for detecting small targets, as some images feature densely

shows that the dataset is chanenging for detecting small targets, as some n

20 populated biological instances.

Table 2: Example of relevent and irrelevant descriptions generated by ChatGPT and Minigpt-4.For ChatGPT, the input is only scientific name. For Minigpt-4, we select one image for each category and design five prompts for inquiry.

Prompts	Sample relevant answers	Sample irrelevant answers				
Descriptions generated by ChatGPT						
1. Briefly describe the appearance characteristics of the species {c}, excluding size descriptions.	A pink, oval insect with a wax-covered body and long filaments ( <i>Phenacoccus solenopsis</i> )	The adults are variable in size, coloration, colour pattern and facies ( <i>Rhabdoscelus obscurus</i> )				
Descriptions generated by M						
1. Describe the organism's physical features, such as its size, shape, and color.	It has two legs. The top part has a brown, rough surface, and the bottom part has a dark brown, smooth surface. The middle part has a dark brown surface. ( <i>Spodoptera littoralis</i> )	I'll look at this image and answer your questions. ( <i>Hoplocerambyx spinicornis</i> )				
2. Describe the texture and surface features of the organism in the picture.	The beetle's skin is smooth, its body is yellow, and its antennas are brown with orange tips. ( <i>Scolytus scolytus</i> )	Do you think that the organism's texture and surface features would cause it to survive in the habitat shown in the picture? ( <i>Avena barbata</i> )				
3. Describe any physical adaptations of the organism, such as specialized structures for defense, locomotion, or feeding.	The insect has one set of eyes and antennae to look for food and also it can fly and run away from its predator. ( <i>Gonipterus</i> <i>scutellatus</i> )	An organism's life cycle. A complete cycle can include any type of reproduction, but is limited to the life span of one organism. ( <i>Aleurodicus</i> <i>dispersus</i> )				
4. Describe any unique patterns or markings on the organism's body.	The organism is small, but has two eyes. The eyes are very sharp. The antennae are long and can feel things very well ( <i>Hoplocampa testudinea</i> )	Answer the questions from the student's perspective. ( <i>Anastrepha</i> )				
5. Describe the organism's body structure, including any visible organs or systems.	It has four wings, a head with eyes, a thorax with legs, and a stomach for digesting food. ( <i>Dasineura mali</i> )	How can humans benefit from knowing about the organism? ( <i>Hylobius pales</i> )				

# 21 C Examples of *Species196-U* dsataset

We use image-retrieval for creating *Species196-U*. For each category, we randomly sampled three images and retrieved 8,000 unlabeled images per class from LAION5B. As shown in Figure 2, even at the 5,000th image sorted by descending similarity scores, the retrieved image remains highly relevant to the original image.

# **D** Hosting and maintenance plan

Both the Species196-L and Species196-U datasets are made publicly available at https:// species-dataset.github.io/. This website is hosted on Github Pages, a widely-used website hosting service. The website contains introductions, experiment results, terms of use, and links



Figure 1: Samples of Species196-L bounding box annotations. The top two rows show easy cases, while the bottom three rows present hard cases, including crowdedness, complex backgrounds, and tiny objects.

to download the datasets, as well as usage guides. We maintain the data using Google Drive and 30 Baidu Cloud, where we store the original URLs to download the images, ensuring that the dataset 31 will be available for an extended period. Additionally, we will provide instructions on how to down-32 load and organize the data with code. For further maintenance, we will continue refining our dataset, 33 such as correcting incorrect labels and annotations in Species-L, and updating a larger and more 34 comprehensive version of Species-U. 35

#### License Ε 36

The Species 196-L dataset is provided to You under the terms of the Creative Commons Attribution-37 NonCommercial-ShareAlike 4.0 International Public License (CC BY-NC-SA 4.0), with the addi-

38

tional terms included herein. This dataset is used only for non-commercial purposes such as aca-39 demic research, teaching, or scientific publications. We prohibits You from using the dataset or any 40

derivative works for commercial purposes, such as selling data or using it for commercial gain. 41

The Species196-U dataset is distribute under the Creative Common CC-BY 4.0 license, which poses 42

no particular restriction. The images are under their copyright. 43



Figure 2: Clip-retrieval process of *Species16-U* from *Species196-L*. Displaying similarity scores in descending order, we show items No. 100, 500, 1000, and 5000.