

1 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 224×224 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.

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

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

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 Minigt-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.

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

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

Model	Species196-L				Cars	FGVC Aircraft	Flowers102
	Scientific name	Common name	Description (ChatGPT)	Description (Minigt-4)			
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

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

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

Table 2: Example of relevant and irrelevant descriptions generated by ChatGPT and Minigt-4. For ChatGPT, the input is only scientific name. For Minigt-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 Minigt-4		
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 antennae 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 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 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* dataset

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

26 D Hosting and maintenance plan

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

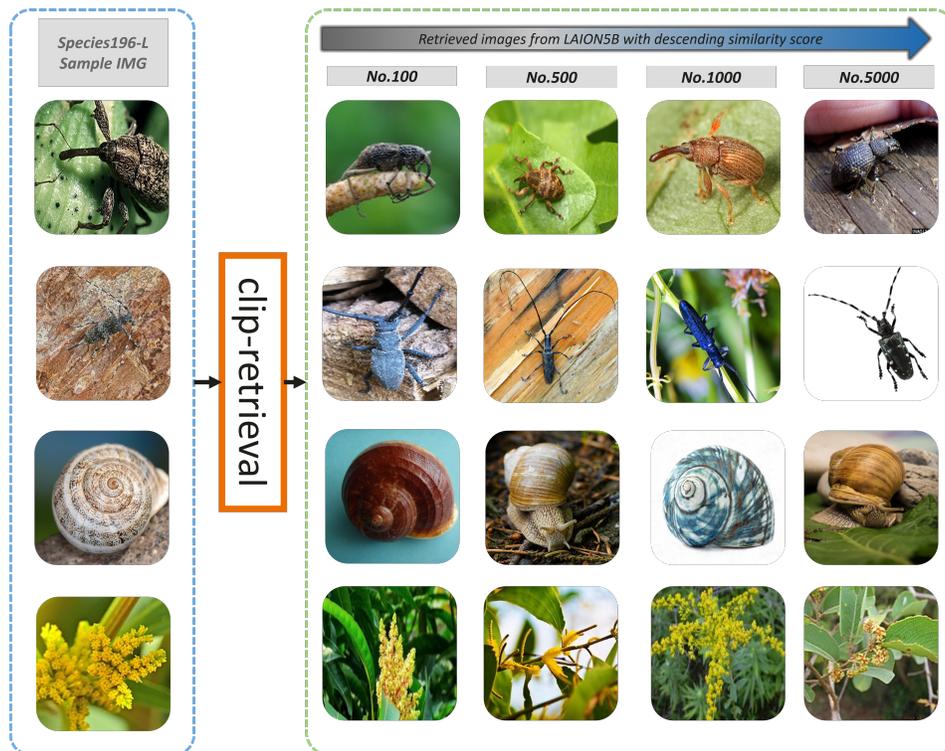


Figure 2: Clip-retrieval process of *Species16-U* from *Species196-L*. Displaying similarity scores in descending order, we show items No. 100, 500, 1000, and 5000.