EcoWikiRS: Learning Ecological Representation of Satellite Images from Weak Supervision with Species Observations and Wikipedia

Supplementary Material

Fine-tuning strategy	OA	F1
Last transformer block	21.6	10.8
Last projection layer	28.5	18.8
EMA	28.9	20.1
Positional encoding + last projection layer (our)	**30.1**	**20.4**

Table 4. Comparison of fine-tuning strategies for the SkyCLIP model trained with WINCEL on the EcoWikiRs dataset

Text type	Number of sentences	Number of unique sentences	Average number of sentences per location
habitat	2'998'305	18'693	10.9
keywords	3'728'644	21'832	13.6
random	19'642'735	103'065	71.6

Table 5. Number of sentences and unique sentences in the different text types of WikiRS dataset

8. Study of fine-tuning strategies.

Table 4 compares the performance of SkyCLIP under various fine-tuning configurations with WINCEL. The text encoder remains frozen, while different layers of the visual encoder undergo fine-tuning: the final transformer block, the last projection layer with or without additional projection layers, and fine-tuning with Exponential Moving Average (EMA). By fine-tuning the last transformer block, the performance significantly deteriorates compared to the other strategies. This may result from the loss of original pretrained representations, as the fine-tuning dataset is comparatively small. Both EMA fine-tuning and updating the last projection layer obtain good performances, while being inferior to training the positional encoding and the last projection layer. The latter outperforms all other fine-tuning strategies, both in terms of OA and F1 score. Therefore, we adopted this strategy in this work.

9. Additional visual results

As discussed in Section 4, generalist species, such as the common blackbird, can live in several habitats. Thus, sentences describing their habitat are likely to be uninformative or irrelevant when paired with a given image. To demonstrate that our model accurately identifies sentences relevant to a given image, we study the cross-modal similarity scores between sentences describing the common blackbird habitat, a generalist species, with aerial images depicting different land covers. We computed the text and visual representations for 4 aerial images and 6 sentences from the Wikipedia article of the *Turdus merula* (common blackbird) with SkyCLIP model fine-tuned with the WINCEL loss and present the cross-modal similarity scores in Figure 5. When the image (ex. image (a)) or the text (sentences 5 and 6) is irrelevant to the other modality, the similarity scores are close to zero or even negative. Conversely, pertinent image-text pairs obtain high similarity values.

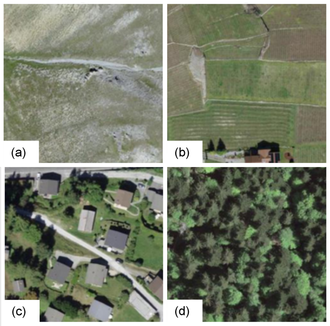
10. Prompting for zero-shot classification

Due to the presence of spurious concept biases, several works highlighted the importance of prompt engineering [2, 39] for attaining high zero-shot classification performances. To select satisfying prompts for the EUNIS ecosystem classification task, we compared manually designed prompt templates. The results are shown in Table 6. Overall, the prompt with EUNIS ecosystem names obtains the highest performance for all backbones, except for RemoteCLIP, whose performances significantly increase by using a prompt template such as "a remote sensing image of {}". Figure 9 presents additional illustrations of cross-modal similarity values, similar to those presented in Figure 4, with the top-5 most similar sentences for the pretrained and the fine-tuned SkyCLIP model for each dataset sample.

11. Dataset construction and statistics

11.1. Wikipedia articles parsing

Extraction of Wikipedia articles. We downloaded a dump of all of Wikipedia through Wikimedia at dumps.wikimedia.org. We processed the dump with the *BeautifulSoup* [42] and *mwparserfromhell* python packages. We extracted articles matching the "Speciesbox" template. We created the species binomial name using the genus and the species string from the Speciesbox. We matched the species article to GBIF species and saved the article content. We cut the articles into sections and extracted sentences from the habitat section, containing keywords from the list mentioned below and also saved all sentences. Statistics on the total number of sentences and



(a)	(b)	(c)	(d)	**Sentences from the Wikipedia article of *Turdus merula***
-0.66	0.19	0.40	0.05	The common blackbird also lives in parks, gardens and hedgerows.
-0.64	0.06	0.42	0.73	Common over most of its range in woodland, the common blackbird has a preference for deciduous trees with dense undergrowth.
-0.49	0.54	0.21	0.16	It eats a wide range of native and exotic fruit, and makes a major contribution to the development of communities of naturalised woody weeds.
-0.47	-0.06	0.24	-0.14	Near human habitation, the main predator of the common blackbird is the domestic cat, with newly fledged young especially vulnerable.
-0.37	-0.01	-0.03	-0.19	The common blackbird breeds in temperate Eurasia, North Africa, the Canary Islands, and South Asia.
-0.25	-0.26	-0.23	0.00	Pairs stay in their territory throughout the year where the climate is sufficiently temperate.

Figure 5. Cross-modal similarities values between sentences from the Wikipedia article of *Turdus merula* (common black bird) and various images from our dataset representing different land cover. High similarity values are shown in green, and low similarity values are depicted in red.

Model	**SkyCLIP**		**CLIP**		**GeoRSCLIP**		**RemoteCLIP**		**Average**	
Prompting approach	**OA**	**F1**	**OA**	**F1**	**OA**	**F1**	**OA**	**F1**	**OA**	**F1**
EUNIS class name	**30.1**	**20.4**	**30.9**	20.1	**29.5**	**20.4**	20.9	11.9	**30.0**	**20.3**
EUNIS class description	21.8	15.4	24.7	16.5	21.8	15.4	25.0	16.0	23.3	15.8
“an aerial image of {}”	28.3	19.3	29.5	20.5	28.3	19.3	24.5	17.5	27.6	19.1
“a remote sensing image of {}”	29.0	18.7	30.5	**20.9**	29.0	18.7	**25.8**	**18.5**	28.6	19.2
“a satellite image of {}”	28.4	18.4	29.0	20.5	28.4	18.4	25.1	17.2	27.7	18.6

Table 6. Comparison of different prompts for the pretrained models and those trained with the WINCEL loss on the EcoWikiRS dataset. OA and F1 values are average metrics values over 5 models, similarly to results from Table 1.

unique sentences are shown in Table 5.

Extraction of keywords sentences. We selected sentences containing at least one of the following strings:

Habitat keywords

urban, city, town, road, railway, rail, highway, port, airport, mineral, dump, construction, green, sport, arable, farmland, irrigated, fruit, berry, plant, tree, olive, crop, pastures, vineyards, cultivation, agriculture, vegetation, forest, forestry, grassland, heathland, moors, woodland, shrub, beach, dunes, sand, rock, bareland, vegetated, inland, marshes, burnt, water, coast, coastal, lagoons, sea, ocean, saline, peatbogs, estuaries, surface, grass, grassland, dry, mesic, littoral, seasonal, wet, alpine, subapline, arctic, scrub, temperate, temperature, mediterranean-montane, plantation, coniferous, deciduous, anthropogenic, coppice, screes, inland, cliffs, outcrops, snow, ice, ice-dominated, garden, park, village, building, transport, hard-surfaced, constructed, runway, airport, road, vehicle, bridge, shrubwood, weed, bareland, fanshaped, ravine, gravel, rectangular, high, low, coastline, cemetery, greenbelt, circular, cloud, dam, terrace, weed, viaduct, wetland, wood, habitat, ecosystem, landcover, eco, supralittoral, zone, area, saline, density, arborescent, hot, cold, thermo, warm, xerophytic, calcareous, broadleaved, leave, mires, pavements, shores, salt, montane, polygon, evergreen, waste, sparse, dense, coastal, atlantic, anthropogenic, reed, shingle, mediterranean, artificial, park, flower, prairie.

11.2. GBIF filtering

The GBIF download [19] included the following criteria:

- *BasisOfRecord* is one of (Observation, Machine Observation, Human Observation, Living Specimen, Occurrence evidence)
- *Country* is Switzerland
- *GadmGid* is CHE
- *TaxonKey* is one of (Animalia, Plantae)
- *Year* 1950-2024

The observations matching one of the following filtering criteria were removed in Python :

- *coordinateUncertaintyInMeters* is larger than $100m$
- *coordinateUncertaintyInMeters* is None
- *species* is None
- *issue flags* is COORDINATE ROUNDED
- The species does not have an article in Wikipedia with a habitat section, based on the species binomial name.

We additionally removed duplicates (multiple records of a species from a particular location).

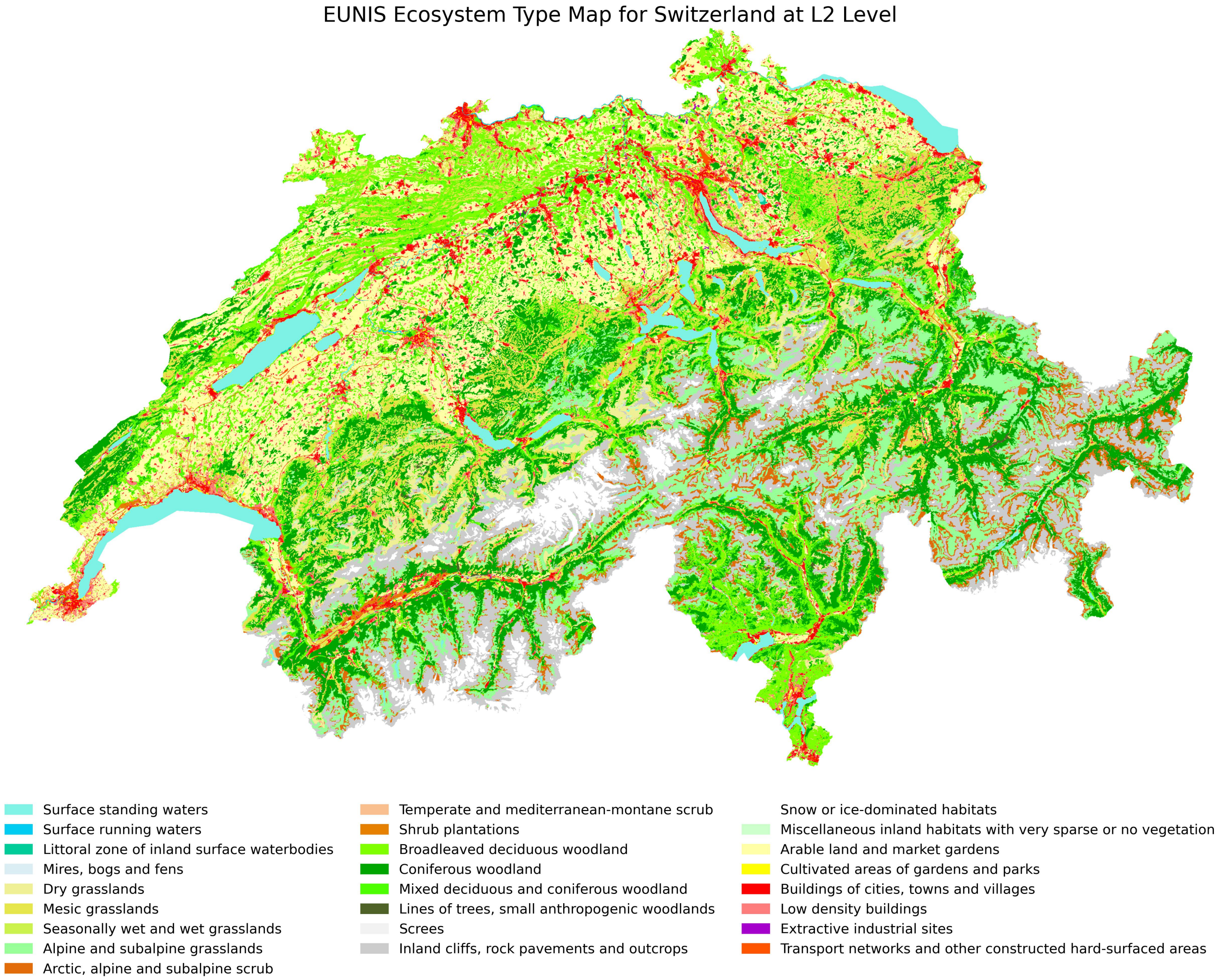

EUNIS Ecosystem Type Map for Switzerland at L2 Level
Surface standing waters
Surface running waters
Littoral zone of inland surface waterbodies
Mires, bogs and fens
Dry grasslands
Mesic grasslands
Seasonally wet and wet grasslands
Alpine and subalpine grasslands
Arctic, alpine and subalpine scrub
Temperate and mediterranean-montane scrub
Shrub plantations
Broadleaved deciduous woodland
Coniferous woodland
Mixed deciduous and coniferous woodland
Lines of trees, small anthropogenic woodlands
Screes
Inland cliffs, rock pavements and outcrops
Snow or ice-dominated habitats
Miscellaneous inland habitats with very sparse or no vegetation
Arable land and market gardens
Cultivated areas of gardens and parks
Buildings of cities, towns and villages
Low density buildings
Extractive industrial sites
Transport networks and other constructed hard-surfaced areas

Figure 6. The EUNIS ecosystem type map in Switzerland at level L2.

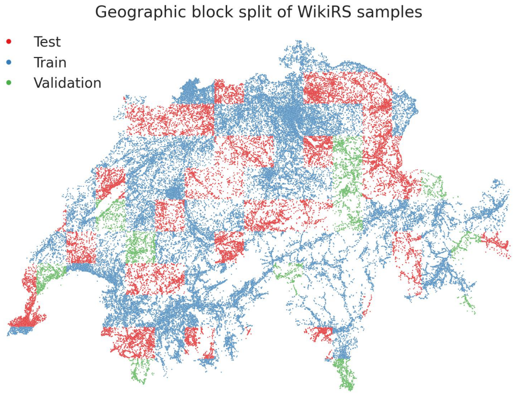


Figure 7. Distribution of our training samples across Switzerland between training, testing and validation sets following a block split approach (with a size of 20 km).

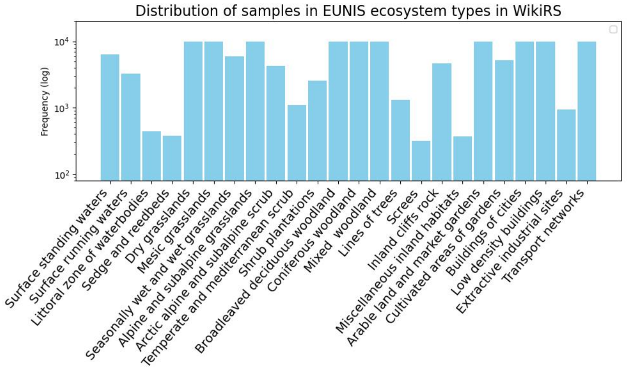
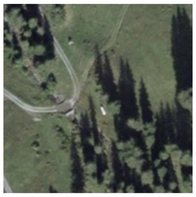
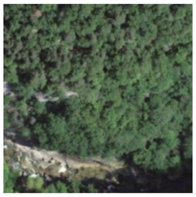


Figure 8. Distribution into EUNIS ecosystem types of samples from the WikiRS dataset on a log scale



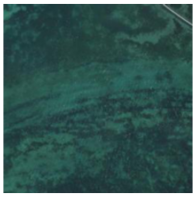
Score	Pretrained SkyCLIP model
0.233	It is becoming rarer, particularly in the north of its distribution, largely due to increasingly intensive agriculture and commercial wildcrafting
0.220	It is rare overall, but may be locally abundant.
0.218	Nevertheless, it is cultivated on a large scale in Estonia.
0.213	Arnica montana grows in nutrient-poor siliceous meadows or clay soils.
0.211	Arnica montana is widespread across most of Europe.

Score	Fine-tuned SkyCLIP model
0.657	Arnica montana grows in nutrient-poor siliceous meadows or clay soils.
0.527	Arnica montana is widespread across most of Europe.
0.033	However Arnica does not grow on lime soil, thus it is an extremely reliable bioindicator for nutrient poor and acidic soils.
-0.105	It is becoming rarer, particularly in the north of its distribution, largely due to increasingly intensive agriculture and commercial wildcrafting .
-0.130	It is rare overall, but may be locally abundant.



Score	Pretrained SkyCLIP model
0.265	This species was once abundant, when forests were used for grazing livestock and trees were coppiced, but is now threatened by overgrowth of larger plants.
0.25	Bjerkandera adusta, Phlebia acerina, Sebacinaceae, Tetracladium sp., and Tomentella sp. Cephelanthera longifolia is vulnerable to grazing by deer.
0.248	Cephalanthera longifolia is common in some parts of its European range, such as southern France and Spain, but endangered particularly in northern areas such as Belgium.
0.237	In 2007 it was listed as a priority species under the UK Biodiversity Action Plan.
0.231	Another investigation indicated 9 mycorrhizal partners .

Score	Fine-tuned SkyCLIP model
0.803	Sword-leaved helleborine usually grows in damp woodland places , forest edges and rocky slopes.
0.639	Cephalanthera longifolia is common in some parts of its European range, such as southern France and Spain, but endangered particularly in northern areas such as Belgium.
0.549	Bjerkandera adusta, Phlebia acerina, Sebacinaceae, Tetracladium sp., and Tomentella sp. Cephelanthera longifolia is vulnerable to grazing by deer.
0.309	This species was once abundant, when forests were used for grazing livestock and trees were coppiced, but is now threatened by overgrowth of larger plants.
0.051	In 2007 it was listed as a priority species under the UK Biodiversity Action Plan.



Score	Pretrained SkyCLIP model
0.250	These dense concentrations of birds are thought to be a defense against attacks by birds of prey such as peregrine falcons or Eurasian sparrowhawks.
0.240	Many birds remain in the same area all year round, but others migrate to spend the winter in mild areas of western Europe or head south as far as Senegal, Gambia and the Red Sea.
0.237	Common starlings in the south and west of Europe and south of latitude 40N are mainly resident. winter in the southwest of the US.
0.236	The Eurasian coot is much less secretive than most of the rail family, and can be seen swimming on open water or walking across waterside grasslands.
0.235	It bobs its head as it swims, and makes short dives from a little jump.

Score	Fine-tuned SkyCLIP model
0.896	It bobs its head as it swims, and makes short dives from a little jump.
0.682	The Eurasian coot is much less secretive than most of the rail family, and can be seen swimming on open water or walking across waterside grasslands.
0.588	In Europe, there are colonies all along the Mediterranean coast, and also on the Atlantic islands and coasts north to Brittany and west to the Azores.
0.400	The goosander is one of the species to which the Agreement on the Conservation of African-Eurasian Migratory Waterbirds applies.
0.318	These dense concentrations of birds are thought to be a defense against attacks by birds of prey such as peregrine falcons or Eurasian sparrowhawks.

Figure 9. Comparison of top-5 sentences scores given by the pretrained and fine-tuned SkyCLIP model on samples of the WikiRS dataset. Scores are cosine similarity values between the text and images features. Scores are ranked by decreasing order of magnitude.

	Observed species : Saxifraga aizoides • It prefers cold and moist well-draining neutral to basic bedrock, gravel, sand, or shale cliff environments. • It is found in North America, including Alaska, across Canada, the Great Lakes region, and Greenland, and in Europe, including the Tatra Mountains, Alps, and Svalbard. • It is a listed threatened species in New York state.
	Observed species : Capra ibex • Alpine ibex are now found in most or all of the Italian and French alpine ranges, southern Germany, Switzerland and Austria. • Alpine ibex are typically absent from woodland areas, Males use lowland meadows during the spring, which is when snow melts and green grass appears. • Alpine ibex tend to live in steep, rough terrain near the snow line. • ...
	Observed species : Sedum acre • It is specially adapted for growing on thin dry soils and can be found on shingle, beaches, drystone walls, dry banks, seashore rocks, roadside verges, wasteland and in sandy meadows near the sea. • Biting stonecrop is a tufted evergreen perennial that forms mat-like stands some tall. • Biting stonecrop spreads when allowed to do so, but is easily controlled, being shallow-rooted. • It is used in hanging baskets and container gardens, as a trailing accent, in borders, or as groundcover. • This plant grows as a creeping ground cover, often in dry sandy soil, but also in the cracks of masonry. • It grows well in poor soils, sand, rock gardens, and rich garden soil, under a variety of light levels. • Biting stonecrop is said to have a peppery taste (hence the name ”biting”) and is sometimes used in herbal medicine. • However, it is considered to be poisonous and consumption is discouraged.
	Observed species : Cardamine heptaphylla • This species is widespread in Central and Southern Europe, from Northern Spain, to Italy and S.W. Germany. • This species grows mainly in mountain woods, especially in beech and spruce forests, but sometimes in plain, at an elevation up to 2,200 metres (7,200 ft) above sea level. • It prefers calcareous soils.
	Observed species : Fulica atra • The coot breeds across much of the Old World on freshwater lakes and ponds, and like its relative the common moorhen, has adapted well to living in urban environments, often being found in parks and gardens with access to water. • It occurs and breeds in Europe, Asia, Australia, and Africa. • The Eurasian coot is much less secretive than most of the rail family, and can be seen swimming on open water or walking across waterside grasslands. • It is an aggressive species, and strongly territorial during the breeding season, and both parents are involved in territorial defence. • They form large flocks on open water in winter. • They are also found on coastal lagoons, shorelines and sheltered ponds. • ...
	Observed species: Fringilla coelebs, Sambucus nigra, Turdus merula • The common chaffinch breeds in wooded areas where the July isotherm is between between 12 and 30 C. • The breeding range includes northwestern Africa and most of Europe and extends eastwards across temperate Asia to the Angara River and the southern end of Lake Baikal in Siberia. • Hedges, waste-ground roadsides, and woods are the typical habitats for the species. • S. nigra is recorded as very common in Ireland in hedges as scrub in woods. • ...

Table 7. Image-text pairs taken taken from our dataset with habitat sentences.