

Leverage Samples with Single Positive Labels to Train CNN-based Models For Multi-label Plant Species Prediction

Notebook for the LifeCLEF Lab at CLEF 2023

Huy Quang Ung^{1,*}, Ryoichi Kojima¹ and Shinya Wada¹

¹KDDI Research, Inc., Fujimino, Saitama, Japan

Abstract

Understanding the geographical distribution of plant species is useful in many scenarios related to biodiversity management and conservation. By associating plant species occurrences with environmental features of each location, we can model the relationship between an environment and the species. However, the cost of multi-label plant species annotation for a large dataset is expensive and time consuming, so it may only be possible to obtain a single positive label for each location. This type of dataset is provided in the GeoLifeCLEF 2023 competition, where learning multi-labels from single positive labels is the main challenge. In this report, we present our proposed models that achieved the best performance in GeoLifeCLEF 2023. We proposed several CNN-based models and a training strategy for learning samples with single positive labels. We conducted experiments to show the effectiveness of our method compared to a simple baseline on the provided dataset.

Keywords

Single positive label, CNNs, three-step training strategy, late fusion

1. Introduction

Modelling species distribution is an essential task for monitoring and making conservation decisions for a wide variety of species. Nowadays, the research community has generated millions of geolocated species observations every year, covering tens of thousands of species, which is a good opportunity to apply machine learning and deep learning based models. A common approach is to build a species distribution model (SDM) [1], which uses the environmental variables of the location (e.g. temperature, elevation, land cover, soil, etc.) to predict the presence of species at that location.

Following the ongoing series of the GeoLifeCLEF competitions [2, 3, 4, 5, 6], the GeoLifeCLEF 2023 [7], which is a part of LifeCLEF 2023 [8], aims to predict the presence of plant species at a given location and their change at different timestamps, providing a large scale dataset of raster and time series based variables. The goal of GeoLifeCLEF 2023 is to predict multi-label plant species for each location, while the last GeoLifeCLEF 2022 is to predict only a single

CLEF 2023: Conference and Labs of the Evaluation Forum, September 18–21, 2023, Thessaloniki, Greece


*Corresponding author.

✉ xhu-ung@kddi.com (H. Q. Ung); ry-kojima@kddi.com (R. Kojima); sh-wada@kddi.com (S. Wada)

🆔 0000-0001-9238-8601 (H. Q. Ung); 0009-0009-3128-7781 (R. Kojima); 0000-0002-9421-8566 (S. Wada)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

label species. Furthermore, the main challenge of GeoLifeCLEF 2023 is that 98 percentages of the samples provided have only single positive labels. Other difficulties include the long-tail distribution of plant species, large-scale multi-modal learning, and a variety of plant species classes.

Convolutional Neural Networks (CNNs) have achieved great performance in computer vision in recent years. In the GeoLifeCLEF 2022 competition, several CNN-based SDM models have been proposed for learning raster-based variables. However, it is difficult to train CNN-based models by samples with single positive labels [9, 10, 11] for multi-label prediction task. Single positive labels for multi-label prediction task are considered as noisy labels due to lack of other positive labels [9].

In this technical report, we present several CNN-based models for multi-label plant species prediction on the provided dataset of GeoLifeCLEF 2023. In addition, we introduce an efficient three-step training strategy for leveraging samples with single positive labels to train our proposed CNN-based model. We also present detailed experiments of our proposed method.

The remaining of this report is organized as follows. Section 2 presents related work of this report. Section 3 describes the provided dataset and our main task in details. Section 4 describes our preprocessing steps. Section 5 introduces our proposed method. Section 6 presents detailed experiments and results. Finally, section 7 concludes this work.

2. Background and related work

2.1. Multi-label learning from single positive labels

To the best of our knowledge, there has been no study of learning multi-label prediction from single positive labels for plant species prediction. However, Cole et al. [9] pointed out this problem for the field of computer vision and proposed a method for estimating unobserved labels. They also considered possible methods such as label smoothing to reduce the negative impacts of negative labels assumed. Zhou et al. [10] introduced a pseudo-labelling method for labelling unobserved positive labels and applied Expectation Maximization loss in their training phase. Xie et al. [11] also proposed a pseudo-labelling method and a regularization term to address this problem. Although these methods showed their effectiveness on learning multi-labels from single positive labels, those were only experimented on benchmarks with less than 200 classes.

2.2. Binary cross-entropy loss

First, let define a setting for a multi-label prediction. We assume that each input x from \mathcal{X} corresponds to a vector label y from the label space $Y = \{y_i\}_{i \in [1, L]} \in \{0, 1\}^L$, where L is the number of classes (i.e. species), an entry $y_i = 1$ if the i -th class is relevant to x (i.e. the species i is present at the location encoded by x) and $y_i = 0$ otherwise (i.e. the species is absent). The main objective is to find a function $f : \mathcal{X} \rightarrow \mathcal{Y}$ that predicts the labels for each x .

The binary cross-entropy (BCE) loss is one of the most common loss for multi-label learning [12]. For an observed data point (x_n, y_n) including full positive and negative classes, the

BCE loss is calculated as follows:

$$\mathcal{L}_{BCE}(f_n, y_n) = -\frac{1}{L} \sum_{i=1}^L [\mathbf{1}_{[y_{ni}=1]} \log(f_{ni}) + \mathbf{1}_{[y_{ni}=0]} \log(1 - f_{ni})] \quad (1)$$

where $f_n = f(x_n) \in [0, 1]$ is the model predicted probability of presence for species i under input x_n , and $\mathbf{1}_{[\cdot]}$ denotes the indicator function, i.e., $\mathbf{1}_k = 1$ if the assertion k is verified, or 0 otherwise.

2.3. Assume negative loss

Suppose that we have partially observed data (x_n, z_n) , the label space is $\mathcal{Z} = \{z_i\}_{i=1, \overline{L}} = \{1, 0, \emptyset\}^L$ and suppose that all of the observed labels are positive i.e., if z_{ni} is known, then $z_{ni} = 1$. Here, we can formulate a loss function with the positive term of $\mathbf{1}_{[y_{ni}=1]} \log(f_{ni})$. For unobserved labels, we cannot simply ignore them in the loss function since it could cause that the trained model always results positive classes. The simple approach is to assume unobserved labels are negative. This ‘‘assume negative’’ (AN) loss is calculated as follows:

$$\mathcal{L}_{AN}(f_n, z_n) = -\frac{1}{L} \sum_{i=1}^L [\mathbf{1}_{[z_{ni}=1]} \log(f_{ni}) + \mathbf{1}_{[z_{ni} \neq 1]} \log(1 - f_{ni})] \quad (2)$$

3. Task and dataset

The final goal of GeoLifeCLEF 2023 is to predict the set of plant species presence in a given location and time using various related features: images and time-series data captured from a satellite, time-series climatic data, and other rasterized environmental data: land cover, human footprint, bioclimatic and soil variables. In this technical report, we present raster-based variables used in our method, i.e., satellite raster images, aggregated human footprint rasters, bioclimatic rasters, and soil-grid rasters. Other variables are described in detail at this competition’s homepage¹.

- Satellite raster images: There are RGB and Near Infra-Red (NIR) captured over a square area of 1280 meter \times 1280 meter. They are formatted in a size of 128 \times 128 pixels. An example is shown in Figure 1.
- Summarized human footprint rasters: A aggregated version of several low-resolution rasters contain seven pressures of the environment which indicate the presence of human being and their activities. The resolution is 1 kilometer per pixel. These rasters were collected in 1993 and 2009. We used the data from 2009 in this report.
- Bioclimatic rasters: They are low-resolution rasters of 19 variables classically used in species distribution modeling. Their resolution is around 1 kilometer per pixel.
- Soil-grid rasters: These consist of low-resolution pedological rasters of nine variables related to soil properties, i.e., pH, clay content, organic carbon, nitrogen, bulk density, sand,

¹<https://www.kaggle.com/competitions/geolifeclef-2023-lifeclef-2023-x-fgvc10/data>

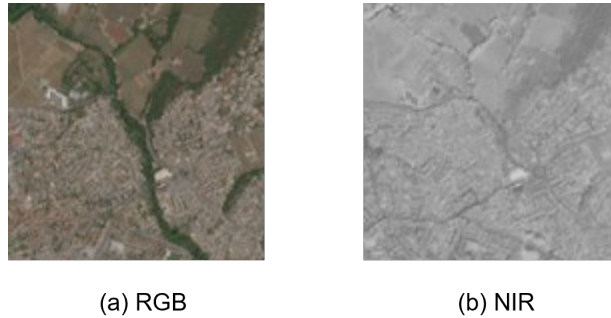


Figure 1: An example of RGB and NIR at (Latitude=43.153, Longitude=6.080).

silt, cation exchange capacity, and coarse fragments. These properties were measured from 5 to 15-centimeter depth. Their resolution is 1 kilometer per pixel.

This competition provides a large-scale training set of about 5 million samples of plant occurrences in Europe. This training set, where only a single positive class is labeled, is so-called presence-only data (PO in short). A validation set consists of 5,948 samples with all the present species (multi-labels of both positive and negative classes), which is so-called presence-absence data (PA in short). Here, we used a part of the PA data to train our models. A testing set consists of 22,404 samples. The testing set is implicitly divided into a public set and a private set for evaluation. The total number of plant species is 10,039 classes.

4. Preprocessing data

For preprocessing data, we used the source code² provided by the organizers. The raster-based variables, which consist of bioclimatic rasters, summarized human footprint rasters, and soi-grid rasters are formed into the size of 128×128 pixels (the same as the sizes of satellite raster images). The resolutions of them are the same as the above-mentioned. Each satellite raster image and raster-based variable are applied the standard normalization before inputting to our models, where their average and standard deviation values are calculated on training samples.

5. Proposed method

This section presents the architectures of our proposed models, a method for combining them, and an efficient training strategy.

5.1. Model architectures

To address this multi-label plant species prediction, we experimented with three CNN-based models with the ResNet [13] backbone, i.e., BioResNet50, FusResNet34, and FusResNet50. The overview of three proposed models is shown in Figure 2.

²<https://github.com/plantnet/GLC>

The BioResNet50 model with the ResNet50 backbone receives bioclimatic rasters as the input. We use available 19 channels of bioclimatic rasters. The FusResNet34 model with three branches of the ResNet34 backbone is a multi-modal late fusion network, which combines bioclimatic rasters (19 channels), satellite imagery (3-channel RGB and 1-channel NIR), a human footprint raster of summary version (1 channel), and soil rasters (9 channels). In FusResNet34, we change the ResNet34 backbone to the ResNet50 one and obtain the FusResNet50 model. Those models are trained in an end-to-end manner.

We simply combine three proposed models by calculating the average output of them. Figure 3 illustrates our ensemble method.

5.2. Three-step training strategy

Training our CNN-based models simultaneously using both the PO and PA data (shown in Figure 4(a)) is not an effective method due to the negative impact of assuming negative labels of PO (shown in our experiment). We propose an efficient training strategy that can improve the performance using the samples with single positive labels (the PO data). Our method is a three-step training strategy as shown in Figure 4(b).

First, we train our models on the PA data only, using the BCE loss, since PA has fully observed labels. This is a kind of warming-up step for multi-label prediction. Secondly, we continue to train the models obtained in the first step using only the PA data with the cross-entropy (CE) loss [14]. The CE loss is one of the most common losses for multi-class classification tasks, where the objective of this task is to obtain a single class per an input sample. In this step, we expect that our models can learn specific features for each class since PO has only single positive labels. Finally, we continue training the models obtained in the second step using only the PA data with the BCE loss. Here, the models were able to adjust their activation units for the multi-label prediction and utilize the specific features learned in the second step.

6. Experimental results

This section presents experiment settings and shows the effectiveness of our models and training strategy. In addition, we perform an ablation study for our proposed training strategy.

6.1. Experimental settings

We divided the training and validation set as described in Table 1. Our baseline model was to train a CNN-based model simultaneously using both PO and PA with the AN loss as described in Figure 4(a).

We implemented our models using the PyTorch [15] framework. The detailed settings of our proposed models and the baseline are described in Table 2. The pre-trained weights of ResNet-34 and ResNet-50 on ImageNet [13] were used to initialize the backbones of our models in the training phase of Step 1 and our baseline case. Due to resources limitation and the time-consuming of the training phase, we only trained our models by 10 and 20 epochs in Step 2 and the baseline, respectively. In steps 1 and 3, we stopped the training phase at epoch 30th due to time constraints, while the validation loss was slightly improved. However, the results could

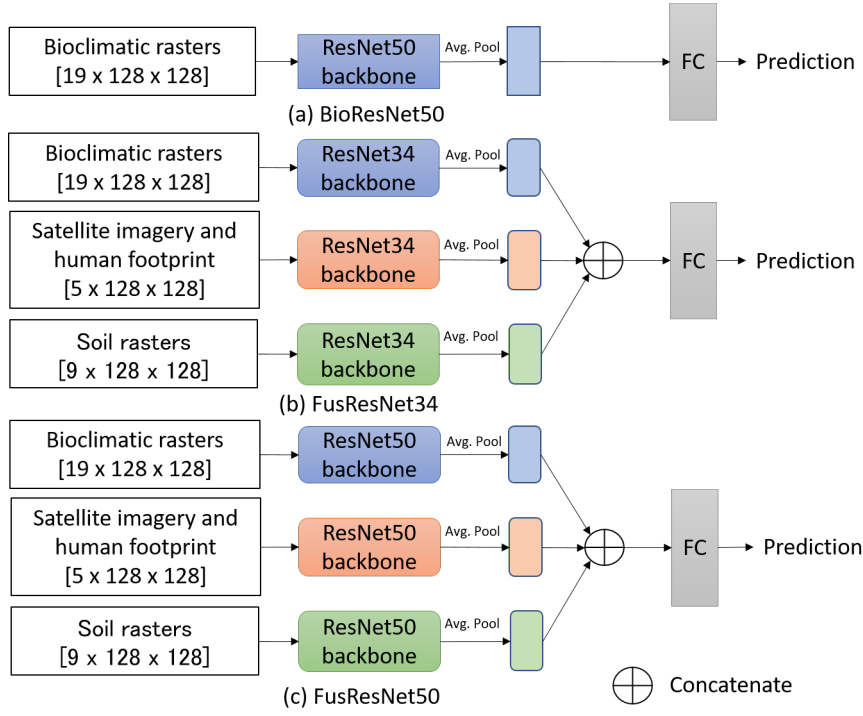


Figure 2: Overview of proposed models' architectures.

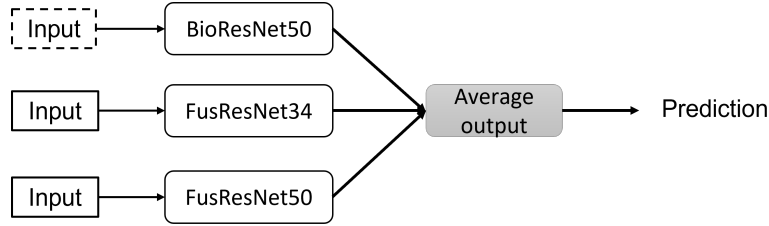


Figure 3: Ensemble method.

be improved if we continue training the models in further epochs. In the inference step, given an input, the models will output the top-20 species with the highest probabilities for evaluation.

The GeoLifeCLEF 2023 competition used the micro F1-score (\uparrow) for evaluation. The micro F1-score (denoted as MF1-score) is calculated as follows:

$$F_1 = \frac{1}{N} \sum_{j=1}^N \frac{TP_j}{TP_j + (FP_j + FN_j)/2} \quad (3)$$

where TP_j , FP_j , and FN_j are the true positive, the false positive, and the false negative of the j -th input sample, respectively. N is the number of samples for evaluation.

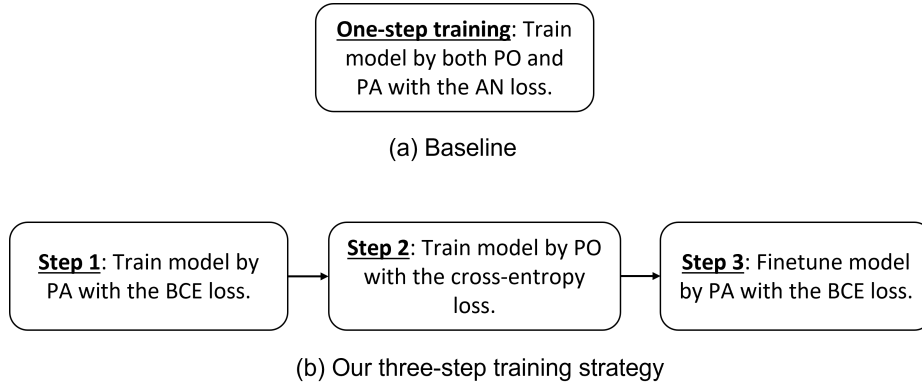


Figure 4: Three-step training strategy.

Table 1

Settings for training and validation sets in training phases.

Subset	Step 1&3		Step 2		Baseline	
	PO	PA	PO	PA	PO	PA
Training set	-	80%	98%	-	100%	80%
Validation set	-	20%	2%	-	0%	20%

Table 2

Hyper-parameters of our proposed models and the baseline model.

Hyper-parameters	Step 1&3	Step 2	Baseline
Batch size	128	96	96
Optimizer	Adam [13]	Adam	Adam
Learning rate (Lr)	0.001	0.003	0.003
Lr scheduler	MultiStep, epoch 15 and 20	-	MultiStep, epoch 15
Lr decay rate	0.1	-	0.1
Maximum #epochs	30	10	20

6.2. Comparison among proposed models

This section presents a comparison among our proposed models and the baseline method which is simultaneously trained on PO and PA using the AN loss. We only implemented the baseline BioResNet50 (denoted as BioResNet50-base) to compare with our BioResNet50 trained by our training strategy. In addition, we tried to train the FusResNet50 in step 2 by 20 epochs (denoted as FusResNet50*) to observe the performance.

Table 3 shows our experimental results. Our BioResNet50 significantly outperforms BioResNet50-base, indicating the effectiveness of our three-step training strategy. Among our proposed models, the multi-modal FusResNet34 and FusResNet50 models achieve better performance than the BioResNet50. The MF1-score values of FusResNet50* are slightly lower

Table 3

Performance of proposed models on the testing set.

Models	One-step training	#Epochs in step 2	MF1-score ↑	
			Public	Private
BioResNet50-base	✓	-	0.060	0.058
BioResNet50	-	10	0.243	0.239
FusResNet34	-	10	0.254	0.249
FusResNet50	-	10	0.254	0.249
FusResNet50*	-	20	0.248	0.242
Ensemble method: BioResNet50 + FusResNet34 + FusResNet50	-	-	0.276	0.270

Table 4

Ablation study of FusResNet50 on the training strategy.

Step 1	Step 2	Step 3	MF1-score ↑	
			Public	Private
✓	-	-	0.200	0.196
-	✓	-	0.059	0.058
✓	✓	-	0.073	0.073
-	✓	✓	0.226	0.221
✓	✓	✓	0.254	0.249

than those of FusResNet50, indicating that further training of the model in step 2 could not improve the performance. The combination of BioResNet50, FusResNet34, and FusResNet50 by the ensemble method achieves the best performance of 0.276 and 0.270 on the public and private testing sets, respectively.

6.3. Ablation study for our proposed training strategy

We conducted an ablation study for our proposed training strategy using the FusResNet50 model. Table 4 presents the detailed results. Overall, applying the three-step training strategy achieves the best performance of around 0.25 on both public and private testing sets. The performance of step 1 alone is around 0.2 on these two testing sets. Without step 3, step 2 alone and the combination of step 1 and step 2 achieve significantly lower performance. Without step 1, the performance of step 2 and step 3 alone is lower 0.03 MF1-core points than that of the combination of these three steps.

7. Conclusion

This technical report presents working notes on the GeoLifeCLEF 2023 competition. For multi-label plant species prediction, we presented our proposed CNN-based models, i.e., BioResNet50,

FusResNet34, and FusResNet50. In addition, we presented a three-step training strategy to improve the prediction performance from learning samples with single positive labels. Our experiments show that FusResNet34 and FusResNet50 achieved the best and comparable performance of around 0.25 on both public and private testing sets. Furthermore, we have also shown the effectiveness of our three-step training strategy.

References

- [1] J. Elith, J. R. Leathwick, Species distribution models: ecological explanation and prediction across space and time, *Annual review of ecology, evolution, and systematics* 40 (2009) 677–697.
- [2] C. Botella, P. Bonnet, F. Munoz, P. P. Monestiez, A. Joly, Overview of geolifeclef 2018: location-based species recommendation, in: *Working Notes of CLEF 2018-Conference and Labs of the Evaluation Forum*, volume 2125, 2018.
- [3] C. Botella, M. Servajean, P. Bonnet, A. Joly, Overview of geolifeclef 2019: plant species prediction using environment and animal occurrences, in: *CLEF 2019 Working Notes-Conference and Labs of the Evaluation Forum*, volume 2380, 2019.
- [4] B. Deneu, T. Lorieul, E. Cole, M. Servajean, C. Botella, P. Bonnet, A. Joly, Overview of lifeclef location-based species prediction task 2020 (geolifeclef), 2020.
- [5] T. Lorieul, E. Cole, B. Deneu, M. Servajean, P. Bonnet, A. Joly, Overview of geolifeclef 2021: Predicting species distribution from 2 million remote sensing images, in: *CLEF 2021-Conference and Labs of the Evaluation Forum*, volume 2936, 2021, pp. 1451–1462.
- [6] T. Lorieul, E. Cole, B. Deneu, M. Servajean, P. Bonnet, A. Joly, Overview of geolifeclef 2022: Predicting species presence from multi-modal remote sensing, bioclimatic and pedologic data, in: *CLEF 2022-Conference and Labs of the Evaluation Forum*, volume 3180, 2022, pp. 1940–1956.
- [7] C. Botella, B. Deneu, J. Estopinan, M. Servajean, D. Marcos, A. Joly, Overview of GeoLife-CLEF 2023: Species presence prediction based on occurrences data and high-resolution remote sensing images, in: *Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum*, 2023.
- [8] A. Joly, C. Botella, L. Picek, S. Kahl, H. Goëau, B. Deneu, D. Marcos, J. Estopinan, R. Chamidullin, M. Šulc, M. Hruz, M. Servajean, B. Kellenberger, E. Cole, H. Glotin, et al., Overview of lifeclef 2023: evaluation of ai models for the identification and prediction of birds, plants, snakes and fungi, in: *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 14th International Conference of the CLEF Association, CLEF 2023, Thessaloniki, Greece, September 18–23, 2023, Proceedings*, Springer, 2023.
- [9] E. Cole, O. Mac Aodha, T. Lorieul, P. Perona, D. Morris, N. Jojic, Multi-label learning from single positive labels, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 933–942.
- [10] D. Zhou, P. Chen, Q. Wang, G. Chen, P.-A. Heng, Acknowledging the unknown for multi-label learning with single positive labels, in: *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXIV*, Springer, 2022, pp. 423–440.

- [11] M.-K. Xie, J. Xiao, S.-J. Huang, Label-aware global consistency for multi-label learning with single positive labels, *Advances in Neural Information Processing Systems* 35 (2022) 18430–18441.
- [12] T. Durand, N. Mehrasa, G. Mori, Learning a deep convnet for multi-label classification with partial labels, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 647–657.
- [13] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [14] I. J. Good, Rational decisions, *Journal of the Royal Statistical Society: Series B (Methodological)* 14 (1952) 107–114.
- [15] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., Pytorch: An imperative style, high-performance deep learning library, *Advances in neural information processing systems* 32 (2019).