Larval Hostplant Prediction from *Luehdorfia japonica* Image using Multi-label ABN

Tsubasa Hirakawa[1⋆], Takaaki Arai[1], Takayoshi Yamashita[1], Hironobu Fujiyoshi[1], Yuichi Oba[1], Hiromichi Fukui[1], and Masaya Yago[2⋆]

[1] Chubu University,
1200, Matsumoto-cho, Kasugai-shi, Aichi 487–8501, Japan
[2] The University Museum, The University of Tokyo,
7–3–1, Hongo, Bunkyo-ku, Tokyo 113–0033, Japan

Abstract. Butterflies are easily recognizable due to their showy coloration, and are a familiar taxon to many enthusiasts. Because of the abundance of specimens collected and the ease of comparison among various species, regional variation in butterfly spots can be observed, which is related to multiple factors such as geological history, topography, climate, and hostplants. In this study, we focus on the relationship between regional variation in butterfly spots and the distribution of larval hostplants and aim to clarify the relationship by classifying the larval hostplants based on images of butterfly spots. Specifically, we focus on the *Luehdorfia japonica*, a species of butterfly with known geographic variation in butterfly spots and a highly understood distribution. We create *Luehdorfia japonica* image dataset based on digital specimens and the metadata about the collection site. We show that the multi-label attention branch network can be trained on the dataset to accurately classify the larval hostplant from the specimen images and that the analysis of the attention map provides the same basis for decision making as the expert knowledge.

Keywords: Attention branch network · Multi-label image classification · Visual explanation · *Luehdorfia japonica* · Regional variation

1 Introduction

Butterflies are highly visible and one of the most familiar species. Because of the abundance of specimens collected, the distribution of their habitat is highly well understood [11]. And, because of the ease of comparison among specimens, many species are known to show regional variation (geographic variation) in their spots. Moreover, due to the high degree of distribution elucidation, it is easy to identify regional extinctions and the timing of extinctions and declines. Since various factors such as geographical history, topography, environment, and larval hostplants, plants eaten during the larval stage, are considered to be related to

⋆ Corresponding authors: hirakawa@mprg.cs.chubu.ac.jp (Tsubasa Hirakawa) and myago@um.u-tokyo.ac.jp (Masaya Yago)

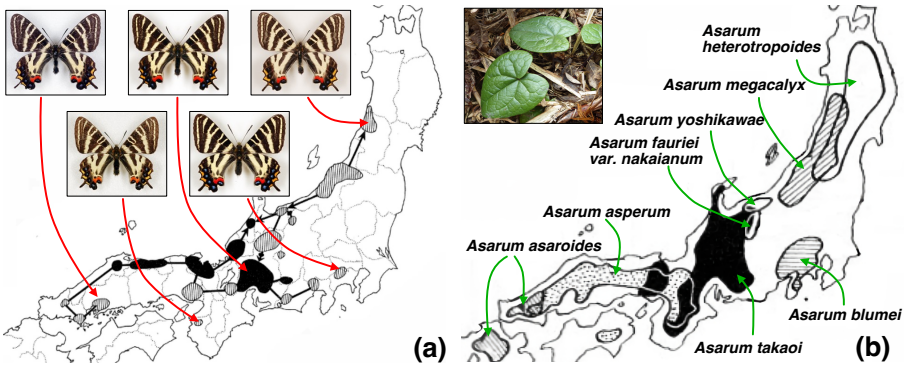


Fig. 1: Overview of the relationship between wing spots and the distribution of the *Asarum*. (a) Schematic illustration of populations based on regional variation in wing shape and spots. The blacker the population, the more developed the black bands of the spots. (b) Conceptual map of the distribution of the *Asarum* used by the *Luehdorfia japonica*. Modified from reference [33], respectively.

regional variation, we expect to elucidate various problems and factors related to butterflies by correlating butterfly specimen data with various data.

Among various species of butterflies, in this study, we focus on *Luehdorfia japonica*, whose distribution is highly known due to its popularity among butterfly lovers. The wing spots of *Luehdorfia japonica* are well known to vary from region to region. As shown in Figure 1, *Luehdorfia japonica* is thought to correspond to its regional variation in the distribution of *Asarum* [11, 24], which is a larval hostplant and has significant speciation.

In this paper, we aim to clarify the relationship between the regional variation of the butterfly spots and their larval hostplants, based on images of *Luehdorfia japonica* specimens and information on the distribution of *Asarum* in each habitat area. Specifically, a *Luehdorfia japonica* image dataset was created from the Tomoo Fujioka Butterfly Collection (Fujioka Collection) and information on the distribution of larval hostplants. We will then analyze the regional variation in the spotted pattern through visualization of the larval hostplant classification and its basis for judgment from the specimen images[3]. To create the dataset, digital specimen images are preprocessed and image samples are collected. At the same time, each image sample is annotated labels for the hostplants used during the larval stage based on the collection site information. In some cases, the hostplant label is not a single label, but multiple labels are assigned depending on the habitat. The dataset is trained and classified using the attention branch network (ABN) [7] to improve the accuracy of visual description and image classification tasks.

[3] The technical report versions of this paper are presented in [35, 36]. In this paper, we construct a larger *Luehdorfia japonica* dataset and report the results of more detailed experiments and their analysis results.

The number of samples in the created dataset is highly dependent on the type of specimens present and the location of the collection. The number of *Luehdorfia japonica* in each collection site and the number of individuals collected varies widely, resulting in a dataset with potential class imbalance. It is known that the recognition accuracy of a dataset with class imbalance decreases in a class with a small number of samples. In this study, we introduce error functions and data sampling techniques that take class imbalance into account to reduce class imbalance. This will not only improve classification accuracy but also confirm changes in the visual explanations acquired, aiming to acquire appropriate gazing regions that support the relationship between mottled patterns and larval hostplants.

2 Related Work

2.1 Application of Computer Vision to Biology

Several biological analyses have been performed by applying computer vision techniques [3,6,20,21,34,38]. Cuthill et al. [3] quantitatively analyzed Müllerian mimicry in butterflies by utilizing the distance between samples in the embedding space obtained by triplet network [28] and categorical cross-entropy loss.

Fan et al. [6] constructed and classified a dataset of 80 swallowtail butterfly species from Yunnan, China, and Lin et al. [20, 21] proposed an image classification model based on deep learning to identify butterfly subfamilies, genus, and subspecies. Both studies identified the biological classification of butterflies as fine-grained image classification tasks. On the other hand, this study deals with a more detailed image identification task because it deals with the regional variation of the spots in a single species of butterflies, *Luehdorfia japonica*. In addition, we will not only classify but also clarify the relationship between larval hostplants and spots using visual explanations (attention maps) obtained in the inference process of the classification model.

2.2 Multi-label Image Classification

This study deals with multi-label image classification tasks. For multi-label image classification, learning methods that take class imbalance into account have been proposed [9, 10, 32]. Major approaches include devising loss functions [14, 25, 27] and sampling methods during training [1, 2, 22, 23]. Since the datasets in this study potentially have a class imbalance problem, we train them considering the class imbalance to improve the classification accuracy.

A typical multi-label classification task in the computer vision field is pedestrian attribute recognition, which estimates attribute information such as gender and clothing from pedestrian images [12,13,26,27,31,37]. For the pedestrian attribute recognition task, approaches that consider class imbalance in multi-label estimation [13,27] and inference models and learning methods that consider the consistency of the gazing region based on the spatial characteristics of the attributes to be estimated [12,26,31,37] have been proposed. Jia et al. [12] propose

a regularization such that the features are spatially and semantically consistent. Specifically, they propose a loss function to make the attention map of samples in a mini-batch whose attributes are positive consistent, and an error to learn the consistency of feature maps (vectors) compressed by weighted global average pooling.

As described above, many pedestrian attribute recognition methods improve recognition accuracy by incorporating semantic information of attributes to be classified, which is prior knowledge possessed by humans, and location information where features of each attribute are likely to appear. On the other hand, we do not utilize prior knowledge possessed by humans or experts, but extract and analyze characteristic micro-regions and patterns from label information that are useful for classification. The objective is to elucidate the causes of regional variation in geographic regions. Furthermore, we will not only develop a model for classification by focusing on the same areas as the experts' knowledge but also reveal new characteristic patterns of spots that have not been revealed before.

2.3 Visual Explanations

Visual explanation is a method for presenting the basis for judgments about the inference results of machine learning models. Major approaches include post-hoc methods such as GradCAM [29,30] and models that embed a mechanism to generate an attention map inside the model [7,27,39]. In this study, we utilize the attention branch network (ABN) [7], which incorporates visual explanations into inference results through an attention mechanism, to classify and analyze each class label (larval hostplants) while indicating the areas that were gazed at during classification.

3 Dataset

3.1 Tomoo Fujioka Butterfly Collection

Tomoo Fujioka Butterfly Collection (*Fujioka Collection*) is one of the most valuable collections of butterfly specimens in the world, with approximately 290,000 specimens of all species from Japan and related neighboring countries, and approximately 1,750 specimen boxes. To make effective use of this collection, a database and virtual museum are being developed [15,16]. From 2019, a reorganization of the butterfly collection record information in this database in a format compliant with GBIF (Global Biodiversity Information Facility) and S-Net (Science Museum Net) is underway [17]. Thus, work is underway to revise the data of the above collections into internationally accessible natural history specimen data.

The use of this collection database and related data such as vegetation and flora will make it possible to elucidate the regional variation of butterfly spots. In this study, we utilize the Fujioka Collection to perform quantitative larval host-plant classification based on images of the *Luehdorfia japonica*. We will focus on

Table 1: Distribution of labels in *Luehdorfia japonica* dataset.

Label (*Asarum*)	Use in exp.	# of samples	
		Male	Female
A. curvistigma	✓	235	45
A. nipponicum	✓	124	29
A. yoshikawae	✓	55	10
A. megacalyx	✓	341	99
A. nipponicum var. *saninense*	✓	54	13
A. rigescens var. *brachypodion*	✓	91	21
A. asaroides	✓	92	24
A. tohokuense	✓	42	10
A. takaoi	✓	2,075	517
A. caulescens		1	0
A. fauriei		0	1
A. asperum	✓	340	148
A. fauriei var. *nakaianum*		3	1
A. ikegamii	✓	145	30
A. blumei	✓	359	74

the Japanese endemic species of the butterfly, *Luehdorfia japonica*, because of its huge mass and remarkable geographical variation of wing patches in the collection. First, we will examine the labels necessary for image classification and build a dataset by utilizing the image data and label data from the Fujioka Collection database. Hereafter, we describe the details of the dataset construction.

3.2 *Luehdorfia japonica* dataset

Digital specimen imaging and preprocessing To capture digital specimen images, we first remove each specimen from the specimen box and photograph it along with its specimen number, collection value label, size scale, and color chart for correction (see the top row in Figure 2). Multiple images are taken at different depths of field, and then depth composited to produce an image in which all parts of the specimen are in focus. The bottom row in Figure 2 shows an example of a digital specimen of *Luehdorfia japonica* taken by the photographer. The image size of the digital specimen is approximately 6,500×4,000 pixels.

As mentioned above, these digital specimens contain color charts and other parts that are unnecessary for image classification. Therefore, we remove them as a preprocessing. Examples of preprocessed images are shown in Figure 3. All cropped image patches were uniformly sized (see the experimental section for details). The total number of created images is 4,434, of which 3,510 are male and 924 are female.

Assigning hostplant label data In addition to the image preprocessing, we annotate larval hostplant labels to each of the *Luehdorfia japonica* image samples. Metadata such as place of collection, date of collection, and collector were

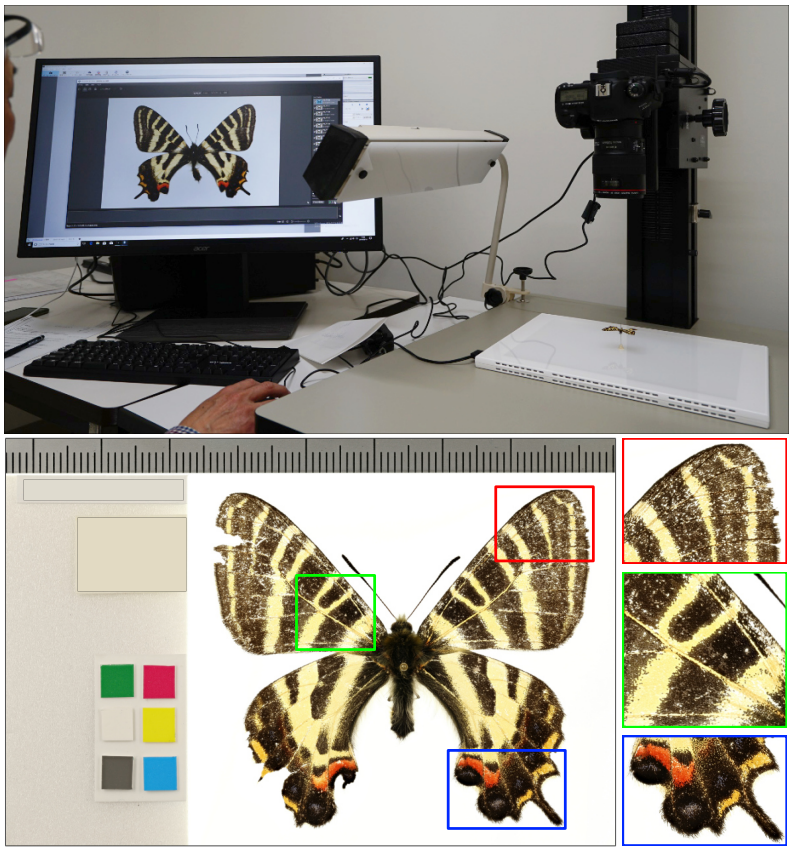


Fig. 2: Photography of the digital specimens of Tomoo Fujioka Butterfly Collection. (Top) Photographing and (Bottom) an example of the digital specimens image data.

assigned to each of the *Luehdorfia japonica* samples. Based on the collection location information and the Japanese vegetation information, we assigned larval hostplant labels to each sample [11, 24].

Table 1 shows a breakdown of the labels in the *Luehdorfia japonica* dataset. Note that the total number of labels does not equal the number of samples, as a single sample may have multiple labels. From the Table 1, there are several labels for which there are not a sufficient number at this time. Therefore, we will limit the number of labels used in the experiment and conduct the experiments.

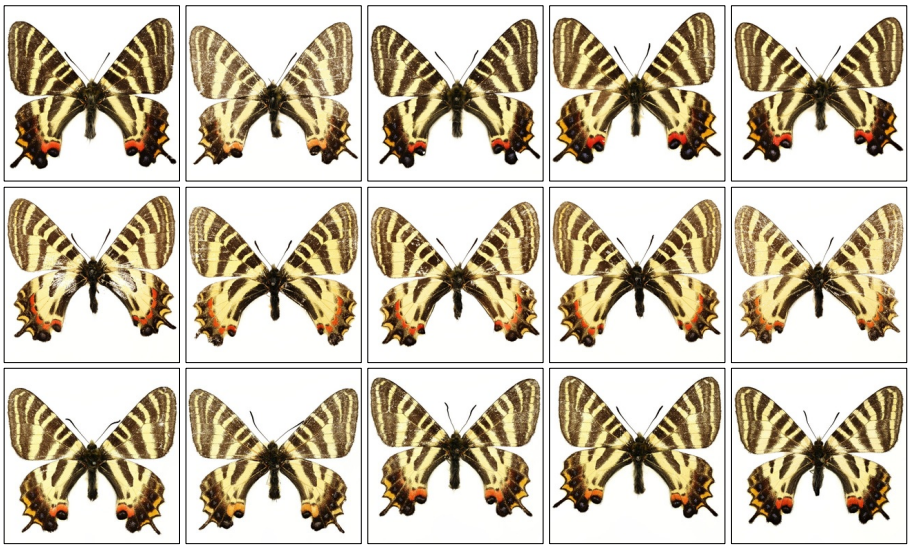


Fig. 3: Examples of pre-processed *Luehdorfia japonica* images.

4 Method

4.1 Network Model

Figure 4 shows the network structure of the multi-label classification ABN. The feature extractor extracts a feature map common to all labels from an input image, and the extracted feature map is input to the attention branch. The attention branch performs global average pooling (GAP) [18] according to each label and obtains the classification result of each label. At the same time, the attention map, which is a feature map obtained from the convolution layer before GAP, is input to the perception branch. After applying weighting (attention mechanism) to the feature maps output from the feature extractor, the final classification result is output. This makes it possible to visualize the regions gazed at during classification, and to perform classification using features that emphasize the features of those regions.

4.2 Model Training Considering Class Imbalance

As described in the previous section, the dataset used in this study has a class imbalance problem. Therefore, if we train a model naively, learning proceeds mainly on labels with a large number of samples, which may bias the classification accuracy. Furthermore, since the objective of this study is not only to achieve high classification accuracy but also to analyze butterfly spots through visualization of the regions gazed at by the model during classification, it is necessary to obtain appropriate classification and reasonable gazing regions. In this

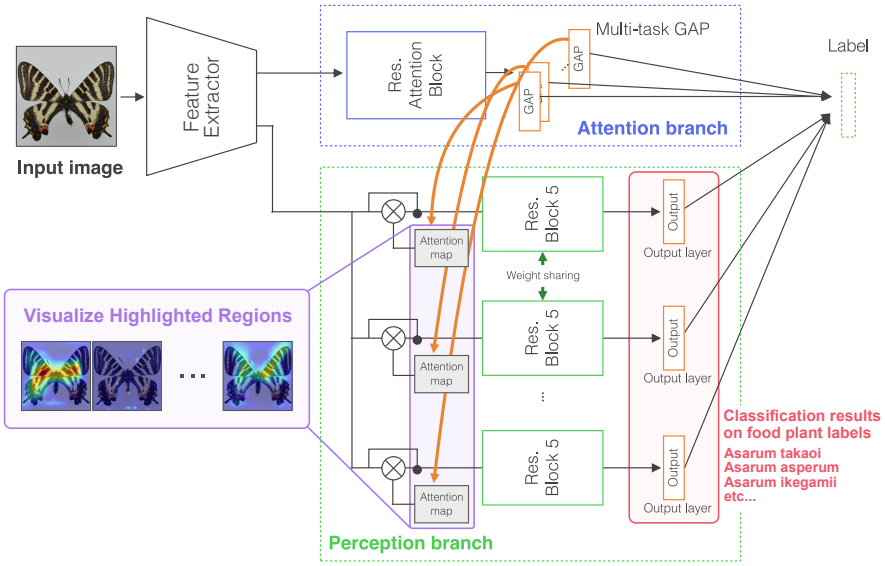


Fig. 4: Network structure of Multi-label classification ABN. The model first extract feature map from an input image via the feature extractor (e.g., ResNet). Then, the extracted feature map is fed into the attention branch, where we further extract feature map and predict classification score passing thrhough global average pooling (GAP). Meanwhile, by using the feature map obtained from the feature extractor and the attention maps by GAP, we compute final prediction results at the perception branch. In addition to the prediction, we can obtain attention map as visual explanation.

study, we experimentally examine changes in accuracy and acquired attention maps by introducing a learning method that takes class imbalance into account. Specifically, we introduce (i) an error function and (ii) data sampling that accounts for class imbalance.

Loss functions The ABN error function L_{abn} is defined as

$$L_{abn} = L_{att} + L_{per}. \tag{1}$$

In other words, in multi-label classification ABN, the errors (L_{att}, L_{per}) between the output of the attention and the perception branches and the correct label is calculated. Conventional multi-label classification ABN uses binary cross-entropy loss (BCE) for each of (L_{att}, L_{per}), but it does not learn enough class labels with insufficient number of samples, which may cause recognition accuracy degradation. In this study, we use the following error functions that take class imbalance into account.

One is a Weighted Focal Loss (WFL) [27]. WFL is a weighted error function based on the prior distribution of labels in the focal loss [19]. Let x be the

logit for a class c and y^c be the positive solution label. Also, when defining the probability $p = \sigma(x)$ using the sigmoid function $\sigma(\cdot)$ and x, WFL is defined as follows:

$$L_{wfl} = -\sum_{c}^{C} w_c \left(y^c (1-p)^\gamma \log(p) + (1-y^c) p^\gamma \log(1-p) \right), \tag{2}$$

where $w_c = e^{-a_c}$ is the weight for class c, calculated from the prior distribution a_c of class c.

The other is an Asymmetric Loss (ASL) [25]. ASL is a loss function that takes into account the balance between positive and negative classes and is defined as follows based on focal loss.

$$L_{asl} = -\sum_{c} \left(y^c (1-p)^{\gamma_+} \log(p) + (1-y^c) p_m^{\gamma_-} \log(1-p_m) \right) \tag{3}$$

where $p_m = \max(p - m, 0)$ is the shifted probability to ignore errors in the easy negative class of classification and m is a parameter representing the margin of probability to ignore errors. By adjusting the balance of the parameter (γ_+, γ_-), the error calculation takes into account the balance between positive and negative classes.

Data sampling Approaches that consider balance when learning and classifying with unbalanced data include "oversampling," which samples more samples from rare classes, and "undersampling," which samples fewer samples from major classes. In this study, in addition to the aforementioned error functions, we consider differences in discrimination accuracy and obtain attention maps by adjusting the data sampling method. Specifically, by referring to the number of male labels shown in Table 1, we undersampled the sample of *Asarum takaoi*, which has a significantly large number of samples, and oversampled the data with a small number of samples such as *Asarum ikegamii*.

Thus, we can take the class balance into account in the training process.

5 Experiments

5.1 Experimental Settings

Dataset As a dataset, we use the *Luehdorfia japonica* dataset described in Section 3.2. The image samples are resized to 448×448 pixels and input to the network. Of this dataset, we use only 3,510 male images. Among them, we use 3,503 images with 12 hostplant labels for the experiment (see the "Use in exp." column in Table 1), as they provide a sufficient number of samples for training and evaluation. The 3,503 samples are divided randomly, and 2,799 are used for training and 704 are used for evaluation. In addition, the following two types of training sample sets are used in the experiment.

Table 2: Classification accuracy [%].

	Training set: unbalanced		Training set: balanced	
	F1-score [%]	mAP [%]	F1-score [%]	mAP [%]
BCE	74.49	92.26	88.68	92.81
WFL [27]	74.61	93.78	88.79	92.74
ASL [25]	81.74	89.23	80.14	92.91

Unbalanced A training set that uses the above 2,799 randomly selected samples as they are. Experiments are performed on the dataset with class imbalance. The training dataset is affected by the class balance of the collected dataset, resulting in a large bias in the ratio of the number of labels per attribute.

Balanced A training set in which the data sampling was performed so that the proportion of the number of attribute labels in the training samples is as equal as possible according to the number of attribute labels, as described in Section 4.2.

The the number of each training sample for each class label are shown in Table 3.

Model As a network model, we use a multi-label classification ABN with a backbone of ResNet-18 [8] pre-trained on ImageNet [4]. The mini-batch size is 32 and the number of training cycles is 1,000 epochs. During training, we apply weak data enhancements such as horizontal flip, random crop, and contrast transformation. Momentum SGD (learning rate= 0.01, momentum= 0.9) is used as the optimization method for all training, and the learning rate is divided by 10 at the 500 and 750 epochs.

Evaluation metrics F1-score and mean average precision (mAP) will be employed as evaluation metrics. Moreover, by visualizing the acquired attention map, a qualitative evaluation will be conducted, and the relationship between the gazing area and the spots will be discussed from a butterfly expert's point of view.

5.2 Accuracy Comparison

Table 2 shows the F1-score and mAP of averages over every larval hostplant label. Comparing the unbalanced dataset (Unbalanced) with the balanced dataset (Balanced), the F1-score improves when using Balanced. In particular, the F1-score is significantly improved when BCE and WFL are used, indicating that sample balance adjustment contributes significantly to the accuracy of the loss function without considering the balance between positive and negative classes.

Table 3: F1-score over each class [%]. The column of "# of samples" indicates the number of labeled samples on training dataset.

Training set	Label (*Asarum*)	# of samples	BCE	WFL	ASL
Umbalanced	*Asarum curvistigma*	188	100.00	100.00	98.94
	A. nipponicum	99	91.30	95.83	93.61
	A. yoshikawae	44	53.33	30.76	77.77
	A. megacalyx	272	66.66	75.96	67.82
	A. nipponicum var. *saninense*	43	30.76	30.76	77.77
	A. rigescens var. *brachypodion*	72	41.66	53.84	57.14
	A. asaroides	73	97.29	94.44	97.29
	A. tohokuense	33	87.50	87.50	94.11
	A. takaoi	1,659	85.30	80.00	74.28
	A. asperum	272	88.37	89.70	85.00
	A. ikegami	115	52.38	57.14	58.53
	A. blumei	287	99.30	99.30	98.59
	mean	–	74.49	74.60	81.74
Balanced	*A. curvistigma*	282	100.00	100.00	100.00
	A. nipponicum	294	97.95	100.00	100.00
	A. yoshikawae	418	73.68	77.77	72.72
	A. megacalyx	408	80.85	76.11	71.42
	A. nipponicum var. *saninense*	408	95.23	73.68	80.00
	A. rigescens var. *brachypodion*	223	63.41	80.00	68.57
	A. asaroides	365	97.43	100.00	88.37
	A. tohokuense	330	100.00	100.00	56.25
	A. takaoi	610	96.17	96.60	93.57
	A. asperum	408	84.76	88.23	71.35
	A. ikegami	436	74.57	73.07	67.69
	A. blumei	576	100.00	100.00	91.71
	mean	–	88.67	88.79	80.14

In addition, focusing on the difference in loss functions, ASL obtains a higher F1-score than the other error functions when using Unbalanced. On the other hand, the F1-score is lower for the Balanced dataset than for BCE and WFL. This confirms that ASL is effective when training on datasets with significantly unbalanced samples.

Next, the F1-score for each larval hostplant label is shown in Table 3. The table also shows the number of training samples for each larval hostplant label. The F1-score of the unbalanced case shows that ASL improves the F1-score of the hostplants with a small number of samples, i.e., *Asarum yoshikawae*, *Heterotropa nipponica* var. *saninense*, and *Asarum tohokuense*. ASL greatly improves the accuracy of attribute classes with a small number of samples, i.e., a large number of negative samples.

On the other hand, the results using Balanced showed no significant change in ASL accuracy, but a large improvement in BCE and WFL accuracy. The reason for this improvement in accuracy can be attributed to the strong co-occurrence of the hostplant labels. Specifically, from Figure 1(b), there is no possibility that both *Asarum blumei* and *Asarum asaroides* are positive, and if *Asarum fauriei* var. *nakaianum* is positive, then *Asarum takaoi* is also positive, and so on. The co-occurrence of positive labels for this larval hostplant classification is considered to be very strong compared to attribute estimation for pedestrian and facial images, which are common computer vision tasks. These results suggest that the distribution of labels is relatively simple and that a simple adjustment of the number of samples greatly improves accuracy. When more complex combinations of positive labels are present, improvements can be expected from loss functions such as ASL, but a detailed analysis of the co-occurrence and distribution of hostplant labels is one of our future works.

5.3 Evaluation of Attention Maps

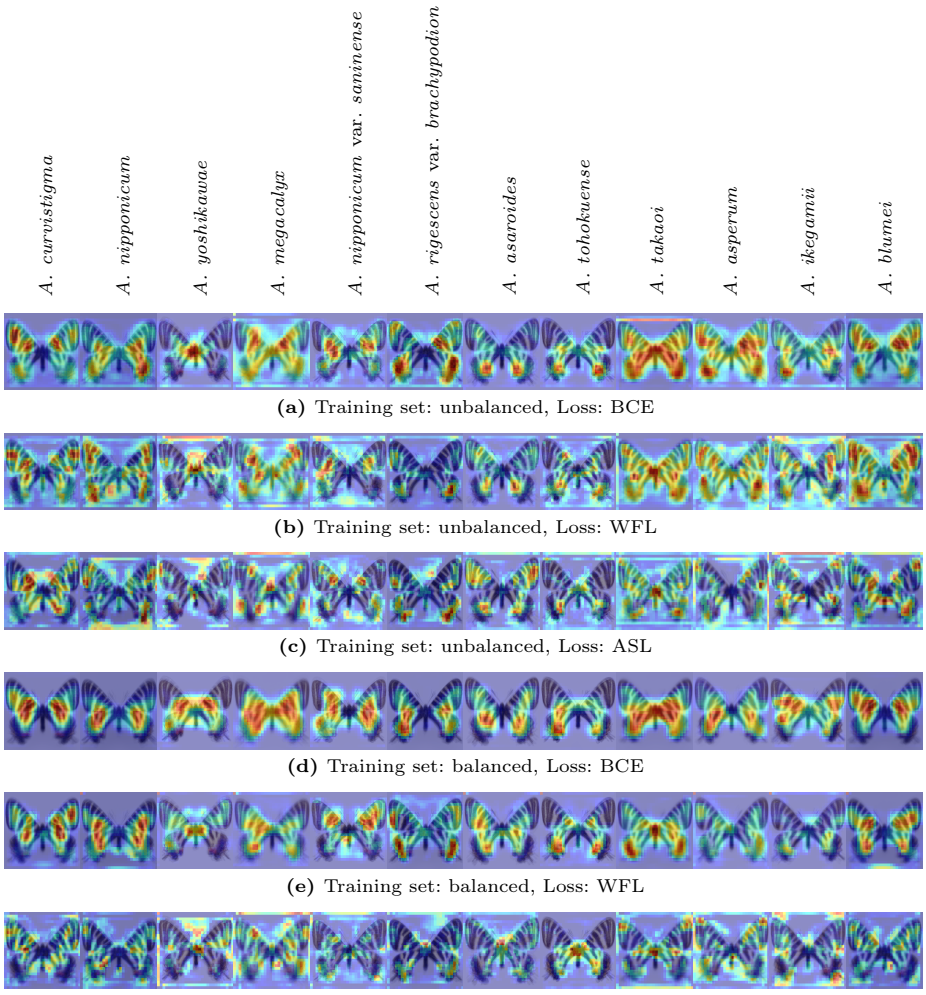
Figure 5 shows the average images of attention maps for each hostplant label. In the attention maps shown in Figure 5(a-c) when using the Unbalanced training set, there are cases where the entire wing is gazed at, and cases where strong attention occurs in the background, which is unnecessary for classification. On the other hand, in the result after the sample balance was adjusted (Figure 5(d-f)), the area of attention to the background was suppressed, and the area of attention existed in the region of the geese. This indicates that the sample balance adjustment contributes to the acquisition of stable attention, and that stable attention maps can be obtained by data sampling.

Next, we compare the differences when the error function is changed in the attention map using the Balance as shown in Figure 5(d-f). BCE and WFL with high F1-score show almost no strong attention to the background area and focus mainly on the body and wings. On the other hand, ASL showed reduced noise compared to the unbalanced case but gazed at areas that should not be recognized compared to BCE and WFL. Consequently, it is clear that sample balance adjustment plays a significant role in obtaining highly explanatory attention maps, and that changing the error function tends to improve the accuracy but not the explanatory power of the attention maps.

5.4 Discussion Based on Expert Findings

Herein, we discuss the relationship between the acquired gazing areas and the spots. Specifically, we present to the butterfly experts the attention map (Figure 5(e)) obtained by using the sample-adjusted WFL with the highest accuracy from the above results and with a stable gazing on the wings of the butterflies.

As a consideration for the overall trend, the attentions are often hit near the upper margin of the forewing (the middle of the costa to the middle of the discoidal cell on the forewing). This is known to be a representative area for experts to check when estimating the collection locality [5]. Therefore, we



(a) Training set: unbalanced, Loss: BCE

(b) Training set: unbalanced, Loss: WFL

(c) Training set: unbalanced, Loss: ASL

(d) Training set: balanced, Loss: BCE

(e) Training set: balanced, Loss: WFL

(f) Training set: balanced, Loss: ASL

Fig. 5: The attention maps taking average over each hostplant label. The top part of the figure shows the name of the hostplant label to which each column of the attention map corresponds.

can say that the gazing area acquired by the multi-label classification ABN is consistent with the general findings of experts and that the explanatory nature presented by the attention map is valid.

In addition to the overall evaluation, we focus on the attention map for the *Asarum ikegamii*. Figure 6 shows the average image of attention maps of *Asarum ikegamii* and several individual attention maps. First, the attention of *Asarum ikegamii* is concentrated around the rearwing (green circle in Figure 6), which corresponds to the conventional expert's finding. Meanwhile, the other attention

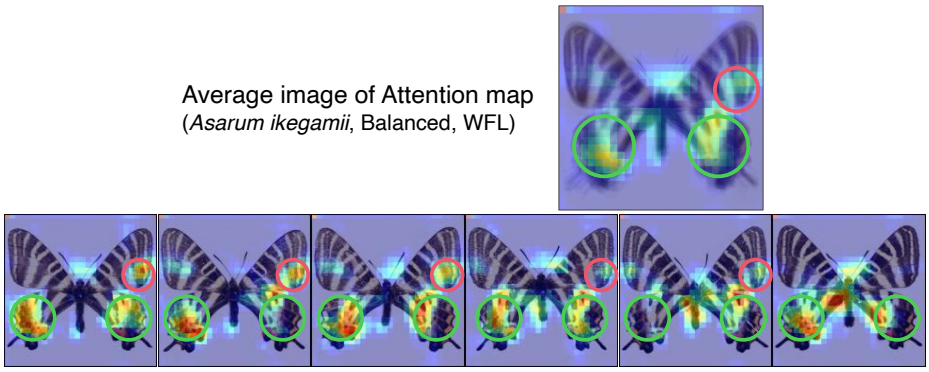


Fig. 6: Attention maps for *Asarum ikegamii*. Top row: the average image of atttention maps for *Asarum ikegamii* label. Bottom row: individual attention maps for *Asarum ikegamii* label. Every attention map images are obtained from a model trained with balanced training set and WFL.

is concentrated near the lower margin of the forewing (red circle in Figure 6). This location differs from the expert's findings and is one of the future issues to be addressed, as a detailed analysis in the future may reveal new features of the spots that have not been recognized before.

6 Conclusion

In this paper, we recognize the larval hostplant classes from *Luehdorfia japonica* images and visualize the basis of decision making on the classification results as an attention map using the multi-label classification ABN. We achieved high classification accuracy by training with a loss function that takes into account class imbalance and data sampling. The analysis of larval hostplant classification and butterfly spots using the attention map revealed that the learned attention map is focused on the areas that match the expert's knowledge.

Our future work includes more detailed validation and the identification of new trends in regional variation beyond the experts' findings. In addition, we will gradually add more *Luehdorfia japonica* samples to the dataset and make the database publicly available.

Acknowledgement This work was supported by JSPS KAKENHI Grant Number JP21H02215 and Collaboration Research Program of International Digital Earth Applied Science (IDEAS), Chubu University (Grant No. IDEAS202006 and IDEAS202105) We sincerely thank Dr. Y. Watanabe for granting us permission to use the distribution maps and for Dr. Y. Okuyama and Dr. H. Ikeda for their valuable advice on larval hostplants, which have greatly enriched the quality of our research.

References

1. Charte, F., Rivera, A., del Jesus, M.J., Herrera, F.: Resampling multilabel datasets by decoupling highly imbalanced labels. In: Hybrid Artificial Intelligent Systems. pp. 489–501 (2015) 3
2. Charte, F., Rivera, A.J., del Jesus, M.J., Herrera, F.: Dealing with difficult minority labels in imbalanced mutilabel data sets. Neurocomputing **326–327**, 39–53 (2019) 3
3. Cuthill, J.F.H., Guttenberg, N., Ledger, S., Crowther, R., Huertas, B.: Deep learning on butterfly phenotypes tests evolution's oldest mathematical model. Science Advances **5**(8), eaaw4967 (2019). https://doi.org/10.1126/sciadv.aaw4967 3
4. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: CVPR. pp. 248–255 (2009) 10
5. Department, C.S.E. (ed.): Tour of 88 Locations on *Luehdorfia japonica* (in Japanese). Choken Shuppan, Osaka (1986) 12
6. Fan, M., Lu, Y., Xu, Q., Zhang, H., Chang, J., Deng, W.: Identification of papilionidae species in yunnan province based on deep learning. In: International Conference on Image, Vision and Computing (ICIVC). pp. 611–614 (2022) 3
7. Fukui, H., Hirakawa, T., Yamashita, T., Fujiyoshi, H.: Attention branch network: Learning of attention mechanism for visual explanation. In: CVPR. pp. 10697–10706 (2019) 2, 4
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. pp. 770–778 (2016) 10
9. Huang, C., Li, Y., Loy, C.C., Tang, X.: Learning deep representation for imbalanced classification. In: CVPR. pp. 5375–5384 (2016) 3
10. Huang, C., Li, Y., Loy, C.C., Tang, X.: Deep imbalanced learning for face recognition and attribute prediction. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**(11), 2781–2794 (2020) 3
11. Inomata, T. (ed.): Atlas of the Japanese Butterflies. Takeshobo (1986) 1, 2, 6
12. Jia, J., Chen, X., Huang, K.: Spatial and semantic consistency regularizations for pedestrian attribute recognition. In: ICCV. pp. 962–971 (2021) 3
13. Jia, J., Gao, N., He, F., Chen, X., Huang, K.: Learning disentangled attribute representations for robust pedestrian attribute recognition. In: AAAI. vol. 36, pp. 1069–1077 (2022) 3
14. Kobayashi, T.: Two-way multi-label loss. In: CVPR. pp. 7476–7485 (2023) 3
15. Kohiyama, K., Fukui, H.: Visualization of regional distribution of organisms (in japanese). Tech. rep., International Digital Earth Applied Science Research Center, Chubu University (2017) 4
16. Kohiyama, K., Tanaka, H., Fukui, H., Kawamura, S.: Development of science communication support system by visualization of regional distribution of organisms (in japanese). Tech. rep., International Digital Earth Applied Science Research Center, Chubu University (2018) 4
17. Kohiyama, K., Tanaka, H., Fukui, H., Kawamura, S.: Development of science communication support system by visualization of regional distribution of organisms in accordance with gbif (in japanese). Tech. rep., International Digital Earth Applied Science Research Center, Chubu University (2019) 4
18. Lin, M., Chen, Q., Yan, S.: Network in network. In: ICLR. pp. 1–10 (2014) 7
19. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: ICCV. pp. 2999–3007 (2017) 8

20. Lin, Z., Jia, J., Gao, W., Huang, F.: Increasingly specialized perception network for fine-grained visual categorization of butterfly specimens. IEEE Access **7**, 123367–123392 (2019) 3
21. Lin, Z., Jia, J., Gao, W., Huang, F.: Fine-grained visual categorization of butterfly specimens at sub-species level via a convolutional neural network with skip-connections. Neurocomputing **384**, 295–313 (2020) 3
22. Liu, B., Blekas, K., Tsoumakas, G.: Multi-label sampling based on local label imbalance. Pattern Recognition **122**, 108294 (2022) 3
23. Liu, B., Tsoumakas, G.: Synthetic oversampling of multi-label data based on local label distribution. In: ECML PKDD. pp. 180–193 (2019) 3
24. Okuyama, Y., Goto, N., Nagano, A.J., Yasugi, M., Kokubugata, G., Kudoh, H., Qi, Z., Ito, T., Kakishima, S., Sugawara, T.: Radiation history of Asian Asarum (sect. Heterotropa, Aristolochiaceae) resolved using a phylogenomic approach based on double-digested RAD-seq data. Annals of Botany **126**(2), 245–260 (2020). https://doi.org/10.1093/aob/mcaa072 2, 6
25. Ridnik, T., Ben-Baruch, E., Zamir, N., Noy, A., Friedman, I., Protter, M., Zelnik-Manor, L.: Asymmetric loss for multi-label classification. In: ICCV. pp. 82–91 (2021) 3, 9, 10
26. Saquib, M.S., Schumann, A., Wang, Y., Stiefelhagen, R.: Deep view-sensitive pedestrian attribute inference in an end-to-end model. In: BMVC. pp. 134.1–134.13 (2017) 3
27. Sarafianos, N., Xu, X., Kakadiaris, I.A.: Deep imbalanced attribute classification using visual attention aggregation. In: ECCV. pp. 680–697 (2018) 3, 4, 8, 10
28. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: CVPR. pp. 815–823 (2015) 3
29. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: ICCV. pp. 618–626 (2017) 4
30. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. International Journal of Computer Vision **128**(2), 336–359 (2020) 4
31. Tang, C., Sheng, L., Zhang, Z., Hu, X.: Improving pedestrian attribute recognition with weakly-supervised multi-scale attribute-specific localization. In: ICCV. pp. 4997–5006 (2019) 3
32. Wang, Y., Gan, W., Yang, J., Wu, W., Yan, J.: Dynamic curriculum learning for imbalanced data classification. In: ICCV. pp. 5017–5026 (2019) 3
33. Watanabe, Y.: Japanese insects 1: *Luehdorfia japonica* (in Japanese). Bun-ichi Sogo Shuppan (1985) 2
34. Wham, D.C., Ezray, B., Hines, H.M.: Measuring perceptual distance of organismal color pattern using the features of deep neural networks. bioRxiv, 736306 (2019) 3
35. Yago, M., Hirakawa, T., Kawamura, S., Oba, Y., Nawa, T., Sugita, S., Kohiyama, K., Fukui, H.: Quantitative consideration of butterfly spotted regional variation using image recognition techniques (in japanese). Tech. rep., International Digital Earth Applied Science Research Center, Chubu University (2020) 2
36. Yago, M., Hirakawa, T., Kawamura, S., Oba, Y., Nawa, T., Sugita, S., Kohiyama, K., Fukui, H.: A quantitative study of geographic variation in butterfly spotted patterns using deep learning image recognition techniques (in japanese). Tech. rep., International Digital Earth Applied Science Research Center, Chubu University (2021) 2
37. Yang, J., Fan, J., Wang, Y., Wang, Y., Gan, W., Liu, L., Wu, W.: Hierarchical feature embedding for attribute recognition. In: CVPR. pp. 13055–13064 (2020) 3

38. Zhao, R., Li, C., Ye, S., Fang, X.: Butterfly recognition based on faster r-cnn. Journal of Physics: Conference Series **1176**(3), 032048 (2019) 3
39. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: CVPR. pp. 2921–2929 (2016) 4