Asymptotic Cross-Entropy Weighting and Guided-Loss in Supervised Hierarchical Setting using Deep Attention Networks

Charles Kantor[1,2,3] **Brice Rauby**[2,3,5] **Léonard Boussioux**[2,3,6]
Emmanuel Jehanno[2,3] **André-Philippe Drapeau Picard**[4] **Maxim Larrivée**[4] **Hugues Talbot**[2,3,7]
[1]Mila Artificial Intelligence Institute, Montreal, Canada [2]Paris-Saclay University, France [3]Ecole CentraleSupélec Paris, France
[4]Montreal Insectarium - Space for Life, Canada [5] Polytechnique Montreal, Canada
[6]Massachusetts Institute of Technology, Operations Research Center, Cambridge, MA, USA [7]Inria Paris, France

Abstract

This article reveals two main techniques for improving fine-grained recognition and classification, defined as executing these tasks between items with similar general patterns, but that differ through small details. First, we build a preliminary automated segmentation algorithm to ignore the image's background for attention-guided classification. To do this, we wield segmentation-issued masks to make the classification network's training easier through an additional loss that penalizes attention given to features outside the mask using convolutional block. Furthermore, we proffer a hierarchical loss based on cross entropy penalizing parent-level classification to leverage the philology of each wildlife species. We applied our approaches in the particular context of butterfly recognition, which is of practical interest to entomologists.

Context and Motivation

In this work, we deal with the issue of accurately identifying large numbers of items in photographs, some of which may differ only in minute details. This is a difficult problem because both large and small differences must be taken into account in order to recognize and classify.

Among the images collected, a high percentage of species remains unidentified and represents a time-consuming labeling task for experts. Identifying an insect on the species level is challenging and depends on the tiniest of details. Citizen scientists can help collect a large amount of data such as insect photographic documentation (Horn et al. 2017; Boussioux et al. 2019), but accurate identification remains restricting. Recent improvements in performance in a wide range of classification tasks with deep learning methods offer population monitoring opportunities and efficient and large scale annotations. We worked with the eButterfly (Prudic et al. 2017) citizen science program, which maintains a fine-grained dataset of observations of all North American butterflies species.

We develop computer vision algorithms and propose fine-grained classification innovations using segmentation tools to encourage the model to focus on areas of an image that are salient for identification. We propose an additional loss using the segmentation masks, penalizing attention given to features outside the mask. We also design a specific loss function that leverages the dataset's hierarchical nature, consequently improving the Family, Genus and Species level accuracy.

AAAI Fall 2020 Symposium on AI for Social Good.

Related Work

Preliminaries

Fine-grained classification is a category of image classification: the task is to distinguish between subtly rather than grossly different items; for example between different species of birds or dogs rather than giraffes vs. trucks. This setting is more complex, requires better annotations, more data and is not yet satisfactorily solved (Xie et al. 2013; Chai, Lempitsky, and Zisserman 2013). A fundamental difficulty is to induce the learning architecture to focus on small but essential details without relying on overly complicated annotations. A recent interesting approach has been to use a deconstruction-reconstruction method to this end (Chen et al. 2019) and bipartite and bi-modal graphs (Zhou and Lin 2016; Song et al. 2020).

Segmentation is a fundamental task in computer vision. Its objective is to find semantically consistent regions that represent objects. Given enough data and annotations, deep recurrent CNN architectures such as ResNet (He et al. 2016) and recurrent auto-encoders like U-Net (Ronneberger, Fischer, and Brox 2015) constitute the current state-of-the-art in segmentation methods. In particular, U-Net and its variants may learn a segmentation task from a few hundred labeled inputs.

The background of macro wildlife photos is typically full of environmental details like grass or leaves that can mislead the classification model and introduce bias. We noticed experimentally via saliency maps that too much attention is generally paid to the background rather than to the insect itself. Consequently, we used automatic segmentation to help focus on the foreground.

Hierarchical Classification

Hierarchical labels of a fine-grained dataset can be leveraged to improve performance (Zheng et al. 2020).

Tree-CNN is an architecture-based approach to this setting in (Roy, Panda, and Roy 2019). Their study aims to

overcome the issue of catastrophic forgetting when fine-tuning a model successively on each task (i.e., pre-training on the task of classifying Families, then Genus and finally Species). The architecture is built with a common trunk and several finer branches corresponding to each task. Hence, the model uses the hierarchy through the trunk while also preserving the memory of each task. This *Tree-CNN* limits the computation costs of retraining and can learn with a smaller training effort than fine-tuning while retaining much of the accuracy.

(Wu, Tygert, and LeCun 2019) present another approach based on the loss instead of the architecture. They propose a new loss that takes the hierarchy into account and is no longer a *flat loss* compared to the cross-entropy (which compares the same level classes). A convenient metric should make all leaves equidistant from the root node.

(Kosmopoulos et al. 2020) develops a different methodology for tuning the loss. They compare different measures presented in the literature and classify them into two subgroups: pair-based and set-based measures. They consider building a hierarchical loss as an optimization problem and propose a pair-based metric, optimized with a max-flow approach. The article also offers a set-based implementation that approximates the sparsity-inducing ℓ_0 norm. Considering the path from the root to the predicted node as a set of nodes and identically for the ground truth node, they can propose a measure that uses the intersection, union and difference between the sets. This measure computes analogous precision, recall and F1 scores. They implement a *Lowest Common Ancestor*, which is a bridge between pair-based and set-based measures. The corresponding new measure performs very well and takes the advantages of both approaches.

Attention and Visualizing CNNs

Attention mechanisms were introduced originally for Neural Machine Translation in (Bahdanau, Cho, and Bengio 2015) using recurrent neural networks. These mechanisms utilize a divide-and-conquer approach to various AI tasks by focusing features on relevant items. These tools have been extensively used in Natural Language Processing tasks (e.g., (Parikh et al. 2016) and (Lin et al. 2017)). (Vaswani et al. 2017) developed a model relying only on attention and achieved state-of-the-art results in machine translation. Later, the use of attention was extended to computer vision tasks such as image classification and segmentation.

The Convolutional Block Attention Module or CBAM (Woo et al. 2018a) proposes a simple attention mechanism for feed-forward Convolutional Neural Network (CNN) architectures. Its lightweight structure and generality make it suitable for many vision tasks that require large numbers of parameters.

Grad-CAM (Selvaraju et al. 2017) is a popular technique to make CNN models more explainable, showing on which areas of the picture they focused on making the prediction, using reverse gradient propagation descent. The discriminative regions are localized through the areas of high gradient flow within the network.

Imbalanced Data

Main source

For our preliminary experiments, we use a data set of pictures submitted across Canada, Mexico and the United States, representing over seven hundred different species as of June 2020. Among these observations, two-thirds have been annotated by experts. The eButterfly program, co-founded by the Montreal Insectarium, allows participants to record sightings by uploading images with date and time information (Prudic et al. 2017).

Highly imbalanced classes

Our data set is organized hierarchically. Each image has three labels: a species belongs to one and only one genus, belonging to one and only one family. This distribution of labels enables us to have different complexity levels for our classification task. Given more than two-thirds labeled images, we anticipate being able to learn the family label with the best precision, provide a slightly less accurate estimate of the genus and a slightly worse again estimate of the species. In the provided dataset, classes are highly imbalanced, meaning we are facing a problem of fine-grained classification with significantly under-represented classes.

Methods

Guided Attention Mechanism

As the shape of butterflies presents a limited variability, it seems feasible to incorporate prior knowledge of the shape of the object of interest. Due to the similarity between butterflies' overall shape, we posit that butterfly segmentation is a simpler task than its fine-grained classification. We propose to use masks obtained through an automatic segmentation pipeline to improve the classification performance. However, even though the segmentation is generally correct, a few failure cases can deteriorate the classification's performance if used during test-time. For this reason, we developed a method to leverage these masks during training through an additional loss later called *guided attention loss*.

Prior automated segmentation We used a pre-trained Mask R-CNN network to generate the segmentation masks and fine-tuned it on a small subset of the dataset. This approach is possible because the butterfly segmentation task is sufficiently similar to the task of segmenting other objects present in a common dataset, and therefore, pre-training is very effective. We annotated a small subset (10%) used for pre-training, and we qualitatively assessed the segmentation performance. The segmentation results obtained were satisfactory to be used in the *guided attention*.

Foreword designed attention-based loss As our goal is to enforce the model's attention on the butterfly, we use a network that was explicitly implementing an attention mechanism. For this reason, we used an attention model based on the generic implementation of CBAM (Woo et al. 2018b).

This architecture is a good candidate for its correct classification results on several benchmarks and because it separates the spatial attention mask from the channel attention. Therefore, we could penalize the high values of the spatial attention mask located outside of the butterfly. Our loss can be written as follows:

$$\mathcal{L}(M, S) = \frac{\sum_{i,k,l} M_i^{k,l}(1 - S^{k,l})}{\sum_{i,k,l} M_i^{k,l}}$$

with $M_i^{k,l}$ the pixel intensity of index (k, l) along the spatial dimension of the attention-issued mask (M) for the channel i and $S^{k,l}$ the pixel's value of index (k, l) along the spatial dimension of the segmentation-issued mask (S).

An attention-based loss is applied at each level to the attention-issued masks (M) computed spatially, at low scale, with the segmentation-issued masks (S) max-pooled to the correct spatial dimension (e.g., for $M \in [28 \times 28]$ and $S \in [250 \times 250]$, a max-pooling is applied to (S)).

Experimental set-up For our preliminary experiments, we use a ResNet (He et al. 2016) pre-trained on Imagenet (Deng et al. 2009). To prevent over-fitting, the weight decay parameter was finetuned and a dropout layer (Srivastava et al. 2014) was added before the last fully-connected layer with a keep-probability. Random rotation, flipping, rescaling and cropping were added for data-augmentation during training. The best weights on the validation set were saved and the training was interrupted when no improvement was noticed for more than 50 epochs. To obtain preliminary results without changing the class-balancing parameters, we restrained ourselves to a reduced dataset that was perfectly balanced containing less than 100 species.

We compare the proposed approach with the model without attention mechanism (original ResNet) to the model with attention mechanism trained without guided attention loss. We witness the importance of the attention mechanism in the classification task, highlighting the potential of the *guided attention* approach. Our proposed approach yields almost as good in top 1 and better in top 3 accuracies than ResNet with CBAM.

Analysis Our top 1 accuracy scores in training are near perfect (better than 99% for the three models) which means our loss is ineffective due to over-fitting. We will use more training data to address the class imbalance in future work as well as using an adaptive sampling strategy. Our strategy with adaptive sampling is to use our uncertainty prediction measure in our approach, called Over-CAM (Kantor et al. 2020): our measure rejects the predictions if the overlap between binarized attention (or transformed saliency maps) and object segmentation is not satisfactory. Indeed, this case implies that the network likely based its prediction on at least some regions outside the butterfly, i.e., both the background and foreground. Then, we determine the overlap distribution on the whole test set. Following that, several thresholds are chosen regarding the distribution curve to determine from which percentage we could ensure a corresponding certainty of prediction.

Indeed, with a correct prediction and a good overlap on a given picture, it is reasonable to under-weigh this sample in our training set. Furthermore, a good prediction with a low overlap would mean the decision is based on irrelevant features and therefore under-weighting the image can even benefit the training. Indeed, we can imagine that the wrongly used features would be forgotten later in training. Finally, we can augment the weight of the images incorrectly predicted, similarly to a hard-negative mining strategy (HNM) (Felzenszwalb et al. 2009): it bases the sampling process on the training results for each class. This is equivalent to providing an uncertainty measure, which we can use to ameliorate the class imbalance problem via an image adaptative sampling.

Regularization The most straightforward solution will be to use more training data (another training set is already at our disposal). Our future work will be to use more training data: one can, for example, use all the training data (with an adaptive sampling strategy in addressing the class imbalance) or pre-train our model on other pre-existing butterflies datasets. If unsuccessful in addressing the overfitting issue, our approach would be to implement stochastic depth as a regularization method, in addition to stronger data-augmentation, such as methods of consistency training applied in a semi-supervised configuration as in MixMatch (Berthelot et al. 2019).

Hierarchical Classification

Our data structure presents a hierarchical property since each label is composed of different but related items. We should make the best use of this knowledge to improve the results. Indeed, classifying other families should be simpler and more robust than classifying over species. Exploiting this hierarchy can improve robustness. For example, if the model is uncertain between two species A and B, which respectively belong to families 1 and 2, while being certain it belongs to family 1, it should predict species A.

Learning underlying structure while preserving flat classification Even if these hierarchies are common in the real world, they are challenging to leverage to improve the classification. On the one hand, using the parents-to-children relation seems critical to extract relevant features and reduce parent-level classification mistakes where the task should be easier. On the other hand, over-penalizing parent-level relationships can cause the classifier to under-perform on leaf classes compared to flat classification. Therefore designing a loss that enforces the learning of the underlying hierarchy while preserving the flat classification performance is a challenge we plan to address.

We thus propose to evaluate the impact of the weighting of the varying elements of a hierarchical loss on the classification performance. In addition, we introduce a loss *WCE* based on cross-entropy that improves the flat classification performance while penalizing parent-level classifi-

Table 1: Resnet model performance with and without Convolutional Block Attention Module and Guided-Attention. We provide the average accuracy obtained over 3 different seeds and the standard deviation between parenthesis. Current best accuracies are in bold.

Accuracy	ResNet	ResNet + CBAM	ResNet + CBAM + Designed Loss
Top 1	79.54 (0.70)	80.95 (0.45)	**81.01 (0.40)**
Top 3	91.72 (0.49)	**93.35 (0.20)**	93.00 (0.29)

cation mistakes. For a given sample, we use the following loss:

$$WCE = -\lambda_s \log p(c_s) - \lambda_g \log p(c_g) - \lambda_f \log p(c_f),$$

with $\lambda_s, \lambda_g, \lambda_f$ being the weighted coefficients for species, genus and family, c_s, c_g, c_f being the species, genus and family class labels and p a probability distribution function.

The general hierarchical loss

Problem definitions In supervised hierarchical classification applied to images, we consider a data-set D containing images whose labels belong to C, a set of classes and we assume the existence and the knowledge of an underlying tree-structures of height $d > 1$. The leaves of the structure represent all the classes of C. More precisely, each leaf is a set containing only one element of C and each element in C is contained in a different leaf. A parent node is defined as the union of its children. We note as C_k the collection of sets composed of the nodes at depth k (each node being the set composed of the classes descending from it). This way, we have C_0, the root of the tree, a collection of sets which union is equal to C. We assume that:

- $\forall c \in C, \exists!\ c' \in C_d,\ c \in c'$
- $\forall c' \in C_d,\ |c'| = 1$
- $\forall 1 \leq i \leq d,\ \forall c \in C_i,\ \forall\ c' \in C_{i-1}, c \cap c' \neq \emptyset \Rightarrow\ c \subseteq c'$
- $\forall 1 \leq i \leq d,\ \forall c \in C_i,\ \exists!\ c' \in C_{i-1},\ c \subseteq c'$

Under these assumptions, it is possible to assess the importance of a classification error. Given an image $I \in D$ and its corresponding labels $c_I \in C$ and considering the prediction $t_I \in C$, we define the importance of the error : $k(I,t) = d - \max\{0 \leq n \leq d \mid \exists c \in C_n,\ t_I \in c \wedge c_I \in c\}$. We note that if $t_I = c_I$, we have $k(I,t) = 0$.

With this setting, we are interested in learning that a classifier reduces the number and the importance of the errors. We will note p the predicted probability; it is defined over the leaves of the structure and can be extended to every node considering each parent node's construction principle. Indeed, at each level, every node has zero intersection with its siblings and therefore, the predicted probability of a parent node is equal to the sum of the predicted probability of its children.

Weighting-Impact method A natural and straightforward approach to learn the hierarchical structure is through a weighted classification loss. We consider the cross-entropy loss for the nodes at each depth in the tree structure. For a depth $k \in \{1, .., d\}$, we consider the cross entropy loss at this depth defined as follow :

$$CE(k, I) = \mathbf{1}_{c_I \in c \wedge\ c \in C_k}(c) \log(p(c))$$

with p the predicted probability of a node c.

Given a tuple of weights $\Lambda = (\lambda_1, \ldots, \lambda_d) \in \mathbf{R}_+^{*d}$, we compute the weighted cross-entropy:

$$WCE_\Lambda(I) = \sum_{j=1}^{d} \lambda_j CE(j, I)$$

This weighted cross entropy loss is differentiable and allows the optimization of the weights of a CNN through gradient descent.

Cross-entropy loss limitation It is critical to tune the Λ parameter properly, which requires a time-consuming optimization or some expert knowledge. Moreover, the cross-entropy loss has inherent limitations that need to be addressed for proper hierarchical learning. Indeed, as the model weights converge during training, the predicted probability of the target class converges to 1. Moreover, given the labels' underlying structure, the parent node's predicted probability is always greater than its children's. Since the cross-entropy loss is expressed as $-\log p(c_i)$, with c_i the label class, its gradient regarding p has a magnitude that decreases as p augments. As a result, the cross-entropy loss naturally under-weighs the optimization of parent-level features with respect to the children and requires a weighting. For this reason, we propose a loss in which gradient magnitude is not decreasing while getting closer to 1. The divergence in 0 implies a small impact of the weighting and the convergence to 0 in 1 implies importance of the species

Loss properties We designed a loss function with the shape shown in Figure 1. When both probabilities are close to 1, it yields 0 and when probabilities are both close to 0, it yields 1. The essence of that idea is that when the gradient magnitude of the loss is close to 0, we have a gradient magnitude higher in the direction of genus rather than families. The reverse is observed when it is close to 1. Such penalization is selected to hinder optimization on the genus if family's optimization is affected negatively.

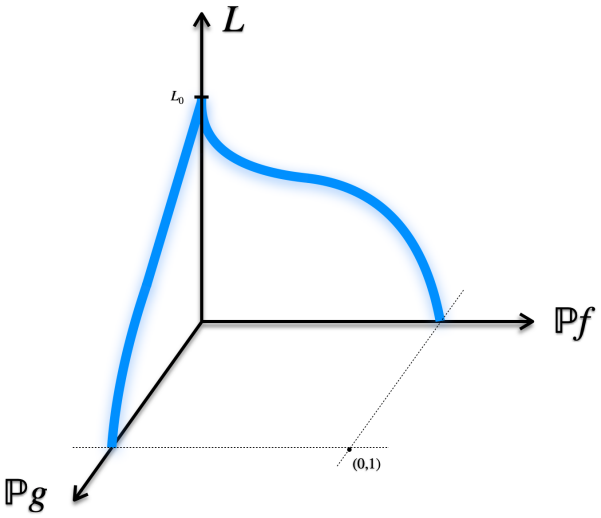

Figure 1: Loss function to be designed: $\mathbf{P}f$ is the family probability and $\mathbf{P}g$ is the genus probability.

Conclusion

In this article, we propose a method for fine-grained recognition and classification of wildlife images. In particular, we propose to guide the convolutional neural networks by leveraging attention masks along with segmentation as a means to being less sensitive to the typical detail-rich environment. This work shows improved results in top-3 accuracy in comparison to the state of the art. Furthermore, we explore the use of a hierarchical loss to leverage species philology. Our approach is general enough to be adapted in broader fine-grained classification contexts. Our methodology can be of great use for large-scale wildlife crowd-sourcing programs that gather crucial census data to understand species demographics and dynamics.

References

Bahdanau, D.; Cho, K.; and Bengio, Y. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. *CoRR* abs/1409.0473.

Berthelot, D.; Carlini, N.; Goodfellow, I.; Oliver, A.; Papernot, N.; and Raffel, C. 2019. MixMatch: A Holistic Approach to Semi-Supervised Learning. URL https://arxiv.org/pdf/1905.02249.pdf.

Boussioux, L.; Giro-Larraz, T.; Guille-Escuret, C.; Cherti, M.; and Kégl, B. 2019. InsectUp: Crowdsourcing Insect Observations to Assess Demographic Shifts and Improve Classification. URL https://arxiv.org/pdf/1906.11898.pdf.

Chai, Y.; Lempitsky, V.; and Zisserman, A. 2013. Symbiotic segmentation and part localization for fine-grained categorization. In *Proceedings of the IEEE International Conference on Computer Vision*, 321–328.

Chen, Y.; Bai, Y.; Zhang, W.; and Mei, T. 2019. Destruction and construction learning for fine-grained image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5157–5166.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*.

Felzenszwalb, P. F.; Girshick, R. B.; McAllester, D.; and Ramanan, D. 2009. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence* 32(9): 1627–1645.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Horn, G. V.; Aodha, O. M.; Song, Y.; Shepard, A.; Adam, H.; Perona, P.; and Belongie, S. J. 2017. The iNaturalist Challenge 2017 Dataset. *CoRR* abs/1707.06642. URL http://arxiv.org/abs/1707.06642.

Kantor, C.; Rauby, B.; Boussioux, L.; Jehanno, E.; and Talbot, H. 2020. Over-CAM : Gradient-Based Localization and Spatial Attention for Confidence Measure in Fine-Grained Recognition using Deep Neural Networks. doi:10.1109/ICCV.2017.322. URL https://hal.archives-ouvertes.fr/hal-02974521. Working paper or preprint.

Kosmopoulos, A.; Partalas, I.; Gaussier, E.; Paliouras, G.; and Androutsopoulos, I. 2020. Evaluation Measures for Hierarchical Classification: a Unified View and Novel Approaches. URL http://www2.aueb.gr/users/ion/docs/dami_final_manuscript.pdf.

Lin, Z.; Feng, M.; Dos Santos, C.; Yu, M.; Xiang, B.; Zhou, B.; and Bengio, Y. 2017. A Structured Self-attentive Sentence Embedding .

Parikh, A.; Täckström, O.; Das, D.; and Uszkoreit, J. 2016. A Decomposable Attention Model for Natural Language Inference. 2249–2255. doi:10.18653/v1/D16-1244.

Prudic, K. L.; McFarland, K. P.; Oliver, J. C.; Hutchinson, R. A.; Long, E. C.; Kerr, J. T.; and Larrivée, M. 2017. eButterfly: Leveraging Massive Online Citizen Science for Butterfly Conservation. *Insects* 8(2). ISSN 2075-4450. doi:10.3390/insects8020053. URL https://www.mdpi.com/2075-4450/8/2/53.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.

Roy, D.; Panda, P.; and Roy, K. 2019. Tree-CNN: A Hierarchical Deep Convolutional Neural Network for Incremental Learning. URL https://arxiv.org/pdf/1802.05800.pdf.

Selvaraju, R. R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; and Batra, D. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, 618–626.

Song, K.; Wei, X.; Shu, X.; Song, R.; and Lu, J. 2020. Bi-Modal Progressive Mask Attention for Fine-Grained Recog-

nition. *IEEE Transactions on Image Processing* 29: 7006–7018. doi:10.1109/TIP.2020.2996736.

Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; and Salakhutdinov, R. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* 15(56): 1929–1958. URL http://jmlr.org/papers/v15/srivastava14a.html.

Vaswani, A.; et al. 2017. Attention is All you Need. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 30*, 5998–6008. Curran Associates, Inc. URL http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf.

Woo, S.; Park, J.; Lee, J.-Y.; and So Kweon, I. 2018a. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, 3–19.

Woo, S.; Park, J.; Lee, J.-Y.; and So Kweon, I. 2018b. CBAM: Convolutional Block Attention Module. In *The European Conference on Computer Vision (ECCV)*.

Wu, C.; Tygert, M.; and LeCun, Y. 2019. A hierarchical loss and its problems when classifying non-hierarchically. URL https://arxiv.org/pdf/1709.01062.pdf.

Xie, L.; Tian, Q.; Hong, R.; Yan, S.; and Zhang, B. 2013. Hierarchical part matching for fine-grained visual categorization. In *Proceedings of the IEEE International Conference on Computer Vision*, 1641–1648.

Zheng, H.; Fu, J.; Zha, Z.; Luo, J.; and Mei, T. 2020. Learning Rich Part Hierarchies With Progressive Attention Networks for Fine-Grained Image Recognition. *IEEE Transactions on Image Processing* 29: 476–488. doi:10.1109/TIP.2019.2921876.

Zhou, F.; and Lin, Y. 2016. Fine-grained image classification by exploring bipartite-graph labels. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1124–1133.