

Metric Weighted Ensemble Focal Loss for Snake Species Identification

Aarti Balana¹

¹Research Consultant, WorldQuant Brain

Abstract

This paper outlines the participation in the LifeCLEF-SnakeCLEF 2023 challenge for automated identification of snake species. A metric-weighted ensemble loss function was introduced. The proposed loss function was put to test with various deep learning architectures, including Data Efficient Image Transformers (DEiT), EfficientNet and DinoV2. The SnakeCLEF 2023 challenge offers a substantial database comprised of images and associated recording locations for 1,784 snake species, serving as a resource to address this identification problem. The DEiT model trained with proposed loss outperformed the EfficientNetB0 model trained with focal loss by 49% in terms of the F1 score (macro). In the context of the SnakeCLEF2023 challenge, highest F1 score (macro) of 0.36 was achieved. Notably, our experiments forego the usage of metadata such as region or country, focusing solely on image based classification. The development of such a system holds significant potential to reduce snakebite induced mortality and morbidity, aligning closely with global health and biodiversity objectives.

Keywords

Metric-Weighted Ensemble Focal Loss, Snake Species Identification, Imbalanced Dataset, Efficient Transformers

1. Introduction

Ophidian Envenomation (OE), recognized as the second most fatal neglected tropical disease, presents a significant global health crisis. Each year OE impacts approximately 1.8 to 2.7 million individuals [1], predominantly those residing in underdeveloped, remote, and politically marginalized tropical regions. The annual death toll from OE ranges from 81,410 to 137,880, while around 400,000 [1] survivors endure permanent physical and psychological trauma, underscoring the urgent necessity for focused attention on this health issue. Ophidian envenomation predominantly impacts individuals in Africa, Asia, and Latin America. Asia alone witnesses up to 2 million cases [1] of snake venom intoxication annually. Concurrently, Africa records an estimated 435,000 to 580,000 [2] instances each year that necessitate medical intervention due to snake bites. The appropriate administration of antivenoms can be a lifesaver, yet it often depends upon the correct taxonomic identification (family, genus, and species) of the snake in

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

✉ aarti.balana@gmail.com (A. Balana)

🌐 <https://github.com/aarti-b/SnakeCLEF2023-Metric-weighted-ensemble-focal-loss-CEUR-WS-proceedings>

(A. Balana)

🆔 0009-0000-6507-974X (A. Balana)

© 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)

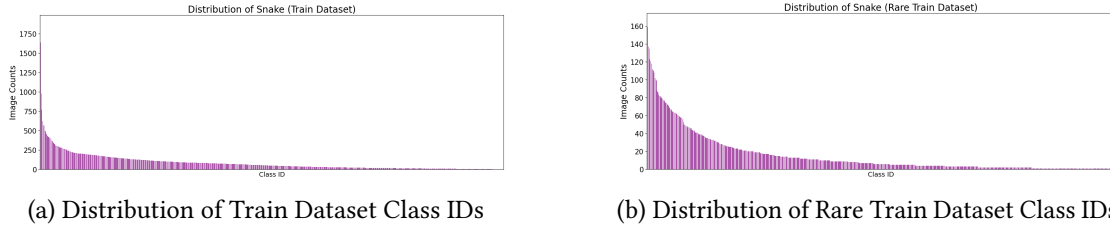


Figure 1: Exponentially Decreasing Distribution of Class IDs in the Train Dataset showcases the long-tailed distribution of example classes. The x-axis represents the unique class IDs, while the y-axis represents the frequency of each class ID in the dataset. It can be observed that a small number of classes have a high frequency but the majority of classes have a significantly lower frequency with many appearing just a few times which indicates a significant class imbalance.

question. An automated system capable of providing a preliminary identification to healthcare professionals using a low quality photograph could expedite the treatment process.

The SnakeCLEF2023 challenge [3, 4, 5] urged participants to experiment upon different deep learning methodologies for an accurate classifier for snake species, adaptable to diverse conditions. This challenge was integrated into the LifeCLEF 2023 research platform, which is dedicated to the automated identification of species [6, 7] across five unique challenges. The approach for loss function implemented in this paper draws inspiration from the metric, venomous species confusion error, mentioned in the challenge [8, 9]. The objective of the SnakeCLEF challenge [3, 4, 5, 8, 9] is to enhance the robust and accurate identification of snake species from photographs. An operational system proficient in snake species recognition from images could substantially enrich the data on snakebite ecoepidemiology, clinical management (like precise antivenom administration), and patient outcomes [10, 11, 12, 13, 14]. Despite the fact that only about 20% of global snake species are medically significant [12], understanding of their epidemiological relevance (like the count of snakebites each species is accountable for in different regions) remains limited. Moreover, the knowledge about the geographical distribution of various snake species is far from complete. It is not uncommon for the majority of images of a particular snake species to come from a small group of countries or even a single country which in turn bias the model and model trained on snake species in one region is not detected in another region because of environmental difference causing the difference in the appearance. On the other hand, several snake species resemble species from other continents, with which they share no overlap in their geographic distribution.

2. Related Work

The automated identification of snake species utilizing machine learning remains an infrequently explored field, primarily due to the scarcity of annotated image datasets. James et al. [15] proposed a semi-automated technique that involved extracting taxonomical features from images to discriminate among six different species. Amiza Amir et al. [16] introduced Image Classification for Snake Species Using Machine Learning Techniques, where they investigated the accuracy of five state-of-the-art machine learning techniques — decision tree J48, nearest

neighbors, k-nearest neighbors (k-NN), backpropagation neural network, and naive Bayes — for image-based snake species identification problem. They showed backpropagation neural network and nearest neighbour are highly accurate with greater than 87 % accuracy. Recent studies have employed automated feature extraction strategies, such as texture features [16] or deep learning methodologies [7, 17, 18, 19] as manual feature extraction of snake species data is time taking and requires some expertise in the feature engineering and in-depth understanding of the dataset.

Image-Based Classification of Snake Species Using Convolutional Neural Network (CNN) by Isa Setiawan et al. [20] experimented with three CNN architectures using a dataset of 415 snake images from five common hazardous venomous snake species in Indonesia where snake images are classified the snake images with a high accuracy of 82%. Patel et al. [17] deployed deep learning strategies to create a smartphone application that could distinguish images of nine snake species native to the Galápagos Islands in Ecuador. They have fused three data sources together resulting a total of 250 images for the training dataset. After testing various model architectures for object detection and image classification, the model based on the Faster Region-based Convolutional Neural Network (Faster R-CNN) [21] ResNet [22] demonstrated the highest classification accuracy at 75%. However, the authors emphasized the necessity for larger training samples for further research in this area. Chamath Abeysinghe et al. [23] experimented classification using Siamese Networks [24] and explored the applicability of single shot learning techniques along with deep neural networks to solve the snake image classification problem. An artificial intelligence (AI) model to identify snakes from across the world: Opportunities and challenges for global health and herpetology, Isabelle Bolon et al., [25], they developed an AI model based on Vision Transformer (ViT) [26]. Snake Detection and Classification using Deep Learning by Zihan Yang et al. [27] used different deep learning models, YOLO v3, Tiny YOLO, Faster RCNN ResNet and others for snake detection and classification problem on Australian snake dataset. The dataset used in [27] paper was a collection from Google Image through web crawling and ImageNet dataset. In SnakeCLEF Challenge 2020, FHDO BCSG[28] used combination of image and location information for snake species identification using object detection and EfficientNets. In SnakeCLEF Challenge 2022, GG [29] introduced a novel architecture CoLKA- Net based on VAN [30] and CoAtNet [31].

3. Dataset

In SnakeCLEF challenge 2023 [5, 9], a comprehensive developed dataset, which comprises 103,404 snake observations with a total of 182,261 photographs. These images represent 1,784 unique snake species sighted across 214 countries. The data was gathered from online biodiversity platforms, namely iNaturalist and Herpmapper. It is important to note that the provided dataset exhibits a pronounced long-tailed class distribution, with the most frequently observed species, *Natrix Natrix*, represented by 1,262 observations (encompassing 2,079 images), while the least observed species have merely 3 recorded observations [8]. This dataset also includes metadata that is country code and name of the species. This dataset is bifurcated into two components, serving as training data where one part includes commonly observed species, while the other contains data pertaining to rarer snake species. Note that in this work, the metadata is not

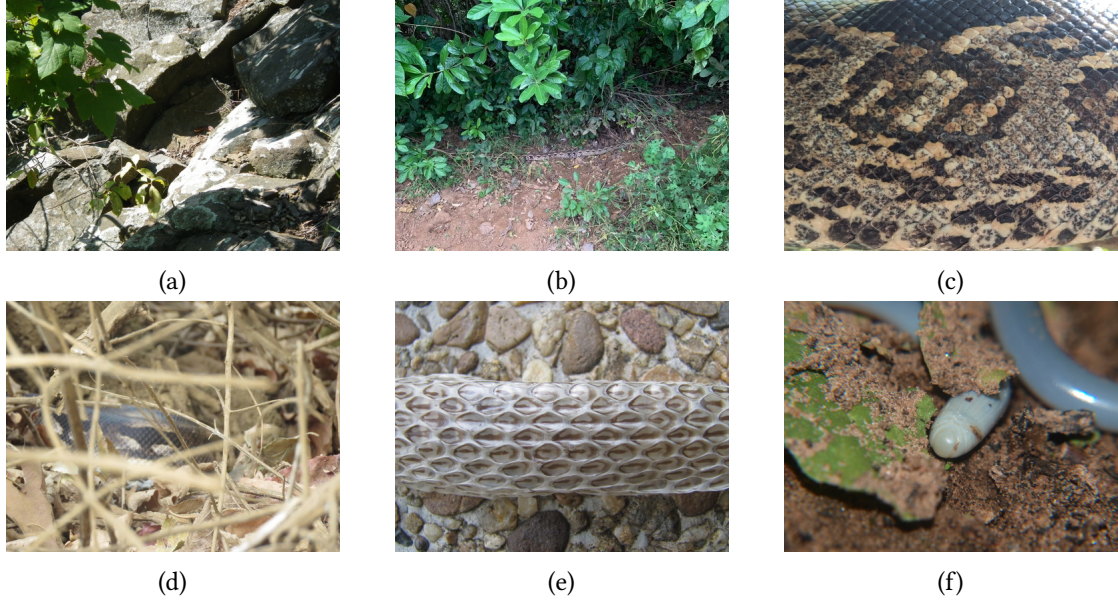


Figure 2: The dataset has variety of images which are quite challenging samples. Some snakes in the image are less distinguishable w.r.t the surface itself. These variations further complicate the classification process.

included to maintain simplicity and considering the scenario when metadata might not be available.

As displayed in Figure 1, the bar plot representing absolute class frequencies, the distribution of snake species in the dataset exhibits a pronounced imbalance. Despite the considerable size of this dataset, it contains numerous challenging examples that make classification a daunting task. For instance, Figure 2 presents images of snakes demonstrating a significant variation in the surfaces on which they are found and some snake species are less distinguishable w.r.t the surface itself. These variations further complicate the classification process.

4. Metric-Weighted Ensemble Focal Loss

Focal loss and equalized focal loss are often used to handle imbalanced datasets because they are designed to give more weight to hard, misclassified examples and less weight to well-classified examples.

Focal loss was introduced in the paper Focal Loss for Dense Object Detection by Lin et al.,[32] as a way to address the class imbalance issue in object detection. It was designed to address the one-stage object detection scenario in which there is an extreme imbalance between foreground and background classes during training. The central idea of the focal loss is to reduce the contribution of well classified examples, allowing the training process to focus more on hard examples. This is achieved by adding a modulating factor $(1 - p_t)^\gamma$ to the cross entropy loss, with tunable focusing parameter $\gamma \geq 0$, that decreases the weight of well-classified examples.

The mathematical formulation of Focal Loss is -

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (1)$$

where $FL(p_t)$ presents Focal loss, α_t is a balancing factor (usually set to 0.25) that balances the importance of positive and negative samples ("positive" and "negative" samples are also often used to refer to true positive and false positive), γ is a focusing parameter (usually set to 2) and p_t is the model's estimated probability for the true class at observation t. In the observations mentioned in Lin et al.,[32], when an example is misclassified and p_t is small, the modulating factor is near 1 and the loss is unaffected. As $p_t \rightarrow 1$, the factor goes to 0 and the loss for well-classified examples is down-weighted. Also, the focusing parameter γ smoothly adjusts the rate at which easy examples are downweighted.

Another variation of Focal loss, Equalized focal loss [33], on the other hand, was designed to address another imbalance, which is the within-class sample imbalance, a common issue in long-tailed classification problems. In a multi-class classification setting, applying a single modulating factor uniformly across all classes (as done in focal loss) might not be the best approach, especially when there is a large class imbalance. This is because the modulating factor does not take into account the relative frequency of the classes in the dataset. In this type of problem, some classes have a lot more examples than others. Equalized focal loss tries to mitigate the issue by estimating class-wise weights to balance the influence of each class. This makes the model focus more on the under represented classes. The Equalized focal loss for the j-th category as-

$$EFL(p_t) = -\alpha_t(1 - p_t)^{\gamma^j} \log(p_t) \quad (2)$$

Here, α_t and p_t are identical to those in the original focal loss. The parameter γ^j acts as the focusing factor for the j-th category, paralleling the role of γ in traditional focal loss. As mentioned in Lin et al.,[32], varying values of γ correspond to different degrees of imbalance between positive and negative instances. For rare categories positive-negative imbalance, a larger γ^j is employed. In cases of categories with minor imbalance, a smaller γ^j is suitable. In equalised focal loss [33], the focusing factor γ^j is decomposed into two parts, namely, a category-agnostic parameter γ_b and a category-specific parameter γ_v^j :

$$\gamma_j = \gamma_b + \gamma_v^j \quad (3)$$

Here, γ_b signifies the focusing factor in a balanced data scenario, regulating the fundamental behavior of the classifier. On the other hand, $\gamma_v^j \geq 0$ is a variable parameter tied to the imbalance degree of the j-th category. It dictates the extent to which the learning focus is on addressing positive-negative imbalance issues.

In this paper, both loss functions are utilized. As mentioned in Eq.(1), focal loss equally treats the learning of all categories with the same modulating factor, so it is element-wise multiplied with the weights obtained by the metric mentioned in [8, 9] for Metric Weighted Ensemble Focal Loss. The metric representation is -

$$L(y, \hat{y}) = \begin{cases} 1 & \text{if } y = \hat{y}, \\ 1.5 & \text{if } y \neq \hat{y} \text{ and } p(y) = 0 \text{ and } p(\hat{y}) = 0 \\ 2 & \text{if } y \neq \hat{y} \text{ and } p(y) = 0 \text{ and } p(\hat{y}) = 1 \\ 2 & \text{if } y \neq \hat{y} \text{ and } p(y) = 1 \text{ and } p(\hat{y}) = 1 \\ 5 & \text{if } y \neq \hat{y} \text{ and } p(y) = 1 \text{ and } p(\hat{y}) = 0 \end{cases} \quad (4)$$

where p is a function and $p(s)=1$ if species is venomous, $p(s)=0$, when it is non-venomous, y is target class and \hat{y} is predicted class. Note that there is a small change done on the original *orig* - $L(y, \hat{y})$ metric for the purpose of using it as a weight. After the calculation the final formulation of the proposed Metric-Weighted Ensemble Focal Loss is -

$$MWEFL = EFL(p_t) + L(y, \hat{y}) \odot FL(p_t) \quad (5)$$

where MWEFL, Metric-Weighted Ensemble Focal Loss.

5. Sampling Strategy

In a normal random sampling scenario, every data point in the dataset has an equal chance of being included in a batch. This can cause problems if the dataset is imbalanced, which is our case, meaning some classes have more samples than others. If a model is trained on these batches, it might perform poorly on the under represented classes simply because it didn't see enough of those examples during training.

This is where WeightedRandomSampler [34] is used in this paper. It assigns each data point in the dataset a weight. The probability that a data point gets included in a batch is proportional to its weight. The idea of weighted random sampling is a general statistical technique which uses multinomial distribution for sampling and is not tied to a specific paper or proposal. That said, many papers in machine learning that deal with imbalanced datasets will make use of weighted sampling or similar techniques e.g., Learning from Imbalanced Data by He and Garcia [35]. Please note that paper, Learning from Imbalanced Data [35], discusses a variety of techniques, not just weighted sampling.

The mathematical steps for the WeightedRandomSampler as mentioned in pytorch are-

- Normalize the weights to create a discrete probability distribution-

$$P(i) = \frac{w_i}{\sum_j w_j} \quad \forall j \in \text{dataset} \quad (6)$$

- Then, draw a sample index i from the discrete distribution:

$$i \sim P(i) \quad (7)$$

- The final output is a sequence of indices, with a total length equal to to the sampled data points. It's worth noting that because this sampling process is inherently probabilistic, different runs will generally yield different results unless a consistent random seed is used.

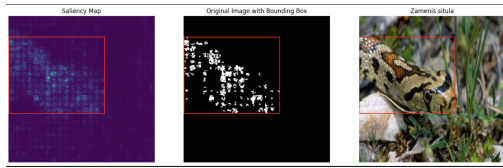


Figure 3: Saliency Object Detection - DEiT model

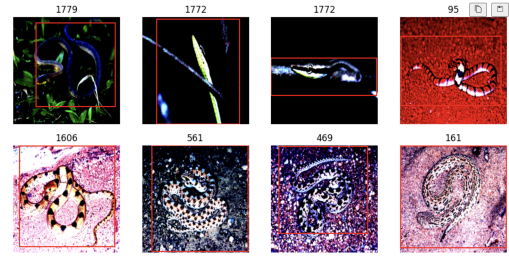


Figure 4: BoundingBox on Augmented Data

6. Implementation

6.1. Object Detection

Before feeding the images into the model, they are pre-processed using a saliency-based object detection method proposed by Rail Chamidullin in 2022, [36, 37]. Given the challenging nature of the dataset, where differentiating a snake from its surrounding environment can be arduous, this process proves beneficial. By eliminating irrelevant sections of the image, the model is more likely to focus on the features of the snake, thus enhancing its performance. Figures 3 and 4 showcase some of the outputs generated by this saliency-based object detection method [36, 37].

6.2. Exploring Different Architectural Designs Through Experimentation

In order to address this imbalanced data classification task, a series of experiments were conducted. These experiments included DEiT [38], EfficientNets [39], and DinoV2 [40]. Among these, DEiT [38] emerged as the most proficient when using an input image size of 384×384 . An exploratory analysis involving DinoV2 [40] was also conducted, primarily focusing on the PCA [41] features, as exhibited in Figure 5.

These figures demonstrate the model's adeptness at learning accurate features. Additionally, DinoV2 [40] was trained for a single epoch, a process that took approximately two and a half hours, yet yielded a commendable macro F1 score at first pass than other experimented models. This performance surpassed the first epoch results of other models. However, due to the substantial time consumption, this model was not further pursued in the current study. Nonetheless, its promising initial results mark it as a potentially worthwhile experiment for future research. These models were trained using the proposed loss function. Individually focal loss [32] and equalized focal loss [33] were also being utilized for training experiments.

The GPUs utilized for this research were the A6000 48 GiB GPU with 45 GiB RAM from Paperspace and the A100 from Colab. It is significant to mention that due to limited resources, the number of training epochs was kept under 30. Apart from the Metric Weighted Ensemble Focal Loss (MWEFL), other loss functions such as the Equalized Focal Loss [33](eFocal) were also explored.

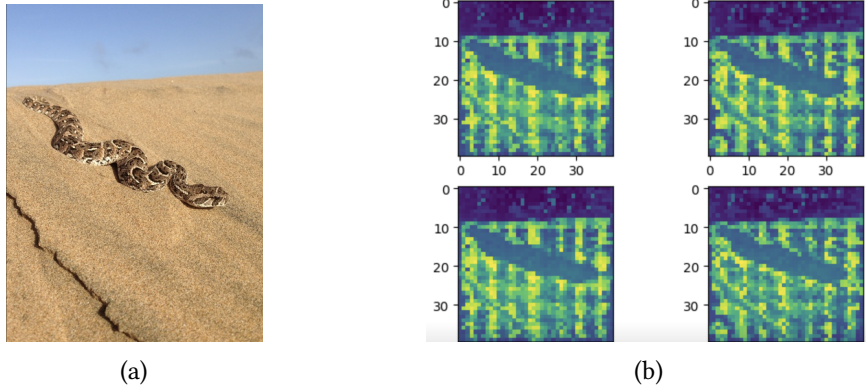


Figure 5: Fig (a) represents Original image of snake on which Dinov2 is executed for feature analysis. Fig (b) represents PCA features components extracted from Dinov2 model exhibiting features learned from the original image on the left side. Here, Dinov2 model feature extraction PCA analysis exhibits well learned spacial and texture features of the original snake image.

Table 1

Table presents training loss, validation loss, F1 scores (macro), top-3 score of training different models (with pre-trained weights) with proposed loss and individual losses

| Model | loss | train loss | val loss | snake score | mac-f1 score | top 3 |
|------------------|--------|------------|----------|-------------|--------------|-------|
| DEit | eFocal | 0.46 | 4.59 | - | 0.36 | 0.72 |
| DEit | MWEFL | 0.64 | 8.06 | 7165 | 0.35 | 0.72 |
| DinoV2 | Focal | 2.04 | 3.44 | 15666 | 0.15 | 0.45 |
| EfficientNet(b0) | Focal | 0.03 | 3.18 | - | 0.24 | 0.55 |

7. Results

Table 1 presents the performance comparison of various model architectures used in the experiments. Each model is evaluated on three distinct metrics – F1 score (macro), snake score (a custom metric employed as a weight in our MWEFL loss function and introduced in the challenge [8, 9]), and top-3 accuracy. The tabulated results provide a clear understanding of the strengths and weaknesses of each architecture in tackling the given classification task. Despite the fact that the highest macro F1-score was achieved by the DEiT model trained using eFocal loss (F1-score (macro) 0.36), the DEiT model trained with MWEFL (F1-score (macro) 0.35) outperformed the others in terms of the lowest snake score.

8. Conclusion and Future Work

The research conducted in this study involved numerous experiments, which could be further improved with access to a larger and more balanced dataset and by running more epochs. Specifically, for those species with a smaller number of samples, one could artificially boost their representation through offline augmentation techniques.

Another promising direction for future work could be to treat this problem as a similarity search task. In this approach, we could construct a database of embeddings for different appearances of all species. Then, for a new image, we calculate the similarity score between the new image and the images in the database. This method could potentially make the model more robust and improve its accuracy by leveraging a more comprehensive understanding of the variability in appearances among species.

9. Acknowledgments

I would like to extend my sincere thanks to the organizing team for providing the private evaluation scores and reviewing this working note.

References

- [1] WHO Team, Who statistics on ophidian envenomation, 2021. URL: <https://www.who.int/news-room/fact-sheets/detail/snakebite-envenoming>.
- [2] M. F. Francis, S. J.-M. Vianney, K. Heitz-Tokpa, K. Kreppel, Risks of snakebite and challenges to seeking and providing treatment for agro-pastoral communities in tanzania, *PLoS one* 18 (2023) e0280836.
- [3] LifeCLEF, Lifeclef-snakeclef, 2023. URL: <https://www.imageclef.org/LifeCLEF2023>.
- [4] A. Joly, H. Goëau, S. Kahl, L. Picek, C. Botella, D. Marcos, M. Šulc, M. Hruží, T. Lorieul, S. S. Moussi, M. Servajean, B. Kellenberger, E. Cole, A. Durso, H. Glotin, R. Planqué, W.-P. Vellinga, H. Klinck, T. Denton, I. Eggel, P. Bonnet, H. Müller, Lifeclef 2023 teaser: Species identification and prediction challenges, in: J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, A. Caputo (Eds.), *Advances in Information Retrieval*, Springer Nature Switzerland, Cham, 2023, pp. 568–576.
- [5] L. Picek, M. Šulc, R. Chamidullin, A. M. Durso, Overview of snakeclef 2023: Snake identification in medically important scenarios, in: *CLEF 2023-Conference and Labs of the Evaluation Forum*, 2023.
- [6] A. Joly, H. Goëau, S. Kahl, B. Deneu, M. Servajean, E. Cole, L. Picek, R. Ruiz de Castañeda, I. Bolon, A. Durso, et al., Overview of lifeclef 2020: a system-oriented evaluation of automated species identification and species distribution prediction, in: *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 11th International Conference of the CLEF Association*, CLEF 2020, Thessaloniki, Greece, September 22–25, 2020, *Proceedings*, Springer, 2020, pp. 342–363.
- [7] A. Joly, C. Botella, L. Picek, S. Kahl, H. Goëau, B. Deneu, D. Marcos, J. Estopinan, R. Chamidullin, M. Šulc, M. Hruží, 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.
- [8] LifeCLEF-HF, Lifeclef snakeclef hf, 2023. URL: <https://huggingface.co/spaces/competitions/SnakeCLEF2023>.

- [9] LifeCLEF, snakeclef2023, 2023. URL: <https://www.imageclef.org/SnakeCLEF2023>.
- [10] I. Bolon, A. M. Durso, S. Botero Mesa, N. Ray, G. Alcoba, F. Chappuis, R. Ruiz de Castañeda, Identifying the snake: First scoping review on practices of communities and healthcare providers confronted with snakebite across the world, *PLoS one* 15 (2020) e0229989.
- [11] R. R. de Castañeda, A. M. Durso, N. Ray, J. L. Fernández, D. J. Williams, G. Alcoba, F. Chappuis, M. Salathé, I. Bolon, Snakebite and snake identification: empowering neglected communities and health-care providers with ai, *The Lancet Digital Health* 1 (2019) e202–e203.
- [12] L. Picek, I. Bolon, A. M. Durso, R. R. de Castañeda, Overview of the snakeclef 2020: Automatic snake species identification challenge, *CEUR Workshop Proceedings*, 2020.
- [13] L. Picek, A. M. Durso, I. Bolon, R. R. de Castañeda, Overview of snakeclef 2021: Automatic snake species identification with country-level focus (2021).
- [14] L. Picek, M. Hruz, A. M. Durso, I. Bolon, Overview of snakeclef 2022: Automated snake species identification on a global scale (2022).
- [15] A. P. James, B. Mathews, S. Sugathan, D. K. Raveendran, Discriminative histogram taxonomy features for snake species identification, *Human-Centric Computing and Information Sciences* 4 (2014) 1–11.
- [16] A. Amir, N. A. H. Zahri, N. Yaakob, R. B. Ahmad, Image classification for snake species using machine learning techniques, in: *Computational Intelligence in Information Systems: Proceedings of the Computational Intelligence in Information Systems Conference (CIIS 2016)*, Springer, 2017, pp. 52–59.
- [17] A. Patel, L. Cheung, N. Khatod, I. Matijosaitiene, A. Arteaga, J. W. Gilkey Jr, Revealing the unknown: real-time recognition of galápagos snake species using deep learning, *Animals* 10 (2020) 806.
- [18] M. G. Krishnan, Impact of pretrained networks for snake species classification., in: *CLEF (Working Notes)*, 2020.
- [19] M. Vasmatkar, I. Zare, P. Kumbha, S. Pimpalkar, A. Sharma, Snake species identification and recognition, in: *2020 IEEE Bombay Section Signature Conference (IBSSC)*, IEEE, 2020, pp. 1–5.
- [20] I. S. Abdurrazaq, S. Suyanto, D. Q. Utama, Image-based classification of snake species using convolutional neural network, in: *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, IEEE, 2019, pp. 97–102.
- [21] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, *Advances in neural information processing systems* 28 (2015).
- [22] 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.
- [23] C. Abeyasinghe, A. Welivita, I. Perera, Snake image classification using siamese networks, in: *Proceedings of the 3rd International Conference on Graphics and Signal Processing*, 2019, pp. 8–12.
- [24] A. He, C. Luo, X. Tian, W. Zeng, A twofold siamese network for real-time object tracking, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4834–4843.
- [25] I. Bolon, L. Picek, A. M. Durso, G. Alcoba, F. Chappuis, R. Ruiz de Castañeda, An artificial

- intelligence model to identify snakes from across the world: Opportunities and challenges for global health and herpetology, *PLoS neglected tropical diseases* 16 (2022) e0010647.
- [26] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, *arXiv preprint arXiv:2010.11929* (2020).
 - [27] Z. Yang, R. Sinnott, Snake detection and classification using deep learning (2021).
 - [28] L. Bloch, A. Boketta, C. Keibel, E. Mense, A. Michailutschenko, O. Pelka, J. Rückert, L. Willemeit, C. M. Friedrich, Combination of image and location information for snake species identification using object detection and efficientnets., in: *CLEF (Working Notes)*, 2020.
 - [29] Y. Shen, X. Sun, Z. Zhu, When large kernel meets vision transformer: A solution for snakeclef & fungiclef, *Working Notes of CLEF* (2022).
 - [30] M.-H. Guo, C.-Z. Lu, Z.-N. Liu, M.-M. Cheng, S.-M. Hu, Visual attention network, *arXiv preprint arXiv:2202.09741* (2022).
 - [31] Z. Dai, H. Liu, Q. V. Le, M. Tan, Coatnet: Marrying convolution and attention for all data sizes, *Advances in Neural Information Processing Systems* 34 (2021) 3965–3977.
 - [32] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
 - [33] B. Li, Y. Yao, J. Tan, G. Zhang, F. Yu, J. Lu, Y. Luo, Equalized focal loss for dense long-tailed object detection, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 6990–6999.
 - [34] pytorch2016, Weightedrandomsampling, Unknown year. URL: <https://pytorch.org/docs/stable/data.html#torch.utils.data.WeightedRandomSampler>.
 - [35] H. He, E. A. Garcia, Learning from imbalanced data, *IEEE Transactions on knowledge and data engineering* 21 (2009) 1263–1284.
 - [36] Rail Chamidullin , objectdetecionsaliency, 2022. URL: https://dspace.cvut.cz/bitstream/handle/10467/99095/F3-DP-2022-Chamidullin-Rail-Master_Thesis.pdf.
 - [37] LifeCLEF, Fine grained object detection saliency method, 2021. URL: <https://github.com/chamidullinr/fine-grained-visual-recognition>.
 - [38] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, H. Jégou, Training data-efficient image transformers & distillation through attention, in: *International conference on machine learning*, PMLR, 2021, pp. 10347–10357.
 - [39] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: *International conference on machine learning*, PMLR, 2019, pp. 6105–6114.
 - [40] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, et al., Dinov2: Learning robust visual features without supervision, *arXiv preprint arXiv:2304.07193* (2023).
 - [41] H. Abdi, L. J. Williams, Principal component analysis, *Wiley interdisciplinary reviews: computational statistics* 2 (2010) 433–459.