Do Not Lose to Losses for SnakeCLEF2024

Matěj Sieber[1], Tomáš Železný[1]

[1]*University of West Bohemia, Faculty of Applied Sciences, Univerzitni 2732/8, 301 00 Pilsen, Czech Republic*

Abstract

This paper presents participation in the SnakeCLEF 2024 challenge, which aims to automate the identification of snake species. We explore various custom loss functions that incorporate the venomousness of snakes. These loss functions are used to train the Swin-v2 tiny model with same training specification as baseline solution to accurately measure the impact of custom loss functions. Swin-v2 tiny model is beneficial due to its low computational demand and opens the possibility for use in handheld devices. Our results show that the best approach for maximising performance on the custom competition metrics is to apply a soft target set according to the venomousness of the snake. The best accuracy is achieved by the model trained with loss, which weights the different classes according to the number of their instances.

Keywords

SnakeCLEF, Snake Bite, Computer Vision, Classification, Snake Species Identification, Imbalanced dataset

1. Introduction

An estimated 5.4 million people worldwide are bitten by snakes each year, resulting in 1.8 to 2.7 million cases of envenomings. This leads to approximately 81,410 to 137,880 deaths annually, with three times as many cases of amputations and other permanent disabilities. Venomous snakebites can cause severe health issues, including paralysis that can inhibit breathing, bleeding disorders leading to fatal hemorrhage, irreversible kidney failure, and tissue damage resulting in permanent disability and limb amputation [1]. To address these challenges, developing accurate and lightweight models for identifying snake species could significantly improve health outcomes. The LifeCLEF-SnakeCLEF 2024 [2, 3] makes it possible to train such models by providing the training data. In addition to the classic metrics used for classification, the competition evaluates on custom metrics that take into account the venomousness of snakes, which is a crucial in a real scenario. This work represents participation in SnakeCLEF2024 with a focus on minimizing custom venomous metrics.

2. Dataset

The SnakeCLEF dataset [4] is a comprehensive collection of snake images used for the classification of snake species. The dataset consists of three parts: training, validation, and a private test set used for competition evaluation. The training and validation sets are publicly available, while the private test set is held back for evaluating the competition entries. The dataset is available in multiple variants, differing in the size of the images, to accommodate various computational capabilities and research needs. As stated in the SnakeCLEF2023 report [5], all subsets combined result in roughly 110,000 real snake observations with community-verified species labels, ensuring high-quality and reliable data for training models. Dataset contains 1,784 snake species, the classes exhibit a long-tailed distribution, meaning that a small number of classes have a large number of images, while many classes have only a few images. Figure 1 shows an illustrative representation of medically important snake species.

Figure 1: Medically-important venomous snakes. Top row: Daboia russelii (Russell's Viper – Asia), Bitis arietans (Puff Adder – Africa), and Crotalus adamanteus (Eastern Diamond-backed Rattlesnake – North America). Bottom row: Bothrops atrox (Common Lancehead – South America), Acanthophis antarcticus (Common Death Adder – Australia), and Vipera ammodytes (Nose-horned Viper – Europe). Courtesy of [5].

3. Evaluation

The competition evaluates competitors' models using four different metrics. The first two are the macro-averaged F1 score and accuracy, which are the standard metrics for classification tasks. In the real world, however, misclassifying different snakes does not have the same consequences. In the worst case scenario, a deadly venomous snake is misclassified as a harmless one, which may result in death. With this in mind, the organisers came up with two other metrics, that take into account whether the snake is venomous or not. These metrics are denoted in Equation (2) and (3). In the real world, however, different venomous snake bites cannot be treated in the same way, as some are more venomous than others and the bite requires different types of serum, which may vary in side-effects, price or location. As it would be very expensive to achieve this level of granularity, a generalisation was made in the form of a universal, harmless, free and always-accessible antivenom.

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

$$L = \sum_i L(y_i, \hat{y}_i), \tag{2}$$

where correct species is y and predicted species is $\hat{y}$. $P(s) = 1$ if species s is venomous, otherwise $p(s) = 0$.

$$M = \frac{w_1 F_1 + w_2 C_{h \to h} + w_3 C_{h \to v} + w_4 C_{v \to v} + w_5 C_{v \to h}}{w_1 + w_2 + w_3 + w_4 + w_5}, \tag{3}$$

where $w_1 = 1$, $w_2 = 1$, $w_3 = 2$, $w_4 = 2$, and $w_5 = 5$ are the weights of individual confusions,
$C_{v \to h}$ is the percentage of wrongly classified venomous species as a harmless species,
$C_{h \to v}$ is the percentage of wrongly classified harmless species as a venomous species,
$C_{v \to v}$ is the percentage of wrongly classified venomous species as another venomous species,
$C_{h \to h}$ is the percentage of wrongly classified harmless species as another harmless species,
and the F1 is the macro averaged F1 score.

4. Experiments

Given the practical limitations and the need for widespread accessibility, we aimed to develop a model that could run on handheld devices such as smartphones and tablets. Real-time identification of snake species has a potential to significantly improve the speed and effectiveness of medical response, potentially saving lives and reducing the incidence of serious snakebite complications. This capability is particularly important in remote areas where snakebite is most common and access to high performance computing resources is limited.

We use the Swin-v2 tiny model [6], which is suitable for this task due to its small size and efficiency. Such lightweight models are less prone to overfitting, which is particularly important when dealing with diverse and unbalanced datasets.

We aim to maximise the performance of the model for those metrics that take into account whether the snake is venomous or not. Primarily, we focus on minimizing the L metric (Equation 2), as this closely matches real-world scenarios where accurate identification of snake species is paramount. The main goal of our work is to optimise the loss functions. Specifically, we created four different custom losses to meet our objectives. All results are presented in Table 1.

To effectively measure impact of custom losses we use the same training parameters as the baseline: `RandResizedCrop` and `RandAugment` augmentations, resolution size 256x256, learning rate 0.01 and SGD optimizer. Full training pipeline for the baseline solution is available at BVRA GitHub[1]. Code for proposed methods can be found at Our GitHub[2].

4.1. Dual-head

The aim of this experiment is to improve the performance of the model by incorporating snake venom information using a combination of two classification losses. We add a second head consisting of one neuron with Sigmoid activation function. In addition to the Categorical Cross Entropy loss, we also train the model on the binary classification of venomous/harmless classes using Binary Cross Entropy loss, resulting in equation denoted in 4.

$$\mathcal{L}_{\text{Dual-head}} = \mathcal{L}_{\text{BCE}} + \mathcal{L}_{\text{CE}} \tag{4}$$

4.2. Rare class boost

Given the nature of the data, there are several strategies to address a long-tailed distribution, such as uneven sampling of training data or assigning greater weight to rarer classes. Our approach utilizes the weight strategy, which demonstrates superior accuracy, aside from ensemble models. In this experiment, we compute the loss with SeeSawLoss [7], multiplying it by the rarity of each class. Although SeeSawLoss was originally developed for long-tailed distribution data, our modification of incorporating dynamic class rarity across the whole dataset further improves the results. These results can be directly compared to the Baseline solution, which also utilizes SeeSawLoss.

$$S_{\text{batch}} = \sum_{i=1}^{B} C_{k_i},$$

$$\beta = \frac{\frac{N}{B}}{S_{\text{batch}}},$$

$$\mathcal{L}_{\text{ClsBoost}} = \beta \times \mathcal{L}_{\text{SeeSaw}}(\mathbf{x}, \mathbf{y}), \tag{5}$$

where C_k is number of instances of the class k, N is the dataset size, and B is the batch size.

4.3. CE + VenomousPenalty

Another approach to take the venom of the snake into account is to simply add a venomous penalty, according to the equation 1, to the classical Categoric Cross Entropy loss. Although this method does not perform notably better than the baseline method, it achieves the best results in the F1 metric.

$$\mathcal{L}_{CE+VP} = \mathcal{L}_{CE} + L, \tag{6}$$

where L is penalty defined in (1).

4.4. Soft target

Instead of classical approach of setting a Cross Entropy target as one-hot vector, we explore using a soft target. In this method, our goal is to set the negative targets to values accordingly to venomous penalties in Equation (1), while ensuring that the sum of the target values remains 1. First, we linearly transform these penalties using Equation (7). This results in a target value of 1 for the least penalized classification, i.e., $(y = \hat{y})$, and a target value of 0 for the most penalized classification, i.e., $(y \neq \hat{y}$ and the venomous snake is classified as harmless). The values are then normalised and used as a soft target.

$$\tau = -0.2 \cdot p + 1, \tag{7}$$
$$\text{Targets} = \text{Norm}(\tau), \tag{8}$$

where p are the penalties in Equation (1), and τ are the targets before normalisation.

This method results in poor performance, as the positive and negative target values are close after normalisation. This motivates us to create a soft target method which sets the positive target to significantly higher value than the negative targets. Our aim is to set the positive target value in the range $\langle 0.5; 1.0 \rangle$. We conducted empirical experiments, which resulted in the target denoted in Equation (9). Models with best performance are denoted as SoftT-3 for temperature parameter $t = 3$ and SoftT-4 for $t = 4$.

$$\text{Target} = -t \cdot \text{Softmax}(\log(p)), \tag{9}$$
$$\text{where } p = 0.1 \text{ for } y = \hat{y},$$
$$p = 1 \text{ for } y \neq \hat{y} \text{ and } p(y) = 0 \text{ and } p(\hat{y}) = 0,$$
$$p = 10 \text{ for } y \neq \hat{y} \text{ and } p(y) = 0 \text{ and } p(\hat{y}) = 1,$$
$$p = 10 \text{ for } y \neq \hat{y} \text{ and } p(y) = 1 \text{ and } p(\hat{y}) = 1,$$
$$p = 100 \text{ for } y \neq \hat{y} \text{ and } p(y) = 1 \text{ and } p(\hat{y}) = 0,$$
$$\text{and } t \text{ is temperature parameter.}$$

4.5. Model Ensemble

The competition sets a limit of 60 minutes for the maximum model inference time on the test set. Since our model is able to process the test set in units of minutes, we decided to use the remaining time for additional experiments. We created an ensemble of our models by averaging the logits. The ensemble of models performed noticeably better. Since there was still a lot of time left, we also tested the ensemble of logits for the given image and its horizontally flipped version for each model as the flip was not part of the augmentations. This doubled our inference time. However, we did not gain any improvement by using this method, so we do not report it.

Table 1

Comparison of results on the public dataset for proposed methods and baseline solution. CE refers to the same method as Baseline, but only trained with the Cross Entropy loss instead of the SeeSaw loss. The ensembles of the models are presented separately.

Method	M metric	L metric	F1	Accuracy
Competition Baseline	67.01	1861	13.34	39.88
CE	62.33	2195	11.69	33.19
Dual-head	67.77	1813	14.56	41.87
ClsBoost	68.04	1788	13.87	**42.75**
CE + VenomousPenalty (CE-VP)	67.10	1863	**14.58**	39.51
SoftT linear	64.62	2024	11.13	36.72
SoftT-4	65.01	2004	12.55	35.54
SoftT-3	**68.15**	**1785**	14.41	41.94
Ensemble SoftT-3 + ClsBoost	**69.92**	**1660**	15.44	**45.77**
Ensemble SoftT-3 + SoftT-4 + CE-VP + ClsBoost	69.88	1668	**16.12**	45.11

Table 2

Comparison of results on the private dataset. The ensembles of the models are presented separately.

Method	M metric	L metric	F1	Accuracy
Competition Baseline	64.30	5067	10.9	36.41
Dual-head	64.94	5004	12.2	37.77
ClsBoost	65.26	4921	11.4	**38.09**
CE + VenomousPenalty (CE-VP)	65.33	**4860**	11.75	36.1
SoftT linear	62.08	5472	10.94	34.53
SoftT-4	64.06	5075	10.39	35.12
SoftT-3	**65.37**	4901	**13.17**	37.77
Ensemble SoftT-3 + ClsBoost	67.00	4611	13.29	41.26
Ensemble SoftT-3 + SoftT-4 + CE-VP + ClsBoost	**67.47**	**4515**	**13.89**	**41.4**

5. Conclusion

This work presents a participation in the competition SnakeCLEF2024. Our approach is based on the use of the compact Swin-v2 tiny model, known for its speed and suitability for running on mobile devices such as smartphones and tablets. Instead of focusing on conventional methods, we decided to experiment exclusively with custom loss functions tailored to our specific scenario. In particular, we focused on the L metric, which is designed to penalise misclassification of snakes based on their venomousness.

The results (see Table 1 and 2) show that the Dual-head approach improved results compared to the baseline solution. And these improvements were stable on both datasets public and private alike. The ClsBoost loss was a viable idea that maintained the best accuracy on both the public and private datasets. Since the loss function did not focus on the custom metrics, but rather aimed to reduce the effect of the long-tailed distribution, the accuracy, even when tied to other metrics, was not sufficient to maintain the best custom metric score. The CE-VP loss function proved that incorporating the L metric into the loss function helped. However, there are notable differences between the performance on the public and test set. The SoftT loss function achieved the best results, namely the M and L metrics on the public dataset and the M and F1 metrics on the private dataset. Since this method uses a hyperparameter chosen by empirical study, this gives the opportunity for future work where the hyperparameter could be tuned within an ablation study with the aim of finding the best performing parameters of this method.

Since the competition allows a maximum inference time of 60 minutes, and our model requires only a few minutes to infer the entire test set, we decided to create an ensemble of our best models, resulting in better scores.

As an extension to our methods, we propose to use location data in the recognition process. By using GPS information, which is available on handheld devices such as smartphones, we can improve the

accuracy of species identification by taking into account the geographical distribution of different snake species.

Acknowledgments

References

[1] World Health Organization, Snakebite envenoming, 2023. Https://www.who.int/news-room/fact-sheets/detail/snakebite-envenoming.

[2] A. Joly, L. Picek, S. Kahl, H. Goëau, V. Espitalier, C. Botella, B. Deneu, D. Marcos, J. Estopinan, C. Leblanc, T. Larcher, M. Šulc, M. Hrúz, M. Servajean, et al., Overview of lifeclef 2024: Challenges on species distribution prediction and identification, in: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, 2024.

[3] L. Picek, M. Hruz, A. M. Durso, Overview of SnakeCLEF 2024: Revisiting snake species identification in medically important scenarios, in: Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, 2024.

[4] LifeCLEF, Snakeclef2024, 2024. Https://huggingface.co/spaces/BVRA/SnakeCLEF2024.

[5] L. Picek, M. Šulc, R. Chamidullin, A. Durso, Overview of snakeclef 2023: snake identification in medically important scenarios, CLEF, 2023.

[6] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, B. Guo, Swin transformer v2: Scaling up capacity and resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 12009–12019.

[7] J. Wang, W. Zhang, Y. Zang, Y. Cao, J. Pang, T. Gong, K. Chen, Z. Liu, C. C. Loy, D. Lin, Seesaw loss for long-tailed instance segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 9695–9704.