UAIC-AI at SnakeCLEF 2021: Impact of convolutions in snake species recognition

Lucia Georgiana Coca, Alexia Theodora Popa, Razvan Contantin Croitoru, Luciana Paraschiva Bejan and Adrian Iftene

"Alexandru Ioan Cuza" University, Faculty of Computer Science, Iasi, Romania

Abstract
Snake identification is crucial in quickly and effectively treating snake bites. With over 2.7 million snake envenomings happening yearly, medical personal are in desperate need of tools that will ease their work as well as save patient lives faster. SnakeCLEF 2021 challenge, which is part of the LifeCLEF laboratory, has exactly the aforementioned goal. This paper presents our team participation at SnakeCLEF 2021. We developed 3 models based on CNN: **GoogLeNet**, **VGG16** and **ResNet-18** and ranked 5^{th} with an F1-Country score of 0.785 using ResNet-18.

Keywords
LifeCLEF, SnakeCLEF, Snake Identification, snake bite, health, CNN, Machine Learning, Snakes

1. Introduction

Human expansion and the amplification of animal habitat destruction is resulting in more and more contact with wildlife in urban areas. Snakes are one species affected by this phenomena, if previously they lived in forests, swamps or even deserts, they are now forced to crawl in urban areas in search of food and shelter. The negative impact is felt not only in animals but also humans, more and more are we in contact with venomous snakes which leads to deadly scenarios that put us on a knife-edge. Being able to quickly identify the species of snakes that came in contact with people will not only give medical personnel precious time but also will give us a better understanding of certain species of snakes and their mobility in this new human habitat.

Annually, according to WHO [1], over 5.4 million people are bitten by snakes (2.7 million are envenomings), of which around 81,000 to 138,000 are deaths not taking into account the many that become disabled or paralyzed. It is hence critical that medical personnel quickly identify the species of snake in order to administer the correct antivenom.

Manual identification is no easy feat, there are more than 3,500 species of snakes, 600 of which are venomous. Training doctors on each and every species is an impossible task, it would be not only time consuming but also very costly. Recent years have brought an exponential growth in A.I. research, of which image recognition seems to be the most benefited. Advances and expansion of the global smartphone market combined with high Internet Penetration Rate in low-income and middle-income countries has lead to an information boom. People now have

CLEF 2021 – Conference and Labs of the Evaluation Forum, September 21–24, 2021, Bucharest, Romania
georgiana.coca@info.uaic.ro (L. G. Coca)

access to sufficient computing power at the tips of their fingers making it possible to classify in real-time almost anything. Snake identification and classification will improve the quality of lives beyond high-income countries and significantly improve epidemiology data and treatment outcomes.

The SnakeCLEF 2021 challenge [2] as part of the LifeCLEF [3] laboratory aims to solve the aforementioned problem by identifying species of snakes from photographs. They provided an image collection of 414,424 photographs belonging to 772 snake species and taken in 188 countries.

This paper describes the participation of team "UAIC-AI", from the Faculty of Computer Science, "Alexandru Ioan Cuza" University of Iasi, Romania, at SnakeCLEF 2021 where we ranked 5th with an F1-Country of 0.785. The remaining of this paper is organized as follows: Section 2 describes state-of-the-art methods in snake identification, Section 3 details the model we developed and the submitted runs and then Section 4 details the results we obtained, finally Section 5 concludes this paper and presents future work.

2. Related work

SnakeCLEF 2020 [4] is the previous edition of the same competition where multiple state-of-the-art systems have been presented. One such system is presented in [5] where the authors have developed a two stage preprocessing method; the first operation implies the transformation of rectangular images to squared ones followed by picture augmentation using location information, which helped them in the snake images recognition(as many species are spatially bound). The image classification algorithm is represented by a family of EfficientNet models [6] that have been extended with a flattening layer, a dense layer with 1000 neurons, a Swish [7] activation function and a dense layer of 783 neurons corresponding to each snake species. Their result was good, ranking 2nd place in the competition with a F1 score of 0.4035.

Last year's best result however was obtain by "Gokuleloop" team [8]. They used a ResNet50-V2 [9], based on CNN architecture and trained the open source models on both ImageNet-1k and ImageNet-21k. The team was focused on domain-specific fine-tuning, experimenting with different pre-trained weights and performance impact. Location information, such as Country and Continent of the snake species, have also been integrated in the model with a final system being comprised of a ResNet-50-V2 architecture fine-tuned from ImageNet-21k weights and a naive probability weighting approach. Authors have found that integrating geographic data improves performance, achieving an F1 score of 0.625.

3. Methods

The developed methods, that address the SnakeCLEF 2021 challenge, are all based on convolutional neural networks. We used 3 models, **GoogLeNet** [10], **VGG16** [11] and **ResNet** [12], each with their own advantages and results. In this section we will take a look in the dataset that is comprised of 414,424 photos, analyze each model and discuss the experiments.

Figure 1: Age differences

Figure 2: Geographic difference

3.1. Dataset

In order to have a clearer picture of the competition, it is mandatory to compare this year's dataset with the previous one. The 2020 dataset was much smaller in comparison to 2021, 245,185 training images were provided split into 783 species comparing to this year where there are now 414,424 images and 772 species. This means that in 2020 there were approximately 313 images per species and in 2021 there are now 536 images, a two-fold increase in images which will nonetheless result in a higher accuracy of the models. Additional geographical metadata (country and continent) for the image is also provided.

Analyzing the dataset yields important information related to the collection. Figure 1 shows age variation between snakes whilst Figure 2 illustrates geographic variations, demonstrating that the dataset has a variety of scenarios.

3.2. GoogLeNet

GoogLeNet is a system that has shown to be powerful in many novel applications. In [13] they used GoogLeNet to remove streak artifacts due to projection missing in sparse-view CT reconstruction and found that the method is practical and reduces artifacts whilst preserving the quality of the image.

Our goal was to experiment with this method and build an architecture that has filters with multiple sizes that can operate on the same level. We kept the starting filters from GoogLeNet which used a (1 x 1) convolution because training is time-consuming and this method was used to compute reductions before the expensive (3 x 3), (5 x 5) convolutions and max-pooling.

The GoogLeNet architecture is 22 layers deep (and 5 pooling layers). These layers are grouped in 9 inception modules, and each of these is connected to the average pooling layer.

The GoogLeNet model used was implemented in PyTorch [14] and trained using Nvidia CUDA on GPU, a massive help was also Google Colab [15] which saved us a lot of time by training on their machines. We used a Cross Entropy Loss and Adam optimizer [16].

We started the training process with a 0.001 learning rate and batch size of 64. Learning rate was adjusted at each epoch (in total we had 10 epochs) until we concluded that 0.001 is the best rate. We also enabled "aux_logits" which adds two auxiliary branches that can improve training.

Table 1
ResNet dataframe example for training

Binomial name	Country	Continent	Genus	Family
Acanthophis antarcticus	Australia	Australia	Acanthophis	Elapidae
Agkistrodon conanti	United States of America	North America	Agkistrodon	Viperidae

3.3. VGG16

VGG16 is a well known model and although it is an older approach we wanted to compare it to the others as see how it performs, as in previous works [17] it showed potential. It is a 16 layer deep convolutional neural network with no residual blocks which improves on other models by replacing large kernel-sized filters with multiple 3×3 kernel-sized filters one after another.

We used the model implemented in Keras that had the pre-trained weights from ImageNet-1k. The model implemented uses Adam optimizer and the hyperparameters were tuned experimentally. Initially the learning rate was set at a base value of 0.001 then initially decreased to 0.0001. The learning decay function used was ReduceLROnPlateau with a factor of 0.6 and patience of 4 while monitoring the validation accuracy. The batch size set was 16 as this gave us a good compromise between training speed and resource consumption.

In order to prevent overfitting we used a Dropout Regularization [18] strategy during the training phase. In an effort to reduce loss, a dropout layer with a probability of 0.5 was added between the last 2 dense layers of the VGG16 model.

3.4. ResNet

ResNet is the third and best model used. A study done in [19] outlines the performance overhead ResNet has over GoogLeNet, there is an almost 15% accuracy difference between the two, in favor of ResNet. The novelty and the performance of the method has lead us to experimenting with the model in order to test it in relationship with other neural networks. As other architectures often face the vanishing gradient problem, ResNet came with a solution called "identity shortcut connection" that skips one or more layers and acts like a much shallower network.

For this task a ResNet18 (from PyTorch) architecture pre-trained on ImageNet-1k [17] was fine-tuned. We decided to keep the pre-trained weights only for the first three layers and freeze them. The final fully connected layer was reset and reshaped to 772 nodes so that it match the current number of classes in the dataset.

The model was trained with 14 epochs with batches of 32 and a learning rate of 0.0002. To prevent overfitting a dropout regularization strategy was chosen during the training phase. We also used Adam Optimizer and cross entropy loss. To reduce loss, inside the basic blocks of ResNet, a dropout with a probability of 0.5 was applied.

Additional information available for each image regarding the country were taken into account in the model. Table 1 illustrate a dataframe fed to the network in the training phase. We included the species name, country, continent, genus and family but also the image path (that is not presented in the table).

Table 2
Official evaluation results

Submission name	F1	F1-Country	Accuracy
uaic_ai_submission1	0.60	0.61	0.79
uaic_ai_submission2	0.21	0.29	0.51
uaic_ai_submission8	0.74	0.78	0.86

4. Evaluation and comparisons

In this section we will discuss the results of the algorithms. Table 2 shows the results of our models. It is clearly seen that the best submission was "uaic_ai_submission_8" with an F1-Country of 0.785 according to the organizers, this ranked us 5^{th}.

Submissions information:

- **uaic_ai_submission1**: This submission was computed using a model with ResNet18 architecture resembling the training techniques from section 3.4 without taking into account the additional country information.

- **uaic_ai_submission2**: This submission was computed with GoogLeNet, consisted by all the information from the 3.2 subsection.

- **uaic_ai_submission8**: This submission was computed using a model with ResNet18 architecture resembling the training techniques from the 3.4 section and also taking into account the additional country information.

The best submission is represented by ResNet18 with country information. Further analysis into the GoogLeNet and VGG16 have revealed that due to aging architecture their performance is limited in comparison to more modern approaches such as ResNet. Scarce hardware resources and limited time have lead to slow progress in increasing ResNet performance and this will be the object of future work. Another research direction aims to address the imbalance of species in the dataset as we noticed that some species have more images than others. We would like to use augmentations from the Albumentation Library [20] for species that have few images in the dataset. The idea is simple, modifications are being done to dataset picture, like change of contrast, saturation, hue or brightness, in order to increase the dataset. Another technique we would like to use is MixUp augmentation [21], the technique is simple and they describe it as follows "mixing up the features and their corresponding labels. Neural networks are prone to memorizing corrupt labels. MixUp relaxes this by combining different features with one another (same happens for the labels too) so that a network does not get overconfident about the relationship between the features and their labels" [21].

5. Conclusions and Future Work

In conclusion, this paper is focused on team UAIC_AI's participation at SnakeCLEF 2021. We had good results with 3 submissions, the best one ranking 5^{th} with an F1-Country of 0.785. For future work we would like to try much novel image algorithms as well as improve the score of the current methods.

Acknowledgments

Special thanks goes to A2 team. This work was supported by project REVERT (taRgeted thErapy for adVanced colorEctal canceR paTients), Grant Agreement number: 848098, H2020-SC1-BHC-2018-2020/H2020-SC1-2019-Two-Stage-RTD.

References

[1] WHO, Snakebite envenoming, www.who.int/news-room/fact-sheets/detail/snakebite-envenoming/, 2021.
[2] L. Picek, A. M. Durso, R. Ruiz De Castañeda, I. Bolon, Overview of snakeclef 2021: Automatic snake species identification with country-level focus, in: Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, 2021.
[3] A. Joly, H. Goëau, S. Kahl, L. Picek, T. Lorieul, E. Cole, B. Deneu, M. Servajean, R. Ruiz De Castañeda, I. Bolon, H. Glotin, R. Planqué, W.-P. Vellinga, A. Dorso, H. Klinck, T. Denton, I. Eggel, P. Bonnet, H. Müller, Overview of lifeclef 2021: a system-oriented evaluation of automated species identification and species distribution prediction, in: Proceedings of the Twelfth International Conference of the CLEF Association (CLEF 2021), 2021.
[4] L. Picek, R. Ruiz De Castaneda, A. M. Durso, P. Sharada, Overview of the snakeclef 2020: Automatic snake species identification challenge, in: Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, 2020.
[5] L. Bloch, A. Boketta, C. Keibel, E. Mense, Combination of image and location information for snake species identification using object detection and efficientnets, in: Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, 2020.
[6] M. Tan, Q. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, in: K. Chaudhuri, R. Salakhutdinov (Eds.), Proceedings of the 36th International Conference on Machine Learning, volume 97 of *Proceedings of Machine Learning Research*, PMLR, 2019, pp. 6105–6114. URL: http://proceedings.mlr.press/v97/tan19a.html.
[7] P. Ramachandran, B. Zoph, Q. V. Le, Searching for activation functions, 2017. arXiv:1710.05941.
[8] M. G. Krishnan, Impact of pretrained networks for snake species classification, in: Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, 2020.
[9] K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks, in: European conference on computer vision, Springer, 2016, pp. 630–645.
[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, CoRR abs/1409.4842 (2014). URL: http://arxiv.org/abs/1409.4842. arXiv:1409.4842.
[11] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2015. arXiv:1409.1556.
[12] 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.
[13] S. Xie, X. Zheng, Y. Chen, L. Xie, J. Liu, Y. Zhang, J. Yan, H. Zhu, Y. Hu, Artifact removal

using improved googlenet for sparse-view ct reconstruction, Scientific reports 8 (2018) 1–9.

[14] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, Pytorch: An imperative style, high-performance deep learning library, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems 32, Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[15] E. Bisong, Google colaboratory, in: Building Machine Learning and Deep Learning Models on Google Cloud Platform, Springer, 2019, pp. 59–64.

[16] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2017. arXiv:1412.6980.

[17] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., Imagenet large scale visual recognition challenge, International journal of computer vision 115 (2015) 211–252.

[18] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, Journal of Machine Learning Research 15 (2014) 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[19] R. U. Khan, X. Zhang, R. Kumar, Analysis of resnet and googlenet models for malware detection, Journal of Computer Virology and Hacking Techniques 15 (2019) 29–37.

[20] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, A. A. Kalinin, Albumentations: Fast and flexible image augmentations, Information 11 (2020). URL: https://www.mdpi.com/2078-2489/11/2/125. doi:10.3390/info11020125.

[21] F. Chollet, et al., Keras, https://keras.io, 2015.