Snake Species classification using Transfer learning Technique

Karthik D[1], Mirunalini P[1] and Jitesh Kumar[2]

[1]Department of Computer Science and Engineering, Sri Sivasubramaniya Nadar College of Engineering, India
[2]Department of Computer Science and Engineering, Sri Venkateshwara College of Engineering, India

Abstract

Transfer learning is a technique that helps to utilise the knowledge of previously trained machine learning models by extending them to solve any related problem. This technique is predominantly used when there is either a scarcity of computational resource or limited availability of labelled data. Categorizing snake at the species level can be instrumental in treatment of snake bites and clinical management. We propose a deep learning model based on transfer learning technique to build a snake species classifier that uses snake photographic images in combination with their geographic location. We have used the Inception ResNet V2 as a feature extractor, extracted the feature vector for each input image and concatenated it with geographic feature information. The concatenated features are classified using a lightweight gradient boost classifier.

Keywords

Transfer Learning, Inception ResNet, Gradient Boosting, Snake Species Classification, Metadata Inclusion

1. Introduction

Snake species identification is essential for biodiversity, conservation and global health. Millions of snake bites occur globally every year, half of which cause snakebite envenoming (SBE), killing people and disabling more in different regions across the globe [1]. Taxonomic identification of the species helps the healthcare providers to articulate the symptoms, responses of the treatment and antivenom efficacy and also aid in clinical management [2, 3]. Identification of the snake species is difficult because of similarity in appearance, situational stress and fear of potential danger [4]. An automatic system that helps in recognizing the snake species from the photographic image and geographic information can be paramount in overcoming the above problems. Hence, we propose an automated system based on transfer learning techniques that utlilizes pre-trained weights of the Inception ResNet V2 [5] to extract input image features. The extracted features, in combination with the geographic features, are classified using a LightGBM [6], a gradient boosting classifier.

The Inception ResNet V2 incorporates residual connections into the inception architecture to perform enhanced feature extraction from images. The Inception ResNet V2 is a convolution neural network which has 164 deep layers where multi-sized convolution filters are combined

CLEF 2021 – Conference and Labs of the Evaluation Forum, September 21–24, 2021, Bucharest, Romania
karthik19047@cse.ssn.edu.in (K. D); miruna@ssn.edu.in (M. P); 2018cse0716@svce.ac.in (J. Kumar)
0000-0001-6433-8842 (M. P)

by residual connections which not only avoids the degradation caused by the deep layers but also reduces training time. The knowledge acquired by the model by training on the ImageNet data set [7] is utilized through transfer learning as a feature extractor.

Gradient boosting [8] is a machine learning technique that can be used for supervised classification problems to produce a prediction model. It is an ensemble of weak prediction models, typically decision trees, known for its prediction speed and accuracy with large and complex data sets. It minimizes the overall prediction error by iteratively generating optimized new models based on the loss function of the previous model. After concatenating the representation vectors of the input images with the geographic information, we trained a lightweight gradient boost classifier to predict the snake species.

2. Dataset

As part of the LifeCLEF-2021 [9], an evaluation campaign aimed at data-oriented challenges related to the identification and prediction of biodiversity, SnakeCLEF-2021 [10] is an image-based snake identification task. For this challenge, a large data set with 414,424 RGB photographic images belonging to 772 distinct snake species, taken in 188 countries is provided. Additionally, geographic metadata comprising of country and continent information is provided to facilitate classification. The data set is split into a training subset with 347,406 images, and a validation sub-set with 38,601 image, both having the same class distribution. The data set is highly imbalanced with a heavy long-tailed distribution. The most frequent class is represented with 22,163 images while the least frequent class by a mere 10 images. A large number of classes in combination with a high intra-class variance (depicted in Figure 1 and low inter-class variance makes this an exigent machine learning classification task.

Figure 1: Four images of the Dispholidus Typus snake species with high visual deviation characterized by age and gender depicting an instance of high inter-class variance in the data set

3. Related Work

An investigation of the accuracy of five machine learning techniques — decision tree J48, nearest neighbors, k-nearest neighbors (k-NN), back-propagation neural network, and naive Bayes — for image-based snake species identification problem was performed in [11]. It revealed the efficacy of back-propagation neural networks which achieved a greater than 87% classification accuracy.

A Siamese network with three main components namely, twin network, similarity function and output neuron was proposed in [12] to classify the snake species. A pair of deep neural networks was proposed where one network extracts features from the test image while the other from a reference image. The features were compared using L1 distance similarity and the final output layer predicted the probability of the test image belonging to same class as the reference image.

Four different region-based convolution neural networks (R-CNN) architectures - Inception V2, MobileNet, ResNet and VGG16 were used in [13] for object detection and image recognition of 9 snake species of the Pseudalsophis genus. Among them, VGG16 and ResNet achieved the highest accuracy of 75%.

A detailed quantitative comparative study between a computer vision algorithm trained to identify 45 species and human experts was performed in [14]. The algorithm used an EfficientNet based model, fine-tuned using preprocessed images to achieve an accuracy between 72 and 87% depending on the test data set. The significant impact of geographic data in addition to visual information for snake species classification was also realized.

4. Methodology

A transfer learning method is adopted to classify the snake species using the data set of snake images and geographic location metadata provided by SnakeCLEF-2021 [9, 10]. The pre-trained Inception ResNet V2, a deep learning convolution neural network is used to extract image features. These features are concatenated with the categorical geographic features and finally classified using a gradient boost classifier.

4.1. Preprocessing

The input images were resized to $299 \times 299 \times 3$ using bi-linear interpolation. To counter the effect of irrelevant factors in the context of the required task such as variation in lighting conditions among the photographs, the images were linearly normalized to values between 0 and 1.

Scale and rotation transformations, along with contrast and saturation variations were performed to make the model more generic, immune to the impact of positional and orientation based features and prevent memorization by enhancing image diversity. RandAugment [15] was used to augment the input images using the aforementioned transformations. RandAugment is parameterized by two values - the number of augmentation transformations to apply sequentially (N), and the magnitude for all the transformations (M). The values used in [15] for the ResNet model i.e N=3 and M=4 were chosen.

4.2. Feature Extraction

The Inception ResNet V2 model was used to perform feature extraction. The model is loaded with weights obtained from pre-training on the ImageNet data set. The fully connected output layer was excluded from the base model. A 2D average pooling layer is appended to produce the representation vector of the input image.

The pre-processed images are fed to the so constructed convolution neural network to produce a feature vector. We have obtained 1536 features for each input image from the output layer. This vector is then augmented with the geographic metadata, containing country and continent information, to perform the snake species classification.

4.3. Gradient Boost Classifier

A decision tree ensemble classifier is trained using the metadata about the geographic location of the photograph along with the image feature vector obtained form the Inception ResNet V2. Gradient boosting algorithm is used to train the classifier. The parameters of the classifier are tuned over several runs to improve classification results.

Five-fold cross-validation is used to obtain a reliable evaluation of model performance for each configuration of parameters. The classifier is trained five times per run, each time selecting a different fold as the cross-validation set and training on the remaining four folds. The average of the performance parameters (accuracy and F1 score) over the five iterations is considered while tuning the parameters. Cross-entropy loss is used to monitor the model's convergence towards the objective in each fold. Early stopping is used to stop the boosting process if the loss starts to diverge.

5. Implementation Details

The training subset consisting of 347,406 images was split into five folds for cross-validation while training the classifier. Since the data-set had a long-tailed distribution across classes, stratified sampling was used to ensure a proportional split and ensure inclusion of images from each class.

The pre-processed images from the training and validation set are fed into the proposed deep learning convolution neural network model. The model produces feature-vectors of size 1536 for each image.

The geographic location describing where the photographs were taken, specifically the continent and country, are encoded into numeric labels. This information is used as categorical features in classifier. For the images in which this data is unavailable, the features are encoded as 'nan'. The classifier imputes the missing values to the mode of the corresponding feature space. The representation vector consisting of 1538 features obtained for each image is used to train the decision tree ensemble classifiers by gradient boosting.

It was observed that learning rates higher than 0.05 lead to quicker divergence, suggesting the suitability of a slower learning rate using with more decision trees. Grid-search was performed by varying the learning rates in the range of 0.001 to 0.05 and the number of decision trees in

the range 100 to 1000. Combinations having the least losses were chosen to further tune the tree-level parameters.

The maximum depth for the tree is left to be determined based on the training progress of the classifier and is not set strictly. This causes the depth to expand until the leaves are pure (has all samples belonging to the same class) or has reached the threshold of minimum number of samples required to split further. Due to the long-tailed distribution of the data set, some classes may require deeper branches to capture more information from the features. The potential over-fitting that may occur is controlled by tuning and setting an upper limit on the number of leaves by performing a grid search over values in range of 32 to 256.

Some other notable tree-level parameters tuned were sub-sampling rate and column-sampling rate. Sub-sampling rate determines the fraction of training samples that are randomly sampled per tree and was tuned between 0.6 and 1.0. Column-sampling rate, on the other hand, specifies the fraction of features used to fit each decision tree and was tuned between 0.5 and 0.9. Both these parameters help prevent over-fitting. They are maintained sufficiently above 0.1 to prevent under-fitting.

6. Results

The country and continent metadata, used as categorical features in the classifier had a significant impact on the classification. Without the categorical data, the testing accuracy of the best run was 40.16%. This improved to 42.96% when contextual data was encoded as categorical features. Country information has the highest impact while continent information also has a notable influence on the classification. Figure 2, depicts the relative importance of the 20 most significant features of the 1538 features used for classification. The feature importance values are normalized and scaled between 0 and 100 to realise the relative impacts. Features named as f1, f2, etc. denote features extracted from the convolution neural network.

Through parameter tuning, the classifier's performance was improved over several runs. The F1-scores macro-averaged across the countries and macro-averaged over all classes were the prescribed the metrics [10]. Eight best runs were selected based on the prescribed metrics evaluated on the prescribed validation set. The metrics are evaluated as an average over the five iterations (for 5-fold cross validation) performed in each run. We have achieved a training accuracy of 71.32%, validation accuracy of 44.16% and a testing accuracy of 42.96% on the best run. The results are summarized in Tables 1 and 2 below:

Table 1
Prediction metrics of the five best runs on the validation set

Run	F1-Score (Country)	F1-Score (Overall)	Accuracy
1	0.455	0.456	0.531
2	0.482	0.469	0.554
3	0.509	0.481	0.569
4	0.522	0.488	0.583
5	0.536	0.497	0.622

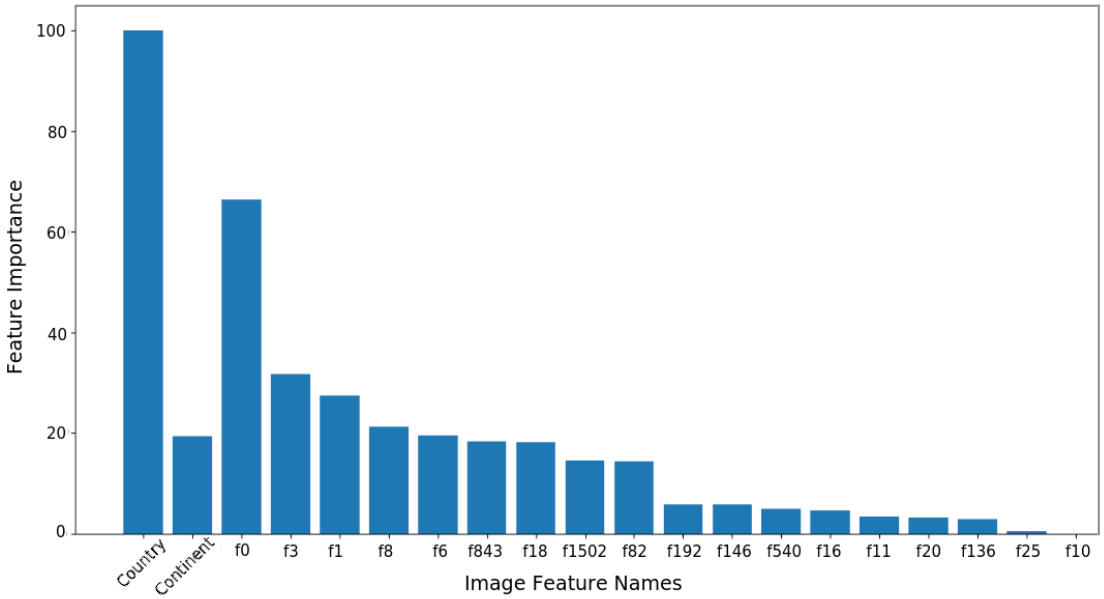

Figure 2: Relative importance on a scale of 0-100 of the 20 most impactful features used to train the classifier. The first two bars represent feature importance of **country** and **continent** respectively

Table 2
Prediction metrics of the five best runs on the test set

Run	F1-Score (Country)	F1-Score (Overall)	Accuracy
1	0.246	0.164	0.428
2	0.247	0.166	0.430
3	0.249	0.159	0.428
4	0.249	0.162	0.432
5	0.252	0.162	0.432

7. Conclusion and Future Work

The results depict the positive impact of integrating contextual country and continent data for snake species classification. Introducing more contextual data such as population counts of various species by region as class-wise probability priors [16], climate information such as temperature and humidity, etc. may contribute to better classification results.

Due to unavailability of sufficient computational resources during the SnakeCLEF-2021 contest period, the results were submitted before complete convergence of the classifier's

training process. Post the deadline, significant improvements in classification accuracy were observed even with a slight increase in the number of iterations applied to train the gradient boost classifier. This suggests that the transfer learning approach adopted here is promising and further parameter tuning and complete training can greatly improve the model performance.

Further efforts to experiment with input image resolutions and alternative pre-trained weights [17] as well as including custom training layers to the frozen base model before extracting features [18] can contribute to the classification performance.

Acknowledgments

Thanks to the Machine Learning Research Group (MLRG), Deptartment of Computer Science and Engineering, Sri Sivasubramaniya Nadar College of Engineering, Chennai, India (https://www.ssn.edu.in/) for providing the GPU resources to implement the model

References

[1] J. M. Gutiérrez, J. J. Calvete, A. G. Habib, R. A. Harrison, D. J. Williams, D. A. Warrell, Snakebite envenoming, Nature reviews Disease primers 3 (2017) 1–21.

[2] YANG, Zihan, Sinnott, Richard., Snake detection and classification using deep learning, in: Proceedings of the 54th Hawaii International Conference on System Sciences, 2021.

[3] A. GARG, D. LEIPE, P. UETZ, The disconnect between dna and species names: lessons from reptile species in the ncbi taxonomy database, Zootaxa 4706 (2019).

[4] B. N. S. B. W. K. H. M. D. C. Stephen W. Corbett, Brian Anderson, Most lay people can correctly identify indigenous venomous snakes, American Journal of Emergency Medicine, The 2698 (2005) A3–A12.

[5] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, Proceedings of the AAAI Conference on Artificial Intelligence 31 (2017). URL: https://ojs.aaai.org/index.php/AAAI/article/view/11231.

[6] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, Lightgbm: A highly efficient gradient boosting decision tree, Advances in neural information processing systems 30 (2017) 3146–3154.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255. doi:10.1109/CVPR.2009.5206848.

[8] J. H. Friedman, Greedy function approximation: a gradient boosting machine, Annals of statistics (2001) 1189–1232.

[9] A. Joly, H. Goëau, S. Kahl, L. Picek, T. Lorieul, E. Cole, B. Deneu, M. Servajean, R. Ruiz De Castañeda, G. H. Bolon, Isabelle, R. Planqué, W.-P. Vellinga, A. Dorso, P. Bonnet, I. Eggel, 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.

[10] L. Picek, A. M. Durso, R. Ruiz De Castañeda, I. Bolon, Overview of snakeclef 2020:

Automatic snake species identification with country-level focus, in: Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, 2021.

[11] A. Amir, N. A. H. Zahri, N. Yaakob, R. B. Ahmad, Image classification for snake species using machine learning techniques, in: S. Phon-Amnuaisuk, T.-W. Au, S. Omar (Eds.), Computational Intelligence in Information Systems, Springer International Publishing, Cham, 2017, pp. 52–59.

[12] C. Abeysinghe, A. Welivita, I. Perera, Snake image classification using siamese networks, in: Proceedings of the 2019 3rd International Conference on Graphics and Signal Processing, ICGSP '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 8–12. URL: https://doi.org/10.1145/3338472.3338476. doi:10.1145/3338472.3338476.

[13] A. Patel, L. Cheung, N. Khatod, I. Matijosaitiene, A. Arteaga, J. W. Gilkey, Revealing the unknown: Real-time recognition of galápagos snake species using deep learning, Animals 10 (2020). URL: https://www.mdpi.com/2076-2615/10/5/806.

[14] A. M. Durso, G. K. Moorthy, S. P. Mohanty, I. Bolon, M. Salathé, R. Ruiz De Castañeda, Supervised learning computer vision benchmark for snake species identification from photographs: Implications for herpetology and global health, Frontiers in Artificial Intelligence 4 (2021) 17.

[15] E. D. Cubuk, B. Zoph, J. Shlens, Q. V. Le, Randaugment: Practical automated data augmentation with a reduced search space (2019). arXiv:1909.13719.

[16] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, W. Xu, Cnn-rnn: A unified framework for multi-label image classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[17] L. Picek, R. Ruiz De Castaneda, A. M. Durso, P. Sharada, Overview of the snakeclef 2020: Automatic snake species identification challenge, CLEF task overview (2020).

[18] M. Zhong, J. LeBien, M. Campos-Cerqueira, R. Dodhia, J. Lavista Ferres, J. P. Velev, T. M. Aide, Multispecies bioacoustic classification using transfer learning of deep convolutional neural networks with pseudo-labeling, Applied Acoustics 166 (2020) 107375. URL: https://www.sciencedirect.com/science/article/pii/S0003682X20304795. doi:https://doi.org/10.1016/j.apacoust.2020.107375.