The first paper

A deep active learning system for species identification and counting in camera trap images

Introduction

Using motion-sensitive cameras hidden throughout the wild, camera traps are used in camera trap surveys to track wildlife. Without requiring direct human observation, researchers may now investigate animal behavior, population levels, and dispersion thanks to these cameras' autonomous movement detection and photo or video capturing capabilities. Because of the slow procedure of decision making in trap surveys, computer vision techniques along with deep learning methods can be used to improve the support of decision making. In this paper, the authors used a deep active learning method combined with computer vision methods for this purpose. The purpose of the paper is to extract and identify different species in an image in different environments based on a novel deep active learning model.

Dataset

In this paper, four different datasets were used. The Snapshot Serengeti dataset features 1.2 million sequences of camera trap images, totaling 3.2 million images, with species and count annotations for 48 animal categories, though about 75% are labeled as empty. The eMammal platform provides over 450,000 images from diverse global locations, covering over 270 species, for machine learning research. The North America Camera Trap Images (NACTI) dataset includes 3.7 million images from five U.S. locations, with labels for 28 species and around 12% empty labels. The Caltech Camera Traps (CCT) dataset comprises 245,000 images from 140 traps in the Southwestern U.S., featuring 22 animal categories and about 70% empty labels. The paper aims to address key challenges in applying deep learning to camera trap images by proposing a novel processing pipeline using the mentioned datasets.

The proposed method

For image classification and object detection in this paper, the following concepts were considered. Classification models for camera trap images typically label an image based on the most likely species, making it difficult to handle images with multiple species. These models also perform poorly on counting tasks compared to classification. They learn from both the animal patterns and backgrounds in training images, which reduces their effectiveness when applied to new locations, leading to decreased accuracy. For instance, a model trained on U.S. images was less accurate when tested on Canadian datasets, highlighting the challenge of transferring models across different regions. Unlike classification models, object detection models can handle numerous species inside a single image, which makes them a good fit for camera trap photos. However, due to their high cost and low importance for ecologists, bounding box annotations for individual animals are generally not accessible in camera trap datasets, which makes them necessary for training these models. Despite this difficulty, subsequent studies have demonstrated that, in the presence of bounding box labels, object detection may be successful. This work investigates the use of transfer learning to use pre-trained models in order to solve the lack of such labels. As seen in the Snapshot Serengeti dataset, object identification models are able to accurately identify and count several species even in the absence of bounding box annotations.

The proposed pipeline starts with applying a pre-trained object detection model—Faster-RCNN for Snapshot Serengeti data and MegaDetector for NACTI data. The model, which identifies animal presence, counts animals, and localizes them, processes images to mark empty ones, count animals, and crop out background. These cropped images are resized to 256 × 256 pixels for further species recognition.
To address the challenges of active deep learning on large datasets, The authors used transfer learning and active learning techniques. They first trained an embedding model to convert crops into 256-dimensional feature vectors, which speeds up the active learning process by reducing the complexity of input data. Then, they used a simple neural network for classification, starting with 1,000 labeled images, and iteratively label additional images based on selection criteria, retraining the model and fine-tuning the embedding model periodically.

Novelty and contribution

Applying active deep learning to large, high-dimensional datasets presents two key challenges: First, limited labeled images make it difficult to train deep neural networks from scratch, necessitating the use of transfer learning from related datasets. Second, active learning requires evaluating the entire unlabeled dataset to identify the most informative samples for labeling, which is slow with millions of high-dimensional images. Random sampling can expedite the process but at the cost of inefficiency, as it may not prioritize the most useful images for labeling. In summary, the authors suggested approach keeps speed while evaluating all data points to choose the most important instances to ask people to categorize.

Results

By decreasing the annotation bottleneck (by 99.5%), The suggested pipeline may make it easier to install large camera trap arrays and boost the effectiveness of studies in wildlife biology.

Downsides

Reliance on pre-trained object detection models can restrict adaptability to new datasets if these models don't fit the specific species or environments well. Additionally, active learning, despite reducing overall labeling effort, still demands significant initial human input and can be computationally heavy, with the risk that selecting the most informative samples might not always yield optimal model performance.

The second paper

Efficient Labeling of EEG Signal Artifacts using Active Learning

Introduction

Electroencephalography, or EEG, is a method that uses sensors applied to the scalp to monitor electrical activity in the brain. It records the brain waves produced by cell activity and offers insightful information about how the brain works. Because EEG may detect irregular brain wave patterns, it is frequently used to diagnose neurological conditions such epilepsy, sleep problems, and brain traumas. It is also essential for tracking brain activity in critical care and surgical environments. Researchers can better understand cognitive processes and how different stimuli affect brain activity by using EEG data. Unwanted signals that might skew brainwave records are known as EEG artifacts, and they can be brought on by a number of circumstances unrelated to real brain activity. These include electrical interference from devices or external sources like appliances, as well as movement artifacts from head or body motions like blinking or eating. Electrode artifacts result from inadequate contact or movement of the EEG electrodes, whereas muscle artifacts are caused by electrical activity in the muscles of the face or neck. To ensure trustworthy analysis of brain activity and to acquire correct EEG data, it is imperative to identify and minimize these artifacts. Along with EMI artifacts from external electromagnetic fields, EEG artifacts also contain EMG artifacts from electrical activity in the muscles of the face and neck. While EMI artifacts originate from electrical equipment or power sources, EMG artifacts can damage EEG recordings with noise from muscle movements. Finding and fixing these kinds of aberrations is essential for accurate brainwave readings since they can both compromise the accuracy of EEG analysis.

Traditional artifact removal from EEG data involves manual identification and removal, which is time-consuming and reduces data quality. Independent Component Analysis (ICA) is an alternative that effectively removes certain artifacts but often requires manual labeling of large EMG artifacts beforehand to ensure accurate results. While automated methods exist for artifact removal, their effectiveness in preserving neural data is still uncertain. Consequently, many researchers still rely on manual artifact removal. Improved and automated initial artifact identification methods could greatly reduce the effort required for processing EEG data.

The main goal of this study is to identify and recognize multi-class artifacts in EEG data using Active Learning (AL) methods. In the current study the user is able to visually identify the EEG signal artifacts contained in the data directly. In this research AL is applied to identify multiple classes of artifacts.

Dataset

Using a 64-channel Biosemi ActiveTwo system, EEG data were acquired at 512 Hz and referred to connected mastoids. Four external EOG channels were used to detect eye movements. EEGLAB and MATLAB were used to down-sample the data to 256 Hz and high-pass filter it at 1 Hz. The data included segments with muscle activity, eye blinks, eye movements, and no artifacts. The VEP dataset had around 1200 epochs per subject, whereas the artifact dataset had 160 epochs per subject with an equitable distribution among artifact classes. Hand-labeled all the data by an EEG specialist. The objective is to categorize artifact occurrences in the VEP dataset by using a classifier trained on the artifact dataset.

The proposed method

The analysis employs an autoregressive-based support vector machine (SVM) classifier, as detailed in a previous study. An AR(2) model is fitted to each EEG channel, generating 136 features per epoch by concatenating AR coefficients from 68 channels (64 EEG + 4 EOG). Classification is performed using a radial basis function support vector machine (RBF-SVM) with training done via LibSVM, and Platt scaling is used to produce probability distributions over the classes for each epoch.

$$Y(t) = \sum_{i=1}^{p} A_i Y(t-i) + \epsilon_t$$

AR Model

Two active learning (AL) strategies are examined in the analysis: the traditional Query by Committee (QBC) technique and the Decision-Confidence Weighted QBC method. By manually labeling the most informative unlabeled data iteratively, both approaches seek to enhance classification performance. K-fold cross-validation is used in the QBC technique to train K classifiers on the original labeled dataset. In the unlabeled data, each classifier then forecasts labels for every epoch. The committee members' level of disagreement determines the ranking of the epochs. Each classifier's confidence in its predictions is taken into account for Decision-Confidence Weighted QBC, and epochs with greater confidence ratings are weighted correspondingly.

Oracle labeling is applied to the epochs that have the highest disagreement or the lowest summed confidence; these labeled epochs are then included in the original dataset. With the new labeled data, the classifier is retrained, and its performance is evaluated on a validation set. Until convergence—defined as reaching a minimum confidence level of 0.6 from each classifier and consistent agreement among committee members—this iterative procedure is continued. The Baseline condition, in which epochs are randomly sampled for labeling without taking any metrics into account, and the full 10-fold cross-validation, which assumes complete knowledge of all labels and represents the theoretical maximum classification performance, are compared in order to assess the efficacy of the AL methods.

The AL procedures are repeated 100 times, sampling from the labeled set before each iteration, to ensure robust performance assessment.

Novelty and contribution

A key difference between Query by Committee (QBC) and Decision-Confidence Weighted QBC is that while QBC may converge if all committee members agree on a prediction, the overall confidence might be low, leading to potential underperformance. Decision-Confidence QBC addresses this by considering the confidence levels, which can lead to improved performance even when committee members agree. Additionally, two measures are used for comparison: a Baseline condition where epochs are randomly sampled for labeling to control for sample size effects, and full 10-fold cross-validation, which provides an upper bound on classification performance by assuming complete label knowledge.

Results

When compared to the Baseline method, Active Learning (AL) techniques such as Query by Committee (QBC) and Decision-Confidence Weighted QBC significantly improve classification results. Less than 25% of labeled data for a number of participants (1, 3, 4, and 7) may obtain almost complete 10-fold cross-validation performance, demonstrating significant time savings when labeling EEG data for artifacts. However, performance can decline with more labeled data for specific individuals (1, 3, 4, and 6), indicating that labeling uninformative data points may impede overall performance. After the first AL iteration, performance usually increases dramatically, suggesting that early labeling aids in the classifier's adaptation to the feature space of the VEP data. The classifier's performance may occasionally deteriorate in subsequent iterations despite ongoing advancements in AL techniques, whereas the Baseline measure may not show any progress. The two AL methods are statistically similar in performance, with only minor improvements noted with Decision-Confidence QBC for certain subjects.

Downsides

The research has some limitations, primarily related to computational efficiency and parameter sensitivity. Current hardware constraints may affect computational times, though future advancements in processors and active learning methods are expected to improve this. The study's findings are based on specific parameters, such as using 5-fold cross-validation and selecting 5 samples per active learning iteration, without exploring other configurations that might yield quicker or better results but could increase the manual labeling effort. Additionally, adjusting parameters like the number of committee members might impact performance and computational demands. Balancing labeling effort, classification performance, and model settings is complex and likely varies depending on the dataset.

The third paper

Generalized batch mode active learning for face-based biometric recognition

Introduction

Biometric face recognition technology verifies identity by analyzing unique facial features, offering enhanced security compared to traditional methods like passwords, which can be forgotten or stolen. Its ability to provide real-time verification is particularly valuable for secure access and surveillance applications. However, it faces challenges such as potential inaccuracies in matching, privacy concerns over the sensitive nature of facial data, and the risk of misuse for mass surveillance. To address these issues, ongoing advancements focus on improving accuracy, reducing error rates, and implementing stringent data protection measures to safeguard privacy and ensure ethical use of the technology.

The biometric image data consists of staggering instances of features. With the large volumes of data created every day and the use of many labeling agents such as biometric devices, traditional active learning approaches update the model one data sample at a time, which can be wasteful. Recent developments center on batch mode active learning, which simultaneously chooses and learns from several data points, to overcome this. In industries where it's critical to efficiently handle massive amounts of redundant data, such robotics and wearable vision systems, this method is very helpful in security and surveillance. In order to manage the considerable redundancy in biometric data, The research presents a unique batch mode active learning system.

The proposed method

In a biometric recognition application analyzing a video stream, an effective batch mode active learning strategy should select images that capture diverse facial expressions to ensure minimal redundancy. To achieve this, an objective function is formulated where the batch is designed to represent various facial appearances and reduce classifier uncertainty. This function includes terms to minimize the entropy of the classifier on unselected images, ensuring that the updated model can handle the remaining data effectively. Additionally, to avoid bias towards high-density image regions, the function incorporates a term to select images from low-density areas, capturing less frequent but important facial expressions. The underlying assumption is that feature vectors in the image space will reflect similarity between similar appearances and distance between different ones, allowing the algorithm to choose a representative and diverse batch.

The batch mode active learning technique is intended to maximize the selection of data samples for updating a classification model. Initializing the Hessian matrix as the identity matrix is the first step in the procedure. After that, the objective function and its gradient are assessed by the algorithm. It solves a quadratic programming (QP) problem iteratively, using the gradient and current Hessian matrix to identify a new potential solution. The Armijo-Goldstein criteria are then used to further refine this potential solution in order to determine the proper step size. The new objective function and gradient are evaluated in order to compute the revised solution and compare it to the prior one.

To ascertain if the change falls below the permissible threshold, the difference between the old and new objective values is evaluated. BFGS equations are used to update the Hessian matrix, making adjustments based on the changes that are observed. In accordance, the gradient and goal function are likewise modified. Until the difference in the goal values is less than a predetermined threshold, signifying convergence, this iterative refining process is continued.

The most illuminating data points are identified when the algorithm has converged and has chosen the top elements in the solution vector based on their values and set these entries to 1. These top entries are then used to choose the final set of data points. By ensuring that the batch that is chosen is representative and diverse, this method successfully reduces redundancy and captures a range of data qualities.

The novelty and contribution

For the automated face recognition problem, a batch mode active learning algorithm based on optimization has been created and implemented. The method that is suggested is general and may be applied widely to pick out notable and model examples from a lot of unlabeled data.

Results

The research had numerous results but in summary, the framework's versatility is demonstrated by its capacity to integrate additional available information. The results from testing on the VidTIMIT and NIST MBGC datasets validate its effectiveness for real-world biometric applications.

Downsides

The optimization-based batch mode active learning algorithm, while effective for biometric recognition, can be computationally intensive and resource-demanding. Its performance depends on high-quality features and can struggle with noisy data. Additionally, incorporating extra information and handling feature distance assumptions can pose challenges, potentially limiting generalizability.

References

1- Norouzzadeh, Mohammad Sadegh, et al. "A deep active learning system for species identification and counting in camera trap images." *Methods in ecology and evolution* 12.1 (2021): 150-161.
2- Lawhern, Vernon, et al. "Efficient labeling of EEG signal artifacts using active learning." *2015 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2015.
3- Chakraborty, Shayok, Vineeth Balasubramanian, and Sethuraman Panchanathan. "Generalized batch mode active learning for face-based biometric recognition." *Pattern Recognition* 46.2 (2013): 497-508.