Autonomous landing scene recognition based on transfer learning for drones

Dr.ECCLESTON

[1] Assistant Professor, ,Malla Reddy Engineering College for Women (Autonomous Institution), Hyderabad, maheshh118@gmail.com[1]

[2] ,[3],[4]Student, , Malla Reddy Engineering College for Women (Autonomous Institution), Hyderabad

thotaashmitha@gmail.com[2], ksanjanareddy1306@gmail.com[3], harikakummari08@gmail.com[4]

ABSTRACT—In this paper, we study autonomous landing scene recognition with knowledge transfer for drones. Considering the difficulties in aerial remote sensing, especially that some scenes are extremely similar, or the same scene has different representations in different altitudes, we employ a deep convolutional neural network (CNN) based on knowledge transfer and fine-tuning to solve the problem. Then, LandingScenes-7 dataset is established and divided into seven classes. Moreover, there is still a novelty detection problem in the classifier, and we address this by excluding other landing scenes using the approach of thresholding in the prediction stage. We employ the transfer learning method based on ResNeXt-50 backbone with the adaptive momentum (ADAM) optimization algorithm. We also compare ResNet-50 backbone and the momentum stochastic gradient descent (SGD) optimizer. Experiment results show that ResNeXt-50 based on the ADAM optimization algorithm has better performance. With a pre-trained model and fine-tuning, it can achieve 97.8450% top-1 accuracy on the LandingScenes-7 dataset, paving the way for drones to autonomously learn landing scenes.

INTRODUCTION

Remote sensing can be divided into ground remote sensing, aerial remote sensing, and aerospace remote sensing according to different working platforms.

For aerial remote sensing, sensors are mounted on aircraft such as balloons, drones, and helicopter. It has become possible for drones to recognize the landing scenes with the development and application of convolutional neural network (CNN) and graphic processing unit (GPU). Huge progress has been made in object detection in aerial images and scene classification in remote sensing images stems from the rise of CNN and the public datasets recently, such as geospatial object detection, scene classification of satellite images, and aerial scene classification. Anand et al. applied deep learning and a label "H" to achieve automatic landing for the drone. However, in emergency landing scenarios, the landing mark may not be arranged in advance. Tian et al. employed the Inception V3 model for landing scene recognition on the ImageNet dataset, and proposed a learning rate decay method with computational verb theory to improve the recognition accuracy. Lu et al. proposed an algorithm for augmenting the l-channel to the traditional RGB images. Experimental results show the effectiveness of the l-channel, and the performance is significantly improved in scene classification and object recognition tasks. Scene recognition based on CNN is inseparable from the support of datasets, such as Places365 dataset, dataset Sconce Understanding (SUN) and SUN attribute dataset. Yang et al.proposed a simultaneous localization and mapping (SLAM) method based on a monocular vision to realize autonomous landing in an emergency and unknown environment for drones. He et al. presented an ingenious backbone named Reset and won the championship on the task of common object in context (COCO) detection and segmentation, ImageNet detection, and localization in ImageNet large scale visual recognition challenge (ILSVRC)-2015 and COCO-2015 competitions. Xie et al. used group convolution inherited by Reset, and showed a ResNeXt-50 with group = 32 and bottleneck width *d*=4, however, the network showed better accuracy than the Reset counterparts only when a building block is "bottleneck". Namely, it has better accuracy on ResNeXt-50, ResNeXt-101, or ResNeXt-152. Xie et al. reviewed many popular and effective approaches to scene recognition. Although there is much literature on scene recognition in the past, little research is conducted on landing scene categories recognition for the drone. To achieve landing scene recognition, we need to face its unique representation. For example, in a since where lots of lotus leaves appear on the water, the scene may be classified as a "water area" rather than "wilderness". In a scene where a bridge spans across the Yangtze River, the scene may be classified as "water area" or "road". In other words, object detection focuses on the foreground of the image, while landing scene recognition focuses more on the background. This is one of the thorny problems of landing scene categories recognition. It has always been the goal of robotics researchers to realize complete drone autonomy and intelligence. At present, autonomous obstacle avoidance and flight for drones have been realized following the planned route. However, these autonomous flights are based on non-emergency conditions. If a drone encounters a sudden drop in battery power or detects a fault and needs an emergency landing, it is necessary to judge whether the current scene below the drone is suitable for an emergency landing. Consequently, the research motivation of this paper is to explore a landing scene recognition that is suitable for the drone. In this case, landing scene recognition can improve the flight safety of the drone.

The key contributions of this paper is as follows: First, we employ a knowledge transfer method based on the Resent, use pre-trained weights to initialize the network, and retrain the model with fine-tuning. Second, dataset LandingScenes-7 is created where images come from drone aerial photography, dataset Places365 and the Internet. We divide the dataset into seven categories, namely "crowded place", "lawn", "road", "vehicle_ intensive place", "wilderness", "wheatfield", and "water area". First, we train the model based on CNN backbone with transfer learning on dataset LandingScenes-7. Second, we judge whether the result of the test stage is correct through the novelty detection module. A certain category is output directly if its probability is high; otherwise, further judgments are made through the thresholding.

RELATED WORK

Effective distributed convolutional neural network architecture for remote sensing images target classification with a pre-training approach

How to recognize targets with similar appearances from remote sensing images (RSIs) effectively and efficiently has become a big challenge. Recently, convolutional neural network (CNN) is preferred in the target classification due to the powerful feature representation ability and better performance. However, the training and testing of CNN mainly rely on single machine. Single machine has its natural limitation and bottleneck in processing RSIs due to limited hardware resources and huge time consuming. Besides, overfitting is a challenge for the CNN model due to the unbalance between RSIs data and the model structure. When a model is complex or the training data is relatively small, overfitting occurs and leads to a poor predictive performance. To address these problems, a distributed CNN architecture for RSIs target classification is proposed, which dramatically increases the training speed of CNN and system scalability. It improves the storage ability and processing efficiency of RSIs. Furthermore, Bayesian regularization approach is utilized in order to initialize the weights of the CNN extractor, which increases the robustness and flexibility of the CNN model. It helps prevent the overfitting and avoid the local optima caused by limited RSI training images or the inappropriate CNN structure. In addition, considering the efficiency of the Naïve Bayes classifier, a distributed Naïve Bayes classifier is designed to reduce the training cost. Compared with other algorithms, the proposed system and method perform the best and increase the recognition accuracy. The results show that the distributed system framework and the proposed algorithms are suitable for RSIs target classification tasks.

Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities

Remote sensing image scene classification, which aims at labelling remote sensing images with a set of semantic categories based on their contents, has broad applications in a range of fields. Propelled by the powerful feature learning capabilities of deep neural networks, remote sensing image scene classification driven by deep learning has drawn remarkable attention and achieved significant breakthroughs. However, to the best of our knowledge, a comprehensive review of recent achievements regarding deep learning for scene classification of remote sensing images is still lacking. Considering the rapid evolution of this field, this article provides a systematic survey of deep learning methods for remote sensing image scene classification by covering more than 160 papers. To be specific, we discuss the main challenges of remote sensing image scene classification and survey: first, autoencoder-based remote sensing image scene classification methods; second, convolutional neural network-based remote sensing image scene classification methods; and third, generative adversarial network-based remote sensing image scene classification methods. In addition, we introduce the benchmarks used for remote sensing image scene classification and summarize the performance of more than two dozen of representative algorithms on three commonly used benchmark datasets. Finally, we discuss the promising opportunities for further research.

Improving object recognition with the ℓ-channel

We augment a new channel called the ℓ-channel to conventional *RGB* images, and propose its application in multiple classification and recognition tasks. The new *RGB*-ℓ image records the same scene using the colour and *frosted light channel*, which are simultaneously captured using a binocular camera with a low-cost frosted glass placed in front of one of the cameras. Due to the light scattering property of the frosted glass, the acquired frosted light channel is imprecise. In this paper we propose a novel optimization that is guided by the RGB channel to refine the ℓ-channel to preserve edges due to scene radiance. Extensive experimental results have demonstrated the effectiveness of our *RGB*-ℓ images, where significant improvements are reported in a variety of scene classification and object recognition tasks.

Places: A 10 million Image Database for Scene Recognition

The rise of multi-million-item dataset initiatives has enabled data-hungry machine learning algorithms to reach near-human semantic classification performance at tasks such as visual object and scene recognition.

Here we describe the Places Database, a repository of 10 million scene photographs, labelled with scene semantic categories, comprising a large and diverse list of the types of environments encountered in the world. Using the state-of-the-art Convolutional Neural Networks (CNNs), we provide scene classification CNNs (Places-CNNs) as baselines, that significantly outperform the previous approaches. Visualization of the CNNs trained on Places shows that object detectors emerge as an intermediate representation of scene classification. With its high-coverage and high-diversity of exemplars, the Places Database along with the Places-CNNs offer a novel resource to guide future progress on scene recognition problems.

SUN Database: Exploring a Large Collection of Scene Categories

Progress in scene understanding requires reasoning about the rich and diverse visual environments that make up our daily experience. To this end, we propose the Scene Understanding database, a nearly exhaustive collection of scenes categorized at the same level of specificity as human discourse. The database contains 908 distinct scene categories and 131,072 images. Given this data with both scene and object labels available, we perform in-depth analysis of co-occurrence statistics and the contextual relationship. To better understand this large-scale taxonomy of scene categories, we perform two human experiments: we quantify human scene recognition accuracy, and we measure how typical each image is of its assigned scene category. Next, we perform computational experiments: scene recognition with global image features, indoor versus outdoor classification, and “scene detection,” in which we relax the assumption that one image depicts only one scene category. Finally, we relate human experiments to machine performance and explore the relationship between human and machine recognition errors and the relationship between image “typicality” and machine recognition accuracy.

Adam: A Method for Stochastic Optimization

We introduce Adam, an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments. The method is straightforward to implement, is computationally efficient, has little memory requirements, is invariant to diagonal rescaling of the gradients, and is well suited for problems that are large in terms of data and/or parameters. The method is also appropriate for non-stationary objectives and problems with very noisy and/or sparse gradients. The hyper-parameters have intuitive interpretations and typically require little tuning. Some connections to related algorithms, on which Adam was inspired, are discussed. We also analyse the theoretical convergence properties of the algorithm and provide a regret bound on the convergence rate that is comparable to the best-known results under the online convex optimization framework. Empirical results demonstrate that Adam works well in practice and compares favourably to other stochastic optimization methods. Finally, we discuss Ada Max, a variant of Adam based on the infinity norm.

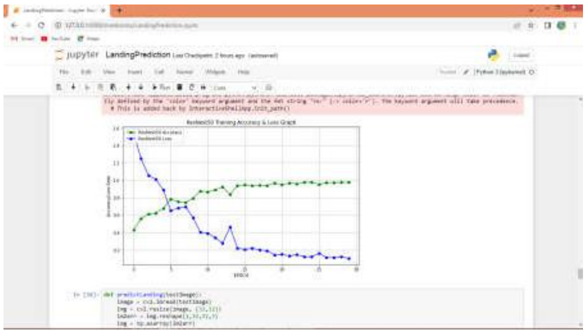
METHODOLOGY

1. **get label:** using this module we will get the label

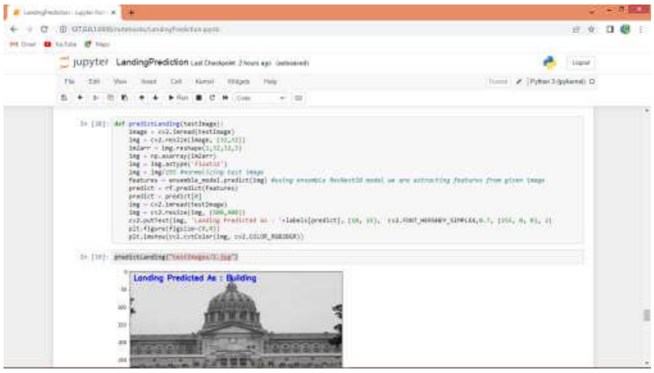
2. **calculate metrics:** using this module, metrics can be calculated

3. **Predict landing:** using this module we will predict the landing position

RESULT AND DISCUSSION



Above is ResNext50 training graph where x-axis represents training epoch and y-axis represents accuracy and loss values where green line represents accuracy and blue line represents loss and with each increasing epoch accuracy got increase and reached closer to 1 and loss got decrease



In above graph defining predict function and this function will take input image path and then using extension ensemble object it will classify given image scenes and in above image scene classify as building

CONCLUSION

In this paper author study autonomous landing scene recognition with knowledge transfer for drones. Considering the difficulties in aerial remote sensing, especially that some scenes are extremely similar, or the same scene has different representations in different altitudes, we employ a deep convolution neural network (CNN) based on knowledge transfer and fine-tuning to solve the problem. Then, LandingScenes-7 dataset is established and divided into seven classes. Moreover, there is still a novelty detection problem in the classifier, and we address this by excluding other landing scenes using the approach of thresholding in the prediction stage. We employ the transfer learning method based on ResNeXt-50 backbone with the adaptive momentum (ADAM) optimization algorithm. We also compare ResNet-50 backbone and the momentum stochastic gradient descent (SGD) optimizer.

REFERENCES

[1] LI B Q, HU X H Effective distributed convolutional neural network architecture for remote sensing images target classification with a pre-training approach Journal of Systems Engineering and Electronics, 2019, 30 (2): 238- 244.

[2] XIA G S, BAI X, DING J, et al DOTA: a large-scale dataset for object detection in aerial imagesProc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, 3974- 3983.

[3] CHENG G, XIE X X, HAN J W, et al Remote sensing image scene classification meets deep learning: challenges, methods, benchmarks, and opportunities

IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13, 3735- 3756.

[4] ANAND A K, BARMAN S, PRAKASH N S, et al Vision based automatic landing of unmanned aerial vehicleProc. of the International Conference on Information Technology and Applied Mathematics, 2019, 102- 113.

[5] TIAN C, HUANG C B An algorithm for unmanned aerial vehicle landing scene recognition based on deep learning and computational verbs Proc. of the IEEE International Conference on Civil Aviation Safety and Information Technology, 2019, 180- 184.

[6] LU C W, TSOUGENIS E, TANG C Kim proving object recognition with the l-channel

Pattern Recognition, 2016, 49, 187- 197.

[7] ZHOU B L, LAPEDRIZA A, KHOSLA A, et al Places: a 10 million image database for scene recognition IEEE Trans. on Pattern Analysis and Machine Intelligence, 2018, 40 (6): 1452- 1464.

[8] XIAO J X, HAYS J, EHINGER K A, et al SUN database: large-scale scene recognition from abbey to zoo Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2010, 3485- 3492.

[9] XIAO J X, EHINGER K A, HAYS J, et al SUN database: exploring a large collection of scene categories International Journal of Computer Vision, 2016, 119 (1): 3- 22.

[10] PATTERSON G, XU C, SU H, et al The SUN attribute database: beyond categories for deeper scene understanding International Journal of Computer Vision, 2014, 108 (1/2): 59- 81.

[11] YANG T, LI P Q, ZHANG H M, et al. Monocular vision SLAM-based UAV autonomous landing in emergencies and unknown environments. Electronics, 2018, 7(5): 73.

[12] HE K M, ZHANG X Y, REN S Q, et al Deep residual learning for image recognition Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, 770- 778.

[13] XIE S N, GIRSHICK R, DOLLAR P, et al Aggregated residual transformations for deep neural networksProc. of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017, 5987- 5995.

[14] XIE L, LEE F F, LIU L, et al Scene recognition: a comprehensive survey Pattern Recognition, 2020, 102, 107205.