A Survey on Fine –grained Visual Recognition and Classification Systems

Haripriya S[1], Dr. Teena Joseph[2]

[1]P G Scholar, Department of CSE, Musaliar College of Engineering & Technology, Pathanamthitta, Kerala, India.
[2]Professor and Hod, Department of CSE, Musaliar College of Engineering & Technology, Pathanamthitta, Kerala, India.

Abstract: *Image recognition is an important area of functional application for implementing algorithms for deep learning. The methods of image recognition for classification include two levels: general-level image classification and classification of fine-grained images. The general classification of images aims to classify images into various key categories, such as ships, vehicles, aeroplanes, etc.Under certain basic-level categories, fine-grained image classification typically tends to identify sub-categories. Fine-grained image classification is a very difficult task, unlike the general-level object classification dilemma, due to the high degree of similarities between subcategories. In recent years, models of fine-grained image recognition have made significant strides. In this article, we address various types of fine-grained classification models focused on different fields of machine learning and artificial intelligence to describe the visual recognition characteristics and variations between them.*

Key words: Deep learning, ResNet, bilinear CNN model, RMA, VLAD, AlexNet

1. INTRODUCTION

Fine-grained classification is difficult since only subtle and local distinctions will distinguish between groups. Typically, variances in the posture, size or rotation make the issue more complicated. The pipeline of finding foreground object or object pieces (where) to obtain discriminative characteristics is pursued by most fine-grained classification systems.As a sub-field of object identification, fine-grained categorization attempts to differentiate subordinate divisions within categories at the entry level[4].In various vision tasks such as image recognition, object identification, and semantic segmentation, the deep learning technology has demonstrated amazing results. In particular, recent developments in deep learning techniques are promoting success in the classification of fine-grained images aimed at separating subordinate groups, such as bird species or dog breeds [11].Due to high intra-class and low inter-class variation, this job is highly difficult. In this article, we study various forms of fine-grained image classification approaches based on deep learning, including general convolutionary neural networks (CNNs), component detection-based, network-based ensemble, and fine-grained image classification approaches based on visual focus.

The rest of the paper is organized as follows. Section 2 provides a review of the relevant methods; Sec. 3 details of technologies and features; Sec. 4 explains result analysis; Sec. 5 conclusions.

2. METHODS

2.1 Visual Recognition Based On Deep Learning for Navigation Mark Classification.

An important field for researching smart ships and intelligent navigation is recognising objects from camera images. In maritime transport, this paper focuses on navigation marks indicating the features of navigational environments (e.g. channels, special areas, wrecks, etc.).A fine-grained classification model named RMA (ResNet-Multiscale-Attention) based on deep learning is proposed to analyse the subtle and local differences among navigation mark types for the recognition of navigation marks[1]. An attention system based on the fusion of feature maps with three scales is suggested in the RMA model to find areas of attention and catch discriminatory characters that are necessary to

differentiate the subtle discrepancies between identical navigation marks. In recent years, with the advancement of AI technology, different intelligent systems have also been used to study smart ships and intelligent navigation[5].Usually, the general level image classification networks, like ResNet-50, use only top-level classification functions, so they do not perform so well with fine-gained level tasks[2]. In this article, a multiple scale focus mechanism is proposed to assemble the information missing from the top-level characteristics. In order to form an attention matrix, features obtained from various stages of ResNet are combined, which is used to notice the favourable area for classification by element-wise multiplication with the input image. In order to complete the final grouping, the improved photographs are then entered into the second ResNet. A model structure called RMA (ResNet-Multiscale-Attention)[1].

2.2 Bilinear CNN Models for Fine-grained Visual Recognition

They suggest bilinear models, an architecture of recognition consisting of two attribute extractors whose outputs are multiplied at each image position using the outer product and pooled to produce an image descriptor[8]. This architecture can model translationally invariant local pair wise function interactions, which is especially useful for fine-grained categorization. The bilinear type simplifies gradient computing and only uses picture labels to facilitate end-to-end training on all networks. Their key achievement is an architecture of recognition that solves many disadvantages to both part-based and texture models.It consists of two CNN dependent feature extractors whose outputs are multiplied at each location of the image using the outer product and pooled through locations to produce an image descriptor[8]. For eg, if one of the networks are a component detector and the other a local feature extractor, the outer product captures pair wise correlations between the feature channels and can model part-feature interactions.Several commonly used orderless texture descriptors are also generalised by the bilinear model, such as Bag-of-Visual-Words[17], VLAD[16], Fisher vector[14], and second-order pooling (O2P). In comparison, unlike these texture definitions, the architecture can be quickly trained end-to-end to contribute to major performance changes.Although they do not further discuss this relation, their architecture is linked to the two visual processing stream theories in the human brain, where there are two major pathways, or "streams." The ventral stream (or, "what pathway") is interested in distinguishing and understanding objects. The dorsal stream (or "where path") is involved in processing the spatial position of the target. As their model is linear in the outputs of two CNNs, their solution is called bilinear CNNs. Figure 1 represents the model.

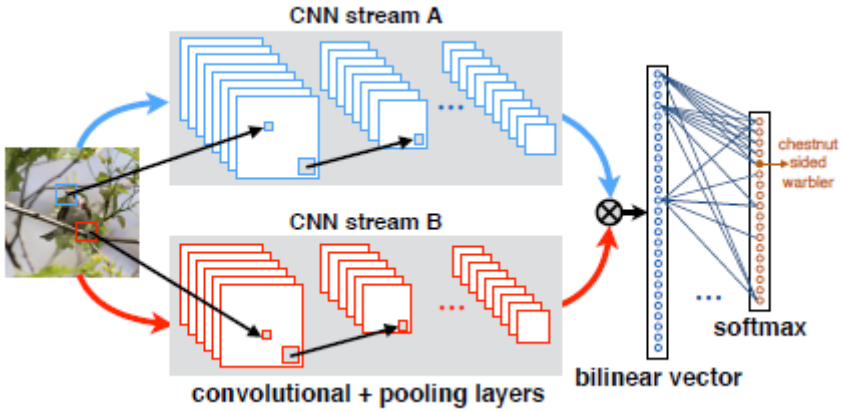

Fig-1: Bilinear CNN model for image classification

A picture is passed through two CNNs, A and B, at the test time, and their outputs are multiplied at each image position using the outer product and pooled to produce the bilinear vector. To receive forecasts, this is transmitted through a classification layer[2]

2.3 Fast CNN Surveillance Pipeline for Fine-Grained Vessel Classification and Detection in Maritime Scenarios

For several vision benchmarks in object detection and classification tasks, deep convolutionary neural networks (CNNs) have proven very successful. However, in wide-view video surveillance settings where artefacts are small, the computational complexity and object resolution requirements of CNNs restrict their applicability. This paper provides a CNN Vessel Positioning and Classification Monitoring Pipeline[6].The proposed pipeline, with three key phases, is based on the GPU implementation of Fast-R-CNN. In the first point, it uses a weak HOG-based object detector for successful area recommendation. The second step utilises the VGG16 maritime fine-tuned network to remove high-level CNN functionality using a linear SVM for off-the-shelf classification. The final stage offers area verification and assigns final class labels using a clear trust scheme[7]. The whole framework is configured by multi-core parallel operation of area proposal and GPGPU computation for CNN evaluation for near real-time efficiency. The major contributions to this paper are: i) the development of a deep CNN pipeline appropriate for identification and classification.(ii) Use of basic object detectors in wide-view images based on hand-engineered features for area recommendation. (iii) Review of state-of-the-art advanced deep CNN pipelines on the latest Annapolis Maritime Monitoring Dataset[10] for surveillance use.

2.4 The Application of Two-level Attention Models in Deep Convolutional Neural Network for Fine-grained Image Classification

They recommend a fine-grained grouping pipeline in this paper that incorporates bottom-up and two top-down attentions[9]. Three forms of attention are introduced into the pipeline: the bottom-up attention that recommends candidate patches, the top-down attention at the object-level that chooses appropriate patches for a particular object, and the top-down attention at the part-level that locates discriminatory pieces. These attentions are often combined to train domain-specific deep networks, and used to optimise facets of both what and where[15]. The classifier does not work on the raw picture for this to work, but rather its patches. The most significant object that is applicable to the identification measures should also be maintained by such patches. A object level FilterNet is implemented to determine whether to progress to the next steps with a fix suggested by the bottom-up process. The FilterNet only takes care of whether a patch is connected to the group of the simple level and aims context patch filtering.Then the filters in the DomainNet display particular interests in unique parts of the object and the pattern of clustering can be seen among filters according to their parts concerned. To find the groups, they use spectral clustering, then use the filters in a group to act as a component detector[9]. The bottom-up ideas are subject to two layers of top-down focus. To pick bird-relevant patches to feed into the classifier, .One performs object-level filtering. To detect sections for classification, the other performs part-level detection. For the part-level process, DomainNet will provide the part detectors and also the function extractor for each of the two classifiers. In the later point, the estimation effects of the two classifiers are combined to incorporate the gains of the attentions of the two levels[13].

2.5Maritime Ship Targets Recognition with DeepLearning

The structure and theory of the Faster RCNN algorithm are studied in this article, and the validity and accuracy of the Faster RCNN algorithm are tested in the V0C2007 dataset, and the data set necessary for marine ship recognition is

then generated and the ship recognition tests are performed under the current Faster RCNN network model[3].By Pick Quest, the RCNN algorithm takes up around 2000 regions, which takes up a lot of time and space overhead. In order to reach a near real-time detection rate, Fast RCNN uses a deep network to neglect the time to suggest the regions, which significantly restricts the efficacy of the target recognition algorithm[3].And Faster RCNN uses the RPN network to propose regions and combines the convolution layer with the Fast RCNN network to share, which greatly reduces the enormous overhead given by the proposed algorithm, such as selective search, and greatly increases the overall running speed of the algorithm[12].A deep learning technique is applied by introducing the Faster RCNN algorithm to the identification of marine vessels in order to increase the precision of the intellectual recognition of marine vessels in addition to the realisation of the algorithm.

3. FEATURES & TECHNOLOGIES

Methods	Technology Used	FeatureExtraction Methods
Visual Recognition Based On Deep Learning for Navigation Mark Classification	ResnetMultiscale Attention model based on Renet 50 model	Automated feature extraction
Bilinear CNN Models for Fine-grained Visual Recognition	Bilinear convolutional neural network with softmax classifier	Bilinear vector, Gaussain mixture model(GMM)
Fast CNN Surveill-ance Pipeline for Fine-Grained Vessel Classification &Dete-ction in Maritime Scenarios	Convolutional neural network pipeline with Support vector machine classifier	Histogram of oriented gradients(HOG)
The Application of Two-level Attention Models in Deep Convolutional Neural Network	Part level and object level attention model in deep convolutional neural network	Object level FilterNet, DomainNet
Maritime Ship Targets Recognition with DeepLearning	Faster Region based convolutional neural network	Softmax function, bounding box regression algorithm

4. RESULT ANALYSIS

Methods	Advantages	Disadvantages
Visual Recognition Based On Deep Learning for Navigation Mark Classification	Automated feature extraction	Unable to recognise marks during night time.
Bilinear CNN Models for Fine-grained Visual Recognition	Process images of an arbitary size in a single forward propagation	Poor feature extraction in large dataset
Fast CNN Surveill-ance Pipeline for Fine-Grained Vessel Classification &Dete-ction in Maritime Scenarios	High accuracy on complex maritime scenarios	High cost and complex network
The Application of Two-level Attention Models in Deep Convolutional Neural Network	High performance in classification	Expensive and Non scalable
Maritime Ship Targets Recognition with DeepLearning	High accuracy for large ships.	Not good for small and weak ships.

5. CONCLUSION

Fine grained image classification aims to recognise subordinate level categories under some basic level category. The paper surveys some recent developments in fine-grained image recognition and semantic segmentation focused on deep learning. Like AlexNet, VGG net, and GoogLeNet, some general convolutionary neural networks were first implemented. They can be specifically tailored to the classification of find-grain pictures. Since there are typically subtle variations in some common sections of visually similar fine-grained artefacts, many approaches turn to deep learning technologies to improve part localization efficiency, while some approaches incorporate part localization into the deep learning system and can be trained end-to-end. In this paper, different applications of fine grained image classification and their comparison are discussed. To achieve further classification power for fine-grained images, some fine-grained classification approaches often incorporate multiple neural networks. Some visual attention-based methods will instantly find the most discriminatory regions of the fine-grained images by incorporating the attention mechanism, without using any bounding box or component annotation.

6. REFERENCES

[1] Mingyang Pan, Yisai Liu, Jiayi Cao, A. Koc and Chi-Hua Chen, `Visual recognition based on Deep Learning for Navigation Mark Classification''IET Comput. Vis., vol. 12, no. 8, pp. 1121_1132, Feb. 2020.

[2] S.-J. Lee, M.-I. Roh, H. Lee, J.-S. Ha, and I.-G. Woo, ``Image-based ship detection and classification for unmanned surface vehicle using real-time object detection neural networks,'' in Proc. Process. 28th Int. Ocean PolarEng. Conf., Sapporo, Japan, Jun. 2018, pp. 726_730.

[3] H. Fu, Y. Li, Y. Wang, and P. Li, ``Maritime ship targets recognition with deep learning,'' in Proc. 37th Chin. Control Conf. (CCC), Wuhan, China, Jul. 2018, pp. 9297_9302.

[4] X. He and Y. Peng, ``Fine-grained image classification via combining vision and language,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recog-nit. (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 7332_7340.

[5] Y.-L. Tang and N.-N. Shao, ``Design and research of integrated information
platform for smart ship,'' in Proc. 4th Int. Conf. Transp. Inf. Saf. (ICTIS),
Banff, AB, Canada, Aug. 2017, pp. 37_41.

[6] F. Bousetouane and B. Morris, ``Fast CNN surveillance pipeline for fine grained vessel classification and detection in maritime scenarios,'' in Proc. 13th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS), Colorado Springs, CO, USA, Aug. 2016, pp. 242_248.

[7] F. Bousetouane and B. Morris. Off-the-shelf cnn features for fine-grained classification of vessels in a maritime environment. In Advances in Visual Computing, pages 379–388.Springer, 2015.

[8] T.-Y. Lin, A. Roy Chowdhury, and S. Maji, ``Bilinear CNN models for fine grained visual recognition,'' in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Santiago, Chile, Dec. 2015, pp. 1449_1457.

[9] T. Xiao, Y. Xu, K. Yang, J. Zhang, Y. Peng, and Z. Zhang, ``The application of two-level attention models in deep convolutional neural network for fine-grained image classification,'' in Proc. IEEE Conf. Comput. Vis.Pattern Recognit. (CVPR), Boston, MA, USA, Jun. 2015, pp. 842_850.

[10] K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. In Computer Vision–ECCV 2014, pages 346–361. Springer,
2014.

[11] S. Branson, G. Van Horn, P. Perona, and S. Belongie, ``Improved birdspecies recognition using pose normalized deep convolutional nets,'' inProc. Brit. Mach. Vis. Conf., Nottingham, U.K., Sep. 2014, pp. 1_14.

[12] N. Zhang, J. Donahue, R. Girshick, and T. Darrell, ``Part-based R-CNNs for fine-grained category detection,'' in Proc. Eur. Conf. Comput.Vis. (Lecture Notes in Computer Science), Zürich, Switzerland, vol. 8689,Sep. 2014, pp. 834_849.

[13] N. Zhang, J. Donahue, R. Girshick, and T. Darrell.Part-based r-cnns for fine-grained category detection.In ECCV. 2014.

[14]Y. Gong, L. Wang, R. Guo, and S. Lazebnik. Multi-scaleorderless pooling of deep convolutional activation features.In ECCV, 2014.

[15] Y. Chai, V. Lempitsky, and A. Zisserman.Symbioticsegmentation and part localization for fine-grainedcategorization.In ICCV, 2013.
[16] H. J´egou, M. Douze, C. Schmid, and P. P´erez.Aggregatinglocal descriptors into a compact image representation.InCVPR, 2010.

[17] G. Csurka, C. R. Dance, L. Dan, J. Willamowski, and C. Bray.Visual categorization with bags of keypoints. InECCV Workshop on Stat. Learn. in Comp. Vision, 2004.

BIOGRAPHIES

Haripriya S is currently pursuing M Tech in Computer Science and Engineering at Musaliar College of Engineering and Technology, Pathanamthitta, Kerala, INDIA. She has received B Tech degree in Computer Science and Engineering from College of Engineering Adoor, Pathanamthitta, Kerala, INDIA. Her research area of interest includes the field of Artificial Intelligence, machine learning and Internet of Things.

Dr.Teena Joseph, Associate Professor , Computer Science & Engineering Department, Musaliar College of Engineering , Pathanamthitta, Kerala. India. She Completed her B.Tech & M.E 2006, 2010 respectively. Her area of interest are Network Security and Machine Learning, She completed her Ph.D in 2018 .She started her carrier in 2007 at Mar Baselious College of Engineering Trivandrum. She is having of 12 years of teaching experience in various colleges. She published 10 articles in various SCI & Scopus Journals.