Neurocomputing 257 (2017) 97–103

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A model for fine-grained vehicle classification based on deep learning

Shaoyong Yu[a,b], Yun Wu[b], Wei Li[b], Zhijun Song[c], Wenhua Zeng[a,*]

[a] Department of Cognitive Science, Xiamen University, Xiamen 361000, China
[b] School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China
[c] The 28th Research Institute of China Electronics Technology Group Corporation, Nanjing 210007, China

ARTICLE INFO

Article history:
Received 7 May 2016
Revised 29 June 2016
Accepted 7 September 2016
Available online 7 February 2017

Keywords:
Fine-graind classification
Deep learning
Vehicle detection
Network collabrative annotation

ABSTRACT

A model for fine-grained vehicle classification based on deep learning is proposed to handle complicated transportation scene. This model comprises of two parts, vehicle detection model and vehicle fine-grained detection and classification model. Faster R-CNN method is adopted in vehicle detection model to extract single vehicle images from an image with clutter background which may contains serval vehicles. This step provides data for the next classification model. In vehicle fine-grained classification model, an image contains only one vehicle is fed into a CNN model to produce a feature, then a joint bayesian network is used to implement the fine-grained classification process. Experiments show that vehicle's make and model can be recognized from transportation images effectively by using our method. Furthermore,in order to build a large scale database easier, this paper comes up with a novel network collaborative annotation mechanism.

1. Introduction

With the promotion of Intelligent Transportation System(ITS) in intelligent city, core technologies applied in ITS develop rapidly and have been updated constantly. In 1970s, only magnetic coils were used to detect vehicles, but now, other technologies like radar, ultrasonic, infrared rays and video image are very popular in practice [1]. As more and more digital video surveillances have been equipped to transportation roads, so visual vehicle detection methods have become research issues of computer vision scientists recently [2].

As an application domain of object detection, vehicle detection plays an important role in ITS, unmanned intelligent car and public security [3]. After detection, we can further classify them in more detail, if applied in public security, it can helps to arrest criminals quickly. Even if the criminals drive away, according to car make, model, color and plate number, we can start all the cameras in the city, which can automatically detect, recognize and locate the car. In this scene, fine-grained classification of vehicle is indispensable.

But in fact, object's intra-class difference is subtle, even sometimes intra-class difference is bigger than inter-class [4] [5], so the research subject of fine-grained classification is very challenging, which can advance the development of face detection [6], action recognition [7] and automatic scene description [8] and so on.

If fine-grained classification of vehicle is applied in transportation and public security, we can acquire more meta information like vehicle make, model, logo, production year, max speed and acceleration and so on [9]. By acquiring these information dynamically, we can build a large intelligent transportation system that can monitor the whole city's road. Further, we can analyse the vehicles on the road at different time to find the discipline of people's going out, then we can schedule transportation rules accordingly, these will make cities more smart and intelligent.

2. Related work

In this paper we focus on three issues, they are how to build a large scale vehicle dataset; how to detect vehicles in natural images; how to fine-grained classify vehicles.

2.1. Dataset generating method

With the availability of large scale training dataset, deep convolutional neural networks based approaches have recently been substantially improving upon the state of the art in image classification [10–15],object detection [12,16,17],and many other recognitions tasks[18–20] [21]. But in early times, lack of datasets and limited computation ability of CPU/GPU restrict CNN only to be applied in small problem domain like digital hand written digit recognition. So we can conclude that training dataset is essential to CNN model. There are a lot of datasets publicly available now, from small scale to large scale. Small image datasets like Caltech101/256

* Corresponding author.
E-mail address: syyu@xmut.edu.cn (W. Zeng).

http://dx.doi.org/10.1016/j.neucom.2016.09.116
0925-2312/

[22,23], MSRC [24], PASCAL [25] have served as training and evaluation benchmarks for most of todays computer vision algorithms. As computer vision research go further, larger datasets are needed. So, datasets like TinyImage [26] which has 80 million images, all these images are acquired from image search engines like Google, Baidu, Bing and so on by using keywords. Other larger datasets like LableMe [27] , Lotus Hill [28] and ImageNet [29] provide 30k, 50k and 50 million labeled and segmented images respectively, which need massive people to annotate. Despite of large quantity of images, it is still not enough for deep learning model. Neural network architecture usually has millions of parameters, existed datasets turn out to be insufficient to learn so many parameters without considerable overfitting. So Researchers take some technical methods like cropping [30], resizing [10] [31], mirror reflection [32] [33] to augment the existed datasets. So we can jump to the conclusion that to build a large scale dataset, one should first collect massive images by internet search engines, and then employ lots of people to annotate them, after that technical methods are used to augment the datasets. Even though, images in the datasets cannot cover all the views of a specified type object.

2.2. Vehicle detection method

With the development of computer image processing technologies, vehicle detection develops rapidly. In the case of whole vehicle detection, a good result has been obtained, in[16,34–37], Girshick R has achieved a detection accuracy of 49.1 percentage under complicated background in VOC2007. He, Zhang [34] improved [16]'s work, has promote the speed to 20–60 times without reducing the detection accuracy. Girshick R[35] further improved the detection accuracy to 66 percentage as well as the speed to 10 times on the basis of Ren[a] s, He[34]'s work,the processing speed can achieve 3 images per second.Once again, Ren S, He K advance the processing speed to 5 images per second in [36]. Though a lot of work has been done in vehicle detection, [16,34–37] aim mainly at generic object detection problem, not specialize in vehicle detection, its detection accuracy and speed is far away from 30 images per second in engineering application with nearly 100 percentage accuracy.

2.3. Fine-grained vehicle classification method

As concerns of fine-grained object classification, current researches generally focus on domain of bird [15,38] [39–43], cat, flower, aeroplane [43], dog [15,38,44], pedestrian [45] and action recognition [46]. For the problem of fine-grained vehicle detection,in [40,43,44] only cover a little, we only found [47] specialize in this field.In [47], a large dataset for fine-grained vehicle detection was first builded, and then a CNN model was used to achieve a Top-1 accuracy of 76.7 percentage,top-5 accuracy of 91.7 percentage. In [47], Linjie got a good result on fine-grained vehicle detection, which use a dataset consists of 161 car models and 136,727 images and all these images were manually collected. But for deep learning network, this quantity of images is not enough, furthermore, we notice that the images fed into the model contains only single vehicle, while in reality, images are usually have clutter background and uncertain number of vehicles. So to get that clean dataset is time-consuming and costly.

In this paper, we come up with a novel method which can do three following things. Firstly, we use a network collaborative annotation mechanism to generate massive annotated vehicle images that can build a continuously growing dataset. Secondly, we use a Faster R-CNN based model to detect vehicles in the existing dataset and then generate images with only one vehicle. At last, we use a CNN model and joint bayesian network to classify vehicles in fine-grained [9].

3. Overall architecture

Natural transportation images usually contain uncertain number of vehicles. If we want to do fine-grained classification on these images, we must first extract all the vehicles. And then, for these extracted images, we use a CNN model to compute features, with which we can classify vehicles easily. The overall architecture is as Fig. 1. This fine-grained vehicle classification model takes an original image with complicated background as input, which is first fed into the vehicle detection model, then a series of sub images contain only single vehicle will be produced by detection model. All the sub images are then transferred to the next classification model, at last all the meta information like make,model of the vehicle will be acquired.

4. Implementation detail

To implement effective fine-grained vehicle classification by using deep learning method, three following question should be solved. (1) How to build a large scale dataset suitable for fine-grained vehicle classification. (2) In what way can we get detect vehicles in images with cluttered background, and then extract them to provide clean input to the subsequent classification model. (3)How to recognize different vehicle key parts first, and then joint all parts together to make a fine-grained classification.

4.1. Build a large scale dataset

Effective vehicle dataset is the source of knowledge acquirement for the network model [48], in order to collect more precise dataset, an automatic dataset collection and network collaborative annotation mechanism is employed in this work. The process of this mechanism is as Fig. 2. In this paper, we implement a software which can automatically acquire vehicle images from internet by using open source search engine Nutch. Images then be stored in a B/S collaborative annotation platform's database, users can access this platform by web browser, the platform will provide users with some un-annotated images for annotation. If an image contains no vehicle, this image will be annotated as negative sample, or the whole vehicle regions and vehicle part regions like front windshield, rear windshield, ceiling, side, logo, headlamps, taillight, fog lamp, air inlet and so on will be marked. Different regions with different color, the platform will store a coordinate and a label for each region. Besides, users can also upload images themselves for annotation. In this way, we can get a continuous increasing dataset. In this work we also propose a novel method that can generate massive images automatically, and these generated images need no annotation. An AutoCAD 3D model of a car of specified make and model is imported into our system, and then images of different views of the car are generated, these images can describe all the details of a car [49] [50] [51]. For fine-grained vehicle detection, we use an AutoCAD 3D model of a car of specified make and model. By changing the camera distance, direction angle and over angle we can get tens of thousands of images of car with different appearance. This sort of images needs no annotation, which can save our effort. First, 3D car model was put on the place where car floors center point coincide with the original point. Then we use three parameters to adjust the camera view, which can generate different 2D image. These three parameters are distance, direction angle and over angle , where ranges from 0° to 360°, changes between 0° and 90° because of symmetry [52]. So we can get the camera coordinate as follows, refer to Fig. 3.

$$x = dis\tan ce * \cos(\alpha/180.0 * \pi) * \sin(\beta/180.0 * \pi)$$
$$y = dis\tan ce * \cos(\alpha/180.0 * \pi) * \cos(\beta/180.0 * \pi)$$
$$z = dis\tan ce * \sin(\alpha/180.0 * \pi)$$

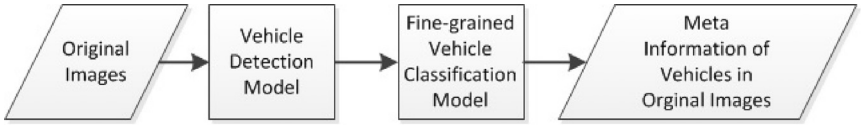

Fig. 1. The overall architecture.

Fig. 2. Process of dataset automatic collection and collaborative annotation mechanism.

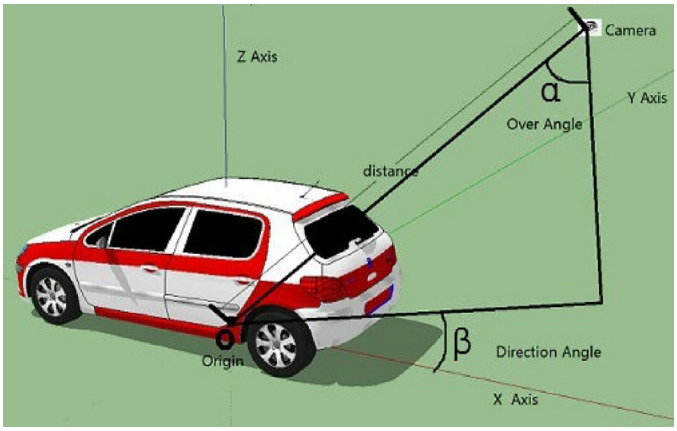

Fig. 3. The principle of generating car images of all views.

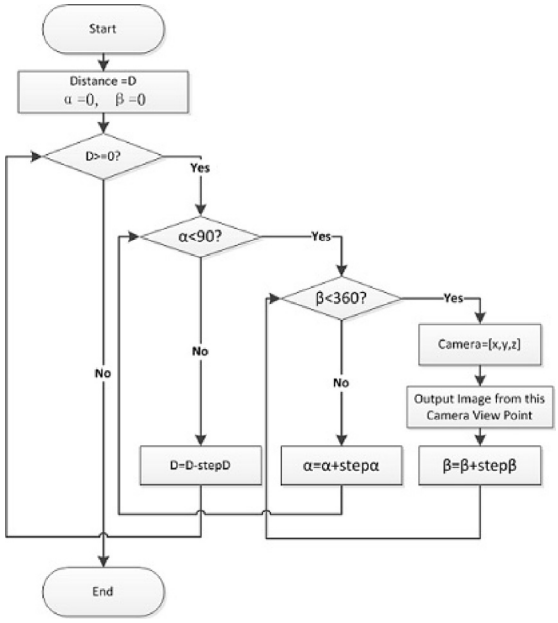

Fig. 4. Algorithm Process of Generating 2D Images from 3D Model.

And then we get massive images of this car model by the following steps as Fig. 4 shows. We import a 3D car model, and then set Camera distance as constant D, which decrease at a step of stepD, set over angle α and direction angle β as 0, which increase at a step of stepα and stepβ respectively. In each loop, when D >0, α <90 and β <360, we can get from above Equation an coordinate, make it cameras position, and then output a 2D image of this car model. To be easily distinguished, we name the image file in the form of "D-α – β.jpg". Some images generated by this algorithm are as Fig. 5.

4.2. Vehicle detection model

Vehicle detection model can detect all the vehicles in images with complicated background and then extract them as sub images. The process of this model is as Fig. 6. Original image is first fed into a convolutional network which uses VGG16 network structure as depicted in [36] [53]. And then feature maps of the original image will be generated, on which a RPN network is applied to acquire region proposals. After that we use a ROI Pooling layer to obtain region proposals on the original image accordingly, which then transferred to a vehicle SVM classifier to judge whether these region proposals are vehicles or not.

4.2.1. Convolutional neural network

Convolutional neural network take original image as input and output corresponding feature maps. The number of layers of network will directly influence on effectiveness of data processing, but this argument is an empirical value, it will reduce the ability of processing if set to a small one, on the other way, if set to a big one, it will make the whole network too complicated. So in this paper, we employ 5 convolutional layer nodes, where each node is a form of stack followed by a MaxPooling layer. The structure of convolutional neural network is as Fig. 7. In this structure, each convolutional layer adopts a small 3*3 region as receptive field with a step of 1 pixel. So a convolutional stack which contains 3 convolutional layer will have a receptive of 7*7 with the reduction of network parameters.

4.2.2. Region proposal network

RPN takes in the feature maps of the original image, and then output a series of proposal regions. Each proposal region denoted with two parameters, that is a coordinate and a probability. RPN use a fully connected network and then scan on the feature maps with a small network which fully connected with a n*n window, then mapped to a 256 vector, next this vector is send to two fully connected classification layer and regression layer. Probability will be obtained in classification layer and coordinate will be gained in regression layer. As depicted in Fig. 8.

5. Vehicle fine-grained classification model

Vehicle fine-grained classification model use an image which contains only a single vehicle as an input to another classification CNN model to generate features, then with a joint bayesian network, the contained vehicle will be correctly classified [54]. As shown in Fig. 9. CNN feature and joint bayesian network have been proved to be effective in face recognition as described in [55]. In this paper, we treat a vehicle as two parts, one part is the inter-difference of different vehicle models, another part is the intra-difference of the same vehicle model, vehicle images taken from

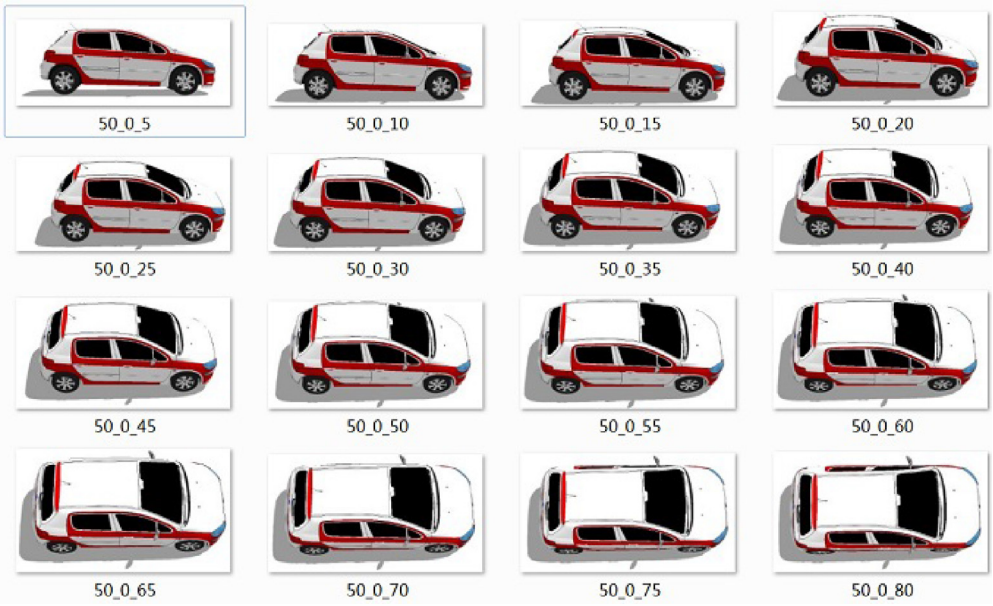

Fig. 5. Images Generated by 3D Model.

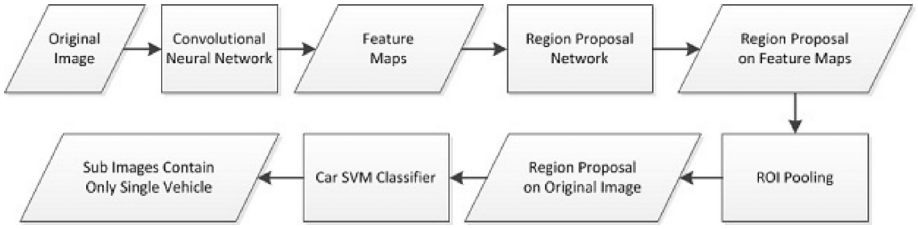

Fig. 6. Process of Vehicle Detection Model.

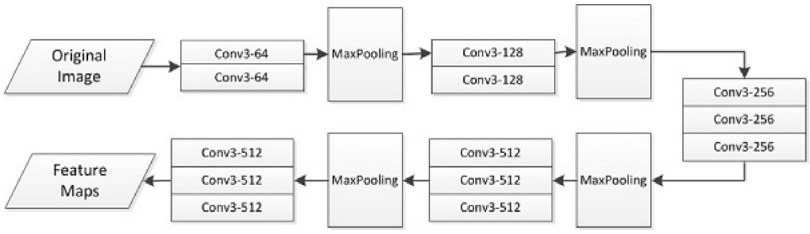

Fig. 7. Architecture of Convolutional Neural Network.

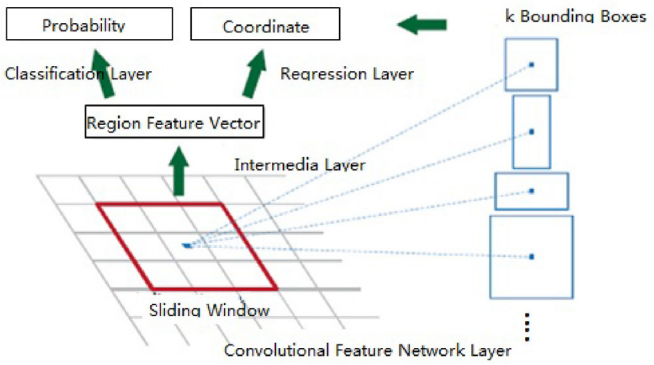

Fig. 8. Architecture of Region Proposal Network.

different views for example. Feature x is represented as the sum of two Gaussian variables.

$$x = \mu + \varepsilon$$

In this equation,

$$\mu \sim N(0, S_\mu)$$
$$\varepsilon \sim N(0, S_\varepsilon)$$

The first variable μ indicates the inter-difference of different vehicle model, and the second variable ε is the intra-difference of the same vehicle model. Given the Gaussian hypothesis of the inter-difference and intra-difference, a joint bayesian network can be used to calculate joint probability of these two features which also follow gaussian distribution.

$$\sum I = \begin{pmatrix} S_\mu + S_\varepsilon & S_\mu \\ S_\mu & S_\mu + S_\varepsilon \end{pmatrix}$$
$$\sum E = \begin{pmatrix} S_\mu + S_\varepsilon & 0 \\ 0 & S_\mu + S_\varepsilon \end{pmatrix}$$

S_μ and S_ε can be computed by EM algorithm.

$$r(x_1, x_2) = \log \frac{P(x_1, x_2|H_I)}{P(x_1, x_2|H_E)}$$

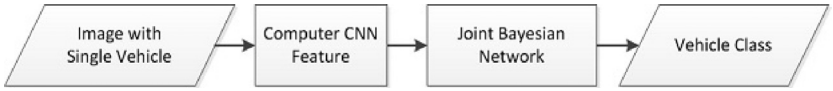

Fig. 9. Process of fine-grained vehicle classification.

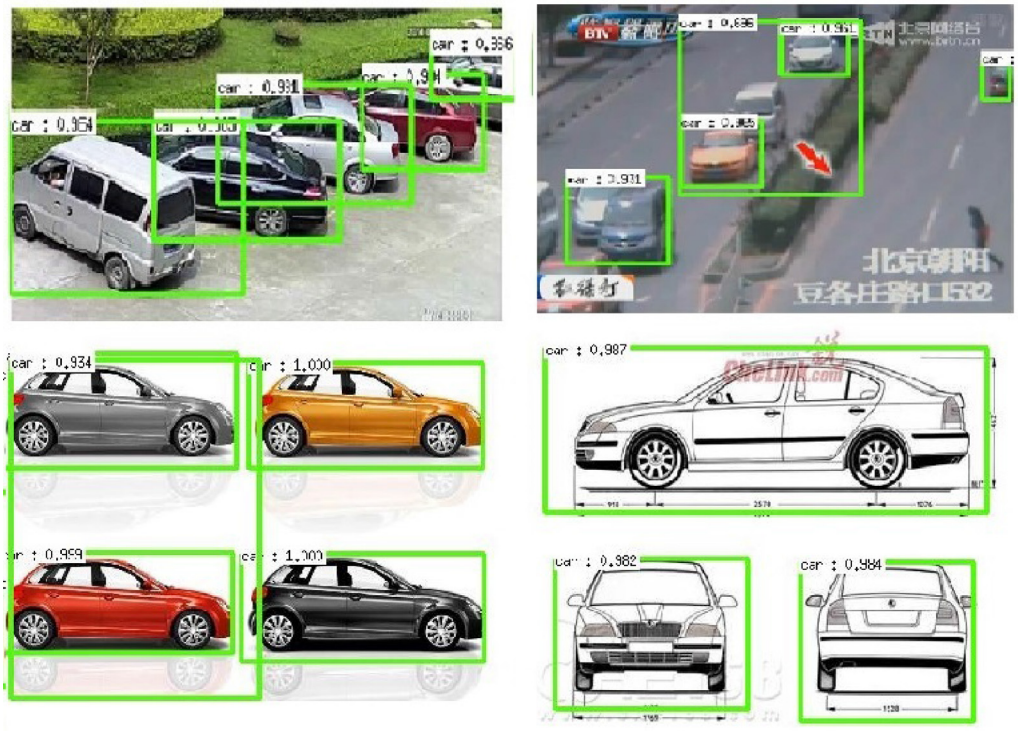

Fig. 10. Some good vehicle detection result.

By using the above equation, the similarity of two images can be computed to judge whether they belong to the same vehicle class.

6. Experimental result

Dataset used in this paper consists of three parts, whole vehicle dataset, vehicle parts dataset and non-vehicle dataset. The whole dataset is used to train faster R-CNN to build a vehicle detection model, including 73 different vehicle makes and 208 different vehicle models, 174,008 images in all; vehicle parts dataset is employed to train fine-grained classification network which can classify different models of vehicles, this dataset totally contains 33,023 images of headlamp, tail light, fog lamp, logo and air inlet; non-vehicle dataset contains 100,000 images download from the internet.

6.1. Vehicle detection result

The dataset of vehicle detection experiments mainly consists of whole vehicle dataset and non-vehicle dataset. For whole vehicle dataset, we divide it into two parts, 120,000 of them are used as positive training dataset, 54,008 of them are positive test dataset. Non-vehicle dataset also splitted into two subsets, 70,000 of them are treated as negative training dataset, 30,000 of them are negative test dataset. So, we have 190,000 images of training dataset and 84,008 images of test dataset in all. In our experiment, we use a single GTX TITAN X graphic card to get an accuracy of 85 percentage of vehicle detection at the speed of 5 images per second. Some good result are as Fig. 10 .

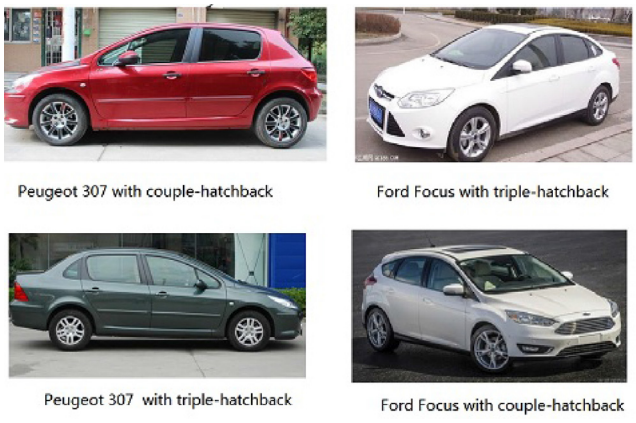

Fig. 11. Some good classification result.

6.2. Fine-grained classification result

From the generated images by detection model which contain only one vehicle per image,we select 200,000 images that basically cover all 208 car models, and then make 180,000 of them as training data, 20,000 of them as test data, plus 33,023 images of vehicle parts as training data,we did our fine-grained classification experiment, at last we got an accuracy of 89 percentage of classification. Some classification result is shown in Figs. 11 and12. As shown in Fig. 11, our fine-grained classification model can tell vehicles make, model and structure under complicated background. We also notice that, in Fig. 12, the vehicle on the left hand side is

Fig. 12. Some false classification result.

Volkswagen Touareg indeed, but on the right hand side is Zongtai T600 which is looks very similar to Volkswagen Touareg,and our model made a mistake.

7. Conclusion

In this paper, a faster R-CNN based vehicle detection model is first used to detect vehicles in images with complicated background, and then the detection result is fed into a fine-grained vehicle classification model which can classify vehicles in more detail. While the vehicle detection model cannot detect all the vehicles in images at 100 percentage, and sometimes it will even deem non-vehicle regions as vehicles, so that has influence on the accuracy of fine-grained classification. At the meantime, the quantity of training dataset will also impact the classification accuracy, overfitting phenomenon will occurs when the dataset is small, which make the model generalize badly. We notice that faster R-CNN is a model aimed at generic object detection, can we take more priors into consideration to improve the network structure to promote the accuracy and speed of detection? In the experiment of fine-grained classification, our model cannot tell vehicle models which appearance is too similar to each other, but in fact, there are some obvious small feature that can distinguish them, let's say vehicle logo for example. Can we add some notable feature classifiers to the classification model so as to advance the accuracy? All these problems are what we will concern in the future.

Acknowledgment

This work is supported by Natural Science Foundation of Fujian Province of China(Grant No.2013J05103 and No. 2016J01325 and No. 2015J05015) and High-level Personnel of Support Program of Xiamen University of Technology (Grant NO. YKJ14014R) and B Program of Education Department of Fujian Province(Grant No.JB13152).

References

[1] Editorial, Special issue on big data driven intelligent transportation systems, Neurocomputing 181 (2) (2016) 1–3.
[2] J. Yu, D. Tao, Y. Rui, M. Wang, Learning torank using user clicks and visual features for image retrieval, IEEE Trans. Cybern. 45 (4) (2015) 767–779.
[3] Q. Ge, T. Shao, C. Wen, R. Sun, Analysis on strong tracking filtering for linear dynamic systems, Math. Probl. Eng. 2015 (6) (2015) 1–9.
[4] Z. Lu, L. Wang, J. Wen, Image classification by visual bag-of-words refinement and reduction, Neurocomputing 173 (2016) 373–384.
[5] L. Xie, J. Wang, B. Zhang, Q. Tian, Incorporating visual adjectives for image classification, Neurocomputing 182 (2016) 48–55.
[6] S.A.A. Shah, M. Bennamoun, F. Boussaid, Iterative deep learning for image set based face and object recognition, Neurocomputing 174 (2016) 866–874.
[7] N. Nedjah, F.P. Silva, A.O. Sa, L. M.Mourelle, D. A.Bonilla, A massively parallel pipelined reconfigurable design for m-pln based neural networks for efficient image classification, Neurocomputing 183 (C) (2016) 39–55.
[8] Q. Ge, D. Xu, C. Wen, Cubature information filters with correlated noises and their applications in decentralized fusion, Signal Process. 94 (1) (2014a) 434–444.
[9] Q. Ge, C. Wen, S. Duan, Fire localization based on range-range-range model for limited interior space, IEEE Trans. Instrument. Meas. 63 (9) (2014b) 2223–2237.
[10] K. Alex, I. Sutskever, G. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2) (2012) 248–255.
[11] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: ECCV, 2014, pp. 818–833.
[12] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Overfeat: integrated recognition, localization and detection using convolutional networks, Eprint Arxiv (2013) 114–130.
[13] A. Chatfield, K. Simonyan, A.Zisserman, Return of the devil in the details: delving deep into convolutional nets, Eprint Arxiv (2014) 3531–3544.
[14] C. Hong, J. Yu, J. Wan, D. Tao, M. Wang, Multimodal deep autoencoder forhuman pose recovery, IEEE Trans. Image Process. 24 (12) (2015) 5659–5670.
[15] D. Lin, X. Shen, C. Lu, Deep lac:deep localization ,alignment and classification for fine-grained recognition, in: CVPR, 2015, pp. 1666–1674.
[16] G. Ross, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: CVPR, 2014, pp. 580–587.
[17] Z.W. Y, W. X, S. M, Generic object detection with dense neural patterns and regionlets, Eprint Arxiv (2014) 662–678.
[18] Razavian, A. Sharif, Cnn features off-the-shelf: an astounding baseline for recognition, in: Computer Vision and Pattern Recognition, 2014, pp. 512–519.
[19] R.M. Taigman Y Yang M, Deepface: closing the gap to human-level performance in face verification, in: Computer Vision and Pattern Recognition, 2014, pp. 1701–1708.
[20] R.M. Zhang N Paluri M, Deepface: closing the gap to human-level performance in face verification, in: Computer Vision and Pattern Recognition, 2014, pp. 1637–1644.
[21] G. Y, W. L, G. R, Multi-scale orderless pooling of deep convolutional activation features, Lecture Notes in Computer Science, 2014, pp. 392–407.
[22] L. Fei-Fei, R. Fergus, P. Perona, One-shot learning of object categories, in: PAMI, 2006, pp. 594–611.
[23] G. Grifn, A. Holub, P. Perona, Caltech-256 object category dataset, Techical Report 7694,Caltech(2007) 100–132.
[24] J. Shotton, J. Winn, C. Rother, A. Criminisi, Textonboost: joint appearance, shape and context modeling for multi-class object recognition and segmentation, in: ECCV, 2006, pp. 1–15.
[25] M. Everingham, L.V. Gool, C.K.I. Williams, J. Winn, A. Zisserman, The pascal visual object classes challenge 2008 results, (2008) http://www.pascal-network.org/challenges/VOC/voc2008/workshop/ 101–112.
[26] A. Torralba, R. Fergus, W. Freeman, 80 million tiny images: a large data set for nonparametric object and scene recognition, in: PAMI, 2008, pp. 1958–1970.
[27] B. Russell, A. Torralba, K. Murphy, W. Freeman, Labelme:a database and web-based tool for image annotation, Int. J. Comput. Vis. (2008) 157–173.
[28] B. Yao, X. Yang, S. Zhu, Introduction to a large-scale general purpose ground truth database: Methodology, annotation tool and benchmarks, in: CVPR, 2007, pp. 169–183.
[29] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: a large-scale hierarchical image database, in: CVPR, 2009, pp. 201–220.
[30] H. Kaiming, Spatial pyramid pooling in deep convolutional networks for visual recognition, in: TPAMI, 2015, pp. 1904–1916.
[31] Chmidhuber, Jurgen, Multi-column deep neural networks for image classification, in: CVPR, 2012, pp. 3642–3649.
[32] J. Yu, Y. Rui, D. Tao, Click prediction for web imagereranking using multimodal sparse coding, IEEE Trans. Image Process. 23 (5) (2014a) 2019–2032.
[33] J. Yu, Y. Rui, Y. Tang, D. Tao, High order distance based multiview stochastic learning in image classification, IEEE Trans. Cybern. (2014b) 2431–2442.
[34] H. K, Z. X, R. S, Spatial pyramid pooling in deep convolutional networks for visual recognition, in: ECCV, 2014, pp. 346–361.
[35] G. R, Fast r-cnn, in: CVPR, 2015, pp. 301–309.
[36] R. S, H. K, G. R, Faster r-cnn: Towards realtime object detection with region proposal networks, in: NIPS, 2015, pp. 442–451.
[37] R. J, D. S, G. R, You only look once:unified, real-time object detection, arXiv preprint (2015) 1222–1231.
[38] T. Xiao, Y. Xu, K. Yang, The application of two-level attention models in deep convolutional neural network for fine-grained image classification, in: CVPR, 2015, pp. 842–850.
[39] Z. Akata, S. Reed, D. Walter, Evaluation of output embeddings for fine-grained image classification, in: CVPR, 2015, pp. 2927–2936.
[40] J. Krause, HailinJin, J. Yang, L. Fei-Fei, Fine-grained recognition without part annotations, in: CVPR, 2015, pp. 5546–5555.
[41] G.V. Horn, S. Branson, R. Farrell, S. Haber, Building a bird recognition app and large scale dataset with citizen scientists:the fine pint in fine-grained dataset collection, in: CVPR, 2015, pp. 595–604.
[42] N. Zhang, J. Donahue, R. Girshick, T. Darrell, Part-based r-cnns for fine-grained category detection, in: LNCS, 2014, pp. 834–849.
[43] T.-Y. Lin, A. Chowdhury, S. Maji, Bilinear cnn models for fine-grained visual recognition, in: ICCV, 2015, pp. 1449–1457.
[44] S. Xie, T. Yang, X. Wang, Y. Lin, Hyper-class augmented and regularized deep learning for fine-grained image classification, in: CVPR, 2015, pp. 2645–2654.
[45] D. Hall, P. Perona, Fine-grained classification of pedestrians in video: benchmark and state of the art, in: CVPR, 2015, pp. 5482–5491.
[46] Y. Zhou, B. Ni, R. Hong, M. Wang, Q. Tian, Interaction part mining:a mid-level approach for fine-grained action recognition, in: CVPR, 2015, pp. 3323–3331.
[47] L. Yang, P. Luo, C.C. Loy, X. Tang, A large-scale car dataset for fine-grained categorization and verification, in: CVPR, 2015, pp. 3973–3981.
[48] Q. Ge, T. Shao, Q. Yang, X. Shen, C. Wen, Multisensor nonlinear fusion methods based on adaptive ensemble fifth-degree iterated cubature information filterfor biomechatronics, IEEE Trans. Syst., Man Cybert.:Syst. 46 (7) (2015) 1–14.
[49] C. Hong, J. Yu, Y. Jane, Z. Yu, X. Chen, Three-dimensional image-based human pose recovery with hypergraph regularized autoencoders, Multim. Tools Appl. 2016 (1) (2016) 1–19.

[50] C. Hong, X. Chen, X. Wang, C. Tang, Hypergraph regularized autoencoder for image-based 3d human pose recovery, Signal Process. 2015 (1) (2015a) 7–19.
[51] C. Hong, J. Yu, D. Tao, M. Wang, Image-based 3d human pose recovery by multi-view locality sensitive sparse retrieval, IEEE Trans. Ind. Electron. 2015 (1) (2015b) 3742–3751.
[52] J. Yu, D. Tao, J. Li, J. Cheng, Semantic preserving distance metric learning and applications, Inf. Sci. 281 (2014) 674–686.
[53] Y. Zhang, L. Zhang, P. Li, A novel bilogically inspired elm-based network for image recognition, Neurocomputing 174 (2016) 286–298.
[54] S. Miao, J. Wang, Q. Gao, F. Chen, Y. Wang, Discriminant structure embedding for image recognition, Neurocomputing 174 (PB) (2016) 850–857.
[55] S. Yi, W. Xiaogang, T. Xiaoou, Deep learning face representation from predicting 10,000 classes, in: CVPR, 2014, pp. 1891–1898.

Shaoyong Yu, Male, a PhD student in Xiamen University. His research interests lie in the areas of computer vision and deep learning. Contact him at syyu@xmut.edu.cn.

Yun Wu, female, PhD. She received her PhD from Xiamen University in 2007. Her research interests lie in the areas of artificial intelligence and big data. His scientific contribution to the AI has more to do with soft computing and the clustering algorithms.

Wei Li is an associate professor in the School of Computer and Information Engineering at Xiamen University of Technology. His research interests include artificial intelligence, computer graphics. He has a Ph.D. in Basic Theory of Artificial Intelligence from Xiamen University. Contact him at weili@xmut.edu.cn.

Song Zhijun, Male, PhD. He received his PhD from Xiamen University in 2013. His research interests lie in the areas of artificial intelligence and big data. His scientific contribution to the AI has more to do with machine consciousness and the logic of mental self-reflection.Contact him at jason@xmu.edu.cn.

Wenhua Zeng is a professor in the Department of Cognitive Science at Xiamen University. His research interests include neural network, grid computing and embedded system. He has a Ph.D. in Industry Automation from Zhejiang University in 1986. Contact him at whzeng@xmu.edu.cn.