INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

IJARIIT

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 8, Issue 2 - V8I2-1344)
Available online at: https://www.ijariit.com

Recognition of objects in Adverse Weather conditions using Dual Subnet Network in Comparison with CNN Network

K. Pavan Kumar
pavankumarroyal612@gmail.com
Annamacharya Institute of Technologies and Sciences, Rajampeta, Andhra Pradesh

K. Bhanu Teja
bhanuteja875@gmail.com
Annamacharya Institute of Technologies and Sciences, Rajampeta, Andhra Pradesh

B. Deepa
deeparajasekhar14@gmail.com
Annamacharya Institute of Technologies and Sciences, Rajampeta, Andhra Pradesh

M. Sagar
sagarmasani23@gmail.com
Annamacharya Institute of Technologies and Sciences, Rajampeta, Andhra Pradesh

C. Venkata Sai Rakesh
rakeshroyal14354@gmail.com
Annamacharya Institute of Technologies and Sciences, Rajampeta, Andhra Pradesh

I. Suneetha Rani
itikalasuneetha@gmail.com
Annamacharya Institute of Technologies and Sciences, Rajampeta, Andhra Pradesh

ABSTRACT

The purpose of this Research work is to Detect objects in Adverse weather conditions using Dual Subnet Network and comparing it with CNN network. Object detection algorithms based on convolutional neural networks have been intensively explored and successfully implemented in numerous computer vision applications during the last half-decade. However, due to poor visibility, recognising things in rainy weather remains a considerable challenge. In this study, we introduce an unique dual-subnet network (DSNet) that can be trained end-to-end and jointly perform three tasks: visibility improvement, object categorization, and object localisation, to handle the object identification problem in the presence of fog. By incorporating two subnetworks: detection and restoration, DSNet achieves complete performance enhancement. RetinaNet is used as a backbone network (also known as a detection subnet) for learning to categorise and find objects. The restoration subnet shares feature extraction layers with the detection subnet and uses a feature recovery (FR) module to improve visibility. Our DSNet outperformed many state-of-the-art object detectors and combination models between dehazing and detection methods while maintaining a high speed, obtaining 50.84 percent mean average precision (mAP) on a synthetic foggy dataset that we composed and 41.91 percent mAP on a public natural foggy dataset (Foggy Driving dataset).

***Keywords**: DualSubnet Network, CNN Network, Foggy data set, Backbone Network, Enhancement, Feature Extraction.*

1. INTRODUCTION

Object detection is critical in IAVs because it not only decides which category each object belongs to and locates things in a given image, but it also helps the system navigate safely in complex driving scenarios. Object detection algorithms based on deep convolutional neural networks (CNNs) [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12] have been introduced and have obtained outstanding results in recent years . Enhancing the visibility of hazy images has traditionally been employed as a preprocessing step to improve object detection performance in the presence of haze. Image dehazing benefits both human vision and numerous technologies that must work in a variety of weather situations, such as traffic surveillance systems, autonomous cars, and outdoor object identification systems. However, as an input to the object detector, employing images preprocessed by a superior dehazing model in terms of restoration performance does not always ensure enhanced object identification performance [13]. Furthermore, due of the added dehazing effort, merging the models employed by dehazing and detection approaches can reduce detection speed.

The anchor-based method makes temporal interest suggestions to determine and localise the representative contents of video sequences, whereas the anchor-free method skips the temporal proposals and forecasts importance scores and segment positions directly. Our interest detection framework is the first attempt to leverage temporal consistency via the temporal interest detection formulation, in contrast to existing supervised video. We first provide a dense sampling of temporal interest proposals with multi-scale intervals that accommodate interest fluctuations in length, and then extract their long-range temporal properties for interest proposal location regression and importance prediction using the anchor-based technique. Notably, both positive and negative segments are attributed to the resulting summary' accuracy and completeness information. We address the disadvantages of temporal suggestions by directly predicting significance scores of video frames and segment positions in the anchor-free approach. The interest detection approach, in particular, may be easily integrated into commercially available supervised video summarising systems. On the SumMe and TVSum datasets, we compare anchor-based versus anchor-free techniques. The effectiveness of both the anchor-based and anchor-free techniques has been demonstrated in experiments [14].

Object identification using deep convolutional neural networks (CNN) has been extensively researched in recent years, with impressive results. However, due to poor visibility, object identification in the presence of fog is still a work in progress. To solve the difficulty of recognising objects in foggy weather circumstances, this study introduces DFONet, a novel CNN-based object identification model. The DFO-Net is divided into two subnets: a defogging subnet and a detection subnet. The defogging subnet is in charge of extracting clear features from cloudy images and passing them along to the detection subnet. Object classification and object localisation are performed by the detection network using these resultant features as input [15].

2. LITERATURE SURVEY

Shih-Chia Huang [16] et al. proposed a unique dual-subnet network (DSNet) that may greatly increase object identification accuracy in bad weather while keeping a short prediction time. Our method does this by learning three tasks at the same time: visibility augmentation, object classification, and object location. There are two subnetworks in the proposed DSNet: 1) a detection subnet and 2) a restoration subnet. The detection subnet is powered by RetinaNet, a CNN that shares a common block (CB) module with the restoration subnet and is in charge of object classification and localisation. As illustrated in Fig. 1, the restoration subnet is created by attaching a proposed feature recovery (FR) module to the CB module for visibility enhancement. The CB module is shared by these two subnetworks so that the clean features (fC2) generated at this module can be used in both subnetworks during joint learning. The detection subnet can be used to train DSNet from start to finish and forecast items.

Lotfi Abdi [17] and colleagues developed a new framework for semantic labelling and object detection that combines principles from deep Convolutional Neural Network (CNN) for object detection and fully-connected Conditional Random Field (CRF) for segmenting and labelling. In particular, we create a new framework that leverages global image features to forecast detection and dramatically lowers background detection errors, as well as a paired CRF as a post-processing step to ensure spatial consistency in the structured prediction. Our unified framework can effectively use the advantages of leading approaches such as CRF and CNN for these two tasks by combining the consistency between final detection results of CNN and CRF based graphical models. Extensive tests on the PASCAL VOC 2007/2012 data sets show that our system is effective for Visual Develop competence.

Baohua Qiang [18] et al. suggested a semantic segmentation (SSOD)-based object detection system for pictures. To begin, we build a feature extraction network that combines the hourglass design network with the attention mechanism layer to extract and fuse multi-scale features, resulting in high-level features with rich semantic information. Second, the feature extraction job is used as an auxiliary task to enable multi-task learning by the system. Finally, multi-scale characteristics are employed to estimate the object's location and categorization. The results of the experiments reveal that our approach significantly improves object detection capability and consistently outperforms the other three comparison algorithms, and that the detection speed can exceed real-time, allowing for real-time identification.

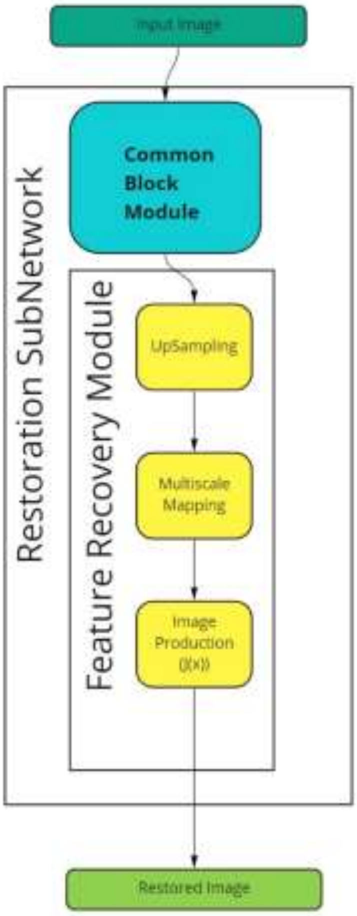
Jun Deng et al. aim for one of the most significant computer vision problems, target identification, which has become a hotspot for study in the last 20 years and is widely employed. Its goal is to quickly and reliably detect and locate a huge number of items in an input dataset that fall into predetermined categories. The algorithms can be classified into two types based on the model training method: single-stage detection algorithms and two-stage detection algorithms. The typical algorithms for each level are described in depth in this publication. Following that, numerous sample algorithms are evaluated and contrasted in this sector, as well as public and special datasets often utilised in target detection. Finally, potential target detection challenges are discussed.

HyperNet, developed by Kong et al. [23], integrates the generation of candidate areas with the detection task, resulting in fewer candidate regions with a greater recall rate. To solve the problem of overfitting and quality mismatch, Cai and Vasconcelos [24] suggested Cascade R-CNN. Deep learning-based object detection techniques may be classified into two types: two-stage detection algorithms and one-stage detection algorithms. The two-stage approach generates a region proposal first [26], then targets the region proposal's border box and category prediction. By utilising a deep ConvNet to categorise object proposals, Girshick et al. [19] introduced the traditional regions with convolutional neural networks (CNN) features (R-CNN) to obtain good object identification accuracy, but it is time-consuming. To address this issue, Girshick et al. [19,21] presented Faster R-CNN, an improved version of R-CNN that employed the region proposal network (RPN) to directly classify the region proposal in the convolutional neural network, achieving the entire detection framework's end-to-end goal. On the basis of Faster R-CNN, He et al. [22] presented Mask R-CNN, which added a branch for semantic segmentation tasks and employed detection and segmentation tasks to extract image characteristics to increase detection accuracy.

3. BLOCK DIAGRAM

The Restoration subnetwork's flowchart. The Feature Recovery Module is only triggered during training to create recovered images for the proposed multi-task learning technique, because the framework only concentrates on the object detection job.

The restoration subnetwork of our DSNet is in charge of creating the fC2 and sharing these characteristics with the detection subnetwork during cooperative learning to increase object detection accuracy in low-light circumstances.



This subnetwork completes the goal using two modules: 1) a CB module, and 2) a FR module, as illustrated in the diagram.

Haze can damage the features of the input image retrieved by the CB module, resulting in poor object detection performance. The restoration network suggests employing a FR module to restore fC2, which are shared to effectively reason about object recognition and object classifi- cation during joint learning. [38] inspired this FR module, which is made up of three submodules: 1) submodule for upsampling, 2) multiscale mapping (MM) submodule, and 3) image production (IP) submodule

4. EXISTING SYSTEM

In most cases, vision algorithms for driver assistance systems must meet stringent real-time limitations. As a result, we pay special attention to the real-time capabilities of the algorithms assessed here[13]. Deep CNN-based object detection algorithms have lately achieved significant progress on a number of object detection benchmarks. This work investigates ways to extract objects of interest without relying on handcrafted characteristics or sliding window techniques. Our strategy includes two concepts: To begin, the network predicts the positions and sizes of all objects from the entire image, and then segments the objects accurately based on the localizations.

Fig. A

Fig.B

Fig.C

Fig.A Image Testing
Fig.B Feature extraction using CNN Network
Fig.C Combining object detectors with CRFs for semantic picture segmentation

In the existing system by using CNN networks we can fail to detect the images in Adverse weather conditons these techniques are failed to detect the objects in inclement weather conditions because of that we are moving to Dual subnet network techniques.

5. PROPOSED SYSTEM

We present a unique dual-subnet network (DSNet) that significantly improves object detection accuracy in bad weather while maintaining a short prediction run time. Our method achieves this goal by teaching three tasks at the same time: visual enhancement, object classification, and object localization. There are two subnetworks in the proposed DSNet: 1) a detection subnet and 2) a restoration subnet. The detection subnet makes use of a CNN called RetinaNet [15], which shares a common block (CB) module with the restoration subnet and is in charge of object classification and localisation. As illustrated in Fig. 1, the restoration subnet is created by attaching a proposed feature recovery (FR) module to the CB module for visibility enhancement. These two subnetworks have a lot in common.

Fig.1 We create a FOD dataset by collecting photos from the public Foggy Cityscapes dataset [27] and categorising two classes, namely human and car, to conduct all of our tests. The FOD collection contains 6, 000 synthetic foggy photos carefully chosen from three Foggy Cityscapes versions: heavy fog, medium fog, and light fog, with attenuation coefficient values of 0.02, 0.01, and 0.005, respectively. According to the Cityscapes [28] annotation protocol, a total of 119, 979 object instances have been annotated, with 60, 279 person instances and 59, 700 car instances. We divided the FOD dataset into three sets: training, validation, and test; no photos were repeated in any group. Figure 3 shows example photos with bounding box annotations overlaid and statistics from the FOD dataset.

Fig1. Images and annotations from the FOD dataset as examples.

6. CONCLUSION

We introduced an innovative strategy to improving object identification performance in bad weather in this research. The detection and restoration subnetworks of our DSNet model, which may be trained end to end for combined learning of visibility improvement, object classification, and object localisation, are made up of two subnetworks. The detection subnet is created with RetinaNet, and the restoration subnet is created by connecting the FR module to the detection network's third residual block's last feature extraction layer. Our proposed approach had the best detection performance on both synthetic and natural foggy datasets, according to the experimental results. Our DSNet is much more accurate than the other models while retaining a high speed, according to qualitative and quantitative evaluations of the contrasted approaches.

7. REFERENCES

[1] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[2] R. Girshick, "Fast r-cnn," in *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards realtime object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[4] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," in *Advances in neural information processing systems*, 2016, pp. 379–387.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779– 788.

[6] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," *arXiv preprint*, 2017.

[7] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[8] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[9] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal loss´ for dense object detection,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[10] T.-H. Le, D.-W. Jaw, I.-C. Lin, H.-B. Liu, and S.-C. Huang, “An efficient hand detection method based on convolutional neural network,” in *2018 7th International Symposium on Next Generation Electronics (ISNE)*. IEEE, 2018, pp. 1–2.

[11] T.-H. Le, S.-C. Huang, and D.-W. Jaw, “Cross-resolution feature fusion for fast hand detection in intelligent homecare systems,” *IEEE Sensors Journal*, vol. 19, no. 12, pp. 4696–4704, 2019.

[12] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, “End-to-end united video dehazing and detection,” in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[13] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, “Benchmarking single-image dehazing and beyond,” *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2018.

[14] Wencheng Zhu, Jiwen Lu,” DSNet: A Flexible Detect-to-Summarize Network for Video Summarization”, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 30, 2021.

[15] Trung-Hieu Le," JOINT OPTIMIZATION FOR OBJECT DETECTION IN FOGGY WEATHER CONDITION, Vol. 27 (2020): UTEHY Journal of Science and Technology.

[16] Shih-Chia Huang, “DSNet: Joint Semantic Learning for Object Detection in Inclement Weather Conditions,IEEE,2019.

[17] Lotfi Abdi, “Semantic Recognition: Unified Framework for Joint Object Detection and Semantic Segmentation”,researchgate, 2017.

[18] Baohua Qiang,” Convolutional Neural Networks-Based Object Detection Algorithm by Jointing Semantic Segmentation for Images”, 7 September 2020,MDPI.

[19] Girshick, R. Fast R-CNN. In Proceedings of the 2015 Ieee International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 1440–1448.

[20] Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 1137–1149. [CrossRef] [PubMed]

[21] He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. IEEE Trans. Pattern Anal. Mach. Intell. 2020, 42, 386–397. [CrossRef] [PubMed]

[22] He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In Proceeding of the 13th European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; pp. 346–361.

[23] Kong, T.; Yao, A.; Chen, Y.; Sun, F. HyperNet: Towards accurate region proposal generation and joint object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016.

[24] Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving into high quality object detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.

[25] Zeiler, M.D.; Taylor, G.W.; Fergus, R. Adaptive deconvolutional networks for mid and high level feature learning. In Proceeding of the 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 6–13 November 2011; pp. 2018–2025.

[26] Shi, G.; Suo, J.; Liu, C.; Wan, K.; Lv, X. Moving target detection algorithm in image sequences based on edge detection and frame difference. In Proceedings of the 2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference, Chongqing, China, 3–5 October 2017; pp. 740–744.

[27] C. Sakaridis, D. Dai, and L. Van Gool, “Semantic foggy scene understanding with synthetic data,” *International Journal of Computer Vision*, pp. 1–20, 2018.

[28] S. Li, I. B. Araujo, W. Ren, Z. Wang, E. K. Tokuda, R. H. Junior, R. Cesar-Junior, J. Zhang, X. Guo, and X. Cao, “Single image deraining: A comprehensive benchmark analysis,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3838–3847.

