



CISTER

Research Centre in
Real-Time & Embedded
Computing Systems

Technical Report

DriftNet: Aggressive Driving Behavior Classification using 3D EfficientNet Architecture

Alam Noor

Bilel Benjdira

Adel Ammar

Anis Koubaa

CISTER-TR-200710

DriftNet: Aggressive Driving Behavior Classification using 3D EfficientNet Architecture

Alam Noor, Bilel Benjdira, Adel Ammar, Anis Koubaa

CISTER Research Centre

Polytechnic Institute of Porto (ISEP P.Porto)

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail:

<https://www.cister-labs.pt>

Abstract

Aggressive driving (i.e., car drifting) is a dangerous behavior that puts human safety and life into a significant risk. This behavior is considered as an anomaly concerning the regular traffic in public transportation roads. Recent techniques in deep learning proposed new approaches for anomaly detection in different contexts such as pedestrian monitoring, street fighting, and threat detection. In this paper, we propose a new anomaly detection framework applied to the detection of aggressive driving behavior. Our contribution consists in the development of a 3D neural network architecture, based on the state-of-the-art EfficientNet 2D image classifier, for the aggressive driving detection in videos. We propose an EfficientNet3D CNN feature extractor for video analysis, and we compare it with existing feature extractors. We also created a dataset of car drifting in Saudi Arabian context this [https URL](https://www.cister-labs.pt). To the best of our knowledge, this is the first work that addresses the problem of aggressive driving behavior using deep learning.

DriftNet: Aggressive Driving Behavior Classification using 3D EfficientNet Architecture

Alam Noor¹, Bilel Benjdira¹, Adel Ammar¹, Anis Koubaa^{1,2}

¹Robotics and Internet-of-Things Lab (RIOTU), Prince Sultan University, Riyadh, Saudi Arabia.

²CISTER, INESC-TEC, ISEP, Polytechnic Institute of Porto, Portugal.

nalam@psu.edu.sa, bbenjdira@psu.edu.sa, aammar@psu.edu.sa, akoubaa@psu.edu.sa

Abstract

Aggressive driving (i.e., car drifting) is a dangerous behavior that puts human safety and life into a significant risk. This behavior is considered as an anomaly concerning the regular traffic in public transportation roads. Recent techniques in deep learning proposed new approaches for anomaly detection in different contexts such as pedestrian monitoring, street fighting, and threat detection. In this paper, we propose a new anomaly detection framework applied to the detection of aggressive driving behavior. Our contribution consists in the development of a 3D neural network architecture, based on the state-of-the-art EfficientNet 2D image classifier, for the aggressive driving detection in videos. We propose an EfficientNet3D CNN feature extractor for video analysis, and we compare it with existing feature extractors. We also created a dataset of car drifting in Saudi Arabian context¹. To the best of our knowledge, this is the first work that addresses the problem of aggressive driving behavior using deep learning.

1. INTRODUCTION

The problem of aggressive driving (also known as car drifting) represents a significant safety problem in several cities. It is often related to youngsters driving cars at high speed and drifting between other vehicles in the road, which put the safety and lives of other drivers into a significant threat. In this paper, we propose to use artificial intelligence for automated anomaly detection of abnormal driving behavior, specifically drifting.

Related works A variety of attempts were made in previous works to identify suspicious cars in videos automatically. Approaches based on video anomaly have mainly two feature extraction types (manual and deep learning). Vari-

ous techniques have historically been carried out in statistical methods for the identification of anomalies. Irregular events, motion patterns, space, time, and texture of optical flow were detected using Hidden Markov Models [1], Gaussian Mixture Models, and Markov Random Fields [2, 3]. Various classification algorithms, such as Support Vector Machine[4], Nearest neighbor Clustering [5], and K-means clustering [6] are used as a binary classification for anomaly detection to extract features and optical flow of regular and abnormal patterns of vehicles.

On the other hand, the approaches to deep learning differ from conventional methods. In the areas of object detection, the rapid implementation and growth of deep learning are primarily performed using a convolution neural network. CNN accomplished its achievement in detection by applying the divisional rule in one-stage to different small anchor boxes (grid) of various sizes [7] and by using region proposal networks for the two-stage approaches [8], then they apply classification to labels and regression to bounding boxes. Different approaches of deep learning have been adopted at the bases of one-stage and two-stage methods to solve the anomaly detection problems. Wang et al. proposed the novel approach of Gaussian mixture models with the region of interest of the YOLOv3 to identify anomaly candidate and remove noises of the background. In last TrackletNet Tracker has been employed to obtain anomaly trajectory [9]. Ammar et al. proposed the convolutional neural network based car detection method using Aerial Images with comparison of F-RCNN and YOLOv3 [10]. Shuai et al. presented the perspective method with the backend of ResNet50 and unsupervised traffic flow analysis to eliminate the moving car and detect the state objects on the road and the space-time matrix used to achieve anomaly car locations. [11]. Belil et al., presented the Generative Adversarial Networks (GAN) architecture to reduce the domain shift with increasing the ability to new targeted domains [12], and using Faster R-CNN and YOLOv3 to detect the cars from drones with high precision [13]. Jianfei et al. proposed an unsupervised framework with multi-

¹<https://www.youtube.com/watch?v=vLzgye1-d1k>

object tracking to combine single-object tracking to solve the tracks and detection problems with efficiency and accuracy [14]. Ding et al. employed 3D ConvNet’s end-to-end architecture to detect violence in videos [15]. Ji et al. presented 3D CNN using bottleneck technique with adopting the DenseNet method of several dense layers and in the last global average pooling layer for feature mapping [16]. Google researchers Mingxing et al. proposed a new and scalable architecture EfficientDet to use a weighted bi-directional feature pyramid network with a combination of EfficientNet as a backbone following the one-stage detector design. Additionally, adopting optimize multi-scale feature fusion for interpretation of resulting feature network [17].

In this paper, we leverage the use of the EffientNet scaling approach, and we adopt it as a backbone multimodal architectural framework for car drifting detection.

The rest of the paper is organized as follows. In Section 2, we present the DrifNet architecture, where we presented the dataset collection process and the proposed 3D EfficientNet neural network model.

2. DriftNet Architecture

In this section, we provide the details of the pipeline of the proposed method for driving anomaly detection. Anomalies of traffic in the AI City Challenge issue are primarily vehicles that have different behaviors from other cars like drifting², stopping, crashing on the road, etc. First, we created a dataset (described below) for vehicle drifting. Second, we proposed a 3D EfficientNet method to detect and recognize aggressive car driving efficiently. Using 3D EfficientNet to detect car drifting, it should be able to identify the pattern of moving vehicles in sliding modes and motions. We considered EfficientNet because it is currently the state-of-the-art algorithm for object detection. Efficiency and accuracy are an essential part of the recognition tasks, and the models of EfficientNet achieve higher accuracy and performance over an extensive range of precision and resource constraints than previous detectors.

2.1. Drifting Dataset

The compilation and annotation of large-scale datasets are painstaking, expensive and time consuming. We created a car drifting clips dataset to advance computer vision and video understanding of aggressive driving detection. 100 video samples of drifting in Saudi Arabia roads were collected from YouTube, and added to 100 videos of normal traffic selected from the AI Challenge dataset. We performed weakly labeling of the video frames as drifting and normal driving, and vocabulary annotation has been designed carefully. We extracted a total of 4900 frames for normal driving and 4900 frames for vehicle drifting. We

Table 1. Anomaly car driving detection.

Categories	Training set	Testing set	Total
Number of images	7840	1960	9800
Percentage	80%	20%	100%
Number of drift cars frames	3920	980	4900
Number of non-drift cars frames	3920	980	4900

split the dataset into 80% for training, and 20% for validation. Table.1 shows the number and percentage of frames of the training and testing sets. Figure 1 shows sample frames of normal driving from the AI City Challenge dataset. Each



Figure 1. From top left to bottom right, all the cars shown are in normal driving positions.

frame has a size of 640x360. We pre-processed all frames to make them suitable for input size 112x112 for the EfficientNet feature extractor. This dataset will be made open-source to drive new research on car drifting detection so that it can effectively contribute to AI City challenges to increase human safety and mitigate the risks of aggressive driving behavior. Figure 2 shows sample frames of the aggressive driving car from the anomaly dataset.



Figure 2. From top left to bottom right the black car shows different positions of drifting.

²<https://www.youtube.com/watch?v=vLzgye1-d1k>

2.2. 3D EfficientNet Network

We evaluated EfficientNet using transfer learning for the drifting dataset. We borrow the same training settings from EfficientNet, which takes the ImageNet pre-trained model and finetune it to the new dataset. Figure 3 represents the workflow of the EfficientNet architecture. The model optimization algorithm 'SGD' was used during the training process with the following parameters: the starting learning ratio is 0.0002, the momentum is 0.5, and the decay rates are $1e-7$. After every convolution neural network, Batch Normalisation is applied with a 0.997 decay batch norm. Besides, a batch of samples was 32 for batch testing and model validation. The sample size was 112×112 . Besides, the early stop training technique was used to avoid overfitting of the model during the training cycle.

3. Experimental results

This section details the experimental results of the proposed method deep transfer learning method. The experiment was conducted on Nvidia GeForce RTX 2070 GPUs and used with PyTorch. We have used four different architecture for effectiveness and efficiency of EfficientNet at drift dataset.

The preparation of input shape dimensions are $N \times C \times T \times H \times W$. Where N and C are the batch size and Channels respectively. H and W are the height and width of the frames, while T is the duration of video of the dataset. Frames of each video has been sampled into 32 and resized to an input size of 112×112 . It is necessary to have a large dataset to minimize the over fitting problems during training, and data augmentation is a very effective method to avoid over fitting. Horizontal flipping, lighting, cropping, slicing and salt and pepper noise techniques have been used for data augmentation. We used the Stochastic Gradient Descent (SGD) optimizer for training with momentum 0.5 and learning rate decay $1e-7$. We split the drifting dataset into 80% for training and 20% for testing. Figure 5 and 5 depict the training and validation loss and accuracy of EfficientNet3D-B0. We have tested other scaling factors for EfficientNet (up to B7) but they did not produce better results.

We compared our proposed model to three other existind models, namely, ConvLSTM, DenseNet, and C3D. C3D [18] is a traditional 3D CNN video descriptor, while ConvLSTM [19] and DenseNet are state-of-the-art models for the identification of anomalies. The C3D model's parameters are initialized in the initial convolution layer, while ConvLSTM uses pre-trained AlexNet for extracting features, and DenseNet takes sequentially multiple layers with batch normalization to extract the features of the object. We have taken pre-trained and applied transfer learning because the scale of driving car dataset is still limited to allow for a complete training of all the parameters of these deep neu-

ral networks. The efficiency and parameter number of these models is compared in Table 2. The proposed model shows a higher accuracy in classification with comparatively less parameters. Our model's validation accuracy is 92.5%

4. Conclusion

This paper proposes an efficient network architecture design to improve the internal architecture as a transfer learning to efficiently detect anomaly vehicles. In order to increase accuracy and efficiency of the anomaly car detecting, the proposed model has comparatively lower parameters with a custom compound scaling process. The experimental results of car drift data sets are based on these optimizations and demonstrate the improvements of our model over framework. The proposed approach is to computing resource-saving and to achieve consistently greater precision and efficiency over a wide variety of limitations of resources than the existing state-of-the-art detection methods.

References

- [1] L. Kratz and K. Nishino. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 2009, pp. 1446-1453.
- [2] H. Nallaivarothayan, C. Fookes, S. Denman, and S. Sridharan. An mrf based abnormal event detection approach using motion and appearance features. 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Seoul, 2014, pp. 343-348.
- [3] D. Ryan, S. Denman, C. Fookes, and S. Sridharan. Textures of optical flow for real-time anomaly detection in crowds. 2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Klagenfurt, 2011, pp. 230-235.
- [4] T. Wang and H. Snoussi. Detection of abnormal visual events via global optical flow orientation histogram. in IEEE Transactions on Information Forensics and Security, vol. 9, no. 6, pp. 988-998, June 2014.
- [5] R. V. H. M. Colque, C. A. C. Jnior, and W. R. Schwartz. Histograms of optical flow orientation and magnitude to detect anomalous events in videos. 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images, Salvador, 2015, pp. 126-133.
- [6] Zhouyu Fu, Weiming Hu, and Tieniu Tan. Similarity based vehicle trajectory clustering and anomaly detection,". IEEE International Conference on Image Processing 2005, Genova, 2005, pp. II-602.
- [7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788.
- [8] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic

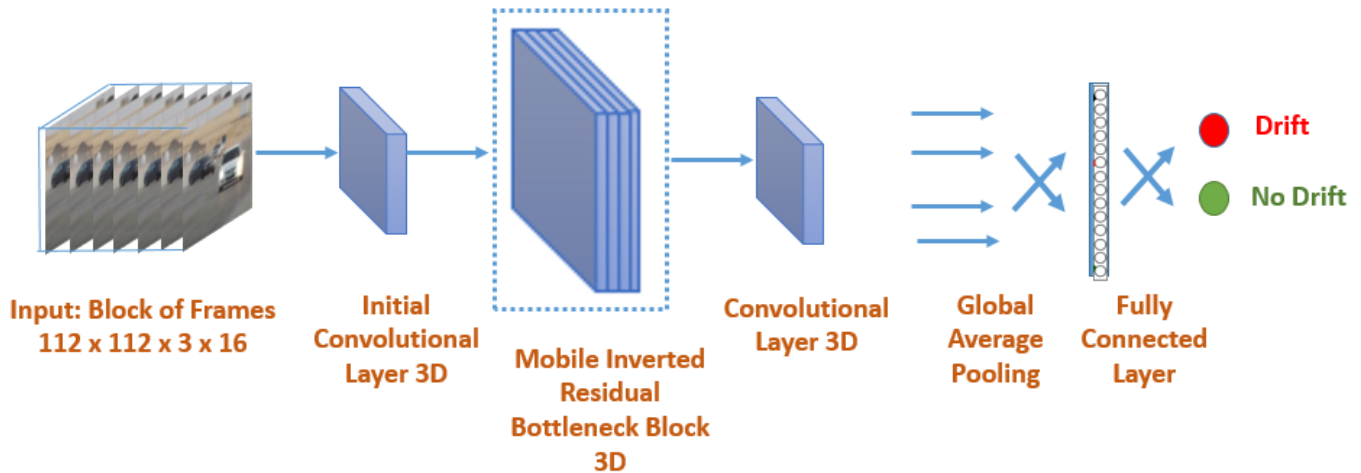


Figure 3. EfficientNet3D-B0 Architecture.

Table 2. Experiment results on car drifting dataset.

Models	Param	Training Accuracy (%)	Validation Accuracy (%)
EfficientNet3D-B0	4.69M	93.75	92.5



Figure 4. Training and validation loss of EfficientNet3D-B0.

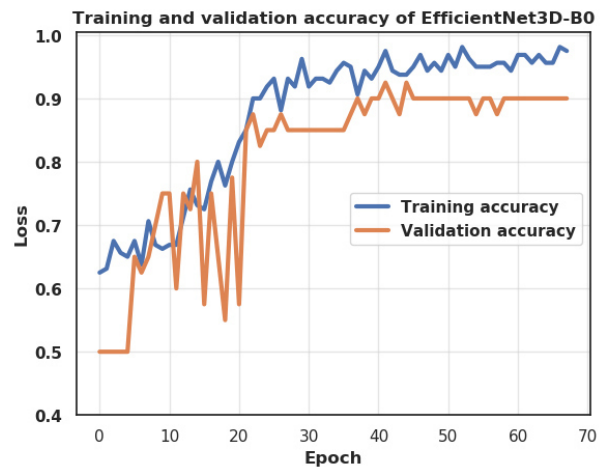


Figure 5. Training and validation accuracy of EfficientNet3D-B0.

segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 580-587.

- [9] Gaoang Wang, Xinyu Yuan, and Jenq-Neng Hwang Aotian Zheng, Hung-Min Hsu. Anomaly candidate identification and starting time estimation of vehicles from traffic videos. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019, pp. 382-390.

- [10] Adel Ammar, Anis Koubaa, Mohammed Ahmed, and Abdulrahman Saad. Aerial images processing for car detection using convolutional neural networks: Comparison between faster r-cnn and yolov3. arXiv:1910.07234v1.



Figure 6. Output of the proposed architecture.

- [11] Yu Lei-Wei Wu Chengkai Zhu Ming Sun Junjie Yan Shuai Bai, Zhiqun He. Traffic anomaly detection via perspective map based on spatial-temporal information ma-

- trix. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019, pp. 117-124.
- [12] Bilel Benjdira, Adel Ammar, Anis Koubaa, and Kais Ouni. Data-efficient domain adaptation for semantic segmentation of aerial imagery using generative adversarial networks. *Appl. Sci.* 2020, 10, 1092.
- [13] B. Benjdira, T. Khursheed, A. Koubaa, A. Ammar, and K. Ouni. Car detection using unmanned aerial vehicles: Comparison between faster r-cnn and yolov3. 2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS), Muscat, Oman, 2019, pp. 1-6.
- [14] Siyang Pan-Yanyun Zhao Zhicheng Zhao Fei Su Bojin Zhuang Jianfei Zhao, Zitong Yi. Unsupervised traffic anomaly detection using trajectories. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019, pp. 133-140.
- [15] Zhu M.-Feng W. Jia B Ding C., Fan S. Violence detection in video by using 3d convolutional neural networks. In: *Bebis G. et al. (eds) Advances in Visual Computing. ISVC 2014. Lecture Notes in Computer Science*, vol 8888. Springer, Cham.
- [16] T. Sun J. Li, X. Jiang and K. Xu. Efficient violence detection using 3d convolutional neural networks. 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Taipei, Taiwan, 2019, pp. 1-8.
- [17] Quoc V. Le Mingxing Tan, Ruoming Pang. Efficientdet: Scalable and efficient object detection. <https://arxiv.org/abs/1911.09070>, 2020.
- [18] R. Fergus L. Torresani D. Tran, L. Bourdev and M. Paluri. Learning spatiotemporal features with 3d convolutional networks. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 4489-4497.
- [19] S. Sudhakaran and O. Lanz. Learning to detect violent videos using convolutional long short-term memory. 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, 2017, pp. 1-6.