ISSN: 0976-0172

Journal of Bioscience And Technology
www.jbstonline.com

Fusion of Deep Learning Models for Video-Based Detection of Risky Driving Behavior

***Dr. Kanaka Durga Returi*[1]*., Ch.Jaya Sindhu*[2]**

1 Professor, Department of CSE, Malla Reddy College of Engineering for Women.,
Maisammaguda., Medchal., TS, India
2, B.Tech CSE (20RG1A05K5),
Malla Reddy College of Engineering for Women., Maisammaguda., Medchal., TS, India

Article Info

Received: 29-06-2023 *Revised: 17-07-2023* *Accepted: 28-07-2023*

ABSTRACT

For the time being, video-based anomalous driving behavior detection is growing in popularity due to its importance in guaranteeing the safety of drivers and passengers in the vehicle and as a necessary step toward attaining automated driving. Modern advancements in deep learning techniques have made it possible to greatly simplify this difficult detection task by providing highly sophisticated models with a large amount of training data in the form of video clips, making it possible to train these models to a high level of accuracy with minimal human intervention. To _rst-time full the entire video-based abnormal driving behavior detection task, this paper places an emphasis on deep learning fusion techniques and introduces three novel deep learning-based fusion models influenced by the recently proposed and popular densely connected convolutional network (Dense Net). The wide group densely (WGD) network, the wide group residual densely (WGRD) network, and the alternative wide group residual densely (AWGRD) network are the three new deep learning-based fusion models.

INTRODUCTION

It is widely acknowledged that, high-resolution videos are more and more commonly seen within a great number of visual applications at the current stage. For instance, in video surveillance, multiple high-resolution cameras are necessary to be placed at different locations. They work together to identify [1], [2], re-Identity [3], [4], and track the moving target [5], [6], making the later high-level analyses based on the moving target (e.g., behavior or even potential intention) more feasible. In emotional computation, high-resolution cameras need to be utilized to capture both obvious and _ne changes of emotions of the target person in real-time [7], [8], which have significant impacts in security issues nowadays. It is easy to perceive from the above descriptions that, acquiring and storing a large volume of high-resolution videos are often not difficult to be realized for the time being. However, the main challenge resides in how to efficiently and effectively make correct high-level decisions based on those low-level video clips of large volumes. In this study, high-resolution videos of drivers recorded within vehicles are emphasized. The high-level decision here is to correctly detect abnormal driving behavior (i.e., patterns) of drivers. Automatic abnormal driving behavior detection is generally accepted as the _rest issue in realizing the popular fully autonomous driving task. It is certain that, for the autonomous driving task, safety issues are undoubtedly _rest priorities. It is widely known that, behavior of drivers need to be well restricted in order to avoid any potential accident. Therefore, multiple high-resolution cameras equipped within the driver's vehicle can be utilized to monitor the driver's status in real time. Generally speaking, videos captured by high-resolution cameras also need to be processed immediately, in order to determine whether the current status of the driver is normal or not. It can be acknowledged from the above descriptions that, both the effectiveness (i.e., the detection accuracy) and the efficiency (i.e., the detection speed) of abnormal driving behavior detection are highly demanded. Also, high-speed wireless transmissions are necessary to realize the swift and reliable transmission of high-quality videos, which further facilitates the above automatic abnormal driving behavior detection task [9]_ [23]. In order to detect abnormal behavior of drivers, an official and precise dentition of abnormal driving is often necessary. According to the International Organization for Standardization (ISO), abnormal driving is denned as the phenomenon that a driver's ability to drive is impaired due to her / his own focus on other activities unrelated to normal driving.

Generally speaking, abnormal driving behavior can be mainly divided into three categories. The _rest one belongs to distracting driving behavior that meets the driver's physical comfort requirements, including smoking, drinking, eating, con_guring the aircon, etc. The second one is to meet the driver's need for distracting driving behavior, including makeup, shaving, chatting, using mobile phones or other unnecessary devices, etc. The third one contains distracted driving behavior caused by the surrounding environment, including caring for children, long-term attentions to unexpected events outside the vehicle, etc. Among the above-mentioned abnormal driving behavior, it is necessary to highlight that, the use of mobile phones has already become a major factor in contemporary abnormal driving. In a recent simulation, researchers have found that making a call while driving can cause the driver to distract 20% of her / his attention. More seriously, if the content of the call is important, it will lead to a distraction up to 37%, which will make the driver 23 times more likely to have an accident than normal drivers [24]. Therefore, the use of mobile phones is also considered as one important abnormal driving behavior for automatic detection in this study. In this study, a single visible-light camera is utilized to record high-resolution videos of the driver, and three novel deep learning-based fusion models are proposed to full the video-based abnormal driving behavior detection task. The basic architecture of deep learning models introduced in this study is mainly motivated by densely connected convolutional networks (Dense Net), which were proposed in 2017 and won the best paper of CVPR the same year [25]. Generally speaking, Dense Net can be regarded as a relatively new convolutional neural network (CNN)-based deep learning architecture, and it has signi_cant merits of reaching the state-of-the-art performance in several well-known classi_-cation challenges (e.g., CIFAR, SVHN, Image Net databases) using less parameters. Also, it is not difficult to be trained even within a tremendously deep model's structure because

Of its intensive utilization of the residual network [26]. Additionally, deep learning fusion techniques are utilized in this study based on the original DenseNet, in order to obtain three novel fusion models for realizing the video-based abnormal driving behavior detection task for the _rst time. The three new fusion models proposed in this study are named as the wide group densely (WGD) network, the wide group residual densely (WGRD) network, and the alternative wide group residual densely (AWGRD) network, respectively. Technically, WGD takes important issues of deep learning models, i.e., the depth, the width and the cardinality, into consideration when designing its model structure based on DenseNet. For WGRD and AWGRD, they are more sophisticated as the important idea of residual networks with superpositions of previous layers is incorporated.

RELATED WORKS

In the following, abnormal driving detection and deep learning techniques, which are closely related to this study, are emphasized. Recent developments in the two aspects are briery reviewed, with pros and cons been discussed.

A. ABNORMAL DRIVING DETECTION

It can be summarized based on literatures of automatic abnormal driving behavior detection that, there are often three commonly used detection schemes. The _rest one is based on the detection of human physiological signals (i.e., electrooculogram, electro-encephalogram, respiratory, blood _own changes, etc.) using diverse kinds of sensors [27], [28]. The second one is based on facial details [29] (i.e., changes in eye movement, mouth movement, head movement, hand features, etc.). The third one is based on motion characteristics of the steering wheel, which is capable to detect the driver's hand pressure [30], the steering time, the brake behavior [31], etc. It is also necessary to point out that, detecting human physiological signals has good real-time performance and high precision, but its main advantage of affecting drivers' normal driving cannot be neglected, either. Furthermore, physiological signals of human beings vary greatly due to the physiological difference in each individual person and her / his environmental conditions. Therefore, it is challenging to provide quantitative and objective standards for detecting human beings' physiological signals as well. For detections based on facial details, eye regions are often emphasized as the gaze direction of eyes are closely related to normal / abnormal driving patterns. Among eyes-based detection methods, the percentage of eyelid closure over the pupil over time (PERCLOS) [32] is popular. Technically,

The percentage of closed eye time per unit time is utilized and it can be explicitly represented in Equation (1).

$$PERCLOS = \frac{Eye\ closing\ time}{Detection\ period} \times 100\% \qquad (1)$$

When the percentage of time that the eyes' closure reaches 70% or even higher, the driver is normally considered to be in an abnormal driving state [32]. Although the PERCLOS there are unfortunately several serious problems with it. First, eyes of drivers with different physiques and habits vary largely. An extreme case is that some people even do not close their eyes when sleeping, making the false positive of PERCLOS inevitably high. Second, some tough challenges, including unexpected movements of the head, will lead to detection failures of eyes. The above situations are certainly not bene_cial for falling abnormal driving behavior detection based on eyes [33]. For detections based on steering wheel, they are similar towards detections based on human physiological signals

B. RECENT DEVELOPMENTS IN DEEP LEARNING AND ITS POPULAR UTILIZATIONS

It is interesting to notice that, deep learning techniques receive vast popularity when powerful computational hardware and large-scale data become more and more available nowadays. Generally speaking, most contemporary

Deep learning models can be categorized into two types, i.e., deep generative learning models and deep discriminant

Learning models. To be septic, deep generative learning models mainly aim to replicate ``fake-but-realistic" data based on real data, and popular deep generative learning models include but not limited to VAE (i.e., variation auto-encoder) [34], GAN (i.e., generative adversarial network) [35], GLOW (i.e., generative _own) [36], etc. Deep discriminative learning models, on the other hand, are mainly utilized for discrimination / classi_cation purposes.

Well-known deep discriminative learning models are often winners of noticeable worldwide vision-based competitions (e.g., ILSVRC, COCO, etc.). Typical deep discriminative learning models include but not limit to Alex Net [37], VGG [38], Google Net [39], Resent, etc. Recently, contemporary deep learning models demonstrate the following trends. First, more and more deep learning models become tremendously deep for guaranteeing outstanding generalization capabilities. Second, their model structures become more and more sophisticated. For instance, the width of many contemporary deep learning models increases is ni_cantly. In [40], it is reported that a wide 40-layer Resent model can gain the similar generalization capability as a conventional ``narrow" single-channel 1001-layer Resent model, but the wide model only costs 1/8 training time of the narrow one. Also, the cardinality of deep learning models increases greatly as well. Cardinality is often regarded as the number of isolated paths in a deep learning model, and those paths often share the same topological structure [41], making many contemporary deep learning models become multichannel- based in their architectures.

METHODOLOGY

In this section, technical details of the three new deep learning fusion models for automatically detecting abnormal driving behavior are elaborated. Since the three new fusion models are inspired by Dense Net, the very model as well

As other conventional deep learning models will be introduced in Section III-A _rest. It is worthy to mention that, all these conventional and popular deep learning models will be implemented and compared with three new deep learning-based fusion models in this study. After introducing conventional deep learning models in Section III-A, the three new deep learning-based fusion models will be elaborated in Section III-B. The energy function to be optimized in deep learning-based fusion models will be introduced in Section III-C.

A. CONVENTIONAL DEEP LEARNING MODELS

In the following, _ve conventional and popular deep learning models, including convolutional neural network (CNN), wide convolutional neural network (Wide CNN), group convolutional neural network (Group CNN), deep residual network (ResNet), and densely connected convolutional network (Dense

Net), are introduced one by one. Details of their model structures utilized in this study are emphasized.

1) CONVOLUTIONAL NEURAL NETWORK (CNN)

One of the earliest CNN models, i.e., the LeNet-5, was originally proposed for falling the recognition and classi_cation of handwritten characters, and its accuracy is satisfactory [52]. Generally speaking, the main architecture of CNN consists of the convolutional layer, the pooling layer, and the full connected layer. To be septic, the convolutional layer and the pooling layer work together to form multiple convolution groups, in order to extract latent features through a layer-by-layer model architecture. Then, the classi_cation task can be completed based on latent features via fully connected layers. In this study, the model structure of CNN is depicted in Figure 1.

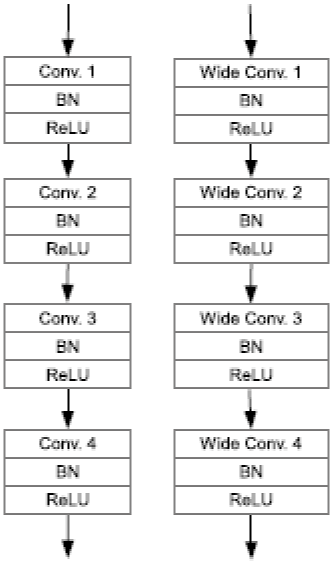


FIGURE 1. An illustration of model architectures in CNN (left) and Wide CNN (right) in this study.

2) WIDE CONVOLUTIONAL NEURAL NETWORK (WIDE CNN)

The idea of Wide CNN actually comes from the wide residual network (WRN) [40], which is on the basis of the deep residual network but further increases the number of layer-based convolution kernels. Figure 1 demonstrates the difference between the conventional CNN model and the Wide CNN model utilized in this study. It can be observed that, one signi_cant difference between CNN and Wide CNN is that, wide convolution layers instead of the traditional ``narrow'' convolution layers are incorporated in Wide CNN. The motivation can be explained as follows. It is widely acknowledged that, it is challenging for gradients to be back-propagated when a deep learning model becomes tremendously deep, and such a tremendously deep model is often hard to be comprehensively
Trained.

3) GROUP CONVOLUTIONAL NEURAL NETWORK (GROUP CNN)

Group CNN mainly counts on group convolutions, which are quite different from traditional convolutions adopted in the vast majority of CNN-based deep learning models. To be septic, each individual convolution _later in CNN operates on all channels, while each individual convolution _later in Group CNN is active only on partial channels. An illustration of the difference between traditional convolution _liters in CNN and group convolution _liters in Group CNN is depicted in Figure 2. It can be noticed that, Figure 2 describes the case of 2-channel. For traditional convolution _liters in CNN (i.e., left in Figure 2), C traditional convolution _liters are executed on N feature maps, in order to obtain C feature maps. For group convolution _liters in Group CNN (i.e., right in Figure 2), N feature maps are divided into two parts (i.e., each part contains *N* 2 feature maps for balanced considerations). Each individual part is then fed into *C* 2 convolution _liters, in order to generate *C*2 feature maps. In this way, different convolution _liters are actually executed on different channels. The idea of Group CNN is important, as different feature maps can be generated using different GPUs and the _nil result can be fused based on them, making the model more efficient to be trained with multiple GPUs. Additionally, the recently popular Resent model also adopts the

Above group convolution idea in its residual network-based architecture [41].

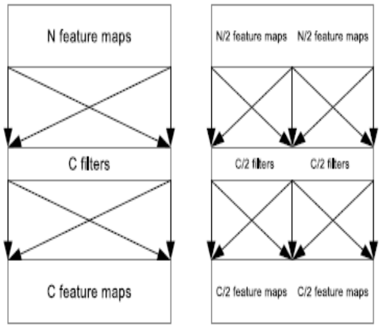


FIGURE 2. An illustration of the difference between traditional convolution filters in CNN (left) and group convolution filters in Group CNN (right).

B. THREE NOVEL DEEP LEARNING-BASED FUSION MODELS: WGD, WGRD, AND AWGRD

In the following, three novel deep learning-based fusion models inspired by DenseNet are introduced to tackle the video-based abnormal driving behavior detection problem for the _rst time. The three new models are named as the wide group dense (WGD) network, the wide group residual dense (WGRD) network, and the alternative wide group residual dense (AWGRD) network, respectively. Technically, WGDtakes important issues of deep learning models, i.e., the depth, the width and the cardinality, into consideration when designing its model structure based on Dense Net. Forward and AWGRD, they are more sophisticated as the important idea of residual networks with superposition's of previous layers is incorporated. Technical details are as follows.

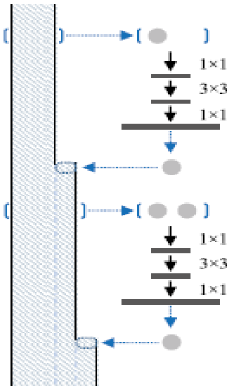


FIGURE 4. An illustration of the tree-like connection architecture in Dense Net.

1) WIDE GROUP DENSELY NETWORK (WGD)

The model architecture of WGD is demonstrated in Figure 5. It can be noticed that, the conventional convolution in Dense Net is replace by the group and wide convolution in WGD. The merit is that, the generalization capability of WGD can be improved via group and wide convolutions in WGD, while the number of parameters in WGD will not increase much. Also, the important enhancement of width and cardinality in WGD can be realized, therein.

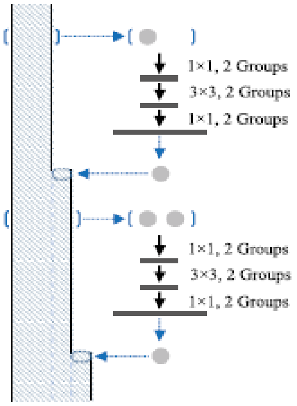


FIGURE 5. The model architecture of WGD.

2) WIDE GROUP RESIDUAL DENSELY NETWORK (WGRD)

The model architecture ofWGRDis depicted in Figure 6. The most signi_cant change of WGRD with respect to WGD is that, the idea of residual networks is incorporated in WGRD. Details of WGRD and its affinity with other deep learning models can be explained as follows. Provided an input image *x*0 transmitted through an *L*-layer network, and the *l*-layer can be represented via a non-linear transformation *Hl* (_) (i.e., composed of BN, Rely, Conk, etc.). Let the output of the *l*-layer be *XL*. Then, *XL* in a conventional feed forward network

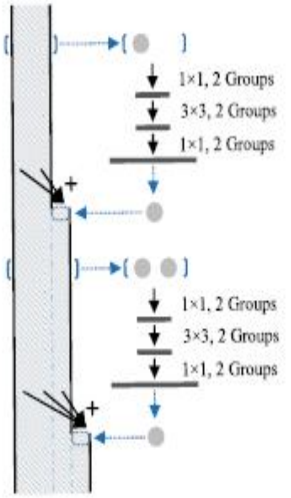


FIGURE 6. The model architecture of WGRD. Can be represented in Equation 2.

$$x_l = H_l(x_{l-1}) \qquad (2)$$

$$x_l = H_l(x_{l-1}) + x_{l-1} \qquad (3)$$

For Dense Net and WGD, the situation becomes more sophisticated. Since Resent only takes the input and the output of the *l*-layer into consideration, intermediate outputs from other previous layers (i.e., *x*1 from the 1*st*-layer, *x*2 from the 2*nd*-layer, *xl*□2 from the (*l* □ 2)-layer) will be totally neglected. In order to tackle the above problem, DenseNet and WGD take all above-mentioned information into consideration when representing *XL,* which can be represented in Equation 4 (i.e., the operator [_] represents the parallel operation).

$$x_l = H_l[x_0, x_1, \cdots, x_{l-1}] \qquad (4)$$

$$x_l = H_l[x_0, x_0 + x_1, \cdots, \sum_{i=0}^{l-1} x_i] \qquad (5)$$

For WGRD, a more sophisticated constitution of *XL* is further applied. As described in Equation 5, the superposition of previous layers (i.e. *xi*) is utilized on the *I*-layer. The reason is because that, more complex features are added as the input of the *l*-layer and the learning capability of the network can be strengthening, therein. In this way, the important idea of residual networks with superpositions of previous layers can be realized in WGRD.

3) ALTERNATIVE WIDE GROUP RESIDUAL DENSELY NETWORK (AWGRD)

In this study, an alternative WGRD (i.e., AWGRD) is also introduced to fully the video-based abnormal driving behavior detection task. The model architecture of AWGRD is illustrated in Figure 7, and its main idea is described in

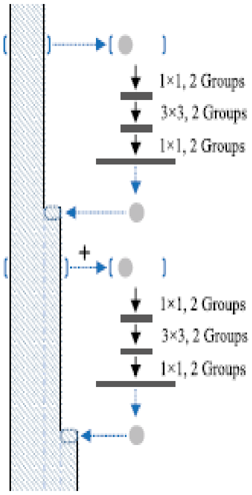


FIGURE 7. The model architecture of AWGRD.

$$x_l = H_l[\sum_{i=0}^{l-1} x_i] \qquad (6)$$

C. THE ENERGY FUNCTION TO BE OPTIMIZED IN DEEP LEARNING-BASED FUSION MODELS

The video-based abnormal driving behavior detection task in this study can be regarded as a general multi-class classi_cation problem, in which the classic cross entropy is utilized to constitute the energy function to be optimized.

Provided the *i*-th image (i.e., frame) of a video clip as *xi*, and its label information as *yi* (i.e., *yi* is represented via a *c*-dimensional feature vector in this study, while *c* indicates the number of classes). Let *y*0 *i* D *P*(*xi*) denote the probability that a deep learning-based fusion model assigns *xi* to one particular class (i.e., *P* (_) indicates the whole mapping of the deep learning-based fusion model), and the energy function based on cross entropy in the three deep learning-based fusion models can be represented in Equation 7.

$$L = -\sum_{i=1}^{m}\sum_{j=1}^{c} y_{ij} log(y'_{ij}) = -\sum_{i=1}^{m}\sum_{j=1}^{c} y_{ij} logP(x_{ij}) \qquad (7)$$

EXPERIMENTS

A. DATABASE AND EXPERIMENTAL SETTINGS

To verify the effectiveness of newly proposed deep learning-based fusion models in automatically detecting abnormal driving behavior of this study, the Cagle state farm distracted driver detection database was utilized [54]. To be septic, there are totally 22,424 color frames (i.e., images) of drivers in video clips of this database. Each individual image has a _axed spatial resolution of 640_480, and all images can be categorized into 10 classes, which indicate 10 different driving patterns. These driving patterns contain safe driving, texting (using right hand), talking on the phone (using right hand), texting (using left hand), talking on the phone (using left hand), operating the radio, drinking, reaching behind, hair and makeup, talking to passenger, etc. Example images of them are displayed in Figure 8.

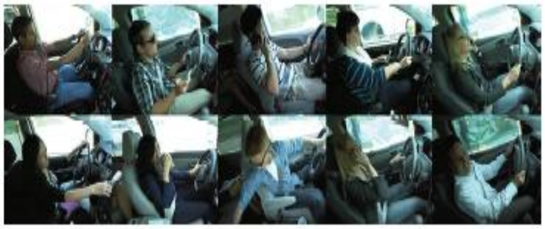
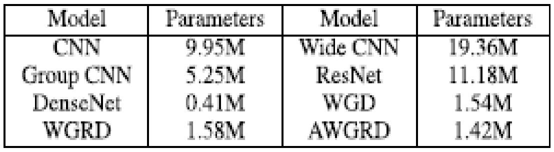


FIGURE 8. Example images of 10 driving patterns in the utilized Cagle state farm distracted driver detection database (from left to right, up to bottom: Safe driving, texting (right hand), talking on the phone (right hand), texting (left hand), talking on the phone (left hand), operating the radio, drinking, reaching behind, hair and makeup, talking to passenger).

TABLE 1. Numbers of parameters to be learned in all deep learning models of this study.



Model	Parameters	Model	Parameters
CNN	9.95M	Wide CNN	19.36M
Group CNN	5.25M	ResNet	11.18M
DenseNet	0.41M	WGD	1.54M
WGRD	1.58M	AWGRD	1.42M

B. EXPERIMENTAL RESULTS AND STATISTICAL ANALYSES

Figure 9 demonstrates the trend of accuracies increasing with respect of training epochs in all compared deep learning models. First, it can be noticed that, accuracies of all deep learning models keep on increasing and then become stable when their training epochs further increase, which is a significant indicator of the thorough training and convergence of all deep learning models. Second, three deep learning-based fusion models, Dense Net, as well as Resent outperform other conventional CNN-based models (i.e., CNN, Wide CNN, and Group CNN) as revealed in Figure 9. For comparisons between three deep learning-based fusion models and Dense Net, it is interesting to notice that, the former reaches the stable stage faster (i.e., fewer epochs) than Dense Net, and significant robustness can be obtained from new deep learning-based fusion models.

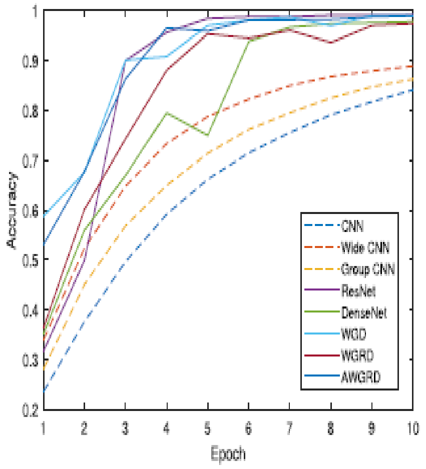


FIGURE 9. The trend of accuracies increasing with respect of training epochs in all deep learning models. It can be observed from Figure 10 that, AWGRD achieves the highest MAP among all compared models. For conventional CNN-based models (i.e., CNN, Wide CNN, Group CNN), their P-R curves are significantly lower than those of others, which indicates that more sophisticated model architectures (e.g., Resent-based, Dense Net-based, etc.) are beneficial for correctly detecting abnormal driving behavior in this study. Another interesting observation from Figure 10 is that, Dense Net and its derivatives (i.e., three novel deep learning-based fusion models) outperform Resent regarding their P-R curves, which suggests that the superposition of previous layers in Dense Net is superior to incorporating only one previous layer in Resent for automatically detecting abnormal driving behavior in this study.

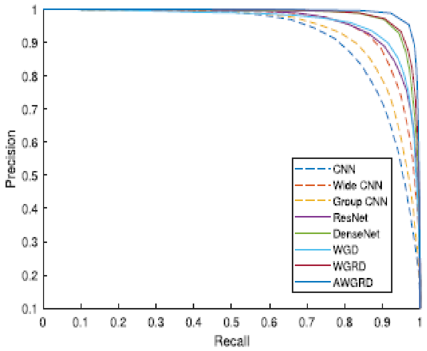


FIGURE 10. Precision-recall curves of all deep learning models in this study. Precision and recall outcomes are then utilized to further calculate the unbiased F-measure (i.e., *F* □ *measure* D 2_*precision_recall precision recall*), and the box-and-whisker plot of F-measures calculated from precision and recall outcomes of all deep learning models are illustrated in Figure 11. In each individual box of Figure 11, the red horizontal line in each box

represents the median of F-measure, while the upper and lower quartiles of F-measure is represented by blue lines

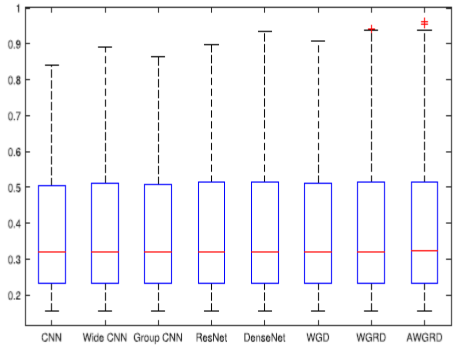
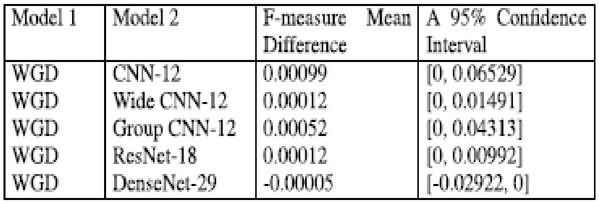


FIGURE 11. The box-and-whisker plot of F-measures calculated from precision and recall outcomes of all deep learning models.

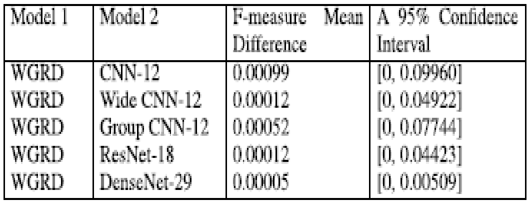
Above and below the median in each box. A vertical dashed line is drawn from the upper quartile and the lower quartile to their most extreme data points, which are within the 1.5 inter-quartile ranges (IQR). Each individual data beyond the 1.5 IQR is marked via a plus sign. Furthermore, a more detailed quantitative analysis made up of one-way analysis of variance (ANOVA) and multiple comparison tests are conducted based on unbiased F-measure outcomes. In one-way ANOVA, F-measure results obtained from all deep learning models are compared to test a hypothesis (*H*0) that ``F-measure means of all deep learning models are equivalent", against the general alternative that these means cannot be all the same. The p-value is utilized here as an indicator to reveal whether *H*0 holds or not. In this study, p-values calculated from all F-measure results are nearly 0, which is a strong indication that all these models cannot share the same F-measure mean. Therefore, the next step is to conduct more detailed paired comparisons. The reason to do so is because that, the alternative against *H*0 is too general. Information about which deep learning model is superior from the statistical perspective cannot be perceived by one-way ANOVA alone. There are two kinds of evaluation after applying multiple comparison tests on calculated F-measure of all models, and quantitative evaluation results are shown in Tables 2, 3, and 4. For the two kinds of evaluation, one is estimated F-measure mean difference, which is a single-value estimator of F-measure mean difference. Another is a 95 % condense interval (CI). In statistics, a CI is a special form of interval estimator for a parameter (i.e. F-measure mean difference in this experiment). Generally speaking, instead of estimating the parameter by a single value, CI is capable to provide an interval estimation which is likely to include the estimated parameter within a specie interval.

TABLE 2. Multiple comparison tests between WGD and other conventional models based on F-measure in this study.



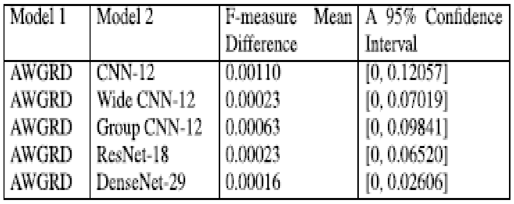
Model 1	Model 2	F-measure Mean Difference	A 95% Confidence Interval
WGD	CNN-12	0.00099	[0, 0.06529]
WGD	Wide CNN-12	0.00012	[0, 0.01491]
WGD	Group CNN-12	0.00052	[0, 0.04313]
WGD	ResNet-18	0.00012	[0, 0.00992]
WGD	DenseNet-29	-0.00005	[-0.02922, 0]

TABLE 3. Multiple comparison tests between WGRD and other conventional models based on F-measure in this study



Model 1	Model 2	F-measure Mean Difference	A 95% Confidence Interval
WGRD	CNN-12	0.00099	[0, 0.09960]
WGRD	Wide CNN-12	0.00012	[0, 0.04922]
WGRD	Group CNN-12	0.00052	[0, 0.07744]
WGRD	ResNet-18	0.00012	[0, 0.04423]
WGRD	DenseNet-29	0.00005	[0, 0.00509]

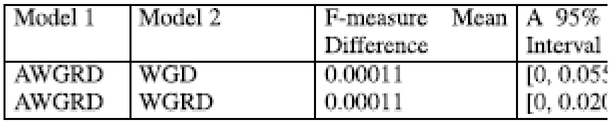
TABLE 4. Multiple comparison tests between AWGRD and other conventional models based on F-measure in this study.



Model 1	Model 2	F-measure Mean Difference	A 95% Confidence Interval
AWGRD	CNN-12	0.00110	[0, 0.12057]
AWGRD	Wide CNN-12	0.00023	[0, 0.07019]
AWGRD	Group CNN-12	0.00063	[0, 0.09841]
AWGRD	ResNet-18	0.00023	[0, 0.06520]
AWGRD	DenseNet-29	0.00016	[0, 0.02606]

as Method 1 in above-mentioned three tables (i.e., WGD in Table 2, WGRD in Table 3, and AWGRD in Table 4), those positive entries in tables substantiate the superiority of Method 1 to other compared conventional deep learning models. It is also necessary to point out that, for the comparison between WGD and DenseNet-29 in Table 2, the single-value estimation is negative and its corresponding 95% confidence interval is also negative. It suggests that DenseNet-29 outperforms WGD based on F-measure in this study, which complies well with the fact that the P-R curve of Dense Net is above that of WGD in Figure 10. Another detailed comparison is carried out regarding the three new deep learning-based fusion models, and statistical outcomes are demonstrated in Table 5. It can be concluded that, AWGRD outperforms WGD and WGRD based on F-measure in this study. Since AWGRD also incorporates the important idea of residual networks with superposition's of previous layers but with less parameters (i.e., 1.42M parameters compared with 1.54M parameters of WGD and 1.58M parameters of WGRD as indicated in Table 1), it _ends a good balance between the effectiveness and efficiency among all three newly proposed models in this study.

TABLE 5. Multiple comparison tests among AWGRD, WGD and WGR based on F-measure in this study.



Model 1	Model 2	F-measure Mean Difference	A 95% Interval
AWGRD	WGD	0.00011	[0, 0.05
AWGRD	WGRD	0.00011	[0, 0.02

CONCLUSION

Video-based research into the identification of anomalous driving behavior is becoming vital, since it provides a robust and automated means of protecting motorists. It's a big deal since it's a necessary stepping stone to completely automated driving (especially in Level-3 and Level-4 phases according to the "autonomous driving" definition supplied by the US Department of Transportation's National Highway Traffic Safety Administration). In order to complete the video-based anomalous driving behavior detection challenge, this research introduces three innovative deep learning-based fusion models for the first time. They take their technical cues from the popular Dense Net model, which was developed relatively recently. WGD places an emphasis on the depth, breadth, and cardinality—three crucial factors in the construction of contemporary deep learning models. In it, WGD becomes much wider and has more cards. Incorporating the crucial concept of residual networks that are superpositions of earlier layers makes WGRD and AWGRD more advanced. Since superpositions of earlier layers allow for a complete description of both temporal and spatial latent information, this technique is particularly useful in the video-based anomalous driving behavior detection job. Extensive testing on the gold-standard Cagle state-farm distracted driver detection dataset and thorough comparisons with many other prominent deep learning models point to the superiority of the newly suggested deep learning-based fusion models in terms of both efficacy and efficiency.

REFERENCES

[1] W. Cao, J. Yuan, Z. He, Z. Zhang, and Z. He, ``Fast deep neural networks with knowledge guided training and predicted regions of interests for real-

time video object detection," *IEEE Access*, vol. 6, pp. 8990_8999, 2018.

[2] H. Shay, Q. Liu, K. Zhang, J. Yang, and J. Deng, ``Cascaded regional patio-temporal feature-routing networks for video object detection," *IEEE Access*, vol. 6, pp. 3096_3106, 2018.

[3] A. Nanda, P. K. Sa, S. K. Choudhury, S. Bash, and B. Mache, ``A neuromorphic person re-identi_cation framework for video surveillance," *IEEE Access*, vol. 5, pp. 6471_6482, 2017.

[4] L. Sun, Z. Jiang, H. Song, Q. Lu, and A. Men, ``Semi-coupled dictionary learning with relaxation label space transformation for video-based person re-identi_cation," *IEEE Access*, vol. 6, pp. 12587_12597, 2018.

[5] Y. Wu, Y. Sui, and G. Wang, ``Vision-based real-time aerial object localization and tracking for UAV sensing system," *IEEE Access*, vol. 5, pp. 23969_23978, 2017.

[6] S.-H. Lee, M.-Y. Kim and S.-H. Bae, ``Learning discriminative appearance models for online multi-object tracking with appearance discriminability measures," *IEEE Access*, vol. 6, pp. 67316_67328, 2018.

[7] M. S. Hossain and G. Muhammad, ``An emotion recognition system for mobile applications," *IEEE Access*, vol. 5, pp. 2281_2287, 2017.

[8] Z. Pan, X. Yi, and L. Chen, ``Motion and disparity vectors early determination for texture video in 3D-HEVC," *Multimedia Tools Appl.*, to be published. doi: 10.1007/s11042-018-6830-7.

[9] J. Wang, Z. Zhang, B. Li, S. Lee, and R. S. Sherratt, ``An enhanced fall detection system for elderly person monitoring using consumer home networks," *IEEE Trans. Consum. Electron.*, vol. 60, no. 1, pp. 23_29, Feb. 2014.

[10] Z. Zhang, X. Guo, and Y. Lin, ``Trust management method of D2D communication based on RF _ngerprint identi_cation," *IEEE Access*, vol. 6, pp. 66082_66087, 2018.

[11] J.Wang,Y. Cao, B. Li, H.-J. Kim, and S. Lee, ``Particle swarm optimization based clustering algorithm with mobile sink for WSNs," *Future Gener. Comput. Syst.*, vol. 76, pp. 452_457, Nov. 2017.

[12] Z. Xue, J. Wang, G. Ding, Q. Wu, Y. Lin, and T. A. Tsiftsis, ``Deviceto- device communications underlying UAV-supported social networking," *IEEE Access*, vol. 6, pp. 34488_34502, 2018.