Liu, P., Chi, H. L., Li, X., & Li, D. (2020, March). Development of a fatigue detection and early warning system for crane operators: A preliminary study. In Construction Research Congress 2020 (pp. 106-115). Reston, VA: American Society of Civil Engineers.

Development of a Fatigue Detection and Early Warning System for Crane Operators: A Preliminary Study

ABSTRACT

Fatigue of operators due to intensive workloads and long working time is one of the significant constraints lead to inefficient crane operations and safety issues. It can be potentially prevented through early warnings of fatigue for further appropriate work shift arrangements. Recently, many deep neural networks have been developed for the fatigue detection of vehicle drivers, through training and processing the facial image or video data of the drivers from available datasets. However, these datasets are difficult to be directly used for the fatigue detections under crane operation scenarios due to the variations of facial features and movement patterns between crane operators and vehicle drivers. Furthermore, there is no representative and public dataset with the facial information of crane operators under construction scenarios. Therefore, this study aims to explore and analyze the features of available datasets and the corresponding data acquisition methods suitable for crane operators' fatigue detection, further providing collection guidelines on crane operators dataset for early warning system development. A hybrid learning architecture is proposed by combining convolutional neural networks (CNN) and long short-term memory (LSTM) for fatigue detection. In order to establish a unified evaluation criterion, the effort of the study includes relabeling three available vehicle drivers datasets, NTHU-DDD, UTA-RLDD, and YawnDD, with human-verified labels at the minute segment level, and to train three hybrid fatigue detection models separately. Then the three trained models are used to evaluate the fatigue status on facial videos of licensed crane operators under simulated crane operation scenarios. The results show that the average test losses are 0.78458, 0.32191, and 0.20294 individually. One of the three datasets with the pretending facial fatigue features is comparatively more accurate in detecting operators' status than the rest of those with subtle facial fatigue features. Further comparisons in terms of labeling interval, environment, and accuracy are discussed for future dataset collections.

INTRODUCTION

With the development of prefabricated buildings, prefabricated products become more and more complicated. These products evolve from light-weight components, large and heavy modules to more substantial and more cumbersome pre-acceptance integrated units. Given this course of prefabricated product evolution, cranes perform a decisive role in the assembly of prefabricated products by lifting them vertically and horizontally (Chi *et al.* 2012). Although cranes are crucial components in many construction operations, they are also accompanied by a significant fraction of construction deaths. Estimates suggest that cranes are involved in up to one-third of all construction and maintenance fatalities (Neitzel *et al.* 2001).

Furthermore, operators' unsafe behavior is the main reason leading to crane safety issues, especially inadequate training and fatigue of practitioners causing unsafe practices of tower crane operations (Tam and Fung 2011). About 60.5% of the crane operators will continue to work even feeling fatigued as a result of long working hours, lack of rest breaks, and demanding physical works. 52.6% of the crane operators have not been arranged breaks every working day (Tam and Fung 2011). Therefore, the

operations and judgment of the crane operator will be a crucial factor for improving safety and productivity, particularly, in the construction site due to the high level of congestion and dynamics (Li *et al.* 2019). It is potentially beneficial to automatically detect and warn the fatigue or drowsiness of crane operators, which can provide supports for not only crane operators but also site superintendents and safety directors to make the proper shifts and breaks.

Although fatigue or drowsiness detection is an import research topic and successful solutions have been applied in domains such as driving and the workplace, there are few studies on developing fatigue detection and warning systems for the crane operators. Nowadays, there are several techniques for fatigue detection (Tadesse *et al.* 2014, Reddy *et al.* 2017, Ngxande *et al.* 2017): performance measurements, physiological measurements, and behavioral measurements. For instance, in the crane operations, performance measurements and physiological measurements include trolley movement speed, loads path deviation, jib rotation speed, heart rate, electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), electrocardiogram (ECG) and so on (Li *et al.* 2019). Although physiological measurements are highly correlated with the operator's mental state and are most sensitive to fatigue detection, they require operators to wear necessary sensory devices, which is an extra burden and inconvenience for operators. As for performance measurements, they are greatly affected by external factors and the operator's own operation habits. Therefore, such methods are intrusive and might not easy to be implemented under crane operation scenarios. Behavioral measurements are obtained from facial movements and expressions using none-intrusive sensors like cameras. The fatigue detection based on computer vision technologies to recognize facial expressions like eye blinks, yawning and nodding has not only high accuracy but also no impact on the operators' work. This kind of approach can analyze the facial features extracted from the facial videos/images. It performs a high accuracy after the boosting of the development of various deep neural networks. These deep learning approaches facilitate the computer to learn by itself for capturing the key features. For example, Zhang *et al.* (2016) adopted the convolutional neural network (CNN) to detect the yawning by using the features in nose region instead of mouth area due to the head turnings of vehicle drivers.

Previous works on fatigue or drowsiness detection focus on extreme fatigue with apparent signs such as yawning, nodding off, and prolonged eye closure (Ghoddoosian *et al.* 2019). However, for crane operators, such explicit signs may not appear until only moments before the accident. Therefore, it is necessary to detect fatigue at an early stage to provide more time for crane operators, site superintendents, and safety directors to make proper reactions. On the other hand, previous works on fatigue detection produced results on datasets that were either acted fatigue, like NTHU-DDD (Weng *et al.* 2016) (in simulated driving environment) and YawDD (Abtahi *et al.* 2014) (in real driving environment), or real fatigue, like UTA-RLDD (Ghoddoosian *et al.* 2019) (in realistic daily life). By "acted" means data where subjects were instructed to simulate fatigue or drowsiness, compared to "realistic" data. Besides, different datasets have various collection methods, testing environments, and label modes. It is difficult to compare the accuracy in fatigue detection among the available datasets. Furthermore, there is no large, public, and realistic dataset on crane operators.

The primary challenge is to determine which kind of available datasets and the collection method are most suitable for crane operators fatigue detection.

To address this issue, for exploring and analyzing which kind of available datasets and the corresponding data acquisition method are suitable for crane operators' fatigue detection, this study develops a hybrid learning architecture. It is designed by combining CNN and long short-term memory (LSTM) for fatigue detection. Firstly, this hybrid learning architecture is adopted for training on three available datasets re-labeled by the authors: NTHU-DDD, UTA-RLDD, and YawDD. Then the three trained models are evaluated on crane operators' facial videos. The objectives of this study are: (1) to accurately detect the facial regions along with critical fatigue features; (2) to compare and verify the training loss and accuracy on different datasets; (3) to explore and analyze which kind of datasets and the corresponding data collection method is suitable for crane operators' fatigue detection.

RELATED WORK

In a fatigued state, crane operators execute the repetitive lift tasks in a complex construction project may lead to catastrophic casualties as those of the vehicle drivers. There are apparent signs to tell whether an operator/driver is fatigue, such as repeatedly yawning, inability to keep eyes open, swaying the head forward, face complexion changes due to blood flow (Ngxande *et al.* 2017). As the facial features of the operator/driver in a fatigued state are significantly different from those of the conscious state, the real-time monitoring the operator/driver's face by the camera can be an efficient, non-invasive and practical approach. In the rest of this section, a review of the available datasets and existing detection methods will be provided.

Datasets

There are several datasets available for training and evaluate fatigue detection approaches, such as UTA-RLDD, NTHU-DDD, and YawDD. Each dataset has its own collection method and environment, label modes, dataset size, and whether the fatigue is "acted" or not. As a result, it is difficult to compare what kind of dataset is suitable for crane operators' fatigue detection. Furthermore, there is no such a large, public, and realistic dataset.

The University of Texas at Arlington Real-Life Drowsiness Dataset (UTA-RLDD) (Ghoddoosian *et al.* 2019) was created for the task of multi-stage drowsiness detection, targeting not only extreme and easily visible cases but also subtle cases when subtle micro-expressions are the discriminative factors. It consists of around 30 hours of RGB videos for 60 healthy participants. Each video was self-recorded by the participant, using a cell phone or a webcam. Detection of these subtle cases can be important for detecting drowsiness at an early stage, so as to activate drowsiness prevention mechanisms.

The NTHU driver drowsiness detection dataset (NTHU-DDD) (Weng *et al.* 2016) is a public dataset collected by Computer Vision Lab, National Tsing Hua University, which contains 36 IR videos under a variety of simulated driving scenarios, including normal driving, yawning, slow blink rate, falling asleep, burst out laughing, and so on. These videos are taken under day and night illumination conditions.

The Yawning Detection Dataset (YawDD) dataset (Abtahi *et al*. 2014) is collected by Distributed and Collaborative Virtual Environment Research Laboratory (DISCOVER Lab), University of Ottawa. It contains two available datasets that specifically target driver yawning. The first dataset contains 322 RGB videos and the second one includes 29 RGB videos, consisting of both male and female drivers, with and without glasses/sunglasses, from different ethnicities, and with 3 different mouth conditions: (1) normal driving with mouth closed (no talking); (2) talking or singing while driving; and (3) yawning while driving.

Fatigue Detection Methods

Fatigue is a risk factor at work as it may lead to decreased motivation, and vigilance, as well as potential accidents and injuries. With the development of machine learning techniques, more and more fatigue detection algorithms have adopted them as underlying learning architecture to analyze the facial features extracted from the video/images. It performs a high accuracy as it facilitates the computer to learn by itself for capturing the key features. For example, Mbouna et al. (2014) developed an approach to extract the visual features from the eyes and head pose of the drivers, and then support vector machines (SVMs) were used to classify the fatigue levels. Park et al. (2016) presented the Driver Drowsiness Detection (DDD) network consisting of three existing networks by SVMs to classify the categories of videos. Choi et al. (2016) trained the hidden Markov models (HMMs) to model the temporal behaviours of head pose and eye-blinking for identifying whether the driver is drowsy or not. Concurrently, features learned from unlabelled data based on deep neural networks, such as the convolutional neural network (CNN), have been proved to have a significant advantage over hand-crafted features in real-time monitoring of fatigue (Zhang et al., 2017). Lyu *et al.* (2018) improved the learning model on the temporal information by integrating CNN with LSTM network. This is a type of recurrent neural network (RNN) that can distinguish the states with long-term dynamical features over sequential frames.

PROPOSED SOLUTION

In this study, we proposed the architecture of hybrid deep neural networks taking the form of Lyu *et al.* (2018) and Li *et al*. (2019). The architecture of the networks and workflow can be seen in Figure 1 and 2. It has been proved to achieve high accuracy in fatigue detection. The architecture consists of three main modules: (1) Face Detector; (2) Spatial Feature Extractor; and (3) Temporal Feature Modeling. They are connected through several steps. Firstly, using multi-task cascaded convolutional neural networks (MTCNNs) to locate the bounding box of the facial area and the corresponding facial landmarks in each frame of the video, and extract the eyes, mouth, head area from the facial area. Secondly, the customized MobileNet (Efficient Convolutional Neural Networks for Mobile Vision Applications) is designed to extract the facial features from the individual-frame images. Finally, due to the fatigue features follow a pattern over time, an LSTM network is used to leverage the temporal pattern from a sequence of features within a specific time interval. The final output of this architecture is the fatigue level. Each step of the proposed module is detailed in the following sections.

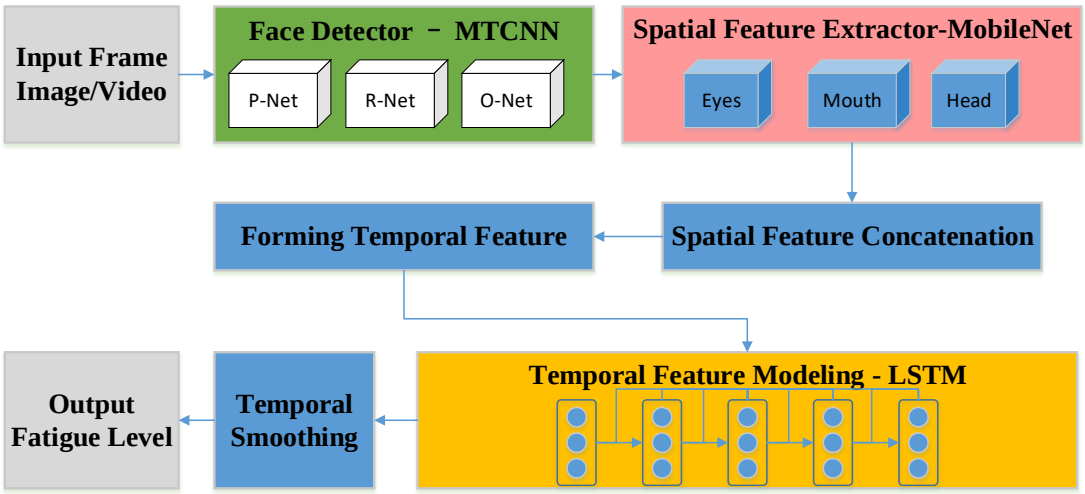


Figure 1. The architecture of the hybrid deep neural networks for fatigue detection

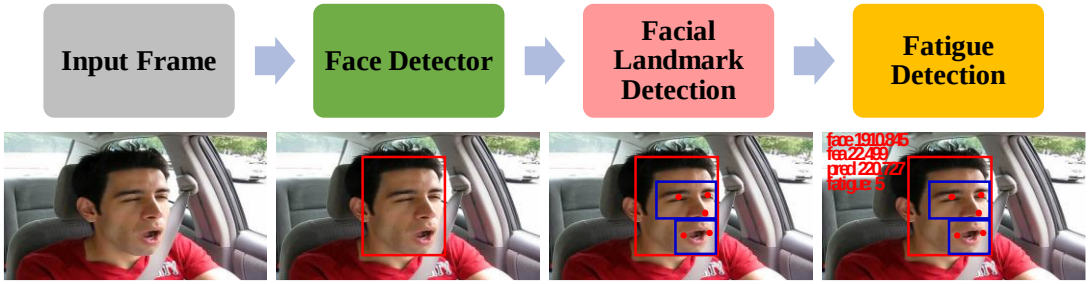


Figure 2. The workflow of the hybrid deep neural networks for fatigue detection

Face Detection

Crane operator fatigue detection in videos can be challenging because facial areas detection and alignment are affected by many factors, such as the lighting conditions, operator's gestures, low resolution, facial angles, expressions, occlusions, and so on. Therefore, it is critical to achieving precise facial detection before the facial feature extraction and fatigue detection. In a specific facial area, positioning the landmarks of mouth and eye areas is also very important. These areas contain significant fatigue features of the operators. The challenges to extract them could be exacerbated for crane operators' cases because of significant pose variations of the operator. He or she would change pose along with the moving loads, extreme lightings or darkness in operation cabin, and occlusions in front of the face (Li *et al.* 2019).

Table 1. Part of extracted facial landmarks by MTCNN

Frame	Sx	Sy	Ex	Ey	lx0	ly0	lx1	ly1	lx2	ly2	lx3	ly3	lx4	ly4
0	320	84	518	346	421	184	494	183	491	239	422	281	483	278
1	317	83	521	347	421	182	495	187	489	239	413	276	479	278

2	321	73	520	344	421	181	502	183	490	234	416	280	490	280
3	313	75	514	344	426	179	505	180	493	232	419	279	492	277
4	314	75	513	348	426	186	504	187	490	239	421	289	487	288

The MTCNN proposed by Zhang *et al.* (2016) is known as one of the fastest and most accurate face detectors. In order to solve the challenges mentioned above, MTCNN is applied to conduct the face detection and face alignment tasks with several stages. MTCNN consists of three network architectures (P-Net, R-Net, and O-Net) to obtain the facial bounding box and facial landmarks. Table 1 shows a part of the extracted facial landmarks, where Sx, Sy, Ex, and Ey represent the bounding boxes of head area; lx0, ly0, lx1, and ly1represent eyes locating points; lx2 and ly2 represent the nose locating points; and lx3, ly3, lx4, and ly4 represent the mouth locating points. Given that there is a correlation between extracted facial landmarks and fatigue levels, the goal of the proposed hybrid architecture is to find such correlations precisely.

Spatial Features Extraction

The objective of the spatial feature extraction is to learn a CNN-based feature extraction model for extracting the facial features from the individual-frame images. It includes head, eyes, and mouth determined by the facial landmarks through MTCNN in Face Detection module. In this study, we adopted MobileNet (Howard *et al.* 2017) as our primary approach to enable a fast and stable training process to generate the feature extraction model. The model has achieved good performance in image recognition of various datasets. Figure 3 demonstrates the improved MobileNet architecture, which includes 13 convolutional layers (grouped into Conv 1-13), 5 max-pooling layers (Max Pool 1-5), one average-pooling layer (Ave Pool) and one fully connected feedforward network layer (FC). In Figure 4, confusion matrices by CNN-based model show that the spatial features extraction model already achieves a high accuracy of detecting fatigue even though its prediction is merely based on a single frame.

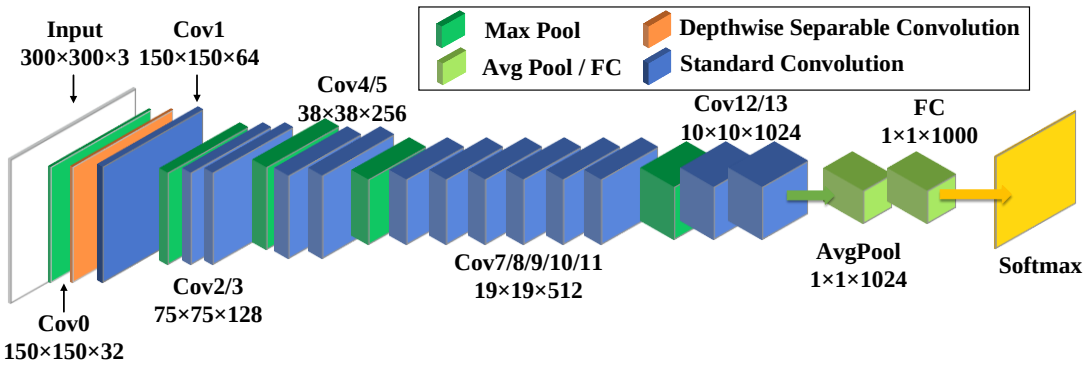
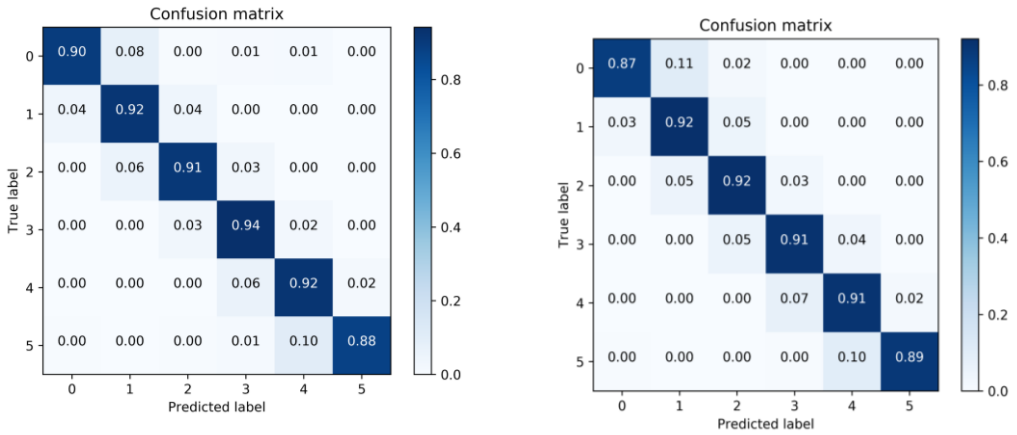


Figure 3. The architecture of CNN-based feature extraction model



(a) (b)

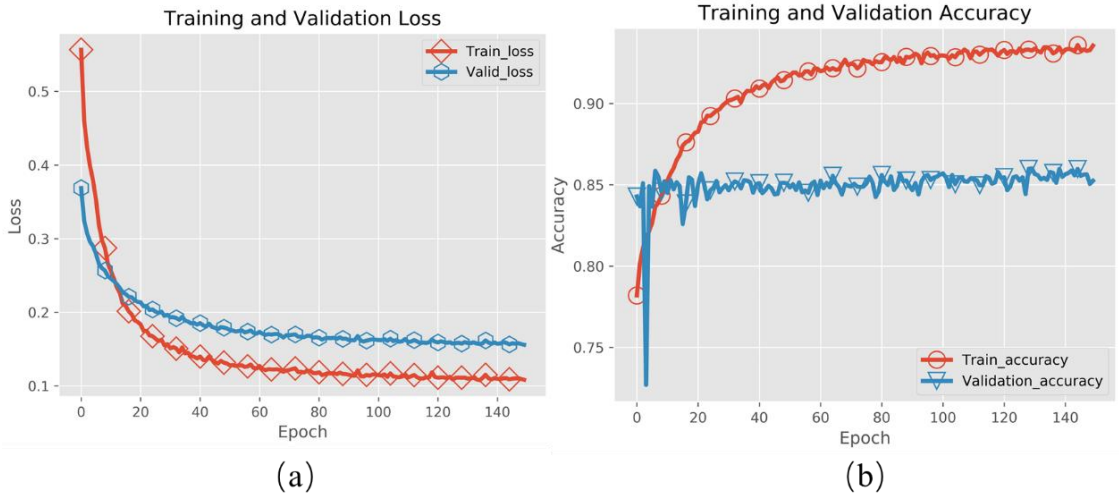
Figure 4. Confusion matrices by the CNN-based model: (a) train (b) test

Temporal Features Extraction

Although the feature extractor model has already enabled to predict the fatigue score of each frame based on the spatial features, sometimes it is still hard to discriminate the slight dynamic variations that have strong temporal dependencies, such as yawning and talking. Therefore, it can be beneficial to consider temporal information in the sequential frames. To this end, the deep LSTM (Greff *et al.* 2015) is applied to model the temporal features. LSTM has emerged as an effective and scalable model for several learning problems related to sequential data.

RESULTS AND DISCUSSIONS

Figure 5 demonstrates the average loss and accuracy on the training set (red line) and the validation set (blue line). The architecture of the proposed hybrid deep neural network with CNN and LSTM achieves 93.55% accuracy on the training set and 85.42% accuracy on the validation set in fatigue detection. Furthermore, as for fatigue level detection, according to the experiment results shown in Figure 6, the prediction values generated by LSTM align well with real label value.



(a) (b)

Figure 5. Loss and accuracy curve of the proposed architecture for both training set and validation set: (a)loss curve; (b)accuracy curve

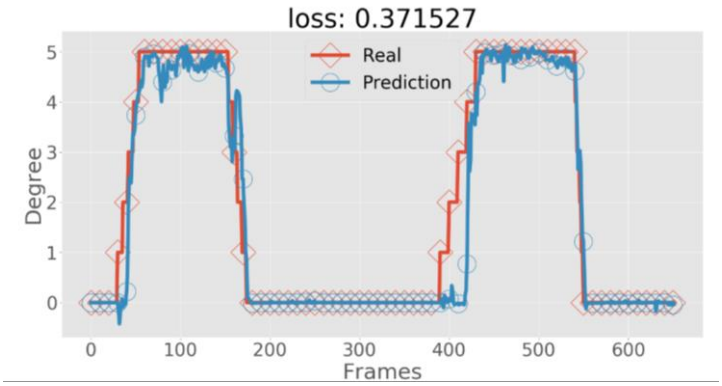


Figure 6. Loss curve of the proposed architecture on detecting fatigue level

In order to compare the accuracy of the trainsets of the three available datasets, we re-label a part of the three datasets. Due to the massive amount of work expected on manual annotation at the frame level, we decided to target a minute segment level for the human-verified re-labeling. 20-mins videos are selected as training data, and 10-min videos are used as the testing dataset. Table 2 represents the average losses and accuracies on the three available datasets. In terms of accuracy, the proposed hybrid architecture works well on NTHU-DDD (77.54%) and YawDD (72.63%). Both of the datasets contain pretending fatigue facial features in driving environments. However, the proposed architecture works less effective on UTA-RLDD (48.05%). The dataset contains subtle fatigue facial features in indoor real-life environments.

Table 2. Average losses and accuracies for different datasets

Datasets	Re-label Accuracy	Loss	Accuracy
UTA-RLDD	min	0.5129	0.4805
NTHU-DDD	min	0.1930	0.7754
YawDD	min	0.2128	0.7263

Next, the three trained hybrid learning models are evaluated on crane operator videos, as shown in Figure 7, in order to explore which dataset is more suitable as train set for operators' fatigue detection. Table 3 and Figure 8 represent the average losses on the crane operator dataset. The average loss based on train set of YawDD is 0.20294, which is smaller than that of NTHU-DDD (0.32191) and UTA-RLDD (0.78458).

(a) (b)

Figure 7. Simulation crane operation by expert operators: (a) fatigue detection for crane operators; (b) crane operation simulation

Table 3. Average losses of crane operator dataset through the proposed hybrid models on the trainsets of UTA-RLDD, NTHU-DDD, and YawDD

	Operator1	Operator2	Operator3	Operator4	Operator5	Average
UTA-RLDD	0.65110	0.60540	0.70331	0.97802	0.98511	0.78458
NTHU-DDD	0.55846	0.43907	0.01371	0.59764	0.00068	0.32191
YawDD	0.16595	0.14256	0.26288	0.19922	0.24410	0.20294

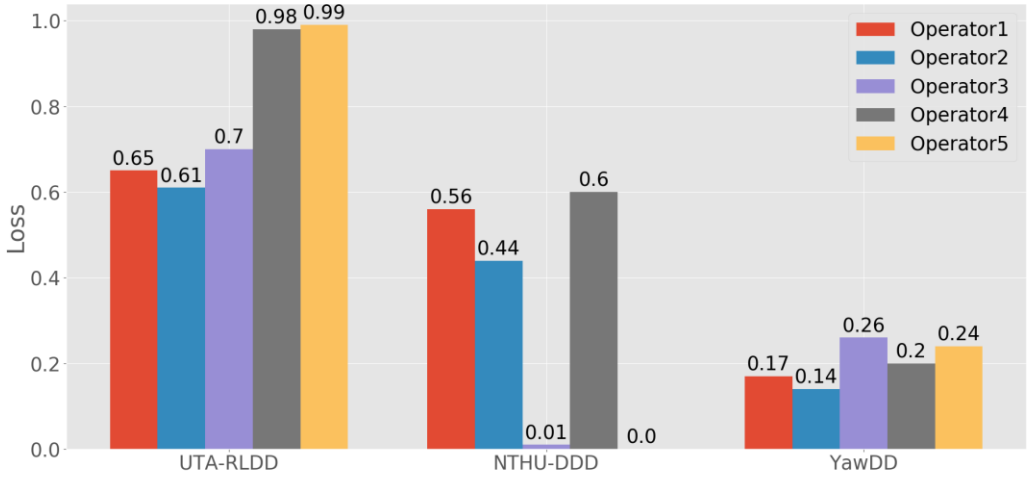


Figure 8. The comparison of losses for crane operator dataset through hybrid architectures on the trainsets of UTA-RLDD, NTHU-DDD, and YawDD

CONCLUSION

In this study, a hybrid learning architecture is proposed to combine CNN and LSTM to detect the fatigue status of the crane operators through facial videos. The improvements and contributions of this study are threefold: (1) expand the fatigue detection approaches for vehicle drivers to that for crane operators; (2) the trained hybrid learning models showed excellent performance in accurately detecting the facial regions with critical fatigue features; and (3) the exploration and analysis on which datasets and the corresponding data collection methods are suitable for crane operators' fatigue detection. The results of the experiment indicate that the proposed learning architecture works with high effectiveness on the crane operators' fatigue detection. As for available datasets, the dataset with apparent fatigue facial features in driving environments is comparatively easier for the detection than those with subtle fatigue facial features in indoor real-life environments. Also, labeling accuracy significantly affects detection. Still, there are some limitations to this study. Firstly, the testing results are based on simulation of crane operation. The realistic fatigue dataset for crane operators should be established for further testing. Secondly, three available datasets should be relabeled completely at the frame level to achieve a unified evaluation criterion. Thirdly, due to the limitation of the available datasets, the characteristics of the population like number, age, and years of experience are not considered in this study, which could be a possible future work.

REFERENCES

Abtahi, S., Omidyeganeh, M., Shirmohammadi, S., & Hariri, B. (2014, March). YawDD: A yawning detection dataset. In Proceedings of the 5th ACM Multimedia Systems Conference (pp. 24-28). ACM.

Chi, H. L., Chen, Y. C., Kang, S. C., & Hsieh, S. H. (2012). Development of user interface for tele-operated cranes. Advanced Engineering Informatics, 26(3), 641-652.

Ghoddoosian, R., Galib, M., & Athitsos, V. (2019). A Realistic Dataset and Baseline Temporal Model for Early Drowsiness Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.

Li, X., Chi, H. L., Zhang, W., & Shen, G. Q. (2019). Monitoring and Alerting of Crane Operator Fatigue Using Hybrid Deep Neural Networks in the Prefabricated Products Assembly Process. In Proceedings of the International Symposium on Automation and Robotics in Construction (Vol. 36). University of Alberta.

Neitzel, R. L., Seixas, N. S., & Ren, K. K. (2001). A review of crane safety in the construction industry. Applied occupational and environmental hygiene, 16(12), 1106-1117.

Ngxande, M., Tapamo, J. R., & Burke, M. (2017, November). Driver drowsiness detection using behavioral measures and machine learning techniques: A review of state-of-art techniques. In 2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech) (pp. 156-161). IEEE.

Reddy, B., Kim, Y. H., Yun, S., Seo, C., & Jang, J. (2017). Real-time driver drowsiness detection for embedded system using model compression of deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 121-128).

Tadesse, E., Sheng, W., & Liu, M. (2014, May). Driver drowsiness detection through HMM based dynamic modeling. In 2014 IEEE International conference on robotics and automation (ICRA) (pp. 4003-4008). IEEE.

Tam, V. W., & Fung, I. W. (2011). Tower crane safety in the construction industry: A Hong Kong study. Safety Science, 49(2), 208-215.

Weng, C. H., Lai, Y. H., & Lai, S. H. (2016, November). Driver drowsiness detection via a hierarchical temporal deep belief network. In Asian Conference on Computer Vision (pp. 117-133). Springer, Cham.

Zhang, K., Zhang, Z., Li, Z., and Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10):1499–1503.

Lyu, J., Yuan, Z., & Chen, D. (2018). Long-term multi-granularity deep framework for driver drowsiness detection. arXiv preprint arXiv:1801.02325.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. IEEE transactions on neural networks and learning systems, 28(10), 2222-2232.