**Related work for stress detection**

**1.XGBoost Algorithm-Based Monitoring Model for Urban Driving Stress: Combining Driving Behaviour, Driving Environment, and Route Familiarity:**

Stress is considered by many studies to affect traffic safety, and many researchers have attempted to monitor the dynamics of driving stress. Previous research has relied excessively on the positive effects of psychological indicators to improve the accuracy of stress monitoring models. However, psychological data collection sensors have not been widely used in conventional vehicles, which makes it impossible to apply the results of that research to actual driving tasks on a daily basis, even if the accuracy is high. This study designs a real driving task to extract data and proposes a driver’s driving stress monitoring model based on driving behaviour, driving environment, and route familiarity. The driving behaviour is described by the speed and acceleration of the vehicle, and the driving environment is quantified by a dilated residual networks (DRN) model that divides the video image from the full region into subregions according to the distribution of the driver’s attention. Based on the psychological data and driver stress inventory (DSI) results, the study used a K-means 3D cluster analysis to obtain the evaluation method of driving stress and constructed an extreme gradient boosting (XGBoost) model to monitor driving stress. Comparisons of performance with other models show that the XGBoost model significantly outperforms the other three mainstream machine learning algorithms and exceeds most traditional models without the use of psychological data. The model’s performance indicators, accuracy, sensitivity, and precision, reached 91.18%–93.25%, 84.13%–89.37%, and 90.25%–91.34%, respectively. The study also summarises the ranking of effects of different scene elements on driving stress for each visual field. The results could make it possible to apply stress monitoring on a large scale to real driving situations, providing urban designers with advice on how to reduce driver stress and directing their attention to those visual areas and visual scene elements that have a higher impact on driving stress and need improvement. This pilot of this study was conducted in Shenzhen, China, from May to December 2019. To eliminate the specificity of the experimental data, the experimental sites were selected in six regions with different community landscapes and road network conditions in Luohu District, Futian District, and Nanshan District, Shenzhen. Regarding the urban development features in Shenzhen, Luohu District, Futian District, and Nanshan District correspond to old town, CBD, and suburban areas, respectively. A closed experimental route is defined in each region and the road sections and intersections are numbered in advance to facilitate later data processing, and it is forbidden to inform the participant of the defined route before the experiment. The six selected experimental routes contained a total of 458 sections and intersections.

**2.** **Improving the Performance of Deep Learning in Facial Emotion Recognition with Image Sharpening:**

CNN is a class of Deep Neural Network which performs well for image related tasks as they tend to capture spatial information. The input to the CNN is a 48\*48 grayscale image. These input images are the output of the Image Data Generator which produces rescaled, rotated, flipped, and sharpened images. The output of the CNN is a probability of seven categories of facial expressions. It is composed of five convolutional layers, three max pooling layers, and three fully connected layers as shown in Fig. 3. All the convolutional layers use “same” padding and “ReLU” activation function. The first two fully connected layers use “ReLU” as the activation function and the last fully connected layer uses “Softmax” activation function. The first, third, and fifth convolutional layers are followed by batch normalization, max pooling of size 2x2 and stride 2x2, and dropout of 20%. The first convolutional layer is composed of a 5x5 kernel with 64 filters. The second and third convolutional layers are composed of 3x3 kernels with 128 filters. The fourth and fifth convolutional layers are composed of 3x3 kernels with 256 filters. The first fully connected layer consists of 1024 neurons and the second fully connected layer consists of 512 neurons followed by a drop out 20%. The final fully connected layer, which produces the output, consists of 7 neurons representing the 7 categories. The model uses the “categorical crossentropy” loss function and “adam” optimizer. The total trainable parameters in the model are 11, 075, 847 and non-trainable parameters are 896. Keras callback ReduceLROnPlateau has been used to tweak the learning rate if the validation accuracy does not increase by a certain amount for 10 epochs. The model was trained for 100 epochs using a batch size of 128.

**3.Deep Learning Model Comparison for Vision-Based Classification of Full/Empty-Load Trucks in Earthmoving Operation:**

Earthmoving is an integral civil engineering operation of significance, and tracking its productivity requires the statistics of loads moved by dump trucks. Since current truck loads’ statistics methods are laborious, costly, and limited in application, this paper presents the framework of a novel, automated, non-contact field earthmoving quantity statistics (FEQS) for projects with large earthmoving demands that use uniform and uncovered trucks. The proposed FEQS framework utilizes field surveillance systems and adopts vision-based deep learning for full/empty-load truck classification as the core work. Since convolutional neural network (CNN) and its transfer learning (TL) forms are popular vision-based deep learning models and numerous in type, a comparison study is conducted to test the framework’s core work feasibility and evaluate the performance of different deep learning models in implementation. The comparison study involved 12 CNN or CNN-TL models in full/empty-load truck classification, and the results revealed that while several provided satisfactory performance, the VGG16-FineTune provided the optimal performance. This proved the core work feasibility of the proposed FEQS framework. Further discussion provides model choice suggestions that CNN-TL models are more feasible than CNN prototypes, and models that adopt different TL methods have advantages in either working accuracy or speed for different tasks. The main contributions of the paper can be summarized as: (1) The framework of an automated, non-contact FEQS applying vision-based deep learning is presented, which has advantages toward existed FEQS methods in terms of manual effort, costs, and errors; (2) the core work of the framework, i.e., the classification of full/empty-load trucks in earthmoving operations, is assessed in terms of feasibility through a comparison study that involves multiple deep learning models; and (3) the comparison study results are further discussed to give model choice suggestions for future implementation of the proposed FEQS. The rest of the paper is as follows: Related work of the current state of art and practice in the domain of earthmoving FEQS and vison-based deep learning is first reviewed to identify the gaps in knowledge and thus to explain the proposed FEQS framework and the authors’ research. Then, the methodology is described, followed by the comparison study results and discussion. Finally, the conclusions of this study are provided in terms of its contributions to knowledge and practice, along with limitations and future work of the study.

**4.Detection of Stress Using Image Processing and Machine Learning Techniques:**

Different people may behave or express differently under stress and it is hard to find a universal pattern to define the stress emotion. And it is not easy train a model offline that classifies and predicts if a person is stressed .To ease the problem, we have evolved with an algorithm that relates to the facial area of interest in stress recognition. Eyebrow which shows a rigid transformation is used as the facial area of interest in stress recognition. The eyebrow movement has not got any universal pattern but the variation corresponding to a particular person is used to detect the stress involved. We consider that stress is detected if a person shows high variation of the eyebrows constantly within a fixed time interval.

The face acquisition module processes the video sequences captured by the camera. The image frames are extracted and the pre-processing of the images for subsequent analysis in the further modules is done. Pre-processing of the images includes two transformation of extracted frame. First one being the pixel transformation and other one is binary transformation of the modified image after pixel transformation. Several approaches are applied to extract the discriminative features to learn the pattern of the different facial features. We investigate approaches based on the pixel analysis of the images which are normalized to standard scale (200 x 200) pixels. Pixel value analysis is the method of analysing an input image from the extreme left top, this involves the analysis of every pixel it encounters in every row of the normalized image. The first sub-module offline displacement calculation, calculates the displacement of the eyebrow position using the obtained coordinates in the previous steps, with respect to its mean position. The coordinates in the first image gives the x and y co-ordinates of the eyebrow as (20, 181) and the next image gives the co- ordinates (21, 202) , this sub-module gives the mean displacement of the eyebrow and this way the displacement calculation for each image with respect to first image is calculated.

**5.Modified Convolutional Neural Network Architecture Analysis for Facial Emotion Recognition:**

The dataset used to train and test the different architecture of hidden layer in the deep neural network of the convolutional neural network is the Karolinska Directed Emotional Faces (KDEF) [10] developed by the Emotion Lab at Karolinska Institutet Sweden. The dataset consist a total of 4900 images of 562\*762 categorized in to seven different emotions i.e. Afraid, Angry, Disgusted, Happy, Neutral, Sad and Surprised. The dataset is divided into test and train dataset in 80% - 20% split. The train dataset consist of 3920 images divided into seven categories and the test dataset consist of 980 images divided into seven categories. The images are scaled down from 562\*762 pixels to 256\*256 pixels before feeding it into the CNN model. The Convolution Neural Network model is trained on NVIDIA GEFORCE GTX 960M Graphics Processing Unit (GPU) with 4 GB of dedicated graphic memory and Intel Core i7 6700 HQ CPU 2.60GHz with 8 GB of RAM on Asus ROG GL552VW. The working environment is Spyder editor for Python 3.5. Keras and NumPy libraries were used with TensorFlow as the backend. The models were trained for 25 epochs with 3920 steps per epochs for training set and 980 steps for validation set. The images present in the dataset were preprocessed by using the Image Data Generator class which generates batches of tensor images. In this method the images were re-scaled by a factor up to 1/255 . The images were randomly flipped in horizontal direction in order to generate randomness in the input image while training the model. Images were sheared in counter clockwise direction up to 0.2 degrees and the zoom range for the images were set to be about 0.2 to provide random zoom. Venturi Architecture is our proposed architecture for the hidden layer of the deep neural network in the convolutional neural network. The architecture consist of 6 layers in the hidden layer with one output layer consisting of 7 nodes based on the 7 different categories in which the facial emotions are classified.

**6. Real-time Facial Expression Recognition “In The Wild” by Disentangling 3D Expression from Identity:**

The proposed system Motivated by the progress in the 3D facial reconstruction from images and the rich dynamic information accompanying videos of facial performances, we collected a large-scale dataset of facial videos from the internet (section III-B) and recovered the per-frame 3D geometry thereof with the aid of 3D Morphable Models (3DMMs) [3] of identity and expression (section III-A). The annotated dataset was used to train the proposed DeepExp3D network, in a supervised manner for regressing the expression coefficients vector e f from a single input image If (section III-C). As a final step, a classifier was added to the output of the DeepExp3D to predict the emotion of each estimated facial expression, and was trained and tested on standard benchmarks for FER. Combined Identity and Expression 3D Face Modelling: we model the 3D face geometry using 3DMMs and an additive combination of identity and expression variation. In more detail, let x = [x1, y1,z1,..., xN, yN,zN] T ∈ R 3N be the vectorized form of a 3D facial shape consisting of N 3D vertices. We consider that any facial shape x can be represented using the following model of shape variation:

x(i, e) = x¯ +Uidi+Uexpe (1)

where x¯ ∈ R 3N is the overall mean shape vector, given by x¯ = x¯id +x¯ exp, where x¯id and x¯ exp are the mean identity and mean expression shape vectors respectively. Uid ∈ R 3N×ni is the orthonormal basis with ni = 157 principal components (ni 3N) , Uexp ∈ R 3N×ne is the orthonormal basis with the ne = 28 principal components (ne 3N), and i ∈ R ni , e ∈ R ne are the identity and expression parameters. In the adopted model (1), the 3D facial shape x is a function of both identity and expression coefficients (x(i, e)). Additionally, the expression variations are effectively represented as offsets from a given identity shape. The identity part of the model, {x¯id,Uid}, originates from the LSFM [4] built from approximately 10,000 scans of different people, the largest 3DMM ever constructed, with varied demographic information. In addition, the expression part of the model, {x¯ exp,Uexp} originates from the work of Zafeiriou et al. [66], who built it using the blendshapes model of Face warehouse [6] and adopting Nonrigid ICP [9] to register the blendshapes model with the LSFM model.

**7. Video-Based Stress Level Measurement Using Imaging Photoplethysmography**

In this section, we describe a method for measuring stress- level

making full use of pulse rate variability features from video

recordings. The flow of our algorithm . First, pulse wave is extracted from recorded facial videos using seven methods in total for benchmark. Second, five features are extracted based on the pulse rate variability. Third, stress levels are classified using k-nearest neighbor method, denoted as k-NN, from extracted features. Details in each step and experimental setup are described as the following subsections. Post processing are applied into all extracted signals for detrending and smoothing. The pulse wave is produced for detrending using smoothness prior [30] (O=2000) and bandpass filtering as the cutoff frequency of from [0.75Hz 3.0Hz] corresponding a range of heart rate in adults. After the bandpass filtering, in order to refine the estimation of PRV, a cubic spline interpolation is applied into filtered signal for up sampling from 30Hz to 500Hz. And then, interpolated signal is smoothed again using five points moving averaging filter . For the analysis of pulse rate variability, peak to peak interval is obtained by calculating the interval time of neighboring peaks in

time series. Lomb-Scale periodogram of peak to peak intervals is obtained referred in [31, 32]. Since papers are proposed for investigating the fact that pulse rate variability is representative indicators of heart rate variability proposed by [33]. Referred in [33],we used pulse rate variability as representative indicators of heart rate variability in this paper. With regard to the features of PRV for measuring mental stress, Task force of European Society of Cardiology and the North American Society of Pacing and Electrophysiology have proposed the standards of measurement and the interpretation of heart rate variability [34]. As an example of making full use of these features, a paper was proposed for estimating stress levels of MDs during surgeons [18]. Especially, fourteen features indicated the significance between relaxation and stress conditions. In this paper, we selected five features for classifying stress levels based on Relief-F algorithm. K-nearest neighbor method is a type of classical approach for classification of data. The algorithm describes as following two steps. K-samples of nearest neighbors are selected around from a sample for test data. A class which includes the most classes in K-samples is determined as classification result. In this paper, five features are selected from extracted fourteen features of PRV parameters based on the Relief-F algorithm which calculates the importance of each feature [35]. We note that a sampling number k is an important parameter which reflects the accuracy of classification. Through an investigation of sampling number k which are range of from one to ten, we set the sampling number k as eight

in this paper. The classifier is trained by cross validation. The accuracy of classification is defined by taking ratio of a number of samples which classify correctly and total samples.