Respiration Rate Detection for In-Cabin Passenger Monitoring

Aaron Brennan, Dr. Amr Elrasad, Dr. Ihsan Ullah

a.brennan55@nuigalway.ie

Ollscoil na Gaillimhe University of Galway · XPERI · Insight SFI Research Centre for Data Analytics

INTRODUCTION

Road safety is a critical issue globally, with driver fatigue being a significant contributor to accidents. According to the European Road Safety Observatory, an estimated one in five fatal crashes is attributed to drowsy driving each year. In this context, our study proposes a novel deep learning model for detecting drivers' respiration rates using a thermal camera. Unlike traditional methods that require signal extraction from specific facial regions of interest, our approach simplifies the detection process by predicting respiration rates directly using deep learning. Our model was rigorously evaluated using new data acquired in a simulated driving environment, additionally we introduced a unique data augmentation technique in the training loop to reduce over-fitting, a common challenge in deep learning models utilizing temporal data.

NOVEL DATASET

The data acquisition conducted by Xperi involved designing an experimental setup that simulated a realistic driving experience, enabling the collection of thermal images and respiration data. The data acquisition process was conducted with participants in a mock driving environment, which consisted of a stationary car equipped with three monitors placed in front of the driver. These monitors provided a 180-degree field of vision to simulate an immersive driving experience. A thermal camera was positioned in front of the participants to capture the heat changes around their faces, and the subjects wore a respiration belt to record their respiration rates throughout the experiment.

The data collection was divided into two distinct sections: the non-noisy dataset and the noisy dataset. In the non-noisy dataset, subjects were instructed to remain still for two minutes without engaging in any driving activity. This allowed baseline respiration data and thermal images to be collected in a controlled setting. In the noisy dataset, participants were asked to drive normally in the simulated environment, creating a more realistic and dynamic scenario that emulated real-world driving conditions.

This study involved a total of 76 adult participants, comprising 45 males and 31 females. Of these, 63 participants were assigned to the training group and 13 to the testing group. The training group consisted of 38 males and 25 females, while the testing group comprised 7 males and 6 females. All participants were used for both the non-driving and driving experiments

IMPLEMENTATION

Using a pre-trained ResNet18 model for our deep learning experiments, we adjusted the final layer of this model to generate a single output, representing the predicted respiration rate in breaths per minute. The input layer of the model was tailored to accommodate both the duration and the frame rate of the input frames. Given that respiration typically occurs at a low frequency (0.15-0.35 Hz), we experimented with down-sampling the frame rate to decrease the amount of data necessary for the model.

The input size to the model was calculated using the formula:

Input Size = Time × Frame Rate

where time is measured in seconds and the frame rate is in Hz.

For our loss function, we utilized the Mean Squared Error (MSE) criterion, as this problem sought to estimate respiration rates in breaths per minute through regression.

During the training process, we implemented an innovative data augmentation technique to enhance the robustness of the model. First, we calculated the standard loss **(loss1)** by passing the normal input through the model and comparing the output to the ground truth label.

$$normal\_output = model(normal\_input)$$

$$\text{loss1} = criterion(label, normal\_output)$$

Subsequently, we reversed the order of the frames in the normal input to create a backward input, representing a sequence of frames moving backward in time. We then passed this reversed input through the model and compared the backward output to the normal output, resulting in a second loss **(loss2).**

$$backward\_input = reverse(normal\_input)$$

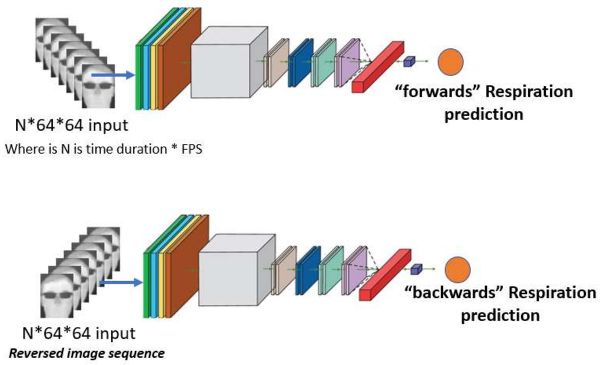
$$backward\_output = model(backward\_input)$$

$$loss2 = criterion(normal\_output, backward\_output)$$

This technique was predicated on the assumption that the model should predict the same respiration values, irrespective of whether the input frames were moving forward or backward in time. By incorporating this assumption into the loss function, we encouraged the model to learn temporal invariance and improve its ability to generalize across various scenarios.

To compute the final loss, we combined loss1 and loss2 with a weighting factor *a*. We selected the value of *a* based on the desired trade-off between fitting the ground truth labels **(loss1)** and ensuring temporal invariance **(loss2)**. A higher value of *a* would prioritize fitting the ground truth, while a lower value would emphasize temporal invariance. In our experiments, we found that an *a* value of 0.5 provided a suitable balance between these competing objectives.

$$\text{final\_loss} = \alpha \cdot \text{loss1} + (1 - \alpha) \cdot \text{loss2}$$



RESULTS AND CONCLUSIONS

The model's ability to detect respiration rates was evaluated under both non-driving and driving conditions. The performance was assessed using different lengths of time (30, 45, and 60 seconds) and frame rates (3, 5, and 10 FPS) for capturing thermal images. The results, represented as the mean absolute error (MAE) in breaths per minute, showed that higher frame rates and longer time durations yielded better results. This suggests that the model can learn and identify respiratory patterns more effectively with more data and a longer observation period. In addition, the use of a backwards loss strategy significantly improved the model's performance in the driving experiments, where the data is inherently noisier. This indicates that advanced loss functions or regularization techniques can enhance the model's robustness in handling noisy data.

NON-DRIVING EXPERIMENT RESULTS

Length of Time	3 FPS	5 FPS	10 FPS
30 seconds	1.64	1.57	1.44
45 seconds	1.38	1.25	1.15
60 seconds	1.26	1.17	1.03

DRIVING EXPERIMENT RESULTS

Length of Time	3 FPS	5 FPS	10 FPS
30 seconds	1.80	1.68	1.64
45 seconds	1.56	1.37	1.28
60 seconds	1.42	1.21	1.19

EFFECT OF USING BACKWARD LOSS AUGMENTATION

Length of Time	using backward loss	without backward loss
30 seconds	1.64	1.89
45 seconds	1.28	1.64
60 seconds	1.19	1.43

While the ResNet-18 model shows promise in detecting drowsiness through respiration rate prediction, it is not sufficient as a standalone drowsiness detection system. However, it could potentially contribute to a more comprehensive detection system when combined with other physiological indicators, such as heart rate and blink rate/length. Our network currently needs around 20ms to infer the respiration rate making it suitable to work in real-time. One limitation is the model's inability to predict respiration rate variability effectively, a crucial factor in detecting drowsiness.

ACKNOWLEDGEMENTS

This publication has emanated from research supported by grant from Science Foundation Ireland under Grant number [SFI/12/RC/2289 P2]