# An Embedded Bed-Exit Monitoring System Using Deep Learning

IEEE ICASI 2024

**Chi-Huang Shih, Pei-Jung Lin, Yeong-Yuh Xu ***

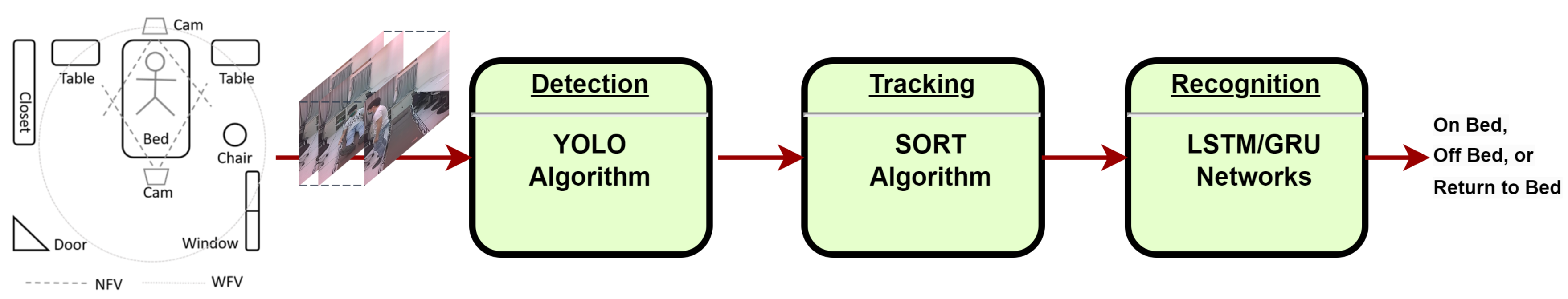
yyxu@ncut.edu.tw (*Corresponding Author)


## Abstract

- **For patients at high risk of falling, detecting bed-exit behaviors is considered the first defense against falls.**
- **This study presents a practical, cost-effective, and easy-to-use bed-exit monitoring system for healthcare settings, designed to preserve privacy and ensure accuracy by detecting bed-exit behavior through a narrow field of view.**
- **The proposed system utilizes a comprehensive three-stage approach. At first, the YOLO algorithm detects the human body trunk within the scene. Following detection, the SORT algorithm tracks the detected body trunk objects across different frames. The final stage leverages deep learning models, notably LSTM or GRU networks, to precisely categorize the movements of the human body trunk.**
- **The results of the experiment show that the proposed system is very efficient in getting an accurate result, 97.97%, with regard to the speed of processing, 7.1 frames per second, which points out great advancement in precise bed-exit monitoring for patient care in healthcare.**


## System Overview

The prototype is powered by the Jetson Xavier NX and uses a CSI camera for video data acquisition. To meet the practical requirements of caregiving environments, our research utilizes an RGB camera with infrared night vision strategically placed at either the head or foot of the bed. This enables the acquisition of near-field vision (NFV) images around the clock.

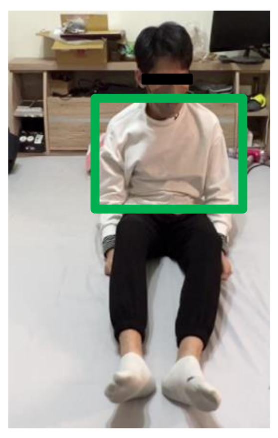



Upon capturing NFV images, the following steps are undertaken to recognize the patient's behavior.

1. The YOLO (You Only Look Once) algorithm is utilized to detect the human body trunk within the scene.
2. The SORT (Simple Online and Realtime Tracking) algorithm is employed to track the detected body trunk objects across frames.
3. The deep learning techniques such as LSTM (Long Short-Term Memory) or GRU (Gated Recurrent Unit) networks are used to classify the tracked objects' actions into three categories: getting off the bed, being on the bed, and returning to the bed.

## Experiments

To verify the proposed method's accuracy and effectiveness, we established a dataset of NFV images depicting bedside behaviors. To align with real-world care settings, the dataset's images are divided into daytime and nighttime categories. Experiments were conducted focusing on object detection and behavior recognition.



## Statistics in the dataset

For both daytime and nighttime categories, the data ratio for training, validation, and testing is 7:2:1, with the test data being independent of training and validation purposes. The dataset comprises 6,937 images.

| Image Mode | Train | Validate | Test |
|---|---|---|---|
| Day | 2983 | 658 | 329 |
| Night | 2017 | 634 | 316 |
| All | 5000 | 1292 | 645 |

## Object Detection Results

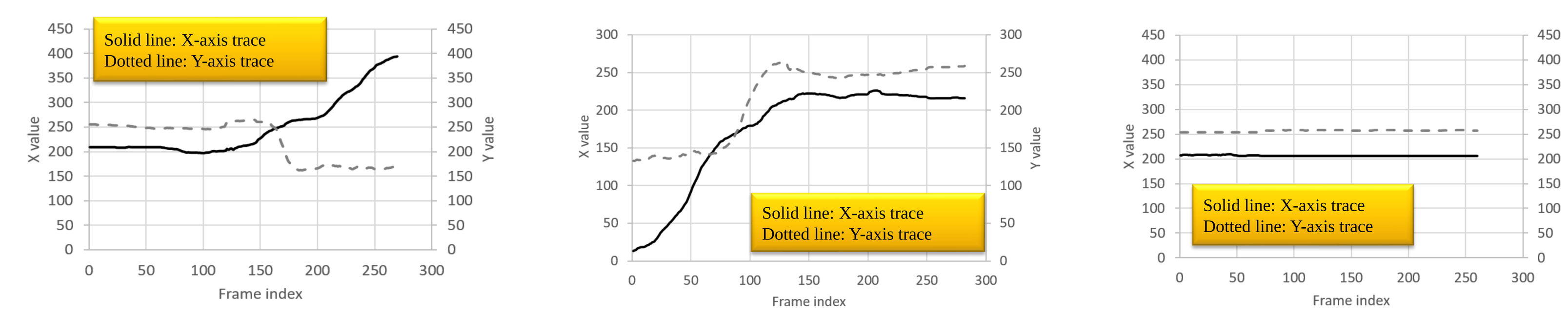
We annotated the human trunk regions in the NFV images as our detection targets. The performance metrics adopted for evaluation include precision, recall, F1 score (F1), and average precision (AP).

| | Validate | | | | Test | | | |
|---|---|---|---|---|---|---|---|---|
| Mode | Precision | Recall | F1 | AP | Precision | Recall | F1 | AP |
| Day | 99% | 99% | 99% | 99.68% | 99% | 99% | 99% | 99.17% |
| Night | 98% | 95% | 96% | 97.36% | 100% | 96% | 98% | 97.98% |
| All | 99% | 97% | 98% | 98.57% | 100% | 98% | 98% | 99.30% |

- The results show that our system achieves a precision of 99% for daytime images and over 98% precision for nighttime images.
- These outcomes indicate that effective human detection obtained during the object detection phase can facilitate subsequent behavior recognition.

## Behavior Recognition Results

During the behavior recognition phase, we initially employ the SORT algorithm to track objects' movement trajectories across consecutive image frames, generating a set of discrete time-series data.




(a) right exit (b) left return (c) lie still

In this experiment, 159 videos of bedside behavior were utilized to evaluate the recognition performance for three types of behaviors: on-bed, off-bed, and returning-to-bed. Experimental data were generated using various combinations of time windows and step sizes, and five different deep learning techniques were employed.

*Table 1. Accuracy and time results for varied window and step sizes*

| | Window size | | | | | |
|---|---|---|---|---|---|---|
| Step size | 50 | | 100 | | 200 | |
| | Acc. | Time(s) | Acc. | Time(s) | Acc. | Time(s) |
| 5 | 87% | 31.29 | 97% | 30.54 | 100% | 32.75 |
| 10 | 80% | 33.47 | 100% | 32.02 | 97% | 32.02 |
| 20 | 70% | 33.47 | 90% | 34.93 | 100% | 32.02 |
| 40 | 43% | 36.39 | 97% | 32.02 | 97% | 34.93 |

*Table 2. The selected hyperparameters for five deep-learning network models*

| Model | Hyper parameters |
|---|---|
| CNN | Filter = 16, Kernel size = 3, Dense = 32; |
| LSTM | Neurons = 32, Dense = 16 |
| GRU | Neurons = 32, Dense = 64 |
| CNN+LSTM | Filter = 8, Kernel size = 7, Neurons = 8, Dense = 64 |
| CNN+GRU | Filter = 8, Kernel size = 5, Neurons = 8, Dense = 64 |

*Table 3. Performance comparison among deep-learning network models*

| Model | Accuracy | Loss | Parameters | FLOPs |
|---|---|---|---|---|
| CNN | 96.87% | 0.1053 | 25,077 | 30,719 |
| LSTM | 97.97% | 0.0615 | 477,061 | 134,400 |
| GRU | 97.89% | 0.0586 | 59,269 | 102,368 |
| CNN+LSTM | 97.25% | 0.0882 | 48,653 | 79,279 |
| CNN+GRU | 97.39% | 0.0835 | 112,149 | 297,727 |

*Table 4. Behavior recognition results*

| Mode | On bed | Off bed | Return |
|---|---|---|---|
| Day | 100%(22/22) | 100%(31/31) | 100%(34/34) |
| Night | 100%(15/15) | 100%(28/28) | 100%(29/29) |
| All | 100%(37/37) | 100%(59/59) | 100%(63/63) |

As shown in Table 3, all models reach 97%- 98% accuracy, with LSTM performing best. CNN+GRU achieves slightly lower accuracy with fewer parameters and less computation, which is more suitable for embedded system operations. Table 4 shows that the proposed bed-exit behavior recognition system achieves a 100% recognition rate, proving the system's feasibility.


National Chin-Yi University of Technology
NCUT