Proceedings of the INMM & ESARDA Joint Virtual Annual Meeting
August 23-26 & August 30-September 1, 2021

A Novel Detection Approach to Preventing Theft of Nuclear Materials Using Deep Learning-Based Object Detection and Human Pose Estimation

Yuki Yokochi	**Kazuyuki Demachi**	**Shi Chen**
University of Tokyo	University of Tokyo	University of Tokyo

ABSTRACT

According to the Incident and Trafficking Database (ITDB) of IAEA, nuclear security incidents occur in the world about once every three days, with theft of nuclear and radioactive materials being the most common. Since the theft of nuclear materials takes place under the disguised as normal operations, currently there are few effective approaches to monitoring and preventing the theft of nuclear materials. To this end, we propose a method to detect theft disguised as normal operations from surveillance cameras based on a temporal sequence of unit actions that not include typical actions, while considering diversity of sequence. First, we employing deep learning-based object detection, human pose estimation models, and human motion recognition model. We first deploy the YOLOv3 model for the object detection task to recognize and localize the storage containers of radioactive materials (e.g., polythene bottles for medical radioisotopes storage), lockers (e.g., cupboards, refrigerators), and tools (e.g., scissors, screwdriver). Then, we detect persons and extract two-dimensional joint positions of the persons in the image using a pre-trained human pose estimation model, i.e., OpenPose. We then dynamically measure the distance between the objects and the key points of the detected persons and set the thresholds to detect the theft. For the human motion recognition task, we deploy the LSTM model, which detects motion based on the temporal sequence data of the 2D skeleton detected by OpenPose. Finally, the unit actions identified from the above three models, and they are coded in a temporal sequence. This sequence is compared with a predetermined sequence to judge as theft, and this sequence is determined by applying unique rules. The proposed method can detect theft that is disguised as a normal operation, and it can overcome the case where a typical theft unit action is not included and various theft operation procedures. A future work is to implement and evaluate the proposed method for detecting unit actions and the theft detection method using the sequenced data.

INTRODUCTION

According to the IAEA's Incident and Trafficking Database (ITDB) [1] (Figure.1), there are approximately 140 nuclear security-related incidents worldwide each year. The largest proportion of these cases is the theft of nuclear materials and radioactive materials by insiders, demonstrating that the current insider detection system is insufficient. Material accountancy currently applied as a countermeasure for nuclear materials management. However, it can be difficult or delayed detecting theft since insiders can forge records. Nuclear facilities address this issue by increasing the frequency of material accountancy and the number of people in charge but this can also lead to increasing cost for facilities. At present, there is already some intelligent monitoring software (e.g., VAAKEYE [2]) that can be deployed in potential on-site theft such as warehouses. In the case of detecting theft in a warehouse, it is possible to identify whether objects that should not be removed are theft by observing each action, so it is sufficient to identify theft when a typical unit theft action is detected. In contrast, to detect theft in nuclear facilities, theft disguised as normal operations needs to be detected as well. If the theft is disguised as normal operations, it can not be detected by identifying each typical unit theft action, thus it is necessary to detect theft from a temporal sequence of unit actions. Furthermore, the sequence of unit actions may vary depending on the person performing the theft. Therefore, this study aims to detect thefts disguised as normal operations from surveillance videos based on a temporal sequence of unit actions that not include typical actions, while considering the diversity of sequences. The proposed detection approach consists of the following three steps:

1. Clarify the difference of temporal sequences between normal operations and theft of nuclear materials at nuclear facilities.
2. Propose a method to detect the on-site theft based on the deference.
3. Implementation and evaluation of the proposed detection method.

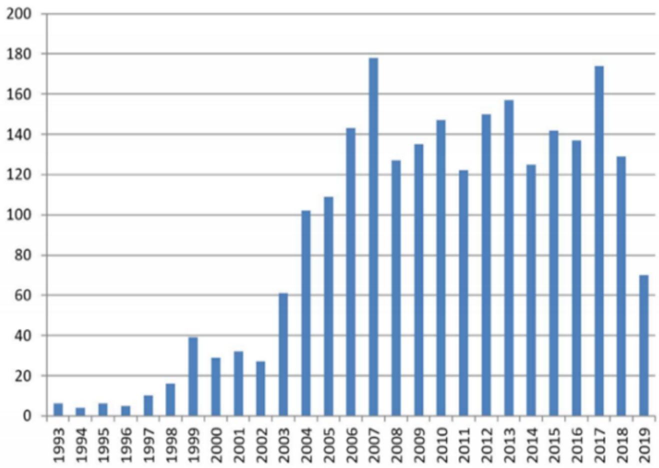

Figure.1 Changes in the number of nuclear security-related incidents reported to ITDB [1]

METHOD

In this section, the proposed approach to detect thefts disguised as normal operations is detailed. First, since there are a variety of normal operations at the nuclear facilities, we propose to increase the reliability of theft detection by combining object detection and pose estimation.

Subsequently, by performing temporal sequence analysis of unit actions, we can identify the theft.

I. Definition of unit actions

As an example, the action of opening a safety box to remove nuclear or radioactive material may lead to theft, but if the material is returned to the safety box in next step, it should not be identified as theft. To distinguish such differences, it is necessary to define each action as a 'unit action'. We label the actions and objects related to identifying theft as 'a, b, c…' and '1, 2, 3…', respectively (Figure.2), which can be obtained by the deep learning-based motion recognition model and object detection model. Additionally, we detect persons and extract two-dimensional joint positions of the persons in the image using the human pose estimation model. Since all of the currently assumed actions with labels 'a, b, c…' are performed with hands, we calculate the distances between the coordinates of both hands and objects and identify their relationships based on the distance threshold. The models deployed in this study are listed as follows:

- Pose estimation model: OpenPose [3], which it is capable of detecting 2D key points of multiple human bodies in real time.
- Human motion recognition model: Long-Short Term Memory (LSTM) [4], which provides appropriate output for long-term temporal sequence, such as the skeletal coordinates of a moving person.
- Object detection model: YOLOv3 [5], which achieves well trade-off between speed and accuracy on a multi-class object detection task.

Finally, we integrate the labels of the related actions and objects into codes (e.g., 'a1', 'b2', see Figure.3) frame by frame to obtain temporal sequence of the unit actions.

a. Open
b. Close
c. Put
d. Have
e. Steal
f. Use

1. Bottle
2. Scissor
3. Driver
4. Safety Box

A. By Safety Box

Figure.2 Examples of action label / object label

Figure.3 Examples of codes indicating unit actions

Ⅱ. Theft identification based on temporal sequence analysis of unit actions

The obtained coded unit action data is constantly compared with the predetermined sequences to identify theft. Once a match is confirmed, then the theft is identified. As mentioned earlier, the detection of nuclear material theft is based on the temporal sequence of unit actions, which varies depending on the person who theft the material. Based on these considerations, a predetermined sequence to determine theft is determined according to the following two candidate decision rules.

1. Detect when specific unit actions appear in a specified order (including cases where there are unrelated actions in between)
2. Confirm whether a unit action that cancels out the possibility of theft appears when a certain unit action has possibility of theft raises.

RESULT

1. Object detection

The number of training data sets for YOLOv3 is shown in Table 1. The shape, center position, width and height information of the bounding box are shown in Figure 4.

Table.1 Number of training datasets for YOLOv3

	Web	Real World	Total
scissor	1250	150	1400
driver	1650	150	1800
bottle	2500	200	2700

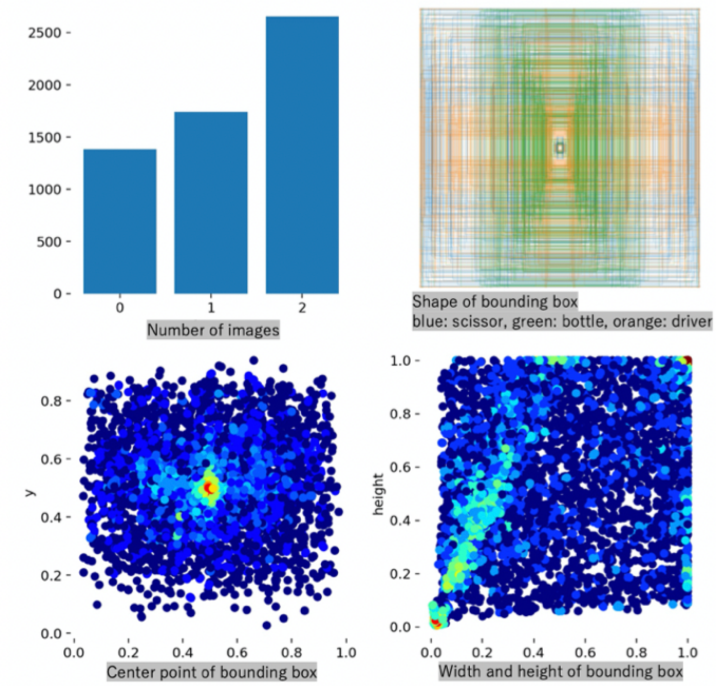

Figure.4 Number of images (upper left), bounding box shape (upper right), bounding box position (lower left), bounding box width and height (lower right)

Implementation details of the object detection model is shown in Table.2.

Table.2 Implementation details of the object detection

Model	YOLOv3
Framework	PyTorch 1.7.1
Batch Size	8
Optimizer	SGD
Learning Rate	0.01
Epoch	300
Platform	Intel Core i7 7820X NVIDIA GeForce GTX 1080Ti/ 11GB Ubuntu 18.04

We train the model for 300 epochs using SGD optimizer with initialized learning rate of 0.01. 95% of the data was used for training and 5% for validating. The training results for object detection are shown in Figure 5. The upper part of this figure shows the training results and the lower part shows the validation results. The mAP@0.5 can be achieved around 0.9.

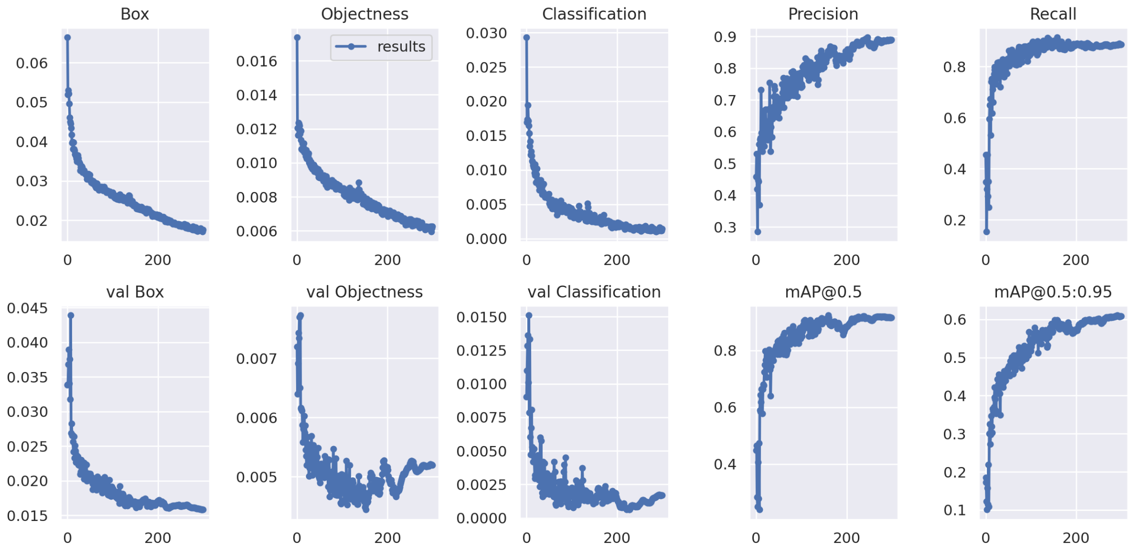

Figure.5 Training and validation results of the object detection model

An example of the results of object detection is shown in Figure 6. It can be seen that even such a small object can be detected with the correct label. However, the certainty is low depending on the angle, and is insufficient in most scenes at about 0.4-0.5.

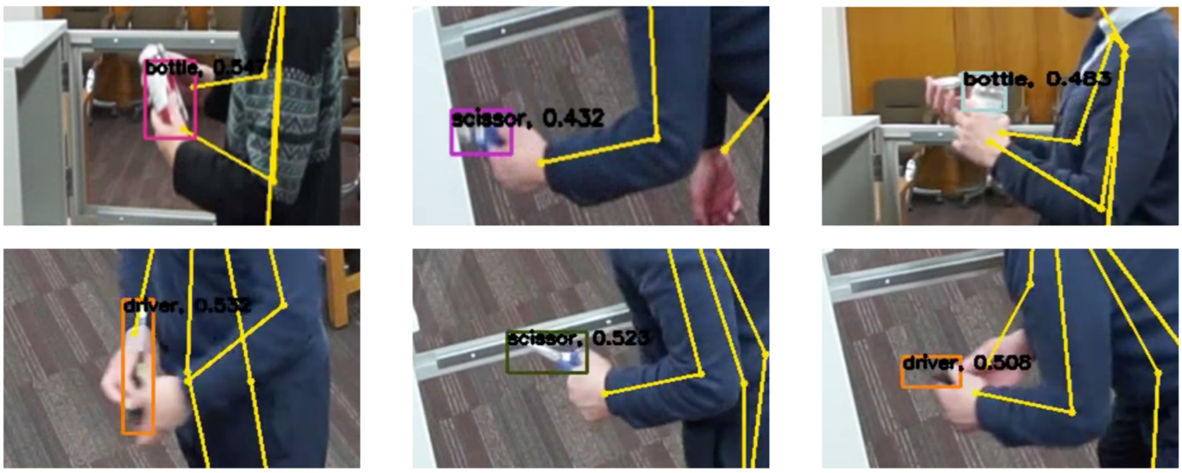

Figure.6 Example of object recognition results

2. Pose Estimation

Figure 7 shows the pose estimation results obtained by OpenPose. The skeleton of each movement can be roughly estimated.

Figure.7 Examples of pose estimation results

3. Human motion recognition

In this study, only the actions of opening and closing doors were collected from STAIR Actions [6] as video data of human actions to be used as training data for the human motion recognition model. The summary of the video data is shown in Table 3 and Table 4.

Table.3 The summary of the video data of opening doors

number	Short	Medium	Long
Backward	26	29	14
oblique backward	21	16	21
Lateral	19	45	13

Table.4 The summary of the video data of closing doors

number	Short	Medium	Long
Backward	15	13	6
oblique backward	19	37	17
Lateral	21	37	26

An example of the temporal sequence data of human skeleton in the video is shown in Figure 8.

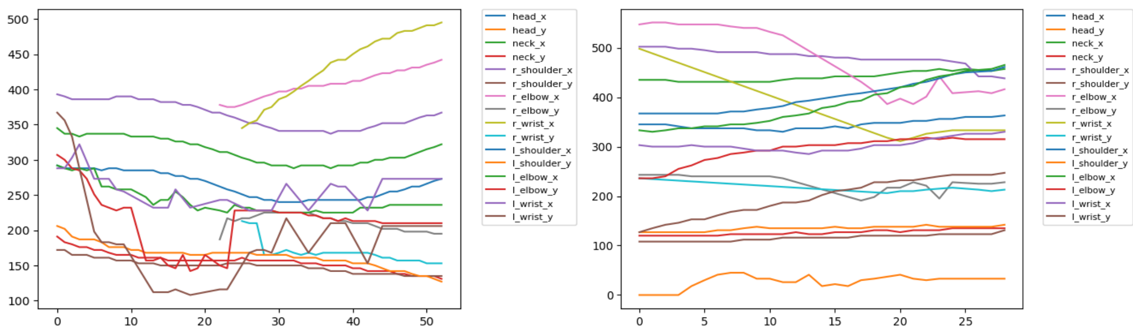

Figure.8 Examples of temporal sequence data of 2D skeletal coordinate (Left: door opening motion Right: door closing motion, vertical axis: pixels, horizontal axis: number of frames)

4. Relationship between pose and objects

As shown in Figure. 9 & 10, the relative distances between the object, the safety box, and both hands were obtained in real time. However, the loss of distance information due to the occlusion is reflected in the graph.

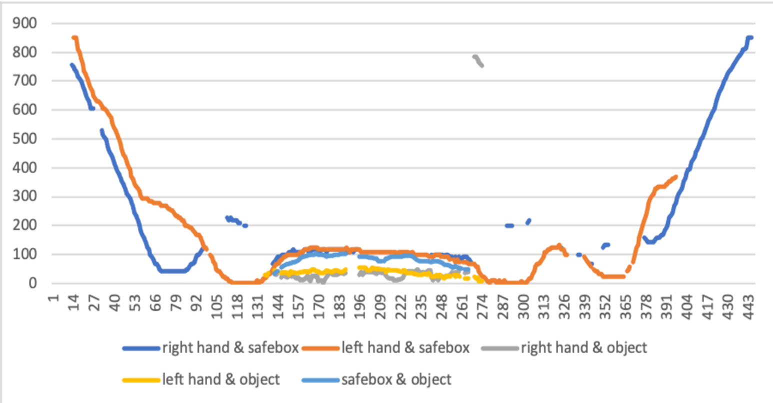

Figure.9 Temporal sequence of distance between bottle-hand-safety box when the bottle is taken out and put back (vertical axis: pixels, horizontal axis: frame)

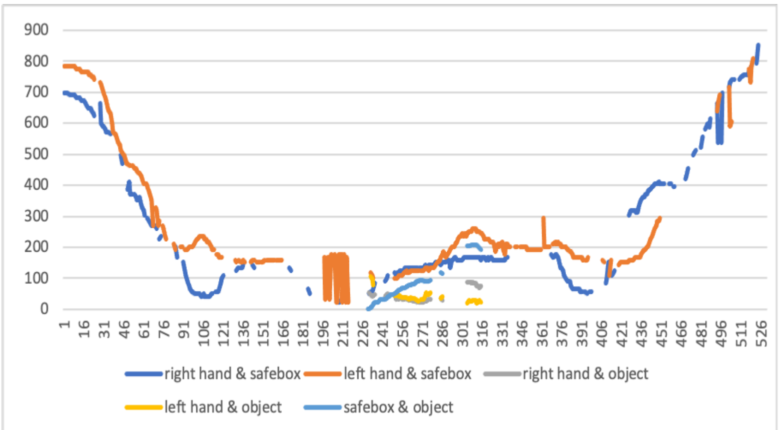

Figure.10 Temporal sequence of the distance between the bottle-hand- safety box when stealing the bottle (vertical axis: pixels, horizontal axis: frame)

Considerations based on these results are as follows. First, the lack of certainty in object detection is due to the high usage of images on the Web. only a few images on the Web are hand-held or taken from various angles. Therefore, there is a high degree of certainty when the object is facing forward, and a low degree of certainty when the object is held or tilted. it will be solved by using images that have been independently prepared for training. As for the fact that some incomplete graphs were observed in the time series data of 2D skeletal coordinates, this was due to overlook the case that there are more than two skeletal data on the screen when drawing the graphs. It will be solved by changing the program so that multiple skeletal data are output as separate temporal sequence data. The lack of distance information between the hand and the object is solved by interpolating the distance and supplementing the data with obstacles in the object and person skeletons in the video.

CONCLUSION

As a method to detect the moment of theft, we proposed following 2 elements for theft detection by sequence of unit actions.

1. The definition of unit actions
2. Method for theft identification based on temporal sequence analysis of unit actions.

In addition, detection model for unit actions were implemented by following procedure.

1. Implemented Pose Estimation model and Object Detection model
2. Proposed LSTM model as human motion recognition model

FUTURE WORKS

Implement and verify the sections that could not be reported as results this time.

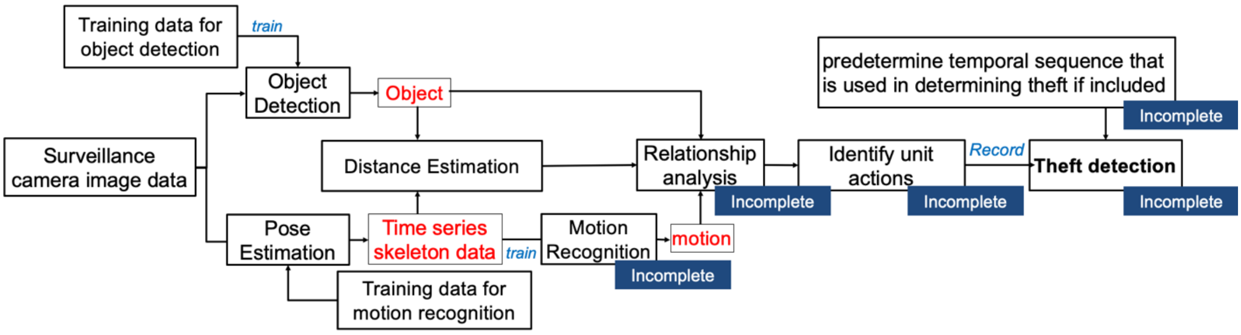

Figure.11 Future works summary

REFERENCES

[1] IAEA Incident and Trafficking Database: https://www.iaea.org/resources/databases/itdb

[2] VAAKEYE Factory DX: https://vaak.co/vaakeye-factory/

[3] Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh, Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, The Robotics Institute, Carnegie Mellon University, 14 Apr, (2017).

[4] Sepp Hochreiter, Fakultat fur Informatik Technische Universitat Munchen 80290 Munchen, Germany, Jurgen Schmidhuber, IDSIA Corso Elvezia 36 6900 Lugano Switzerland : LONG SHORT-TERM MEMORY,(1997).

[5] Joseph Redmon, Ali Farhadi, YOLOv3: An Incremental Improvement, University of Washington, 8 Apr, (2018).

[6] STAIR Actions: https://actions.stair.center