

## Robust individual pig tracking

Aggaluck Jaoukaew, Watcharapan Suwansantisuk, Pinit Kumhom

Department of Electronic and Telecommunication Engineering, King Mongkut's University of Technology Thonburi, Bangkok, Thailand

### Article Info

#### Article history:

Received Jun 23, 2023

Revised Jul 16, 2023

Accepted Jul 17, 2023

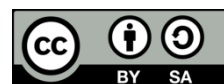
#### Keywords:

Multi-object tracking  
Performance evaluation  
Pig detection  
Pig localization  
Pig tracking

### ABSTRACT

The locations of pigs in the group housing enable activity monitoring and improve animal welfare. Vision-based methods for tracking individual pigs are noninvasive but have low tracking accuracy owing to long-term pig occlusion. In this study, we developed a vision-based method that accurately tracked individual pigs in group housing. We prepared and labeled datasets taken from an actual pig farm, trained a faster region-based convolutional neural network to recognize pigs' bodies and heads, and tracked individual pigs across video frames. To quantify the tracking performance, we compared the proposed method with the global optimization (GO) method with the cost function and the simple online and real-time tracking (SORT) method on four additional test datasets that we prepared, labeled, and made publicly available. The predictive model detects pigs' bodies accurately, with F1-scores of 0.75 to 1.00, on the four test datasets. The proposed method achieves the largest multi-object tracking accuracy (MOTA) values at 0.75, 0.98, and 1.00 for three test datasets. In the remaining dataset, the proposed method has the second-highest MOTA of 0.73. The proposed tracking method is robust to long-term occlusion, outperforms the competitive baselines in most datasets, and has practical utility in helping to track individual pigs accurately.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



### Corresponding Author:

Watcharapan Suwansantisuk

Department of Electronic and Telecommunication Engineering, King Mongkut's University of Technology Thonburi

126 Pracha Uthit Road, Thung Khru, Bangkok 10140, Thailand

Email: watcharapan.suw@kmutt.ac.th

## 1. INTRODUCTION

The global pig industry is valued at US\$ 254 billion in 2022 and is estimated to reach US\$ 418 billion by 2028 [1]. In commercial farming, pigs are raised in closed pens and are subjected to stress and illness. Pig locations across time can reveal the pigs' activities and well-being [2]–[8] and enable a farm to detect a disease early [9]–[14]. The accurate tracking of individual pigs benefits the billion-dollar industry.

The tracking of individual farm pigs is challenging. Pigs of similar size and age were raised in the same pen to manage their growth [10]. However, similar pigs are difficult to differentiate and track individually [3]. In addition, the outline shape of a pig changes according to the pig's activity. A dynamic shape complicates a vision-based tracker, which identifies an object as a pig based on its appearance. Finally, the trajectory of the pig was random. Two pigs that meet momentarily may depart in unpredictable directions and cause the tracker to switch the pigs' identities incorrectly. Individual pig tracking aims to accurately detect and track each pig over time. Detection accuracy is commonly measured by the false positive ratio (FPR), false negative ratio (FNR), precision, recall, and F1-score. Tracking accuracy is commonly measured by FPR, FNR, mostly tracked trajectory (MTT), number of identity switches (NIS), and

multi-object tracking accuracy (MOTA) [15]–[19]. To be accurate, a tracker must overcome appearance similarity, dynamic pig shape, and random pig movement.

Tracking individual pigs can be handled using the framework of multi-object tracking (MOT) [18]–[23], which is divided into invasive and noninvasive methods. Invasive methods implant a unique radio frequency identification (RFID) tag or place a specific mark on an animal, and install an RFID reader to interrogate tags or a video camera to capture the marks [24], [25]. For farms that place distinct ear tags on the pigs, a previous study [24] applies a convolutional network to recognize the tags and track individual pigs. RFID and marks increase the tracking accuracy and reduce identity switching. However, they disturb animals and become futile if tags or marks are detached from the animals [26]. Noninvasive methods require the installation of one or more video cameras in a pen to identify and localize individual pigs in each video frame and track the pigs over time [15], [27]–[29]. The locations of the pigs are relative to the pixel coordinates and, if needed, can be transformed into the physical coordinates of the pen. If the video processing and the tracking occur at a centralized unit, a large amount of video data can be compressed using convolutional neural networks (CNN) [30] or other techniques before transmission. Noninvasive methods minimize animal suffering and can be applied to large farms. Considering animal welfare, noninvasive methods are appealing.

Several variations of noninvasive methods exist for tracking pigs and other animals. In [15], a fingerprinting technique was proposed that mitigates identification (ID) switching in unmarked individual fish during their interaction and grouping. The fingerprinting technique uses both the position and direction of the fish to confirm its identity. The fish's position is extracted from the background, whereas the fish's direction is extracted from a style of motion including exclusion, occlusion, and interaction. The tracking method is appropriate for fishes and animals that occlude one another for a short duration, but is unsuitable for pigs, which engage in long-term occlusion. To cope with occlusion, the method [31] automatically and dynamically builds representations that enable the robust and effective tracking of animals. This method uses a Gaussian mixture model and expectation maximization for background subtraction and tracking, respectively. Images for this method must conform to a specific color model, namely the YUV color model, which separates the luminance component (Y) from the two chrominance components (U and V). The YUV color model with appropriate lighting control is required to minimize false negatives during background subtraction. Without lighting control, this method faces the problem of disconnected foreground blobs for a single object. More importantly, the method is efficient only when the animals have simple shapes and move in regular motion patterns. This method is unsuitable for pigs with complex shapes. To address illumination changes, long-term occlusion, and ID switching, a study [32] developed a collaborative tracking algorithm to track multiple objects in the presence of inter-occlusion using a color-based particle filter. The position of each detected object is represented by a blob, that is, a rectangular bounding box, which is used to build the tracker model. The distance between the blobs is used to determine the type of occlusion, including overlapping, partial occlusion, and full occlusion. An appropriate tracking method was selected based on the occlusion type. A drawback of this method is that it requires several complex algorithms to deal with multiple tasks. To develop a tracking method that is simple, and robust to illumination changes and long-term occlusions, machine learning approaches are used to detect animals or objects of interest in video frames. Common methods include faster region-based convolutional neural network (faster R-CNN) [33]–[38], you only look once (YOLO) [39]–[43], single-shot multi-box detector (SSD) [44]–[47], and feature pyramid network (FPN) [48]–[51]. After training with a dataset, these methods recognize the external appearance of an object or animal of interest, and predict the blob position, size, and boundary of each animal. In [43], YOLOX-S and YOLO v5s detectors have been used in detecting and classifying pig activities. In [36], the Faster R-CNN model was applied to automatically localize individual pigs from a side-view video and track them in a situation where, owing to a detection error, some pigs were lost from a video frame. Faster R-CNN can process frames at a speed of 0.2 sec per frame [36]. With this speed, faster R-CNN achieves near real-time detection and is attractive.

Although much work has been done on animal and pig tracking, existing work is fundamentally limited. The previous work constructed a rectangular bounding box around a pig and represented the pig's location by the centroid of the box. As two pigs are nearby and aligned diagonally, the centroid of a bounding box for one pig may be located on the body of another pig, leading to ID switching after the two pigs depart. A new method that resolves these limitations will increase tracking accuracy.

In this paper, we propose a method that tracks individual pigs and is robust to long-term occlusions. We prepared datasets and trained a faster R-CNN to recognize pigs in a farm environment. The model detects the pigs' bodies and heads in each video frame and differs from existing models that detect a combination of the body and head as one unit. The proposed method represents each pig's location using a bounding box around the pig's body, as opposed to a single centroid in the existing work. The enlarged representation of the location alleviates ID switching. We evaluated the performance of the proposed tracking method against state-of-the-art methods and found that the proposed method outperformed existing methods. The main contributions are the following:

- Labeled datasets, which are publicly available [52] to measure the accuracy of pig detection and tracking.
- An individual-pig tracking method, which is robust to a long-term pig-to-pig occlusion.
- Performance evaluation of the proposed method, showing superior performance compared to state-of-the-art methods.

The proposed method can detect and track pigs accurately, even for pigs in an environment different from that used in training the predictive model. These contributions may improve farm practices and pig welfare.

The remainder of this paper is organized: section 2 describes the problem statement, the dataset for training a pig-recognition model, and the proposed pig-tracking method. Section 3 describes the datasets for testing the model, evaluates the performance of the proposed method for detection and tracking, discusses the results, and suggests future research directions. Section 4 concludes the paper and summarizes important findings.

## 2. METHOD

The research method entails problem formulation (subsection 2.1), preparation of the training dataset (subsection 2.2), and design of the tracking method (subsections 2.3–2.5). As an overview, the proposed pig-tracking method identifies the head and body of each pig in video frames using a faster R-CNN model specifically trained for pig recognition. The proposed method matches the bounding boxes over the pigs' bodies across frames and repeats the matching for the bounding boxes over the heads of the pigs. Finally, it matches the body and head of the pig in each frame. Each part of the research method is described in detail below.

### 2.1. Problem statement

The system model is illustrated in Figure 1. The pen is closed, has a video camera attached to the ceiling, and contains  $N$  pigs of similar size and appearance. The video camera provides  $K$  frames of red-green-blue (RGB) images, denoted by  $F [1..K]$ , for pig tracking. Each frame contains all  $N$  pigs, possibly with pig-to-pig occlusions. The multi-tracking system takes video frames  $F [1..K]$  as input, detects, and tracks individual pigs in video frames. The output is a unique identifier,  $i$ , of each pig, where  $1 \leq i \leq N$ , and the bounding box with respect to the pixel coordinates in each frame of the  $i$ th identified pig. Our aim is to design a method to track individual pigs accurately.

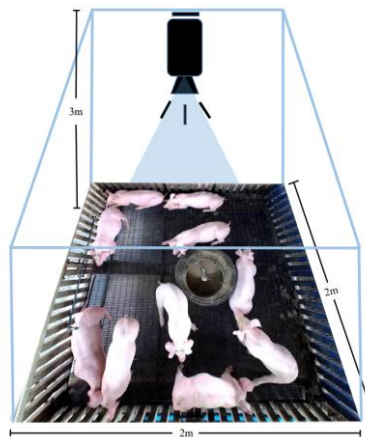


Figure 1. A system to track individual pigs takes a video as an input

### 2.2. Training a model

To train a model to recognize pigs' bodies and heads, we prepared a training video dataset. The videos were taken from a  $2 \times 2$  square-meter pen consisting of  $N=10$  pigs. The breed of the eight pigs was a mix of Landrace and Large White. The breed of the two pigs was a mix of Duroc and Hampshire. Each pig weighed 8-16 kg. Training videos were collected in the morning, noon, and afternoon with 696 video frames in total, covering different activities and lighting conditions.

To identify an object as a pig, we manually annotated the pig's head and body boundaries in the video frames using the Pascal Visual Object Classes Challenge 2007 as shown in Figure 2. Pigs that are heavily occluded are not annotated to improve boundary detection. The annotated head and body boundaries

assume a rectangular shape parallel to the image sides and cover the major head and body areas. For pigs that are oriented in roughly the same direction, the length and width of their boundaries are labeled as approximately the same size. This approach helps improve the accuracy of faster R-CNN boundary detection. In our prepared dataset, the total number of bounding boxes over either the head or body is approximately 6,500, a large quantity, coming from 10 pigs/image $\times$ 696 images, subtracted by the number of heads and the number of bodies, respectively, that are heavily occluded.

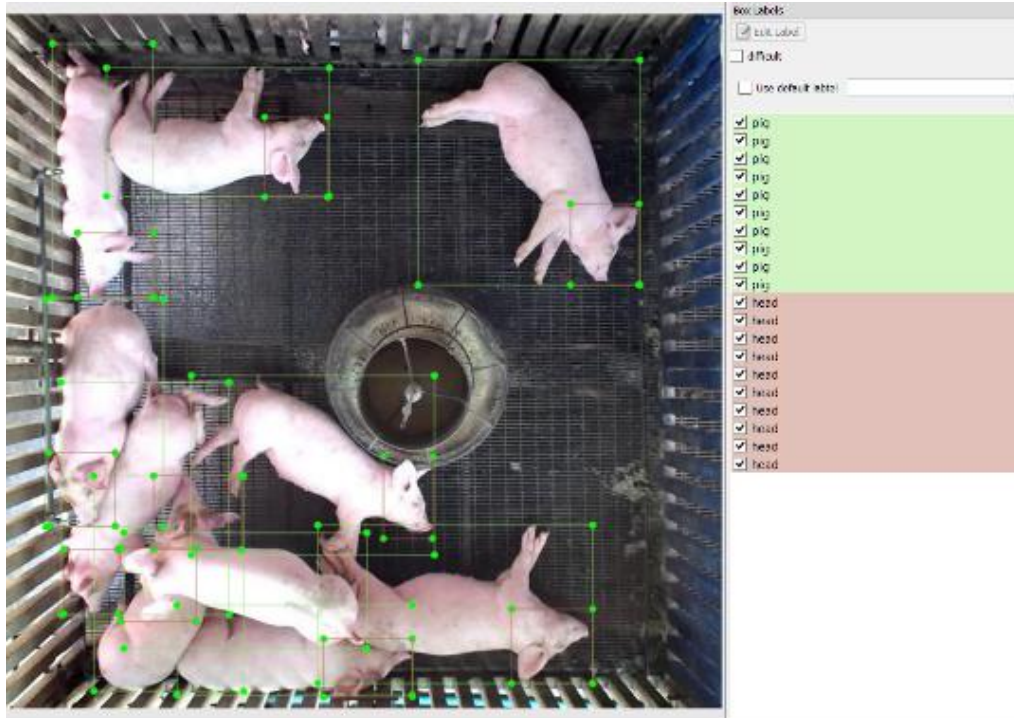


Figure 2. Video frames were labelled and used for training a neural network to recognize pigs' bodies and heads

Seventy percent of the 696 images were used for training the faster R-CNN model, while the 30% remaining images were used for testing the model. In the training stage, the learning rate was set to 0.0002 and the batch size was set to 1. The learning rate was small and appropriate because when pigs stayed in a group, their appearances were similar. The batch size can be increased to suit the computing power of the training machine. In the training and testing stages, the lost value obtained from 200,000 rounds of processing was used to determine the accuracy of the generated faster R-CNN model. We use appropriate parameters to train the model.

### 2.3. Tracking algorithm

The tracking algorithm is shown in Algorithm 1. The algorithm uses video frames  $F [1..K]$  and the number  $N$  of pigs as inputs, where  $F [k]$  denotes the image at the  $k$ th video frame for  $1 \leq k \leq K$ . The output was an array  $D[1..N, k_{st}..K]$  of structures, where  $D[n, k]$  contains information about the bounding boxes over the head and body of pig  $n$  in frame  $k$ . Line 1 applies the faster R-CNN model to predict the boundaries  $B_{body}$  and  $B_{head}$  around the pigs' bodies and heads, respectively. The variables  $B_{body} [1..M, 1..K]$ ,  $B_{head} [1..M, 1..K]$ ,  $n_{body}[1..K]$ , and  $n_{head}[1..K]$  are arrays, where  $M$  is the maximum number of detectable pigs in video frames. For a given frame  $1 \leq k \leq K$ , the variables  $n_{body}[k]$  and  $n_{head}[k]$  are the numbers of detected pigs' bodies and heads, respectively, in the  $k$ th video frame. Variable  $B_{body} [n, k]$  is the structure of the bounding boxes over the body of the  $n$ th detectable pig in the  $k$ th video frame, where  $1 \leq n \leq n_{body}[k]$ ; and variable  $B_{head} [n, k]$  is the analogous structure over the pig's head, where  $1 \leq n \leq n_{head}[k]$ . Structures  $B_{body} [n, k]$  and  $B_{head} [n, k]$  contain the points (xmin, ymin) in the upper-left corner, (xmax, ymax) in the lower-right corner, and (cx, cy) in the centroid of the bounding box. Line 2 determines the index  $k_{st}$  of the first video frame that contains  $N$  pigs' bodies and  $N$  pigs' heads. We consider that the number of frames is sufficiently large such that  $k_{st}$  exists.

**Algorithm 1. Individual pig tracking**

**Input:**  $F[1..K]$ ,  $N$   
**Output:**  $D[1..N, k_{st}..K]$   
1:  $(B_{body}, n^{body}, B_{head}, n^{head}) = \text{Faster-R-CNN}(F)$   
2:  $k_{st} = \text{the minimum index such that both } n^{body}[k_{st}] \text{ and } n^{head}[k_{st}] \text{ equal } N$   
3:  $(C_{body}, C_{head}) = \text{BlobRepairing}(B_{body}, n^{body}, B_{head}, n^{head}, k_{st})$   
4:  $\text{match} = \text{Body-Head Matching}(C_{body}, C_{head}, k_{st})$   
5: Initialize  $D[n, k]$  to contain information of the body  $C_{body}[n, k]$  and head  $C_{head}[\text{match}[n, k], k]$ , for each  $k_{st} \leq k \leq K$  and  $1 \leq n \leq N$

Next, the algorithm calls the method of blob repair, which maps the bodies of the same pig across different frames to the same identification number and maps the heads of the same pig across different frames to the same identification number. The output variables  $C_{body}[1..N, k_{st}..K]$  and  $C_{head}[1..N, k_{st}..K]$  are arrays of structures containing information about the bodies and heads, respectively, of  $N$  pigs in each frame, starting from the  $k_{st}$ th frame to the last frame. Fields of structures  $C_{body}[n, k]$  and  $C_{head}[n, k]$  are (xmin, ymin), (xmax, ymax), (cx, cy), and a new field id, which is the unique identification of the pig's body and head, respectively. For flexibility in blob repair, the unique identification of the pig's body may differ from that of the pig's head, even though they identify the same pig. This discrepancy will be resolved in the next step, line 4, which matches the identification number of the pig's body to the corresponding identification number of the pig's head for the same pig. Finally, line 5 consolidates the body and head of the same pig into the same unit and outputs array  $D[1..N, k_{st}..K]$  of structures. The structure  $D[n, k]$  contains the body and head positions in frame  $k$  of the pig, whose unique identification is  $n$ . This algorithm terminates and completes the tracking task.

**2.4. Blob repairing**

Body-blob repair aims to match a pig's body in a given frame with the corresponding pig's body in the next frame. Head-blob repair aims to perform analogous matching for pigs' heads. See Figure 3 for an illustration. The word "repairing" emphasizes the most important step in body-to-body and head-to-head matching: to repair a lost body blob, a lost head blob, an excessive body blob, and an excessing head blob. Blob repairing matches either the heads or the bodies across different frames.

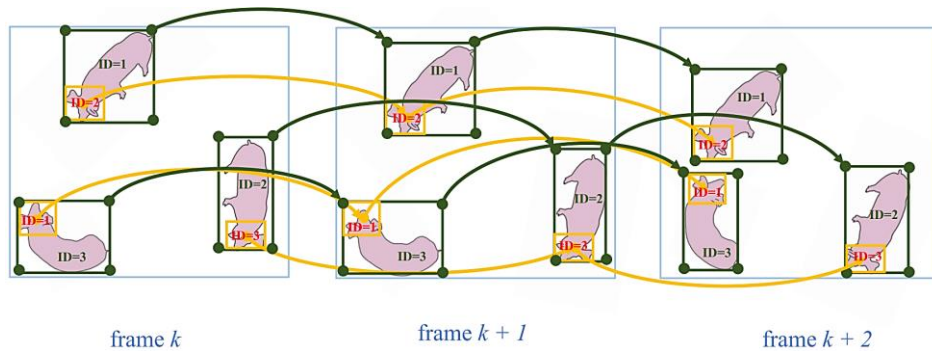


Figure 3. Blob repairing maps body and head blobs among consecutive frames, shown for  $N=3$  pigs. The lines indicate body or head blobs that belong to the same pig

The proposed method of blob repair appears in algorithm 2 and is based on the following ideas. First, the start frame, that is, the  $k_{st}$ th frame, has  $N$  head blobs and  $N$  body blobs by construction, and has already been repaired. Blob repairing progressively matches pigs' bodies and heads in a current frame,  $k$ , with the bodies and heads in the previous frame,  $k-1$ , for  $k_{st}+1 \leq k \leq K$ . Second, if the number of body blobs in the current frame does not equal  $N$ , pigs in the current frame must be heavily occluded so that the faster R-CNN model either fails to detect or excessively detects a pig's body. In this case, the body blobs in the current frame are unreliable. Blob repairing will equate the positions of the pigs' bodies and heads in the current frame to those in the previous frame. See lines 3 and 4 of Algorithm 2. Third, if the number of body blobs in the current frame is  $N$ , the body blob in the current frame is matched to the nearest body blob in the previous frame. In line 8 of the algorithm, the notion of "nearest" is measured by the average displacement between the bounding-box coordinates:

$$f(n, j, k) = \frac{1}{4} \sum_s (B_{body}[n, k].s - B_{body}[j, k-1].s) \quad (1)$$

where the summation index  $s$  covers fields  $s \in \{x_{min}, y_{min}, x_{max}, y_{max}\}$ . In the algorithm, the set of unmatched indices is stored in variable  $U$ , which initially equals a full set  $\{1, 2, 3 \dots, N\}$  and is reduced by one element at a time to the empty set-in lines 6-12. Fourth, the number of detectable pigs' heads in the current frame may differ from that in the previous frame. Blob repairing matches pigs' heads in the previous frame with the nearest ones in the current frame. Here, the notation of "nearest" for pigs' heads is the distance between the centroids of the bounding boxes (line 15 of the algorithm).

$$g(n, j, k) = \sqrt{\sum_t (B_{head}[n, k].t - B_{head}[j, k-1].t)^2} \quad (2)$$

The summation index  $t$  covers fields  $t \in \{cx, cy\}$ . Using different measures  $f$  and  $g$  for the body and head blobs is appropriate because the pig's body may be rotated within the blob, making the centroid an unfit representation of the body's blob position. In addition, the pigs' heads are smaller than their bodies and will not rotate much within the blob, making a centroid a suitable choice to capture the head position. Fifth, if the number of detectable pigs' heads in the current frame is smaller than that of the previous frame, the unmatched head blobs and their identities in the previous frame are copied to the current frame, as shown in lines 20-24. On the other hand, if the number of detectable pigs' heads in the current frame is larger than that of the previous frame, the unmatched head blobs and their identities in the current frame are discarded. The outputs of the algorithms are arrays  $C_{body} [1..N, 1..K]$  and  $C_{head} [1..N, 1..K]$ , where  $C_{body} [n, k]$  and  $C_{head} [n, k]$  are the structures of the  $n$ th blob for the pig's body and head, respectively, at the  $k$ th frame, where  $1 \leq n \leq N$  and  $k_{st} \leq k \leq K$ .

#### Algorithm 2. Blob repairing

```

Input:  $B_{body}[1..M, 1..K]$ ,  $n^{body}[1..K]$ ,  $B_{head}[1..M, 1..K]$ ,  $n^{head}[1..K]$ ,  $k_{st}$ 
Output:  $C_{body}[1..N, k_{st}..K]$ ,  $C_{head}[1..N, k_{st}..K]$ 
1: Initialize  $C_{body}[n, k_{st}] = B_{body}[n, k_{st}]$ ,  $C_{body}[n, k_{st}].id = n$ ,  $C_{head}[n, k_{st}] = B_{body}[n, k_{st}]$ ,
   and  $C_{head}[n, k_{st}].id = n$ , for each  $1 \leq n \leq N$ 
2: for  $k = k_{st} + 1$  to  $K$  do
3:   if  $n^{body} = N$  then
4:     Initialize  $C_{body}[n, k] = C_{body}[n, k-1]$  and  $C_{head}[n, k] = C_{head}[n, k-1]$ , for each
        $1 \leq n \leq N$ 
5:   else
6:      $U = \{1, 2, 3, \dots, N\}$ 
7:     for  $n = 1$  to  $N$  do
8:        $J = \arg \min_{j \in U} f(i, j, k)$ 
9:        $C_{body}[n, k] = B_{body}[n, k]$ 
10:       $C_{body}[n, k].id = C_{body}[J, k-1].id$ 
11:       $U = U \setminus \{J\}$ 
12:    end for
13:     $V = \{1, 2, 3, \dots, n^{head}[k]\}$ 
14:    for  $n = 1$  to  $n^{head}[k-1]$  do
15:       $J = \arg \min_{j \in V} g(n, j, k)$ 
16:       $C_{head}[J, k] = B_{head}[J, k]$ 
17:       $C_{head}[J, k].id = C_{head}[n, k-1].id$ 
18:       $V = V \setminus \{J\}$ 
19:    end for
20:     $m = N - n^{head}[k]$ 
21:    if  $n^{head}[k] < N$  then
22:      Let  $j_1, j_2, \dots, j_m$  denote the elements of  $V$ 
23:       $C_{head}[n^{head}[k] + i, k] = C_{head}[j_i, k]$ , for each  $1 \leq i \leq m$ 
24:    end if
25:  end if
26: end for

```

### 2.5. Matching of pig's body to its head

The matching stage aims to match a body's blob to a head's blob in the same frame such that the same pig has its body and head paired together. The key idea in the proposed matching is to examine the intersection of union (IoU) as well as to minimize the distance between the head and body's blobs. The top three IoU values are used to find the head blob with the closest distance to the body blob. Next, we describe the algorithm used for the matching stage in detail.

The algorithm for the matching stage appears in algorithm 3 and is illustrated in Figure 4. The inputs are the index of the start frame  $k_{st}$ , and arrays  $C_{body}$  and  $C_{head}$  of structures, containing information about the body's blobs and head's blobs. The output is an array  $match [1..N, 1..K]$  of integers to indicate the matching. For each frame  $k$ , the value of  $match [n, k] = j$  indicates that the body blob  $C_{body} [n, k]$  and the head blob  $C_{head} [j, k]$  belong to the same pig. The algorithm matching the body's blobs and head's blobs in the same frame is

shown in lines 1–33. The index of the frame under matching is denoted by variable  $k$ , which runs from  $k_{st}$  to  $K$ , as shown in line 1. The inner loop of the algorithm iterates through variable  $i$ , which is the index of the body's blob and ranges from 1 to  $N$ , as shown in lines 2–32. In line 5, set  $U$  of unmatched head indices is initialized to equal a full set  $\{1, 2, 3, \dots, N\}$ . Lines 6–9 calculate the value of IoU from each pair of blobs of the body and head. The three largest IoUs, namely  $\text{IoU}[j_1] \geq \text{IoU}[j_2] \geq \text{IoU}[j_3]$ , were selected for further matching, as shown in line 10. If frame  $k$  is the start frame, the matching is based only on  $\text{IoU}[j_1]$  because the minimum distance to the previous frame is unavailable, as shown in lines 14–16. In line 14, the index of the already-matched head is removed from set  $U$ . For a subsequent frame  $k > k_{st}$ , the distances of the head's blobs in the current and previous frames are considered, as presented in lines 17–19. Figure 4(a) illustrates the pigs in the previous frame, where the bodies have already been matched to the heads. The distance between the head's blobs in the current and previous frames is calculated. Figure 4(b) illustrates the pigs in the current frame, where each pig's body is to be matched based on the distance with a head. The distances under consideration have three values stored in variables  $\text{dist}_1$ ,  $\text{dist}_2$  and  $\text{dist}_3$ , where

$$\text{dist}_\ell = \sqrt{\sum_u (C_{\text{head}}[j_{\text{last}}, k-1].u - C_{\text{head}}[j_\ell, k].u)^2} \quad (3)$$

for  $u \in \{\text{xmin}, \text{xmax}\}$  and  $\ell=1, 2, 3$ . All three values are further compared to determine the closest distance based on the three conditions presented in lines 21–30. In the first condition, if  $\text{dist}_1$  is smaller than both  $\text{dist}_2$  and  $\text{dist}_3$ , then index  $j_1$  is matched with the body's blob at index  $i$ , and the already-matched index  $j_1$  is removed from set  $U$ . The second and third conditions are for cases where  $\text{dist}_2$  and  $\text{dist}_3$ , respectively, are the minima among the three distances. Figure 4(c) is an example of a pig's head that best matches the pig's body under consideration. After the algorithm finds the best body-to-head matching, a single ID is assigned to the body and head blobs of the same pig, as shown in Figure 4(d). Overall, the algorithm processes each frame and maps the IDs of the body and head blobs.

### Algorithm 3. Body-head matching

**Input:**  $k_{st}$ ,  $C_{\text{head}}$ ,  $C_{\text{body}}$   
**Output:**  $\text{match}[1..N, 1..K]$  array of integers to indicate the index of the head that matches to the index of the body

```

1: for  $k=k_{st}$  to  $K$  do
2:   for  $n=1$  to  $N$  do
3:     Initialize  $\text{IoU}[j]=0$ , for each  $j=1, 2, 3, \dots, K$ 
4:      $\text{body}=\text{rectangle}$  with the corners  $(C_{\text{body}}[n, k].\text{xmin}, C_{\text{body}}[n, k].\text{ymin})$  and
       $(C_{\text{body}}[n, k].\text{xmax}, C_{\text{body}}[n, k].\text{ymax})$ 
5:      $U=\{1, 2, 3, \dots, N\}$ 
6:     for  $j=1$  to  $N$  do
7:        $\text{head}=\text{rectangle}$  with the corners at  $(C_{\text{head}}[j, k].\text{xmin}, C_{\text{head}}[j, k].\text{ymin})$ 
      and  $(C_{\text{head}}[j, k].\text{xmax}, C_{\text{head}}[j, k].\text{ymax})$ 
8:        $\text{IoU}[j]=\text{area}(\text{body} \cap \text{head})/\text{area}(\text{body} \cup \text{head})$ 
9:     end for
10:    Pick the indices  $j_1, j_2, j_3$  such that  $j_1 \in U, j_2 \in U, j_3 \in U$ , and
       $\text{IoU}[j_1] \geq \text{IoU}[j_2] \geq \text{IoU}[j_3] > 0$  are the three largest positive IoUs
11:    if (cannot find such  $j_1, j_2, j_3$ ) then
12:      Unable to match the head-and-body for this frame. Go to line 1
      for the next value of  $k$ 
13:    end if
14:    if ( $k$  equals  $k_{st}$ ) then
15:       $\text{match}[n, k]=j_1$ 
16:       $U=U \setminus \{j_1\}$ 
17:    else
18:       $j_{\text{last}}=\text{match}[n, k-1]$ 
19:      Calculate  $\text{dist}_\ell$  according to (3) for  $\ell=1, 2, 3$ 
20:       $\text{dist}_{\min}=\min\{\text{dist}_\ell: \ell=1, 2, 3\}$ 
21:      if  $\text{dist}_{\min}=\text{dist}_1$  then
22:         $\text{match}[n, k]=j_1$ 
23:         $U=U \setminus \{j_1\}$ 
24:      else if  $\text{dist}_{\min}=\text{dist}_2$  then
25:         $\text{match}[n, k]=j_2$ 
26:         $U=U \setminus \{j_2\}$ 
27:      else
28:         $\text{match}[n, k]=j_3$ 
29:         $U=U \setminus \{j_3\}$ 
30:      end if
31:    end if
32:  end for
33: end for

```

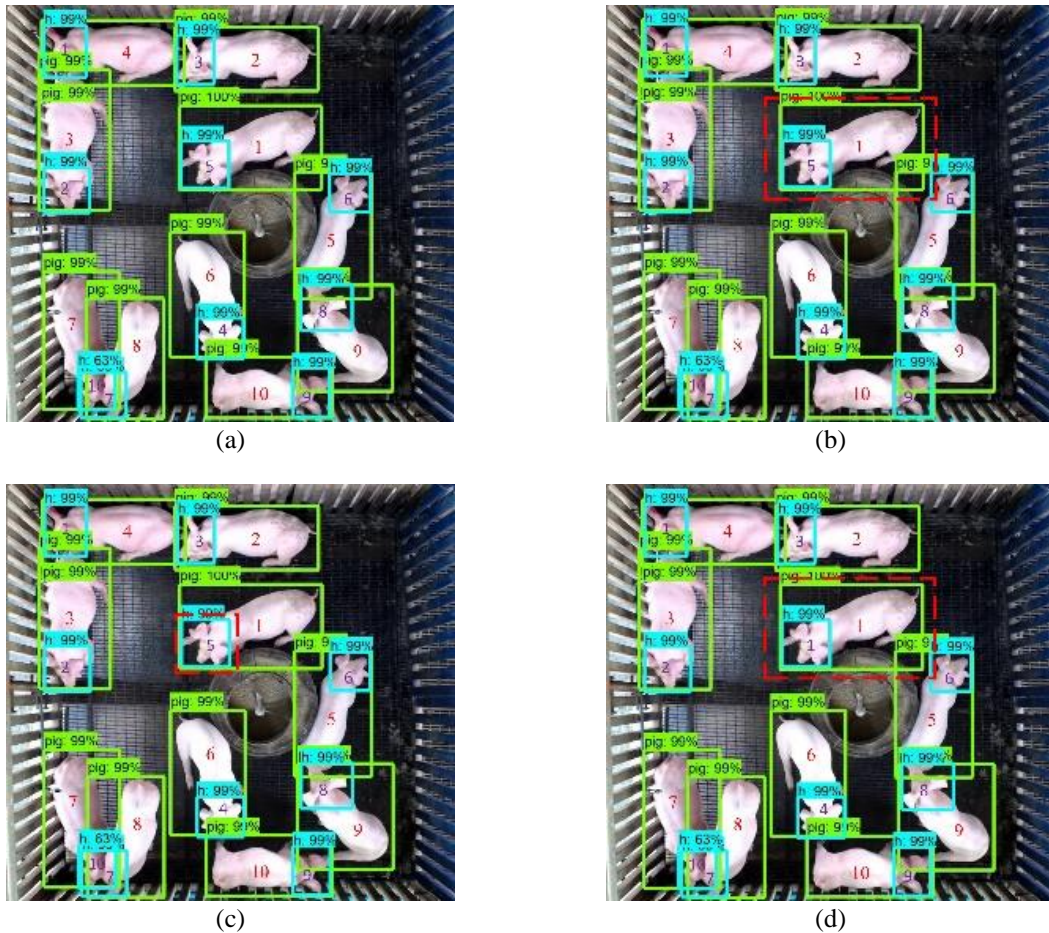


Figure 4. The matching stage (a) considers the previous frame, (b) selects each body bounding box in the current frame, (c) matches the body to a head, and (d) assigns the same ID to the matching head and body

### 3. RESULTS AND DISCUSSION

To evaluate the performance of the proposed tracking method, we took steps to ensure fairness. First, we prepared test datasets that were disjointed from the dataset used in section 2.2 for training the pig-detection model. Then, we evaluated the performance of the proposed detection model and the proposed pig tracking method on the test datasets.

#### 3.1. Test datasets

We prepared four datasets (Videos 1-4) to evaluate the pig identification and tracking methods. Each video contained 10 frames, down-sampled from 5,000 continuous frames, of pigs of the same number and mixed breeds. Videos 1-3 were captured in the morning, midday, and evening, respectively. The pig, pen, and camera setup in videos 1-3 are the same as those in the videos used in training the Faster R-CNN model (section 2.2). Video 4 has a different camera setup, pen, and pigs from the video used in training the Faster R-CNN model. Videos 1-3 were taken from an environment familiar to the pig detection model, while video 4 was not.

Table 1 lists the characteristics of the datasets used for the pig detection. A different time of the day leads to a different pig behavior and the occlusion ratio (OR), which measures the degree to which a pig's part overlaps with the same part of another pig. The head OR equals  $\frac{\sum_{k=1}^K d_k}{\sum_{k=1}^K n_k}$  where  $d_k$  is the number of pigs' heads that overlap with the heads of any other pig in frame  $k$ , and  $n_k=N$  is the total number of pigs in frame  $k$ . The body OR is defined similarly, but on the pigs' bodies. The higher the OR, the greater the overlap, and the more difficult it is to detect a pig. The body-to-body OR is largest at 0.40 on video 2, because at noon, pigs tend to lay down, rest, and cause body-to-body occlusion. Videos 1-4 test the abilities of the detection and tracking methods under various conditions. We made the test datasets available to other researchers [52].



Table 1. Test datasets consist of four videos of different characteristics (bbox=bounding box)

Dataset	Capturing Environment	Head		Body	
		# of bboxes	OR	# of bboxes	OR
VDO 1	Morning on familiar pen and pigs	100	0.04	100	0.26
VDO 2	Midday on familiar pen and pigs	100	0.13	100	0.40
VDO 3	Afternoon on familiar pen and pigs	100	0.12	100	0.24
VDO 4	Afternoon on unfamiliar pen and pigs	100	0.06	100	0.22

### 3.2. Performance metrics

We considered six performance metrics for pig detection: true positive ratio (TPR), FPR, FNR, precision, recall, and F1-score. In addition, we considered six performance metrics for pig tracking: TPR, FPR, FNR, MTT, NIS, and MOTA. The performance metrics for detection and tracking were obtained from the predicted bounding boxes and ground truth, without human intervention, for reproducibility. In this section, we describe the method to obtain these performance metrics.

During detection, in a given video frame  $f$ , the ground truth bounding boxes  $g_1, g_2, \dots, g_N$  are matched with the predicted bounding boxes  $p_1, p_2, \dots, p_n$ , where  $n$  is the number of predicted bounding boxes in frame  $f$ . The matching method is the greedy maximum-weight bipartite-graph matching, where the two disjoint sets of vertices are  $\{g_1, g_2, \dots, g_N\}$  and  $\{p_1, p_2, \dots, p_n\}$ . The weight  $w_{i,j}$  between vertices  $g_i$  and  $p_j$  favors, first, the IoU and, second, the centroid between bounding boxes. In particular,  $w_{i,j}$  is a tuple  $(a_{i,j}, b_{i,j})$  where  $a_{i,j}$  is the IoU between  $g_i$  and  $p_j$ , if the IoU is  $\geq 0.6$ ; and  $a_{i,j}$  is zero, otherwise. The element  $b_{i,j}$  is the distance between the centroids of the bounding boxes  $g_i$  and  $p_j$ . An IoU threshold of 0.6 is suitable, meaning that the two bounding boxes are significantly overlapped [36]. A comparison between the two weights begins with the IoU comparison and, in the case of a tie, is settled by the distance comparison. Following the greedy implementation, we repeatedly add an edge to the matching, starting from the maximum-weighted edge to a smaller-weighted edge, as long as the added edge preserves bipartite matching. The matching process establishes the pairs  $(g_i, p_{m(i)})$  between the ground truth and the predicted bounding box, where  $m(i)$  is the corresponding matched vertex. After the matching, true positive (TP), false positive (FP), FN are obtained from the IoU between each bounding-box pair  $(g_i, p_{m(i)})$ , using standard definitions of these metrics [53], where the IoU threshold is set to 0.6. TPR, FPR, and FNR are the ratios of TP, FP, and FN, respectively, to the number of ground truth bounding boxes, which is 100 as shown in Table 1. Maximum matching ensures that the TPR is at the largest possible value, and that the FPR and FNR are the smallest. The TPR, FPR, and FNR in the detection can serve as the ultimate limits for the analogous tracking metrics.

To evaluate the tracking performance, we rearranged the predicted pig IDs to match those in the ground truths; and then evaluated the TP, FP, FN, MTT, and NIS. The step to rearrange the predicted pig IDs ensures fairness because the tracker and ground truth may name the same pig by two different numbers consistently through the video frames. Matching occurs between the predicted bounding boxes and the ground truth at the earliest ground-truth video frame where the predicted model detects all  $N$  pigs. The matching is the greedy maximum-weight bipartite-graph matching with the same weight construction used by the detection metrics. Having  $N$  predicted pigs means that the matching is perfect: every pig's predicted ID is matched to a unique ground-truth pig ID. The predicted pig IDs were renamed to match the ground truth IDs.

After perfect matching, the ground truth bounding box  $g_i$  and predicted bounding  $p_i$  for the same index  $i$  are deemed to identify the same pig. Then, the TP, FP, and FN are obtained in each frame using the standard definitions of these metrics [53], with an IoU threshold of 0.6. In computing the MTT, the predicted trajectory of a given pig is considered mostly tracked if 80% of the predicted bounding boxes significantly overlap with the ground truths. Again, the two boxes overlap significantly if their IoU is  $\geq 0.6$  [36]. To obtain an NIS for a test video, we sum the NISs of the individual pigs. For example, Table 2 contains the predicted pig IDs, the NIS for each pig, and the NIS for video 1 and the proposed tracker. To obtain an NIS of a given pig, we match the pig's predicted bounding boxes to the ground-truth's bounding boxes in each ground-truth video frame, using the greedy maximum-weight bipartite-graph matching; and count the number of ID switches. Finally, MOTA is obtained from  $1-x$  [36], where  $x$  is the ratio between the sum of FP, FN, and NIS to the sum of the total ground truth bounding boxes and the maximum possible NIS value. The total number of ground-truth bounding boxes is  $MN$ , and the maximum possible NIS value is  $(M-1)N$ , where  $M=10$  is the total number of ground-truth frames as shown in Table 1. The method for computing the detection and tracking metrics is appropriate.

### 3.3. Detection evaluation

Figure 5 shows examples of bounding boxes that were generated by our model to detect the heads and bodies of pigs. Images of pigs in subfigures 5(a)–5(d) are taken from videos 1 to 4, respectively, and show a group of pigs in similar positions. A green bounding box is the detected pig's body, while a cyan box

is the detected pig's head. The proposed method accurately detects pigs' bodies even when the pigs are congregated in a group and experience pig-to-pig occlusion. Detecting a pig's head is more difficult than detecting a pig's body. In Figure 5(d), video 4 contains mistakes in head detection: two bounding boxes intended for pigs' heads appear at the buttocks; and one bounding box is missing from the pig's head. Visual inspection showed that the proposed pig detection model made few mistakes and performed well in these video frames.

Table 2. To obtain the NIS, predicted pig IDs are matched to ground-truth's, shown for video 1 and the proposed tracker. The ID of zero means that the tracker loses track of a pig

Ground-truth frame	Ground-truth pig ID									
	1	2	3	4	5	6	7	8	9	10
500	1	2	3	4	5	6	7	8	9	10
1,000	1	2	3	4	5	6	7	8	9	10
1,500	1	2	3	8	5	6	7	4	9	10
2,000	7	2	3	8	5	6	1	4	9	10
2,500	7	2	3	8	5	6	1	4	9	10
3,000	7	0	0	8	5	6	1	4	9	10
3,500	7	3	0	8	5	6	1	4	9	10
4,000	7	8	3	0	5	6	1	4	9	10
4,500	7	8	3	4	5	6	1	0	9	10
5,000	7	0	3	8	5	6	1	4	9	10
NIS for each pig	1	4	2	4	0	0	1	3	0	0
Sum=NIS	15									

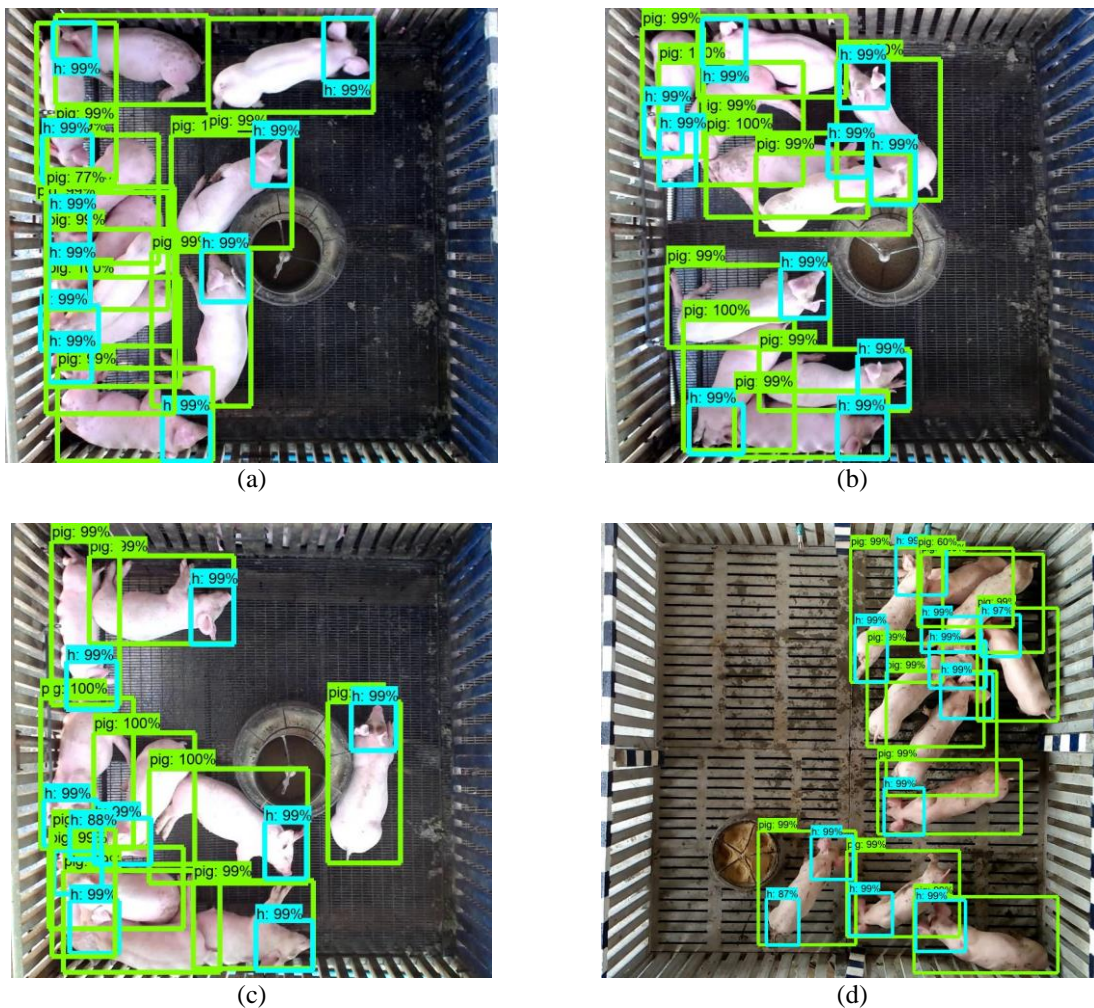


Figure 5. The bounding boxes cover the detected pigs' bodies and heads, shown as an example for frames taken from (a) VDO1, (b) VDO2, (c) VDO3, and (d) VDO4

Table 3 shows the overall detection performance of the pigs' heads and bodies. The ORs reported in Table 1 affect the FPR and FNR of pig detection. A high OR tends to lead to an error in boundary detection and, hence, a large FPR. The FNR is zero because the predictive model outputs  $N$  bounding boxes for heads and bodies, although some bounding boxes are not at the correction positions, and hence a non-zero FPR. The different lighting conditions in videos 1-3 did not significantly affect the F1-score, showing the robustness of the detection model to the lighting condition. The precision, recall, and F1-score for body detection are generally larger than those for head detection. For example, in video 3, the F1-scores are 0.99 on body detection and 0.83 on head detection. Furthermore, for unfamiliar video 4, the F1-score for body detection had the best value of 1.00, whereas the F1-score for head detection was 0.29. Bodies are larger and hence easier to detect than heads. The proposed algorithm appropriately uses the bodies of pigs for tracking.

Table 3. Detecting the pigs' bodies is more accurate than detecting the pigs' heads (Prec.=Precision)

Part	Dataset	TPR	FPR	FNR	Prec.	Recall	F1
Body	VDO1	0.94	0.06	0.00	0.94	1.00	0.97
	VDO2	0.95	0.05	0.00	0.95	1.00	0.97
	VDO3	0.99	0.01	0.00	0.99	1.00	0.99
	VDO4	1.00	0.00	0.00	1.00	1.00	1.00
Head	VDO1	0.78	0.22	0.00	0.78	1.00	0.88
	VDO2	0.72	0.28	0.00	0.72	1.00	0.84
	VDO3	0.71	0.29	0.00	0.71	1.00	0.83
	VDO4	0.17	0.83	0.00	0.17	1.00	0.29

Table 4. In tracking, the proposed method outperforms the state-of-the-art methods in most test datasets

Dataset	Method	TPR	FPR	FNR	MTT	NIS	MOTA
VDO1	Proposed	0.64	0.36	<b>0.00</b>	5	<b>15</b>	0.73
	SORT	<b>0.90</b>	<b>0.00</b>	0.10	<b>10</b>	20	<b>0.84</b>
	GO	0.55	0.45	<b>0.00</b>	4	23	0.66
VDO2	Proposed	<b>0.70</b>	0.30	<b>0.00</b>	<b>5</b>	<b>18</b>	<b>0.75</b>
	SORT	0.63	<b>0.27</b>	0.10	4	23	0.68
	GO	0.53	0.47	<b>0.00</b>	3	36	0.58
VDO3	Proposed	<b>0.99</b>	<b>0.01</b>	<b>0.00</b>	<b>10</b>	<b>2</b>	<b>0.98</b>
	SORT	0.92	0.08	<b>0.00</b>	8	3	0.94
	GO	0.77	0.23	<b>0.00</b>	6	22	0.78
VDO4	Proposed	<b>1.00</b>	<b>0.00</b>	<b>0.00</b>	<b>10</b>	<b>0</b>	<b>1.00</b>
	SORT	0.82	0.18	<b>0.00</b>	8	2	0.89
	GO	0.05	0.95	<b>0.00</b>	0	34	0.34

### 3.4. Tracking evaluation

We compared the proposed method with the state-of-the-art methods [17], [36] which track moving animals of a similar appearance without lighting control. The global optimization (GO) method in [17] is a deterministic method using the Hungarian algorithm, whereas the simple online and real-time tracking (SORT) method in [36] is a probabilistic method using a combined Hungarian algorithm and Kalman filter. The existing methods are competitive trackers.

The results of the individual pig tracking are shown in Table 4, where the best performance for each metric is indicated in bold. The FNRs of the proposed method and the GO method are zero for every test video, while the FNRs of the SORT method are 0.10 for videos 1-2 and 0.0 for videos 3-4. The proposed and GO methods track all  $N$  pigs in each video frame, while the SORT method fails to track 10% of the pigs, i.e., 10 out of the 100 pigs in total as shown in Table 1, in either video 1 or video 2. With a large FPR and a small FNR, the proposed method places correct bounding boxes on the areas where the pigs of the intended IDs are located. The TPR of the proposed methods are 0.70, 0.99, and 1.00, which are the largest values in test videos 2, 3, and 4, respectively. In Video 1, the TPR of the proposed method is 0.64, which is the second largest value after the TPR of 0.90, achieved by the SORT method. The overall tracking accuracy was captured by MOTA. The proposed method has the largest MOTAs in three out of four test videos and a perfect MOTA of 1.00 in video 4. The proposed method performs exceptionally well on video 4, which contains different pigs in a different pen from those used by the Faster R-CNN. The exceptional performance indicates that the proposed method is a robust pig tracker in an unfamiliar environment. The proposed method outperformed the state-of-the-art methods for most videos.

In the GO method, tracking pigs from their heads is inaccurate. As shown in Table 3, detecting a pig's head is more difficult than detecting a pig's body, which is confirmed by the low TPR of head detection. For example, in Video 4, the TPR for detecting the pigs' heads was 0.17. If the pigs must be both detected and tracked, the tracking TPR will not exceed 0.17. In Table 4, the tracking TPR of the GO method

was 0.05, which is  $\leq 0.17$ . The detection TPRs in Table 3 are the upper bounds of the tracking TPRs. As shown in Table 4, tracking pigs by their bodies, as in the proposed and SORT methods, generally outperforms tracking the pigs by their heads.

Figures 6 and 7 are examples of the trajectories obtained from fine manually labeled ground truth (dashed line) and the various trackers (solid lines) on selected pigs and videos. The x- and y-axes are the image x- and y-coordinates in pixels. The trajectories of the ground truth, proposed, and SORT methods are taken from the centroids of the pig's bodies. On the other hand, the trajectory of the GO method is from the centroids of the pig's heads. The proposed method tracks a pig accurately in Figures 6(a) and 7(a), as the ground truth's trajectories agree with the trajectories produced by the proposed method. A broken line on the trajectory indicates a loss of tracking, which contributes to a false negative (FN). A sharp transition on the trajectory indicates an unusual pig's movement and is caused by ID switches. The proposed method does not have a tracking loss on these exemplary trajectories. In contrast, the SORT method suffers from a tracking loss in Figures 6(b) and 7(b). The tracking loss of pig ID 4 in either Figure 6(b) or Figure 7(b) does not contradict the FNR of 0.00 in Table 4 for the SORT method on Videos 3-4, because the FNR in Table 4 is the average of the FNRs of  $N$  pigs and is rounded to two decimal places, due to the 100 available ground-truth frames in the test datasets. Furthermore, the GO method has a low tracking accuracy and produces a trajectory that is far from the ground truth. A discrepancy comes from ID switches, i.e., a different pig was tracked in Figure 6(c) (also observed in Figures 6(b) and 7(c)), and from an error in the positions of the head bounding boxes in Figure 7(c). A poor trajectory in Figure 7(c) is consistent with a large FPR of head detection on video 4 as shown in Table 3. Indeed in Table 4, the GO method has an MTT of zero; it does not track any trajectory correctly for any pig on 100 ground-truth frames as shown in Table 1. The proposed method is the most accurate tracker.

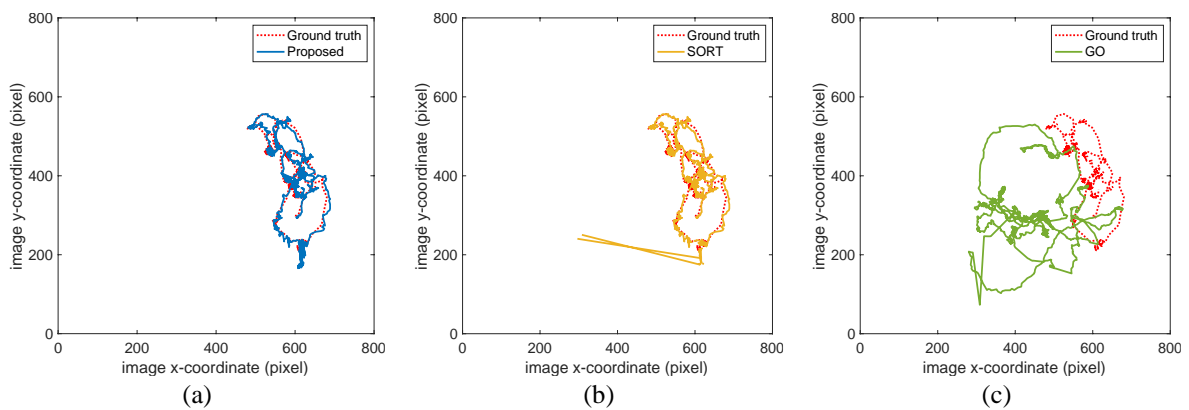


Figure 6. The trajectories of pig ID 6 on test Video 3 show that (a) the proposed method is most accurate than (b) the SORT and (c) GO methods

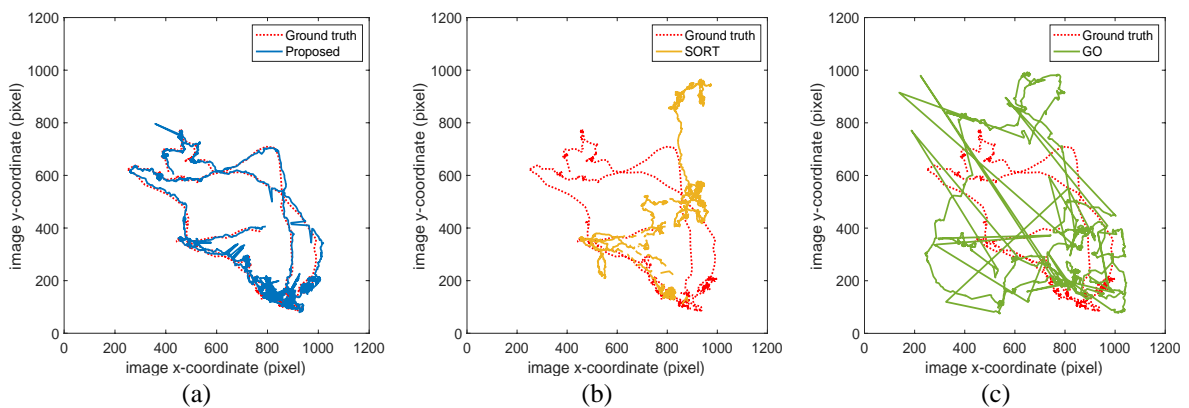


Figure 7. The trajectories of pig ID 4 on test video 4 show that (a) the proposed method is most accurate than (b) the SORT and (c) GO methods

Several factors play a role in increasing the tracking accuracy of the proposed method. The proposed method uses a rectangle to mark the boundary of each pig. In contrast, existing methods mark each pig by using its centroid alone. As a pig moves, its centroid may become closer to one another and cause an ID switch. A rectangular boundary captures a larger portion of the body and is more robust to ID switching. These advantages improved the tracking performance of the proposed method.

There are several directions to extend this research. The proposed method for pig tracking is sufficiently general to be applied to other animals, provided that the animal's head can be distinguished from its body. Future research can entail tracking other animals that have great economic, societal, or cultural importance. In addition, the process of head matching and body matching can be achieved simultaneously across several frames, as opposed to being done on each pair of adjacent frames. To reduce the complexity of the matching, a metaheuristic algorithm such as the giant trevally optimizer (GTO) [54] can be applied. Moreover, the number of animals in each pen was fixed and known in this study. Future research may cover unknown or changing numbers of animals. Finally, future research can use knowledge of the animal's tracked location as a feature to determine its activity. These future studies will extend the proposed tracking method to a broader context.

#### 4. CONCLUSION

Farm animals living in closed pens have high opportunities to fight or be injured. Individual pig tracking can increase animal welfare and provide a basis for behavioral monitoring and disorder diagnosis. However, individual pig tracking is difficult to achieve accurately because of the similarity in pig appearance and the tendency of pigs to remain in a group and create an occlusion. A method that can accurately track individual pigs has an advantage for animals and farm owners.

Using a top-view video, this study developed a method to track each pig in a realistic farm environment. To detect pigs in a given video, we created and labeled a dataset of pigs on an actual farm and trained a Faster R-CNN to recognize an object as a pig. The key idea in dataset preparation is to label only the visible pigs to increase detection accuracy. To evaluate the performance of the proposed detection method, we tested the model on separate videos and found that detection performance increased significantly when the pigs' bodies were used for identification. This finding matches the intuition that the pig's body occupies a large portion and serves to better identify a pig. The developed model detected pigs well across all test videos and stipulated that the body was the main feature for pig identification.

The proposed tracking method builds on the strength of the pig detection model and specific tracking ideas. The position of a pig in the next frame is difficult to forecast. To mitigate this difficulty, the proposed method detects both the head and body of each pig and uses only the frames in which the pig under consideration is detectable. Furthermore, to improve tracking accuracy, the proposed method uses a rectangle, as opposed to a centroid, to locate each pig. The proposed method is superior to state-of-the-art methods, namely, the SORT and GO methods. The identity of an individual pig can be tracked, even in the case of pig-to-pig occlusion. Given this advantage, the complete trajectory of each pig can be obtained and used for behavior monitoring, pig activity classification, and disease detection.

#### ACKNOWLEDGEMENTS

This work is supported in part by the Research Strengthening Project of the Faculty of Engineering, King Mongkut's University of Technology Thonburi.

#### REFERENCES




- [1] "Global pork market, size, global forecast 2023-2028, industry trends, growth, share, outlook impact of inflation, opportunity company analysis," *Research and Markets*, 2023. <https://www.researchandmarkets.com/report/pork> (accessed Apr. 07, 2023).
- [2] M. A. Kashiha *et al.*, "Automatic monitoring of pig locomotion using image analysis," *Livestock Science*, vol. 159, pp. 141–148, Jan. 2014, doi: 10.1016/j.livsci.2013.11.007.
- [3] S. Ott *et al.*, "Automated video analysis of pig activity at pen level highly correlates to human observations of behavioural activities," *Livestock Science*, vol. 160, pp. 132–137, Feb. 2014, doi: 10.1016/j.livsci.2013.12.011.
- [4] C. Munsterhjelm, M. Heinonen, and A. Valros, "Effects of clinical lameness and tail biting lesions on voluntary feed intake in growing pigs," *Livestock Science*, vol. 181, pp. 210–219, Nov. 2015, doi: 10.1016/j.livsci.2015.09.003.
- [5] S. G. Matthews, A. L. Miller, T. Plötz, and I. Kyriazakis, "Automated tracking to measure behavioural changes in pigs for health and welfare monitoring," *Scientific Reports*, vol. 7, no. 1, Dec. 2017, doi: 10.1038/s41598-017-17451-6.
- [6] R. B. D'Eath *et al.*, "Automatic early warning of tail biting in pigs: 3D cameras can detect lowered tail posture before an outbreak," *Plos One*, vol. 13, no. 4, Apr. 2018, doi: 10.1371/journal.pone.0194524.
- [7] A. Alameer *et al.*, "Automated detection and quantification of contact behaviour in pigs using deep learning," *Biosystems Engineering*, vol. 224, pp. 118–130, Dec. 2022, doi: 10.1016/j.biosystemseng.2022.10.002.
- [8] F. Hakansson and D. B. Jensen, "Automatic monitoring and detection of tail-biting behavior in groups of pigs using video-based deep learning methods," *Frontiers in Veterinary Science*, vol. 9, Jan. 2023, doi: 10.3389/fvets.2022.1099347.

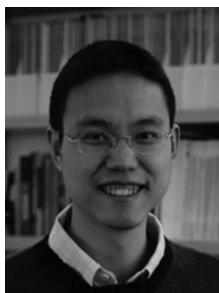
- [9] I. Kyriazakis and J. Houdijk, "Food intake and performance of pigs during health, disease and recovery," *Paradigms in pig science*, pp. 493–513, 2007.
- [10] E. Fernández-Carrión, M. Martínez-Avilés, B. Ivorra, B. Martínez-López, Á. M. Ramos, and J. M. Sánchez-Vizcaino, "Motion-based video monitoring for early detection of livestock diseases: The case of African swine fever," *Plos One*, vol. 12, no. 9, Sep. 2017, doi: 10.1371/journal.pone.0183793.
- [11] P. Statham, L. Green, M. Bichard, and M. Mendl, "Predicting tail-biting from behaviour of pigs prior to outbreaks," *Applied Animal Behaviour Science*, vol. 121, no. 3–4, pp. 157–164, Dec. 2009, doi: 10.1016/j.applanim.2009.09.011.
- [12] Y. Seddon, "Development of improved disease monitoring tools and management strategies to promote health in finishing pigs," Newcastle University, 2011.
- [13] M. H. Rostagno, S. D. Eicher, and D. C. Lay, "Immunological, physiological, and behavioral effects of salmonella enterica carriage and shedding in experimentally infected finishing pigs," *Foodborne Pathogens and Disease*, vol. 8, no. 5, pp. 623–630, May 2011, doi: 10.1089/fpd.2010.0735.
- [14] S. T. Ahmed, H.-S. Mun, H. Yoe, and C.-J. Yang, "Monitoring of behavior using a video-recording system for recognition of Salmonella infection in experimentally infected growing pigs," *Animal*, vol. 9, no. 1, pp. 115–121, 2015, doi: 10.1017/S1751731114002213.
- [15] A. Pérez-Escudero, J. Vicente-Page, R. C. Hinz, S. Arganda, and G. G. de Polavieja, "idTracker: tracking individuals in a group by automatic identification of unmarked animals," *Nature Methods*, vol. 11, no. 7, pp. 743–748, Jul. 2014, doi: 10.1038/nmeth.2994.
- [16] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, and T.-K. Kim, "Multiple object tracking: A literature review," *Artificial Intelligence*, vol. 293, Apr. 2021, doi: 10.1016/j.artint.2020.103448.
- [17] M. Wang, M. L. V Larsen, D. Liu, J. F. M. Winters, J.-L. Rault, and T. Norton, "Towards re-identification for long-term tracking of group housed pigs," *Biosystems Engineering*, vol. 222, pp. 71–81, Oct. 2022, doi: 10.1016/j.biosystemseng.2022.07.017.
- [18] L. Zhang, H. Gray, X. Ye, L. Collins, and N. Allinson, "Automatic individual pig detection and tracking in pig farms," *Sensors*, vol. 19, no. 5, Mar. 2019, doi: 10.3390/s19051188.
- [19] M. Wutke *et al.*, "Detecting animal contacts—a deep learning-based pig detection and tracking approach for the quantification of social contacts," *Sensors*, vol. 21, no. 22, Nov. 2021, doi: 10.3390/s21227512.
- [20] M. Kashiha *et al.*, "Automatic identification of marked pigs in a pen using image pattern recognition," *Computers and Electronics in Agriculture*, vol. 93, pp. 111–120, Apr. 2013, doi: 10.1016/j.compag.2013.01.013.
- [21] Q. Guo *et al.*, "Enhanced camera-based individual pig detection and tracking for smart pig farms," *Computers and Electronics in Agriculture*, vol. 211, Aug. 2023, doi: 10.1016/j.compag.2023.108009.
- [22] M. Mittek, E. T. Psota, J. D. Carlson, L. C. Pérez, T. Schmidt, and B. Mote, "Tracking of group-housed pigs using multi-ellipsoid expectation maximisation," *IET Computer Vision*, vol. 12, no. 2, pp. 121–128, Mar. 2018, doi: 10.1049/iet-cvi.2017.0085.
- [23] E. Psota, M. Mittek, L. Pérez, T. Schmidt, and B. Mote, "Multi-pig part detection and association with a fully-convolutional network," *Sensors*, vol. 19, no. 4, Feb. 2019, doi: 10.3390/s19040852.
- [24] E. T. Psota, T. Schmidt, B. Mote, and L. C. Pérez, "Long-term tracking of group-housed livestock using keypoint detection and MAP estimation for individual animal identification," *Sensors*, vol. 20, no. 13, Jun. 2020, doi: 10.3390/s20133670.
- [25] L. E. van der Zande, O. Guzhva, and T. B. Rodenburg, "Individual detection and tracking of group housed pigs in their home pen using computer vision," *Frontiers in Animal Science*, vol. 2, Apr. 2021, doi: 10.3389/fanim.2021.669312.
- [26] E. Leslie, M. Hernández-Jover, R. Newman, and P. Holyoake, "Assessment of acute pain experienced by piglets from ear tagging, ear notching and intraperitoneal injectable transponders," *Applied Animal Behaviour Science*, vol. 127, no. 3–4, pp. 86–95, Nov. 2010, doi: 10.1016/j.applanim.2010.09.006.
- [27] N. M. Lind, M. Vinther, R. P. Hemmingsen, and A. K. Hansen, "Validation of a digital video tracking system for recording pig locomotor behaviour," *Journal of Neuroscience Methods*, vol. 143, no. 2, pp. 123–132, 2005, doi: 10.1016/j.jneumeth.2004.09.019.
- [28] B. T. Morris and M. M. Trivedi, "A survey of vision-based trajectory learning and analysis for surveillance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1114–1127, Aug. 2008, doi: 10.1109/TCSVT.2008.9271109.
- [29] T. Fukunaga, S. Kubota, S. Oda, and W. Iwasaki, "GroupTracker: Video tracking system for multiple animals under severe occlusion," *Computational Biology and Chemistry*, vol. 57, pp. 39–45, Aug. 2015, doi: 10.1016/j.compbiolchem.2015.02.006.
- [30] H. T. Sadeeq, T. H. Hameed, A. S. Abdi, and A. N. Abdulfatah, "Image compression using neural networks: a review," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 17, no. 14, pp. 135–153, Dec. 2021, doi: 10.3991/ijoe.v17i14.26059.
- [31] V. Papadourakis and A. Argyros, "Multiple objects tracking in the presence of long-term occlusions," *Computer Vision and Image Understanding*, vol. 114, no. 7, pp. 835–846, Jul. 2010, doi: 10.1016/j.cviu.2010.02.003.
- [32] J. Xiao and M. Oussalah, "Collaborative tracking for multiple objects in the presence of inter-occlusions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 2, pp. 304–318, Feb. 2016, doi: 10.1109/TCSVT.2015.2406193.
- [33] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.
- [34] M.-C. Roh and J. Lee, "Refining faster-RCNN for accurate object detection," in *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, May 2017, pp. 514–517, doi: 10.23919/MVA.2017.7986913.
- [35] H. Jiang and E. Learned-Miller, "Face detection with the faster R-CNN," in *12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017)*, May 2017, pp. 650–657, doi: 10.1109/FG.2017.82.
- [36] J. Cowton, I. Kyriazakis, and J. Bacardit, "Automated individual pig localisation, tracking and behaviour metric extraction using deep learning," *IEEE Access*, vol. 7, pp. 108049–108060, 2019, doi: 10.1109/ACCESS.2019.2933060.
- [37] K. Srijakkot, I. Kanjanasurat, N. Wiriyakrieng, and C. Benjangkaprasert, "The comparison of Faster R-CNN and Atrous Faster R-CNN in different distance and light condition," *Journal of Physics: Conference Series*, vol. 1457, no. 1, Jan. 2020, doi: 10.1088/1742-6596/1457/1/012015.
- [38] D. Avola *et al.*, "MS-Faster R-CNN: multi-stream backbone for improved faster R-CNN object detection and aerial tracking from UAV images," *Remote Sensing*, vol. 13, no. 9, Apr. 2021, doi: 10.3390/rs13091670.
- [39] M. Ju *et al.*, "A kinect-based segmentation of touching-pigs for real-time monitoring," *Sensors*, vol. 18, no. 6, May 2018, doi: 10.3390/s18061746.
- [40] A. Alameer, I. Kyriazakis, and J. Bacardit, "Automated recognition of postures and drinking behaviour for the detection of compromised health in pigs," *Scientific Reports*, vol. 10, no. 1, Aug. 2020, doi: 10.1038/s41598-020-70688-6.
- [41] A. Bhujel, E. Arulmozhi, B.-E. Moon, and H.-T. Kim, "Deep-learning-based automatic monitoring of pigs' physico-temporal activities at different greenhouse gas concentrations," *Animals*, vol. 11, no. 11, Oct. 2021, doi: 10.3390/ani11113089.




- [42] M. Kim, Y. Choi, J. Lee, S. Sa, and H. Cho, "A deep learning-based approach for feeding behavior recognition of weanling pigs," *Journal of Animal Science and Technology*, vol. 63, no. 6, pp. 1453–1463, Nov. 2021, doi: 10.5187/jast.2021.e127.
- [43] S. Tu *et al.*, "Automated behavior recognition and tracking of group-housed pigs with an improved DeepSORT method," *Agriculture*, vol. 12, no. 11, Nov. 2022, doi: 10.3390/agriculture12111907.
- [44] W. Liu *et al.*, "SSD: single shot multibox detector," in *Computer Vision ECCV 2016*, Springer International Publishing, 2016, pp. 21–37.
- [45] N. Chengcheng, Z. Huajun, S. Yan, and T. Jinhui, "Inception single shot multibox detector for object detection," in *2017 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, Jul. 2017, pp. 549–554, doi: 10.1109/ICMEW.2017.8026312.
- [46] Y. Wang, C. Wang, and H. Zhang, "Combining a single shot multibox detector with transfer learning for ship detection using sentinel-1 SAR images," *Remote Sensing Letters*, vol. 9, no. 8, pp. 780–788, Aug. 2018, doi: 10.1080/2150704X.2018.1475770.
- [47] Y. Zheng and G. Wu, "Single shot multibox detector for urban plantation single tree detection and location with high-resolution remote sensing imagery," *Frontiers in Environmental Science*, vol. 9, Nov. 2021, doi: 10.3389/fenvs.2021.755587.
- [48] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 936–944, doi: 10.1109/CVPR.2017.106.
- [49] T. W. Cenggoro, A. H. Aslamiah, and A. Yunanto, "Feature pyramid networks for crowd counting," *Procedia Computer Science*, vol. 157, pp. 175–182, 2019, doi: 10.1016/j.procs.2019.08.155.
- [50] H. Wang and T. Wang, "Multi-scale residual aggregation feature pyramid network for object detection," *Electronics*, vol. 12, no. 1, Dec. 2022, doi: 10.3390/electronics12010093.
- [51] K. Min, G.-H. Lee, and S.-W. Lee, "Attentional feature pyramid network for small object detection," *Neural Networks*, vol. 155, pp. 439–450, Nov. 2022, doi: 10.1016/j.neunet.2022.08.029.
- [52] A. Jaoukaew, W. Suwansantisuk, and P. Kumhom, "Test datasets for individual pig tracking." 2023, Accessed: May 08, 2023. [Online]. Available: [https://github.com/aggaluck/GT\\_and\\_VDOs\\_Pigs](https://github.com/aggaluck/GT_and_VDOs_Pigs).
- [53] Z.-M. Qian, X. E. Cheng, and Y. Q. Chen, "Automatically detect and track multiple fish swimming in shallow water with frequent occlusion," *Plos One*, vol. 9, no. 9, Sep. 2014, doi: 10.1371/journal.pone.0106506.
- [54] H. T. Sadeeq and A. M. Abdulazeez, "Giant trevally optimizer (GTO): A novel metaheuristic algorithm for global optimization and challenging engineering problems," *IEEE Access*, vol. 10, pp. 121615–121640, 2022, doi: 10.1109/ACCESS.2022.3223388.

## BIOGRAPHIES OF AUTHORS






**Aggaluck Jaoukaew**    received a B.Sc. degree in physics from King Mongkut's University of Technology Thonburi (KMUTT), Thailand (2005); and an M.Eng. degree in electrical and information engineering from KMUTT, Thailand (2008). He is currently pursuing a doctoral degree in electrical and information engineering technology at the Department of Electronic and Telecommunication Engineering, KMUTT, Thailand. His main research interests include image processing, object detection, and behavior recognition. His general interests include machine learning, the Internet of things, and their applications in computer vision. He can be contacted at [aggaluck.hin@mail.kmutt.ac.th](mailto:aggaluck.hin@mail.kmutt.ac.th).



**Watcharapan Suwansantisuk**    received B.S. degrees in electrical and computer engineering and in computer science from Carnegie Mellon University, Pennsylvania, in 2002, and M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology in 2004 and 2012, respectively. He is currently an assistant professor at King Mongkut's University of Technology Thonburi (KMUTT), Thailand. Before joining KMUTT, he spent summers at the University of Bologna, Italy, as a visiting research scholar, and at the Alcatel-Lucent Bells Laboratory, NJ, as a research intern. His main research interests include wireless communications, synchronization, and statistical signal processing. Dr. Suwansantisuk serves on the technical program committees for various international conferences and served as the symposium co-chair for the IEEE Global Communications Conference in 2015. He received the Leonard G. Abraham Prize in the field of communications systems from the IEEE Communications Society in 2011, jointly with Prof. M. Chiani and Prof. M. Win, and the Best Paper Award from the IEEE RIVF International Conference on Computing and Communication Technologies in 2016, jointly with N. Chedoloh. He can be contacted at [watcharapan.suw@kmutt.ac.th](mailto:watcharapan.suw@kmutt.ac.th).



**Pinit Kumhom**    received a B.Eng. degree in electrical engineering from King Mongkut's Institute of Technology Thonburi, Thailand, in 1988, and a Ph.D. degree in electrical and computer engineering from Drexel University, Pennsylvania, in 2000. He is currently an assistant professor with the King Mongkut's University of Technology Thonburi. His research interests include the Internet of things and its applications, digital system design and implementation, and signal and image processing. He can be contacted at [pinit.kumhom@mail.kmutt.ac.th](mailto:pinit.kumhom@mail.kmutt.ac.th).