Warm Liquid Spill Detection and Tracking Using Thermal Imaging

Ghazal Rouhafzay
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, Canada
0000-0003-3762-0900

Haitao Tian
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, Canada
0000-0002-8464-7973

Pierre Payeur
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, Canada
0000-0003-3103-9752

***Abstract*—** Detection of liquid spill is a crucial and effective task to maintain safety and protection in various environments. Thermal imaging as a passive imaging modality working in different lighting conditions and even through smoke can be advantageously used to detect liquid spill in challenging conditions. Deep learning-based object detectors are well-established techniques to detect and localize different objects or phenomena in a variety of image modalities, however they require large scale databases with bounding box annotation in order to be trained from scratch. In this work, we present, evaluate, and compare three different methods to address the unavailability of substantial datasets dedicated to liquid spill detection from thermal images in the context of health and safety prevention. A Flir A35 thermal camera is used to collect data for the experiments. The three methods are based respectively on a conventional image processing algorithm using watershed segmentation, a weakly supervised approach using Gradient Class Activation Mapping, and an unsupervised deep learning approach for salient object detection guided by motion. No pixel level annotation is required for the proposed approaches. The work demonstrates that a conventional image processing approach, achieving an average precision and an average recall as high as 0.83 and 0.72 respectively, can reliably detect and localize warm liquid spill in sequences of thermal images.

Keywords—Thermal imaging, liquid spill detection, deep object detector, weakly supervised learning, unsupervised learning.

I. INTRODUCTION

Generally speaking, liquid spill can be an indicator of a serious threat in different situations and environments. Leakage in oil pipelines, water leakage from warm water units in a variety of industrial machines, and bleeding detection following an injury are some examples. Many research efforts are devoted to detecting liquid spill in a timely manner. While most of the existing leakage detection techniques rely on the use of sensors in direct contact with the liquid surface [1], noncontact sensing approaches such as image and video processing frameworks can effectively increase the operational efficiency.

Infrared thermal imaging cameras have found their ways in a variety of domains such as industrial inspection, medicine, security, protection, etc., to analyze scenes or particular objects. In thermal imaging, the surface temperature of objects within a scene is estimated by interpreting the intensity of infrared radiometric signals received at the camera sensor which is known as a microbolometer. Heat sensing elements in a microbolometer are sensitive to radiometry signals within a wavelength of 7 to 14 µm. The electrical resistance of the sensing elements changes in response to the incident radiometry signals. These resistance changes are then measured and mapped into temperature values.

While detection and segmentation of objects with solid shapes and unique temperature signatures within a scene is a well-established task in the field of computer vision for thermal imaging, a variety of situations impose more ambiguity on the operational conditions. Spilling liquid detection with no specific shape signature is one of these challenging cases, specially if the temperature of the liquid does not differentiate distinctively from other objects in the scene. A possible application could be early detection of blood on clothes or floor as a result of an injury. According to the fact that the emissivity of liquids in general is lower than the emissivity of a perfect emitter, the temperature reading using a thermal camera would not match the exact real temperature, adding uncertainty to the detection task. Often, when thermal cameras are used for temperature reading from surfaces with lower emissivity, a series of calculation are in place to compensate the effect of lower infrared emission captured by the camera. However, taking into consideration that for some monitoring and detection applications we do not have any prior knowledge about the existing materials in the scene, the proposed framework needs to rely on other characteristics of the sequence of captured frames rather than on pure thermography.

In this paper, we focus on the problem of warm liquid spill detection using thermal imaging. The case study targets the detection of injuries leading to blood spill. More specifically we investigate three solutions to detect bleeding as a result of an injury in thermal videos. The lack of any publicly accessible annotated dataset for liquid spill/bleeding detection, together with the challenges and laboriousness of creating a large dataset with pixel level annotation in such a sensitive context, motivated us to consider three different methods. One is based on conventional image processing algorithms, another one is using a weakly supervised deep learning approach, while the last method leverages an unsupervised deep network with saliency estimation driven by optical flow.

The rest of the paper is structured as follows. Section II discusses the data acquisition setup. Section III-A introduces the bleeding detection algorithm using conventional image processing approaches. The weakly supervised technique is detailed in section III-B. The working principle and setups for the unsupervised deep learning approach are explained in Section III-C. Section IV presents and evaluates the results and Section V concludes the paper.

II. Data Acquisition

In this research a Flir A35 thermal camera with a focal length of 9 mm is used to acquire thermal images. The camera streams 14-bit 320×256 radiometry data, with a maximum framerate of 60 Hz. The Temperature Linear Mode of the thermal camera is activated for data acquisition, therefore a signal to temperature mapping is performed as $T_k = 0.04 \times S$, where S is the 14-bit radiometry data and T_k is the temperature read at each pixel in kelvin. For data acquisition the framerate is reduced to three frames per second to simplify data processing.

To simulate the liquid spill/bleeding, a water bag filled with 37±2℃ water is used, and the participant is asked to pour water gradually on her clothes and on the floor.

III. Data Processing

A. Conventional segmentation based detection

The thermal signature of the monitored scene can be decomposed mainly into three levels to distinguish among the background, human body surface and regions on the body with higher temperature. The latter include the flow of warm liquid, the forehead, cheeks, and areas between joints, such as a flexed elbow, where heat can be trapped. For this purpose, a three level Otsu's thresholding algorithm is applied to divide the temperature spectrum of the scene into three clusters. Since these threshold values are determined in an offline phase with no object warmer than the person in the frame, the acquired threshold value remains acceptable for cases where objects with higher temperature are imaged, however a recalculation of the threshold will be required if the environment changes. Once the threshold values are determined, the framework illustrated in Figure 1 is implemented to perform detection and generate an alarm if warm liquid spill is detected.

For any acquired thermal image frame, noise is firstly reduced by median filtering over a neighborhood window of size 5x5 pixels. The previously mentioned threshold values then cluster the image into three main regions based on the temperature level. The clusters at this stage mainly mark the background, human body and warmer regions on the body surface. However, such an approach is not enough to perform segmentation because of the noisy nature of thermal images. The Sobel edge detection algorithm is then applied on the thermal map to find edges. We perform the main segmentation using the watershed algorithm [2] with the Sobel edge detection as the input image and the thermal thresholding map as the markers for the three regions. The watershed algorithm then gives a fine segmentation of the thermal image. Next step is to find an exact contour around the regions marked by the highest threshold.

Once the contours of the regions of interest are detected, a bounding box encountering the region in a rectangular shape is calculated. At this step, a series of false detections can be present. These false detections mainly include warmest regions of the body such as the face or the waterbag itself when it becomes visible to the camera. Elimination of false detections based on precise temperature reading is not practically feasible since the lower emissivity of water compared to human skin prevents to distinguish between them.

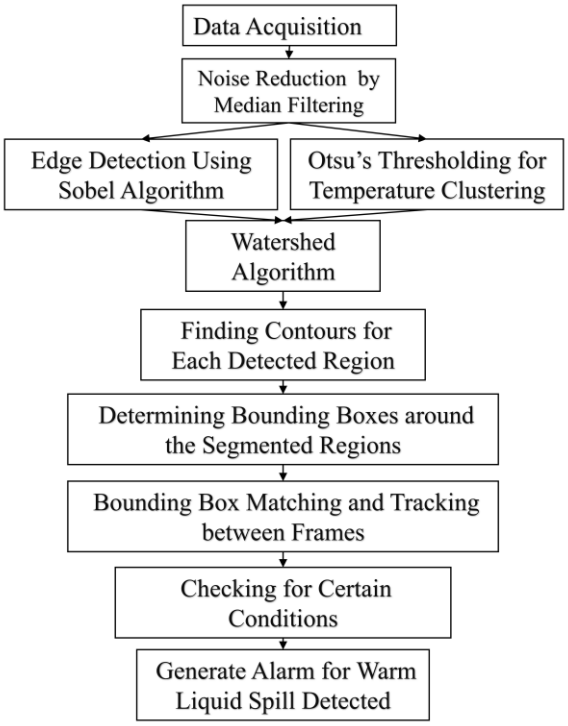

Fig. 1. Liquid spill detection framework using conventional image processing.

In order to remove false detections, a matching and tracking algorithm is proposed to examine the evolution of suspicious regions between the consecutive frames. The lack of any shape signature in liquids as well as the lack of color signature in thermal images make standard matching criteria useless. Therefore, for each bounding box in frame *i*, we search for a matching bounding box in frame *i-1* with shortest Euclidean distance between the centers of the two-bonding boxes.

The pseudo-code of the proposed algorithm for bounding box matching is as follows:

```
for all detected bounding boxes BB_j^i in current frame i
  find the center of the BB_j^i as C_BB_j^i
  find the size of the BB_j^i as S_BB_j^i
  find the Euclidean distances between all BB_j^i and BB_j'^(i-1)
  if  m ≤ n
    Match the bounding boxes with minimum distance between C_BB_j^i and C_BB_j'^(i-1)
  else
    find the difference between all S_BB_j^i and S_BB_j'^(i-1)
    Match the bounding boxes with minimum distance between C_BB_j^i and C_BB_j'^(i-1) and minimum differences between S_BB_j^i and S_BB_j'^(i-1)
    Mark the remaining BB_j^i as new detection.
```

where $0 \leq j' < n$ denotes the bounding box indices in frame *i-1*, and $0 \leq j < m$ represents the bounding boxes in frame *i*.

Once the detected bounding boxes are matched between successive frames, a series of conditions are checked to generate an alarm on warm liquid spill detection. These conditions are developed empirically to distinguish among normal and abnormal regions with similar temperature

signatures and rely on the fact that spilling liquid area grows over time. The conditions are as follows:

1- If a candidate region has an area smaller than 20 pixels, it is removed as it should possibly refer to small heat trapping between human joints.
2- If a candidate region has a bounding box with a height value twice longer than its width, and the center of the associated bounding box is positioned on the lower vertical half of the frame, a "warning" alarm is sent meaning that there is a possibility of liquid spill. This condition is determined using the prior knowledge about the position of the camera and considering the fact that liquids flow in the direction of gravity.
3- If a region is gradually increasing in size within 10 consecutive frames, a more powerful alarm is sent as "warm liquid spill detected".

B. *Weakly supervised detection*

In recent years deep learning has achieved a great success for analysis and processing of a variety of image modalities and for many different tasks such as classification, regression, detection, and semantic segmentation.

Most of the relevant approaches for detection and segmentation require large image datasets with either pixel level annotation or at least bounding boxes around the regions of interest. YOLO [3] and its newer versions represent state-of-the-art methods for object detection and give very promising results for different image modalities, however, the tedious nature of collecting and annotating a large dataset for training has encouraged us to consider alternative weakly supervised approaches. Unlike supervised object detection techniques that require bounding box annotations for training, weakly supervised object detectors [4] estimate the objects location by only image level labeling.

In this paper we take advantage of the Gradient Class Activation Mapping (Grad-CAM) [5] to roughly estimate the location of liquid spill in a thermal frame, if the frame is classified as containing liquid spill.

In order to train a Grad-CAM based liquid spill detector, a binary classifier is firstly trained by fine-tuning a pretrained Resnet 101 on a balanced dataset with 1200 frames without any liquid spill in it and 1200 frames containing liquid spill. These 2400 frames are sampled from video streams that are collected using the Flir A35 camera and annotated as either containing liquid spill or not. The binary classifier achieves an accuracy of 98.33% on a 80/20 split for training and validation. For evaluation and comparison of the detection results using the three methods in the paper, we employ a test dataset including 252 consecutive frames of a video sequence as detailed in section IV. The binary classifier reaches an accuracy of 99.2 on these 252 frames.

The idea of Grad-CAM is to compute the gradient of the final classification score of the wining class with respect to the final convolutional layer in the network. Given a test image, if it is classified as containing liquid spill by the binary classifier, the classification score will be used to compute gradients with respect to the last convolutional layer in Resnet 101. The gradient values are linearly combined with the activations from the last convolutional layer and resized back to the size of the original image to create a color map highlighting the class-specific regions in the input image. An example of a colormap generated by Grad-CAM and superposed on the input image is depicted in Figure 2.

The last step in this approach is to convert the class activation map into a binary image using a threshold (here 0.85) and computing a bounding box around it to mark the location of the detected liquid spill on the thermal image. The threshold value is determined empirically to maximize the performance of the detector. Finally, an alarm is set by visualizing the bounding box and a text message. Figure 2 summarizes the framework of liquid spill detection using Grad-CAM as a weakly supervised approach.

It is worth mentioning that the accuracy values reported in this section evaluate the performance of the backbone classifier for the binary classification task. An evaluation of the weakly supervised detector for the detection task comparing the detected bounding box against the ground truth will be provided in section IV.

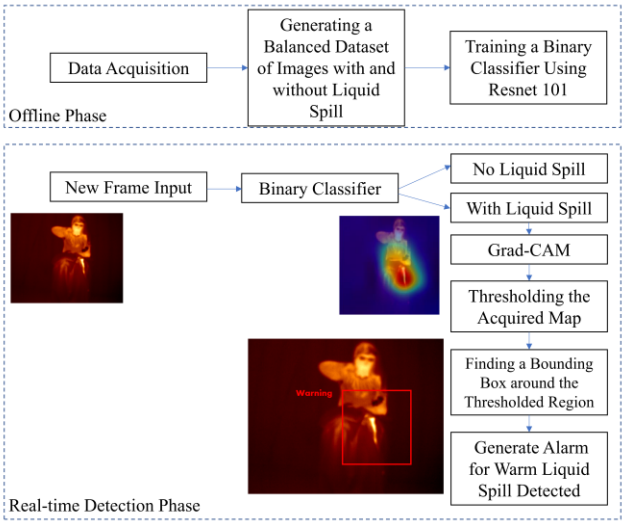

Fig. 2. Weakly supervised framework for liquid spill detection.

C. *Unsupervised detection based on motion guided attention*

Salient object/region detection in videos aims to identify and localize noteworthy regions that are of interest and of importance in consecutive video frames. Such an approach attempts to imitate the attention mechanism in human visual perception. This strategy is therefore considered in this paper for warm liquid spill detection on thermal frames where the regions containing liquid flow are considered salient from the point of view of an attention mechanism.

More specifically, we adopt the unsupervised saliency detection approach presented in [6] to highlight regions containing liquid flow. This is justified by the observation that the region corresponding to liquid spill across frames demonstrates significant and fast changes, which are dominant over other human body parts movement as well as the background. Figure 3 illustrates some examples of the saliency maps computed for thermal image frames.

With the saliency-aware maps of thermal frames, we define objectiveness bounding boxes denoting liquid flow regions with thresholding. It is worth noting that this thresholding process is built upon the saliency values of the

saliency-aware map, which is distinct from the conventional image thresholding strategy, which utilizes temperature-wise thresholding on the entire thermal frames. Since the saliency-aware representation is invariant to temperature changes but rather sensitive to the spread of liquid, the corresponding thresholding strategy is expected to be reliable.

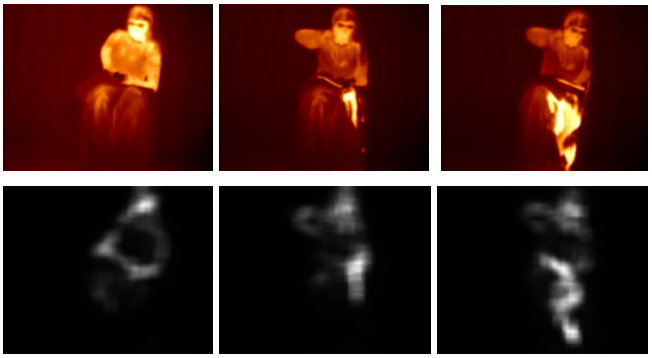

Fig. 3. Input thermal frames (upper row) and corresponding saliency-aware representations (lower row).

IV. Results and Discussions

In order to quantitively evaluate each of the proposed methods, we adopted the LabelImg [7] annotation tool to manually annotate bounding boxes around the spilling liquid regions for 252 test frames. As presented in Table I, five metrics are computed for each method with respect to the ground truth. They are computed as follows:

$$Average\ IoU = \frac{1}{n}\sum_{i=1}^{n}\frac{area(A\cap B)}{area(A\cup B)} \quad (1)$$

$$Average\ IoMin = \frac{1}{n}\sum_{i=1}^{n}\frac{area(A\cap B)}{min\,(areaA,\ areaB)} \quad (2)$$

$$Average\ Precision = \frac{1}{n}\sum_{i=1}^{n}\frac{TP}{TP+FP} \quad (3)$$

$$Average\ Recall = \frac{1}{n}\sum_{i=1}^{n}\frac{TP}{TP+FN} \quad (4)$$

where A represents the detected bounding box, B represents the ground truth bounding box, n is the number of test frames, i.e., 252, TP is the number of pixels in each frame that are correctly detected as liquid, FP is the number of pixels in each frame that are falsely detected as liquid, and FN represents the number of pixels in each frame that are missed to be detected as liquid. Also in Table I, the fifth metric is the number of missed frames that represents the number of frames in which liquid spill was present, but the algorithms failed to detect it.

While the weakly supervised approach shows a higher performance in terms of identifying that liquid spill exists in a thermal frame with fewer missed frames, the conventional segmentation-based approach performs better to localize the liquid spill region in the image. The higher average precision and higher average recall of the conventional approach confirm the superiority of this approach in minimizing both the false positives and false negatives. These results are justifiable by the fact that deep learning approaches call for large databases to be efficiently trained while the available dataset for this study is of a moderate size.

TABLE I. Evaluation Metrics

Evaluation Metric	Method		
	Conventional segmentation-based detection	***Weakly supervised detection***	***Unsupervised detection with motion attention***
Average IoU	0.6878	0.5626	0.0911
Average IoMin	0.8628	0.7138	0.2627
Average Precision	0.8307	0.6168	0.2256
Average Recall	0.7226	0.6785	0.1389
Number of missed frames	12	2	0

Figure 4 illustrates samples of detection performed using respectively the conventional, the weakly supervised, and the unsupervised saliency-aware approaches for 20 consecutive frames. The frames are selected in a way to represent the early start of the liquid spill until the flow has significantly expanded. The green bounding box highlights the ground truth. Detections by the conventional, the weakly supervised and the unsupervised approaches are depicted by blue, red and magenta bounding boxes respectively. One can notice that in the first two frames of the sequence the conventional approach fails to detect any liquid spill, which is due to the fact that the algorithm checks for the shape and the area conditions of the bounding box to trigger the alarm. Bounding boxes with an area smaller than 20 pixels are removed. Also, the conventional algorithm waits to find a growth in the area of the bounding box to set the alarm. These conditions were set to prevent false detection of small regions with constant area corresponding to heat traps between joints or the subject's face. The overall twelve missed frames correspond only to four seconds at the beginning of the thermal images sequence.

Conversely, the weakly supervised approach misses on detecting the liquid spill on only two out of the 252 frames. However, it underperforms the conventional segmentation approach in terms of localization of the liquid flow, progressively diverging from the region of interest as the spread of warm liquid expands. The unsupervised approach achieves the lowest performance for localizing the liquid flow over the entire test video sequence with the saliency being dragged toward warmer areas on the upper part of the body until the detection becomes more accurate when the flow of warm liquid spill increases, as was also demonstrated in Figure 3. The weak definition of saliency and the lack of consideration for physics-driven characteristics of liquid spill in the two deep learning methods are coherent with the experimental results.

V. Conclusion

In this research we propose and compare three approaches for warm liquid spill detection with a thermal camera in the context of health and safety prevention. The first approach makes use of the conventional watershed segmentation approach with Otsu's thresholding as the markers to segment regions within the temperature map of warm liquid, and then checks for the shape and area growth in consecutive frames to generate an alarm. The second approach relies on a Resnet 101 fine-tuned as a binary classifier on thermal images and Grad-

CAM to localize the liquid spill. The third approach is an unsupervised approach for detection of salient regions in video sequences that is adapted to detect expanding liquid spills. The conventional segmentation-based approach achieves an average precision and average recall of 0.83 and 0.72 and outperforms the weakly supervised and unsupervised approaches in terms of liquid spill localization. Such a conclusion is explained by the fact that deep learning requires large datasets to be efficiently trained and does not inherently encode physics laws in the intermediate representation of the detection stage. The work demonstrates a case where conventional image processing methods can prove superior to modern deep learning approaches and provide an advantage for applications where substantial and realistic image datasets are virtually impossible to acquire.

REFERENCES

[1] A. MacLean, J. McCormack, and B. Culshaw, "Distributed Sensing for Liquid Leaks and Spills," Proceedings, Volume 7677, *Fiber Optic Sensors and Applications VII*; 767703, 2010.

[2] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, PWS Publishing, Pacific Grove, CA, 1999.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 779–788, 2016.

[4] D. Zhang, J. Han, G. Cheng, and M.-H. Yang, "Weakly Supervised Object Localization and Detection: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[5] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-Octob, pp. 618–626, 2017.

[6] H. Li, G. Chen, G. Li, and Y. Yu, "Motion Guided Attention for Video Salient Object Detection," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019.

[7] Tzutalin. LabelImg. Git code (2015). "https://github.com/tzutalin/labelImg"

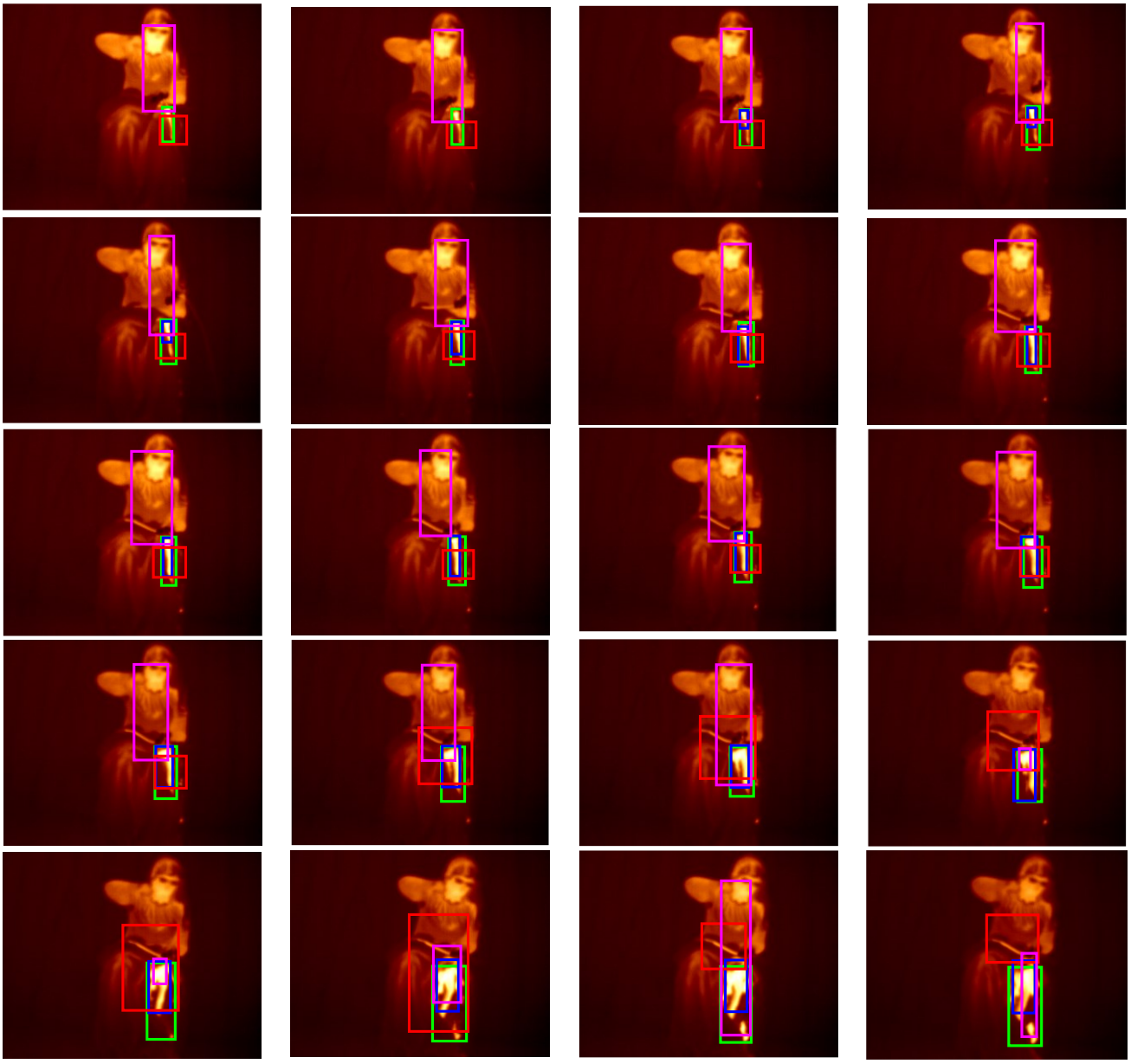

Fig. 4. Simulated bleeding detection results in consecutive frames where the ground truth is represented in green, the conventional image processing approach is highlighted in blue, the weakly supervised approach is highlighted in red, and the unsupervised approach is highlighted in magenta.