applied sciences

MDPI

Article

Anthropometric Ratios for Lower-Body Detection Based on Deep Learning and Traditional Methods

Jermphiphut Jaruenpunyasak [1], Alba García Seco de Herrera [2] and Rakkrit Duangsoithong [3,*]

[1] Department of Biomedical Sciences and Biomedical Engineering, Faculty of Medicine, Prince of Songkla University, Songkhla 90110, Thailand; jjermphi@medicine.psu.ac.th
[2] School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK; alba.garcia@essex.ac.uk
[3] Department of Electrical Engineering, Faculty of Engineering, Prince of Songkla University, Songkhla 90110, Thailand
* Correspondence: rakkrit.d@psu.ac.th

Abstract: Lower-body detection can be useful in many applications, such as the detection of falling and injuries during exercises. However, it can be challenging to detect the lower-body, especially under various lighting and occlusion conditions. This paper presents a novel lower-body detection framework using proposed anthropometric ratios and compares the performance of deep learning (convolutional neural networks and OpenPose) and traditional detection methods. According to the results, the proposed framework helps to successfully detect the accurate boundaries of the lower-body under various illumination and occlusion conditions for lower-limb monitoring. The proposed framework of anthropometric ratios combined with convolutional neural networks (A-CNNs) also achieves high accuracy (90.14%), while the combination of anthropometric ratios and traditional techniques (A-Traditional) for lower-body detection shows satisfactory performance with an averaged accuracy (74.81%). Although the accuracy of OpenPose (95.82%) is higher than the A-CNNs for lower-body detection, the A-CNNs provides lower complexity than the OpenPose, which is advantageous for lower-body detection and implementation on monitoring systems.

Keywords: anthropometric ratio; lower-body detection; deep learning; OpenPose

check for updates

Citation: Jaruenpunyasak, J.; Seco de Herrera, A.G.; Duangsoithong, R. Anthropometric Ratios for Lower-Body Detection Based on Deep Learning and Traditional Methods. *Appl. Sci.* **2022**, *12*, 2678. https://doi.org/10.3390/app12052678

Academic Editor: Antonio Fernandez-Caballero

Received: 5 November 2021
Accepted: 30 December 2021
Published: 4 March 2022

1. Introduction

For daily exercises, many people tend to focus dominantly or solely on cardiovascular exercises to burn calories. Lower-body strength, however, is also important not only for achieving perfect physical condition but also for maintaining total body health. Moreover, by strengthening the lower-body, one can improve one's agility and balance, helping to avoid falls and injuries during both daily activities and workouts. In addition, many studies [1,2] have found that lower-body strength and power are correlated and required for performing high-intensity, short-duration activities, such as jumping, sprinting, or carrying a load. Nevertheless, proper exercises should be performed to minimize the risk of injury from strengthening the lower-body. To prevent falls and injuries from these activities, the ability to detect the lower-body is crucial for monitoring the postures of participants during workouts.

With advances in computer technology, human body detection has become crucial in diverse applications such as surveillance systems, vehicle navigation, and posture recognition. Human body detection can also be applied to study human behavior and activities of daily living (ADLs) [3]. Thus, this human body detection can observe unusual signs in an activity sequence [4]. Moreover, it is a valuable indicator and threshold for monitoring systems in a workplace to identify inappropriate tasks and enhance injury prevention [5]. It is even used to automatically control home devices such as light sources and air conditioning to maintain suitable living conditions [6]. However, variations in human pose, clothing, and uncontrollable environmental characteristics can decrease the accuracy of human body detection [7].

There are two main types of methods for human body detection, namely, methods using wearable or non-wearable sensors. Methods of the first type rely on one or multiple sensors attached to the human body. These sensors can directly detect changes in a person's orientation. The main benefits of wearable sensors are their speed and high accuracy. They do not require specific environmental conditions, but they might need a long time to set up and cause inconvenience to people wearing them. Furthermore, sensor placement and dislocation can be problematic [8]. Another problem is associated to an absence of interoperability among a variety of sensor deployments [9]. Methods of the second type rely on non-wearable sensors consisting of one or multiple cameras to detect the human body. These methods need only low-cost equipment that is easy to set up to monitor the human body. However, without control over specific conditions such as illumination and shadows, their detection accuracy may be reduced [10–13]. If the distance between the camera and an individual is inappropriate, the image information might be distorted owing to motion blur [14].

Based on non-wearable sensors, there are two main popular classes of methods in computer vision for detecting people in an image [10–15]. The methods in the first class are referred to as traditional methods. A feature vector is extracted from the input image using techniques such as object-based approaches. Then, this feature vector is used to train a classification model. The second class of methods is based on deep learning [16]. In the last few years, deep learning has become a very popular machine learning approach for detecting objects and monitoring human poses and activities [17]. Deep learning has the ability to learn local and global features from an image by means of convolutional neural networks for human detection [16]. These features are also customized by using mapping, fusion, and selection techniques to significant success in human posture recognition [18].

In the case of occlusion [19], it might be either intra-class occlusion or inter-class occlusion. The intra-class occlusion happens when the interesting object is hidden by the same category of object such as crowded people. The inter-class occlusion refers to the object which is occluded by an object of another category such as the vehicle in pedestrian detection. Consequently, computer vision techniques might be troublesome for detection because the object to be detected may not appear similar to the objects in the training data set. To deal with the occlusion [20], it can be the generated de-occluded image using the generative adversarial network (GAN) to reconstruct the occluded object in the image. However, this method requires of a large various of object categories and labeling a dataset with feasible occlusions of every category. In term of occluded human image, if the computer vision techniques [10–12] can detect some parts of the body of human such as hand or face, these body parts may be possible to determine the position of the lower-body from the locations of the detected body parts by referencing anthropometric data [21].

Anthropometrics is the study of measurements of the human body taken from diverse populations such as age, gender or nationality. Applications of using the anthropometric consist of a suitable design of clothing [22] and workstation [23] from this human body reference. Anthropometric data involves human body measurements such as weight, height, and length. For human body detection, measurements of physical limb and body proportions can be used to investigate various associations with the height and width of the lower-body in an image [24,25]. Anthropometric data include age, gender, nationality and human body measurements such as weight, height, and girth. These data are useful for applications such as the suitable design of clothing, machines, and workspaces based on human body references. Moreover, These data can be beneficial to examine differences in the anthropometric characteristics and physical capacity in padel players concerning their competing level [26].

This paper presents novel anthropometric ratios that can be used in combination with both deep learning and traditional methods for lower-body detection. The proposed ratios can be applied to human images to detect either certain parts of the body or the full human body. The detected lower-body will then be indicated on the output image, as shown in Figure 1.

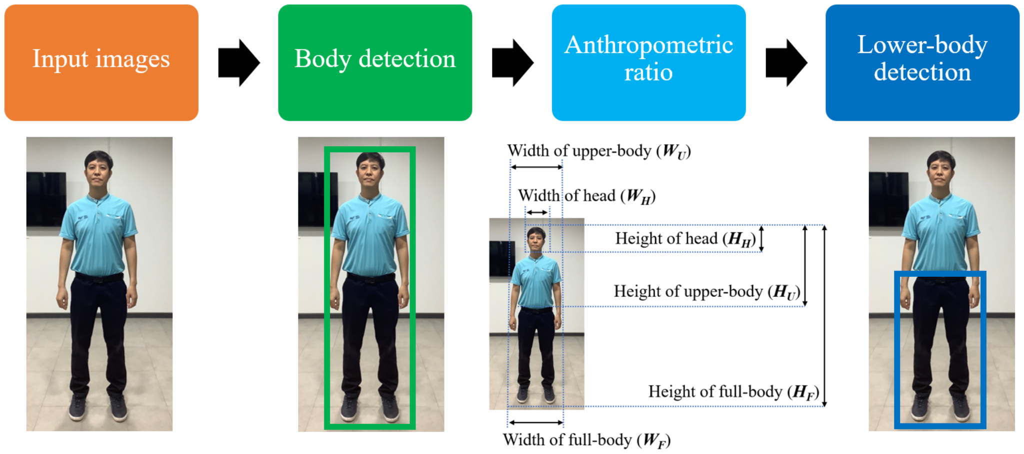


Figure 1. Conceptual framework for lower-body detection using the proposed anthropometric ratios.

The rest of this paper is organized as follows: Section 2 briefly describes the related work in the literature. The proposed method is introduced in Section 3. Experimental results are reported in Section 4 and discussed in Section 5. Finally, the conclusion is presented in Section 6.

2. Related Work

Several human detection algorithms have been studied and developed for many applications, e.g., person identification [27], automatic vehicles [28], and human gait analysis [29]. This section discusses the two main approaches to human body detection: traditional methods and deep learning methods.

2.1. Traditional Methods

In traditional methods, a high-dimensional image is transformed into low-dimensional data in the form of a feature vector. Then, the feature vectors extracted in this way are used to train a classification algorithm. In general, the traditional methods for human detection can be divided into two types: background subtraction and object-based methods.

Background subtraction [30] was introduced for object detection to identify moving objects based on the differences between the current frame and a background frame in either a pixel—by-pixel or window-by-window fashion [31]. This background subtraction was also combined the depth information to extract an object with higher success and capability [32]. A human bounding box can be detected by a refinement algorithm by matching the contour of the shadow of the human body. In addition, the Gaussian mixture model (GMM)-based background learning technique was applied to separate the human object from the background [33]. Iazzi et al. [34] also applied the background subtraction with a support vector machine (SVM) classifier to detect a fall in elderly people. The results showed that this method can gain a high accuracy of fall detection. However, the main problem with this background subtraction is that is not robust against changes in brightness or camera motion in the case of a non-stationary frame because the background frame is not updated [35,36]. Thus, Chiu et al. [37] proposed an idea of color category entropy to approximate the number of essential background groups and initiate acceptable representative background groups to accommodate dynamic background.

An object-based method for detecting faces was introduced by Mena et al. [38], who presented the Viola-Jones (VJ) algorithm. The VJ algorithm relies on an integral image representation and simple rectangular features such as Haar-Like features, based on which cascaded classifiers are used to detect faces. Adeshina et al. [39] also applied Haar-Like features with local binary patterns (LBP) to customize classroom face classification. As a result, the proposed algorithm showed a lower value of false-negative rate (FNR) of face classification than other methods. However, the VJ algorithm still has disadvantages when confronted with varying lighting conditions and occlusion. The histograms of oriented gradients (HOG) features [11] proposed in 2005 are high-dimensional features based on edges, which can be used in combination with a SVM classifier to detect human regions. When a SVM classifier is trained on the HOG features of both positive (human) and negative (non-human) images, the resulting HOG-SVM method can successfully detect human targets even in dark or occluded images. Moreover, Patel et al. [40] proposed a fusion of HOG features for human action recognition in video. The result showed that this fusion of HOG features and meta-cognitive neural network classifier archived high accuracy to detect human action. However, the HOG feature takes a long time to calculate the sliding and scaling windows needed to extract HOG features covering the entire input image. To deal with these problems, He et al. [41] proposed a fully-convolutional neural network for semantic regions of interest to detect pedestrians based on HOG and a SVM classifier. In addition, this proposed method can increase the speed of the algorithm. Additionally, Yang et al. [42] presented the parallel feature fusion based on Choquet integral between HOG and LBP features for pedestrian detection. This proposed algorithm improved the accuracy of detection and reduced the time of pedestrian detection.

2.2. Deep Learning Method

Deep learning methods consider both local and global features of the input image by using kernel filters for human body detection. Jammalamadaka et al. [43] presented a human body recognition approach using deep learning. They found that convolutional neural networks (CNNs) can successfully extract and detect human body parts. They also constructed a suitable pose estimation method by using a similar mapping between dimensions. However, this system requires optimizing a low-dimensional space for the human pose search. Qin and Josef [44] proposed a pedestrian detection method using improved CNNs with multi-feature fusion selection. Local parts of the human body were considered individually for the extraction of local features. Then, they merged these local features for full-body human detection. The ability to detect people achieved through multi-feature fusion was superior to what could be achieved based on the individual original features. However, the human anatomical proportions used to divide the parts of the human body were designed without referring to anthropometric data drawn from real human body measurements. Furthermore, the complexity of the multi-feature fusion process remained higher than that of using only one feature. Considering the human parts, Cao et al. [45] proposed the OpenPose method which is applied the CNNs model to map between the Part Affinity Fields (PAFs) and body joints. The results showed the high accuracy of body part detection of the human for multi-person detection. Lin et al. [46] also deployed the OpenPose method to detect human movement through detecting the keypoints of human joint changes. They applied the series recurrent neural network, long- and short-term memory (LSTM), and gated recurrent unit (GRU) models to detect a human fall. Nonetheless, the length of keypoints might be related to the human anatomical proportion from anthropometric data of the real human body.

While previous approaches can achieve adequate pedestrian detection performance, they may be limited to full-body detection based on human images. Consequently, in some applications, such as lower-body activities, these previous algorithms cannot perform well due to insufficient information. Our work investigates an indirect detection method for lower-body detection using anthropometric ratios applied to human body images.

3. Materials and Methods

This section describes the data set used in this study, the feature extraction processes with conventional and deep learning methods, the classifiers used, and our experimental evaluations. The experimental framework is illustrated in Figure 2. There are three main phases: image input, feature extraction, and classification.

This paper presents a framework for lower-body detection using proposed anthropometric ratios, as illustrated in Figure 2. A person is captured or recorded by a camera, producing the input image. Then, the human body is detected and scaled using the proposed anthropometric ratios in combination with either traditional techniques (A-Traditional), such as the VJ and HOG-SVM algorithms, or the CNNs technique (A-CNNs). Finally, the lower-body area in the image is detected.

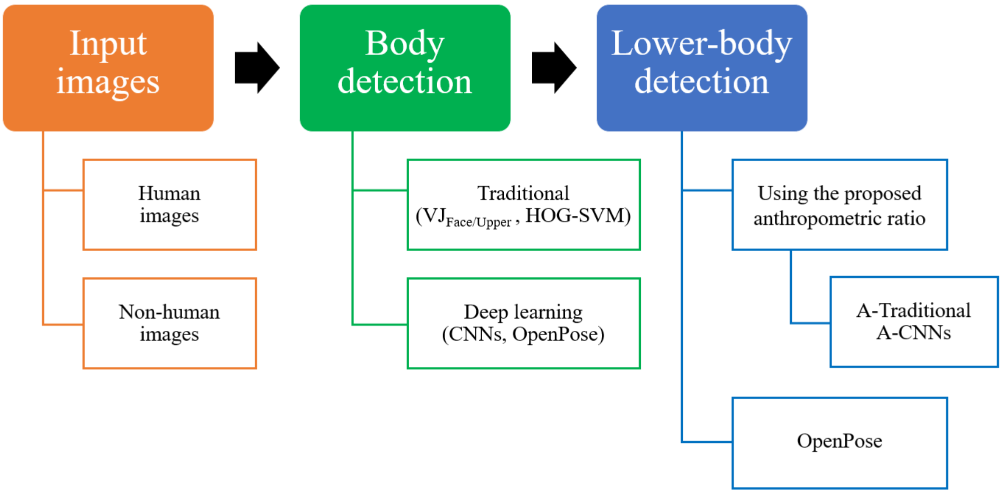

Input images
Body detection
Lower-body detection
Human images
Non-human images
Traditional ($VJ_{Face/Upper}$, HOG-SVM)
Deep learning (CNNs, OpenPose)
Using the proposed anthropometric ratio
A-Traditional
A-CNNs
OpenPose

Figure 2. Detailed framework for lower-body detection using the proposed anthropometric ratios.

For human body detection, there are two popular methods based on sliding windows for overall positioning and scaling in images [11,12,16]: traditional methods and deep learning methods.

3.1. Traditional Methods

As the basis for the application of the proposed anthropometric ratios, two traditional methods are used in this study to detect the human body: the VJ algorithm and the HOG-SVM algorithm. The VJ algorithm is used to perform frontal face or upper-body detection. The HOG algorithm is also used for frontal full-body detection.

3.1.1. Viola-Jones Algorithm (VJ)

The VJ algorithm [12,38] extracts simple features based on the notion of cascaded classifiers. Some instances of contrast detection are performed in cells in specific locations in an image, such as the human eye. The VJ algorithm solves the complex learning problem by using an enormous number of positive and negative training images together with a cascade of simple classifiers. The corresponding process is summarized in Figure 3.

As shown in Figure 3, the first step of the VJ Algorithm 1 for face detection is to convert the input image into a greyscale image. Next, the integral image representation is rapidly generated by calculating the value at pixel (x, y) as the sum of the pixels above and to the left of (x, y). Then, the sum of the values of all pixels in rectangle D, as shown in Figure 4, can be computed as $4 + 1 - (2 + 3)$. Subsequently, the entire image is scanned to calculate Haar-like features by subtracting the sum of the pixels under white rectangles from the sum of the pixels under black rectangles in patterns similar to those shown in Figure 5. Adaptive boosting (AdaBoost) is then used as a machine learning algorithm to find the best such rectangle features among the approximately 160,000 possible features in a window of 24×24 pixels in order to construct a linear combination of corresponding classifiers. In the final phase of the VJ algorithm, the input image is fed into these cascaded classifiers; if the input image passes all stages of the classifier cascade, the input image is identified as a human face image, whereas if the image does not pass any stages, it is not a human face image, as shown in Figure 6.

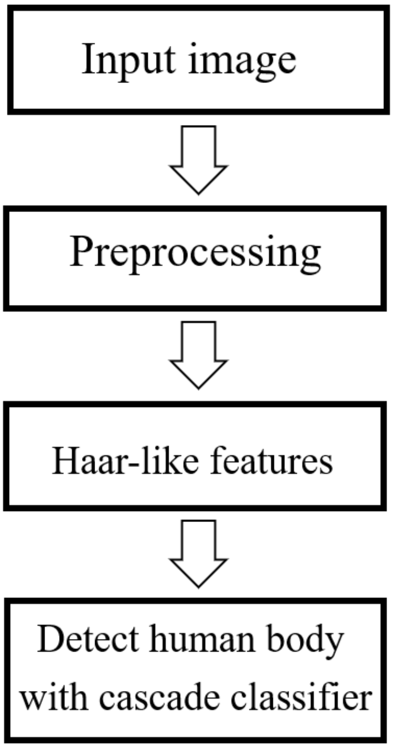


Figure 3. Viola-Jones algorithm.

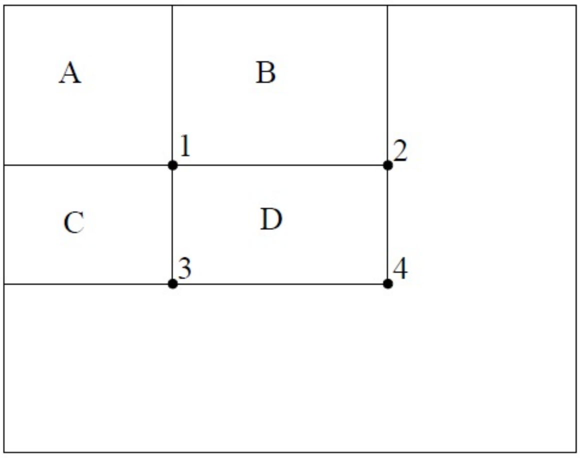


Figure 4. The summation of the pixels can be computed with four reference location.

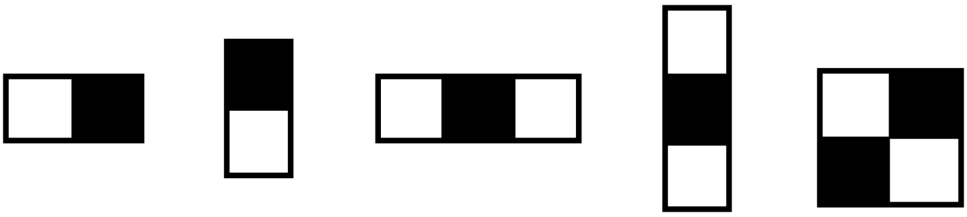


Figure 5. The example of rectangle images of Haar-like features.

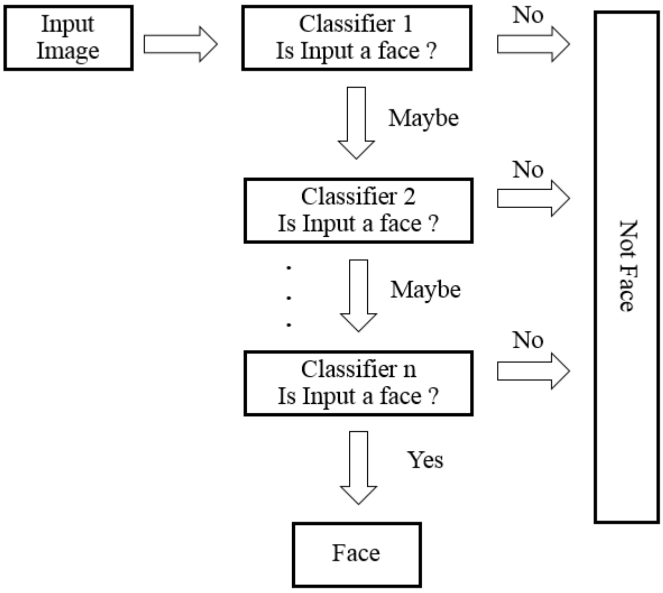


Figure 6. The cascade classifier for face detection.

Algorithm 1: Viola-Jones for face detection algorithms

```
Data: $P$ of the image is more than zero
Result: Result of face detection
Convert color image to gray image;
while $N_{scale} > 0$ do
    Down-sample image by one scale;
    Compute integral image for current scale;
    while $N_{sliding} > 0$ do
        while $N_{cascade} > 0$ do
            while $N_{filter} > 0$ do
                Filter the detection window;
            Accumulate filter outputs within this stage;
            if accumulation fails to pass per-stage threshold then
                Reject this window as a face;
                Break the while loop;
    if this detection window passes all $N_{cascade}$ of thresholds then
        Accept this window as a face;
    else
        Reject this window as a face;
where $P$ is the pixel sizes of an image size, $N_{scale}$ is the number of scales in image
  pyramid, $N_{sliding}$ is the number of the sliding detection window, $N_{cascade}$ is the
  number of stage in the cascade classifier, and $N_{filter}$ is the number of filter in
  the stage.
```

3.1.2. Support Vector Machine Classification Based on Histograms of Oriented Gradients (HOG-SVM)

Dalal and Triggs [11] proposed the HOG method for human detection, as demonstrated in Figure 7. In this method, the HOG features of both positive images (human images) and negative images (non-human images) are extracted and used to fine-tune a pre-trained linear SVM classifier for human detection. The overall process of HOG-SVM detection is summarized in Figure 8.

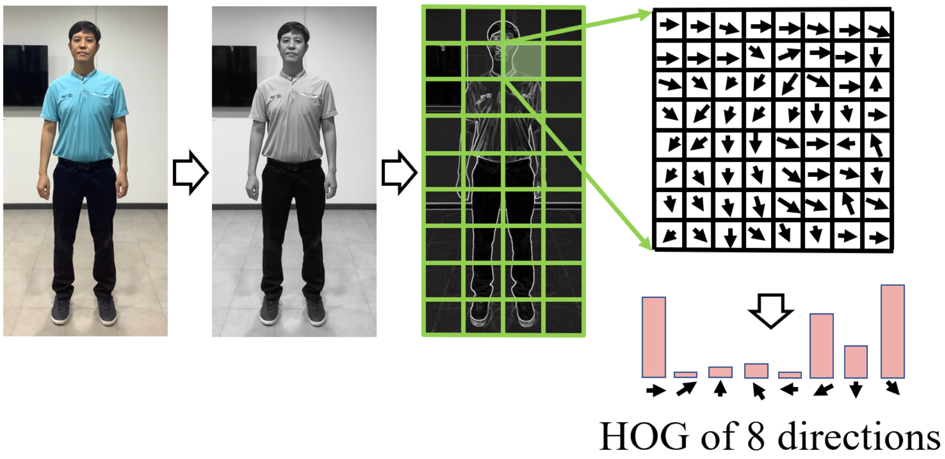


Figure 7. The example of HOG process.

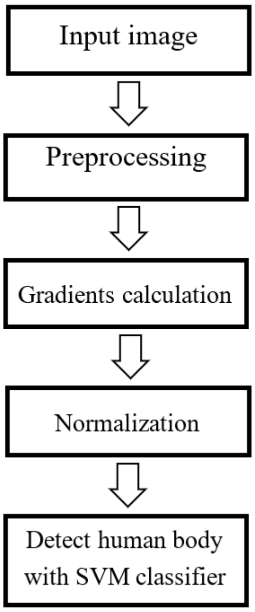


Figure 8. The HOG-SVM for human detection.

The HOG-SVM algorithm for human detection is presented in Algorithm 2. The HOG-SVM process is initialized with configuration parameters such as the sizes of cells, blocks, and bins of a sliding window. The image is then converted to greyscale. The sliding window is calculated as the HOG feature for whole the image.

For the sliding window, the gradients on the x-axis and y-axis are calculated using Equations (1) and (2), and the edge angles are computed using Equation (3).

$$G_x(x,y) = f(x+1,y) - f(x-1,y) \tag{1}$$

where $G_x(x,y)$ is the gradients of x-axis and $f(x,y)$ is the pixel value of gray scale image at (x,y) coordinates.

$$G_y(x,y) = f(x,y+1) - f(x,y-1) \tag{2}$$

where $G_y(x,y)$ is the gradients of y-axis and $f(x,y)$ is the pixel value of gray scale image at (x,y) coordinates.

$$Direction(x,y) = arctan(G_y(x,y)/G_x(x,y)) \tag{3}$$

where $Direction(x,y)$ is the angle of gradients at (x,y) coordinates. The magnitude of the gradients is presented in Equation (4).

$$Magnitude(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \tag{4}$$

where $Magnitude(x,y)$ is the magnitude of gradients at (x,y) coordinates. The edge histogram is created by gradient voting as shown in Equation (5)–(7).

$$\alpha = (n + 0.5) - \frac{N_{bin} * Direction(x,y)}{\pi} \tag{5}$$

where α is the weight of gradient vote, N_{bin} is the number of bins, and $Direction(x,y)$ is the angle of gradients by Equation (3).

$$m_n = (1 - \alpha) * Magnitude(x,y) \tag{6}$$

where m_n is the magnitude of gradient vote at the n bin, α is the weight of gradient vote by Equation (5), and $Magnitude(x,y)$ is the angle of gradients by Equation (4).

$$m_{nearest} = (\alpha) * Magnitude(x,y) \tag{7}$$

where $m_{nearest}$ is the magnitude of gradient vote which is near the n bin, α is the weight of gradient vote by Equation (5), and $Magnitude(x,y)$ is the angle of gradients by Equation (4).

Algorithm 2: Support vector machine classification based on histograms of oriented gradients

```
Data: $P$ of the image is more than zero
Result: Result of human detection
Configure the parameters of sizes of cell, block, bins, and percentage of overlapping;
Convert color image to gray image;
while number of scales in image pyramid do
    Downsample image by one scale;
    while $N_{scale} > 0$ do
        while $N_{block} > 0$ do
            while $N_{cell} > 0$ do
                Calculate the magnitude gradients of $x$-axis and $y$-axis;
                Calculate the edge degree;
                while $N_{bin} > 0$ do
                    Build a histogram from edge orientations and gradients level;
        Vote the gradients level in each the edge orientations;
        Normalize the histogram by neighbour cells;
        Flattening 2D features into a vector of features;
        Test this vector in SVM classifier;
        if detection window passes the thresholds then
            Accept this window as a human;
        else
            Reject this window as a human;
where $P$ is the pixel sizes of an image size, $N_{scale}$ is the number of scales in image
  pyramid, $N_{block}$ is the number of the block in each window image, $N_{cell}$ is the
  number of cell in each block, and $N_{bin}$ is the direction in each cell.
```

Subsequently, the HOG features are normalized as shown in Equation (8) to be suitable for a variety of lighting conditions [11].

$$M_i = \frac{m_i}{\sqrt{\sum_{j=1}^{K} m_j^2 + e^2}} \tag{8}$$

where M_i is the normalized magnitude of gradient vote at i bin when $i = 1$ to K, K is the number of cell in one block multiplied by the number of bins (N_{bin})and e is a small constant value.

The 2D features of the sliding window extracted in this way are converted into a single vector of features. Finally, this vector of features is tested in a SVM classifier. If the sliding window passes the threshold, it is detected as a human.

3.2. Deep Learning

Deep learning [16] is a technique for machine learning that can consider both the low-level and high-level information in a large data set. A deep learning architecture is generally similar to that of an artificial neural network but has greater numbers of hidden layers and nodes. In this study, a CNN is used for frontal full-body detection. A CNN model typically consists of convolutional layers, pooling layers, and fully connected layers, as shown in Figure 9:

1. Convolutional layers: These layers are the core of the model, consisting of filters or kernels to calculate image features such as lines, edges, and corners. Generally, a filter consists of a mask matrix of numbers moved over the input image to calculate specific features. The convolution operations of filters consist of dot products and summations between the filters and the input image. The output of these operations is usually passed through an activation function designed for a particular purpose, such as the rectified linear unit (ReLU) activation function for non-linear input.
2. Pooling layers: These layers generally reduce the dimensionality of the features. They represent pooled feature maps or new sets of features, moving from the local scale to the global scale. There are several possible pooling operations, including taking the maximum, average, or summation of each corresponding cluster of data from the previous layer.
3. Fully connected layers: A fully connected layer, in which every neuron in the previous layer is connected to every neuron in the current layer, is typically used as a final layer. The softmax activation function is commonly used in a fully connected output layer to classify the input image into one of several classes based on the training images.

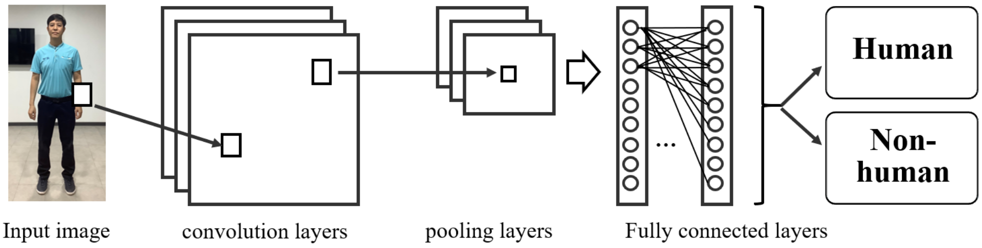


Figure 9. The diagram of CNNs for human and non-human detection.

Moreover, the experiment was compared with OpenPose method [45] which is a pre-trained model for human detection based on the PAFs relating to human body joints. In the OpenPose technique, there are three main procedures for human detection:

1. Keypoints localization: The input image is located and predicted all the possible keypoints as human body joints based on a confidence map. This map is also beneficial of one person pose estimation.
2. Part Affinity Fields: The keypoints are mapped to the 2-dimensional vector field for location and orientation of the associated human limbs.

3. Greedy Inference: The 2-dimensional vector field is generated the pose keypoints for all the people in the image.

To maintain runtime performance, the OpenPose method [45] is the limited computation to a maximum of 6 stages, allocated differently procedures across the part affinity fields and keypoints localization.

3.3. Proposed Lower-Body Detection Framework Using Anthropometric Data

In this section, anthropometric data [21,25,47] representing scaling relations for the human body are introduced. This section also illustrates a method of using anthropometric data to transform three regions of interest (ROIs) of the human body, namely, the full-body, the upper-body, and the face, into the lower-body ROI.

3.3.1. Anthropometric Data

In this section, anthropometric data [48] representing human body information are introduced. A survey of anthropometric data is generally related to the size, motion, and mass of the human body. Such survey data can be applied to design suitable clothing, ergonomic devices, or workspaces. In this study, human body size data from the NASA Anthropometry and Biomechanics Facility [21] are selected to provide information on the height and width of various body parts in a standing posture from the frontal view. This information was collected from healthy adults with an average age of approximately 40 years and from a wide range of ethnic and racial backgrounds. The example of anthropometric dimensional data is shown in Figure 10. There are three dominant parts of the human body considering in this experiment:

1. Full-body: the width of the full-body is similar to the width of the upper-body, and the height of the full-body is measured from the foot to the top of the head.
2. Upper-body: the width of the upper-body is recorded from the edge of the left hand to the edge of the right hand with the hands resting on the body, and the height of the upper-body is measured from the waist to the top of the head.
3. Head: the width of the head is measured from the left ear to the right ear, and the height of the head is measured from the chin to the top of the head.

To address occlusion problems affecting the lower-body, this research aims to detect the lower-body indirectly by using anthropometric data. To apply anthropometric data for lower-body detection, a suitable human ROI ratio can be used to transform an ROI corresponding to any other part of the body into the lower-body ROI. In this study, three main ROIs are considered for transformation to the lower-body:

- Full-body: the HOG-SVM or CNNs algorithm is used to detect the full-body of the target, as shown in Figure 11.
- Upper-body: the VJ algorithm for upper-body detection (VJ_{Upper}) is applied to detect the upper-body of the target, as illustrated in Figure 12.
- Head: the head of the target is detected by using the VJ algorithm for face detection (VJ_{Face}), under the assumption that the head ROI is close to the face ROI, as demonstrated in Figure 13.

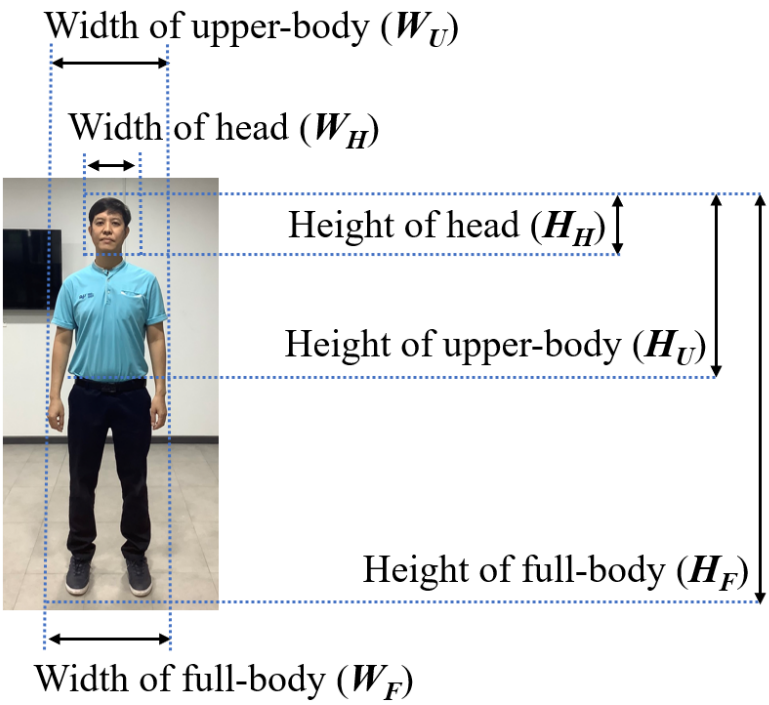

Width of upper-body (W_U)
Width of head (W_H)
Height of head (H_H)
Height of upper-body (H_U)
Height of full-body (H_F)
Width of full-body (W_F)

Figure 10. The example of the anthropometric dimensional data.

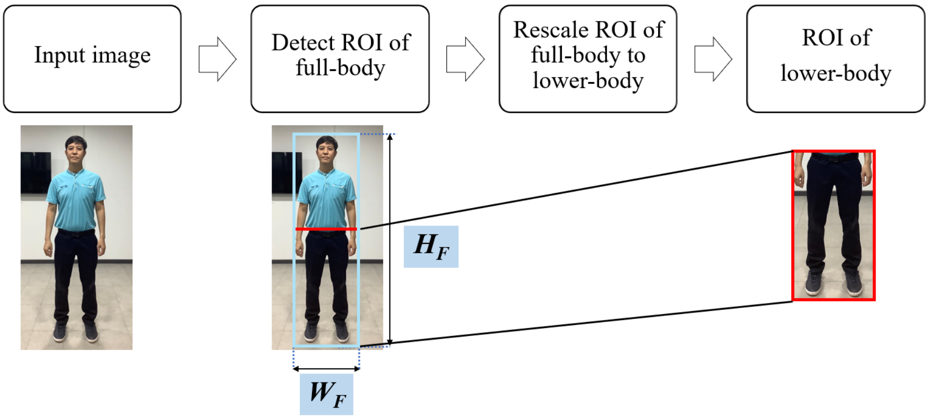


Figure 11. The process for lower-body detection using full-body ratio.

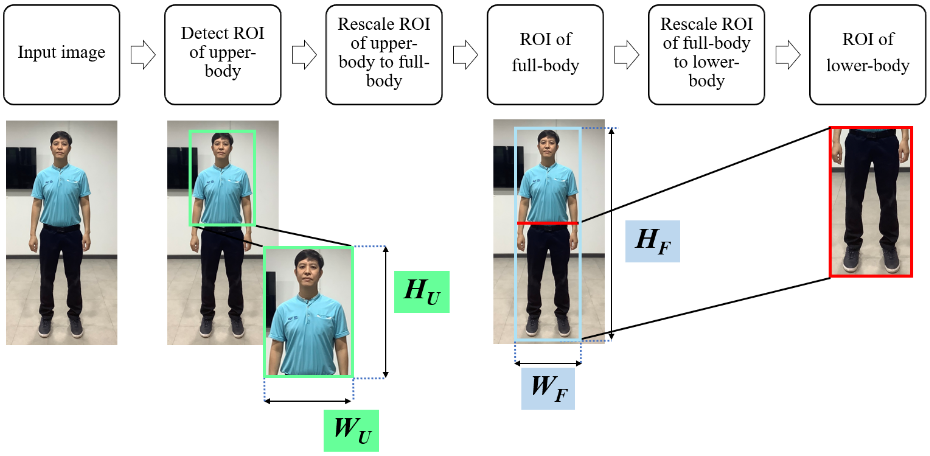


Figure 12. The process for lower-body detection using upper-body ratio.

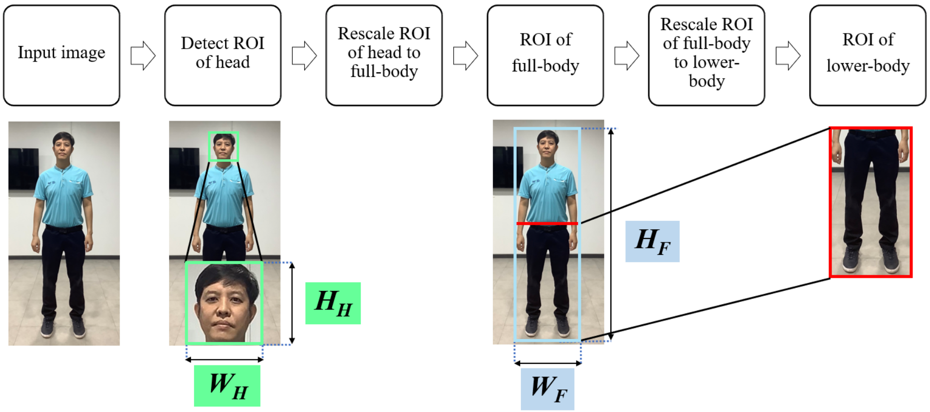


Figure 13. The process for lower-body detection using head ratio.

The anthropometric data ratios are constructed from the median scaled sizes of the head, upper-body, lower-body, and full-body in a standing posture, as shown in Tables 1 and 2. However, the anthropometric data collected from female subjects are not sufficiently comprehensive; therefore, in this study, only male anthropometric data are selected for lower-body detection.

Table 1. The anthropometric data of upper-body in cm (centimetres), NA = Not Available where H_U is the height of upper-body, W_U is the width of upper-body, H_F is the height of full-body, W_F is the width of full-body, R_{HU} is the ratio of H_U:H_F, and R_{WU} is the ratio of W_U:W_F.

Gender	H_U	W_U	H_F	W_F	R_{HU}	R_{WU}
Male	71.6	55.1	179.9	55.1	0.398	1
Female	60.3	NA	157.0	NA	0.384	NA

Table 2. The anthropometric data of head in cm (centimetres), NA = Not Available where H_H is the height of head, W_H is the width of head, H_F is the height of full-body, W_F is the width of full-body, R_{HH} is the ratio of H_H:H_F, and R_{WH} is the ratio of W_H:W_F.

Gender	H_H	W_H	H_F	W_F	R_{HH}	R_{WH}
Male	24.4	15.7	179.9	55.1	0.136	0.285
Female	21.8	15.6	157.0	NA	0.139	NA

3.3.2. Transformation of the Full-Body ROI into the Lower-Body ROI

The HOG-SVM or CNNs algorithm can be used to directly detect the full-body ROI of a human in an image. This full-body ROI can then be cropped to obtain the lower-body ROI as shown in Equations (9)–(11). This process is also illustrated in Figure 14. In addition, the information of the anthropometric data of upper-body is shown in Table 1.

$$R_{HU} = H_U / H_F \tag{9}$$

where R_{HU} is the ratio height of upper-body per full-body, H_U is the height of upper-body, and H_F is the height of full-body.

$$H_{Lower} = H_F * (1 - R_{HU}) \tag{10}$$

where H_{Lower} is the length of lower-body ROI, H_F is the length of full-body ROI, and R_{HU} is the ratio height of upper-body per full-body as shown in Table 1.

$$W_{Lower} = W_F \tag{11}$$

where W_{Lower} is the width of lower-body ROI and W_F is the width of full-body ROI.

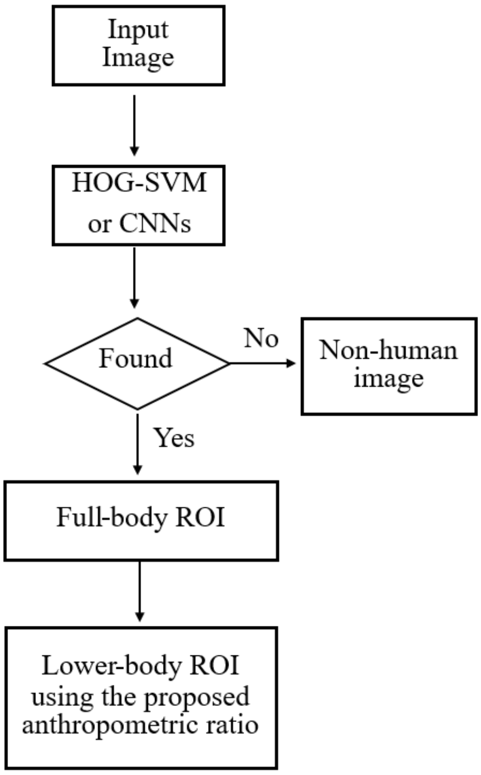


Figure 14. The framework of the HOG-SVM or CNNs for lower-body detection.

3.3.3. Transformation of the Upper-Body ROI into the Lower-Body ROI

As illustrated in Figure 15, VJ_{Upper} is used to detect the ROI of the upper-body; then, this ROI is converted into the full-body ROI by using R_{HU} and R_{WU} as shown in Equations (9), (12)–(14).

$$H_F = H_{upper}/R_{HU} \tag{12}$$

where H_F is the height of full-body ROI and H_{upper} is the height of upper-body detection ROI by VJ algorithm and R_{HU} is the ratio height of upper-body per full-body as shown in Table 1.

$$R_{WU} = W_U/W_F \tag{13}$$

where R_{WU} is the ratio width of upper-body per full-body, W_U is the width of upper-body, and W_F is the width of full-body.

$$W_F = W_{upper}/R_{WU} \tag{14}$$

where W_F is the width of full-body ROI and W_{upper} is the width of upper-body detection ROI by VJ algorithm and R_{WU} is the ratio width of upper-body per full-body as shown in Table 1.

Subsequently, the estimated full-body ROI is cropped to obtain the lower-body ROI as shown in Equations (10) and (11).

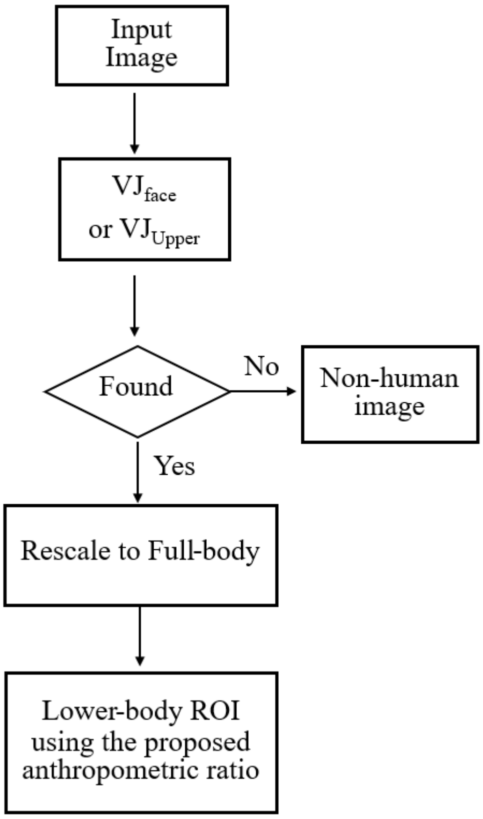

Input Image
VJ$_{face}$ or VJ$_{Upper}$
Found
No
Non-human image
Yes
Rescale to Full-body
Lower-body ROI using the proposed anthropometric ratio

Figure 15. The framework of the VJ_{Face} or VJ_{Upper} for lower-body detection.

3.3.4. Transformation of the Face ROI into the Lower-Body ROI

In case of using VJ_{Face} to find the ROI of the human face, as demonstrated in Figure 15, the face ROI is converted into the full-body ROI by means of R_{HH} and R_{WH} as shown in Equation (15)–(18). Moreover, the information of the anthropometric data of head is shown in Table 2.

$$R_{HH} = H_H / H_F \tag{15}$$

where R_{HH} is the ratio height of head per full-body, H_H is the height of head, and H_F is the height of full-body.

$$R_{WH} = W_H / W_F \tag{16}$$

where R_{WH} is the ratio width of head per full-body, W_H is the width of head, and W_F is the width of full-body.

$$H_F = H_{Face} / R_{HH} \tag{17}$$

where H_F is the height of full-body ROI and H_{Face} is the height of face detection ROI by VJ algorithm and H_{head} ratio is the ratio height of head per full-body as shown in Table 2.

$$W_F = W_{Face} / R_{WH} \tag{18}$$

where W_F is the width of full-body ROI and W_{Face} is the width of face detection ROI by VJ algorithm and R_{WH} is the ratio width of head per full-body as shown in Table 2.

Then, this full-body ROI is cropped to obtain the lower-body ROI as shown in Equations (10) and (11). To summarize, diagrams of the frameworks for using the HOG-SVM or CNNs algorithm and the VJ_{Face} or VJ_{Upper} algorithm for lower-body detection are shown in Figures 14 and 15, respectively.

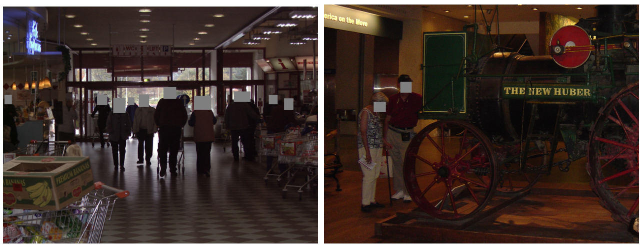
3.4. Dataset

Experiments were conducted using the INRIA Person Dataset [11], which consists of upright human images (positive images) and general background images (negative images). This data set is challenging for lower-body detection methods because it consists of images captured under various lighting conditions and containing occluding objects, such as vehicles and furniture, close to the human targets of interest.

In these experiments, 2416 positive images and 1218 negative images were used for training, where the negative images were obtained by randomly cropping the background images. Similarly, the data set used for testing included 1126 positive images and 453 randomly cropped negative images.

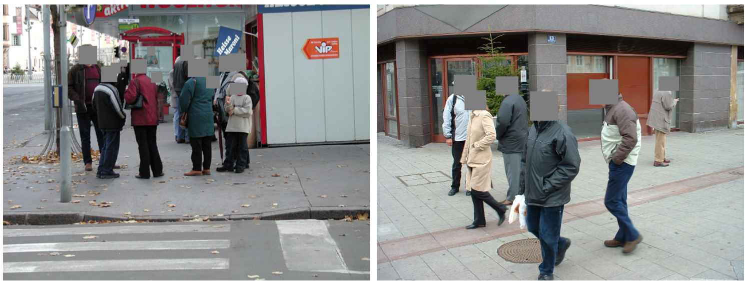
To analyse the results under different image conditions, five cases of human detection were investigated and analysed. Example images for cases 1–5 are shown in Figures 16–20, respectively. Five cases of scenario [49–53] are described as:

1. Case of challenging lighting conditions: The light level in an image may not be sufficient to clearly reveal the presence of humans [49,51]. In particular, this may occur in indoor and night-time scenes, resulting in low image quality.
2. Case of occlusion: Occlusion refers to overlapping either between a human and another human or between a human and another object in the image [49–51]. This can affect the ability to identify complete human shapes, such as in the case of a group of standing people.
3. Case of multiple people: There may be more than one person in an image [49,51], such as in public sightseeing images or shopping mall images. Some algorithms can support multiple detection [11,12].
4. Case of a difference in pose between the training and test images: A pose refers to the gesture or posture of a human in an image. For a test image depicting a person in a pose that does not appear in the training images [52], it may be difficult to detect whether the ROI is human or not human because it is not sufficiently similar to the training images [11].
5. Case of different clothes: People in images may wear clothes of many different colors, sizes, and styles as well as different accessories [53]. Sometimes, certain clothing characteristics may make it difficult to identify a human shape.



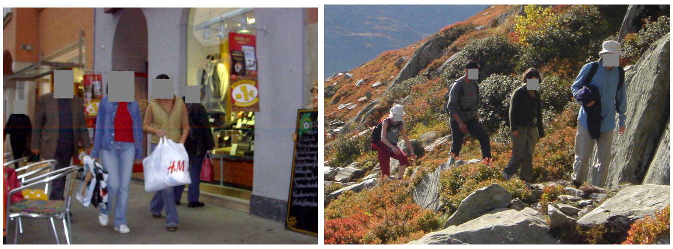
(**a**) shopping mall (**b**) museum

Figure 16. Example images of challenging lighting conditions.



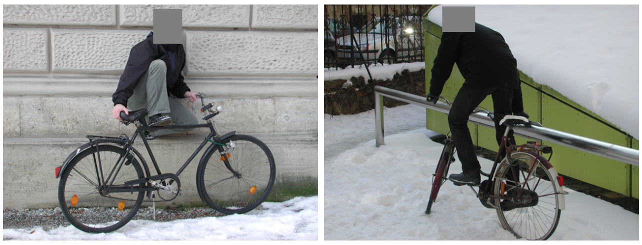
(**a**) group of people (**b**) walking people in the street

Figure 17. Example images of the occlusion.



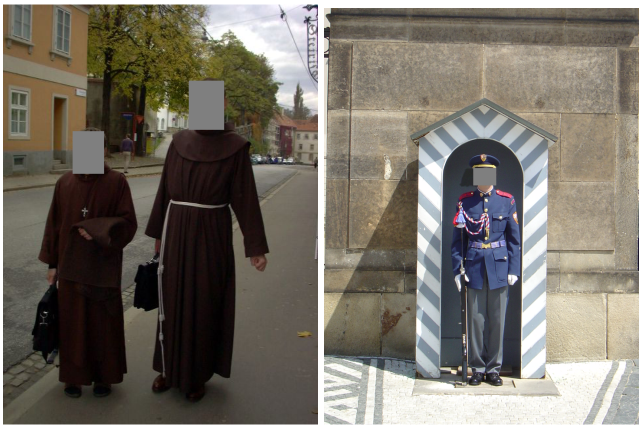
(**a**) street (**b**) mountain

Figure 18. Example images of multiple people.



(**a**) sitting on a bicycle (**b**) riding on a bicycle

Figure 19. Example images of a difference in pose between the training and test images.



(**a**) long gown (**b**) solider uniform

Figure 20. Example images of challenging clothes.

3.5. Evaluation

To evaluate the lower-body detection performance of the frameworks, the confusion matrix [54,55] and complexity were used as performance measures. The difference image cases listed above were also analysed to investigate their influence on the detection ability.

Table 3 presents the confusion matrix used for the evaluation of the frameworks. The columns represent a framework's detection results, and the rows represent the actual class. The entries in Table 3 are defined as follows:

- TP denotes the number of images in the human data set that are correctly detected to contain at least one lower-body ROI.
- FN denotes the number of images falsely identified as non-human images in the human data set.
- FP denotes the number of images in the non-human data set that are falsely detected to contain at least one lower-body ROI.
- TN denotes the number of images correctly identified as non-human images in the non-human data set.

Table 3. Confusion matrix for two classes' detection.

	Detected Human	Detected Non-Human
Human	TP	FN
Non-human	FP	TN

The performance of a framework on detection problems can be measured based on the confusion matrix. This paper focuses on three measures: sensitivity, specificity, and accuracy. Their equations are given in Table 4. The first common measure of detection performance is the accuracy. It can be used to evaluate the overall efficiency of a framework. Meanwhile, the sensitivity measures the accuracy of human detection in human images (positive images), whereas the specificity measures the accuracy of non-human detection (negative images).

Table 4. Measures of detection performance based on the confusion matrix [54,55].

Measurement	Equation
Sensitivity	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$

The complexity of each framework, reflecting the complexity of the algorithm used for detection, was also investigated.

The parameters used for VJ detection in these experiments were in the same scale range as in a previous experiment [12]. For face detection, the minimum window size was 20 × 20 pixels, and for upper-body detection, the minimum window size was 60 × 60 pixels. For the extraction of HOG features, the window size was 64 × 128 pixels. The CNNs model was modified from Chakrabarty and Chatterjee experiment [56]. The customized model comprised two pairs of convolutional and pooling layers, each with 32 filters with dimensions of 3 × 3. A final fully connected layer with 64 neurons was used to classify each input image as human or non-human based on the softmax function, as illustrated in Figure 21. In the case of the OpenPose, a pre-trained model was deployed to human pose estimation with six stages as in a previous experiment [45].

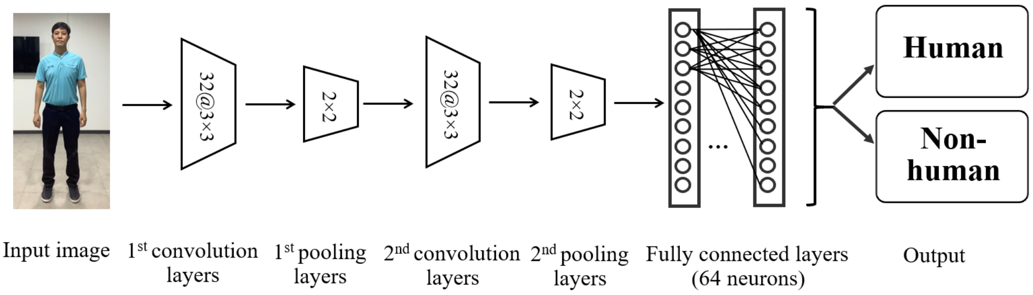


Figure 21. The diagram for configuration of CNNs model, modified from Chakrabarty and Chatterjee [56].

4. Experimental Results

In this section, three perspectives are considered for the evaluation of lower-body detection with the proposed anthropometric ratios: the performance of different frameworks, their complexity, and their sensitivity to different image conditions such as lighting conditions, an occlusion, multiple people, a difference in pose between the training and test images, and challenging clothes (The detail of image conditions are explained in Section 3.4).

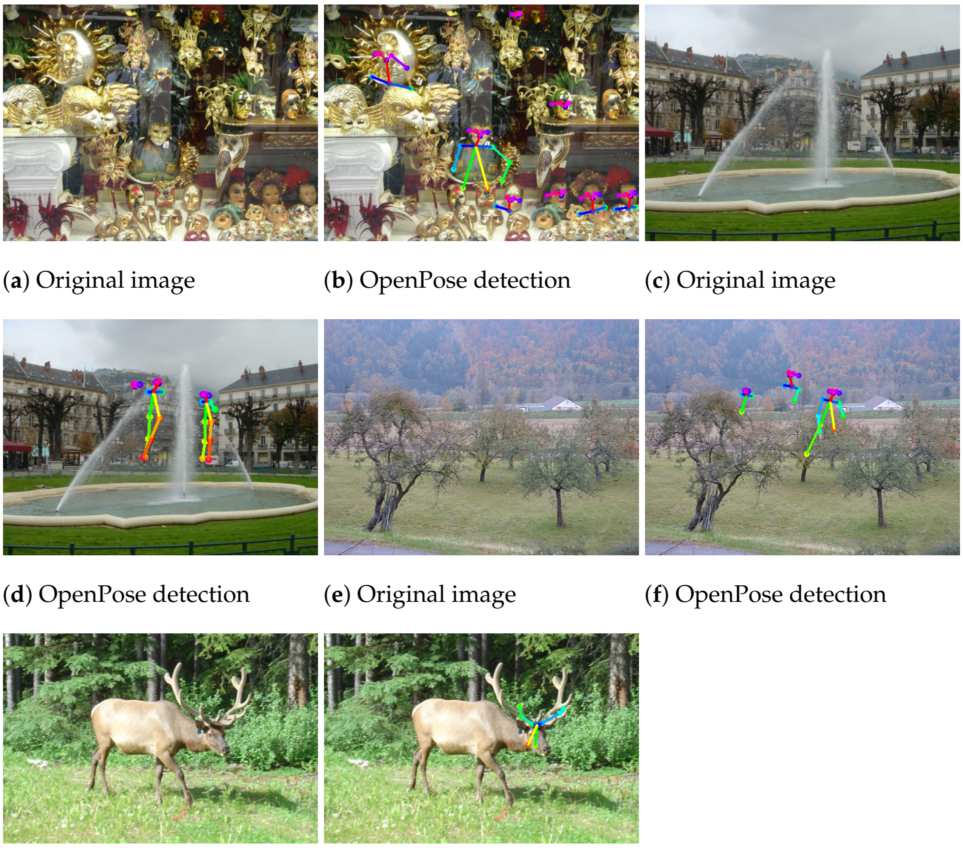
4.1. Accuracy

Table 5 summarizes the performance of different frameworks for lower-body detection with the proposed anthropometric ratios. It is clear that in the case of the sensitivity measure tested on the human image, the HOG-SVM framework again shows the highest performance among the traditional algorithms for detecting human images (72.22%), while VJ_{Face} and VJ_{Upper} can only achieve human detection with an accuracy of less than 36%. Regarding the deep learning techniques, OpenPose achieves higher sensitivity (99.73%) than the A-CNNs method (75.94%).

Table 5. Performance of frameworks for lower-body detection with the proposed anthropometric ratios.

Method	A-Traditional (%)			Deep Learning (%)	
Algorithm	VJ_{Face}	VJ_{Upper}	**HOG-SVM**	**OpenPose**	**A-CNNs**
Sensitivity	34.03	35.76	72.22	99.73	75.94
Specificity	99.56	91.83	85.43	86.09	99.30
Accuracy	74.09	70.04	80.30	95.82	90.14

On the non-human data set, all frameworks achieve a total specificity of more than 85% for detecting background images. The VJ_{Face} method achieves the highest specificity (99.56%), while the specificity of HOG-SVM is the lowest (85.43%). For the deep learning methods, the specificity of A-CNNs (99.30%) is higher than that of OpenPose (86.09%). The specificity of OpenPose is decreased by its false positive detections on background images, as shown in Figure 22.



(**a**) Original image (**b**) OpenPose detection (**c**) Original image

(**d**) OpenPose detection (**e**) Original image (**f**) OpenPose detection

(**g**) Original image (**h**) OpenPose detection

Figure 22. Result of false positive detection using OpenPose method. Result of false positive detection using OpenPose method.

An averaged accuracy of A-Traditonal methods is around 74.81%. The HOG-SVM algorithm also provides higher overall accuracy (80.30%) than the other traditional methods, while the accuracy of VJ_{Face} is higher than that of VJ_{Upper}. In terms of deep learning, both the A-CNNs and OpenPose methods achieve overall accuracies of greater than 90%.

4.2. Complexity

Table 6 lists the complexity of each algorithm (VJ_{Face}, VJ_{Upper}, HOG-SVM, and A-CNNs algorithms) for lower-body detection. The VJ_{Face} and VJ_{Upper} algorithms have low complexity ($O((P+4N_r)MT)$, whereas the HOG-SVM method has high complexity ($O(N_rWN_f^2)$. A-CNNs and OpenPose methods also have a high complexity of $O(N_T \sum_{l=1}^{d}(n_{l-1}C_l))$, but the A-CNNs has only one stage of N_T while the OpenPose is configured as six stages of N_T.

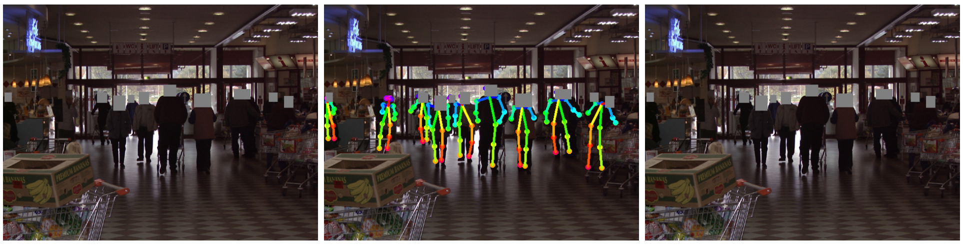
Table 6. Complexity of algorithm for lower detection.

Method	**A-Traditional**		**Deep Learning**
Algorithm	VJ_{Face}, VJ_{Upper}	**HOG-SVM**	**OpenPose, A-CNNs**
Integral	$O(P+4N_r)$	-	-
Casdcade	$O(MT)$	-	-
HOG	-	$O(N_rWN_f)$	-
SVM	-	$O(N_f)$	-
CNNs	-	-	$O(N_T \sum_{l=1}^{d}(n_{l-1}C_l))$
Total	$O((P+4N_r)MT)$	$O(N_rWN_f^2)$	$O(N_T \sum_{l=1}^{d}(n_{l-1}C_l))$

Note: P is the pixel sizes of an image size, N_r is the number regions of interest covers W pixels, M is the number of stage in the cascade classifier ($N_{cascade}$) in Algorithm 1, T is the number of filter in the stage (N_{filter}) in Algorithm 1, N_f is the number of features that is calculate by $N_{cell} * N_{block} * N_{bin}$ in Algorithm 2, N_T is the number state of CNNs or OpenPose, n_l is the number of convolution layer, C_l is the convolution layer complexity which equals to $s_l^2 \cdot n_l \cdot m_l^2$, s_l is the size of kernel filter, m_l is the size of pooling layer, and d is total of deep learning layers.

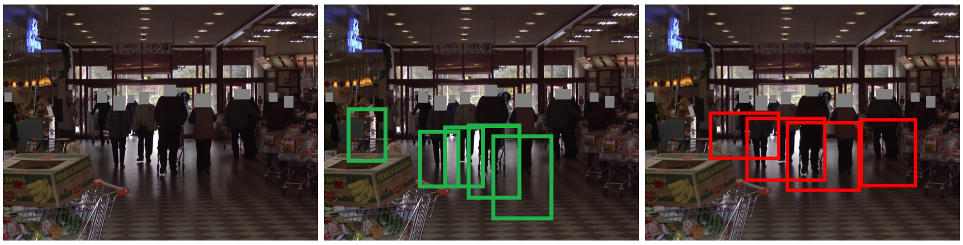
4.3. Different Image Conditions

Examples of results for cases 1–5 are illustrated in Figures 23–27, respectively. In most of these cases, detection is achieved by A-CNNs, OpenPose and HOG-SVM; however, the HOG-SVM framework cannot detect the person in the non-standing pose depicted in Figure 26. VJ_{Face} and VJ_{Upper} are able to achieve detection in cases 3–5, while in case 1, neither version of the VJ algorithm can detect any humans, as shown in Figure 23.



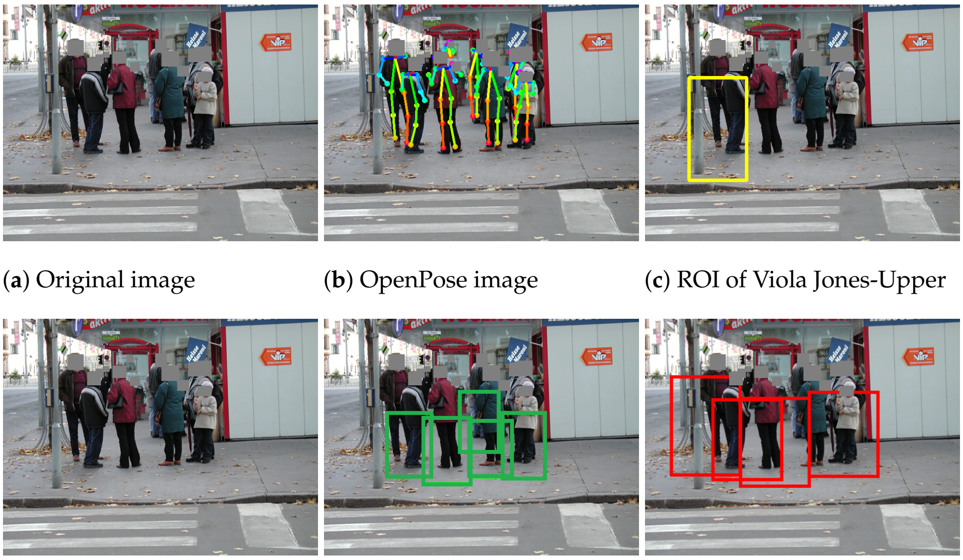
(**a**) Original image (**b**) OpenPose image (**c**) ROI of Viola Jones-Upper

Figure 23. *Cont.*



(**d**) ROI of Viola Jones-Face (**e**) ROI of HOG-SVM (**f**) ROI of A-CNNs

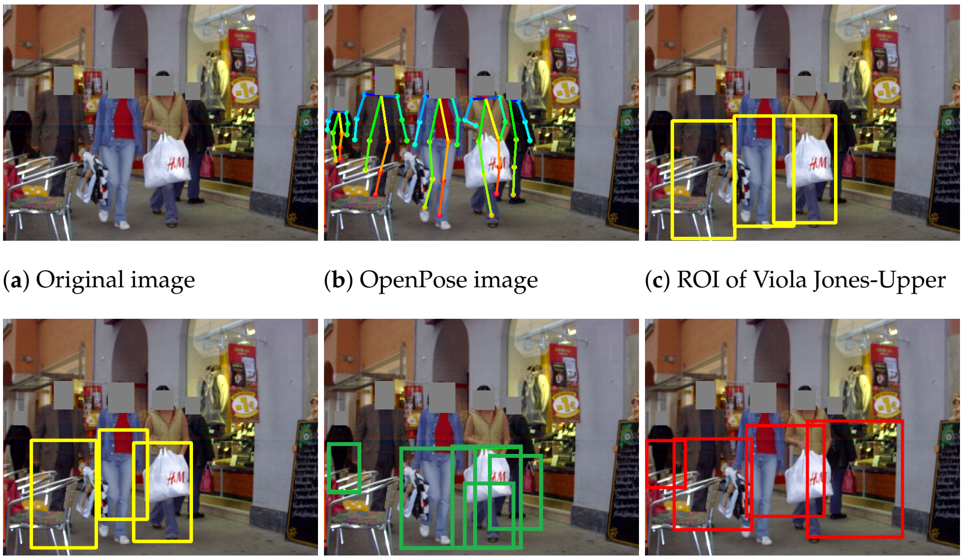
Figure 23. Case of challenging lighting conditions.



(**a**) Original image (**b**) OpenPose image (**c**) ROI of Viola Jones-Upper

(**d**) ROI of Viola Jones-Face (**e**) ROI of HOG-SVM (**f**) ROI of A-CNNs

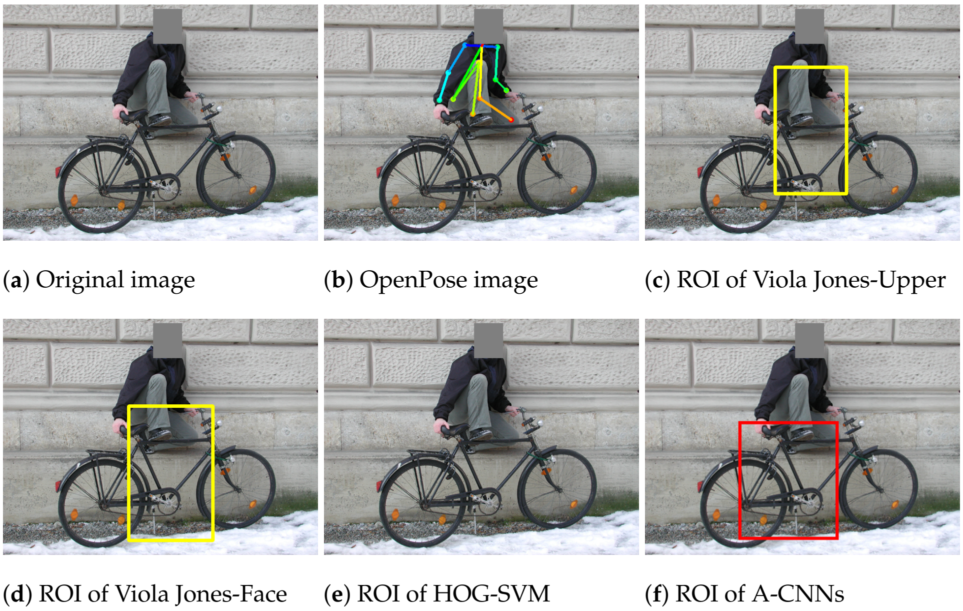
Figure 24. Case of occlusion.



(**a**) Original image (**b**) OpenPose image (**c**) ROI of Viola Jones-Upper

(**d**) ROI of Viola Jones-Face (**e**) ROI of HOG-SVM (**f**) ROI of A-CNNs

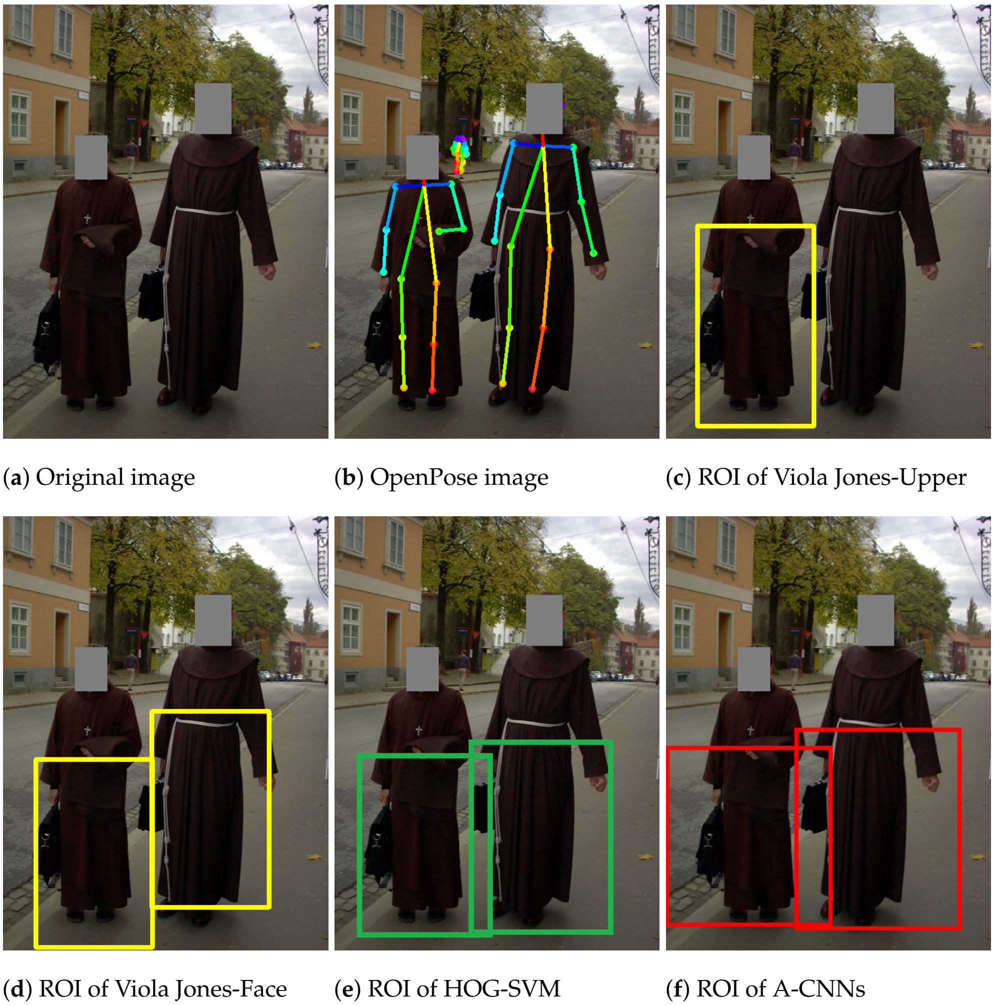
Figure 25. Case of multiple people.



(**a**) Original image (**b**) OpenPose image (**c**) ROI of Viola Jones-Upper

(**d**) ROI of Viola Jones-Face (**e**) ROI of HOG-SVM (**f**) ROI of A-CNNs

Figure 26. Case of a difference in pose between the training and test images.



(**a**) Original image (**b**) OpenPose image (**c**) ROI of Viola Jones-Upper

(**d**) ROI of Viola Jones-Face (**e**) ROI of HOG-SVM (**f**) ROI of A-CNNs

Figure 27. Case of challenging clothes.

5. Discussion

In this section, lower-body detection with the proposed anthropometric ratios is discussed from three perspectives: accuracy, complexity, and different image conditions.

According to the results in Table 5, the proposed anthropometric ratios can be used to scale other detected parts of the human body to obtain lower-body ROIs. In addition, the A-CNNs, OpenPose and HOG-SVM methods achieve success in lower-body detection with high sensitivities of more than 80% because they can successfully detect and transform human shapes under various lighting and occlusion conditions. Regarding specificity, the VJ_{Face} algorithm provides higher specificity than the other methods for detection on the non-human data set because most background images consist of scenes such as sightseeing locations and mountains; therefore, the Haar-like rectangular templates rarely match these backgrounds. Regarding the performance of A-CNNs, it is sometimes not fair to use such background data sets in deep learning unless the background images are further categorized into subclasses, such as trees, appliances and buildings. To enhance the detection performance, a similar problem has been solved by using a one-class classifier based on a CNNs [57]. OpenPose was trained on the COCO data set [58], which contains images of two hundred fifty thousand people with keypoints [45]. Consequently, it seemed to provide the highest detection accuracy on the INRIA data set in this experiment. However, our A-CNNs model, which was trained on human body ROIs, could achieve higher specificity than OpenPose, which was trained on keypoints. According to Figure 22, OpenPose seems to be more suitable for human detection in a plain room than for application in an outdoor environment for exercise monitoring. For the purpose of lower-body detection, lower-body ROIs are based on the proposed anthropometric ratios, while OpenPose focuses on body keypoints, which need more optimization than lower-body ROIs based on the NASA Anthropometry and Biomechanics data [21]. Moreover, the proposed anthropometric ratios can also be modified for use in locating different parts of the body, such as the thigh, leg, and foot, to monitor lower-body activities without any need to retrain the A-CNNs model. In contrast, OpenPose would need to be retrained on a data set containing new keypoints for the detection of different body parts.

According to the results in Table 6, the VJ_{Face} and VJ_{Upper} algorithms have relatively low complexity because they can reduce the complexity of the general algorithm from $O(N_rW)$ to $O(P+4N_r)$ based on the integral image calculation. Then, the complexity of the cascaded classifiers (with M filters and T thresholds) is $O(MT)$. Therefore, the final complexity of the VJ algorithm is $O((P+4N_r)MT)$. In the case of HOG-SVM, the complexity of the HOG feature calculation is $O(N_rW)$, which is multiplied by $O(N_f)$ (the number of features). The $O(N_f)$ complexity of the HOG calculation is high because of the iterations over N_{cell}, N_{block} and N_{bin}. If the dimensionality of the HOG features is not minimized to reduce the scale of $O(N_f)$, this algorithm might be too complex to be suitable for on-line detection. Subsequently, the complexity of the linear SVM is $O(N_f)$. Hence, the total complexity of HOG-SVM is $O(N_rWN_f^2)$. The A-CNNs method has a high classifier complexity, which depends on the number of convolutions. It also requires customizing a model to achieve high accuracy. The OpenPose method uses a CNNs model in its three main procedures for human detection. Moreover, the number of stages of each procedure influences the complexity of OpenPose. The number of stages should be optimized to achieve a suitable trade-off between accuracy and complexity [59].

In the case of different image conditions, although the proposed anthropometric ratios can be used to crop the lower-body regions of human images, this approach is limited to images of humans in a standing posture. In addition, five cases of human detection were discussed as:

1. Case of challenging lighting conditions: HOG-SVM, A-CNNs and OpenPose yield better detection results than VJ_{Face} and VJ_{Upper}. The former methods are not sensitive to lighting conditions when the features are in dark images.
2. Case of occlusion: The HOG-SVM, A-CNNs and OpenPose methods can detect overlapping humans in images. VJ_{Upper} is able to detect some of the human targets

in the image considered in this example, but VJ_{Face} is not because the faces of the humans in this image are rotated around the vertical axis.

3. Case of multiple people: Most methods can detect the lower bodies of the people in this image because the other characteristics of this image are beneficial, such as good lighting, full visibility of the upper bodies and a frontal view of the faces.
4. Case of a difference in pose between the training and test images: HOG-SVM cannot detect the lower-body of the person in this image because it depicts a human sitting on a bicycle and thus is not similar to the positive training images, i.e., standing human images. The A-CNNs uses the softmax function for classification, so the result is expressed in the form of a probability value expressing how close the input image is to the training images, whereas the OpenPose still can detect human keypoints because of the variety of postures used for training from the COCO data set [58].
5. Case of challenging clothes: The HOG-SVM, A-CNNs and OpenPose methods can detect the lower bodies of the people in this image because they still have human-looking shapes. VJ_{Face} can also detect the lower-body regions because there is no occlusion of the faces, while VJ_{Upper} can detect one of the two humans in the image.

6. Conclusions

This paper proposes anthropometric ratios for use in combination with either deep learning or traditional methods for lower-body detection in images captured under various environmental conditions. As seen from the results, the proposed framework can be beneficial for transforming some parts of the human body into corresponding lower-body ROIs; however, it is limited to images of humans in a standing posture captured from a frontal view only. Furthermore, in the deep learning methods, A-CNNs (90.14%) and OpenPose (95.82%) achieve higher accuracy than the averaged A-Traditional methods (74.81%) despite challenging illumination and occlusion conditions. However, the complexity of OpenPose, which depends on the number of nodes, layers, and stages, is higher than A-CNNs. In future work, anthropometric ratios suitable for various human postures will be studied. The specific data set provides the image conditions such as illumination conditions, occlusion, multiple people, the difference in posture, and a variety of clothes that will be tested. Furthermore, the A-CNNs model will be optimized its parameters for human body detection in a wide variety of scenarios. Additionally, the detection framework will be combined with a tracking system for faster monitoring of lower-body activities.

Author Contributions: Conceptualization, J.J., A.G.S.d.H. and R.D.; methodology, J.J.; software, J.J.; validation, J.J., A.G.S.d.H. and R.D.; formal analysis, J.J., A.G.S.d.H. and R.D.; investigation, J.J. and R.D.; resources, R.D.; data curation, J.J.; writing—original draft preparation, J.J.; writing—review and editing, A.G.S.d.H. and R.D.; visualization, R.D.; supervision, A.G.S.d.H. and R.D.; project administration, R.D.; funding acquisition, J.J. and R.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Coordinating Center for Thai Government Science and Technology Scholarship Students (CSTS), National Science and Technology Development Agency (NSTDA), Prince of Songkla University for Graduate School Dissertation Funding for Thesis and the Thailand Research Fund (TRF) and Thaicom under Grant RDG6150001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available INRIA Person Dataset [11] was analyzed in this study. This data can be found here: http://lear.inrialpes.fr/data, accessed on 2 May 2020.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Orr, R.M.; Dawes, J.J.; Lockie, R.G.; Godeassi, D.P. The Relationship Between Lower-Body Strength and Power, and Load Carriage Tasks: A Critical Review. *Int. J. Exerc. Sci.* **2019**, *12*, 1001–1022.
2. Nigro, F.; Bartolomei, S. A Comparison Between the Squat and the Deadlift for Lower Body Strength and Power Training. *J. Hum. Kinet.* **2020**, *73*, 145–152, doi:10.2478/hukin-2019-0139.
3. Pirsiavash, H.; Ramanan, D. Detecting activities of daily living in first-person camera views. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2847–2854, doi:10.1109/CVPR.2012.6248010.
4. Sepesy Maučec, M.; Donaj, G. Discovering Daily Activity Patterns from Sensor Data Sequences and Activity Sequences. *Sensors* **2021**, *21*, 6920, doi:10.3390/s21206920.
5. Tamantini, C.; Cordella, F.; Lauretti, C.; Zollo, L. The WGD—A Dataset of Assembly Line Working Gestures for Ergonomic Analysis and Work-Related Injuries Prevention. *Sensors* **2021**, *21*, 7600, doi:10.3390/s21227600.
6. Yun, J.; Lee, S.S. Human Movement Detection and Identification Using Pyroelectric Infrared Sensors. *Sensors* **2014**, *14*, 8057–8081, doi:10.3390/s140508057.
7. Dang, Q.; Yin, J.; Wang, B.; Zheng, W. Deep learning based 2D human pose estimation: A survey. *Tsinghua Sci. Technol.* **2019**, *24*, 663–676, doi:10.26599/TST.2018.9010100.
8. Wang, Q.; Markopoulos, P.; Yu, B.; Chen, W.; Timmermans, A. Interactive wearable systems for upper body rehabilitation: A systematic review. *J. NeuroEng. Rehabil.* **2017**, *14*, 20, doi:10.1186/s12984-017-0229-y.
9. Hamidi, M.; Osmani, A. Human Activity Recognition: A Dynamic Inductive Bias Selection Perspective. *Sensors* **2021**, *21*, 7278. doi:10.3390/s21217278.
10. Zabri Abu Bakar, M.; Samad, R.; Pebrianti, D.; Mustafa, M.; Abdullah, N.R.H. Computer vision-based hand deviation exercise for rehabilitation. In Proceedings of the 2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 27–29 November 2015; pp. 389–394, doi:10.1109/ICCSCE.2015.7482217.
11. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893, doi:10.1109/CVPR.2005.177.
12. Viola, P.; Jones, M.J. Robust Real-Time Face Detection. *Int. J. Comput. Vis.* **2004**, *57*, 137–154, doi:10.1023/b:visi.0000013087.49260.fb.
13. ElMaghraby, A.; Abdalla, M.; Enany, O.; Nahas, M.Y.E. Detect and Analyze Face Parts Information using Viola- Jones and Geometric Approaches. *Int. J. Comput. Appl.* **2014**, *101*, 23–28, doi:10.5120/17667-8494.
14. Koo, J.H.; Cho, S.W.; Baek, N.R.; Park, K.R. Face and Body-Based Human Recognition by GAN-Based Blur Restoration. *Sensors* **2020**, *20*, 5229, doi:10.3390/s20185229.
15. Moeslund, T.B.; Hilton, A.; Krüger, V.; Sigal, L. (Eds.) *Visual Analysis of Humans*; Springer: London, UK, 2011, doi:10.1007/978-0-85729-997-0.
16. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444, doi:10.1038/nature14539.
17. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. *Comput. Intell. Neurosci.* **2018**, *2018*, 7068349. doi:10.1155/2018/7068349.
18. Khan, S.; Khan, M.A.; Alhaisoni, M.; Tariq, U.; Yong, H.S.; Armghan, A.; Alenezi, F. Human Action Recognition: A Paradigm of Best Deep Learning Features Selection and Serial Based Extended Fusion. *Sensors* **2021**, *21*, 7941, doi:10.3390/s21237941.
19. Saleh, K.; Szenasi, S.; Vamossy, Z. Occlusion Handling in Generic Object Detection: A Review. In Proceedings of the 2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI), Herl'any, Slovakia, 21–23 January 2021; pp. 000477–000484, doi:10.1109/sami50585.2021.9378657.
20. Dong, J.; Zhang, L.; Zhang, H.; Liu, W. Occlusion-Aware GAN for Face De-Occlusion in the Wild. In Proceedings of the 2020 IEEE International Conference on Multimedia and Expo (ICME), London, UK, 6–10 July 2020; pp. 1–6, doi:10.1109/icme46284.2020.9102788. 6–10 July
21. National Aeronautics and Space Administration. Anthropometry and Biomechanics. Available online: https://msis.jsc.nasa.gov/sections/section03.htm (accessed on 19 January 2020).
22. Petrosova, I.; Andreeva, E.; Guseva, M. The System of Selection and Sale of Ready-to-Wear Clothes in a Virtual Environment. In Proceedings of the 2019 International Science and Technology Conference "EastConf", Vladivostok, Russia, 1–2 March 2019; pp. 1–5, doi:10.1109/eastconf.2019.8725390.
23. Realyvásquez-Vargas, A.; Arredondo-Soto, K.C.; Blanco-Fernandez, J.; Sandoval-Quintanilla, J.D.; Jiménez-Macías, E.; García-Alcaraz, J.L. Work Standardization and Anthropometric Workstation Design as an Integrated Approach to Sustainable Workplaces in the Manufacturing Industry. *Sustainability* **2020**, *12*, 3728, doi:10.3390/su12093728.
24. Rativa, D.; Fernandes, B.J.T.; Roque, A. Height and Weight Estimation From Anthropometric Measurements Using Machine Learning Regressions. *IEEE J. Transl. Eng. Health Med.* **2018**, *6*, 1–9, doi:10.1109/jtehm.2018.2797983.
25. Barrón, C.; Kakadiaris, I.A. Estimating Anthropometry and Pose from a Single Uncalibrated Image. *Comput. Vis. Image Underst.* **2001**, *81*, 269–284, doi:10.1006/cviu.2000.0888.
26. Sánchez-Muñoz, C.; Muros, J.J.; Cañas, J.; Courel-Ibáñez, J.; Sánchez-Alcaraz, B.J.; Zabala, M. Anthropometric and Physical Fitness Profiles of World-Class Male Padel Players. *Int. J. Environ. Res. Public Health* **2020**, *17*, 508, doi:10.3390/ijerph17020508.

27. Almasawa, M.O.; Elrefaei, L.A.; Moria, K. A Survey on Deep Learning-Based Person Re-Identification Systems. *IEEE Access* **2019**, *7*, 175228–175247, doi:10.1109/ACCESS.2019.2957336.
28. Grigorescu, S.; Trasnea, B.; Cocias, T.; Macesanu, G. A survey of deep learning techniques for autonomous driving. *J. Field Robot.* **2020**, *37*, 362–386, doi:10.1002/rob.21918.
29. Nieto-Hidalgo, M.; Ferrández-Pastor, F.J.; Valdivieso-Sarabia, R.J.; Mora-Pascual, J.; García-Chamizo, J.M. Gait Analysis Using Computer Vision Based on Cloud Platform and Mobile Device. *Mob. Inform. Syst.* **2018**, *2018*,7381264, doi:10.1155/2018/7381264.
30. Xu, R.; Ueno, S.; Kobayashi, T.; Makibuchi, N.; Naito, S. Human Area Refinement for Human Detection. In *Image Analysis and Processing—ICIAP 2015*; Murino, V., Puppo, E., Eds.; Springer International Publishing: Genova, Italy, 2015; pp. 130–141.
31. Kim, C.; Lee, J.; Han, T.; Kim, Y.M. A hybrid framework combining background subtraction and deep neural networks for rapid person detection. *J. Big Data* **2018**, *5*, 22, doi:10.1186/s40537-018-0131-x.
32. Zhou, X.; Liu, X.; Jiang, A.; Yan, B.; Yang, C. Improving Video Segmentation by Fusing Depth Cues and the Visual Background Extractor (ViBe) Algorithm. *Sensors* **2017**, *17*, 1177, doi:10.3390/s17051177.
33. Han, S.J.; Shin, J.S.; Kim, K.; Lee, S.Y.; Hong, H. Using Human Objects for Illumination Estimation and Shadow Generation in Outdoor Environments. *Symmetry* **2019**, *11*, 1266, doi:10.3390/sym11101266.
34. Iazzi, A.; Rziza, M.; Oulad Haj Thami, R. Fall Detection System-Based Posture-Recognition for Indoor Environments. *J. Imaging* **2021**, *7*, 42, doi:10.3390/jimaging7030042.
35. McIvor, A.; Zang, Q.; Klette, R. The Background Subtraction Problem for Video Surveillance Systems. In *Robot Vision*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 176–183, doi:10.1007/3-540-44690-7_22.
36. Zamalieva, D.; Yilmaz, A. Background subtraction for the moving camera: A geometric approach. *Comput. Vis. Image Underst.* **2014**, *127*, 73–85, doi:10.1016/j.cviu.2014.06.007.
37. Chiu, S.Y.; Chiu, C.C.; Xu, S.S.D. A Background Subtraction Algorithm in Complex Environments Based on Category Entropy Analysis. *Appl. Sci.* **2018**, *8*, 885, doi:10.3390/app8060885.
38. Mena, A.P.; Mayoral, M.B.; Díaz-Lópe, E. Comparative Study of the Features Used by Algorithms Based on Viola and Jones Face Detection Algorithm. In *International Work-Conference on the Interplay between Natural and Artificial Computation*; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2015; pp. 175–183, doi:10.1007/978-3-319-18833-1_19.
39. Adeshina, S.O.; Ibrahim, H.; Teoh, S.S.; Hoo, S.C. Custom Face Classification Model for Classroom Using Haar-Like and LBP Features with Their Performance Comparisons. *Electronics* **2021**, *10*, 102, doi:10.3390/electronics10020102.
40. Patel, C.I.; Labana, D.; Pandya, S.; Modi, K.; Ghayvat, H.; Awais, M. Histogram of Oriented Gradient-Based Fusion of Features for Human Action Recognition in Action Video Sequences. *Sensors* **2020**, *20*, 7299, doi:10.3390/s20247299.
41. He, M.; Luo, H.; Chang, Z.; Hui, B. Pedestrian Detection with Semantic Regions of Interest. *Sensors* **2017**, *17*, 2699, doi:10.3390/s17112699.
42. Yang, R.; Wang, Y.; Xu, Y.; Qiu, L.; Li, Q. Pedestrian Detection under Parallel Feature Fusion Based on Choquet Integral. *Symmetry* **2021**, *13*, 250, doi:10.3390/sym13020250.
43. Jammalamadaka, N.; Zisserman, A.; Jawahar, C.V. Human pose search using deep networks. *Image Vis. Comput.* **2017**, *59*, 31–43, doi:10.1016/j.imavis.2016.12.002.
44. Qin, Q.; Vychodil, J. Pedestrian Detection Algorithm Based on Improved Convolutional Neural Network. *J. Adv. Comput. Intell. Intell. Inform.* **2017**, *21*, 834–839, doi:10.20965/jaciii.2017.p0834.
45. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.E.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 172–186, doi:10.1109/tpami.2019.2929257.
46. Lin, C.B.; Dong, Z.; Kuan, W.K.; Huang, Y.F. A Framework for Fall Detection Based on OpenPose Skeleton and LSTM/GRU Models. *Appl. Sci.* **2021**, *11*, 329, doi:10.3390/app11010329.
47. BenAbdelkader, C.; Yacoob, Y. Statistical body height estimation from a single image. In Proceedings of the 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition, Amsterdam, The Netherlands, 17–19 September 2008; pp. 1–7, doi:10.1109/AFGR.2008.4813453.
48. Li, Z.; Jia, W.; Mao, Z.H.; Li, J.; Chen, H.C.; Zuo, W.; Wang, K.; Sun, M. Anthropometric body measurements based on multi-view stereo image reconstruction. In Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013; pp. 366–369, doi:10.1109/EMBC.2013.6609513.
49. Wong, Y.; Chen, S.; Mau, S.; Sanderson, C.; Lovell, B. ChokePoint Dataset. 2017. Available online: http://arma.sourceforge.net/chokepoint/ (accessed on 24 December 2021). doi:10.5281/ZENODO.815657.
50. Zhou, Q.; Wang, S.; Wang, Y.; Huang, Z.; Wang, X. AHP: Amodal Human Perception Dataset. Available online: https://sydney0zq.github.io/ahp/ (accessed on 24 December 2021).
51. Institute of Computer Graphics and Vision, University of Graz. Datasets. Available online: https://www.tugraz.at/institutes/icg/research/team-bischof/learning-recognition-surveillance/downloads/ (accessed on 24 December 2021).
52. Ding, W.; Hu, B.; Liu, H.; Wang, X.; Huang, X. Human posture recognition based on multiple features and rule learning. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 2529–2540, doi:10.1007/s13042-020-01138-y.
53. Yu, S.; Li, S.; Chen, D.; Zhao, R.; Yan, J.; Qiao, Y. COCAS: A Large-Scale Clothes Changing Person Dataset for Re-Identification. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 3397–3406, doi:10.1109/CVPR42600.2020.00346.

54. O'Brien, C.M. Statistics for Bioengineering Sciences: With MATLAB and WinBUGS Support by Brani Vidakovic. *Int. Stat. Rev.* **2013**, *81*, 471–472, doi:10.1111/insr.12042_12.
55. Tharwat, A. Classification assessment methods. *Appl. Comput. Inform.* **2021**, *17*, 168–192, doi:10.1016/j.aci.2018.08.003.
56. Chakrabarty, N.; Chatterjee, S. A Novel Approach to Age Classification from Hand Dorsal Images using Computer Vision. In Proceedings of the 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 27–29 March 2019; pp. 198–202, doi:10.1109/ICCMC.2019.8819632.
57. Ruff, L.; Vandermeulen, R.; Goernitz, N.; Deecke, L.; Siddiqui, S.A.; Binder, A.; Müller, E.; Kloft, M. Deep One-Class Classification. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; Volume 80, pp. 4393–4402.
58. Lin, T.Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. In Proceedings of the 13th European Conference in Computer Vision Part V (ECCV 2014), Zurich, Switzerland, 6–12 September 2014; pp. 740–755, doi:10.1007/978-3-319-10602-1_48.
59. Groos, D.; Ramampiaro, H.; Ihlen, E.A. EfficientPose: Scalable single-person pose estimation. *Appl. Intell.* **2020**, *51*, 2518–2533, doi:10.1007/s10489-020-01918-7.