Original Research Paper

Gender discrimination, age group classification and carried object recognition from gait energy image using fusion of parallel convolutional neural network

Newlin Shebiah Russel | Arivazhagan Selvaraj

Centre for Image Processing and Pattern Recognition, Department of Electronics and Communication Engineering, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India

Correspondence
R. Newlin Shebiah, Centre for Image Processing and Pattern Recognition, Department of Electronics and Communication Engineering, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India, 626-005.
Email: newlinshebiah@mepcoeng.ac.in

Abstract
Age and gender are the two key attributes for healthy social interactions, access control, intelligence marketing etc. Likewise, carried object recognition helps in identifying owner of the baggage being abandoned or the person littering in the public places. The above-mentioned surveillance task displays discriminative characteristics in gait. Primates can accomplish scene context understanding and reacting to different circumstances with varying reflexes with ease. Human beings achieve this by recollecting prior experiences and adapting to new situations quickly. Modelling the human behaviour, this research work has combined customized and learnable filters so that knowledge database can always be kept up to date, as well as, provides flexibility in learning new contexts. Thus, a specialized parallel deep convolutional neural network architecture with customized filters that extracts intrinsic characteristics and data driven learnable filters are fused to enhance the performance of single convolutional neural network is proposed. From the experimentation it is observed that, the learning is augmented when customized filters and learnable filters are fused together. Results show that the proposed system achieves better performance for CASIA B dataAQ2abase and OU-ISIR gait database-large population dataset with age and real-life carried object.

INTRODUCTION

Understanding human behaviour helps in predicting imminent abnormal activities and consequently assists the security agencies to react immediately. Analysing far-field video sequences for human abnormality detection seems to be a challenging problem because of low-resolution imagery. Thus an alternative choice other than facial features is required for analysing human behaviour at a distance. Gait recognition has emerged as an attractive biometric technology for the identification of people by analysing the way they walk even at a distance and in uncontrolled scenarios. However, the main challenges in real uncooperative environment is the effects of inherent intra-class variations caused by covariate factors such as change in view angles, walking speed variation, object carried condition and type of foot wear. Gait energy image (GEI) the averaged silhouettes of one gait period, integrates both static and dynamic components of human walking including motion frequency, temporal and spatial changes that reflects in the appearance of human body.

The state-of-the-art methodology reported in the literature can be categorized into model-based and model-free approaches. Model-based approaches [1,2] rely on the stride parameters that describe the gait by human body structure. Model based approach is computationally expensive and require high-resolution images. Computationally effective and simple model-free based methods emphasize on the motion parameters extracted from silhouettes. Here, the success lies in extracting discriminative features.

Deep learning methods learn discriminative feature representations or metrics directly from raw data and reported promising results for gait recognition. Wu et al.[3] learned pairwise similarities by a 3D convolutional neural network (CNN) with temporal input. Takemura et al.[4] further improved the pairwise similarity learning methods by optimizing the input,
output and loss functions of CNN-based methods on large scale cross-view datasets. GEINet [5] is an end-to-end network that can discover discriminative information for gait recognition by integrating segmentation and recognition modules. Yu et al. [6,7] constructed Generative adversarial networks that can transform gait images obtained from any viewpoint to the side view irrespective of the view point, carried condition and clothing. He et al.[8] proposed multi-task generative adversarial networks for learning view-specific feature representations with new gait template called period energy image which is an extension of GEI to enrich the spatial and temporal information in cross-view gait recognition.

In this paper, the problem of gender discrimination, age group classification and carried object recognition from gait images is resolved using an ensemble model with parallel multi-scale CNN.

The Contributions of this paper is summarized as follows:

- Pyramidal representation of the input GEI facilitates the CNN to learn the multi-scale signal effectively. It mimics the behaviour of human brain where different areas of brain respond to stimuli with different scales
- Parallel networks with varying number of convolution and pooling layers to automatically learn and recognize the most discriminative changes of gait features
- An extensive empirical evaluation of various tasks like gender discrimination, age group classification and carried object recognition with deep network architecture and score level fusion of customized filters and learnable filters at the layer 1 boost the classification accuracy to a great extent

The paper is structured as follows: Section 2 summarizes related work on gender discrimination, age group classification and carried object recognition. Section 3 describes the proposed method in detail. A comprehensive experimental study on gait large population dataset is detailed in Section 4. The concluding remarks of this paper are given in Section 5.

2 RELATED WORK

This section elaborates few state-of-the-art methods for gait based gender classification, gait based age estimation and carried object recognition.

2.1 Gait-based gender classification

Gender Classification is certainly a simple task for human. However, it is an active research problem due to its interesting applications in surveillance systems where gender specific restricted access to areas (train coach, lavatory), content based image retrieval, targeted marketing etc. Although human face contains important visual information for gender perception it requires the human subject's cooperation or attention and it is very difficult to perceive at long shots. Gender identification with gait signatures has added advantage as it is non-obtrusive and can be captured at a distance [9], without requiring cooperation or even awareness of the subject [10]. Yu et al. [11] learned local features by partitioning GEI into body parts like head, chest, and legs and classified the features using support vector machine (SVM). Lu and Tan [12] presented a view-invariant gait-based gender classification method with a subspace learning approach. Hu et al. [13] proposed a mixed conditional random field method for gait-based gender classification. As human tend to assume certain poses at each part of gait cycle, Issac et al. [14] proposed to delineate the gait instance as a sequence of poses or frames. Castro et al. [15] projected a CNN-based method with optical flow features for human identification in which, human gender is classified as supporting evidence for decision-making. Do et al., [16] used average gait image rather than a GEI and claimed to be computationally efficient and robust against view changes. Additionally, a viewpoint model and distance signal model is constructed to remove carried objects and worn coats from a silhouette to reduce the interference in the resulting classification. Sakata et. al. [17] proposed sequential multi-task learning where each CNN for non-target tasks is trained one by one in sequence and the CNN for the target task is trained last. Since gender classification is easiest compared with age estimation, gender classification is done first.

2.2 Gait based age estimation

Most of the studies on age estimation focused on the face images, which tend to become more wrinkled and sag with age [18,19]. Like Wrinkles, skin, swiftness, head to body ratio, gait and pitch or heaviness of the speech define the baselines for the age estimation. Davis [20] showed the gait difference between an adult and a child in terms of leg length, stride width, and stride frequencies. Ince et al. [21] showed that the head-to-body ratio of a child is different from an adult. Similar findings were also reported in [22] by analysing the most widely used appearance-based gait representation known as GEIs (GEIs) or averaged silhouettes [23], which contains both gait and shape information. There are obvious changes in head-to-body ratio as children grow and in addition, as people get older, a middle-aged spread and stoop appear. Hence, varying appearances in gait representation serve as clues in gait-based human age estimation.

Makihara et al. [24] learned an age regression model for each age group using support vector regression with a Gaussian kernel in conjunction with a manifold learning technique. Lu et al. [25] proposed a multi label-guided subspace to better characterize and correlate age and gender information, because gait features vary depending on a subject's attributes. Sakata [17] proposed three CNN stages: CNN for gender estimation, CNN for age-group estimation and a CNN for age regression. The OU-ISIR gait database comprising the large population dataset with age and gender details is created and its performance is evaluated using benchmarks like Gaussian process regression,
support vector regression, multilabel-guided subspace and ordinary preserving manifold analysis [26].

2.3 Gait based carried object recognition

Several prior approaches have been projected in the literature for detecting carried object, as it is a key component in varied security applications. Haritaoglu et. al. [27] proposed backpack detection framework with the presumption that shape of the human body is symmetric, and individuals display periodic motion while they are moving unhampered. Hydra a real-time framework for detecting and tracking multiple people was proposed by Haritaoglu et. al. [28] that combines head detection by projection histogram and correlation-based matching methods to segment multiple people in a crowd and track them. W4 [29] learns and creates people’s appearance model by locating their parts like head, hands, feet and torso and these body parts can even be tracked through interactions and occlusions. It can also determine whether a person is carrying an object, and segments the object so it can be tracked during exchanges.

View and carried object category invariant detection is by detecting protrusions and comparing with the several exemplars, corresponding to different views of binary silhouettes [30]. Along with temporal templates, the colour information is fused to reduce the false positive detections [31]. The carried object detection is formulated as a superpixel classification problem using the codebook [32]. A human-baggage detector is modelled as the mixture of body parts as head, torso, leg, and baggage parts [33]. Ben Abdelkader [34] portrayed a technique to recognize person carrying an object using motion-based recognition approach that integrates shape and periodicity cues of the human silhouette. The deviation of the periodicity and the amplitude indicates the person carrying an object.
3 | PROPOSED METHODOLOGY

The parallel architecture proposed for gender discrimination, age estimation and carried condition classification using GEIs is shown in Figure 1. The input GEIs are weighted using a Gaussian average and scaled down by pyramid representation. Apart from the input image of size $p \times q$, the parallel architecture receives the down sampled versions of the input image that is, $\left(\frac{p}{2}, \frac{q}{2}\right)$ and $\left(\frac{p}{4}, \frac{q}{4}\right)$. Parallel architecture is designed by minimizing the input image size and the associated number of convolutional layers to keep the number of connections between the last convolutional layer and the fully connected layer fixed. Convolution neural network consists of a stack of layers that takes an input image of size ($h \times w \times c$) and predicts the class or label probabilities at the output.

The input image used for analysis from GEIs is normalized to the size (128, 88, 1). Here, parallel CNNs are trained from the scratch with progressively reduced input size by the factor of 2 thus the CNN A receives the input of size $128 \times 88$ pixels and second net CNN B $64 \times 44$ and third net CNN C with $32 \times 22$. The number of convolution and pooling layers of each CNNs are engineered to keep the input to the fully connected layer fixed ($128$ feature maps of size $4 \times 2$).

For gender discrimination the indicative of walker’s gender are posture sway, waist-hip ratio, and shoulder-hip ratio [35]. Males have a tendency to swing their shoulders and females tend to swing their hips more than their shoulders while walking. Further, males have broader shoulders and females usually have thin waists and wider hips. For age estimation, gait characteristics like regularity, stability and pace that distinguishes young and old adults. Moreover, head-to-body ratio and presence of stoop are the indications from gait characteristics to predict the age of a person. The appearance and dynamics of the person get varied that when carrying objects and is well visualized from GEI.

In all the mentioned applications, the high frequency components of the GEI play a vital role. Thus the weights of the initial convolutional layer are handcrafted so as to detect specific features for gender discrimination, age group classification and carried object recognition.

The 8 filters used at the initial layer of customized stream is shown in Figure 2. Deep learning aims at learning feature automatically from data at multiple levels of abstraction and maps input to the output directly without depending on human-crafted features. Thus, the fusion of stream with customized weights and learnable weights at first convolutional layer aimed at boosting the performance characteristics.

The training of all CNNs in this work has been carried out by optimizing the cross-entropy objective function using the mini-batch Nesterov’s accelerated gradient descent. Backpropagation of the gradient has been performed with an initial learning rate of 0.001 and the momentum of 0.9. The size of a mini-batch has been set to 28 with 30 epochs. The outputs from the six instances are combined in a single ensemble by score level fusion of the corresponding softmax layers.

The number of learnable parameters defines the complexity of the CNN model. To calculate the learnable parameters in the convolutional layer, the kernel size is multiplied with the number of kernels and the bias term is added with it. For example, the first layer has 8 filters with kernel size $3 \times 3$, thus the learnable is $80 (3 \times 3 \times 8) + 8$. Since back propagation is not involved in pooling layer it has no learnable parameters. Batch normalization acts to standardize the mean and variance of each unit and hence adds two trainable parameters for each depth. Fully connected layer has highest number of learnable parameters since every neuron in the previous layer and current layer needs to be connected. The summary of the number of learnable parameters for every layer for CNN A, CNN B and CNN C is tabulated in Table 1.

4 | RESULTS AND DISCUSSION

In this section, robustness of the parallel CNN is evaluated for gender discrimination, age group classification and carried object classification with CASIA B database, OU-ISIR gait database large population dataset with age (OULP-Age)” and OU-ISIR large population gait database with real-life carried object (OU-LP-Bag).

4.1 | CASIA B database

The CASIA gait dataset B (CASIA B) [36] encompasses GEIs of 124 human taken from 11 different view angles ($0^\circ$, $18^\circ$, $36^\circ$, $54^\circ$, $72^\circ$, $90^\circ$, $108^\circ$, $126^\circ$, $144^\circ$, $162^\circ$ and $180^\circ$). There are 10 walking sequences for each subject under each view. The 10 walking sequences include 6 normal walking sequences, 2 carrying-bag sequences and 2 wearing-coat sequences. Sample GEIs under 11 different viewing angles for carried object are shown in Figure 3.

4.2 | OU-ISIR gait database, large population dataset with age (OULP-Age)

The “OU-ISIR gait database, large population dataset with age (OULP-Age)” [26] was used to assess the performance of the proposed gender discrimination and age estimation methodology. OULP-Age is the collection of 63,846 gait images (31,093 males and 32,753 females) with age and gender information. The dataset comprises of GEIs of size $88 \times 128$ pixels with
| Layer | Rep. | Input | Filter | Output | Learnable |
|-------|------|-------|--------|--------|-----------|
| L1    | Input layer | 128 × 88 × 1 | 128 × 88 × 1 | 128 × 88 × 1 | 128 × 88 × 1 |
| L2    | Conv_1 | 128 × 88 × 1 | 8,3 × 3 | 128 × 88 × 8 | 80 |
| L3    | Conv_2 | 128 × 88 × 8 | 8,3 × 3 | 128 × 88 × 32 | 2336 |
|       | Batch_1 | 128 × 88 × 32 | 128 × 88 × 32 | 64 |
|       | Norm1 | 128 × 88 × 32 | 128 × 88 × 32 | 64 |
|       | ReLu_1 | 128 × 88 × 32 | 128 × 88 × 32 | 64 |
|       | Maxpool_1 | 128 × 88 × 32 | 128 × 88 × 32 | 64 |
| L4    | Conv_3 | 64 × 44 × 32 | 32,3 × 3 | 64 × 44 × 32 | 2248 |
|       | Batch_1 | 64 × 44 × 32 | 64 × 44 × 32 | 64 |
|       | Norm2 | 64 × 44 × 32 | 64 × 44 × 32 | 64 |
|       | ReLu_2 | 64 × 44 × 32 | 64 × 44 × 32 | 64 |
|       | Maxpool_2 | 64 × 44 × 32 | 64 × 44 × 32 | 64 |
| L5    | Conv_4 | 32 × 22 × 32 | 64,3 × 3 | 32 × 22 × 32 | 1498 |
|       | Batch_1 | 32 × 22 × 32 | 128 |
|       | Norm3 | 32 × 22 × 32 | 128 |
|       | ReLu_3 | 32 × 22 × 32 | 32 × 22 × 32 | 128 |
|       | Maxpool_3 | 32 × 22 × 32 | 32 × 22 × 32 | 128 |
| L6    | Conv_5 | 16 × 11 × 64 | 128,3 × 3 | 16 × 11 × 128 | 73556 |
|       | Batch_1 | 16 × 11 × 128 | 16 × 11 × 128 | 256 |
|       | Norm4 | 16 × 11 × 128 | 16 × 11 × 128 | 256 |
|       | ReLu_4 | 16 × 11 × 128 | 16 × 11 × 128 | 256 |
|       | Maxpool_4 | 16 × 11 × 128 | 16 × 11 × 128 | 256 |
| L7    | Conv_6 | 8 × 5 × 128 | 128,3 × 3 | 8 × 5 × 128 | 147584 |
|       | Batch_1 | 8 × 5 × 128 | 8 × 5 × 128 | 256 |
|       | Norm5 | 8 × 5 × 128 | 8 × 5 × 128 | 256 |
|       | ReLu_5 | 8 × 5 × 128 | 8 × 5 × 128 | 256 |
|       | Maxpool_5 | 8 × 5 × 128 | 8 × 5 × 128 | 256 |
| L8    | Fully connected | 4 × 2 × 128 | 1 × 1 × # | 5125 |
|       | 1 × 1 × # | 5125 |
| L9    | SoftMax | 1 × 1 × # | 1 × 1 × # | 5125 |
|       | Class output | 1 × 1 × # | 1 × 1 × # | 5125 |

**TABLE 1** Summary of learnable parameter in CNN A, CNN B and CNN C
age ranges from 2 to 90 years. Figure 4 shows some sample gait energy images from OULP-age dataset with varying ages and gender. The age group class for age group estimation by ensemble of CNNs is divided based on the human height. The kids and teens have rapid growing phase compared with adult and thus, finer scale of age interval for the people is allocated below 15 years. The people in early adulthood and midlife shows stabilized height thus grouped under a single category. The people in late adulthood experience a loss of height due to the compression of spine, and softening of muscle and bone tissues and hence form another category.

The performance of the proposed carried object recognition is assessed using OU-ISIR large population gait database with real-life carried object (OU-LP-Bag) [37]. The OU-LP-Bag dataset is meant for vision-based gait recognition with carried objects (COs) and to estimate the position of the carried object. The dataset includes size-normalized (i.e. $128 \times 88$) GEIs annotated with seven class labels based upon the carried object with respect to the human body.

Some of the sample human blobs and the GEIs with varying carried state are shown in Figure 5. The categories include: subjects without any carried object (NoCO), subjects with COs in multiple body parts (MuCO), subjects changing the position of CO's within a GEI gait period (CpCO), the subjects with COs at the front (FrCO) and back regions (BaCO), side bottom region (SbCO) and side middle region (SmCO).
TABLE 2  Dataset details of pedestrian with varied carrying condition

| Carried Condition                  | Representation | Number of Samples |
|-----------------------------------|----------------|-------------------|
| No carried object                 | NoCO           | 10323             |
| Object in front region            | FrCO           | 10255             |
| Object in back region             | BaCO           | 10862             |
| Object in side bottom region      | ShCO           | 6736              |
| Object in side middle region      | SmCO           | 3392              |
| Object in multiple regions        | MuCO           | 16224             |
| Object with position being changed from one region to another within a gait period | GpCO | 2657 |

The database comprises of about 60,450 samples under seven varying carried condition and the number GEI under each label is displayed in Table 2.

4.4  Gender discrimination

In CASIA B, database there are 124 subjects including 93 males and 31 females. Since gender classification is a two-class problem, it is better to use equal number of subjects for different genders to avoid bias. As of the state-of-the-art works, all the 31 females and 31 males are used for experimental analysis.

Table 3 shows the performance measures like precision, specificity, recall, false positive rate, and accuracy for gender discrimination using CASIA B database. The discriminating features are learned simultaneously by six parallel CNNs with the input size progressively reduced. The proposed method with multiple CNN stages is trained more intensively and efficiently for the gender discrimination task. From the result, it is observed that, the decision level fusion of the scores from the characteristics features trained by CNN A, CNN B and CNN C yields virtuous results. CNN with learnable filters achieves marginal increase in performance compared with customized filters given the flexibility to adapt to the changing scenario.

In Yu et al [11], GEI is parsed into body parts like head, chest and legs and classified the features using SVM reporting 95.97% accuracy. Zhang et al. [38] represented the GEI as Tensor with multi-linear principal component analysis reporting 98.1% accuracy. Do et al. [39] used aggregate GEI with SVM reporting 98.8%. Isaac et al. [40] obtained 100% accuracy with Elliptic Fourier descriptors and linear discriminant analysis with Bayes’ rule.

For OU-ISIR gait database, large population dataset with age (OULP-Age), collection of 63,846 gait images (31,093 males

TABLE 3  Performance measure for gender discrimination using CASIA B database

|                                      | Precision | Specificity | Recall | FPR  | Accuracy |
|--------------------------------------|-----------|-------------|--------|------|----------|
| CNN A _ customized                   | 99.42     | 99.43       | 99.88  | 0.57 | 99.65    |
| CNN B _ customized                   | 99.77     | 99.77       | 99.42  | 0.23 | 99.59    |
| CNN C _ customized                   | 99.53     | 99.50       | 92.91  | 0.50 | 96       |
| CNN A & CNN B (customized)          | 100       | 100         | 100    | 0    | 100      |
| CNN B & CNN C (customized)          | 99.88     | 99.88       | 99.53  | 0.12 | 99.71    |
| CNN A & CNN C (customized)          | 99.88     | 99.88       | 99.65  | 0.12 | 99.77    |
| CNN A, CNN B & CNN C (customized)  | 100       | 100         | 99.88  | 0    | 99.94    |
| CNN A _ learnable filter             | 99.77     | 99.77       | 99.88  | 0.23 | 99.83    |
| CNN B _ learnable filter             | 99.65     | 99.65       | 99.65  | 0.35 | 99.65    |
| CNN C _ learnable filter             | 99.42     | 99.40       | 95.40  | 0.60 | 97.33    |
| CNN A & CNN B (learnable)           | 100       | 100         | 100    | 0    | 100      |
| CNN B & CNN C (learnable)           | 99.88     | 99.89       | 100.00 | 0.11 | 99.94    |
| CNN A & CNN C (learnable)           | 100       | 100         | 99.88  | 0    | 99.94    |
| CNN A, CNN B & CNN C (learnable)   | 100       | 100         | 100    | 0    | 100      |
| CNN A _customized + CNN A _learnable| 99.77     | 99.77       | 100    | 0.23 | 99.88    |
| CNN B _customized + CNN B _learnable| 100       | 100         | 100    | 0    | 100      |
| CNN C _customized + CNN C _learnable| 99.77     | 99.76       | 97.60  | 0.24 | 98.67    |
| CNN A (learnable & customized) + CNN B (learnable & customized) + CNN C (learnable & customized) | 100 | 100 | 100 | 0 | 100 |

Yu et al. (2009) [11] – – – – 95.97
Zhang et al. (2015) [38] – – – – 98.10
Do et al. (2019) [39] – – – – 98.80
Isaac et al. (2019) [40] – – – – 100.00
TABLE 4  Performance measure for gender discrimination using OU-ISIR gait database

| Model Description                                      | Precision | Specificity | Recall  | FPR   | Accuracy |
|--------------------------------------------------------|-----------|-------------|---------|-------|----------|
| CNN A _ customized                                     | 99.33     | 99.45       | 95.79   | 0.55  | 97.76    |
| CNN B _ customized                                     | 98.99     | 99.17       | 95.31   | 0.83  | 97.38    |
| CNN C _ customized                                     | 95.13     | 96.18       | 97.75   | 3.82  | 96.86    |
| CNN A & CNN B (customized)                             | 100.00    | 100.00      | 98.51   | 0     | 99.33    |
| CNN B & CNN C (customized)                             | 99.66     | 99.73       | 99.50   | 0.27  | 99.63    |
| CNN A & CNN C (customized)                             | 100       | 100         | 98.51   | 0     | 99.33    |
| CNN A, CNN B & CNN C (customized)                      | 100       | 100         | 99.33   | 0     | 99.70    |
| CNN A _ learnable filter                                | 98.66     | 98.90       | 96.55   | 1.10  | 97.83    |
| CNN B _ learnable filter                                | 94.96     | 96.02       | 96.58   | 3.98  | 96.26    |
| CNN C _ learnable filter                                | 94.96     | 95.96       | 94.96   | 4.04  | 95.52    |
| CNN A & CNN B (learnable)                              | 99.66     | 99.73       | 99.66   | 0.27  | 99.70    |
| CNN B & CNN C (learnable)                              | 98.82     | 99.06       | 98.82   | 0.94  | 98.95    |
| CNN A & CNN C (learnable)                              | 99.66     | 99.73       | 99.16   | 0.27  | 99.48    |
| CNN A, CNN B & CNN C (learnable)                       | 100       | 100         | 99.83   | 0     | 99.93    |
| CNN A _ customized + CNN A _ learnable                  | 99.83     | 99.86       | 98.67   | 0.14  | 99.33    |
| CNN B _ customized + CNN B _ learnable                  | 99.83     | 99.86       | 99.17   | 0.14  | 99.55    |
| CNN C _ customized + CNN C _ learnable                  | 98.99     | 99.20       | 99.49   | 0.80  | 99.33    |
| CNN A (learnable & customized) + CNN B (learnable & customized) + CNN C (learnable & customized) | 100 | 100 | 100 | 0 | 100 |
| Sakata et al. (2019) [17]                              | 97.42     | 97.49       | 98.01   | 2.51  | 97.74    |

4.5 Gender specific age estimation

This sub section focuses on performance of age group classification after gender discrimination by dividing the age group into 5 age groups. Six parallel CNNs are trained from the base with varying image resolution and it is observed that specific networks had better favor certain categories.

For gender specific age group classification, female class CNN A customized outperforms CNN B learnable stream, whereas, in other cases CNN learnable stream performs better than CNN customized. For male gender, CNN customized learnable stream performs better than CNN learnable. Additionally, the average recognition rate of male gender is larger than that of female gender in gender specific age group classification. This is due to the fact that, female show a large range of appearance variations due to variations in dress type, hair style and foot wear and this is greater than that in males.

Further, while classifying people in age group between (0–5) greater confusion occur between group (0–5) and (6–10) age groups since some of the samples share the characteristics from the adjacent age group. Significant confusion is reported between group (0–5), (6–10) and (11–15) age groups while processing for (6–10) age group category. Additionally, confusion between the group (>60) and (16–60) pull down the accuracy of the proposed architecture to some extent.

Table 5 exhibits the comparison plot of the proposed architecture with Sakata et al. [17] and Li et al. [24] for gender specific age group classification and it implies that the proposed method outperforms the existing method with a large margin.

4.6 Carried object recognition

In this section, classification of carried condition based on the gait feature is analysed. There are copious applications such as intrusion of person with baggage in prohibited area, tracing person with a knapsack etc. In CASIA B database, each 124 person has 2 sequence with carried condition viewed at 11 different orientations. Here, the carried state is classified as a two class problem carrying object or not. The performance measures for carried object detection for CASIA B database is shown in Table 6.

In case of OSU database, for evaluating the performance of the proposed architecture, the number of subjects under each
TABLE 5  Performance measure for age group classification using OU ISIR gait database

| Model Type | Female | Male | Combined Male and Female |
|------------|--------|------|-------------------------|
| CNN A _ customized | | | |
| | | | |
| CNN B _ customized | | | |
| | | | |
| CNN C _ customized | | | |
| | | | |
| CNN A & CNN B (customized) | | | |
| CNN B & CNN C (customized) | | | |
| CNN A & CNN C (customized) | | | |
| CNN A, CNN B & CNN C (customized) | | | |
| CNN A_ learnable filter | | | |
| CNN B_ learnable filter | | | |
| CNN C_ learnable filter | | | |
| CNN A & CNN B (learnable) | | | |
| CNN B & CNN C (learnable) | | | |
| CNN A & CNN C (learnable) | | | |
| CNN A & CNN B & CNN C (learnable) | | | |

Sakata et al. (2019) [17] | | | |
Li et al. (2018) [24] | | | |
TABLE 6  Performance measure for carried object recognition for CASIA B database

|                  | Precision | Specificity | Recall | FPR | Accuracy |
|------------------|-----------|-------------|--------|-----|----------|
| CNN A _ customized | 91.90     | 93.35       | 98.27  | 6.65| 95.49    |
| CNN B _ customized | 89.99     | 91.91       | 98.39  | 8.09| 94.66    |
| CNN C _ customized | 88.51     | 90.48       | 94.73  | 9.52| 91.79    |
| CNN A & CNN B (customized) | 98.09     | 98.37     | 100    | 1.63| 99.11    |
| CNN B & CNN C (customized) | 94.99     | 95.83     | 99.85  | 4.17| 97.61    |
| CNN A & CNN C (customized) | 96.02     | 96.65     | 99.54  | 3.35| 97.95    |
| CNN A, CNN B & CNN C (customized) | 96.91     | 97.38   | 99.85  | 2.62| 98.50    |
| CNN A _ learnable filter | 95.29 | 96.00     | 97.88  | 4   | 96.85    |
| CNN B _ learnable filter | 88.81   | 91.12 | 99.50  | 8.88| 94.60    |
| CNN C _ learnable filter | 87.04   | 89.70 | 97.20  | 10.30| 92.82    |
| CNN A & CNN B (learnable) | 97.64     | 98     | 100    | 2   | 98.91    |
| CNN B & CNN C (learnable) | 95.43     | 96.17     | 99.23  | 3.83| 97.54    |
| CNN A & CNN C (learnable) | 96.47     | 97.00     | 99.09  | 3   | 97.95    |
| CNN A, CNN B & CNN C (learnable) | 97.20     | 97.63     | 100    | 2.37| 98.70    |
| CNN A customized + CNN A learnable | 98.09     | 98.36     | 99.85  | 1.64| 99.04    |
| CNN B customized + CNN B learnable | 95.73     | 96.42     | 99.85  | 3.58| 97.95    |
| CNN C customized + CNN C learnable | 95.29     | 96.00     | 97.88  | 4   | 96.85    |
| CNN A (learnable & customized) + CNN B (learnable & customized) + CNN C (learnable & customized) | 99.56 | 99.62 | 100 | 0.38 | 99.79 |

Leon et al, 2014 [42]

TABLE 7  Confusion matrix CNN A learnable

|        | BaCO | CpCO | FrCO | MuCO | NoCO | SbCO | SmCO |
|--------|------|------|------|------|------|------|------|
| BaCO   | 85.22| 0.53 | 0.27 | 5.99 | 2.93 | 0.40 | 4.66 |
| CpCO   | 5.57 | 41.27| 25.30| 5.42 | 4.37 | 7.98 | 10.09|
| FrCO   | 0.13 | 5.17 | 84.35| 0.80 | 3.98 | 2.25 | 3.32 |
| MuCO   | 16.18| 2.25 | 9.81 | 58.22| 2.12 | 4.91 | 6.50 |
| NoCO   | 0.79 | 0.53 | 3.43 | 0.13 | 86.53| 5.02 | 3.57 |
| SbCO   | 0.66 | 0.92 | 4.49 | 1.06 | 18.23| 68.82| 5.81 |
| SmCO   | 10.85| 5.07 | 16.04| 2.95 | 13.21| 9.32 | 42.57|

As for the classification accuracy of each label, NoCO, FrCO and BaCO worked well because there was no carried object in NoCO, the shape and position of the carried object were stable in BaCO and in FrCO, the appearance change is well grasped and can be easily distinguished from other labels. However, SbCO was considerably confused with NoCO because of the shape similarity with respect to the upper part of the GEIs and with SmCO due to slight variation in the carried position. The classes of SmCO, MuCO and CpCO, the GEI features were unstable, hence sometimes samples of these labels were misclassified as other labels. Because of the occlusion of COs with the subject’s body for SmCO, the GEI feature was confused with that of SbCO, NoCO, MuCO and BaCO, depending on the part of the COs that was occluded. Similarly, for the case of MuCO, it was confused with BaCO, subjects typically carried, for example, a backpack in the back region together with a small object.
TABLE 8  Confusion matrix CNN A customized

|       | BaCO | CpCO | FrCO | MuCO | NoCO | SbCO | SmCO |
|-------|------|------|------|------|------|------|------|
| BaCO  | 81.76| 2.00 | 0.27 | 8.92 | 3.99 | 0.40 | 2.66 |
| CpCO  | 3.77 | 64.91| 10.54| 6.48 | 3.31 | 4.97 | 6.02 |
| FrCO  | 0.40 | 9.95 | 69.50| 3.45 | 6.90 | 3.58 | 6.23 |
| MuCO  | 15.12| 4.77 | 7.43 | 63.79| 1.86 | 3.05 | 3.98 |
| NoCO  | 1.19 | 1.06 | 1.85 | 0.92 | 86.13| 6.87 | 1.98 |
| SbCO  | 0.53 | 2.91 | 1.45 | 0.53 | 14.66| 74.64| 5.28 |
| SmCO  | 11.56| 14.39| 7.19 | 5.07 | 9.55 | 9.20 | 43.04|

TABLE 9  Confusion matrix CNN B learnable

|       | BaCO | CpCO | FrCO | MuCO | NoCO | SbCO | SmCO |
|-------|------|------|------|------|------|------|------|
| BaCO  | 74.57| 2.00 | 0.53 | 5.46 | 2.93 | 0.27 | 14.25|
| CpCO  | 4.07 | 62.50| 10.09| 3.92 | 0.75 | 2.56 | 16.11|
| FrCO  | 0.13 | 10.08| 69.89| 0.93 | 3.05 | 1.72 | 14.19|
| MuCO  | 12.86| 6.63 | 10.34| 53.71| 0.80 | 1.86 | 13.79|
| NoCO  | 0.92 | 1.19 | 3.70 | 0.79 | 76.22| 1.98 | 15.19|
| SbCO  | 0.26 | 5.94 | 3.70 | 1.59 | 11.76| 47.69| 29.06|
| SmCO  | 5.42 | 8.37 | 7.31 | 1.42 | 5.54 | 2.59 | 69.34|

TABLE 10  Confusion matrix CNN B customized

|       | BaCO | CpCO | FrCO | MuCO | NoCO | SbCO | SmCO |
|-------|------|------|------|------|------|------|------|
| BaCO  | 79.49| 1.46 | 0.13 | 9.05 | 3.6  | 1.33 | 4.93 |
| CpCO  | 3.61 | 54.22| 18.67| 8.43 | 3.31 | 5.87 | 5.87 |
| FrCO  | 0.27 | 6.37 | 83.55| 0.93 | 3.98 | 2.52 | 2.39 |
| MuCO  | 15.38| 1.99 | 9.02 | 60.74| 1.46 | 5.44 | 7.15 |
| NoCO  | 1.06 | 0.92 | 2.64 | 0.26 | 86.92| 6.34 | 1.85 |
| SbCO  | 0.66 | 1.98 | 4.23 | 0.79 | 17.44| 70.54| 4.36 |
| SmCO  | 10.26| 9.55 | 15.8 | 4.72 | 13.21| 11.08| 35.38|

TABLE 11  Confusion matrix CNN C learnable

|       | BaCO | CpCO | FrCO | MuCO | NoCO | SbCO | SmCO |
|-------|------|------|------|------|------|------|------|
| BaCO  | 74.83| 1.33 | 0.00 | 9.99 | 1.73 | 1.07 | 11.05|
| CpCO  | 3.46 | 52.71| 12.05| 7.08 | 1.66 | 8.13 | 14.91|
| FrCO  | 0.27 | 9.68 | 68.70| 1.33 | 2.52 | 4.91 | 12.60|
| MuCO  | 11.27| 3.45 | 6.90 | 61.14| 0.27 | 5.44 | 11.54|
| NoCO  | 1.59 | 1.45 | 1.45 | 0.66 | 64.20| 19.42| 11.23|
| SbCO  | 0.40 | 2.38 | 1.32 | 0.40 | 8.32 | 80.32| 6.87 |
| SmCO  | 6.96 | 7.55 | 6.60 | 2.83 | 5.07 | 12.03| 58.96|

TABLE 12  Confusion matrix CNN C customized

|       | BaCO | CpCO | FrCO | MuCO | NoCO | SbCO | SmCO |
|-------|------|------|------|------|------|------|------|
| BaCO  | 79.49| 0.93 | 0.13 | 9.59 | 3.06 | 1.2  | 5.59 |
| CpCO  | 5.42 | 40.51| 20.78| 9.34 | 3.46 | 7.68 | 12.8 |
| FrCO  | 0.4  | 4.91 | 73.21| 2.79 | 4.38 | 3.71 | 10.61|
| MuCO  | 13.38| 1.99 | 9.02 | 60.74| 1.46 | 4.38 | 7.03 |
| NoCO  | 1.45 | 1.06 | 2.11 | 0.79 | 72.39| 14.53| 7.66 |
| SbCO  | 0.4  | 1.85 | 2.64 | 1.19 | 10.17| 96.17| 0.79 |
| SmCO  | 12.03| 4.83 | 11.08| 4.48 | 8.73 | 10.5 | 48.35|

TABLE 13  Confusion matrix fusion of CNN A CNN B and CNN C customized and learnable filter

|       | BaCO | CpCO | FrCO | MuCO | NoCO | SbCO | SmCO |
|-------|------|------|------|------|------|------|------|
| BaCO  | 97.74| 0.00 | 0.00 | 1.20 | 0.53 | 0.00 | 0.53 |
| CpCO  | 1.51 | 82.23| 9.34 | 1.96 | 0.30 | 2.56 | 2.11 |
| FrCO  | 0.00 | 0.80 | 98.28| 0.13 | 0.27 | 0.13 | 0.40 |
| MuCO  | 4.77 | 0.27 | 2.25 | 90.45| 0.53 | 0.80 | 0.93 |
| NoCO  | 0.00 | 0.00 | 0.13 | 0.00 | 98.55| 0.92 | 0.40 |
| SbCO  | 0.00 | 0.13 | 0.13 | 0.13 | 2.64 | 96.17| 0.79 |
| SmCO  | 5.78 | 2.71 | 4.01 | 1.42 | 3.42 | 3.66 | 79.01|

While considering the static shape and dynamic motion of the gait feature SbCO and SmCO, subjects frequently carried small and lightweight COs, which were occluded by the subject’s body very often. Thus, SbCO and SmCO are better discriminated by the scaled down input versions that is, CNN C. For the case of BaCO, subjects typically carried a large CO, such as a backpack and the position of the CO was fixed and stable within a gait period. However, the large CO heavily affected the shape and posture, which is better defined by CNN A. Similarly, for MuCO, person carries a big backpack together with other types of carried objects that were carried in other regions of the human body. Here, the position of the carried object in the back region is fixed, other locations were arbitrary. The GEI samples for MuCO were largely affected not only by shape but also by motion which is well defined by CNN B. Regarding FrCO, the subjects typically carried a lightweight object by hand in the front region Therefore, the GEI samples of FrCO were affected slightly by shape and fairly affected by motion. Regarding CpCO, the carried location was arbitrary in any region of the body within a gait period. Due to the indiscriminately changing position of the object from one region to another the GEI samples for CpCO were rigorously affected by the motion feature along with shape features and better discriminated by CNN A.

Table 14 shows the comparison of the proposed carried object detector with MSVM [37] and SIAME [37]. MSVM is a benchmark, with a third-degree polynomial kernel for the classification of the carried objects and Siamese is based on CNN network architecture, in which two input GEI features are used to train the two parallel CNN networks with shared parameters for gait recognition.
**TABLE 14**  Performance measure for OU-ISIR large population gait database with real-life carried object

|                        | BaCO | CpCO | FrCO | MuCO | NoCO | SbCO | SmCO | Average |
|------------------------|------|------|------|------|------|------|------|---------|
| CNN A _ customized      | 81.76| 64.91| 69.50| 63.79| 86.13| 74.64| 43.04| 68.72   |
| CNN B _ customized      | 79.49| 54.22| 83.55| 64.06| 86.92| 70.54| 35.38| 67.40   |
| CNN C _ customized      | 80.42| 58.28| 73.87| 64.06| 84.15| 75.43| 47.52| 69.20   |
| CNN A & CNN B (customized) | 89.35| 73.64| 88.73| 77.72| 94.58| 84.15| 43.40| 78.26   |
| CNN B & CNN C (customized) | 89.21| 59.79| 91.51| 77.98| 90.75| 83.09| 48.58| 77.07   |
| CNN A & CNN C (customized) | 82.42| 58.28| 73.87| 64.06| 84.15| 75.43| 47.52| 69.20   |
| CNN A _ learnable filter | 85.22| 41.27| 84.35| 58.22| 86.53| 68.82| 42.57| 66.72   |
| CNN B _ learnable filter | 74.57| 62.50| 69.89| 53.71| 76.22| 47.69| 69.34| 64.96   |
| CNN C _ learnable filter | 74.83| 52.71| 68.70| 58.96| 64.20| 80.32| 58.96| 65.94   |
| CNN A & CNN B (learnable) | 92.54| 62.65| 89.92| 91.55| 71.99| 68.75| 50.12| 78.26   |
| CNN B & CNN C (learnable) | 86.15| 71.54| 83.02| 71.62| 72.94| 82.96| 77.71| 79.05   |
| CNN A & CNN C (learnable) | 90.55| 57.53| 89.39| 72.94| 89.17| 86.53| 60.14| 78.07   |
| CNN A, CNN B & CNN C (learnable) | 95.34| 75.75| 94.43| 80.11| 94.19| 87.58| 80.31| 86.89   |
| CNN A _ learnable filter | 85.22| 41.27| 84.35| 58.22| 86.53| 68.82| 42.57| 66.72   |
| CNN B _ learnable filter | 74.57| 62.50| 69.89| 53.71| 76.22| 47.69| 69.34| 64.96   |
| CNN C _ learnable filter | 74.83| 52.71| 68.70| 58.96| 64.20| 80.32| 58.96| 65.94   |
| CNN A & CNN B (learnable) | 92.54| 62.65| 89.92| 91.55| 71.99| 68.75| 50.12| 78.26   |
| CNN B & CNN C (learnable) | 86.15| 71.54| 83.02| 71.62| 72.94| 82.96| 77.71| 79.05   |
| CNN A & CNN C (learnable) | 90.55| 57.53| 89.39| 72.94| 89.17| 86.53| 60.14| 78.07   |
| CNN A, CNN B & CNN C (learnable) | 95.34| 75.75| 94.43| 80.11| 94.19| 87.58| 80.31| 86.89   |
| CNN A _ learnable filter | 85.22| 41.27| 84.35| 58.22| 86.53| 68.82| 42.57| 66.72   |
| CNN B _ learnable filter | 74.57| 62.50| 69.89| 53.71| 76.22| 47.69| 69.34| 64.96   |
| CNN C _ learnable filter | 74.83| 52.71| 68.70| 58.96| 64.20| 80.32| 58.96| 65.94   |
| SIAME [37]              |     |     |     |     |     |     |     |        |
|                        | 78.90| 46.20| 72.80| 62.60| 76.80| 73.00| 32.70| 63.29   |
| MSVM [37]               |     |     |     |     |     |     |     |        |
|                        | 74.70| 26.50| 60.20| 40.10| 72.60| 58.40| 36.00| 52.64   |

5  | CONCLUSION

In this paper, gender discrimination, age group classification and carried object recognition is performed by analysing the traits displayed in GEIs. A unified model with six parallel CNNs learning the discriminative features from pyramidal decomposition of the input GEI is suggested. The individual learning by each separate components and decision level score fusion lead to obvious performance enhancement over separate learning. From the conducted comprehensive experiment for gender discrimination, age group classification and carried object recognition, the parallel Network architecture significantly outperforms the state-of-the-art methods. In future, more efficient and effective network structures and metric learning strategies can be explored for finely dividing the age group labels and age estimation.

**ORCID**

Newlin Shebiah Russel(1) https://orcid.org/0000-0002-0835-5848

**REFERENCES**

1. Bobick, A.F., Johnson, A.Y.: Gait recognition using static, activity-specific parameters. Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, Kauai, HI, USA, Dec. 2001, pp. I.423–I.430.
2. Tafazzoli, F., Safabakhsh, R.: Model-based human gait recognition using leg and arm movements. Eng. Appl. Artif. Intell. 23(8), 1237–1246 (2010)
3. Wu, Z., et al.: A comprehensive study on cross-view gait based human identification with deep CNNs. IEEE Trans. Pattern Anal. Mach. Intell. 39(2), 209–226 (2016)
4. Takemura, N., et al.: On input/output architectures for convolutional neural network-based cross-view gait recognition. IEEE Trans. Circuits Syst. Video Technol. 29(9), 2708–2719 (2018)
5. Shiraga, K., et al.: GEINet: View-invariant gait recognition using a convolutional neural network. Proc. IEEE Int. Conf. Biometrics, Halmstad, Sweden, June 2016, pp. 1–8.
6. Yu, S., et al.: GaitGAN: Invariant Gait Feature Extraction Using Generative Adversarial Networks. Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, July 2017, pp. 532–539
7. Yu, S., et al.: Neurocomputing invariant feature extraction for gait recognition using only one uniform model. Neurocomputing 239, 81–93 (2018)
8. He, Y., et al.: Multi-task GANs for view-specific feature learning in gait recognition. IEEE Trans. Inf. Forensics Secur. 14(1), 102–113 (2018)
9. Boulgouris, V., et al.: Recognition: A challenging signal. IEEE Signal Process. Mag. 22, 78–90 (2005)
10. Boyd, J.E., Little, J.J.: Biometric gait recognition, Lect. Notes Comput. Sci. 3161, 19–42 (2005)
11. Yu, S., et al.: A study on gait-based gender classification. IEEE Trans. Image Process. 18(8), 1905–1910 (2009)
12. Lu, J., Tan, Y.P.: Uncorrelated discriminant simplex analysis for view-invariant gait signal computing. Pattern Recognit. Lett. 31(5), 382–393 (2010)
13. Hu, M., et al.: Gait-based gender classification using mixed conditional random field. IEEE Trans. Syst. Man, Cybern. Part B Cybern. 41(5), 1429–1439 (2011)
14. Isaac, E.R.H.P., et al.: Multiview gait-based gender classification through pose-based voting. Pattern Recognit. Lett. 126, 41–50 (2018)
15. Castro, F.M., et al.: Automatic learning of gait signatures for people identification. Lect. Notes Comput. Sci. 10366, 257–270 (2017)
16. Do, T.D., et al.: Real-time and robust multiple-view gender classification using gait features in video surveillance. Pattern Anal. Appl. 23, 399–413 (2019)
17. Sakata, A., et al.: Gait-based age estimation using multi-stage convolutional neural network. IPSJ Trans. Comput. Vis. Appl. 11, 4 (2019)
18. Antipov, G., et al.: Minimalistic CNN-based ensemble model for gender prediction from face images. Pattern Recognit. Lett. 70, 59–65 (2016)
19. Antipov, G., et al.: Effective training of convolutional neural networks for face-based gender and age prediction. Pattern Recognit. 72, 15–26 (2017)
20. Davis, J.W.: Visual categorization of children and adult walking styles. Lect. Notes Comput. Sci. 2091, 295–300 (2001)
21. Ince, O.F., et al.: Child and adult classification using ratio of head and body heights in images. Int. J. Comput. Commun. Eng. 3(2), 120–122 (2014)
22. Makhtara, Y., et al.: Gait-based age estimation using a whole-generation gait database. Proc. IEEE Int. Joint Conf. Biometrics, Washington, DC, USA, pp. 1–6 Oct. (2011)
23. Liu, Z., Sarkar, S.: Simplest representation yet for gait recognition: Averaged silhouette. Proc.—Int. Conf. Pattern Recognit. 37, 211–226 (2004)
24. Li, X., et al.: Gait-based human age estimation using age group-dependent manifold learning and regression. Multimed. Tools Appl. 77(21), 28333–28354 (2018)
25. Lu, J., Tan, Y.P.: Gait-based human age estimation. IEEE Trans. Inf. Forensics Secur. 5(1), 761–770 (2010)
26. Xu, C., et al.: The OU-ISIR gait database comprising the large population dataset with age and performance evaluation of age estimation. IPSJ Trans. Comput. Vis. Appl. 9(1), 24 (2017)
27. Haritaoglu, I., et al.: Hydra: multiple people detection and tracking using silhouettes. Proc. Second IEEE Workshop on Visual Surveillance, Fort Collins, CO, USA, June 1999, pp. 6–13
28. Haritaoglu, I., et al.: Carried object detection in videos using color information. IEEE Trans. Inf. Forensics Secur. 8(1), 1620–1631 (2013)
29. Haritaoglu, I., et al.: W4: Real-time surveillance of people and their activities. IEEE Trans. Pattern Anal. Mach. Intell. 22(8), 809–830 (2000)
30. Damen, D., Hogg, D.: Detecting carried objects from sequences of walking pedestrians. IEEE Trans. Pattern Anal. Mach. Intell. 34(6), 1056–1067 (2012)
31. Tzamidou, G., et al.: Carried object detection in videos using color information. IEEE Trans. Inf. Forensics Secur. 8(10), 1620–1631 (2013)
32. Ghadiri, F., et al.: From superpixel to human shape modelling for carried object detection. Pattern Recognit. 89, 134–150 (2019)
33. Wahiyo Hariyono, J., Jo, K.H.: Body part boosting model for carried bag-gage detection and classification. Neurocomputing 228, 106–118 (2017)
34. BenAbdelkader, C., Davis, L.: Detection of people carrying objects: A motion-based recognition approach. Proc.—5th IEEE Int. Conf. Autom. Face Gesture Recognition, FGR. 2002, 378–383 (2002)
35. Mather, G., Murdoch, L.: Gender discrimination in biological motion displays based on dynamic cues. Proc. R. Soc. B Biol. Sci. 258(1353), 273–279 (1994)
36. Yu, S., et al.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. Proc.—Int. Conf. Pattern Recognit. 4, 441–444 (2006)
37. Uddin, M.Z., et al.: The OU-ISIR large population gait database with real-life carried object and its performance evaluation. IPSJ Trans. Comput. Vis. Appl. 10(1), 5 (2018)
38. Zhang, D., Wang, Y.: Using multiple views for gait-based gender classification. Proc. 26th Conf. Chinese Control and Decision, Changsha, China, May-June 2014, pp. 2194–2197.
39. Do, T.D., et al.: Real-time and robust multiple-view gender classification using gait features in video surveillance. Pattern Anal. Appl. 23, 399–413 (2019)
40. Lamar-Leon, J., et al.: Persistent homology-based gait recognition robust to upper body variations. Proc. 32nd Int. Conf. Pattern Recognition, Cancun, Mexico, Dec. 2016, pp. 1083–1088

How to cite this article: Shebiah RN, Arivazhagan S. Gender discrimination, age group classification and carried object recognition from gait energy image using fusion of parallel convolutional neural network. *IET Image Process.* 2021;15:239–251. https://doi.org/10.1049/ipr2.12024