



Article Aquila Optimization with Transfer Learning Based Crowd Density Analysis for Sustainable Smart Cities

Mesfer Al Duhayyim ^{1,*}, Eatedal Alabdulkreem ², Khaled Tarmissi ³, Mohammed Aljebreen ⁴, Bothaina Samih Ismail Abou El Khier ⁵, Abu Sarwar Zamani ⁶, Ishfaq Yaseen ⁶ and Mohamed I. Eldesouki ⁷

- ¹ Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj 16273, Saudi Arabia
- ² Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
- ³ Department of Computer Sciences, College of Computing and Information System, Umm Al-Qura University, Mecca 24382, Saudi Arabia
- ⁴ Department of Computer Science, Community College, King Saud University, P.O. Box 28095, Riyadh 11437, Saudi Arabia
- ⁵ Department of Architectural Engineering, Faculty of Engineering and Technology, Future University in Egypt, New Cairo 11845, Egypt
- ⁶ Department of Computer and Self Development, Preparatory Year Deanship,
- Prince Sattam Bin Abdulaziz University, Al-Kharj 16278, Saudi Arabia
 ⁷ Department of Information System, College of Computer Engineering and Sciences, Prince Sattam Bin
- Abdulaziz University, Al-Kharj 16278, Saudi Arabia
- Correspondence: m.alduhayyim@psau.edu.sa

Abstract: Video surveillance in smart cities provides efficient city operations, safer communities, and improved municipal services. Object detection is a computer vision-based technology, which is utilized for detecting instances of semantic objects of a specific class in digital videos and images. Crowd density analysis is a widely used application of object detection, while crowd density classification techniques face complications such as inter-scene deviations, non-uniform density, intra-scene deviations and occlusion. The convolution neural network (CNN) model is advantageous. This study presents Aquila Optimization with Transfer Learning based Crowd Density Analysis for Sustainable Smart Cities (AOTL-CDA3S). The presented AOTL-CDA3S technique aims to identify different kinds of crowd densities in the smart cities. For accomplishing this, the proposed AOTL-CDA3S model initially applies a weighted average filter (WAF) technique for improving the quality of the input frames. Next, the AOTL-CDA3S technique employs an AO algorithm with the SqueezeNet model for feature extraction. Finally, to classify crowd densities, an extreme gradient boosting (XGBoost) classification model is used. The experimental validation of the AOTL-CDA3S approach is tested by means of benchmark crowd datasets and the results are examined under distinct metrics. This study reports the improvements of the AOTL-CDA3S model over recent state of the art methods.

Keywords: sustainability; smart cities; deep learning; crowd density; video surveillance

1. Introduction

The proliferation of smart cities (SC), smart security, and smart community have made anomalous behavior analysis a hot topic in crowd event research. Security is a highly significant component in city; it becomes vital to ensure a safe ecosystem for the data that is produced and for the people [1]. Cities should enforce security measures for guaranteeing complete security of the individuals' information, and the information produced by the sensors and urban infrastructure, among others. In the present era, it is very significant to enable compliance with precautionary measures that safeguard people; for instance, if they have adequate social distancing measures in public places [2]. In this context, imaginative systems that enable us to find those places where there are less pedestrians



Citation: Al Duhayyim, M.; Alabdulkreem, E.; Tarmissi, K.; Aljebreen, M.; El Khier, B.S.I.A.; Zamani, A.S.; Yaseen, I.; I. Eldesouki, M. Aquila Optimization with Transfer Learning Based Crowd Density Analysis for Sustainable Smart Cities. *Appl. Sci.* **2022**, *12*, 11187. https://doi.org/10.3390/ app122111187

Academic Editors: Junchi Yan and Minghao Guo

Received: 8 October 2022 Accepted: 2 November 2022 Published: 4 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). is of considerable significance, in order to simplify the leisure of the family who need to go out [3]. The SC construction is a dynamic procedure, and the management platform should always be ready not only for ingesting the real-time dataset, but for incorporating the information from various sources and bringing about these data for analysis, along with creating various visualization approaches and integrating several data [4].

Currently, intelligent monitoring is increasingly implemented in common places (shopping malls, hospitals, and campuses) [5]. Of these, the crowding degree is the main problem that influences the study of the abnormal performance of crowds. It can be realized by crowd counting and crowd density (CD) estimation. CD prediction allures the interest of researchers abroad and at home in increasing numbers [5]. Currently, researchers abroad and at home have performed research relating to crowd abnormality with the focus on track path abnormality, crowd violence, crowd crowding, and crowd panic. The density and count of the crowd were frequently employed to reflect the degree of crowd crowding. There are two types of crowd analysis which rely upon crowd density prediction and crowd count. Crowd density prediction is assessing the crowd dispersion and specific groups of people [6].

Crowd analysis is gaining considerable attention among scholars in recent times due to several factors. The massive growth in urbanization and the global population has led to a rise in events such as political rallies, public demonstrations, and sporting events, among others. Similar to other computer vision (CV) problems, crowd analysis will face several complexities such as an uneven distribution of people, inter-scene variations in appearance, high clutter, non-uniform illumination, an unclear viewpoint, intra-scene and scale issues, and occlusions; such issues are very tough to solve. The uncertainty of the problem in addition to the wide array of applications for analyzing crowds resulted in an augmented focus among researchers in recent times. Convolution neural network (CNN) achieved fruitful results in image processing and in CD estimation. On the other hand, the hyperparameters of the DL models play a vital role in attaining enhanced performance. Owing to continual deepening of the model, the number of parameters of DL models also increases quickly which results in model overfitting [7]. At the same time, different hyperparameters have a significant impact on the efficiency of the CNN model. Particularly, the hyperparameters such as epoch count, batch size, and learning rate selection are essential to attain an effectual outcome. The trial-and-error method for hyperparameter tuning is a tedious and erroneous process; metaheuristic algorithms can be applied. The hyperparameter tuning can be considered an NP hard problem which can be solved using metaheuristic algorithms such as genetic algorithm, hunger games search, and memetic algorithm, among others [8]. The metaheuristic is a high-level problem-independent algorithmic model which offers a collection of strategies to design heuristic optimization algorithms [9,10]. Metaheuristics can be utilized for combinatorial optimization where an optimal solution is required over a discrete search-space.

This study presents an Aquila Optimization with Transfer Learning based Crowd Density Analysis for Sustainable Smart Cities (AOTL-CDA3S). This AOTL-CDA3S technique aims to identify different kinds of crowd densities in the SC environment. In order to accomplish this, the proposed AOTL-CDA3S technique initially applies the weighted average filter (WAF) technique for improving the quality of the input frames. Next, the AOTL-CDA3S technique employs an AO algorithm with the SqueezeNet model for feature extraction. In this work, we have used the Aquila optimization algorithm for hyperparameter tuning due to its faster optimization speed, global exploration ability, high search efficiency, and fast convergence speed. Finally, to classify crowd densities, an extreme gradient boosting (XGBoost) classification mechanism is used. The experimental validation of the AOTL-CDA3S technique is tested by means of benchmark crowd datasets and the results are examined under distinct metrics.

2. Literature Survey

Ding et al. [11] introduced new encoder-decoder CNNs that merge the feature maps in both decoding and encoding sub-networks for generating a more reasonable density map and predicting the people count very precisely. Moreover, the authors present an innovative evaluation technique termed the Patch Absolute Error (PAE) which is a suitable method to measure the accurateness of density maps. Alrowais et al. [12] modelled an MDTL-ICDDC method to detect objects. This MDTL-ICDDC method focused on the effectual classification and identification of CD on video surveillance systems. The MDTL-ICDDC method mainly uses an SSA in addition to the NASNetLarge method as a feature extractor where the hyper-parameter tuning will be executed by the SSA. In addition, for CD and classification process, a weighted ELM (WELM) technique was used.

Wang et al. [13] modelled a trivial CNN related CD estimation method through the merging of the dilated convolution modified and MobileNetv2. In [14], a novel testing process that depends on features from accelerated segment test (FAST) techniques was presented for detecting the crowd features from drone imageries captured through different camera positions and orientations. In [15], utilizing the Hough circle transformation, a CD estimation technique was modelled. In the technical background and foreground, data was divided by leveraging the ViBe technique along with the segmentation of foreground data. Such segmented foreground data can be employed in the Hough circle transformation for CD estimation.

In [16], the authors were interested in implementing ML for crowd management for monitoring populated areas and to thwart congestion circumstances. The authors devised a Single-CNN including the Three Layers (S-CNN3) method for counting people in a scene and make a conclusion of crowd estimation. Afterward, a comparative analysis for density counting accomplishes the efficiency of this presented technique towards the CNNs with Switched CNN (SCNN) and four layers (single-CNN4). Zhou et al. [17] present a multilinear rank support tensor machine (MRSTM) considering a tensor collection as input to the issue of predicting the CD level. Moreover, an alternative SVM technique was modelled for training an MRSTM method.

Zhao et al. [18] present a new approach utilized for correctly analyzing the crowd stability dependent upon images achieved in a real-time video surveillance system (VSS) from dense crowd conditions. For enhancing the accuracy of human head detection to the crowd count and CD estimate, the authors enhance the CNN approach with further columns and merged features, achieving a four-column CNN (4C-CNN). Liu et al. [19] project a model which is integrated dual-modal data; the video surveillance streams and transportation scheduling data for predicting the future CD in the transportation buildings. This technique employs the temporal convolutional layer for extracting the time dependence of video streams and transportation schedules. The predictions fuse the data or utilize the GRU layers for predicting the CD.

In [20], a Multi-Step CD Predictor (MSCDP) for fusing video frame structures and equivalent density heatmaps is presented for correctly forecasting future CD heatmaps. For capturing long-term periodic movement features, the long-term optical flow context memory (LOFCM) element was planned for storing learnable patterns. In [21], the authors initially established a WiFi monitor recognition which captures smartphone passive Wi-Fi signal data comprising MAC address and RSSI. Afterward, the authors present a positioning technique dependent upon the smartphone passive WiFi probe and a dynamic fingerprint management approach. Therefore, the authors designed a computing model for the probability of a user creating one WiFi signal for identifying people populations. Lastly, the author presents a CD estimate solution dependent upon a WiFi probe packet positioning technique.

3. The Proposed Model

For accurate crowd density classification, we have developed a new AOTL-CDA3S technique for sustainable smart cities. The presented AOTL-CDA3S technique aims to

identify different kinds of crowd densities in the SC environment, and encompasses WAF noise removal, SqueezeNet feature extraction, AO based hyperparameter tuning, and XGBoost classification. Figure 1 defines the block diagram of AOTL-CDA3S system.



Figure 1. Block diagram of the AOTL-CDA3S system.

3.1. Image Pre-Processing

In the first phase, the presented AOTL-CDA3S technique eliminates the existing noise. The WAF was intended to pre-process to suppress noise and enhance spatial domain features effectively [22]. The filter W^{η} was defined by the matrix, whereby η represents the odd number. Each component value of matrix is determined by the distance between the center of matrix and the present place, as illustrated below. The center of matrix was defined by $w_{(\eta+1)/2,(\eta+1)/2} = 2/\eta^2$. The proposed filter continues to the edges; however, the filter suppresses the speckle noise associated with other filters, such as the mean filter, and maintain the continuity of an image.

$$w_{i}j = \frac{1}{\eta^{2}\sqrt{\left(\frac{\eta+1}{2}-i\right)^{2}+\left(\frac{\eta+1}{2}-j\right)^{2}}}; i = 1, 2, \eta; j = 1, 2\eta;$$
(1)

In Equation (1), $I_1, I_2 \in \mathbb{R}^{N_r \times N_c}$, the convolution? of each image using W^{η} is accomplished to acquire 2 images, $I_1^w(\eta) = I_1 * W^{\eta}$ and $I_2^w(\eta) = I_2 * W^{\eta}$, while the * indicates the 2D convolution function.

3.2. Feature Extraction

To derive features, the presented AOTL-CDA3S technique utilized the SqueezeNet model to produce feature vectors. In general, CNN includes convolutional, pooling, and full connected (FC) layers [23]. Initially, the feature was extracted by employing several pooling and convolutional layers. Afterward, the mapping feature in the last convolution layers is changed to a 1D vector followed by the resultant layer classifying the input image. The network decreases the square variance between the predictive and classifier results and alters the weighted variable by BP. The neuron in every layer is orderly in three dimensions (width, depth, and height) so that depth defines the count of input feature mapping or

channel count of input images, and height and width signify the size of neurons. The convolutional layer comprises many convolutional filters and extracting features in the image with convolution approach. The convolution filters of the current layer convoluted the input feature mapping for removing local features and realizing the resultant feature mapping. Then, the non-linear feature mapping was realized by the activation function. The pooling or subsampling layers are the last convolutional layers. This applies a down sampling approach that is a specific value as an outcome in particular region.

As the variable count for AlexNet and VGGNet enhances, the SqueezeNet network infrastructure is established that is minimally variable but maintains accuracy. The fire elements develop an important element in SqueezeNet. This element was separated to expand and squeeze infrastructure. The 1×1 convolutional layer obtained considerable attention from the network infrastructure, thereby attaining data integration on the channel and a linear combination of many features mapping. If the count of input and output channels is superior, the convolution kernels develop well. Next, adding 1×1 convolutional to all the single inception systems decreases the count of input channels, and the complex function and convolution kernel variable were reduced. Lastly, 1×1 convolutional was added for increasing the extracting feature and the count of channels. When the sampling reduction approach can be delayed, a superior activation graph was offered to convolutional layer, so the higher activation graph reserves more data and offers superior classifier efficacy.

For the hyperparameter optimization process, the AO algorithm is utilized. The AO approach mimics Aquila's social activity to catch the prey [24]. AO is a populationbased optimized method which is based on other metaheuristic algorithms and begins by establishing an initial population X with N agent. The succeeding equation is exploited to implement these processes as explained in Algorithm 1.

$$X_{ij} = r_1 \times (UB_j - LB_j) + LB_j, \ i = 1, 2, \ Nj = 1, 2, \ \dots, \ Dim$$
(2)

Here, UB_j and LB_j illustrate the restriction of search space. $r_1 \in [0, 1]$ indicates the arbitrary number and Dim shows the dimension of agents. In this study, the subsequent stage is to implement exploitation and exploration until the optimum solution is determined. There are 2 phases in exploration and exploitation. The X_b optimum agent and (X) the average agent is employed in the exploration, and the mathematical expressions are given below:

$$X_i(t+1) = X_b(t) \times \left(\frac{1-t}{T}\right) + (X_M(t) - X_b(t) * rand),$$
(3)

$$X_M(t) = \frac{1}{N} \sum_{i=1}^{N} X(t), \ \forall j = 1, 2, \dots, Dim$$
(4)

The exploration phase can be controlled by means of using $\left(\frac{1-t}{T}\right)$. The maximal quantity of generations is demonstrated by *T*. The exploration process makes use of the Levy flight (*Levy* (*D*)) distribution and *X*_b for upgrading the solution in the following:

$$X_i(t+1) = X_b(t) \times Levy(D) + X_R(t) + (y-x) * rand,$$
(5)

Levy
$$(D) = s \times \frac{u \times \sigma}{|v|^{\frac{1}{\beta}}}, \sigma = \left(\frac{\Gamma(1+\beta) \times sine\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right) \times \beta \times 2^{\left(\frac{\beta-1}{2}\right)}}\right)$$
 (6)

where, s = 0.01 and $\beta = 1.5$. *u* and *v* specify the arbitrary values. X_R denotes arbitrarily selected agents. In addition, *y* and *x* represent two parameters employed to stimulate the spiral shape:

$$y = r \times cos(\theta), \ x = r \times sin(\theta)$$
 (7)

$$r = r_1 + U \times D_1, \ \theta = -\omega \times D_1 + \theta_1, \ \theta = \frac{3 \times \pi}{2}$$
(8)

Algorithm 1: Pseudocode of AO algorithm					
Parameter initialization					
WHILE (Termination criteria is unsatisfied) do					
Determine value of fitness function					
$X_{best}(t) = $ Compute optimal attained solution based on fitness value					
for $(i = 1, 2, N)$ do					
Upgrade mean value of present solution $X_M(t)$.					
Upgrade the x, y, G_1, G_2 , Levy(D), etc.					
if $ au \leq \left(\frac{2}{3}\right) * T$ then					
if $rand \leq 0.5$ then					
Expanded exploration (X_1)					
Upgrade present solution					
if Fitness $(X_1(t+1)) < Fitness(X(t))$ then					
$X(t) = (X_1(t+1))$					
if Fitness $(X_1(t+1)) < Fitness(X_{best}(t))$ then					
$\lambda_{best}(I) = \lambda_1(I+I)$					
end if					
ella					
Narrowed exploration (X_2)					
Upgrade current solution.					
if Fitness $(X_2(t+1)) < Fitness(X(t))$ then					
$X(T) = (X_2(T+1))$					
if Fitness $(X_2(t+1)) < Fitness(X_{best}(t))$ then					
$X_{best}(t) = X_2(t+1)$					
end if					
end if					
end if					
else					
$\int Fx randod exploitation (Y_{c})$					
Ungrade present solution					
if Fitness $(X_2(t+1)) < Fitness(X(t))$ then					
$X(t) = (X_3(t+1))$					
if Fitness $(X_3(t+1)) < Fitness(X_{hest}(t))$ then					
$X_{hest}(T) = X_3(T+1)$					
end if					
end if					
else					
Narrowed exploitation (X_4)					
Upgrade present solution					
if Fitness $(X_4(t+1)) < Fitness(X(t))$ then $X_4(T) = (X_4(T+1))$					
$\mathbf{X}(I) = (\mathbf{X}_4(I+1))$					
If fitness $(X_4(t+1)) < Fitness(X_{besf}(t))$ then					
$X_{best}(t) = X_4(t+1)$					
end if					
end if					
end if					
ena II and for					
end ror end while					
return optimal solution (X_{t-i})					
indest).					

$$X_i(t+1) = (X_b(t) - X_M(t)) \times \alpha - rnd + (UB \times rnd + LB) \times \delta$$
(9)

Now UB = (UB - LB), α , and δ indicate the exploitation adjustment variable. *rnd* \in [0, 1] represents a random value as follows:

$$X_i(t+1) = QP \times X_b(t) - GX - G_2 \times Levy(D) + rnd \times G_1$$
(10)

$$GX = (G_1 \times X(t) \times rnd)$$
$$QP(t) = t^{\frac{2 \times rnd(\)-1}{(1-T)^2}}$$
(11)

Furthermore, G_1 represents the motion employed for tracking the optimal individual solution in the following:

$$G_1 = 2 \times rnd() - 1, \ G_2 = 2 \times \left(1 - \frac{t}{T}\right)$$

$$(12)$$

Here, *rnd* denotes a random value. Moreover, G_2 denotes a parameter that minimized from two to zero:

$$G_2 = 2 \times \left(1 - \frac{t}{T}\right) \tag{13}$$

3.3. Crowd Density Classification

To classify crowd densities, the XGBoost model was utilized in this study. XGBoost [25] is an enhanced model stimulated from the XGBoost decision tree and could build boosted tree effectively and operate simultaneously. The boosted tree in XGBoost is classified into regression and classification trees. The aim is to enhance the values of objective function. Frequency is a simplified version of gain, which is the amount of features in each constructed tree. Gain is the major reference factor of the importance of a feature in the tree branch.

$$w_{\ell}^{2}(T) = \sum_{t=1}^{J-1} \hat{\tau}$$
(14)

For a single decision tree *T*, Breiman developed a score of importance for every predictor feature X_{ℓ} . The DT is comprised of J—l internal node, and partitions the region into two sub-regions at *t* nodes by the prediction feature X_{ℓ} as follows:

$$w_{\ell}^{2}(T) = \frac{1}{M} \sum_{m=1}^{M} \hat{\tau}_{1}^{2}(T_{m})$$
(15)

The significant feature is based on the predictive performance variations when the feature is replaced with random noise, and hen attain how every feature contributed toward the predictive performance during training. The electricity load is sensitive to temperature variables. Furthermore, the supplement features are significant features for forecasting loads.

4. Performance Evaluation

The proposed model is simulated using the Python 3.6.5 tool. The proposed model is tested on the PC i5-8600k, GeForce 1050Ti 4 GB, 16 GB RAM, 250 GB SSD, and 1TB HDD. The parameter settings are given as follows: learning rate: 0.01, dropout: 0.5, batch size: 5, epoch count: 50, and activation: ReLU. The experimental validation of the AOTL-CDA3S model is carried out using a crowd dataset with 1000 samples as represented in Table 1. The dataset holds 250 samples under every class. Figure 2 demonstrates some sample images.

Labels	Class	Sample count
C1	Dense Crowd	250
C2	Medium Dense Crowd	250
C3	Sparse Crowd	250
C4	No Crowd	250
Total Number of Samples		1000



Figure 2. Sample images.

The crowd density analysis of the AOTL-CDA3S model in the form of confusion matrix is portrayed in Figure 3. With the entire dataset, the AOTL-CDA3S model has recognized 240 samples to C1, 220 samples to C2, 240 samples to C3, and 228 samples to C4. Additionally, with 70% of the TR database, the AOTL-CDA3S method has recognized

167 samples to C1, 156 samples to C2, 161 samples to C3, and 156 samples to C4. Furthermore, with 30% of the TS database, the AOTL-CDA3S technique has recognized 73 samples to C1, 64 samples to C2, 79 samples to C3, and 72 samples to C4.



Figure 3. Confusion matrices of AOTL-CDA3S system (a) Entire database, (b) 70% of TR database, and (c) 30% of TS database.

Table 2 and Figure 4 illustrate brief crowd classification results of the AOTL-CDA3S model on the entire data. The AOTL-CDA3S model has shown enhanced results under each class. For instance, in class-1, the AOTL-CDA3S model has obtained $accu_y$ of 97%, $prec_n$ of 92.31%, $sens_y$ of 96%, $spec_y$ of 97.33%, and F_{score} of 94.12%. Meanwhile, in class-3, the AOTL-CDA3S approach has gained $accu_y$ of 97.60%, $prec_n$ of 94.49%, $sens_y$ of 96%, $spec_y$ of 98.13%, and F_{score} of 95.24%. Concurrently, in class-4, the AOTL-CDA3S method has gained $accu_y$ of 95.70%, $prec_n$ of 91.57%, $sens_y$ of 91.20%, $spec_y$ of 97.20%, and F_{score} of 91.38%.

Entire Dataset						
Labels	Accuracy	Precision	Sensitivity	Specificity	F-Score	
Class-1	97.00	92.31	96.00	97.33	94.12	
Class-2	95.30	92.83	88.00	97.73	90.35	
Class-3	97.60	94.49	96.00	98.13	95.24	
Class-4	95.70	91.57	91.20	97.20	91.38	
Average	96.40	92.80	92.80	97.60	92.77	

Table 2. Crowd classification outcome of the AOTL-CDA3S system with different measures under the entire database.





Table 3 and Figure 5 exemplify detailed crowd classification results of the AOTL-CDA3S approach on 70% of the TR database. The AOTL-CDA3S technique has exhibited enhanced results under each class. For example, in class-1, the AOTL-CDA3S methodology has attained *accu_y* of 96.43%, *prec_n* of 90.76%, *sens_y* of 95.43%, *spec_y* of 96.76%, and *F_{score}* of 93.04%. Meanwhile, in class-3, the AOTL-CDA3S approach has gained *accu_y* of 97.14%, *prec_n* of 94.15%, *sens_y* of 94.15%, *spec_y* of 98.11%, and *F_{score}* of 94.15%. Concurrently, in class-4, the AOTL-CDA3S technique has achieved *accu_y* of 94.86%, *prec_n* of 90.17%, *sens_y* of 89.14%, *spec_y* of 96.76%, and *F_{score}* of 89.66%.

Training Phase (70%)						
Labels	Accuracy	Precision	Sensitivity	Specificity	F-Score	
Class-1	96.43	90.76	95.43	96.76	93.04	
Class-2	94.43	90.70	87.15	96.93	88.89	
Class-3	97.14	94.15	94.15	98.11	94.15	
Class-4	94.86	90.17	89.14	96.76	89.66	
Average	95.71	91.45	91.47	97.14	91.43	

Table 3. Crowd classification outcome of the AOTL-CDA3S system with different measures under 70% of the TR database.



Training Phase (70%)

Figure 5. Crowd classification outcome of the AOTL-CDA3S system under 70% of the TR database.

Table 4 and Figure 6 demonstrate the brief crowd classification results of the AOTL-CDA3S algorithm on 30% of the TS database. The AOTL-CDA3S approach has displayed enhanced results under each class. For example, in class-1, the AOTL-CDA3S technique has gained *accu_y* of 98.33%, *prec_n* of 96.05%, *sens_y* of 97.33%, *spec_y* of 98.67%, and *F_{score}* of 96.69%. Meanwhile, in class-3, the AOTL-CDA3S approach has reached $accu_y$ of 98.67%, $prec_n$ of 95.18%, $sens_y$ of 100%, $spec_y$ of 98.19%, and F_{score} of 97.53%. Concurrently, in class-4, the AOTL-CDA3S technique has reached *accu_y* of 97.67%, *prec_n* of 94.74%, *sens_y* of 96%, *spec*_{*y*} of 98.22%, and *F*_{*score*} of 95.36%.

Testing Phase (30%)						
Labels	Accuracy	Precision	Sensitivity	Specificity	F-Score	
Class-1	98.33	96.05	97.33	98.67	96.69	
Class-2	97.33	98.46	90.14	99.56	94.12	
Class-3	98.67	95.18	100.00	98.19	97.53	
Class-4	97.67	94.74	96.00	98.22	95.36	
Average	98.00	96.11	95.87	98.66	95.93	

Table 4. Crowd classification outcome of the AOTL-CDA3S system with different measures under30% of the TS database.



Testing Phase (30%)

Figure 6. Crowd classification outcome of the AOTL-CDA3S system under 30% of the TS database.

The training accuracy (TR_{acc}) and validation accuracy (VL_{acc}) acquired by the AOTL-CDA3S approach under test database is shown in Figure 7. The simulation value denoted by the AOTL-CDA3S method has attained maximal values of TR_{acc} and VL_{acc} ; the VL_{acc} is greater than TR_{acc} .

The training loss (TR_{loss}) and validation loss (VL_{loss}) reached by the AOTL-CDA3S method under the test database are displayed in Figure 8. The simulation values displayed by the AOTL-CDA3S technique have exhibited minimal values of TR_{loss} and VL_{loss} . Seemingly, the VL_{loss} is less than TR_{loss} .



Figure 7. *TR_{acc}* and *VL_{acc}* analysis of the AOTL-CDA3S system.



Training and Validation Loss

Figure 8. *TR*_{loss} and *VL*_{loss} analysis of the AOTL-CDA3S system.

A clear precision-recall inspection of the AOTL-CDA3S technique under the test database is depicted in Figure 9. The figure denotes that the AOTL-CDA3S approach has resulted in enhanced values of precision-recall values in every class label.



Figure 9. Precision-recall analysis of the AOTL-CDA3S system.

A brief ROC study of the AOTL-CDA3S approach under the test database is represented in Figure 10. The outcomes exemplified by the AOTL-CDA3S method have revealed its capability in classifying different classes in the test database.



Figure 10. ROC curve analysis of AOTL-CDA3S system.

Lastly, a comprehensive comparative study of the AOTL-CDA3S model was made with recent models in Table 5 and Figure 11 [12]. These results affirm that the AOTL-CDA3S

model has reached improved performance over other models. For instance, based on $accu_y$, the AOTL-CDA3S model has gained a higher $accu_y$ of 98%, whereas the Gabor, BoW-SRP, BoW-LBP, GLCM-SVM, GoogleNet, VGGNet, and MDTL-ICDDC models have reached reduced $accu_y$ of 71.95%, 79.95%, 84.26%, 80.20%, 84.51%, 84.49%, and 96.94% respectively.

_	-	-		
Methods	Accuracy	Precision	Sensitivity	F-Score
AOTL-CDA3S	98.00	96.11	95.87	98.66
Gabor	71.95	61.88	61.84	61.78
BoW-SRP	79.95	68.25	67.75	68.32
BoW-LBP	84.26	74.36	74.37	74.74
GLCM-SVM	80.20	75.57	73.63	87.99
GoogleNet	84.51	83.16	84.94	80.92
VGGNet	84.49	85.72	82.74	84.76
MDTL-ICDDC	96.94	93.24	93.11	93.29

Table 5. Comparative analysis of the AOTL-CDA3S system with recent approaches.



Figure 11. Comparative analysis of the AOTL-CDA3S system with recent approaches.

Additionally, based on $prec_n$, the AOTL-CDA3S method has obtained a higher $prec_n$ of 96.11%, whereas the Gabor, BoW-SRP, BoW-LBP, GLCM-SVM, GoogleNet, VGGNet, and MDTL-ICDDC methods have attained a reduced $prec_n$ of 61.88%, 68.25%, 74.36%, 75.57%, 83.16%, 85.72%, and 93.24% correspondingly.

The AOTL-CDA3S model has attained maximum crowd density classification performance. Since the SqueezeNet model has 50 times less parameters than AlexNet with effective accuracy, the proposed model reaches better performance over other models. At the same time, the SqueezeNet model derives the significant features from the preprocessed input images for the classification process. In addition, the DL model suffers from several issues: namely, gradient vanishing, overfitting, hyperparameter tuning, and extensive computation. The DL model encompasses several hyperparameters and the selection of optimal configuration for these parameters in such a high dimensional space is challenging. These hyper-parameters act as knobs, which can be tweaked during the training of the model. Because of continual deepening of the model, the number of parameters of DL models also increases quickly which results in model overfitting. At the same time, different hyperparameters (learning rate, epoch count, and batch size) have a significant impact on the efficiency of the CNN model, particularly the learning rate. It is also necessary to modify the learning rate parameter for obtaining better performance. Therefore, in this study, we employed the AO technique for the hyperparameter tuning of the SqueezeNet model, which in turn enhances the classification performance. In addition, the unique characteristics of the AO algorithm, such as faster optimization speed, global exploration ability, high search efficiency, and fast convergence speed helps to attain improved crowd classification results over other models.

5. Conclusions

For accurate crowd density classification, we have developed a new AOTL-CDA3S technique for sustainable smart cities. The presented AOTL-CDA3S technique aims to identify different kinds of crowd densities in the SC ecosystem. To accomplish this, the presented AOTL-CDA3S approach initially used the WAF technique for improving the quality of the input frames. Next, the AOTL-CDA3S model employs the AO algorithm with the SqueezeNet model for feature extraction. Finally, to classify crowd densities, the XGBoost classification model is used. The experimental validation of the AOTL-CDA3S model is tested using benchmark crowd datasets and the results are examined under distinct metrics. The comparative outcomes report the improvements in the AOTL-CDA3S method over recent state of the art techniques. In the future, the crowd density classification results can be improvised using ensemble learning with fusion based DL models. Hybrid metaheuristic algorithms can be designed to enhance the characteristics of the AO algorithm.

Author Contributions: Conceptualization, E.A. and K.T.; methodology, M.A.D.; software, I.Y.; validation, E.A., M.A.D. and M.A.; formal analysis, B.S.I.A.E.K.; investigation, E.A.; resources, B.S.I.A.E.K.; data curation, A.S.Z.; writing—original draft preparation, E.A., K.T. and M.A.; writing—review and editing, M.I.E. and E.A.; visualization, A.S.Z.; supervision, E.A.; project administration, M.A.D.; funding acquisition, E.A. All authors have read and agreed to the published version of the manuscript.

Funding: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R161), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: (22UQU4331004DSR14). Research Supporting Project number (RSP2022R459), King Saud University, Riyadh, Saudi Arabia.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article as no datasets were generated during the current study.

Conflicts of Interest: The authors declare that they have no conflict of interest. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

References

- Fan, Z.; Zhang, H.; Zhang, Z.; Lu, G.; Zhang, Y.; Wang, Y. A survey of crowd counting and density estimation based on convolutional neural network. *Neurocomputing* 2022, 472, 224–251. [CrossRef]
- Garcia-Retuerta, D.; Chamoso, P.; Hernández, G.; Guzmán, A.S.R.; Yigitcanlar, T.; Corchado, J.M. An efficient management platform for developing smart cities: Solution for real-time and future crowd detection. *Electronics* 2021, 10, 765. [CrossRef]

- 3. Sreenu, G.; Durai, S. Intelligent video surveillance: A review through deep learning techniques for crowd analysis. *J. Big Data* **2019**, *6*, 1–27. [CrossRef]
- Solmaz, G.; Baranwal, P.; Cirillo, F. CountMeIn: Adaptive Crowd Estimation with Wi-Fi in Smart Cities. In Proceedings of the 2022 IEEE International Conference on Pervasive Computing and Communications (PerCom), Pisa, Italy, 21–25 March 2022; pp. 187–196.
- Minoura, H.; Yonetani, R.; Nishimura, M.; Ushiku, Y. Crowd density forecasting by modeling patch-based dynamics. *IEEE Robot. Autom. Lett.* 2020, 6, 287–294. [CrossRef]
- 6. Fitwi, A.; Chen, Y.; Sun, H.; Harrod, R. Estimating interpersonal distance and crowd density with a single-edge camera. *Computers* **2021**, *10*, 143. [CrossRef]
- Fu, X.; Pace, P.; Aloi, G.; Li, W.; Fortino, G. Toward robust and energy-efficient clustering wireless sensor networks: A double-stage scale-free topology evolution model. *Comput. Netw.* 2021, 200, 108521. [CrossRef]
- 8. Fu, X.; Pace, P.; Aloi, G.; Yang, L.; Fortino, G. Topology optimization against cascading failures on wireless sensor networks using a memetic algorithm. *Comput. Netw.* **2020**, *177*, 107327. [CrossRef]
- Premkumar, M.; Jangir, P.; Kumar, C.; Sundarsingh Jebaseelan, S.D.T.; Alhelou, H.H.; Madurai Elavarasan, R.; Chen, H. Constraint estimation in three-diode solar photovoltaic model using Gaussian and Cauchy mutation-based hunger games search optimizer and enhanced Newton–Raphson method. *IET Renew. Power Gener.* 2022, *16*, 1733–1772. [CrossRef]
- Premkumar, M.; Jangir, P.; Sowmya, R.; Alhelou, H.H.; Mirjalili, S.; Kumar, B.S. Multi-objective equilibrium optimizer: Framework and development for solving multi-objective optimization problems. *J. Comput. Des. Eng.* 2022, *9*, 24–50. [CrossRef]
- Ding, X.; He, F.; Lin, Z.; Wang, Y.; Guo, H.; Huang, Y. Crowd density estimation using fusion of multi-layer features. *IEEE Trans. Intell. Transp. Syst.* 2020, 22, 4776–4787. [CrossRef]
- Alrowais, F.; Alotaibi, S.S.; Al-Wesabi, F.N.; Negm, N.; Alabdan, R.; Marzouk, R.; Mehanna, A.S.; Al Duhayyim, M. Deep Transfer Learning Enabled Intelligent Object Detection for Crowd Density Analysis on Video Surveillance Systems. *Appl. Sci.* 2022, 12, 6665. [CrossRef]
- 13. Wang, S.; Pu, Z.; Li, Q.; Wang, Y. Estimating crowd density with edge intelligence based on lightweight convolutional neural networks. *Expert Syst. Appl.* **2022**, 206, 117823. [CrossRef]
- 14. Almagbile, A. Estimation of crowd density from UAVs images based on corner detection procedures and clustering analysis. *Geo-Spat. Inf. Sci.* 2019, 22, 23–34. [CrossRef]
- 15. Purwar, R.K. Crowd density estimation using hough circle transform for video surveillance. In Proceedings of the 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 7–8 March 2019; pp. 442–447.
- Alashban, A.; Alsadan, A.; Alhussainan, N.F.; Ouni, R. Single Convolutional Neural Network with Three Layers Model for Crowd Density Estimation. *IEEE Access* 2022, 10, 63823–63833. [CrossRef]
- 17. Zhou, B.; Song, B.; Hassan, M.M.; Alamri, A. Multilinear rank support tensor machine for crowd density estimation. *Eng. Appl. Artif. Intell.* **2018**, *72*, 382–392. [CrossRef]
- Zhao, R.; Dong, D.; Wang, Y.; Li, C.; Ma, Y.; Enríquez, V.F. Image-Based Crowd Stability Analysis Using Improved Multi-Column Convolutional Neural Network. *IEEE Trans. Intell. Transp. Syst.* 2021, 23, 5480–5489. [CrossRef]
- Liu, W.; Yang, Y.; Zhong, J. Towards Dual-Modal Crowd Density Forecasting in Transportation Building. In Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 18–23 July 2022; pp. 1–8.
- 20. Wang, S.; Lyu, Y.; Xu, Y.; Wu, W. MSCDP: Multi-Step Crowd Density Predictor in Indoor Environment. 2022. [CrossRef]
- 21. Tang, X.; Xiao, B.; Li, K. Indoor crowd density estimation through mobile smartphone wi-fi probes. *IEEE Trans. Syst. Man Cybern. Syst.* **2018**, *50*, 2638–2649. [CrossRef]
- Li, P.; Liu, X.; Xiao, H. Quantum image weighted average filtering in spatial domain. Int. J. Theor. Phys. 2017, 56, 3690–3716. [CrossRef]
- Ucar, F.; Korkmaz, D. COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. *Med. Hypotheses* 2020, 140, 109761. [CrossRef]
- Abualigah, L.; Yousri, D.; Abd Elaziz, M.; Ewees, A.A.; Al-Qaness, M.A.; Gandomi, A.H. Aquila optimizer: A novel meta-heuristic optimization algorithm. *Comput. Ind. Eng.* 2021, 157, 107250. [CrossRef]
- Zheng, H.; Yuan, J.; Chen, L. Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation. *Energies* 2017, 10, 1168. [CrossRef]