SRTTU

Journal of Electrical and Computer Engineering Innovations
(JECEI)

Journal homepage: http://www.jecei.sru.ac.ir

JECEI

Research paper

Deep Neural Network with Extracted Features for Social Group Detection

A. Akbari, H. Farsi[*]*, S. Mohamadzadeh*

Department of Communication Engineering, Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran.

Article Info

Article History:
Received 24 May 2020
Reviewed 24 July 2020
Revised 09 September 2020
Accepted 12 November 2020

Keywords:
Social group detection
Deep neural network
Feature extraction
Video processing

[*]Corresponding Author's Email Address:
h.farsi@birjand.ac.ir

Abstract

Background and Objectives: Video processing is one of the essential concerns generally regarded over the last few years. Social group detection is one of the most necessary issues in crowd. For human-like robots, detecting groups and the relationship between members in groups are important. Moving in a group, consisting of two or more people, means moving the members of the group in the same direction and speed.

Methods: Deep neural network (DNN) is applied for detecting social groups in the proposed method using the parameters including Euclidean distance, Proximity distance, Motion causality, Trajectory shape, and Heat-maps. First, features between pairs of all people in the video are extracted, and then the matrix of features is made. Next, the DNN learns social groups by the matrix of features.

Results: The goal is to detect two or more individuals in social groups. The proposed method with DNN and extracted features detect social groups. Finally, the proposed method's output is compared with different methods.

Conclusion: In latest years, the use of deep neural networks (DNNs) for learning and detecting has been increased. In this work, we used DNNs for detecting social groups with extracted features. The indexing consequences and the outputs of movies characterize the utility of DNNs with extracted features.

Introduction

The importance of video processing has increased over time and the expanded use of a camera to detect suspicious activity in a crowd and social anomalies [1]-[9]. Seeing social groups is concerned by governments for detecting dangerous situations. To identify abnormal behavior, recognizing social groups is a prerequisite. Contextual abnormal human behavior detection is presented in [10], [11]. People's interest in moving in groups and tracking in the groups are detected by multiple cameras [12]. To walk in the group means to be a subsystem in the group; in other words, a group of two or more individuals may be considered in the same direction of motion. Walking in a group is known to be moving through crowds through the personal control of someone or other men. An individual joining a group is affected by the group, so that the group suits the person's pace and direction. Social signal processing examines social relationships, conversational relationships, and even the position of people during the conversation. In addition, it shows the importance of identifying a social group by a robot. Here, it is essential to identify social groups to match these relations. Detection of social groups to aid the behavior of the robot in teamwork with humans was reported in [13], where linear extrapolation of inter-event intervals implemented an anticipation method. Skeletal data from participants detected social groups, and the method of

Doi: 10.22061/JECEI.2020.7451.392

anticipating events was used to transfer robots among the human group [14], [15]. The technique described in [16], [17] detects the robot's conversation and social classes, using people's direction, and lower-body orientation. Clustering games are applied for detecting conversation groups in image and video [8], [18]. Deep matching-based pairwise potential with a conventional spatiotemporal relation-based pairwise potential was used to track many people in the video [19]. In this article, social group is detected by deep neural network (DNN).

The rest of this paper is structured as follows: A summary of the literature is discussed in 2th Section. The fundamentals of a DNN are outlined in 3th Section. The proposed method is defined in 4th Section. The experimental findings are shown in 5th Section. The premise is eventually set out in 6th Section.

Related Work

Human analysis applications in a crowded scene are divided into four categories: visual surveillance, crowd management, public space and entertainment design [20]. Visual surveillance is a system of monitoring activity in a building or area. Crowd management is important to manage crowd areas like stadiums. Public space is generally accessible to people. Human analysis is important for Public space and Entertainment design.

In this article, social group detection is classified into three categories: Support Vector Machine (SVM)-based, clustering-based, and deep neural network-based detection.

In SVM-based methods, sociological features between pairs of trajectories and supervised learning are applied for group detection. Body and head orientation features extracted from the social scene are used for detecting the social groups. An electric dipole and each person's eyesight are used for this purpose. If a connection is found between the eyes of the peoples, this group of people is put within a social group [21]. The relationship between people is identified using graph-based clustering in the method reported in [22], and the SVM based classification detects further social group activity. Spatial proximity prior, similarity properties prior, and spatiotemporal closure prior are used to detect social group in the crowd. Features are used to train SVM for human group detection [23].

Hierarchical clustering is used to classify groups within society. Tracking of salient points and adaptive clustering is used to identify hierarchical social classes. To detecting social groups, agglomerative hierarchical clusters with pair proximity and speed are applied [24].

The Density-Based clustering algorithm is used to detect the pedestrian groups. Spatiotemporal-oriented energy, slight motion energies, inter-group flow-field distance, and bottom-up hierarchical clustering are calculated for this purpose. If the inter-group flow-field distance between two persons is less than a threshold, then two persons are considered as a social group [25]. Generating coherent filtering clusters, anchor tracklet, seeding tracklets are applied for group detection and crowd understanding [26].

In DNN Based method, data-driven Generation of Socially Acceptable Trajectories (App-LSTM) is employed for detecting small groups. The App-LSTM consists of position LSTM encoder, orientation LSTM encoder, and Group Interaction Module (GIM). A deep affinity network is used based on position and orientation to detect conversational groups. For this purpose, an interaction graph is made based on location and orientation. Then pairwise affinities are computed, and the affinity matrix is made. Finally, the conversational group is detected [27]. Generative adversarial networks and LSTM encoder are used in crowds for the estimation of trajectories and group identification [28].

In recent years, DNNs have been used in many fields of science and proved their applications [29]-[30]. The Relu and Softmax functions are the most commonly used activation functions. In the ReLU activator function, if the value of the input is greater than 0, the output is equal to the input.

On the other hand, if the value of the input is less than 0, the output is 0. The main advantage of using the ReLU activator function is that it has a constant derivative for all inputs greater than 0. Better and faster network learning is the result of this constant derivative. In [31], it is proven that supervised coaching of DNNs is a good deal quicker, if the hidden layers are composed of the ReLU.

Softmax activation functions are usually used in the output layer for classification purposes. The outputs of this function are normalized to a range of 0 to 1 [32]. With the softmax activation function, the categorization becomes very simple and the output is finally probabilistic.

The Proposed Method

The Architecture of the proposed method is shown in Fig. 1.

The DNN in the Proposed Method is designed with input layers, three intermediate layers with 200 nodes, and one output layer. The output of each input layer is the input of the next layer. The structure of the DNN is shown in Fig. 1. The DNN used in the proposed method has five inputs and one output. The number of entries is equal to the number of features between pairs. If the output is equal to 1, two persons are in the same social group, and if the output is 0, two persons are not in the same social group. DNN training is performed on training data. In the training phase, weights and bias of all layers of the DNN are computed.

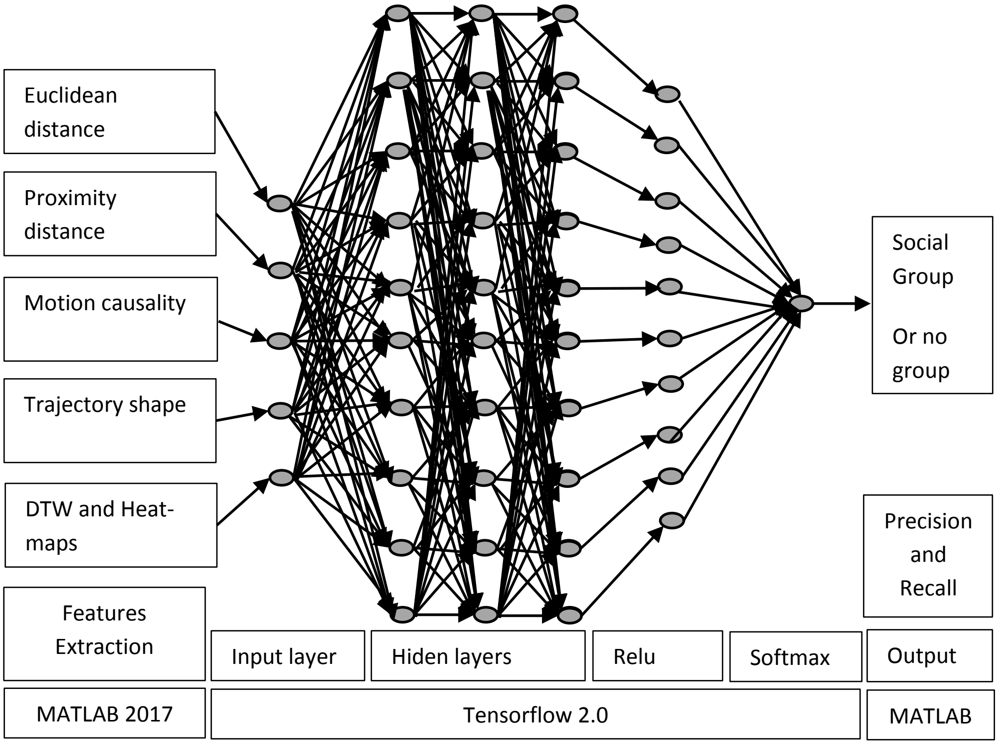

Fig. 1: The Architecture of the proposed method.

Training and experimentation in identifying social groups require extracting attributes between two or more persons. In the following, the features used in the proposed method are defined. In this method, a row of feature matrices are defined for persons 'a' and 'b' as:

$$d = \left[a, b, d_{eu}, d_{ph}, d_{sh}, d_{ca}, d_{he}, i, g\right]_{a,b} \qquad (1)$$

The feature matrix consists of the label number of individuals, five features between two persons, the number of videos in the dataset, and the parameter represent two persons 'a' and 'b' in one group. Here, d_{eu} is Euclidean distance, d_{ph} is the Proximity distance, d_{sh} is motion causality, d_{ca} is trajectory shape, and d_{he} is the heat map between persons 'a' and 'b'. In (1), the number of video database is 'i'. The parameter 'g' represents whether persons 'a' and 'b' are in one group. If 'g' is equal to one, persons 'a' and 'b' are in the same group, and if 'g' is equal to 0, it means that two persons are not in the same social group.

All five features are calculated between the two presented people in the video. In the matrix, the first and second columns of each row are the label number of persons, and the third column to the seventh column shows the features.

There are a large number of people in the video, and finding five features between them cause high computational costs.

For example, in the feature matrix in the first movie of the dataset, 'i', the number of videos is one, and the number of people in the first movie is 48.

Then, 1128 rows of feature matrix represent the features between them. In the second movie of the dataset, 'i' is equal to two, the number of people in the second movie is 49, and then the next 1,176 rows of feature matrix represent the features between them. In the following, the features used in the proposed method are introduced.

Euclidean distance between all pairs is computed for each frame of videos. The mean of Euclidean distance in all frames is defined as the Euclidean distance feature for each pair.

Investigating the proximity distance of individuals in a group is one of the most critical issues of Social Group Representation. Figure 2 illustrates the importance of a Proximity distance feature. Proximity theory describes the theory of proximity based on the physical distance property [21].

According to proximity theory, there are four classes for separating the type of relationship between individuals based on the distance of two individuals. Each class is separated by the boundary shown in Fig. 2. Intimate class is lower than 0.5 meter, between 0.5 and 1.2 meter is personal class, between 1.2 and 3.7 meter is social class and between 3.7 and 7.6 meter is public class.

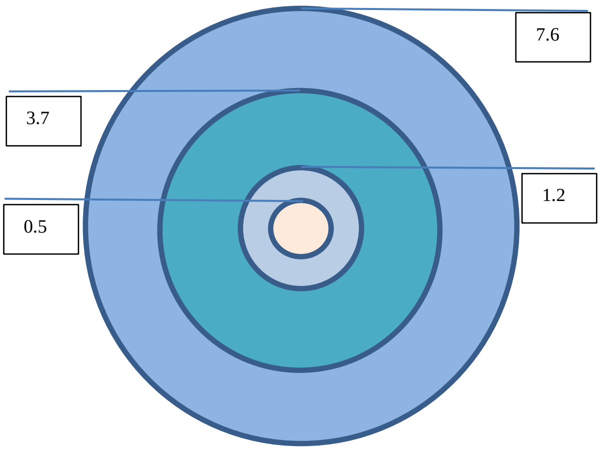

Fig. 2: Proximity theory and the importance of physical distance in identifying social groups.

The proposed method uses a Gaussian mixture model inspired by the proximity theory. The Gaussian mixture model [23] is expressed by (2):

$$GMM(P_a^t - P_b^t) = \frac{1}{4}\sum_{i=1}^{4} N(P_a^t - P_b^t|0,\sigma_i) \qquad (2)$$

The Gaussian mixture model between two persons ‘a’ and ‘b’ is obtained at moment t, where P_a^t is the position of the person ‘a’ at moment t. In this model, N is a simple Gaussian model with zero mean. The variances are the boundaries of the proximity theory, namely 0.5, 1.2, 3.7, and 7.6. Consequently, between two persons ‘a’ and ‘b’, at moment t, the four variances are calculated, and the final output is the mean of four Gaussians with different variances. Next, for the entire consecutive frame in the video between the two people, it is necessary to calculate the average Gaussian mixture model based on the next relation [23].

$$d_{ph} = \frac{1}{T}\sum_{t\epsilon T} GMM(P_a^t - P_b^t) \qquad (3)$$

Equation 3 extracts the proximity distance feature for identifying the social groups, where T is the total video time. The output of the physical distance attribute between the two people will be a numerical value.

Motion causality is extracted from the movement of people. All datasets provide people’s location in each frame. Based on this information, $\bar{V}_a(n)$ is the stored information about the path traveled from (t-n to t-1). $P\left(V_a(t)|\bar{V}_a(n)\right)$ is the trajectory predictor of a person where is calculated based on the ‘n’ previous frames. Figure 3 illustrates the similarities of the movement of individuals in successive frames.

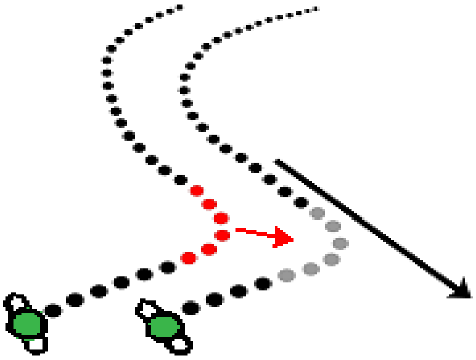

Fig. 3: Motion causality feature.

This problem is formulated as follows [23]:

$$RSS_c = \sum_{t=1}^{T}\left(V_a(t) - P\left(V_a(t)|\bar{V}_a(n)\right)\right)^2 \qquad (4)$$

where $\bar{V}_a(n)$ is the stored information about the path traveled by the person ‘a’ from (t-n to t-1). Trajectory predictor of the person ‘a’ with the stored information of two persons ($\bar{V}_a(n)$ for the person ‘a’ and $\bar{V}_b(n)$ for the person ‘b’) is $P\left(V_a(t)|\bar{V}_a(n),\bar{V}_b(n)\right)$. This trajectory predictor with information of two persons is formulated as follows [23]:

$$RSS_u = \sum_{t=1}^{T}\left(V_a(t) - P\left(V_a(t)|\bar{V}_a(n),\bar{V}_b(n)\right)\right)^2 \qquad (5)$$

Where trajectory predictor is computed with information of one person in the (5) and trajectory predictor is computed with information of two-person in the (6). In (6), the similarity of these trajectory predictors is computed [23].

$$\mathrm{S}_{\mathrm{b}\to\mathrm{a}} = \frac{(\mathrm{RSS}_\mathrm{c} - \mathrm{RSS}_\mathrm{u})/\mathrm{n}}{\mathrm{RSS}_\mathrm{u}/(\mathrm{T} - 2\mathrm{n} - 1)} \qquad (6)$$

In the (6), T is the time of each video. In the (7), Fisher-Snedecor probability function is used to extract the motion causality feature [23].

$$\mathrm{d_{ca}(a,b)} = \max_{\mathrm{S}\in\{\mathrm{S_{b\to a}},\mathrm{S_{a\to b}}\}} \int_0^{\mathrm{S}} \mathrm{F(x|n,K-2n-1)dx} \qquad (7)$$

In the (7), the similarity characteristic of the changes of movement of individuals ‘a’ and ‘b’ in successive frames is computed.

Trajectory shape is extracted from the Dynamic Time Warping algorithm between two-time sequences. Figure 4 presents the importance of resembling the shape of the path taken by individuals in identifying social groups.

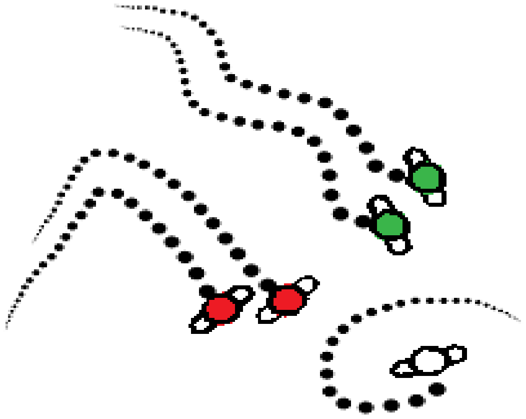

Fig. 4: Trajectory shape feature.

To compute the cumulative cost function $\gamma_{ab}(i,j)$, it is necessary to calculate the Euclidean distance matrix of the path of the person 'a' at the first moment with the path of the person 'b' at the first moment (D_{ab}^{11}). Here, D_{ab}^{ij} is the Euclidean distance matrix of the path of the person 'a' at the moment i, with the path of the person 'b' at moment j.

Dimensions of the Euclidean distance matrix are f and g, which represent the path traveled by the person 'a' and 'b', respectively.

The cumulative cost function is defined as follows [23]:

$$\gamma_{ab}(i,j) = D_{ab}^{ij} + min\{\gamma_{ab}(i-1,j) + \gamma_{ab}(i-1,j-1) + \gamma_{ab}(i,j-1)\} \tag{8}$$

To calculate the cumulative cost function $\gamma_{ab}(i,j)$, it is necessary to calculate D_{ab}^{11}, which is equal to $\gamma_{ab}(1,1)$, based on (8).

Then, $\gamma_{ab}(i,j)$ is obtained based on the location of two individuals in the time sequence. Finally, the similarity property of the shape of the path followed by individuals is obtained from (9).

$$d_{sh}(a,b) = \gamma_{ab}(f,g)/\max(f,g) \tag{9}$$

Since the two numbers of f and g can be different, trajectory shape is divided into the maximum of these two numbers.

Heat map parameter for person 'a' is obtained using (10) based on the average heat map in a small rectangular plot of p and q.

The rectangular segments R and C are derived from the path information obtained by individuals 'a' and 'b', which are equal to the rectangle that two people traveled in and out of the time window [23].

$$H_a(i,j) = \sum_{p=1}^{R}\sum_{q=1}^{C} \bar{E}(p,q)e^{-\tau\|(p-i,q-j)\|} \tag{10}$$

where τ in (10) is the time window of the video. Then, the features of the DTW and Heat-maps between individuals 'a' and 'b' are estimated by [23]:

$$d_{he}(a,b) = \sum_{i=1}^{R}\sum_{j=1}^{C} H_a(i,j)H_b(i,j) \tag{11}$$

The output of (11) is the similarity feature of the DTW and heat map between individuals 'a' and 'b', indicating the sharing of heat maps of individuals 'a' and 'b' in the rectangular segment R and C.

Evaluation

MATLAB 2017 software was used to extract the feature and TensorFlow 2.0 was used for training and testing of deep learning. The output of the TensorFlow is a matrix, and the first and second columns of this matrix are people's number.

If the third column of this matrix is one, two people in the corresponding row are in the same group, and if the third column is zero, the two people in the relevant row are not in the same social group. MATLAB 2017 has been used to identify groups more than two person and to display the output on video.

If the person with number of 40 is in the same group with the person with number of 43, the person with number of 43 and the person with number of 44 are in the same group, and the persons with numbers of 40 and 44 are in the same group, a three-person group consists of people with numbers 40, 43 and 44 are made.

Due to a large number of people in the video as well as people crossing different paths, it is difficult to identify groups. Crossing groups together also complicates the problem. Student003 [8], GVEII [22], ETH [28], Hotel [28], and MPT 20x100 [23] datasets were used to evaluate the efficiency of the proposed method. Table 1 lists the features of these five databases.

In Table 1, the variable of v is the number of videos, p is the number of participants, and g is the number of groups. The variables of d1 and d2 are the minimum and the maximum distances to a person (in meters), respectively. These datasets include people's route data and the number of people. The third column of the table shows the number of people in the five databases. The average distance between groups in the five datasets varies, indicating the need for training.

Table 1: Comparison of five datasets

	v	P	G	d1	d2
ETH	1	117	18	0.99	2.79
Hotel	1	107	11	0.75	2.0
Student 003	20	406	108	0.41	0.70
GVEII	30	630	207	0.77	1.66
MPT 20x100	20	82	10	0.63	1.45

The results are compared using the precision and recall parameters.

The precision is the ratio of the quantity of organizations successfully recognized to the variety of all the agencies recognized.

The recall is the ratio of the variety of agencies efficaciously listed in the database to the number of faulty companies. The standard F-score, F1, is set out as follows:

$$F1 = \frac{2 * \text{precision} * \text{recall}}{\text{precision } + \text{recall}} \tag{12}$$

Because ETH, Hotel, student003 (CBE), GVEII, and MPT-20x100 datasets have unique properties, actual situations are evaluated with them. We current the effects output from the proposed technique in this section, and look at them with distinct methods. The results of the proposed method in each database are compared to existing methods.

In [23], SVM based classification was used. Hierarchical clustering was reported in [24]. In [26], generating coherent filtering clusters was used. In [28], Generative adversarial networks and LSTM encoder were used.

Table 2 shows the results of the proposed method in the ETH and Hotel datasets.

In Table 2, the letter ‘P’ indicates precision, and the letter ‘R’ indicates recall. ETH and Hotel datasets consist of one video and have less information than student003 (CBE), GVEII, and MPT-20x100 datasets. Deep neural network output in ETH and Hotel datasets is much weaker than existing methods.

This means that little information for the deep neural network will not lead to the desired result.

Table 3 shows the results of the proposed method in the CBE dataset.

Table 2: Compare precision and recall for the ETH and Hotel datasets

	ETH		Hotel	
	P	R	P	R
Proposed method	62.5	59.7	65.2	55.9
[23]	91.1	83.4	89.1	91.3
[24]	80.7	80.8	88.9	89.3
[26]	69.3	68.2	67.3	64.1
[28]	91.3	83.5	90.2	93.1

Table 3. Compare precision and recall for the CBE dataset

	Precision	Recall
Proposed method	81.8	80.6
[23]	82.3	74.1
[24]	72.2	65.1
[26]	10.6	76.0
[28]	82.1	63.4

As shown in Table 3, the recall of the technique has an exceptional application, and the precision of the [23] method has the best application. The execution of the proposed method in the CBE dataset is shown in Fig. 5.

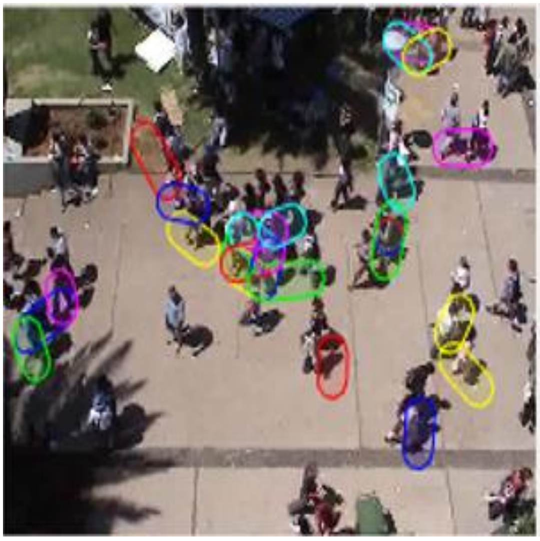

Fig. 5: The output of the social groups detected by the proposed method in the CBE dataset.

Table 4 shows the results of the proposed method in the GVEII dataset. GVEII dataset has 30 video and the most data for training and test of the deep neural network.

Table 4. Compare precision and recall for the GVEII dataset

	Precision	Recall
Proposed method	80.9	81.5
[23]	79.7	77.5
[28]	77.6	73.1

As shown in Table 4, the precision and recall of the proposed method are the best among all methods. The execution of the proposed method in the GVEII database is presented in Fig. 6.

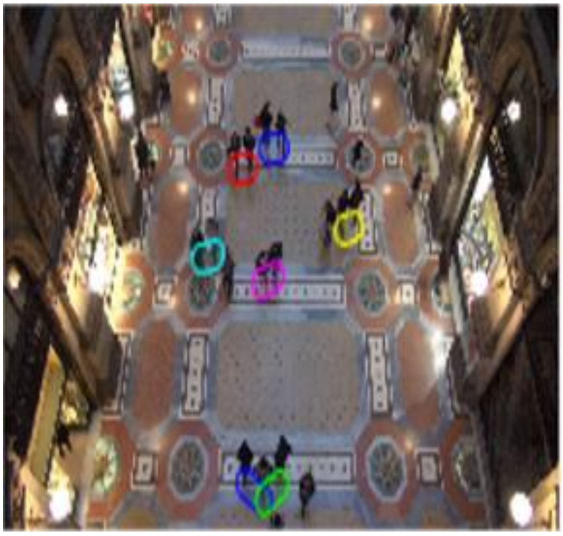

Fig. 6: The output of the social group detection by the proposed method in the GVEII dataset.

In Fig. 7 present the results of the proposed method in the MPT-20x100 dataset.

As shown in Fig. 8, the F-score of the proposed method has the best performance in airport1, 1chinacross4, 1dawei5, 1grand1, 1japancross2, 1manko3, 1manko29, 1shatian3, 2dawei1, 2jiansha5, 2manko2, and 2niurunning2. In Fig. 8, SCSL is [23] and VASG is [24].

The results of the proposed method in the ETH, Hotel, student003 (CBE), GVEII, and MPT-20x100 datasets showed that small datasets like the ETH, and Hotel datasets have little data for deep neural network training.

Large databases like the CBE, and GVEII databases have enough data to train deep neural networks, and the performance of this system in databases with high information is acceptable.

MPT-20x100 dataset has one video in some scenes and two different videos in some scenes. As a result, in videos containing only one video in scene, MPT-20x100 dataset does not perform well in the deep neural network.

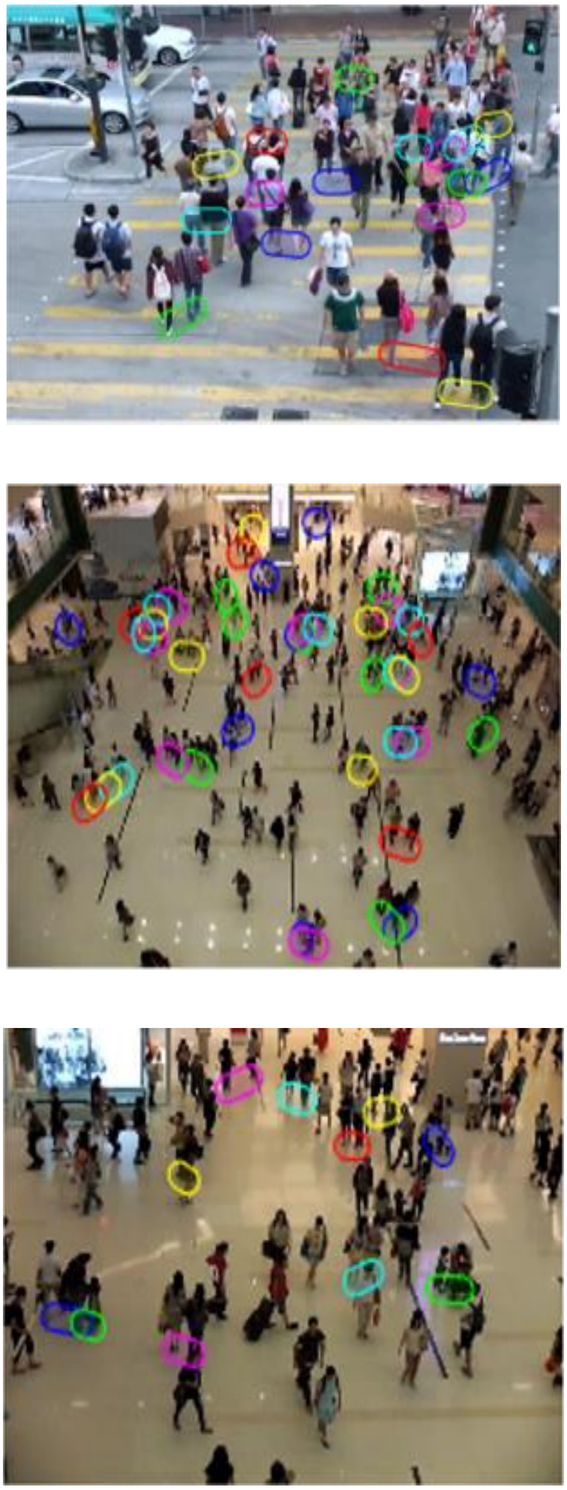

Fig. 7: The output of social group detection by the proposed method in the MPT-20x100 dataset.

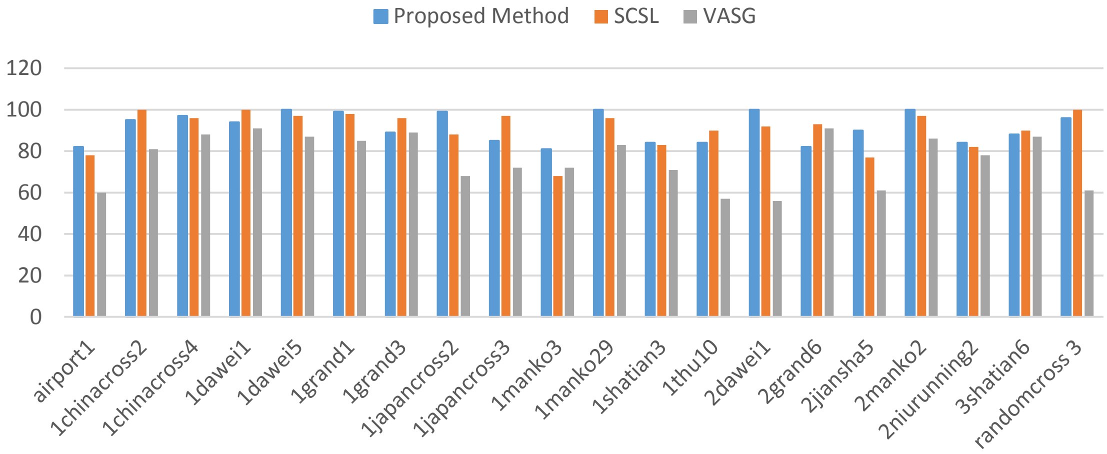

Fig. 8: Compare F-score for the MPT-20x100 dataset.

Conclusion

In this work, we used DNNs for detecting social groups with extracted features. Social group detection is one of the most necessary issues involved these days in the evaluation of interpersonal members in groups. In latest years, the use of deep neural networks (DNNs) for learning and detecting has been increased. The indexing consequences and the outputs of movies characterize the utility of DNNs with extracted features.

Author Contributions

A. Akbari, H. Farsi, and S. Mohamadzadeh designed the experiments. A. Akbari collected the data. A. Akbari carried out the data analysis. A. Akbari, H. Farsi, and S. Mohamadzadeh interpreted the results and wrote the manuscript.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Acknowledgement

The authors gratefully acknowledges the supports provided by Electrical and Computer Engineering department from University of Birjand, Birjand, Iran.

Abbreviations

DNN	Deep neural network
GIM	Group Interaction Module
SVM	Support Vector Machine
LSTM	Long short-term memory

References

[1] S. Guo, Q. Bai, S. Gao, Y. Zhang, A. Li, "An analysis method of crowd abnormal behavior for video service robot," IEEE Access, 7: 169577 – 169585, 2019.

[2] S.M. Hosseini, H. Farsi, H.S. Yazdi, "Best clustering around the color images," Int. J. Comput. Electr. Eng., 1(1): 20-24, 2009.

[3] P. Etezadifar, H. Farsi, "A new sample consensus based on sparse coding for improved matching of SIFT features on remote sensing images," IEEE Trans. Geosci. Remote Sens., 58(8): 5254-5263, 2020.

[4] A. Jalali, H. Farsi, "A new steganography algorithm based on video sparse representation," Multimed. Tool. Appl., 79(3): 1821-1846, 2020.

[5] R. Nasiripour, H. Farsi, S. Mohamadzadeh, "Visual saliency object detection using sparse learning," IET Image Proc., 13(13): 2436-2447, 2019.

[6] M. Hasheminejad, H. Farsi, "Sample-specific late classifier fusion for speaker verification," Multimed. Tool. Appl., 77(12): 15273-15289, 2018.

[7] H. Farsi, R. Nasiripour, S. Mohammadzadeh, "Eye gaze detection based on learning automata by using SURF descriptor," J. Inf. Syst. Telecommun., 6(1): 41-49, 2018.

[8] M. Hasheminejad, H. Farsi, "Frame level sparse representation classification for speaker verification," Multimed. Tool. Appl., 76(20): 21211-21224, 2017.

[9] H. Farsi, R. Nasiripour, S. Mohammadzadeh, "Improved generic object retrieval in large scale databases by SURF descriptor," J. Inf. Syst. Telecommun., 5(2): 128-137, 2017.

[10] O.P. Popoola, K. Wang, "Video-based abnormal human behavior recognition—A review," IEEE Trans. Syst. Man Cybern. Part C Appl. Rev., 42(6): 865-878, 2012.

[11] H. Farsi, "Improvement of minimum tracking in minimum statistics noise estimation method," Signal Process. Int. J. (SPIJ), 4(1): 17-22, 2010.

[12] F. Solera, S. Calderara, E. Ristani, C. Tomasi, R. Cucchiara, "Tracking social groups within and across cameras," IEEE Trans. Circuits Syst. Video Technol., 27(3): 441-453, 2017.

[13] T. Iqbal, M. Moosaei, L. D. Riek, "Tempo adaptation and anticipation methods for human-robot teams," in Proc. RSS, Planning HRI: Shared Autonomy Collab. Robot. Workshop: 1-3, 2016.

[14] T. Iqbal, S. Rack, L.D. Riek, "Movement coordination in human–robot teams: a dynamical systems approach," IEEE Trans. Rob., 32(4): 909-919, 2016.

[15] X.-T. Truong, T.-D. Ngo, "“To approach humans?”: A unified framework for approaching pose prediction and socially aware robot navigation," IEEE Trans. Cognit. Dev. Syst., 10(3): 557-572, 2017.

[16] M. Vázquez, A. Steinfeld, S. E. Hudson, "Parallel detection of conversational groups of free-standing people and tracking of their lower-body orientation," in Proc. Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference,: 3010-3017, 2015.

[17] H. Farsi, M. Mozaffarian, H. Rahmani, "Improving voice activity detection used in ITU-T G. 729. B," presented at the 3rd WSEAS International Conference on Circuits, Systems, Signal and Telecommunications, Ningbo, China, 2009.

[18] S. Vascon, M. Pelillo, "Detecting conversational groups in images using clustering games," in Multimodal Behavior Analysis in the Wild: Elsevier, 12: 247-267, 2019.

[19] S. Tang, B. Andres, M. Andriluka, B. Schiele, "Multi-person tracking by multicut and deep matching," in Proc. European Conference on Computer Vision: 100-111, 2016.

[20] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, S. Yan, "Crowded scene analysis: A survey," IEEE Trans. Circuits Syst. Video Technol., 25(3): 367-386, 2015.

[21] H.S. Park, J. Shi, "Social saliency prediction," in Proc. Computer Vision and Pattern Recognition (CVPR), IEEE Conference: 4777-4785, 2015.

[22] K.N. Tran, A. Bedagkar-Gala, I.A. Kakadiaris, S. K. Shah, "Social cues in group formation and local interactions for collective activity analysis," in Proc. International Conference on Computer Vision Theory and Applications (VISAPP): 539-548, 2013.

[23] F. Solera, S. Calderara, R. Cucchiara, "Socially constrained structural learning for groups detection in crowd," IEEE Trans. Pattern Anal. Mach. Intell., 38(5): 995-1008, 2016.

[24] W. Ge, R.T. Collins, R.B. Ruback, "Vision-based analysis of small groups in pedestrian crowds," IEEE Trans. Pattern Anal. Mach. Intell., 34(5): 1003-1016, 2012.

[25] S. Huang, D. Huang, M.A. Khuhroa, "Social pedestrian group detection based on spatiotemporal-oriented energy for crowd video understanding," KSII Trans. Internet Inf. Syst., 12(8): 3769-3789, 2018.

[26] J. Shao, C.C. Loy, X. Wang, "Learning scene-independent group descriptors for crowd understanding," IEEE Trans. Circuits Syst. Video Technol., 27(6): 1290-1303, 2017.

[27] M. Swofford, J.C. Peruzzi, M. Vázquez, R. Martín-Martín, S. Savarese, "DANTE: deep affinity network for clustering conversational interactants," arXiv preprint arXiv:1907.12910, 2019.

[28] T. Fernando, S. Denman, S. Sridharan, C. Fookes, "GD-GAN: generative adversarial networks for trajectory prediction and group detection in crowds," in Proc. Asian Conference on Computer Vision,: 314-330, 2018.

[29] A. Sezavar, H. Farsi, S. Mohamadzadeh, "A modified grasshopper optimization algorithm combined with cnn for content based image retrieval," Int. J. Eng., 32(7): 924-930, 2019.

[30] A. Sezavar, H. Farsi, S. Mohamadzadeh, "Content-based image retrieval by combining convolutional neural networks and sparse representation," Multimed. Tool. Appl., 78: 20895-20912, 2019.

[31] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning," Nature, 521: 436-444 , 2015.

[32] B. Zhao, J. Feng, X. Wu, S. Yan, "A survey on deep learning-based fine-grained object classification and semantic segmentation," Int. J. Autom. Comput. , 14(2): 119-135, 2017.

Biographies

Ali Akbari received his B.S., M.S. and PhD degree in Communication Engineering from University of Birjand, Birjand, Iran, in 2020. His research interests are the fields of computer vision, image processing, video processing, machine learning and deep learning.

Hassan Farsi received the B.Sc. and M.Sc. degreed from Sharif University of Technology, Tehran, Iran, in 1992 and 1995, respectively. Since 2000, he started his Ph.D. in the Center of Communications Systems Research (CCSR), University of Surrey, Guildford, UK, and received the Ph.D. degree in 2004. He is interested in speech, image and video processing on wireless communications. Now, he works as professor in communication engineering in department of Electrical and Computer engineering, university of Birjand, Birjand, Iran.

Sajad Mohamadzadeh received the B.Sc. degree in electrical engineering from Sistan & Baloochestan, University of Zahedan, Iran, in 2010. He received the M.Sc. degree in telecommunication engineering from university of Birjand, Birjand, Iran, in 2012. He received the Ph.D. degree in telecommunication engineering from university of Birjand, Birjand, Iran, in 2016. He is currently an academic staff in Department of electrical and computer engineering, university of Birjand, Birjand, Iran. His area research includes image processing and retrieval, pattern recognition, digital signal processing and sparse representation.

How to cite this paper:

A. Akbari, H. Farsi, S. Mohamadzadeh, "Deep neural network with extracted features for social group detection," J. Electr. Comput. Eng. Innovations, 9(1): 47-56, 2021.

DOI: 10.22061/JECEI.2020.7451.392

URL: http://jecei.sru.ac.ir/article_1483.html