Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2021) 29: 717 – 729

doi:10.3906/elk-2003-124

An adversarial framework for open-set human action recognition using skeleton data

Özge ÖZTİMUR KARADAĞ*
Department of Computer Engineering, Rafet Kayış Faculty of Engineering, Alanya Alaaddin Keykubat University, Antalya, Turkey

Received: 21.03.2020 • **Accepted/Published Online:** 31.05.2020 • **Final Version:** 30.03.2021

Abstract: Human action recognition is a fundamental problem which is applied in various domains, and it is widely studied in the literature. Majority of the studies model action recognition as a closed-set problem. However, in real-life applications it usually arises as an open-set problem where a set of actions are not available during training but are introduced to the system during testing. In this study, we propose an open-set action recognition system, human action recognition and novel action detection system (HARNAD), which consists of two stages and uses only 3D skeleton information. In the first stage, HARNAD recognizes a given action and in the second stage it decides whether the action really belongs to one of the a priori known classes or if it is a novel action. We evaluate the performance of the system experimentally both in terms of recognition and novelty detection. We also compare the system performance with state-of-the-art open-set recognition methods. Our experiments show that HARNAD is compatible with state-of-the-art methods in novelty detection, while it is superior to those methods in recognition.

Key words: Open-set recognition, novelty detection, human action recognition, adversarial networks

1. Introduction

Human action recognition is the determination of the type of action from a given sequence of images. It is a fundamental problem in various domains such as robotics, human–computer interaction, surveillance systems, and video indexing. It has been widely studied in the last two decades and a great number of studies for human action recognition has been proposed. [1–7]

In the early days of the millennium, the area attracted the attention for security reasons and surveillance systems are thoroughly studied in those days [8, 9]. Later, with the emergence of the deep learning methods robotics systems have advanced considerably and they have come closer to becoming a part of daily life [10]. Thus, researchers pay more attention to the human–robot interaction systems, and human action recognition is a fundamental problem in these interaction systems.

Early studies of human action recognition mostly rely on RGB data. As the data acquisition systems improved, 3D data have become available and recently most promising algorithms for action recognition either directly use 3D data or use it as complementary with the 2D data. 3D data can be in the form of RGB+depth information or it can be in the form of skeleton data which is represented by the 3D positions of body joints [11, 12]. In this study, we are merely interested in action recognition using skeleton data. The first advantage of using skeleton data is that it transforms the action recognition into a view invariant problem. The second

*Correspondence: ozge.karadag@alanya.edu.tr

advantage is that it reduces the time complexity of action recognition especially when compared to pixel level processing of data via deep networks. In the last decade with the emergence of deep learning techniques, the robotic systems advanced considerably, which resulted in a need for better interpretation of human activities. This contributed to the increasing interest of both robotic and human–computer interaction researchers in the action recognition systems.

Classical approaches take the problem as a closed-set problem assuming that all type of activities are available during training. However, this is not the case for the real world problems. Most of the time, a set of activities which are not available during training arises during testing. In this case the systems need to distinguish the new action from a priori known activities and detect it as a novel action. Yang et al. [13] propose an open-set solution for human action recognition using radar signals. The main idea in that study is to create a negative dataset by generating samples using generative adversarial networks (GANs).

The task of classifying a given sample as unseen, in other words, detecting that the sample is in some respect different from the data processed during training is referred as outlier detection or novelty detection. Although researchers propose novelty detection studies for various domains, novelty detection is not studied thoroughly for the human action recognition problem. Majority of the novelty detection problems handle the problem as a one-class classification problem where the training data is used to construct a system which models this data, and it is employed to classify test samples as normal and abnormal based on their conformance with the constructed model.

Generative models are very suitable for the novelty detection problem and they are already used for novelty detection in the one-class classification formulation of the problem. In this study we propose a two-staged framework for simultaneous recognition and novelty detection. In the first stage it performs multiclass classification by means of a generative adversarial network and in the second stage it performs novelty detection by means of a set of deep expert detectors.

We can summarize the main contributions of this study as follows. Firstly, the proposed system solely uses 3D skeleton information, which is fast, and effective for robotic research. Secondly, the proposed system introduces a novelty detection system for action recognition domain, which is not thoroughly studied in the literature. Thirdly, the proposed system proposes a simultaneous solution for action recognition and novelty detection using GANs. As far as our research reveals, this is the first attempt to formulate the two problems together in the action recognition domain.

We start with providing brief information about related studies from the literature in Section 2 and then we give the details of the proposed system in Section 3. We provide our experimental studies in Section 4 and we complete the study with concluding remarks in Section 5.

2. Related work

2.1. Human action recognition

Aggarwal and Xia [5] reviewed action recognition using 3D data. They summarize various methods for gathering 3D information for action recognition and they discuss the advantages and disadvantages of those techniques. In this study, 3D information is in the form of skeleton data. The advantage of using skeleton data is that it is view-invariant and also its execution time is less compared to the RGB+D data.

There are a great number of studies in the human action recognition literature, which use deep learning architectures. Ji *et al.* [14] proposed modeling spatio-temporal information using 3D convolutional neural networks instead of using complex handcrafted features for the classification task. Wang et al. [15] use

convolutional neural networks and long short term memory units with an attention model to learn spatio-temporal features.

In [16], the authors propose utilizing pose data and appearance data separately and combine their action recognition results via a fully-connected layer. Instead of a 3D convolutional model, they incorporate time data into 2D convolutional model via an image like representation.

2.2. Open-set recognition

Support vector machines (SVM) are widely used for one-class classification or novelty detection problem. In [17, 18], the authors focus on SVM models and propose an adaptation of these models to the open-set recognition problem such that the generalization and specialization of these models are improved.

One straightforward way of novelty detection via deep networks is the thresholding of the softmax layer. Researchers also propose to extend the capability of deep networks for the open-set recognition problem [19, 20]. For this purpose, the authors in [19] propose using the fully connected layer before the softmax layer to decide if a sample is from a known or an unknown category and they modify the softmax layer to include novel classes. The authors in [20] propose detecting novel samples by thresholding and then reconstructing the classification layer by adding new predictors for new categories.

The authors in [21] employs a voting-based mechanism to detect novel samples in the action recognition problem using visual features and utilizes zero-shot learning for the classification of the unknown sample.

2.3. Generative adversarial networks (GANs)

GANs are proposed by Goodfellow et al. [22] for supervised learning and then it has been reformulated and adapted to various problems. Radford et al. [23] propose deep convolutional generative adversarial networks for learning representations from unsupervised data. Salimans et al. [24] develop techniques for improving the performance of GANs and they employ those techniques to the semisupervised classification problem. The loss function of GAN is not a good representative on the quality of the generated data. Arjovsky et al. [25] propose to overcome this problem by using Wasserstein GAN loss function. Gulrajani et al. [26] introduce a new gradient term to the Wasserstein GAN loss. Saliman et al. [24] propose using feature matching loss function for the training of the generator. This loss function ensures that the features of the generated data matches the statistics of the real data. For the open-set recognition problem, Yang et al. [13] propose employing GANs to create a negative class dataset by means of synthetic data generation. For novelty detection, Sabokrou et al. [27] propose an adversarial framework for one-class classification.

3. Human action recognition and novel action detection

In this study we propose a human action recognition and novel action detection system (HARNAD) which resolves the human action recognition and novel action detection tasks simultaneously using an adversarial approach. The system architecture of HARNAD is provided in Figure 1. The system first processes given data to obtain 3D pose data and then processes this data in two stages. In the first stage, recognition is achieved by means of a deep network which is employed as a generalist to classify each sample into one of the a priori known classes. In the second stage, novelty detection takes place by means of a set of GANs which are employed as specialists to determine if the classification in the previous step is correct or if the sample is from an unknown class.

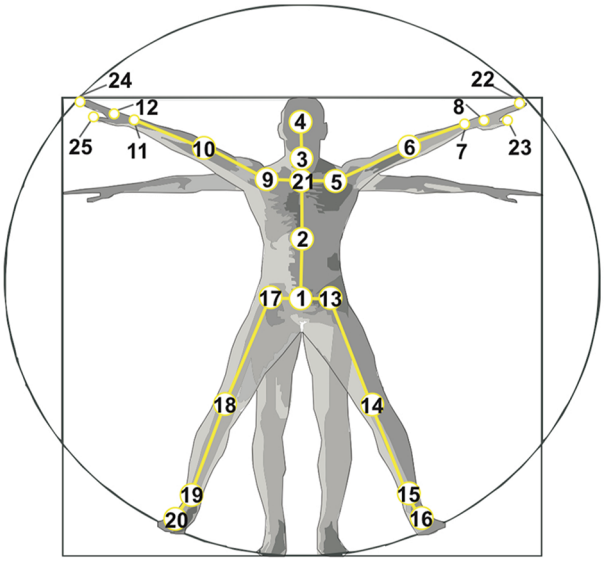

Figure 1. Body joints recorded in each frame [28].

3.1. Human action representation

Action is represented as a set of frames. Each frame is represented by the joint coordinates as provided in Figure 2. Each frame contains the 3D coordinate information from N body joints. Our representation scheme which is similar to that in [16] is presented as "Representation" in the shaded rectangle of Figure 1. In this figure, pose information is represented in an image-like form, where the x,y,z coordinates are represented as the dimension, and the vertical axis encodes different frames and the horizontal axis encodes body joints.

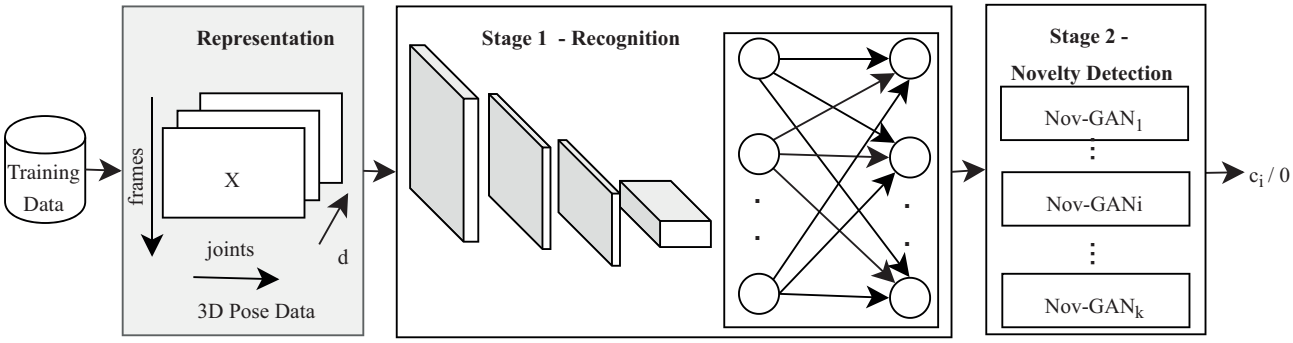

Figure 2. System architecture of HARNAD.

3.2. Human action recognition at stage 1

At stage 1 of HARNAD, human action recognition is handled as a closed-set problem, where a given action is classified into one of a priori known classes by means of a convolutional neural network architecture. The architecture of the CNN is given in Figure 3. The CNN consists of two convolution layers followed by a maxpooling layer and another convolution layer which is also followed by a maxpooling layer. The CNN is completed by three dense layers.

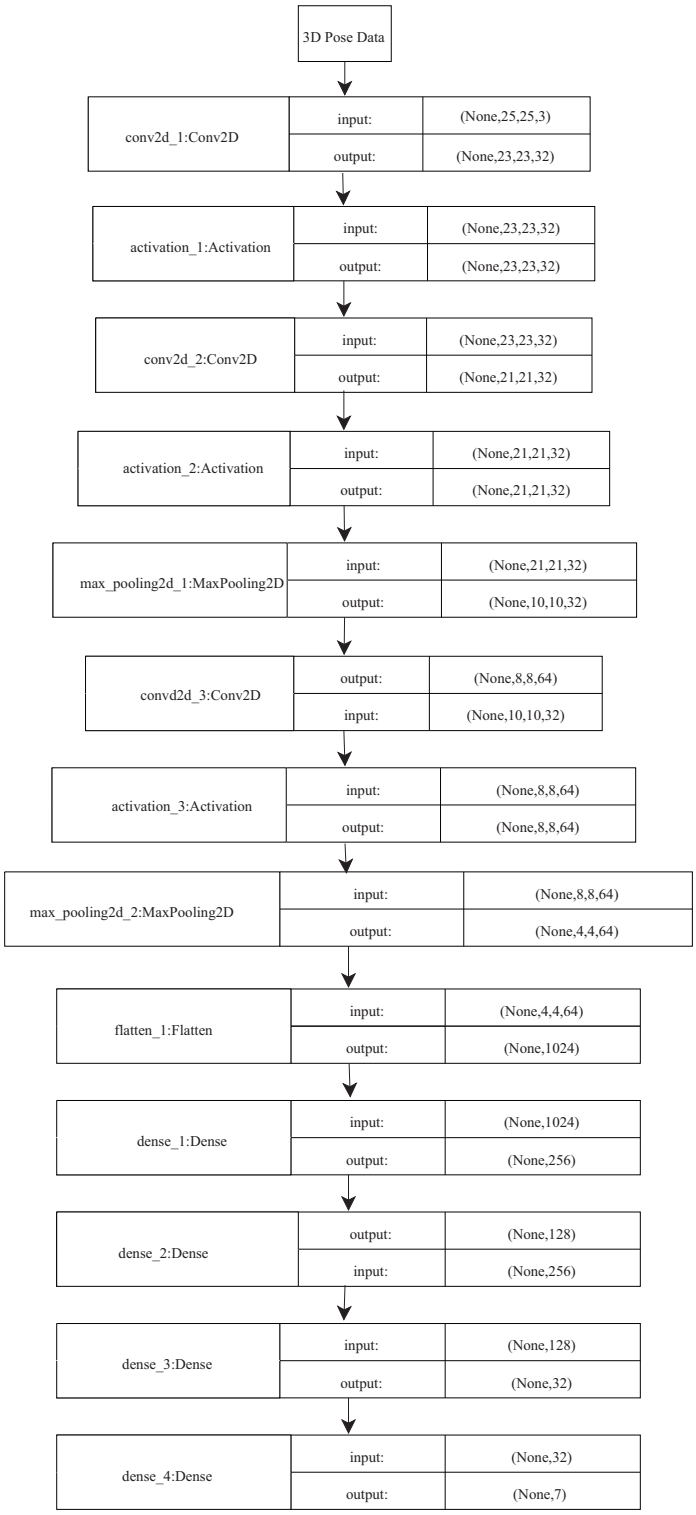

Figure 3. Architecture of CNN at stage 1 of HARNAD.

3.3. Novel action detection at stage 2

3.3.1. Generative adversarial networks (GANs)

GANs consist of two networks. The first network is the generator, G. It has latent variables θ which given noise data Z generates samples $g_\theta(Z)$. The goal of the generator is to find the parameters θ that provides output $g_\theta(Z)$ with a distribution P_θ which is close to X. The second network is the discriminator, D, with parameters w, which given a sample tries to determine if the sample is from real data or from fake data generated by the generator. Ideal discriminator assigns 1 to the sample from real distribution and 0 to the generated samples. The idea behind the GAN training is to get G model data, X, such that it achieves a good generalization performance for the real data, even so good that it can fool D. Meanwhile, during training, D learns how to distinguish between real and fake data. This is formulated as a minmax problem as in Equation 1.

$$\min_{\theta} \max_{w} \mathbf{E}[log(D_w(X)) + log(1 - D_w(g_\theta(Z)))] \tag{1}$$

3.3.2. Novelty detection by GAN

In stage 1, we assigned each sample to an action category. However, there are novel actions that should not be assigned to any of the a priori known categories. In order to detect novel actions, we employ expert detectors, each of which has an expertise in an action category. Thus, for c classes, we train c number of expert detectors. For this purpose, we employ the GAN architecture [27], which consists of a reconstruction network (R) and a discrimination network (D), whose architectures, as employed in this study, are given in Figures 4 and 5, respectively. These architectures are employed by eliminating the last feature, so that 24 body joints are employed at this stage. The GANs trained at this stage are referred as the novelty GANs (Nov-GAN) and they gain expertise in detecting only one class of action. Therefore, we train c number of Nov-GANs, one for each action category. An action labeled as a_i in stage 1 is fed to $Nov - GAN_i$ in stage 2. Nov-GAN first reconstructs the input data by the R network and provides an output for the reconstruction loss, $l_{rec.}$. After that the reconstructed data is fed to the D network to get the output of the discriminator d_i. A novel action has a larger reconstruction loss and smaller discriminator output value. Using the two values, we define the novelty detection rule for an input action at a Nov-GAN as in equation 2.

$$o_i = \begin{cases} c_i & l_{rec} < \tau_1 \wedge d_i > \tau_2 \\ novel\_class & otherwise \end{cases} \tag{2}$$

τ_1 is set as $l_{mean} + (l_{max} - l_{mean})/2$ and τ_2 is set as $d_{min} + (d_{mean} - d_{min})/2$. In these equations, l_{mean} refers to the mean reconstruction loss, while l_{max} is the maximum reconstruction loss over all train inputs. Similarly, d_{mean} is the mean discriminator output and d_{min} is the minimum discriminator output over all train inputs.

4. Experiments

4.1. Dataset

There is a limited number of datasets with 3D data (skeleton) such as Cornell Activity Datasets CAD-120[1], MSR Daily Activity 3D[2], UTKinectAction [29], G3D [30], Human36m [31], and NTU. Most of these datasets

[1]CAD-120. Cornell Activity Dataset [online]. Website http://pr.cs.cornell.edu/web3/CAD-120/ [accessed 29 February 2020]

[2]MSR Daily Activity Dataset [online]. Website https://www.microsoft.com/en-us/download/details.aspx?id=52315 [accessed 29 February 2020]

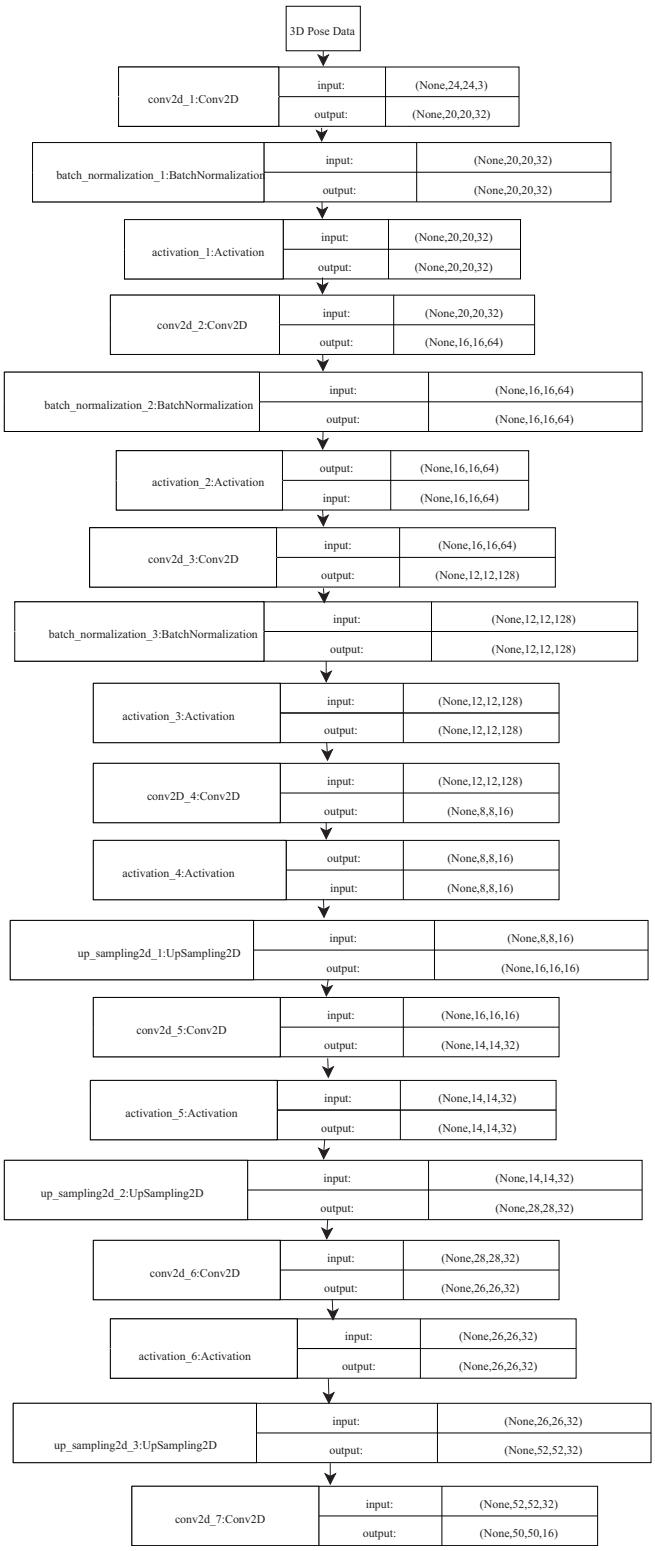

Figure 4. Reconstruction network at stage 2.

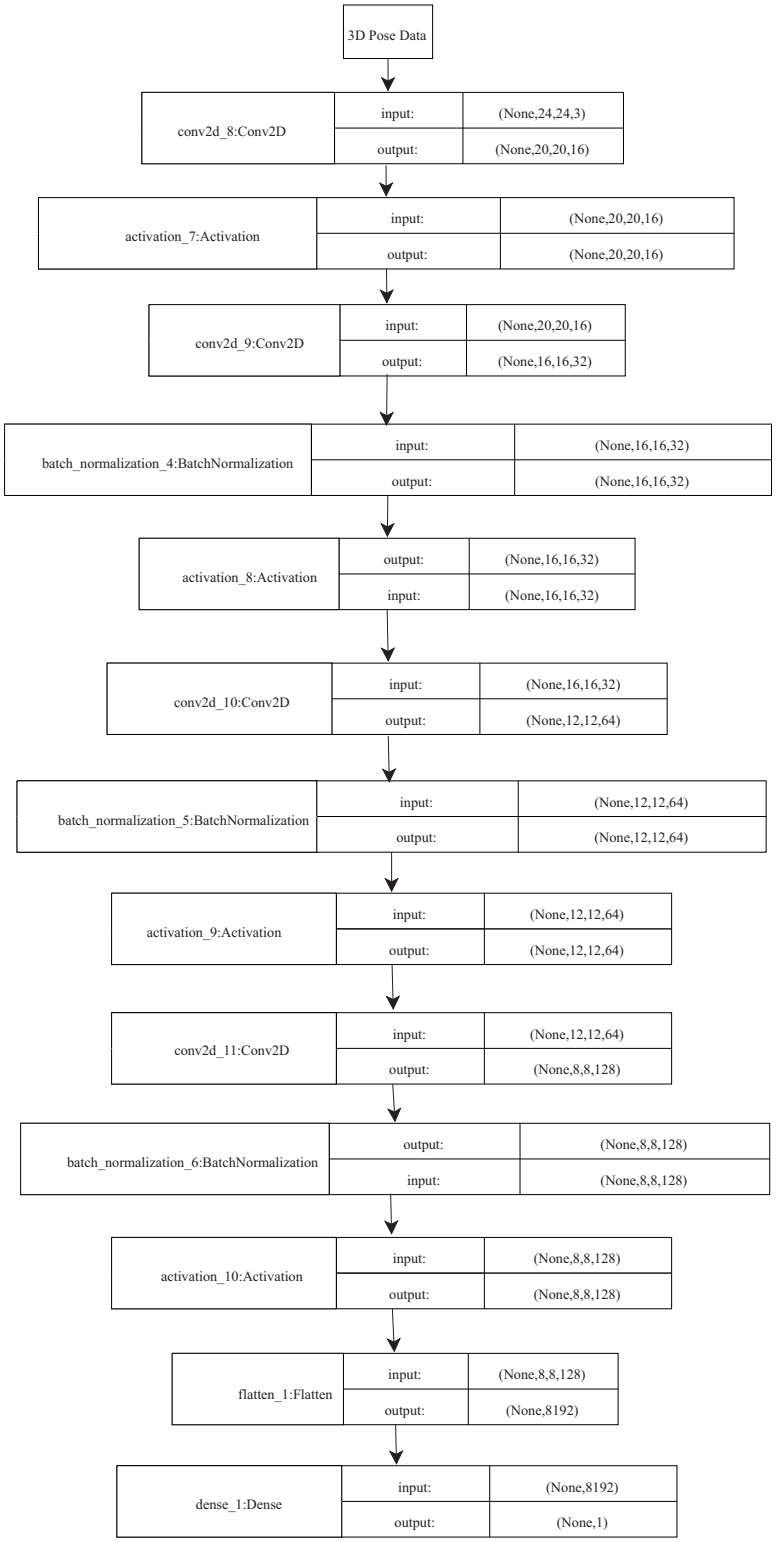

Figure 5. Discrimination network at stage 2.

(Cornell, MSR, UTK, G3D) have a limited number of samples, which makes them unpractical for deep learning techniques. We select a subset of activities from the NTU Daily Actions Dataset [28]. A subset of actions which Wang and Wang [32] report to be effectively classified (accuracy $\geq$ 0.7) while only pose data is selected. The selected actions are pick up, sit down, stand up, wear jacket, take off jacket, wear a shoe, wear glasses, take off glasses, put on hat/cap, take off hat/cap, cheer up, hop, and jump up. Half of the actions, that is 7 actions are selected as seen actions and the rest of the 7 actions are left as unseen actions.

4.2. Experimental setup

In this study, we employ only the skeleton data consisting solely of 3D joints information for representation of human actions. For a given action, the representation is obtained by extracting 3D coordinates of each 25 joints for a group of 25 frames, as shown in the representation part of Figure 1, resulting in a representation of the form $25 \times 25 \times 3$. In this way, the number of training samples for each action varies in the range (1100–1900). Using those features, we construct two experimental setups.

The aim of the first setup is to evaluate the sensitivity of HARNAD to the openness of the problem. For this purpose, we keep the number of training classes fixed, but we increase the number of testing classes by introducing unknown classes. The number of unknown classes is set as {1,3,5,7}. The number of target classes is set as one more than the number of training classes, that class corresponding to the novel class. In this setup, it is assumed that %50 of test data is novel. Hence, the total number of test samples is set as 1400 (100 samples from each known class and 700 samples equally distributed from unknown class(es)).

The aim of the second setup is to evaluate the recognition and novelty detection of HARNAD compared to the state-of-the-art systems. For this purpose, we run HARNAD for ten folds, each time with a different combination of known and novel action grouping. That is, at each run a random set of seven actions are selected as known and the rest of the seven actions are left as novel classes. The number of test samples from each known class is set as 100 and the number of unknown samples is also set as 100. Mean sensitivity and mean specificity values are recorded.

4.3. Evaluation

The openness of an open-set recognition problem is evaluated by equation 3 [17]. For a fixed number of training classes, increasing the number of testing classes and the number of target classes increases the openness.

$$Openness = 1 - \sqrt{\frac{2 \times |TrainingClasses|}{|TestingClasses| + |TargetClasses|}} \tag{3}$$

In evaluation of an open-set problem, the problem can be either considered as a binary classification problem where the attention is on novelty detection or it can be considered as a multiclass classification problem. For a binary classification problem, sensitivity and specificity criteria are mostly used while for the multiclass problem, precision and recall criteria are used for evaluation. Sensitivity and recall are the same and they measure the proportion of actual positives by equation 4, while specificity measures the proportion of actual negatives by equation 5. Sensitivity and specificity are evaluated for each class individually and the average value of all classes is reported as the system performance. Precision measures the proportion of true positives over all positive labeled samples by equation 6. Using the two criteria precision and recall, f-score is evaluated

by equation 7 which enables one to use a single metric for evaluating the classification performance.

$$Sensitivity = \frac{TruePositive}{TruePositive + FalseNegative} \tag{4}$$

$$Specificity = \frac{TrueNegative}{TrueNegative + FalsePositive} \tag{5}$$

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{6}$$

$$Fscore = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{7}$$

4.4. Results

Action recognition accuracy of HARNAD is compared with the state-of-the-art open-set recognition methods. For this purpose, the most widely used open-set recognition methods, isolation forest [33], one-class SVM (OCSVM) [34], and CNN with thresholding of SoftMax, are employed. Initially the CNN in the first stage as given in Figure 3 is employed and image features are extracted from the $'dense'_1$ layer. Initially, those features are provided to the isolation forest. Then, those features are provided to the one-class SVM classifier. And as a third baseline, output of $'dense'_4$ is fed to a SoftMax layer and the most probable classifications with $probability < 0.80$ are labeled as unknown class. In this setup, the sensitivity of HARNAD to the openness parameter is evaluated in terms of the recognition accuracy and it is compared with the state-of-the-art methods. The obtained results are provided graphically in Figure 6. In this figure, the sensitivity of methods as openness increases are provided. Known samples are kept the same at all levels of openness and openness is increased by adding more unknown classes (1, 3, 5, 7, respectively) during testing. It is observed that the performance of HARNAD outperforms the other methods as the openness increases. This difference in performances is mainly caused by novel samples; hence, it can be concluded that HARNAD detects novel samples better than other methods.

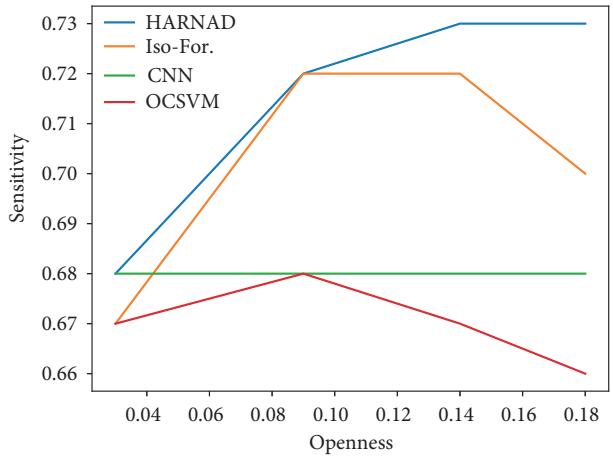

Figure 6. Openness vs. sensitivity.

In the second experimental setup, open-set action recognition performance of HARNAD is compared with The state-of-the-art open-set recognition methods, isolation forests [33], one-class SVM [34], and CNN classification by thresholding the output of SoftMax layer, which are commonly used in the anomaly detection problems. In this setup, the percentage of the number of novel samples to the number of known samples, p, is set as $\{0.1, 0.5, 1\}$. The results are given in Table. In this table, all the performance criteria are given for the three values of p. As p increases, the amount of novel samples increases; for $p = 1$ the number of novel samples and the number of known samples are the same. It is observed that as the number of novel samples increases, all the criteria indicate a decrease in the performance for all systems. HARNAD is more robust to this change and it is able to get the best performance over all methods for $p = 1$.

Table . Comparison of HARNAD with the state-of-the-art methods.

		p	Iso-For. [33]	OCSVM [34]	CNN	HARNAD
Sens.	(Rec.)	0.1	0.68	0.69	0.71	0.74
		0.5	0.70	0.66	0.68	0.73
		1	0.70	0.67	0.69	0.73
Spec.		0.1	0.96	0.96	0.96	0.96
		0.5	0.93	0.93	0.93	0.94
		1	0.92	0.93	0.93	0.94
Prec.		0.74	0.75	0.69	0.69	0.73
		0.5	0.43	0.52	0.48	0.48
		1	0.38	0.36	0.40	0.44
F-score		0.1	0.71	0.72	0.70	0.75
		0.5	0.53	0.58	0.56	0.58
		1	0.49	0.47	0.51	0.55

5. Conclusion

We propose a two-stage architecture for open-set action recognition problem which uses 3D coordinates of joints for action representation. The first stage of the architecture consists of a CNN and the second stage consists of a set of expert GANs each of which has expertise in a certain class. The system first classifies a given sample as one of known classes in the first stage. Then, in the second stage the corresponding expert GAN determines whether the sample really belongs to the assigned class or it is a novel sample. We compare the recognition and novelty detection performance of HARNAD with state-of-the-art methods via two experimental setups. Empirical study reveals that HARNAD is better than the other methods both in novelty detection and recognition and its advantage over the other methods is more clearly observed as the number of novel samples increases.

As a future work, we are planning to integrate a zero-shot learning method to the proposed system so that the system can assign class labels to the detected novel samples as well.

Acknowledgments

This paper used the NTU RGB+D Action Recognition Dataset made available by the ROSE Lab at the Nanyang Technological University, Singapore. We would like to thank Aykut Erdem for his valuable comments throughout this research.

References

[1] Gao Z, Xuan H, Zhang H, Wan S, Choo KR. Adaptive fusion and category-level dictionary learning model for multiview human action recognition. IEEE Internet of Things Journal 2019; 6 (6): 9280-9293. doi: 10.1109/JIOT.2019.2911669

[2] Jegham I, Khalifa AB, Alouani I, Mahjoub MA. Vision-based human action recognition: An overview and real world challenges. Forensic Science International: Digital Investigation 2020; 32: 200901. doi: 10.1016/j.fsidi.2019.200901

[3] Tufek N, Yalcin M, Altintas M, Kalaoglu F, Li Y, Bahadir SK. Human action recognition using deep learning methods on limited sensory data. IEEE Sensors Journal 2020; 20 (6): 3101-3112. doi: 10.1109/JSEN.2019.2956901

[4] Ahmad Z, Khan N. Human action recognition using deep multilevel multimodal (M^2) fusion of depth and inertial sensors. IEEE Sensors Journal 2020; 20 (3): 1445-1455.

[5] Aggarwal JK, Xia L. Human activity recognition from 3D data: A review. Pattern Recognition Letters 2014; 48: 70–80. doi: 10.1016/j.patrec.2014.04.011

[6] Laptev I, Marszalek M, Schmid C, Rozenfeld B. Learning realistic human actions from movies. In: IEEE Conference on Computer Vision and Pattern Recognition; Anchorage, AK, USA; 2008. pp. 1-8.

[7] Ji S, Xu W, Yang M, Yu K. 3D Convolutional neural networks for human action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 2013; 35 (1): 221-231.

[8] Oh S, Hoogs A, Perera A, Cuntoor N, Chen CC et al. A large-scale benchmark dataset for event recognition in surveillance video. In: Computer Vision and Pattern Recognition (CVPR 2011); Providence, RI, USA; 2011. pp. 3153-3160. doi: 10.1109/CVPR.2011.5995586

[9] Popoola OP, Wang K. Video-based abnormal human behavior recognition—A review. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 2012; 42 (6): 865-878. doi: 10.1109/TSMCC.2011.2178594

[10] Wan S, Gu Z, Ni Q. Cognitive computing and wireless communications on the edge for healthcare service robots. Computer Communications 2020; 149: 99-106. doi: 10.1016/j.comcom.2019.10.012

[11] Chen C, Jafari R, Kehtarnavaz N. Improving human action recognition using fusion of depth camera and inertial sensors. IEEE Transactions on Human-Machine Systems 2015; 45 (1): 51-61. doi: 10.1109/THMS.2014.2362520

[12] Shi L, Zhang Y, Cheng J, Lu H. Two-stream adaptive graph convolutional networks for skeleton-based action recognition. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019); Long Beach, CA, USA; 2019. pp. 12018-12027. doi: 10.1109/CVPR.2019.01230

[13] Yang Y, Hou C, Lang Y, Guan D, Huang D et al. Open-set human activity recognition based on micro-doppler signatures. Pattern Recognition 2018; 85: 60-69. doi: 10.1016/j.patcog.2018.07.030

[14] Ji S, Xu W, Yang, M, Yu K. 3D convolutional neural networks for human action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 2013; 35(1): 221-231. doi: 10.1109/TPAMI.2012.59

[15] Wang L, Xu Y, Cheng J, Xia H, Yin J et al. Human action recognition by learning spatio-temporal features with deep neural networks. IEEE Access 2018; 6: 17913-17922. doi: 10.1109/ACCESS.2018.2817253

[16] Luvizon D, Tabia H, Picard D. Multimodal deep neural networks for pose estimation and action recognition. In: Congres Reconnaissance des Formes, Image, Apprentissage et Perception (RFIAP 2018); Marne-la-Vallee, France; 2018. pp.1-20

[17] Scheirer WJ, Rocha AR, Sapkota A, Boult TE. Toward open set recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 2013; 35 (7): 1757-1772. doi: 10.1109/TPAMI.2012.256

[18] Scheirer WJ, Jain LP, Boult TE. Probability models for open set recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 2014; 36: 2317-2324. doi: 10.1109/TPAMI.2014.2321392

[19] Bendale A, Boult TE. Towards open set deep networks. In: IEEE Conference on Computer Vision and Pattern Recognition; Las Vegas, NV, USA; 2016. pp. 1563-1572.

[20] Shu Y, Shi Y, Wang Y, Zou Y, Yuan Q et al. ODN: Opening the deep network for open-set action recognition. In: IEEE International Conference on Multimedia and Expo (ICME); San Diego, CA, USA; 2018. pp.1-6.

[21] Roitberg A, Al-Halah Z, Stiefelhagen R. Informed democracy: Voting-based novelty detection for action recognition. In: British Machine Vision Conference (BMVC); Newcastle, UK; 2018.

[22] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D et al. Generative adversarial nets. In: Advances in Neural Information Processing Systems 27 (NIPS 2014); Montreal, Canada; 2014. pp. 2672-2680.

[23] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. 2015; arXiv.

[24] Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A et al. Improved techniques for training GANs. In: 30th International Conference on Neural Information Processing Systems; Barcelona, Spain; 2016. pp. 2234-2242.

[25] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. In: 34th International Conference on Machine Learning; Sydney, Australia; 2017. PMLR 70: 214-223.

[26] Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville AC. Improved training of Wasserstein GANs. In: International Conference on Neural Information Processing Systems; Long Beach, CA, USA; 2017. pp.5767-5777.

[27] Saboktrou M, Khalooei M, Fathy M, Adeli E. Adversarially learned one-class classifier for novelty detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition; New York, USA; 2018. pp. 3379-3388.

[28] Shahroudy A, Liu J, Ng T, Wang G. NTU RGB+D: A large scale dataset for 3D human activity analysis. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Las Vegas, US; 2016. pp. 1010-1019.

[29] Xia L, Chen CC, Aggarwal JK. View invariant human action recognition using histograms of 3D joints. In: Computer Vision and Pattern Recognition Workshops (CVPRW); Providence, RI, USA; 2012. pp. 20-27.

[30] Bloom V, Makris D, Argyriou V. G3D: A gaming action dataset and real time action recognition evaluation framework. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops); Providence, RI, USA; 2012. pp. 7-12.

[31] Ionescu C, Papava D, Olaru V, Sminchisescu C. Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments. IEEE Transactions on Pattern Analysis and Machine Intelligence 2014; 36 (7): 1325-1339.

[32] Wang H, Wang L. Beyond joints: Learning representations from primitive geometries for skeleton-based action recognition and detection. IEEE Transactions on Image Processing 2018; 27 (9): 4382-4394.

[33] Liu FT, Ting KM, Zhou ZH. Isolation forest. In: The Eighth IEEE International Conference on Data Mining; Pisa, Italy; 2008. pp.413-422.

[34] Schölkopf B, Williamson R, Smola A, Shawe-Taylor J, Platt J. Support vector method for novelty detection. In: 12th International Conference on Neural Information Processing Systems; Perth, WA, Australia; 1999. pp.582-588.