Generalized Adversarial and Hierarchical Co-occurrence Network based Synthetic Skeleton Generation and Human Identity Recognition

Joseph G. Zalameda
Dept. of ECE
Old Dominion University
Norfolk, VA, USA
jzala001@odu.edu

Brady Kruse
Dept. of Computer Science
Mississippi State University
Starkville, MS, USA
bak225@msstate.edu

Alexander M. Glandon
Dept. of ECE
Old Dominion University
Norfolk, VA, USA
aglan001@odu.edu

Megan A. Witherow
Dept. of ECE
Old Dominion University
Norfolk, VA, USA
mwith010@odu.edu

Sachin Shetty
Dept. of CMSE
Old Dominion University
Norfolk, VA, USA
sshetty@odu.edu

Khan M. Iftekharuddin
Dept. of ECE
Old Dominion University
Norfolk, VA, USA
kiftekha@odu.edu

2022 International Joint Conference on Neural Networks (IJCNN) | 978-1-7281-8671-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/IJCNN55064.2022.9892887

***Abstract*—Human skeleton data provides a compact, low noise representation of relative joint locations that may be used in human identity and activity recognition. Hierarchical Co-occurrence Network (HCN) has been used for human activity recognition because of its ability to consider correlation between joints in convolutional operations in the network. HCN shows good identification accuracy but requires a large number of samples to train. Acquisition of this large-scale data can be time consuming and expensive, motivating synthetic skeleton data generation for data augmentation in HCN. We propose a novel method that integrates an Auxiliary Classifier Generative Adversarial Network (AC-GAN) and HCN hybrid framework for Assessment and Augmented Identity Recognition for Skeletons (AAIRS). The proposed AAIRS method performs generation and evaluation of synthetic 3-dimensional motion capture skeleton videos followed by human identity recognition. Synthetic skeleton data produced by the generator component of the AC-GAN is evaluated using an Inception Score-inspired realism metric computed from the HCN classifier outputs. We study the effect of increasing the percentage of synthetic samples in the training set on HCN performance. Before synthetic data augmentation, we achieve 74.49% HCN performance in 10-fold cross validation for 9-class human identification. With a synthetic-real mixture of 50%-50%, we achieve 78.22% mean accuracy, significantly ($p<0.05$) outperforming the baseline HCN performance. The proposed framework demonstrates the feasibility of combining a synthetic data generation architecture with hierarchical co-occurrence feature learning for human identity recognition.**

Keywords—Security, Human Identity Recognition, Skeleton Data, Motion Capture Data, Generative Adversarial Network, Auxiliary Classification, Synthetic Skeleton Data, Co-occurrence

I. Introduction

Modern sensing technologies enable capture of efficient, high-quality representations of human structure and dynamic human motion. Among these representations, articulated human pose, also known as skeleton, provides a rich source of biometric and behavioral information. Sequences of skeleton frames provide an inherently compact and robust-to-noise representation of human pose over time. Recent advances in deep learning have facilitated apposite processing of skeleton sequence data for tasks including fall detection [1], human action recognition [2], and human identity recognition [3]. In this work, we focus on the task of human identity recognition. Skeleton-based identity recognition offers quick, unintrusive biometric authentication for security applications.

While originally designed for image data, convolutional neural networks (CNNs) have also demonstrated outstanding ability to extract high-level features from skeleton data [4]. To use skeleton sequence data with CNNs, the skeleton is presented as an image-like representation with columns for each joint, rows for each video frame, and a channel for each 3D spatial coordinate, i.e., x, y, z. CNNs extract features from local receptive fields and perform global feature aggregation over the channel dimension. Because the joints are typically given as channels when presented to the CNN, standard CNN processing of these image-like representations of skeletons miss important inter-joint correlation. To capture co-occurrence, features that exist between skeleton joints, Li et al. [4] introduce the Hierarchical Co-occurrence Network (HCN) framework for CNN-based action recognition from skeleton data. The HCN framework rearranges the feature map dimensions within the intermediate layers of the network to extract features based on global aggregation over the different joints. Features related to the motion of the joints across frames are also learned explicitly using a parallel network branch that operates on a calculated derivative of the skeleton video over time. Given that HCN captures joint co-occurrence and dynamics of motion features, we adapt the HCN for our task of human identity recognition from skeleton videos.

The stringent reliability and validity requirements of security applications and the tendency of deep learning models to overfit on small data both motivate the need for a large and diverse training set for deep learning-based human identity recognition. However, depending on the acquisition

technology, constructing such datasets may be time consuming and expensive. Real datasets are also prone to missing data, including missing frames and unknown or outlier joints within a frame. This missing data and noise may corrupt data samples, further reducing dataset size or reliability. To overcome these challenges, synthetic skeleton data may be generated to augment smaller training sets. Generative adversarial networks (GANs) [5] have been proposed for synthetic skeleton action generation [6-8]. GANs consist of two networks in competition. A generator network synthesizes samples from noise and a discriminator network aims to distinguish synthesized examples from real examples. Multiple variants of GAN have been proposed, including conditional GAN (cGAN) [9] which enables the generation of class-conditioned samples, and Auxiliary Classifier GAN (AC-GAN) [10], which improves upon cGAN by adding an auxiliary output to the discriminator to classify synthetic examples.

The most widely used quantitative evaluation metrics of synthetic data are built upon the Inception V3 network pretrained on the ImageNet dataset. The Inception V3 network [11] is a deep CNN model trained for classifying over 1000 categories of natural images in the ImageNet [12] database. Metrics such as the Inception Score [13] and Fréchet Inception Distance [14] use the output of an intermediate or final layer of Inception V3 to compute scores quantifying the realism of the generated data. However, it has been noted that using the Inception V3 model for evaluating generative models trained on datasets outside of ImageNet yields misleading results [15, 16]. To overcome this limitation, we propose a within-dataset metric for synthetic skeleton data evaluation inspired by the Inception Score.

To address the need for synthetic skeleton data to improve the robustness of human identification of a given dataset with only a few hundred samples per identity class, we propose an AC-GAN and HCN hybrid framework: the Assessment and Augmented Identity Recognition for Skeletons (AAIRS) framework. This framework generates synthetic skeleton videos and performs realism assessment of these samples. The framework uses the synthetic samples as feedback to the classifier to fine-tune the human identity recognition model. Our AAIRS framework incorporates an HCN model trained for human identity recognition (HCN-ID) and AC-GAN model adapted for identity-conditional synthetic skeleton generation (AAIRS-GAN). To our knowledge, our approach is the first to bring generative adversarial and hierarchical co-occurrence learning together in a single framework. We train and evaluate our framework using a challenging ground truth skeleton dataset extracted from lidar motion capture data [17]. The dataset contains 9 subjects performing 2 trials of a simple walking movement in sequences consisting of a few hundred frames. We consider two experimental pipelines: 1) HCN-ID based realism assessment of AAIRS-GAN generated samples, and 2) training and evaluation of HCN-ID under synthetic data augmentation. In our first pipeline, we use the trained HCN-ID model to compute a within-dataset realism assessment score on synthetic skeleton samples generated by the AAIRS-GAN model. In our second pipeline, we retrain HCN-ID on a mixture of real and synthetic data samples and evaluate the effect of different training set compositions on performance. Our contributions are as follows: A) adapt, train and evaluate HCN for identity recognition from skeleton gait data, B) adapt and train AC-GAN to generate synthetic skeleton data for identity recognition, C) introduce and compute an appropriate within-dataset realism metric for synthetic skeleton data, and D) augment HCN training set with synthetic samples and investigate the effect of the percentage of synthetic samples on identity classification performance.

The remainder of this paper is structured as follows. Section II discusses relevant background on deep learning for skeleton data, the HCN framework, generation of synthetic data using AC-GAN, and Inception Score for evaluation of synthetic data. Section III presents our AAIRS framework. Models and procedures for human identity classification, synthetic skeleton generation, and realism assessment of generated data are detailed. Section IV describes the two experimental pipelines and discusses the results. Section V concludes and suggests directions for future work.

II. Background

A. Deep Learning for Skeleton Data

Most of the prior work on deep learning for skeleton data utilize Recurrent Neural Networks (RNNs) with Long-Short Term Memory (LSTM) for modeling temporal motion. Zhu et al. propose a fully connected LSTM network for learning co-occurrence features [4]. Song et al. introduce hierarchical co-occurrence with spatial and temporal attention modules for selecting dominant joints and assigning importance weights to individual frames, respectively [18]. However, LSTM networks model temporal structure over the input space in isolation without accounting for higher order features that may explain some of the underlying variation in the data [19].

Du et al. [20] and Ke et al. [21] use CNNs to perform skeleton-based action recognition. Both studies perform transformations on the raw skeleton data to obtain an image-like representation prior to feeding the data into the CNN. Du et al. [20] encode temporal movement and skeleton joints as rows and columns, respectively, and x, y, z dimensions are considered as the channels. Ke et al. [21] propose a similar representation but perform a coordinate transformation and rescaling to represent x, y, and z channels as gray-scale images. The modeling of joint co-occurrence features in these methods is limited by the receptive field of the CNN, i.e., joint co-occurrence information is learned for neighboring joints only. The HCN model proposed by Li et al. [4] addresses this limitation by modeling global co-occurrence and achieves state-of-the-art results on action recognition benchmark data.

B. HCN Framework

The HCN model with weights θ_H takes a video with one or more skeleton views $x_1, \ldots, x_N$ where each $x_n \in X \in$

$\mathbb{R}^{J,C,T}$ for J joints, $C = 3$ axes for the 3D coordinates of a joint, and T frames per sample. HCN takes these skeletons as input and produces a vector of estimated probabilities $\hat{y}$ for each class as in equation (1).

$$\hat{y} = [p_{y1}, \dots, p_{y10}] = H(x_1, \dots, x_N | \theta_H) \quad (1)$$

Skeleton motion is modeled explicitly as follows. Given each input $x \in \mathbb{R}^{J,C,T}$ to the HCN network H, the network calculates the derivative w.r.t. time $\dot{x} \in \mathbb{R}^{J,C,T}$ using the differencing formula in eq. (2).

$$\dot{x} = \frac{\partial x}{\partial t} \approx x(:,:,2{:}T) - x(:,:,1{:}T-1) \quad (2)$$

The HCN contains 2 parallel branches that are optimized separately. Branch 1 takes $x_1, \dots, x_N$ and branch 2 takes $\dot{x}_1, \dots, \dot{x}_N$, respectively. In both branches, the HCN performs hierarchical aggregation of feature information. First, a point-level aggregation over the x, y, and z channels is performed using 1×1 and $1 \times J$ convolutions. Then, aggregation is performed globally across all joints. A permutation of the dimensions of x and $\dot{x}$ occurs in each branch to orient the joint dimensions of each sample to the channel dimension of the convolutional filter inputs. The concatenated output of the parallel branches undergoes further convolutional and fully connected operations prior to classification.

C. AC-GAN Framework

AC-GAN is an image synthesis framework from the GAN family of generative models. Like cGAN, AC-GAN is a conditioned model that incorporates label selection as a part of the synthetic data generation. In addition to the noise input, the AC-GAN generator also takes a class label as a conditioned input. The AC-GAN discriminator performs classification of images as real or generated. In addition to the discriminator, an auxiliary classifier is used to classify generated images into image classes. In practice, the discrimination and auxiliary classification tasks are performed by a single discriminator network optimized for both tasks. The goal of AC-GAN is to generate synthetic data that reflects the mix of classes present in the real data. Let eq. (3) represent the distribution of synthetic data X for given label Y.

$$X_{fake} | Y \sim P(X = x | Y = y) \quad (3)$$

The AC-GAN generator with weights θ_G can be described as in eq. (4) where given input class label $y_i \in \{y_1, \dots y_I\}$ with I possible classes and noise vector $\vec{\epsilon} \sim MVN(\vec{\epsilon} \mid \mu = \vec{0},\ \Sigma = \sigma^2 \mathrm{I})$.

$$x_{fake} = G(y_i, \vec{\epsilon} \mid \theta_G) \quad (4)$$

The AC-GAN discriminator with weights θ_D can be described as in eq. (5) where input x is drawn either from real data x_{real} or synthetic generated data x_{fake}. Similar to the HCN output in eq. (1), the discriminator outputs a vector $\hat{y}$ for the estimate of the probability for each class y_i, as well as an estimate $\hat{r}$ of the probability that the source of the sample is real. Note that the estimate $\hat{f}$ of the probability that the source of the sample is fake is implied as $\hat{f} = 1 - \hat{r}$.

$$\hat{y}, \hat{r} = D(x | \theta_D)) \quad (5)$$

The AC-GAN framework uses two separate loss functions. L_S is the log-likelihood of correct source, while L_C is the log-likelihood of correct class. The discriminator attempts to maximize $L_S + L_C$ while the generator is trained to maximize $L_S - L_C$. L_S and L_C can are defined in eq. (6) and eq. (7) respectively. Fig. 1 shows the network architecture for both the generator and discriminator used by the AC-GAN based framework in this paper.

$$L_S = E[\log(\hat{r}) \mid x \in X_{real}] + E[\log(1 - \hat{r}) \mid x \in X_{fake}] \quad (6)$$

$$L_C = E[\log(\hat{y}_i) \mid Y = y_i] \quad (7)$$

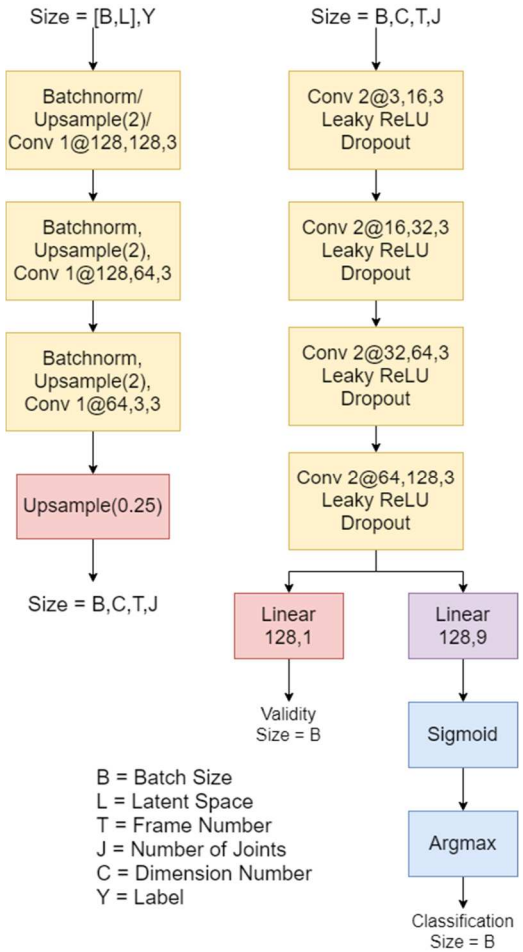


Fig. 1. AC-GAN generator (left) and discriminator (right) architectures.

D. Inception Score

The Inception Score is the most widely adopted metric for quantitative evaluation of the realism of synthetic data. While the Inception Score is computed using the ImageNet pretrained Inception V3 network, the mathematical framework for computing the score is general enough such that any trained classification model may be substituted for Inception V3. In general, a classification network trained on a specific problem domain may be used to evaluate the realism of GAN-generated examples in the same domain. Given a set of synthetic data X_{fake}, generated by a GAN, we evaluate realism using eq. (10). The idea is to compare the conditional distribution given by the classifier, e.g., $\hat{y} = [p_{y1}, \dots, p_{y10}]$ from the classifier H in eq. (1), to the marginal distribution of the labels as given in eq. (8). The marginal distribution is estimated using the expected conditional distribution $\hat{y} = H(x|\theta_H)$ as in eq. (9), where $P(X)$ is simplified assuming a uniform distribution. We use KL divergence as a metric of the similarity of these distributions and then take the expected value by averaging over a large set of N examples of synthetic data as in eq. (10). We take 2 to the power of the final expectation in eq. (10) to convert from units of entropy to perplexity.

$$Y \sim P(Y) = \int_X P(X,Y)dx \tag{8}$$

$$\begin{aligned} P(Y) &= \int_X P(X=x,Y)dx = \\ &\int_X P(Y|X=x)P(X=x)dx \approx \\ \hat{P}(Y) &= \int_X H(x|\theta_H)P(X)dx = \\ E[H(x|\theta_H)] &= \sum_x H(x|\theta_H)/N \end{aligned} \tag{9}$$

$$\begin{aligned} &Inception\ Score = \\ &E\left[KL\left(H(x|\theta_H) \,||\, \hat{P}(Y)\right)\right] = \\ &\sum_x KL(H(x|\theta_H) \,||\, \hat{P}(Y))/N \end{aligned} \tag{10}$$

III. Methods

A. HCN for Human Identity Classification (HCN-ID)

The HCN-ID model follows the same structure as the HCN model architecture in [4]. The number of output units in the classification layer is chosen to reflect the number of identity classes.

Consider the data distribution of real skeleton data in eq. (11). Each sample skeleton sequence is composed of dimensions J joints, $C = 3$ axes for the 3D coordinates of a joint, and T frames per sample such that $X \in \mathbb{R}^{J,C,T}$. Each sample has a corresponding subject identity, $Y \in \{y_1, \dots, y_I\}$ for I subject identity classes.

$$X, Y \sim P_{real}(X = x, Y = y) \tag{11}$$

Given a single subject input, the HCN-ID model estimates the marginal distribution of the real data in eq. (12). The HCN-ID model takes a single input sample sequence x and estimates a vector of probabilities for each of the possible subjects as in eq. (13).

$$Y|X \sim P(Y = y|X = x) \tag{12}$$

$$\hat{y} = [p_{y1}, \dots, p_{y10}] = H(x|\theta_H) \tag{13}$$

B. AAIRS-GAN Synthetic Skeleton Data Generation and Realism Assessment via HCN-ID Score

The overall pipeline for synthetic skeleton generation and realism assessment is summarized in Fig. 2. Our AAIRS-GAN model for synthetic skeleton data generation follows the AC-GAN training framework in [10]. The generator accepts two inputs, noise vector $\vec{\epsilon} \sim MVN(\vec{\epsilon} \mid \mu = \vec{0},\ \Sigma = \sigma^2 \mathrm{I})$ and input class label $y_i \in \{y_1, \dots y_I\}$, to yield $x_{fake} \in \mathbb{R}^{J,C,T}$. As in eq. (5), the discriminator performs discrimination and auxilliary classification tasks. In the discrimination task, the estimate $\hat{r}$ is the probability that discriminator input x corresponds to $x_{real} \in \mathbb{R}^{J,C,T}$. In the auxiliary classification task, $\hat{y}$ corresponds to the predicted class label.

For realism assessment, each generated skeleton sample is input into the pretrained HCN-ID model to output a vector of predicted probabilities as in eq. (13). Then, the HCN-ID outputs for a set of N generated samples are used with eq. (10) to compute the HCN-ID score, a measure of within-dataset synthetic skeleton data realism. The HCN-ID score is an approximation of the AAIRS-GAN generator realism with larger N yielding a better approximation.

C. Synthetic Data Augmentation of HCN-ID

The pipeline for synthetic data augmentation of the HCN-ID model is shown in Fig. 3. A pool of synthetic samples is obtained from the AAIRS-GAN generator. A mixture of real and synthetic samples is formed such that M% synthetic samples are distributed evenly across all classes. HCN-ID is then trained on this synthetic data augmented training set. A 10-fold cross validation scheme is used to assess the performance of the HCN-ID with synthetic data augmentation. Model performance is assessed for different compositions of real and synthetic data in the training set. Steps of 10% are considered as values of M with a minimum synthetic-real mixture of 10%-90% and maximum synthetic-real mixture of 50%-50%

IV. Results and Discussion

A. Experimental Pipelines

To evaluate the proposed AAIRS framework, we consider two experimental pipelines. The first pipeline, shown in Fig. 2, assesses the realism of the synthetic samples generated by the AAIRS-GAN using the HCN-ID score. The HCN-ID score is used to quantify the realism of the synthetic samples being produced by the AAIRS-GAN.

The second pipeline is shown in Fig. 3 and utilizes a mixture of real and synthetic samples for synthetic data

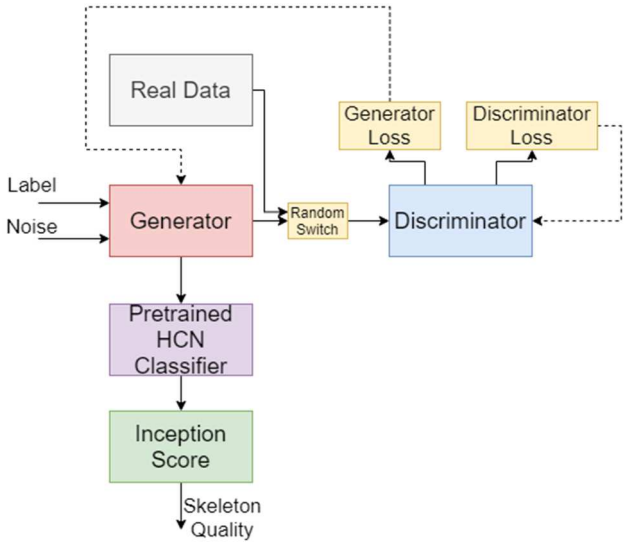


Fig. 2. Skeleton Generation and Realism Assessment Pipeline

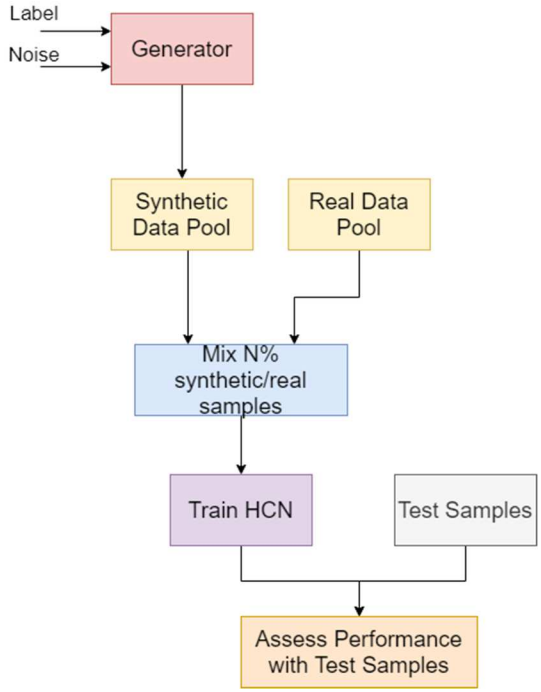


Fig. 3. Pipeline for Synthetic Data Augmentation of HCN-ID

augmented training of the HCN-ID model. This second pipeline represents the AAIRS hybrid training framework incorporating real and synthetic samples to increase sample size with the goal of improving human identity classification.

B. Dataset

For training and evaluation of the proposed AAIRS framework, we consider the challenging ground truth skeleton dataset extracted from lidar motion capture data in [17]. This dataset contains a total of 176 skeleton samples from $I = 9$ subjects following the distribution shown in Fig 4. Each sample consists of $J = 13$ joints animated over a continuous $T = 3$ frame segment from multiple video sequences of the subjects walking in a fixed position, centered at the origin of the 3-dimensional space. The data are normalized between -1 and 1, maintaining the relative lengths between joints over all subjects in the dataset. The samples are distributed into training and testing sets following a 10-fold cross validation scheme.

C. HCN-ID Results

We train the HCN-ID model using the ADAM optimization algorithm with a learning rate of 0.001 for 50 epochs and a batch size B = 64. The 10-fold cross validation results for 9-class classification of human identity for the HCN-ID and Zero-Rule [22] baseline model are shown in Table I. The mean test accuracy over the 10 cross validation test folds is 74.49%, outperforming the baseline model by 56.71% with similar standard deviation. Fig. 5 (a) and Fig. 5 (b) show representative plots displaying the convergence of accuracy and loss, respectively, during training.

D. AAIRS-GAN Results

The AAIRS-GAN framework is trained using the ADAM optimization algorithm and a batch size B = 64. Visual inspection of the synthetic skeleton videos generated by the AAIRS-GAN demonstrates that the generator has learned to produce skeletons with realistic anatomy i.e. head placement above the shoulders. Fig. 6 shows a series of representative sequence of frames produced by the AAIRS-GAN. For

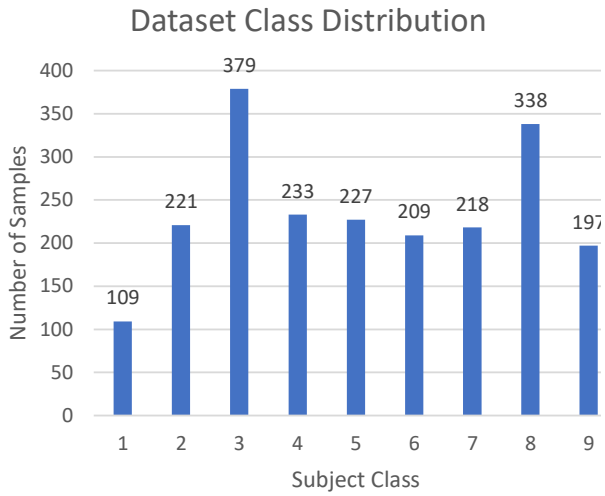
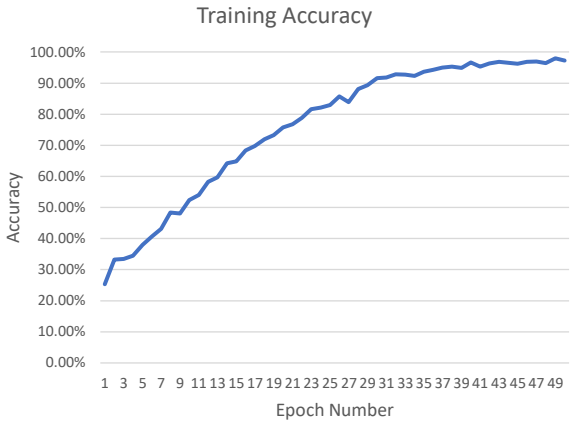


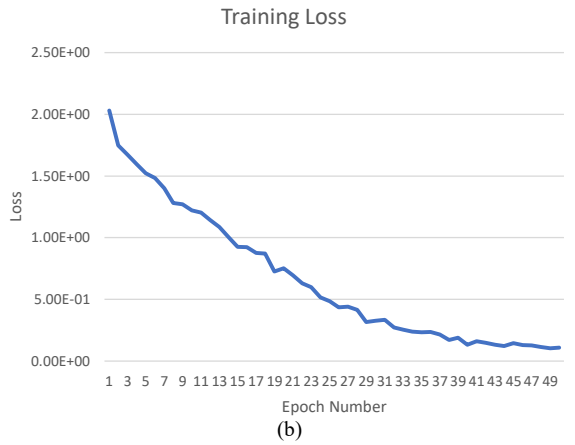
Fig. 4. Dataset Class Distribution

TABLE I. COMPARISON OF HCN-ID AND ZERO RULE BASELINE MODEL 10-FOLD CROSS VALIDATION MEAN ACCURACY

Dataset	Mean Accuracy
Zero Rule (Baseline)	17.78% ± 2.49%
HCN-ID	74.49% ± 2.64%



(a)



(b)

Fig. 5. Representative training (a) accuracy and (b) loss plots during HCN-ID training

comparison, a representative real skeleton sequence is shown in Fig. 7. The synthetic skeletons in general exhibit correct relative anatomical joint locations with legs, arms, and torso positioned correctly.

E. HCN-ID Score for Synthetic Skeleton Realism Assessment

HCN-ID scores are computed on the synthetic data generated based upon the training set for each 10-fold cross validation split. The HCN-ID model is trained on the same training set as the AAIRS-GAN model. The HCN-ID score ranges from 1 (worst score) to 9 (best score) in theory based on the underlying perplexity for 9 classes. We compute the HCN-ID score using 100 generated samples per class or $N = 900$.

Table II shows the average HCN-ID scores of the real and synthetic data. The real data scores an average of ~8 out of 9, and the synthetic data scores an average of ~4. The synthetic data score is greater than the minimum score of 1, demonstrating that the AAIRS-GAN has learned a subset of features underlying the variation of the subject data. Further improvements will seek to increase the HCN-ID score of AAIRS-GAN generated data to come closer to matching the realism of the ground truth data.

It is important to note that the performance of the HCN-ID model will affect the computation of the HCN-ID score.

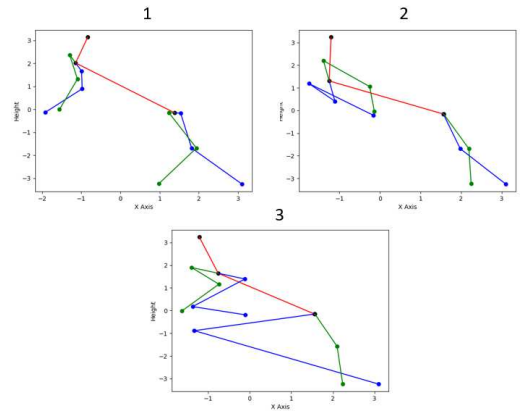


Fig. 6. Representative Synthetic Skeleton Frame Sequence Example

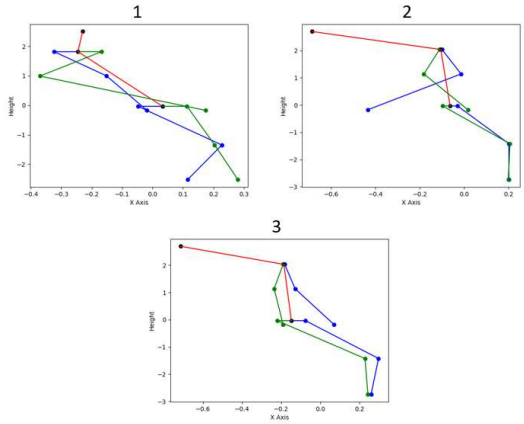


Fig. 7. Representative Real Skeleton Frame Sequence Example

Improving HCN-ID performance will yield a better measurement of synthetic data realism.

F. AAIRS Hybrid Training Framework

The AAIRS hybrid training framework is trained and tested using the same previously mentioned 10-fold cross validation training and test sets. The synthetic data is generated using the AAIRS-GAN trained on the same ground truth as the HCN-ID. This prevents test set information from leaking into the HCN-ID training set through the AAIRS-GAN generator. The synthetic data is then mixed in with the real training data in intervals of 10% ranging from 0% (baseline) to 50% synthetic data. We train the synthetic data augmented HCN-ID models using the ADAM optimizer with a learning rate of 0.001 for 50 epochs.

The AAIRS Hybrid Training results are shown in Table III. The results suggest that augmenting the HCN-ID with AAIRS-GAN generated synthetic data substantially improves performance. The best performance of 78.22% is achieved by 50%-50% synthetic-real mixture. This represents a significant (paired t-test with t = 2.2767, p = 0.0353) improvement over the 74.49% accuracy achieved by the 0%-100% synthetic-real mixture for 10-fold cross validation.

TABLE II. 10-FOLD CROSS VALIDATION MEAN HCN-ID SCORES

Dataset	Mean HCN-ID Score
Real Data	8.352 ± 0.1049
Synthetic Data	3.933 ± 0.8673

TABLE III. AAIRS HYBRID 10-FOLD CROSS VALIDATION MEAN ACCURACIES FOR DIFFERENT PERCENTAGES OF SYNTHETIC SAMPLES IN THE TRAINING SET

% Synthetic Samples in Training Set	Mean Accuracy
0%	74.49% ± 2.64%
10%	75.93% ± 1.61%
20%	76.12% ± 2.37%
30%	75.70% ± 1.89%
40%	77.15% ± 2.32%
50%	78.22% ± 2.37%

V. CONCLUSION AND FUTURE WORK

This work presents the AAIRS framework for hierarchical co-occurrence human identity classification and generation of skeleton data. The proposed framework consists of a co-occurrence classification network (HCN-ID) and a synthetic data generation framework (AAIRS-GAN). We demonstrate and evaluate the AAIRS framework in two pipelines for assessing the realism of generated synthetic data and investigating the effect of synthetic data augmentation on HCN-ID classification performance. The first pipeline simultaneously generates labeled synthetic skeleton data with the AAIRS-GAN and evaluates this data with the HCN-ID score. The second pipeline trains the HCN-ID on a mixture of real and synthetic data by providing a feedback structure from the set of synthetic skeleton data to the HCN-ID training framework.

The AAIRS framework demonstrates hierarchical co-occurrence learning and generation of human identity features from skeleton data with higher than baseline HCN-ID score. Additionally, the AAIRS framework demonstrates significantly higher 10-fold cross validation mean accuracy with synthetic data augmentation when compared to the HCN-ID without data augmentation. The preliminary results reported show the feasibility of our baseline and data-augmented training of the HCN-ID model as well as generation and evaluation of AAIRS-GAN samples. Future work will consider transfer learning from larger benchmark skeleton datasets to improve the performance of the HCN-ID and AAIRS-GAN. Leveraging information learned from a large, diverse, and class-balanced dataset may improve the overall performance of the AAIRS framework. As an extension of the HCN-ID score, we plan to implement an appropriate Fréchet Inception Distance inspired metric that takes the distribution of real images into account when evaluating the synthetic samples. In the future, we plan to compare our framework to other methods for human identification using skeleton data.

ACKNOWLEDGEMENTS

The authors would like to acknowledge partial support of this work by US Army NVESD, CERDEC through Grant No. 100659, DoD Center of Excellence in AI and Machine Learning (CoE-AIML) under Contract Number W911NF-20-2-0277 with the U.S. Army Research Laboratory, and by the National Science Foundation under Grant No. 1828593, Grant No. 1950704, and Grant No. 1753793.

REFERENCES

[1] W. Chen, Z. Jiang, H. Guo, and X. Ni, "Fall detection based on key points of human-skeleton using openpose," *Symmetry,* vol. 12, no. 5, p. 744, 2020.

[2] H.-B. Zhang *et al.*, "A comprehensive survey of vision-based human action recognition methods," *Sensors,* vol. 19, no. 5, p. 1005, 2019.

[3] B. Bhowmick, "Person identification using skeleton information from kinect," 2013.

[4] C. Li, Q. Zhong, D. Xie, and S. Pu, "Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation," *arXiv preprint arXiv:1804.06055,* 2018.

[5] I. Goodfellow *et al.*, "Generative adversarial nets," *Advances in neural information processing systems,* vol. 27, 2014.

[6] W. Xi, G. Devineau, F. Moutarde, and J. Yang, "Generative Model for Skeletal Human Movements based on conditional DC-GAN applied to pseudo-images," *Algorithms,* vol. 13, no. 12, p. 319, 2020.

[7] S. Yan, Z. Li, Y. Xiong, H. Yan, and D. Lin, "Convolutional sequence generation for skeleton-based action synthesis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 4394-4402.

[8] B. Degardin, J. Neves, V. Lopes, J. Brito, E. Yaghoubi, and H. Proença, "Generative Adversarial Graph Convolutional Networks for Human Action Synthesis," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 1150-1159.

[9] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784,* 2014.

[10] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," in *International conference on machine learning*, 2017: PMLR, pp. 2642-2651.

[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818-2826.

[12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, 2009: Ieee, pp. 248-255.

[13] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," *Advances in neural information processing systems,* vol. 29, 2016.

[14] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in neural information processing systems,* vol. 30, 2017.

[15] M. Rosca, B. Lakshminarayanan, D. Warde-Farley, and S. Mohamed, "Variational approaches for auto-encoding generative adversarial networks," *arXiv preprint arXiv:1706.04987,* 2017.

[16] S. Barratt and R. Sharma, "A note on the inception score," *arXiv preprint arXiv:1801.01973,* 2018.

[17] A. Glandon *et al.*, "3d skeleton estimation and human identity recognition using lidar full motion video," in *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019: IEEE, pp. 1-8.

[18] S. Song, C. Lan, J. Xing, W. Zeng, and J. Liu, "An end-to-end spatio-temporal attention model for human action recognition from skeleton data," in *Proceedings of the AAAI conference on artificial intelligence*, 2017, vol. 31, no. 1.

[19] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2015: IEEE, pp. 4580-4584.

[20] Y. Du, Y. Fu, and L. Wang, "Representation learning of temporal dynamics for skeleton-based action recognition," *IEEE Transactions on Image Processing,* vol. 25, no. 7, pp. 3010-3022, 2016.

[21] Q. Ke, M. Bennamoun, S. An, F. Sohel, and F. Boussaid, "A new representation of skeleton sequences for 3d action recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3288-3297.

[22] R. Choudhary and H. K. Gianey, "Comprehensive review on supervised machine learning algorithms," in *2017 International Conference on Machine Learning and Data Science (MLDS)*, 2017: IEEE, pp. 37-43.