Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

PPHAR-GAN: Privacy-Preserving Action Recognition Using Generative Adversarial Network-Based Deep Learning In Surveillance Videos

Bodupally Janaiah[1] , Suresh Pabboju[2]

[1]Research Scholar Osmania University, Department of CSE Hyderabad India
[2]Professor Chaitanya Bharathi Institute of Technology Department of IT , Hyderabad India.
Email: [1]janaiahmvsr@gmail.com [2]plpsuresh@gmail.com

Maintaining public safety and security, particularly in urban areas, now depends on the ability to recognize human behavior from surveillance images. Surveillance cameras in public spaces are becoming increasingly common for the same reason. Video analytics and human activity recognition have been simplified by artificial intelligence (AI). Advances in deep learning and generative adversarial network (GAN) architectures may play a major role in improving research on action recognition. However, preserving the identity of the individuals in the video is necessary for successful identification of human activities. There are instances where safeguarding privacy is essential to upholding the public interest. This means that certain privacy preservation-related limitations exist in the current research. In order to address this issue, we presented in this study a unique GAN architecture called privacy-preserving human action recognition GAN (PPHAR-GAN). This design takes use of effective human action recognition and identity concealment. Our suggestion was for an algorithm called Privacy-Preserving Human Action Recognition (PP-HAR), which makes use of PPHAR-GAN to maximize efficiency in human action recognition while maintaining privacy. The suggested technique was tested using the JHMDB and DALY benchmark datasets. The findings showed that, with the greatest accuracy of 98.51%, the suggested algorithm beats many other deep learning-based models already in use. As a result, real-time applications needing privacy-preserving human action recognition can use our framework.

Keywords - Privacy-Preserving Human Action Recognition, Deep Learning, Artificial Intelligence, Generative Adversarial Network, Public Videos Surveillance

1. INTRODUCTION

Computer vision and video analysis heavily rely on human action recognition from surveillance footage. In video footage shot by security cameras, it automatically recognizes and categorizes human movements or activities. For a number of uses, including behavior analysis, anomaly detection, and security surveillance, this technology is indispensable. Computer algorithms study motion patterns, body positions, and interpersonal relationships between people in the camera frames to identify human behaviors in security footage. Complex human activities may now be successfully recognized by DL techniques especially

CNN and RNNs. The study of human activity detection from surveillance videos with privacy preservation is a crucial field of study that aims to provide strategies and tactics for examining human behavior in video footage while safeguarding the privacy of those filmed. This field entails identifying and comprehending human behavior while protecting individuals' private information by utilizing technologies like computer vision, machine learning, and privacy-enhancing strategies. In order to preserve sensitive data, like faces or identities, while still detecting and classifying human behaviors, researchers are developing algorithms. Striking a balance between the necessity of video analysis for security or surveillance reasons and the value of upholding peoples' right to privacy is the aim.

Numerous studies have been conducted on the identification of privacy-preserving actions, as examined in, among other places, [1], [3], [6], and [9]. Rajput et al. [1] suggested that strong security may be provided by a secure identification technique that combines little data overhead with picture obfuscation. Trust concerns between data owners and suppliers provide a difficulty to the development of cloud-based expert systems. Wu et al. [3] an adversarial training paradigm designed to maximize privacy budgets while preserving privacy for deep learning action recognition. Krishnaswamy and Zhang [6] influenced by the development of silicon phased array technology. Next-generation MIMO technology addresses analog/RF interference, as demonstrated by experiments. Hao et al. [9] suggested WiNN approach, which combines WiFi and video data, improves HAR robustness by overcoming environmental limitations and achieving over 80% accuracy. Based on the literature analysis, it was found that further work has to be done to improve privacy-preserving action recognition in order to balance data utility and privacy.

The following are our contributions to this publication. In this research, we suggested a unique GAN architecture called privacy-preserving human action recognition GAN (PPHAR-GAN). This design makes use of effective human activity recognition and identity concealment. Our suggested technique, called Privacy-Preserving Human Action Recognition (PP-HAR), leverages PPHAR-GAN to maximize efficiency in human action recognition while maintaining privacy. The suggested approach was assessed using the UCF50 benchmark dataset, and the findings showed that it performed better than many deep learning-based models already in use. This is how the rest of the paper is organized. The literature on current DL -based techniques for action recognition that protects privacy is reviewed in Section 2. In Section 3, we propose an architecture and methodology for improving privacy-preserving action recognition. Section 5 concludes our investigation with recommendations for future research directions. Section 4 presents the experimental results.

2. RELATED WORK

Rajput et al. [1] recommended safe identification solution combines image obfuscation with little data overhead to offer excellent security. Emerging cloud-based expert systems have challenges due to trust issues between providers and data owners. Yan et al. [2] improved autonomous video analysis in surveillance and assisted living with a privacy-first mindset. The proposed method maintains accuracy without obscuring the desired data. Wu et al. [3] a DL adversarial training model that preserves privacy for action recognition while optimizing privacy budgets. Imran et al. [4] established a successful deep learning system for real-time violence detection in videos by utilizing ADIs, Mobile Net, and GRU while maintaining privacy. Wang et al. [5] with the use of non-invertible motion characteristics calculated through phase correlation and transformation, lens-free coded aperture cameras are intended to provide privacy-preserving action detection.

Krishnaswamy and Zhang [6] influenced by the development of silicon phased array technology. Next-generation MIMO technology offers an experimentally confirmed solution to analog/RF interference. Zare et al. [7] suggested 2D video clip representation, VSTM, gets over CNN roadblocks and yields superior results for CNN-based action recognition. Kumar and Harikiran [8] enhanced human privacy and surpasses existing methods for privacy-preserving activity recognition with the high action identification accuracy of the suggested OPA-PPAR algorithm. Hao et al. [9] suggested WiNN approach, which combines WiFi and video data, improves HAR robustness by overcoming environmental limitations and achieving over 80% accuracy. Chaudhary et al. [10] Data size and privacy preservation are very important in HAR for computer vision. A unique PDI network addresses these problems by offering effective visual representation and privacy.

Wu et al. [11] popular area of computer vision that deals with issues like occlusions and crowded backgrounds is video-based human action detection. Though they have limitations, deep learning techniques provide potential possibilities. Jaouedi et al. [12] acknowledged human behavior is necessary for several uses. A hybrid deep learning model achieves very high accuracy on the KTH dataset. Khan et al. [13] HAR is solved by combining DNN with multitier features, which improves accuracy across a wide range of datasets. Jin et al. [14] tackled computational challenges by presenting a real-time CNN-based temporal image and human activity recognition method. Pareek et al. [15] explored how machiene leraning and deep learning techniques were applied for HAR between 2011 and 2019, focusing on issues, datasets, applications, and future advances.

Mihanpour et al. [16] used CNN and DB-LSTM on raw video frames, a unique approach to human activity identification achieves excellent accuracy. Dash et al. [17] for human action recognition, a new framework that preserves extended temporal information without complicated CNN input windows combines SIFT and CNN. Gao et al. [18] for remote healthcare monitoring, a recurrent 3D convolutional network (R3D) that combines 3D CNN and LSTM is suggested. Yu et al. [19] used algorithms like CNN+LSTM, 3D CNN, and Two-Stream CNN, human action recognition seeks to recognize activities in videos. Akula et al. [20] focused on IR-based human action recognition for AAL systems, using 2D-CNN architecture to achieve an accuracy.

Hussain et al. [21] presented an accurate HAR method that does not require convolution by employing retrained Vision Transformer for spatial information and LSTM for temporal dependencies. The UCF50 and HMDB51 datasets exhibit enhanced accuracy according to experimental findings. Amrutha et al. [22] used AI, ML, and deep learning to improve video surveillance, authorities may quickly become aware of any suspicious activity and be notified in real-time. Arunnehru et al. [23] surpassed the accuracy of current approaches by proposing 3D-CNN with motion cuboid for real-time action identification in surveillance footage. Koli and Bagban [24] focused on hand gesture detection using CNN to help deaf and mute people communicate. Ahmad et al. [25] used Chi-2 and mutual information, the new HAR approach picks the best fit by extracting features from the VGG19 model and increasing accuracy.

Elharrouss et al. [26] suggested approach is efficient in detecting, identifying, and summarizing various human behaviours by leveraging motion tracking and HOG-based recognition. Surek et al. [27] achieved great accuracy on the HMDB51 dataset, deep learning models such as ViT and ResNet improve human action detection. More deep learning architectures being investigated in future research. Farrajota et al. [28] suggested a technique that combines high-level and low-level characteristics to provide optimal outcomes for human activity identification. Plans for the future call for adding motion data and working with bigger datasets. Parro et al. [29] presented a real-time, edge computing optimized system for video surveillance activity recognition and people detection. For action recognition, it uses a lightweight feature vector with LSTM, attaining SOTA accuracy in challenging situations. Ullah et al.k [30] gathered and examines seventy academic papers with an emphasis on deep learning architectures, techniques, problems, and datasets. HAR uses many input formats and a variety of deep neural architectures, with 3D convolutional networks being a prominent choice. Difficulties including a range of visual appearances, variation within and across classes, dataset constraints, and over fitting.

Serpush and Rezaei [31] approached uses CNN, LSTM, and Softmax-KNN to identify key frames that increase human action recognition. It performs better on the UCF101 dataset than earlier techniques. Although there are trade-offs between speed and precision, it is efficient. Predicting future behaviour is the goal of future research. Wei et al. [32] improved human action recognition by merging visual and inertial inputs through feature- and decision-level fusion. Raval et al. [33] explored optical surveillance's use of human activity recognition (HAR), describing methods, characteristics, models, and datasets. Yu et al. [34] introduced P-RRNNs, two-stream CNNs that improve the recognition of human actions. Results from experiments on the UCF101 and HMDB51 datasets show how effective it is. Deeper P-RRNNs and feature fusion will be investigated in more detail. Chaudhary et al. [35] issued with privacy and data size in computer vision. To tackle this, a Pose Guided Dynamic Image (PDI) network summarizes films to help recognize human actions while maintaining anonymity.

Kumar et al. [36] for the safety fields, human activity recognition is essential. With more computing, the Gated Recurrent Neural Network enhances categorization. Algorithm performance is influenced by feature extraction. The suggested approach has potential across a range of datasets. Mohan et al. [37] automated methods using CNN and PCANet enhance detection accuracy; manual monitoring presents difficulties. Video surveillance is becoming more and more common in public spaces. Meng et al. [38] for action recognition, a novel network called QST-CNN-LSTM combines quaternion spatial-temporal CNN with LSTM. Accuracy on a range of datasets is improved. Snoun et al. [39] presented a unique method based on skeleton extraction from films for human action recognition. There are three methods put forth: body articulations, skeleton superposition, and dynamic skeleton. They are classified by combining them with CNNs. Tested and produced better results on the RGBD-HuDact and KTH datasets. Abdellaoui and Douik et al. [40] outlined a cutting-edge technique for using

Deep Belief Networks (DBNs) to identify human behavior. When evaluated on datasets from KTH and UIUC, they produced accuracy levels higher than 95%. Subsequent research endeavors will investigate the combination of motion capture and unsupervised classification.

3. PROPOSED SYSTEM

The section presents our methodology for privacy-preserving action recognition from surveillance videos, including the proposed framework mechanisms and underlying algorithm.

3.1 Problem Statement

Let's say we are using raw images from camera (X) for training. In addition, we have a budget for privacy (B) and a goal task (T). We mathematically define the aim of privacy-preserving visual identification as follows, where γ is a weight coefficient:

$$\min_{f_T,f_d} L_T(f_T(f_d(X)), Y_T) + \gamma L_B(f_d(X)), \qquad (1)$$

where f_T is a representation of the model's performance on the input data for the target task T. Since supervised tasks like as action recognition and visual tracking are often run on X, a label set Y_T is supplied, and the task performance on T is assessed using a standard cost function L_T (e.g., softmax). But first, we need to develop the $L_B$budget cost function.The likelihood of privacy leaking increases withL_B. This enables us to assess the input data's risk of privacy leakage. Our objective is to find an active degradation function, f_d to alter the original X, which serves as the shared input for L_T and L_B.

– When comparing the target task performance L_T to that of utilizing the raw data, i.e.,

$$\min_{f_T,f_d} L_T(f_T(f_d(X)), Y_T) \approx \min_{f_T'} L_T(f_T'(X), Y_T).$$

– When compared to unprocessed data, i.e., $L_B(f_d(X)) \ll L_B(X)$. the finances for privacy LB is significantly decreased.

The privacy budget cost L_B is not easily defined. Task-driven implementation is often necessary, and it must be practically executed in designated application contexts. In smart homes or offices with video surveillance, for instance, it could be preferable to keep people's faces and identities hidden. As a result, on the edited video$f_d(X)$., a decrease in $L_B$as may be interpreted as a suppression of the identity recognition. This approach may also be used to determine privacy-related characteristics like age, gender, or ethnicity. We transform L_B $f_d(X)$into $L_B(f_b(f_d(X)), Y_B)$, where f_b is the budget model used to predict the related privacy data. For example, we represent identity label and other privacy-related annotations as Y_B. Reduction of L_B as opposed to L_T, will encourage$f_b(f_d(X))$ to diverge from YB to the greatest extent possible.

This kind of task-driven, supervised formulation of L_B poses at least two challenges: Oftentimes, target task labels are more accessible than the annotations linked to privacy budgets, shown by Y_B Particularly, (1) Y_T and $Y_B$accessible on the same X are frequently incompatible; and (2) given the nature of privacy protection, just lowering the success rate of one$f_b$model is insufficient. To guarantee the best feasible privacy protection, define a family of privacy prediction functions $P: f_d(X) \rightarrow Y_B$. Next, think about eliminating all possible models f_b from P. This is different from the typical supervised training goal, which calls for the identification of a single model to do the intended task. We rebuild the generic form (1) of $L_B$using the task-driven definition.

$$\min_{f_T, f_d} L_T(f_T(f_d(X), Y_T) \ + \ \gamma \max_{f_b \in P} L_B(f_b(f_d(X)), Y_B). \qquad (2)$$

A minimum of one f_T function that can correctly predict Y_T from $f_d(X)$; exists, or should exist; and (2) for all (or should exist) f_b functions ∈ P, not even the best one can predict YB from fd(X). These are the two objectives for the solution $f_d$that should be met concurrently. Much of the earlier work used an empirical f_d (like simple downsampling) to solve $\min_{f_T}, f_d \ L_T(f_T(f_d(X), Y_T)$ [9,61]. Essentially, the equation $\min_{f_T}, f_d \ L_T(f_T(f_d(X), Y_T)$ was solved in [47] in order to jointly adapt f_d andf_T. Next, the influence of fd on L_B—which is defined as face recognition mistake rates—was confirmed by the authors through empirical verification. As a result, privacy protection cannot be guaranteed because these techniques do not expressly optimize for budgets.

As opposed to Traditional Adversarial Training Whether the adversarial perturbations are intended to "fool" a single $f_b$or all possible f_bS is the main difference between (2) and earlier studies using traditional adversarial training [43, 38]. We believe that the latter is crucial as it considers generalization capacity to suppress undetected privacy violations. Moreover, the majority of ongoing studies focus on perturbations that, when applied to the pixel domain, have the least potential visual impact on people, such as the ℓ_p norm restriction. That is not in keeping with our objectives. Reducing the disturbance in the (learned) feature domain of the intended utility task might be one approach to conceptualize our model.

3.2 Proposed Framework

To put it briefly, Figure 1 shows a model architecture that may be used to execute the suggested formulation (2). In order to create the anonymized video $f_d(X)$.,., the original video data X is first obtained as input and is subsequently supplied via the active degradation module f_d. The anonymized video is concurrently fed through a privacy prediction model (f_b,). and a target task model (f_T) during training. The three modules—f_d, f_T, and $f_b$can all be learned and used with neural networks. The hybrid loss of L_T and $L_B$is is used to train the whole model. The objective of privacy-preserving visual acknowledgement will be achieved by fine-tuning the whole pipeline$f_d(X)$, which will identify the best task-specific transformation to the benefit of the target task and the drawback of the privacy violation. Following training, incoming video (from a camera, for example) may be anonymized using the obtained active degradation from the local device and transferred to the backend (cloud, for example) for target task analysis.

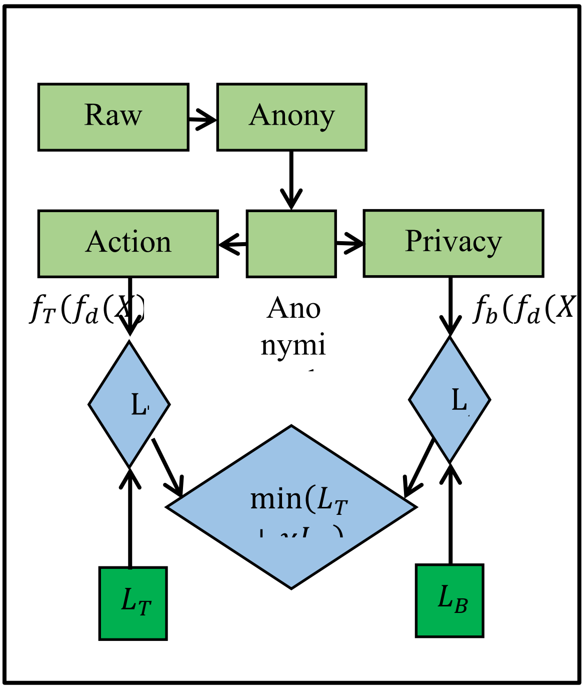


Figure 1: Proposed framework for privacy-preserving human action recognition

A private, end-to-end, scalable, and flexible pipeline for visual identification is produced by the suggested approach. Its method is associated with the recently developed field of feature disentanglement research [64]. Using that approach, factorized latent representations are produced in non-overlapped groups that accurately characterize the data associated with certain desired properties. Generative models [10,51] and reinforcement learning [20] are examples of preliminary applications.

The training exhibits the same vulnerability to inadequate local minimums and/or collapse as other adversarial models, including GANs [16]. Thus, we provide a well-designed training method using a three-module alternating update technique, which may be likened to a three-party game, as detailed in the appendix. Preventing any of the three modules (f_d, f_T, and f_b) from changing "too quickly" is our main goal in theory. Keep an eye on L_T and $L_b$to to see which of the three modules need an update next.

Options pertaining to f_d, f_T, **and** f_b a significant impact on the performance will come from the decisions made in the three modules. One way to produce fd as a nonlinear mapping is by filtering, as was discussed in [47]. A natural picture need not be the output $f_d(X)$; the shape of f_d might be arbitrarily chosen. Let us assume, for convenience, that fd is a 2-D convolutional neural network (CNN) that is trainable. What comes out of it is $f_d(X)$, a 2-D feature map with the same resolution as the input video frame. This is only true for the first concatenation of building blocks; models that have been trained on real pictures are often used to startf_T and f_b.

Additionally, $f_d(X)$ should preferably be light and compact because it will be delivered to the cloud across (limited-bandwidth) channels.

To ensure the impact of f_d, it's critical to choose models that are both f_T and $f_b$sufficiently resilient to engage in competition. We provide improvements to the state-of-the-art CNNs utilized for similar tasks by applying the resilient pre-training technique proposed in [61] for the degraded input $f_d(X)$,

Particular consideration should be given to the budget cost (second term) that is specified in (2). Since it is unknown how we can be positive that f_b is the "best possible" privacy prediction model when we employ it with a certain CNN architecture, we call this "the ∀ Challenge". In other words, even if a f_d function fails one f_bmodel and exposes privacy, is it conceivable for another $f_b^{'}$ $\in P$ to predict Y_B given $f_d(X)$ Selecting a robust privacy prediction model would be a naïve empirical approach, presuming that af_d function that can fool this strong one would also be able to fool other potential functions, even if it is computationally unfeasible to search over P fully. However, the resultant $f_d(X)$ does not generalize and could overfit the features of a single unique f_b. Two more sophisticated and useful recipes are provided in Section 3.3.

Selections for L_T **and** L_B We consider target task f_T and privacy prediction fb as classification models with output class labels, without sacrificing generality. The goal task T may be performed as efficiently as possible by simply selecting L_T as the KL divergence, or $KL(f_T(f_d(X), Y_T)$.

Type equation here.

To maximize the divergence between $f_b(f_d(X))$ andY_B, we need to reduce the privacy budget $L_B(f_b(f_d(X)), Y_B)$. This leads to a non-standard and difficult method of obtaining L_B. Although minimizing a concave function would result in several numerical instabilities that frequently blow up, one potential solution is the negative KL divergence between the projected class vector and the ground truth label. Alternatively, we exploit and minimize the negative entropy function of the projected class vector to promote "uncertain" predictions. We will employ Y_B in the interim to ensure a strong enough $f_b$during commencement. Additionally, for the model to resumeY_B will be necessary.

Type equation here.

3.3 Action Recognition Module

For effective human activity recognition from surveillance footage, we suggested a hybrid DL approach based on GANs. Action recognition as well as feature engineering are supported in the suggested GAN-based scheme. An outline of the proposed module is shown in Figure 2.

In order to acquire features—which are crucial for action detection later on—a DL model called ResNet50 is employed. The action recognition process is aided by the GAN model in the other section of the suggested architecture. In the proposal system, the convolutional LSTM model serves as the foundation for action recognition. An image identification-focused deep neural network is called ConvNet. Through the simulation of the anatomy of the human visual cortex, it detects objects using a network of neurons. This study uses LSTM to solve the problem since video data is sequential and affects subsequent frames. Feedback from the preceding node is sent to the next node in a recurrent network termed LSTM. To prevent conflicts between classes and enable regenerating output and input feedback, the proposed model substitutes an adversarial loss function for the conventional cross-entropy.

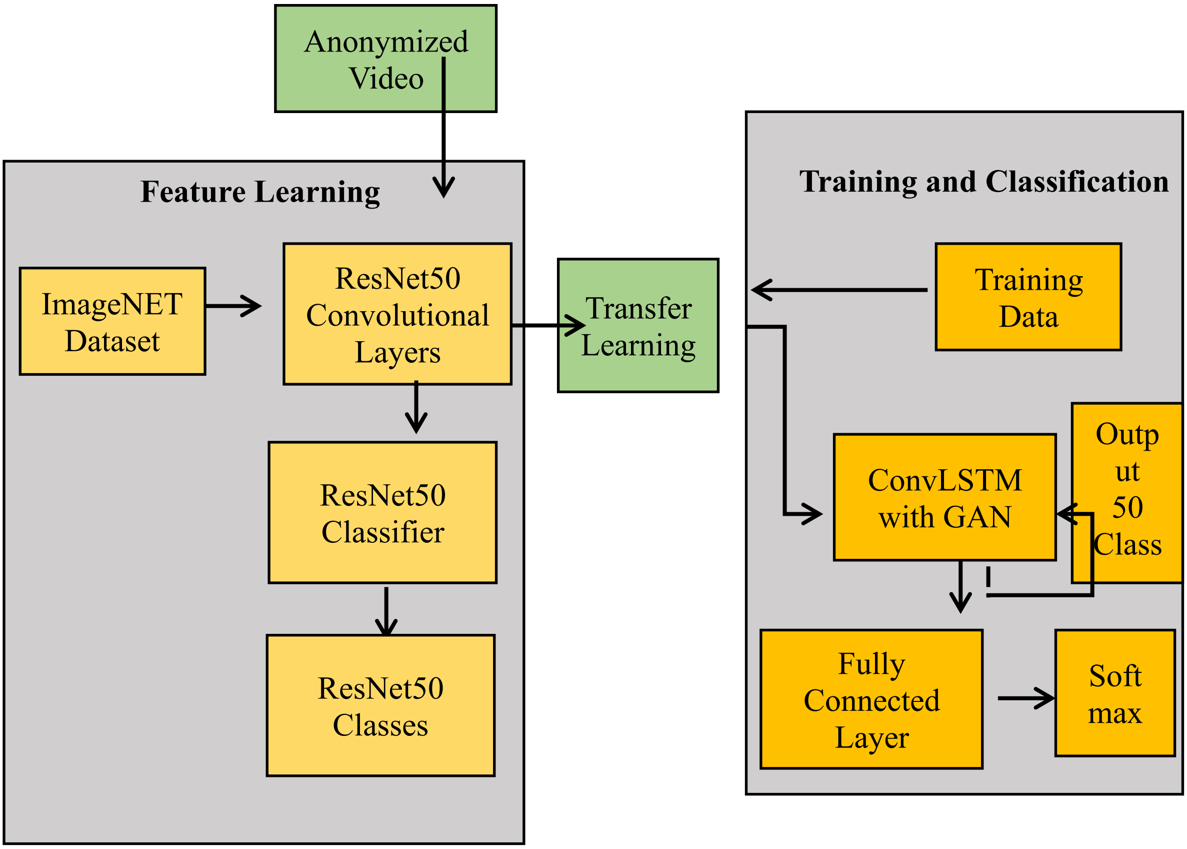


Figure 2: Proposed methodology for action recognition

The ResNet50 model is necessary in order to extract features from an input video. The ImageNet dataset serves as the pre-training data for our deep learning model. But the tagged surveillance films in the training dataset are used to retrain the model. The training and classification modules of ResNet50 employ features derived from the feature learning module. A convolutional LSTM model is part of the GAN architecture, which is utilized by the suggested system. Using a generator and discriminator used for action detection in surveillance footage, the GAN architecture carries out its iterative process. A softmax layer, a fully connected layer, and multi-class classification receive the output of the GAN model.

3.4 Dealing with the ∀ Challenge

Thus, we study two simple and simply implementable strategies to enhance the learnt f_d's generalization across all conceivable $f_b \in P$ (i.e., no model is able to predict privacy with any degree of reliability).

Future research will examine other advanced model resampling or model-search techniques, such as [68].

Resuming the Budget Model We substitute random weights for the current weights in f_b at a predetermined training point (for example, when the privacy budget $L_B(f_b(f_d(X)))$ stops decreasing). By avoiding minor overfitting between $f_b$and fd (i.e., fd is specialized primarily to confuse the current $f_b)$, this random restarting attempts to eliminate unnecessary increases in parameter values. The existing fd(X) is trained to be strongly rivaled by the new model $f_b$by freezing the f_d and f_T training and switching to minimizing $KL(f_b(f_d(X)), Y_B)$, until the new f_b has received significant training to become a formidable rival to the current fd(X):Next, we unfreeze f_d and f_T and replace $f_b$with the reduction in entropy to zero, resuming the adversarial training. There are several ways to repeat it.

Combined Budget Model In the alternative method, the continuous P is approximated by means of a distinct collection of M example functions. Now that we have the budget model ensemble $\{f_b^i\}_{i=1}^{M}$,, we minimize the discretized surrogate of (2) as follows:

$$min_{f_T, f_d} L_T(f_T(f_d(X), Y_T) + \gamma\, max_{i\in\{1,2,\dots,M\}} L_B(f_b^i(f_d(X))). \qquad (3)$$

Restricting (3) to the lowest value will only suppress, at each iteration (mini-batch), the model $f_b^i$with the largestL_B cost, or the one that is "most confident" about its current privacy forecast. An example of the previous fundamental framework is given by (3) with M = 1. It is simple to combine the group technique with restarting.

3.5 Proposed Algorithm

Our suggestion was for an algorithm called Privacy-Preserving Human Action Recognition (PP-HAR), which makes use of PPHAR-GAN to maximize efficiency in human action recognition while maintaining privacy.

```
Algorithm: Privacy-Preserving Human Action Recognition (PP-HAR)
Input: Dataset D
Output: Results of human action recognition R, performance statistics P
    Begin
    (T1, T2)←SplitData(D)
    Train ResNet50 with ImageNet
    Retrain ResNet50 with T1
    featureMaps←FeatureLearningUsingResNet50(T1)
    For each featureMap in featureMaps
        Generator function
        Discriminator function
    End For
    optimizedFeatureMaps←GANWithConvLSTM(featureMaps)
    T2'←Anonymization(T2)
    R←RecognizeActions(FC, T2)
    P←Evaluation(R, ground truth)
    Display R
    Display P
    End
```

Algorithm 1: Privacy-Preserving Human Action Recognition (PP-HAR)

Algorithm 1 accepts a dataset D as input and produces performance statistics P and the results of human action recognition R. Using the function SplitData(D), the program first divides the dataset D into two halves, T1 and T2. After that, it uses ImageNet to train a ResNet50 model and retrains it using the first section of the dataset T1. Next, the ResNet50 model learns feature maps from T1. A generator and discriminator function are applied to every feature map, suggesting that a GAN is being used to improve the feature maps. Then, to improve these feature maps, the technique uses a ConvLSTM network. It is suggested that anonymizing the data used for action recognition preserves privacy, as this is how the second portion of the dataset, T2, is anonymized using the function Anonymization(T2). In order to identify activities in the anonymized dataset T2, the algorithm proceeds with a function called RecognizeActions(FC, T2). Here, FC stands for Fully Connected layers. Comparison of the recognized actions R with the real ground truth labels is how the action recognition performance is assessed; at the end, results R and performance statistics P are shown. In conclusion, the PP-HAR algorithm protects the privacy of the persons in the dataset by securely identifying human behaviors while anonymizing the data. It uses a convolutional long short-term memory (CNV) network for optimization, a GAN for feature learning, and a pre-trained ResNet50 model for feature refinement. The user is shown the results of the algorithm's performance evaluation against ground truth data.

3.6 Evaluation Method

Let us consider the following: evaluation set X^e, target task labels Y_T^e privacy annotations Y_B^e, and training data X. Compared to traditional visual identification exams, our evaluation is substantially harder. After the learning active degradation is implemented, we should assess if the learnt target task model will continue to perform well in the future or whether a random privacy prediction model will perform worse. To determine the classification accuracy A_T one can first proceed as per usual approach, which involves utilizing the acquired f_d(and f_T to X^e, to compare $f_T(f_d(X^e))$ w.r.t. Y_T^e .The larger the number, the better. Should we just note that the ∀ issue causes the learned f_d and $f_b$to to yield subpar classification accuracy on X^e, then the second assessment is unsatisfactory. Stated otherwise, f_d has to generalize both the data space and the$f_b$model space. We provide a novel approach to experimentally confirm that $f_b$prevents trustworthy privacy prediction for alternative models: we first re-sample a distinct set of N models $\{f_b^j\}_{j=1}^N$from P, none of which will overlap with the M budget models that were trained. Using the learnt f_d, we next train each of them to predict privacy information over the degraded training data X; in other words, we minimize $f_b^j\ (f_d(X)), j = 1, \dots, N$. Using each trained instance of f_b^j and f_d we finally compute X^e to get the classification accuracy of the j-th model. To protect f_ds privacy, the best accuracy among the N models on $f_d(X^e)$,, represented as A_b^N, will be selected by default; the lower the better.

4. EXPERIMENTAL RESULTS

5.1 Datasets

In order to create this dataset, we used two datasets: DALY and JHMDB. The DALY dataset was first presented in [21] and is sourced from [22]. Over thirty hours of YouTube videos with annotations in both geographical and temporal domains make up the collection. There are ten human behaviors that are observed daily, totaling 3600 occasions. There are distinct time bounds when considering action classes. To get over the uncertainties brought on by noise, this is crucial. As seen in Figure 3, a few of the activity courses include drinking, using lipstick on the lips, cleaning teeth, making phone calls, snapping pictures, and playing the harmonica.

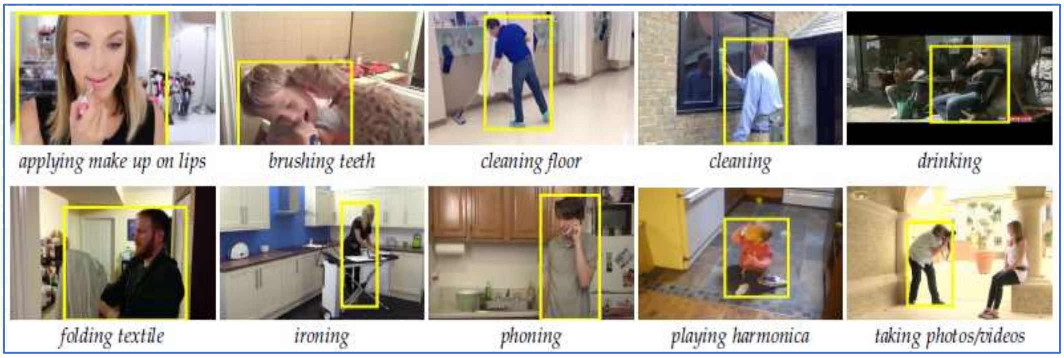


Figure 3: Representative pictures with human action annotated for10 classes of DALY dataset

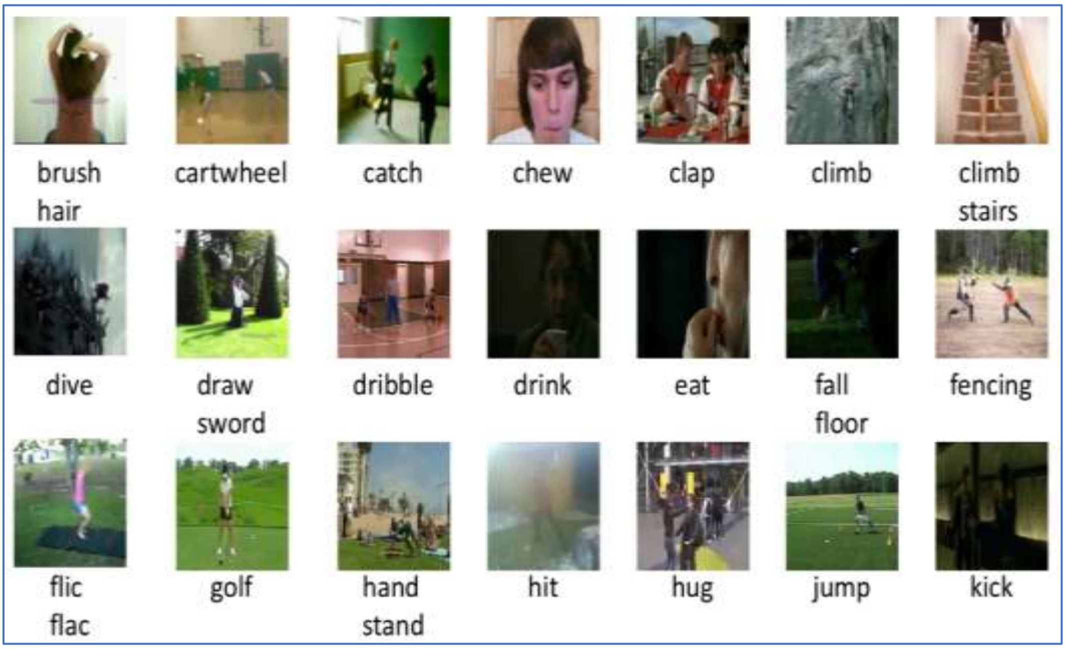


Figure 4: Some of the images from JHMDB dataset with annotated human actions

JHMDB is an additional dataset that was presented in [23] and gathered from [24]. This dataset serves as a standard for posture estimation, human activities detection, and human detection. In reality, JHMDB is created from the HMDB51 dataset [25], which has 5,100 movies with 51 distinct human behaviors in it. As shown in Figure 4, JHMDB is a subset of HMDB51 that has 21 categories involving a single individual doing certain behaviors. A few of the motions include leap, run, kick, embrace, and so on.

5. Results

JHMDB and DALY datasets are used in the research. Face verification error and mean average accuracy are the metrics used to describe the experimental outcomes. Numerous baseline or state-of-the-art methods are contrasted with our anonymization technique. They consist of edge, super-pixel, masked, noise x 3, and blurr x 3. As seen in Figure 5, the outcomes are also visible as altered photographs.

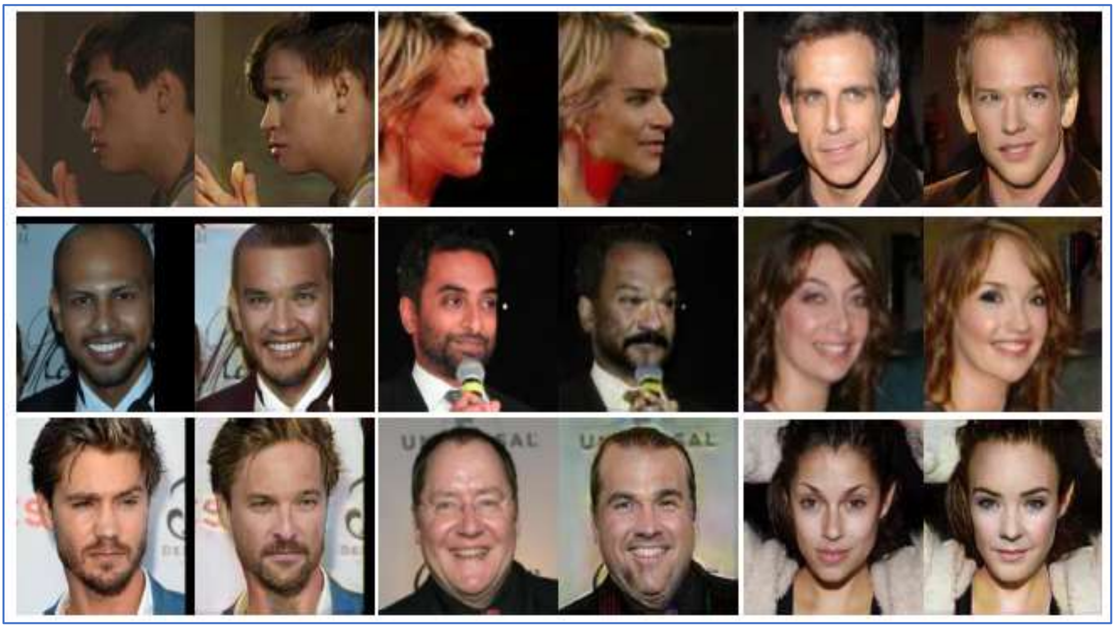


Figure 5: Input images from outside the datasets before and after anonymization

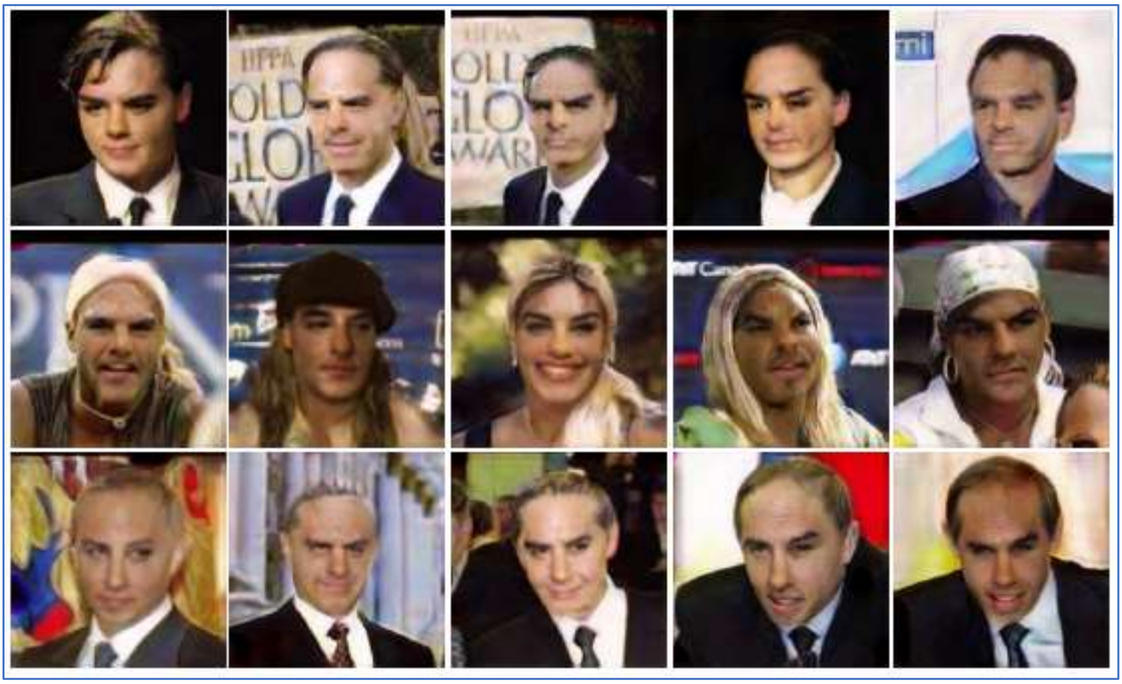


Figure 6: Identifiable photos of the suggested system for user research

The visible difference between the before and after face change is provided by the qualitative findings displayed. Our program first does anonymization before recognizing actions, which is the reason behind this. Our anonymization technique has been proven to work well based on a user research including prominent figures. Figure 6 displays the anonymous samples that were used for the user research. The observations in Figure 7 are the result of experiments conducted using DALY and JHMDB about face verification by discriminator in oppositional learning.

HAR Model	**Accuracy (%)**
CNN	87.34
LSTM	90.54
ConvLSTM	95.23
GAA-HAR	97.73
PP-HAR (Proposed)	98.51

Table 1: Performance comparison among deep learning models in human action recognition

Table 1 shows the accuracy and identification of human behaviors in surveillance footage for all DL models, including the state-of-the-art and suggested models.

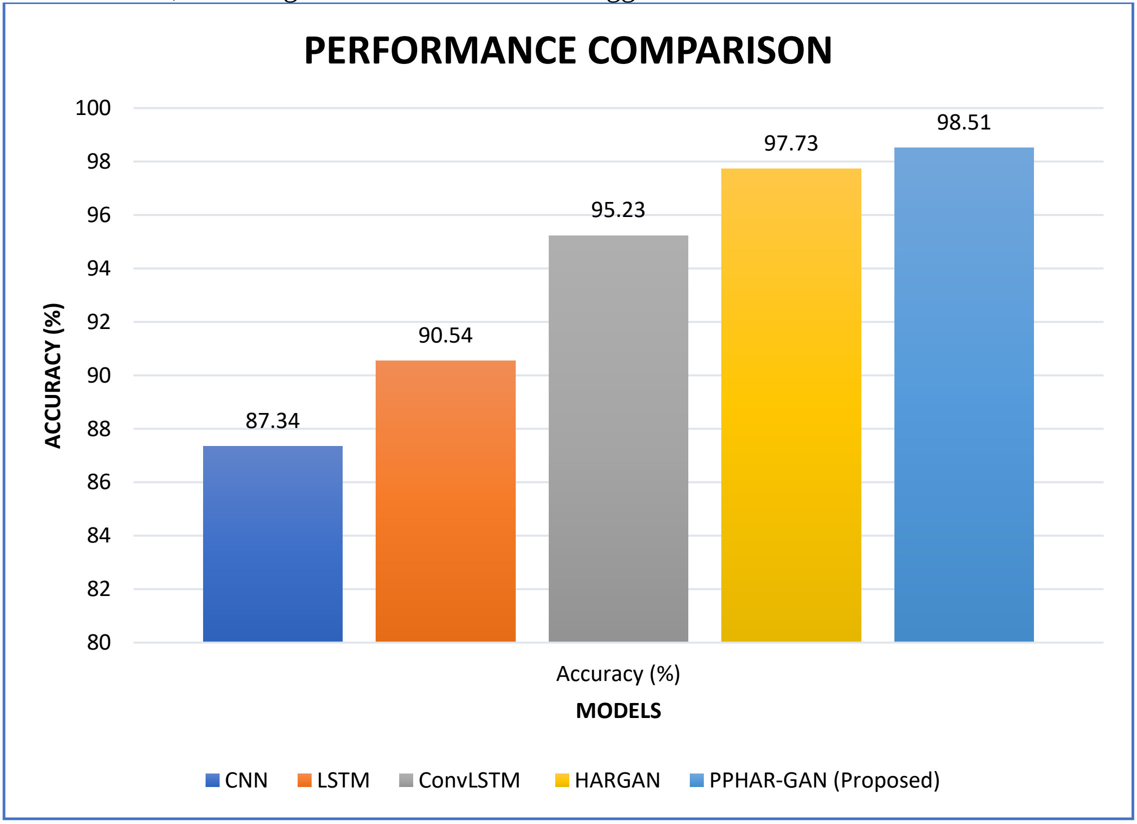
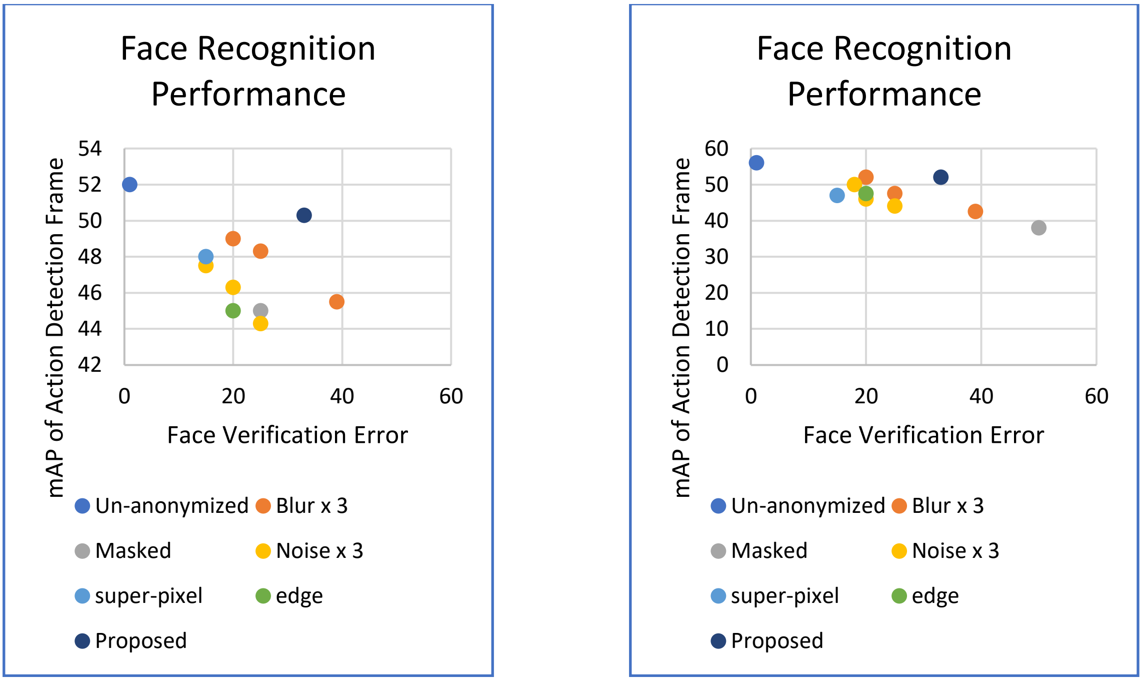


Figure 6: Performance comparison among deep learning models

Figure 6 presents an accuracy comparison of the performance of several DL models in recognising human actions. Owing to the structure and underlying functionality of each deep learning model, varying degrees of accuracy were exhibited. On the horizontal axis, the models' accuracy is indicated, and on the vertical axis, the accuracy value. The baseline CNN model obtained an accuracy of 87.34%, according to the experimental findings. The CNN model's performance was inferior to that of the baseline LSTM model, which attained 90.54% accuracy. Convolutional LSTM, a hybrid DL model, outperformed CNN and LSTM models with an accuracy of 95.23%. In contrast, the HARGAN model outperformed the CNN, LSTM, and Convolutional LSTM models, with an accuracy of 97.73%. The PPHAR-GAN model, which is based on DL, has the greatest accuracy of 98.51%.



(a) (b)

Figure 7: Face recognition performance using JHMDB (left) and DALY (right) (**Note:** high error rate means high performance)

As seen in Figure 7, several anonymization techniques are used to compare the face recognition performance. Both the JHMDB and the DALY databases show the results. Higher performance is indicated by a high recognition mistake rate as face recognition is intended to be inhibited in the adversarial scenario. Consequently, the suggested method has outperformed the state of the art in terms of performance. Action recognition and privacy preservation have both improved, according to the empirical investigation. The project aims to correctly identify human behaviors while avoiding identity revelation. This has been satisfied as the experimental study's observations are made. Additionally, the findings show that the suggested strategy outperforms baselines in terms of anonymization. Both empirical research and user research provide proof for it.

5. CONCLUSION AND FUTURE WORK

The unique GAN architecture we suggested in this study is called privacy-preserving human action recognition GAN (PPHAR-GAN). This design takes use of effective human action recognition and identity concealment. Our suggestion was for an algorithm called Privacy-Preserving Human Action Recognition (PP-HAR), which makes use of PPHAR-GAN to maximize efficiency in human action recognition while maintaining privacy. The suggested framework's action recognition module makes use of a hybrid deep learning strategy that combines action recognition, training, and feature learning. This module is in charge of appropriately identifying human behaviors from anonymised video input. The JHMDB and DALY benchmark datasets were used to assess the suggested method. With the maximum accuracy of 98.51%, the findings showed that it exceeds several deep learning-based models currently in use. These are some ideas about where the research should go in the future. Initially, there is need for improvement in the suggested approach to enable scalable analysis of numerous films at once. Secondly, live-streaming videos from different domains in real time must be used to assess the suggested framework.

References

[1] Rajput Amitesh Singh, Raman Balasubramanian and Imran Javed. (2020). Privacy-preserving human action recognition as a remote cloud service using RGB-D sensors and deep CNN. Expert Systems with Applications, 152, pp.1–15. doi:10.1016/j.eswa.2020.113349

[2] Yan, Jiawei; Angelini, Federico and Naqvi, Syed Mohsen (2020). Image Segmentation Based Privacy-Preserving Human Action Recognition for Anomaly Detection, IEEE, pp.8931–8935. doi:10.1109/ICASSP40776.2020.9054456

[3] Wu, Zhenyu; Wang, Haotao; Wang, Zhaowen; Jin, Hailin and Wang, Zhangyang (2020). Privacy-Preserving Deep Action Recognition: An Adversarial Learning Framework and A New Dataset. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp.1–14. doi:10.1109/TPAMI.2020.3026709

[4] Javed Imran;Balasubramanian Raman;Amitesh Singh Rajput; (2020). Robust, efficient and privacy-preserving violent activity recognition in videos . Proceedings of the 35th Annual ACM Symposium on Applied Computing, pp.1–8. doi:10.1145/3341105.3373942

[5] Wang, Zihao W.; Vineet, Vibhav; Pittaluga, Francesco; Sinha, Sudipta N.; Cossairt, Oliver; Kang, Sing Bing (2019). Privacy-Preserving Action Recognition Using Coded Aperture Videos. IEEE, pp.1–10. doi:10.1109/cvprw.2019.00007

[6] Liang, Chengwu; Liu, Deyin; Qi, Lin; Guan, Ling (2020). Multi-Modal Human Action Recognition With Sub-Action Exploiting and Class-Privacy Preserved Collaborative Representation Learning. IEEE Access, 8, pp.39920–39933. doi:10.1109/ACCESS.2020.2976496

[7] Zare, Amin; Abrishami Moghaddam, Hamid; Sharifi, Arash (2019). Video spatiotemporal mapping for human action recognition by convolutional neural network. Pattern Analysis and Applications, pp.1–15. doi:10.1007/s10044-019-00788-1

[8] Kambala Vijaya Kumar, Jonnadula Harikiran. (2022). Privacy preserving human activity recognition framework using an optimized prediction algorithm. IAES International Journal of Artificial Intelligence (IJ-AI). 11(1), p.254~264.

[9] Yanling Hao;Zhiyuan Shi;Yuanwei Liu; (2020). A Wireless-Vision Dataset for Privacy Preserving Human Activity Recognition . 2020 Fourth International Conference on Multimedia Computing, Networking and Applications (MCNA), pp.1–10. doi:10.1109/mcna50957.2020.9264288

[10] Chaudhary, Sachin; Dudhane, Akshay; Patil, Prashant; Murala, Subrahmanyam (2019). Pose Guided Dynamic Image Network for Human Action Recognition in Person Centric Videos. , IEEE, pp.1–8. doi:10.1109/AVSS.2019.8909835

[11] Di Wu, Nabin Sharma, and Michael Blumenstein. (2017). Recent advances in video-based human action recognition using deep learning: A review. IEEE, pp.1-8.

[12] Jaouedi, Neziha; Boujnah, Noureddine; Bouhlel, Med Salim (2019). fvA New Hybrid Deep Learning Model For Human Action Recognition. Journal of King Saud University - Computer and Information Sciences, pp.1–12. doi:10.1016/j.jksuci.2019.09.004

[13] Khan, Muhammad Attique; Javed, Kashif; Khan, Sajid Ali; Saba, Tanzila; Habib, Usman; Khan, Junaid Ali; Abbasi, Aaqif Afzaal (2020). Human action recognition using fusion of multiview and deep features: an application to video surveillance. Multimedia Tools and Applications, pp.1–27. doi:10.1007/s11042-020-08806-9

[14] Jin, C.-B., Li, S., Do, T. D., & Kim, H. (2015). Real-Time Human Action Recognition Using CNN Over Temporal Images for Static Video Surveillance Cameras. Advances in Multimedia Information Processing -- PCM 2015, 330–339. doi:10.1007/978-3-319-24078-7_33

[15] Pareek, Preksha; Thakkar, Ankit (2020). A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications. Artificial Intelligence Review, pp.1–64. doi:10.1007/s10462-020-09904-8

[16] Mihanpour, Akram; Rashti, Mohammad Javad; Alavi, Seyed Enayatallah (2020). Human Action Recognition in Video Using DB-LSTM and ResNet, IEEE, pp.133–138. doi:10.1109/ICWR49608.2020.9122304

[17] Samarendra Chandan Bindu Dash;Soumya Ranjan Mishra;K. Srujan Raju;L. V. Narasimha Prasad; (2021). Human action recognition using a hybrid deep learning heuristic . Soft Computing, pp.1–14. doi:10.1007/s00500-021-06149-7

[18] Gao, Yongbin; Xiang, Xuehao; Xiong, Naixue; Huang, Bo; Lee, Hyo Jong; Alrifai, Rad; Jiang, Xiaoyan; Fang, Zhijun (2018). Human Action Monitoring for Healthcare based on Deep Learning. IEEE Access, pp.1–8. doi:10.1109/ACCESS.2018.2869790

[19] Zeqi Yu;Wei Qi Yan; (2020). Human Action Recognition Using Deep Learning Methods . 2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ), pp.1–6. doi:10.1109/ivcnz51579.2020.9290594

[20] Akula, Aparna; Shah, Anuj K.; Ghosh, Ripul (2018). Deep Learning Approach for Human Action Recognition in Infrared Images. Cognitive Systems Research, pp.1–19. doi:10.1016/j.cogsys.2018.04.002

[21] Altaf Hussain , Tanveer Hussain , Waseem Ullah , and Sung Wook Baik. (2022). Vision Transformer and Deep Sequence Learning for Human Activity Recognition in Surveillance Videos. Hindawi Computational Intelligence and Neuroscience, pp.1-10.

[22] Amrutha, C.V; Jyotsna, C.; Amudha, J. (2020). Deep Learning Approach for Suspicious Activity Detection from Surveillance Video, IEEE, pp.335–339. doi:10.1109/ICIMIA48430.2020.9074920

[23] Arunnehru, J.; Chamundeeswari, G.; Bharathi, S. Prasanna (2018). Human Action Recognition using 3D Convolutional Neural Networks with 3D Motion Cuboids in Surveillance Videos. Procedia Computer Science, 133(), 471–477. doi:10.1016/j.procs.2018.07.059

[24] Koli, Rashmi R.; Bagban, Tanveer I. (2020). Human Action Recognition Using Deep Neural Networks, IEEE, pp.376–380. doi:10.1109/WorldS450073.2020.9210345

[25] Tariq Ahmad, Jinsong Wu, Imran Khan, Asif Rahim and Amjad Khan. (2021). Human Action Recognition in Video Sequence using Logistic Regression by Features Fusion Approach based on

CNNFeatures. (IJACSA) International Journal of Advanced Computer Science and Applications. 12(11), pp.1-8.

[26] Elharrouss, Omar; Almaadeed, Noor; Al-Maadeed, Somaya; Bouridane, Ahmed; Beghdadi, Azeddine (2020). A combined multiple action recognition and summarization for surveillance video sequences. Applied Intelligence, pp.1–23. doi:10.1007/s10489-020-01823-z

[27] Guilherme Augusto Silva Surek 1 , Laio Oriel Seman and Stefano Frizzo Stefenon. (2023). Video-Based Human Activity Recognition Using Deep Learning Approaches. MDPI, pp.1-15.

[28] Farrajota, M.; Rodrigues, João M. F.; du Buf, J. M. H. (2018). Human action recognition in videos with articulated pose information by deep networks. Pattern Analysis and Applications, pp.1–12. doi:10.1007/s10044-018-0727-y

[29] Antonio Carlos Cob-Parro, Cristina Losada-Gutiérrez, Marta Marrón-Romera, Alf. (2024). A new framework for deep learning video based Human Action Recognition on the edge. Expert Systems With Applications, pp.1-17.

[30] HADIQA AMAN ULLAH, SUKUMAR LETCHMUNAN, M. SULTAN ZIA and UMAIR MUNEER BUTT. (2021). Analysis of Deep Neural Networks for Human Activity Recognition in Videos—A Systematic Literature Review. IEEE Access. 9, pp.1-22.

[31] Fatemeh Serpush and Mahdi Rezaei. (2021). Complex Human Action Recognition Using a Hierarchical Feature Reduction and Deep Learning-Based Method . SN Computer Science, pp.1–15. doi:10.1007/s42979-021-00484-0

[32] Wei, Haoran; Jafari, Roozbeh; Kehtarnavaz, Nasser (2019). Fusion of Video and Inertial Sensing for Deep Learning–Based Human Action Recognition. Sensors, 19(17), pp.1–13. doi:10.3390/s19173680

[33] Raval, R.M.; Prajapati, H.B.; Dabhi, V.K. (2019). Survey and analysis of human activity recognition in surveillance videos. Intelligent Decision Technologies, pp.1–24. doi:10.3233/IDT-170035

[34] SHENG YU, LI XIE, LIN LIU AND DAOXUN XIA. (2019). Learning Long-Term Temporal Features With Deep Neural Networks for Human Action Recognition. IEEE Access. 8(.), pp.1-11.

[35] Chaudhary, Sachin; Dudhane, Akshay; Patil, Prashant; Murala, Subrahmanyam (2019). Pose Guided Dynamic Image Network for Human Action Recognition in Person Centric Videos. IEEE, pp.1–8. doi:10.1109/AVSS.2019.8909835

[36] Badhagouni Suresh Kumar;S. Viswanadha Raju;H.Venkateswara Reddy; (2021). Human Action Recognition Using A Novel Deep Learning Approach . IOP Conference Series: Materials Science and Engineering, pp.1–9. doi:10.1088/1757-899x/1042/1/012031

[37] Mohan, Aiswarya; Choksi, Meghavi; Zaveri, Mukesh A (2019). Anomaly and Activity Recognition Using Machine Learning Approach for Video Based Surveillance, IEEE, pp.1–6. doi:10.1109/ICCCNT45670.2019.8944396

[38] Meng, Bo; Liu, XueJun; Wang, Xiaolin (2018). Human action recognition based on quaternion spatial-temporal convolutional neural network and LSTM in RGB videos. Multimedia Tools and Applications, pp.1–18. doi:10.1007/s11042-018-5893-9

[39] Ahmed Snoun;Nozha Jlidi;Tahani Bouchrika;Olfa Jemai;Mourad Zaied; (2021). Towards a deep human activity recognition approach based on video to image transformation with skeleton data . Multimedia Tools and Applications, pp.1–24. doi:10.1007/s11042-021-11188-1

[40] Mehrez Abdellaoui and Ali Douik. (2020). Human Action Recognition in Video Sequences Using Deep Belief Networks. Traitement du Signal. 37(1), pp.37-44