ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

BACK POSE BASED PERSON RECOGNITION WITH PIX2PIX GENERATIVE ADVERSARIAL NETWORK

MOHANKRISHNA SAMANTULA[1,3*], SRINIVAS GORLA[2] , D. LALITHA BHASKARI [3]

[1]Dept of CSE, GST, GITAM (Deemed to be University), Visakhapatnam, INDIA

[2]Dept of CSE, GST, GITAM (Deemed to be University), Visakhapatnam, INDIA

[3]Dept of CS&SE, Andhra University, Visakhapatnam, INDIA

E-mail : [1]msamantu@gitam.edu , [2]sgorla@gitam.edu , [3]lalithabhaskari@yahoo.co.in

ABSTRACT

Face recognition has attracted greater attention during the past few decades, there were various algorithms that were proposed for the task and many generate accurate results. Our work initiated with a question – can we identify a person when seen from behind? This led to the idea of back pose human recognition (BPGAN). In our work, we introduced a novel approach (BPGAN) which involves generating a person's frontal image when the system is feed with images that were captured from behind and images taken partially like when the person is positioned at different angles like facing right and facing left. This approach enables the AI based system to perform the task of recognition of a person even when seen from behind, where it lacks the data to recognize the individual from frontal data and has some back pose images of the individual. In this paper we are trying to mimic the human behavior of identifying the person when seen from behind relying on the body shape and his historical knowledge of that particular person's facial information. This algorithm helps in recognizing the individuals in the crime scenes where the images are blurred or images from behind are available.

Keywords: *Backpose, Pix2pix, GAN, CNN, Human recognition.*

1. INTRODUCTION

The human recognition has played a very vital role in the field of security and other applications which includes various methods of human recognition that includes but not limited to face, fingerprint, iris and also base on gait recognition. As it is well known that there were problems associated with face and iris recognition methods that pertain to quality of the images. Moreover, these techniques may not be reliable on human recognitions and on partial data or when the images are occluded where the original images are damaged because of motion blur or optical blur that are a result of long-distance shots. The human recognition significantly fails its purpose dues to these types of images. The person recognition based on human body can be broadly classified into two types: gait recognition and texture and shape-based body recognition. The basic idea was here to identify humans from their about-facing view by presenting images of humans mostly male from behind (about-facing)[25,14]. Humans are generally very good at recognizing faces. Even when it comes to recognizing the know person even from far away distance, like one can spot a recognizable face standing far away and talking with other people or when seen crossing the street. Although the face can be seen, it lacks the clarity needed to clearly discern the features and facial expressions that are essential for identification. In the aforementioned circumstances, the body is largely visible and provides the data needed for identification. Since body features are given lower importance than facial features from a visual standpoint, they may occasionally be able perform in a more effective recognition system when compared to facial features in analyzing the person far away from viewpoint. As it was seen Image generation for recognition has given more importance in the past few years, considering the works done [1-3], which involves the synthesizing the facial data from various angles while retaining the identity has attracted major task, as it has gained vast importance in the industrial applications such as face analysis and video analysis stating from crime analysis to other various applications. In current research domains this task has been greatly improvised by various models including

Generative Adversarial Networks (GAN)[18]. Additionally, in [15], the researchers discovered that by learning the entirely nonlinear model, Graph Convolutional Neural Networks and Generative Adversarial Networks can successfully rebuild quality face form and texture, respectively. Non-Euclidean structures like graphs can be directly convolved using CNN images, which can both efficiently acquire crucial information from the edge which allows to lower the computational complexity of the image. These qualities have led to its widespread application recently in 3D face datasets [18, 19] and mesh datasets [16, 17]. In the meantime, GAN has recently demonstrated superhuman abilities in textures and structural characteristics. However, there are still some issues with accurately conveying positive outcomes.

In the proposed technique BPGAN which uses the methodology of Pix2Pix translation with conditional adversarial networks that was proposed by Isola et al. (2017), discusses basically about the approach of learning process by mapping from input images to output images. The Pix2Pix GAN [37], is a method suitable for a deep CNN involving image to image translation problems. As pix2pixGAN model does not fit into application specific this model can be applied to a various task like label maps, generating colorful photos from black and white images, transforming sketches into photos. The production of the images are mostly dependent on the source input, that is mostly the real image captured and used in a GAN models. After obtaining both the Real and Fake images, the discriminator is expected to analyze whether the obtained image is a natural transformation of the source image.

Whereas in Generator phase, it is forced to provide real images of the target domain using adversarial loss training. Generator also computes the L1 loss which is obtained as a result of differentiation between the fake image and the expected real image. When loss generated is more it encourages the generator model to generate correct translations of the real image. Numerous face recognition techniques in use today assume the existence of frontal faces with identical face sizes. Because faces might differ in appearance and because of external influences, this premise might not hold true in actual life. Considering the possibilities of using these methods we have initiated the experiment of identifying the humans when seen from behind. Here we are trying to exhibit the human natural phenomenon of identifying the individuals when seen from behind. As a natural capability of humans, we are good enough to identify the person when seen from back at a reasonable distance where we lack his frontal viewpoint. Here it's an attempt to mimic that behavior of human by training the machine to attain the identification capability even when seen from behind.

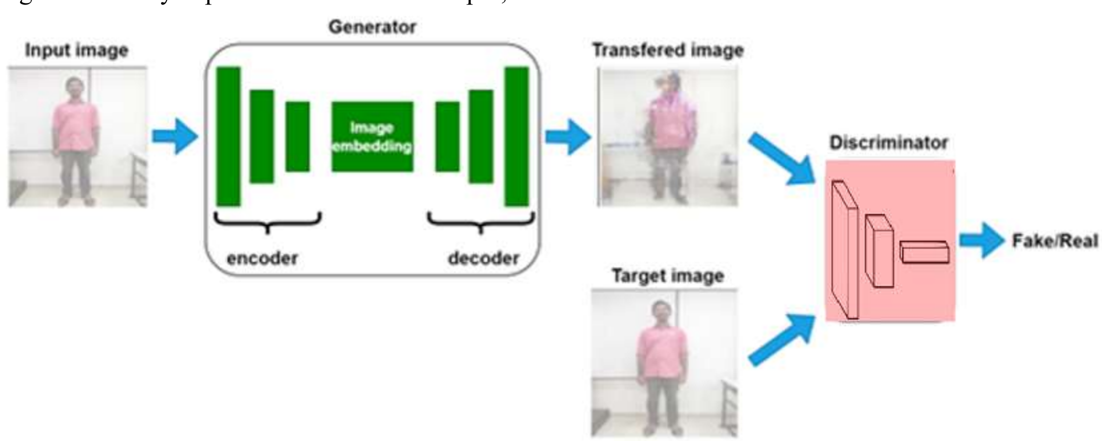


Figure. 1 Flowchart For Analyzing Images Through GAN

we have come with a novel systematic approach to recognize a person based on his back pose, the proposed system is trained with the Front Image of an individual using pixel to pixel matching. Initially a real image is given to the generator of pix2pix which produces a fake image, now the combination of fake image and real image as a pair and real image and real fake are given to the discriminator to classify the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

generated is fake or real based on the discriminator loss. In this process Leaky ReLU and U-Net architectures are used to optimize, which further aids the model's attainment of high fidelity and accurate human recognition. We provide an overview of the architecture, containing both the discriminator and generator models. The discriminator which is considered to be a deep convolutional neural network that carries out picture categorization. specifically, classification of conditional images. It predicts the accuracy of the target image which is a true translation of the source image or a fake one based on the input of both the source image (behind person image) and the target image (front person image).

2. RELATED WORKS:

Recognition of humans by their faces had significant importance for past few decades, to achieve this traditional image processing techniques have been used vastly where it involves mathematical transformation between the pixels on the available information. Given the information without any noise- missing data or pixel data, it was convenient to recognize, there were various methods that handled this missing data where the recognition would be possible even with the partial data ie using the partial facial recognition systems where it is possible to restore the image with the partial facial input data. Our idea is totally quite unique where we are attempting to retrieve the image with no facial information that is recognizing the person when seen from behind. Earlier attempts have been observed but none have been found in the process of retrieving or identifying the person when the input is given as back pose pictures. The study in [1] showed how even when people aren't aware that their bodies might be used for identification, they can still be a vital factor in recognizing people. Here, the authors pointed out that psychological research has mostly concentrated on the function of the body in social communication. Eye movement patterns were observed when an individual is being recognized using the frontal data and when seen without proper frontal data. Here it was also experimented with results obtained with face only images and body only image the human identification process.

2.1 Person recognition based on Posture of the body

In [2] authors discussed about the procedure that are used for person recognition based on body shape and motion, where body information that can be used for authenticating has remained an unanswered puzzle to one understanding of how humans identify and interact socially with the individuals in the real world. The current study shows that, as shown in figure 2, processing of human identity is not entirely limited by the substantial dependence on the face for identification.

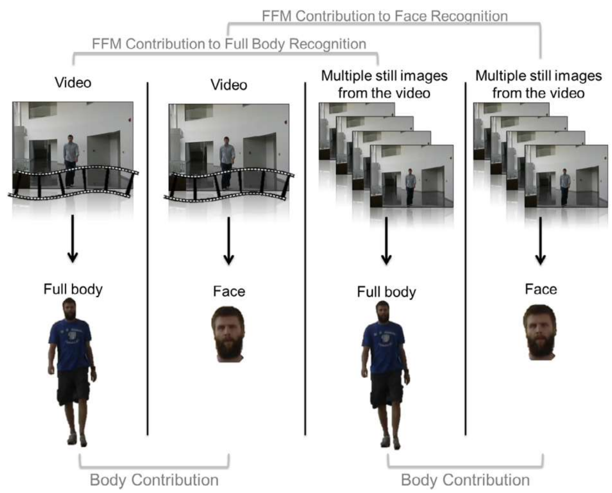


Figure. 2 Motion Process Performance Comparison Between Video And Still Images. (Noa Simhi,2015)

In [33] the authors have reviewed the behavioral, electrophysiological and neurofunctional research on the perception of whole bodies and physical expressions regarding the face perception in current focus of their study. They have gone over every piece of data that is currently available in greater depth than has been done previously, but they also made an argument involving the importance of comparison of faces and bodies that takes into account wider issues than just the category-specific modularity of either faces or bodies. Similarly, in [19], authors developed a single model that works on both human posture and shape of the body. This model made it feasible to mimic muscle deformations effectively both as a function of position and as a factor of the subject's physical features. The model's capacity to produce random human body shapes helps greatly in the creation of highly lifelike character animations. To create a learning-based technique, 114 volunteers' full bodies were scanned using a 3D laser for about 550 scans.

2.2 Person recognition from body shape and gait

In works that we have gone through mostly its been a process of identification of human either through facial information or with the partial information where the dataset contains images of headless images or bodyless images, the back posed images are not identified in any of the related works. In [4] human recognition is discussed based on the shape of the body, The CAESAR (Civilian American and European Surface Anthropometry Resource) initiative had almost gathered 5000 people's traditional measurements as well as 3D scans with 73 anthropometry landmarks. a standard methodology to recognize people based on their body type and gait was proposed by the authors in [5]. Based on methods available for body shapes matching, a viewpoint dependent method is provided. Here in this approach using cyclic gait analysis, crucial frames are extracted that can be used for a test sequence. Closest neighbor matching among correlation scores is used to determine the subjects by employing standardized correlation to compare these frames to training frames.. While in [7][12] the authors describe an intelligent system strategy for distant human identification based on human body shape information. The head, shoulder, and trunk are the physical features that are utilized. This work established image processing approaches for detecting these bodily traits. The features are then detected using a fuzzy logic approach and fed into a multilayer neural network-based recognition system.

While in [8], data that is collected from biometric sources like gait and side face is used and combined at the feature extraction level. From numerous video frames, a side face image of the face in high quality is produced. The physical aspects of human walking are analyzed using the Gait Energy Image and a spatial-temporal compact description of gait in video. Facial traits and gait features are separately retrieved from side angled face images and Gait Energy Image, respectively, using a techniques like Principal Component Analysis and Multiple Discriminant Analysis.

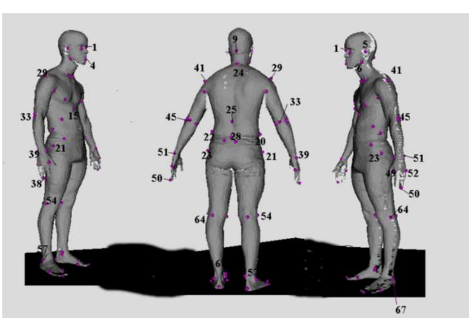


Figure. 3 Caesar Body Feature Representations.(Afzal Godil,2003)

2.3 Human identification from gaze distribution

In [6] an attempt is made to comprehend the distribution of gaze in naturalistic scenarios and how the preferential-looking paradigm was used. In the paper couple of scenarios were shown parallelly, one with a single person present (person-present), and the other without (person-absent). Furthermore, the quantity of time spent examining each part of the scenery during each trial, as well as the duration and orientation of the first fixation, were observed. They looked at eye movements while viewing freely and while completing a task that required them to differentiate between the genders of the shown human figure. A significant preference for glancing at the person in the scene was found, according to the research. They employed stimuli in this study that depicts two situations, in which it had a person in it. Major goal was to monitor random patterns to estimate whether complex visual objects, like people, might attract attention without being a search target. In [26] the authors proposed a method called GazeEMD that may be utilized for HRI applications to determine whether a person is looking at an object or not. We quantify the similarity between the simulated and real gazes at objects using the Earth Mover's Distance (EMD). The similarity score is then applied to ascertain whether the object is receiving the human eye's attention.

3. DATASET DESCRIPTION

In this work , we have collected image of humans reflecting different angles, which can be used to train the deep neural model. In [4] The CAESAR (Civilian American and European Surface Anthropometry Resource dataset have been used for recognition based on body position, here we are proposing a novel method to identify the human from the back posed images where full-frontal information is unavailable that can be used by recognition algorithm. In this novel approach we gather our own dataset by taking pictures of various individual depicting different viewpoints that can be used by the algorithm to properly solve the problem of identifying the person based on back pose.

Journal of Theoretical and Applied Information Technology
29th February 2024. Vol.102. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Front View *Right facing View* *Back pose View* *Left facing View*

Figure 4: Multiple Views Of An Individual

The dataset contains images of around 600 individuals captured from various different angles like front , Right, left and back views. These images were used for training the deep learning model for predicting the individual when the input is mostly given with back pose.

4. ARCHITECTURAL APPROACH OF BPGAN

As an approach towards attaining the capability of backpose recognition, we are proposing a novel approach for human recognition based on his back pose stature. The proposed method is coined as BPGAN (Back Pose Generative Adversarial Network) which works on the concept of Pix2Pix model. This proposed method it facilitates the ideology of identifying the person based on his back pose. Actually, we haven't come across any earlier approaches that describe this feature extraction where it helps in identifying the person based on the information of his back pose. Most of the approaches are either getting trained on the face features or body pose or from gait movements, but none have mentioned the detection of the human based on the back pose, this motivated us in developing a methodology that helps in identifying the humans based on the back pose data. Here we tried to mimic the human capability of identifying the known person when seen from backwards, taking the support of deep neural networks we are proposing this method which involves the concept of pix2pix generative adversarial networks. There are various GAN approaches like a) cycle GAN b) Conditional GAN c) Pix2Pix GAN

The Generative Adversarial Network (GAN) architecture is a basically a pack of (a) Generator and (b) Discriminator models. The discriminator basically classifies the images as original (from the dataset) or hoax (manufactured), with the support of generator phase particularly doing the task of creating the synthetic images. While the generator phase receives updates through the discriminator, the discriminator is now getting updating directly. Here it can be observed that, the two models are concurrently getting trained with adversarial strategy where the generator always tends to deceive the discriminator and discriminator strives to identify the fake images. Here in our approach of identifying the human when seen from behind using the back pose (BPGAN), Pix2Pix model of conditional GAN is used.

4.1 Pix2Pix GAN for Back Pose Detection

The Pix2Pix GAN a variant of conditional GAN (cGAN) proposed by Phillip Isola, which is basically used for picture-to-image translation tasks, in which the derived images that are treated as a conditional on a source images. The Discriminator have access to both source image and the target image as well, it makes use of both of these images to predict if the target is a genuine transformation of the source image. The non discriminator phase undergoes training through adversarial loss, which helps it to produce reliable outputs in the target domain. The L1 loss obtained as a result of the created image and the anticipated output image is used to update the generator as well. The generator model is motivated to produce accurate translations of the source image by this additional loss. In this paper we have used the Pix2Pix GAN [37] for Back Image of person to Front Image of a person as a translation task. In a nutshell the generator both uses real data and noise to learn the mapping.

$$G : \{X, Y\} \rightarrow Y \quad (1)$$

Similarly, In addition to real data, the discriminator also learns representation from labels.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$D\ (X, Y) \qquad (2)$$

4.2 Construction of a Pix2Pix Model

The discriminator and generator models made up the bulk of the pix2pix architecture. Deep convolutional neural networks are used as the discriminator to classify images. conditional-image classification specifically. It predicts whether the output image is likely to be a efficient translation of the Real image or a fake one by using both the source image (the back person image) and the target image (the front person image) as input.

a) The Discriminator Phase

1. Concatenate()([src_image, target_image]) discriminator takes source and target images that are concatenated together

2. The merged model should pass to step by step 6 layers Convolutional Neural network starting from 64 filters to expand multiplier of two up to 512 filters

   a. Sample of 1st Layer
   Ds = Con2D (64, (4, 4), strides= (2, 2), padding='same', kernel_initializer=init) (merged)
   Ds = BatchNormalization ()
   Ds = LeakyReLU (alpha = 0.2)

   b. So what's happening here?
      i. Here we have added 3 things Con2D, BatchNormalization and LeakyReLU
      ii. Conv2D that helps us process the picture and at initial we uses 64 filters which means the program can look for 64 different features or patterns in the image
      iii. (4, 4) This represents the size of each filter. It tells the program how many pixels wide and tall each filter should be to slide and check each patches
      iv. (2,2): This represents the stride of the filters. The stride tells the program how many steps the hands should take between each patch they look at. In this case, the filters will move 2 pixels horizontally and 2 pixels vertically.
      v. So, when the program runs the Conv2D operation, it will apply 64 different filters, each looking at a 4x4 patch of the picture, and moving 2 pixels at a time. This allows the program to capture different features and patterns in the image
      vi. Batch Normalization, the program adds the Normalizer to the neural network. The Normalizer looks at the information from the previous step, which is the picture "d" that was processed by Conv2D and LeakyReLU, and adjusts it so that the neurons in the network are balanced.
      vii. LeakyReLU(alpha=0.2), LeakyReLU looks at each part of the picture "d". If it's bright, it keeps it the same because it's already exciting. But if it's dark, it makes a small sound to show it's still there but not as excited.
      viii. The purpose of the alpha parameter is to control the amount of leakage and fine-tune the behavior of the neuron.
      ix. with an alpha value of 0.2, it means that when the Leaky ReLU receives a negative input, it will leak out 20% of that input as a small sound. So, instead of completely staying quiet, it will make a little noise that is 20% of the negative input value.

   c. patch = Activation('sigmoid')(d): In this line, the program applies an activation function called 'sigmoid' to the output from the previous step. The sigmoid function takes the output values and squashes them between 0 and 1. It's like a special transformation that helps us describe the output as probabilities. In this case, it's used to determine the likelihood of something being present or not.

   d. model = model ([src-image, tar-image], patch): Here, the program creates a model using the inputs 'src-image' and 'itarget_image' and the output 'patch'

   e. model.compile(loss='binary_crossentropy', optimizer="adam", loss_weights =[0.5]): This line compiles the model by specifying the loss function, optimizer, and loss weights. The 'binary_crossentropy' loss function compares the projected output to the anticipated output to see how well the model is performing. The optimizer, 'opt', is the algorithm that will adjust the model's parameters to minimize the loss.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The loss weights, '[0.5]', indicate the relative importance of different losses if there are multiple outputs or objectives

3. Return the Model

b) The Generator Phase

In Pix2pix GAN model the generator is generally represented encoder and decoder model which employs a U-Net design The model is designed to create a target image (such as the front image) from a source image (such as the back image). In order to achieve the above, the input images are first down sampled or encoded to a bottleneck layer, and then the bottleneck representation is up sampled or decoded to the size of the output image. While in U-Net architecture [38], skipped connections are implemented to create a U-shaped pattern between the encoding levels and the appropriate decoding layers.

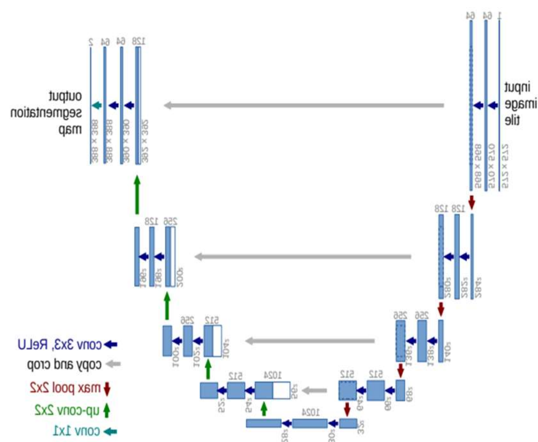


Figure 5: U-net architecture (example for 32x32 pixels)-(Olaf Ronneberger-2015)

Encoder which is a part of generator converts the input image given by the users into a smaller feature representation, and Decoder the other part of the generator, which is a series of transpose-convolution layers that is designed to reverse the actions of encoder layers. The discriminator[31,35], always works on pair of images that is input image and unknown image which can be either target or generated image, instead of identifying a single image as real or fake and output a label for pair as real or fake. . The loss function that is calculated by the discriminator is given by

$$\min_{G}\max_{D} \mathbb{E}_{x,y}\left[logD(x,G(x))\right] + \mathbb{E}_{x,y}\left[log(1 - D(G(x,y)))\right] \quad (3)$$

In pix2pix the generator model is trained through the discriminator model, whereas the discriminator model is trained directly on fake and real images. The process of updating will continue until it minimizes the loss predicted by the discriminator for the generated images. Therefore, it is suggested to produce more authentic real images. Simultaneously the generator also gets updated with the loss to minimize the L1 loss or mean absolute error (MAE) between the Fake and the target image. The adversarial loss and the L1 loss are weighted together to update the generator.

4.3 Generating Pix2PixGAN Model for Back Pose Detection

a) Defining BPGAN

1. Define Input Shape: insrc = Input(shape=image_shape)
2. Pass the src image to the generator and generate the fake output
    a. genout = g_model(in_src)
3. Pass the Source image and generator output to the discriminator to generate the discriminator output
    a. disout = d_model([insrc, genout])
4. The code creates a new model (model) that takes the source image (insrc) as input. The output of the model consists of two parts: the discriminator's classification output (disout) and the generated image (genout).
    a. model = Model(insrc, [disout, genout])
5. "compile model":
    a. The code compiles the model by specifying the loss functions, optimizer, and loss weights.
    b. The loss functions are specified as a list: 'binary_crossentropy' and 'mae'. The 'binary_crossentropy' loss measures the difference between the predicted and true labels for the discriminator's classification output.
    c. The 'MAE' loss (mean absolute error) measures the difference between the generated image and the target image.
    d. The optimizer used is Adam with a learning rate (lr) of 0.0002 and a beta_1 value of 0.5.
    e. The loss weights are set as [1, 100], indicating the relative

importance of the two losses in the overall loss calculation.

f. optm= Adam(lr=0.0002, beta_1=0.5)

g. model.compile(loss=['binary_cross entropy', 'mae'], optimizer=opt, loss_weights=[1,100]

6. Finally, the defined model is returned as the output of the function.

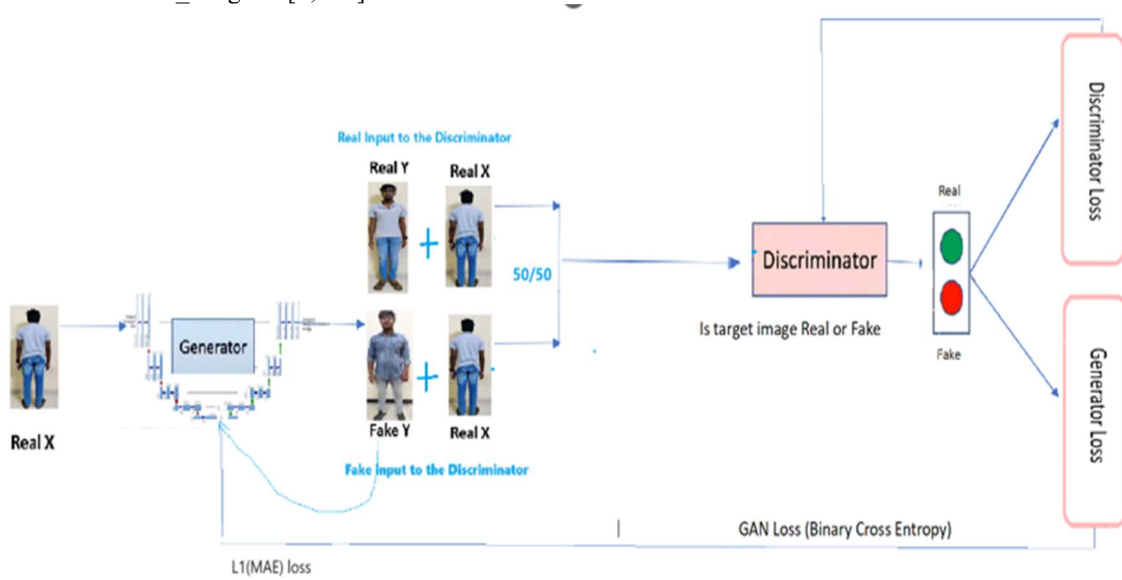


Figure. 6 Flowchart for back pose image analysis through Pix2PixGAN

4.3 Generating Pix2PixGAN Model for Back Pose Detection

a) Defining BPGAN

7. Define Input Shape: insrc = Input(shape=image_shape)
8. Pass the src image to the generator and generate the fake output
   a. genout = g_model(in_src)
9. Pass the Source image and generator output to the discriminator to generate the discriminator output
   a. disout = d_model([insrc, genout])
10. The code creates a new model (model) that takes the source image (insrc) as input. The output of the model consists of two parts: the discriminator's classification output (disout) and the generated image (genout).
    a. model = Model(insrc, [disout, genout])
11. "compile model":
    a. The code compiles the model by specifying the loss functions, optimizer, and loss weights.
    b. The loss functions are specified as a list: 'binary_crossentropy' and 'mae'. The 'binary_crossentropy' loss measures the difference between the predicted and true labels for the discriminator's classification output.
    c. The 'MAE' loss (mean absolute error) measures the difference between the generated image and the target image.
    d. The optimizer used is Adam with a learning rate (lr) of 0.0002 and a beta_1 value of 0.5.
    e. The loss weights are set as [1, 100], indicating the relative importance of the two losses in the overall loss calculation.
    f. optm= Adam(lr=0.0002, beta_1=0.5)
    g. model.compile(loss=['binary_cross entropy', 'mae'], optimizer=opt, loss_weights=[1,100]

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

12. Finally, the defined model is returned as the output of the function.

b) BPGAN Module Training

1. generate real samples from dataset
    a. [BACK, FRONT], LABEL_REAL = generate_real_samples(dataset, n_batch, n_patch)
2. generate fake samples using gan model predict function
    a. FAKE_FRONT, LABEL_FAKE = generate_fake_samples(g_model, X_realA, n_patch)
3. Train the discriminator with real output and calculate d1 Loss
    a. d_model.train_on_batch([BACK, FRONT], LABEL_REAL)
4. Train the discriminator with generated fake output calculate d2 Loss
    a. d_model.train_on_batch([BACK, FAKE_FRONT], LABEL_FAKE)
5. Train BPGAN model with Real Image and get the Gan Loss
    a. gan_model.train_on_batch(BACK, [LABEL_REAL, FRONT])
6. Plot the output between FAKE_FRONT and FRONT
7. Repeat these steps till number of epochs batch size specified.

5. RESULTS- ANALYSIS:

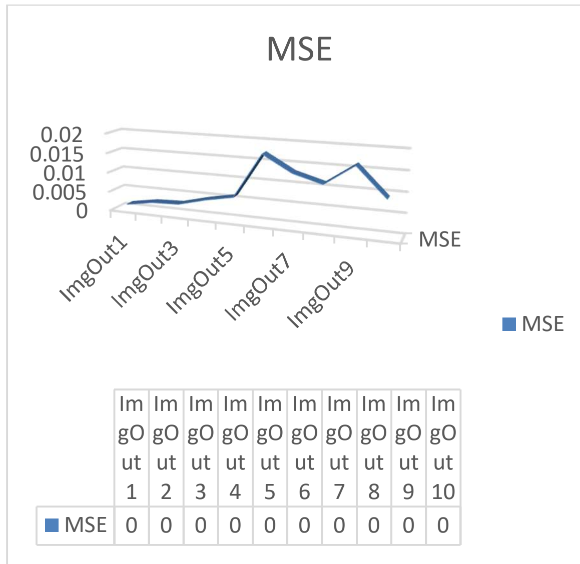
In result analysis we projected the evaluation of the proposed model BPGAN , where it has shown decent achievement of identifying a person from back pose when provided with the frontal image information as an input, here in our model we have used the variant to Generative Adversarial network model, Pix2PixGan which is conditional GAN , that uses the image to image translation.

5.1 Evaluation Metrics

Here, we show the findings from BPGAN's evaluation of the back pose identification using the structural similarity index and mean square error , two evaluation indicators. These are the most common metrics used to gauge image quality.

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[y(i,j)-x(i,j)\right]^2 \qquad (4)$$

In the above equation x indicates the generated image, Y denotes the back posture image, and m n denotes the image size. The cumulative squared error between the original and compressed images is represented by the term "MSE," or mean square error, between x and y. A greater disparity between the original image and the fake is indicated by a higher MSE score.



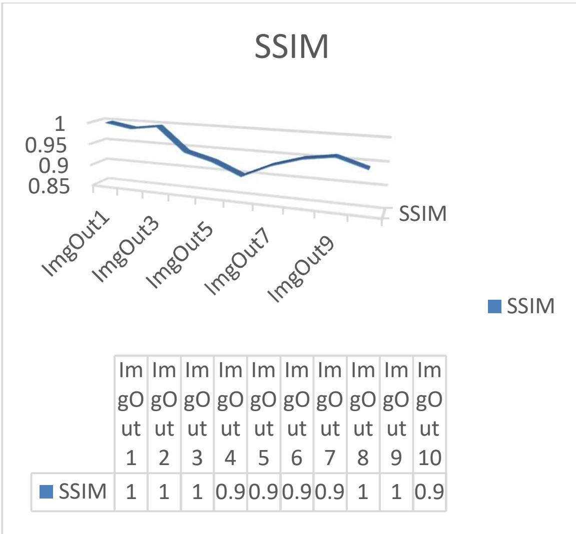
Mean Square distribution among the Outputs

SSIM is used to assess the similarity of two photographs. The SSIM is denoted as

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \qquad (5)$$

The parameters μx,μy are represented as the means of x and y, $\sigma^2 x$ and $\sigma^2 y$ are the representation of variances of x and y, σ_{xy} is the covariance between x and y, $c_1 = (k_1L)^2$ and $c_2 = (k_2L)^2$ are the constants used to maintain stability and $k_1 = 0.01$ and $k_2 = 0.03$ are the default values. The value range of SSIM is in-between 0 and 1. It is been observed that when the SSIM value is closer to 1, the probability of two images being similar is high.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



Structural Similarity Index Distribution Among The Outputs

5.2 GAN Loss Functions:

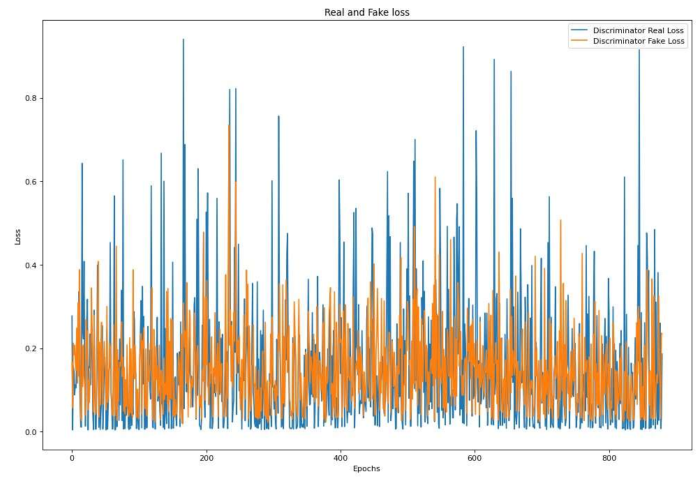
A GAN basically contains two loss functions: One loss function for generator training and the other for discriminator training. As the discriminator is trained, it attains the capability of classifying both the original data and the fake data from the generator. By optimizing the following function, it degrades itself for classifying a real entity as fake or a fake instance (produced by the generator) as real.

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(x^{(i)}\right) + \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right) \right] \quad (6)$$

From the above equation it's been observed that optimizing the value obtained from log(1-D(G(z)) allows the generator to consistently label the fake image which are a result of the generator. Log(D(x)) represents the correctness of the generator in classifying the genuine images.

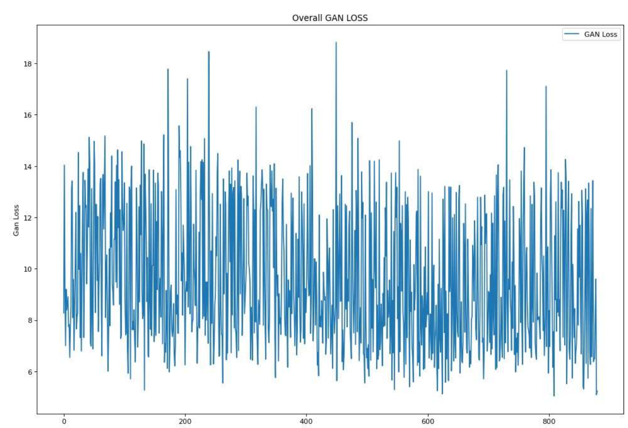
In generator training process a noise is sampled to generate the outputs. The output generated is given as an input to the discriminator, where the determination of Real" or Fake" is evaluated. The loss obtained in generator phase is subsequently analysed using the discriminator's categorization, if output generated deceives the discriminator, it is rewarded; otherwise, it is penalized. In order to attain the minization in the generator phase the below equation is considered.

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right) \quad (7)$$



Graph Depicting Discriminator Real Loss Vs Fake Loss

In the above line graph a clear differentiation between discriminator real loss vs fake loss is shown as the number of epochs are been increased. The GAN loss is observed to be around 6% which is typically the discriminators output that was generated for the samples.



Graph Depicting Overall GAN Loss

6. EXPERIMENTAL RESULTS:

Output images of Various individuals

Mean-Squared-Error (MSE)	Structural-Similarity- Index (SSI)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

	0.0013912517471839747	0.9995006340132713
	0.0025930191802267643	0.9890818866161037
	0.002883947254181658	0.9976506656291382
	0.0047	0.9443
	0.0061	0.9279
	0.0173	0.9010
	0.0133	0.9310
	0.0113	0.9510

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

	0.0163	0.9610
	0.0092	0.9410

Sample results generated for Person Recognition based on Back Pose

7. CONCLUSION

We presented a novel recognition model BPGAN for back pose-based detection of an individual using Pix2pixGAN which is a benchmark approach image to image conversion tasks. A novel model has been proposed for human recognition with backpose information. In this paper we are successful in mimicking the human behaviour of identifying a person when seen from behind. In order to exhibit this behavior in AI based machines we trained the model with Pix2Pix GAN for image-to-image matching which is between Back Image of person to that of Front Image of a person. However, processing data containing images that are occult, image with partial facial information still presents some challenges for the current image identification system in practical applications. The pix2pixGAN network architecture serves as the foundation in achieving the task with overall GAN loss of 6%. According to the findings of our experiments, proposed technique is effective at identifying individuals based on their back posture. Though Only male photographs are trained in this model, we are also in the process of adapting this article's strategy to work for other genders.

REFERENCES:

[1] Rice et al," Unaware Person Recognition from the Body When Face Identification Fails", Psychological Science, Vol. 24, No. 11 (NOVEMBER 2013), pp. 2235-2243.

[2] Noa Simhi , Galit Yovel, "The contribution of the body and motion to whole person recognition", Vision Research 122 (2016) 12–20.

[3] Axelsson et al. "Body Inversion Effects with Photographic Images of Body Postures: Is It about Faces?", Frontiers in Psychology, November 2019, Volume 10, Article 2686.

[4] Afzal Godil, Patrick Grother and Sandy Ressler "Human identification from body shape",Fourth International Conference on 3-D Digital Imaging and Modeling, 2003.3DIM 2003. Proceedings. IEEE.

[5] Robert T Collins, Ralph Gross and Jianbo Shi," Silhouette-based Human Identification from Body Shape and Gait", Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition 0-7695-1602-5/02.

[6] S Fletcher-Watson, J M Findlay, S R Leekam, V Benson," Rapid detection of person information in a naturalistic scene" Perception, 2008, volume 37, pages 571-583.

[7] N K A M Rashid, M F Yahya and A Shafie," Human Identification at a Distance Using Body Shape Information" , IOP Conf. Series: Materials Science and Engineering 538(2013) 012058

[8] Xiaoli Zhou and Bir Bhanu," Feature Fusion of Face and Gait for Human Recognition at a Distance in Video" , The 18th International Conference on Pattern Recognition (ICPR'06) 0-7695-2521-0/06,2006..

[9] Minh-Tuan Nguyen, Guo-Shiang Lin "A Vision-based Human Identification Method',2016 International Conference on Consumer Electronics-Taiwan.

[10] https://www.palizafzar.com/en/articles/human-identification-from-body-shape/

[11] Leonid Pishchulin Stefanie Wuhrer, Thomas Helten, Christian Theobalt, Bernt Schiele,"Building statistical shape spaces for 3D human modeling, Pattern Recognition, Volumec67, July 2017, Pages 276-286

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[12] Seung-Yeob BaekKunwoo Lee," Parametric human body shape modeling framework for human-centered product design", Journal of Computer-Aided Design, Volume 44, Issue 1,January 2012, Pages 56-67.

[13] Brett Allen Brian Curless Zoran Popovic," The space of human body shapes:reconstruction and parameterization from range scans" , ACM Transactions on Graphics,Volume 22,Issue 3July 2003 pp 587–594.

[14] Brett Allen Brian Curless Zoran Popovic," Articulated Body Deformation from Range Scan Data", ACM Transactions on Graphics, Volume 21, Issue 3July 2002 pp 612–619.

[15] Sungmin Kim and Chang Kyu Park," Parametric Body Model Generation for Garment Drape Simulation", Fibers and Polymers,5(1)-2004, pp 12-18.

[16] Azuola, F., Badler, N. I., Ho, P., Kakadiaris, I., Metaxas, D., & Ting, B. (1994). Building Anthropometry- Based Virtual Human Models, 1994 IEEE. Reprinted from Proceedings of the IMAGE VII Conference, June 1994.

[17]Kasap M, Magnenat-Thalmann N. Parameterized human body model for real-time applications. In: Proceedings of the 2007 international conference on cyberworlds, 2007,pp. 160–167.

[18] A. Hilton, D. Beresford, T. Gentils, R. Smith, W. Sun, J. Illingworth Whole-body modeling of people from multi-view images to populate virtual worlds Visual Computer,16 (7) (2000), pp. 411-436.

[19] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, H.-P. Seidel,"A statistical model of human pose and body shape Computer Graphics Forum, 2 (28) (2009), pp. 337-346.

[20] N. Hasler, C. Stoll, B. Rosenhahn, T. Thormählen, H.-P. Seidel, "Estimating body shape of dressed humans".Comput. Graph., 33 (3) (2009), pp. 211-216.

[21] O. van Kaick, H. Zhang, G. Hamarneh, D. Cohen-Or, "A survey on shape correspondence" Computer Graphics forum Volume 30 (2011), number 6 pp. 1681–1707

[22] H. Seo, N. Magnenat-Thalmann "An automatic modeling of human bodies from sizing Parameters" I3D '03: Proceedings of the 2003 symposium on Interactive 3D graphicsApril 2003Pages 19–26

[23] Shih-Chung Hsu ,Yu-Wen Wang, Chung-Lin Huang," Human Object Identification for Human-Robot Interaction by using Fast R-CNN, 2018 Second IEEE International Conference on Robotic Computing.

[24] Sue Fletcher-Watson, John M Findlay, Susan R Leekam, Valerie Benson," Rapid detection of person information in a naturalistic scene", Perception, 2008, volume 37,pages 571 – 583.

[25] Zhanfeng Yue, Rama Chellappa," Synthesis of Silhouettes and Visual Hull Reconstruction for Articulated Humans", IEEE Transactions on Multimedia, VOL. 10, NO. 8, DECEMBER 2008.

[26] Shari Matzner, Evelyn J. Boettcher, Darrell Lochtefeld and Timothy Webb," Standoff Human Identification Using Body Shape 2015 IEEE International Symposium on Technologies for Homeland Security (HST),2015

[27] Minh-Tuan Nguyen, Guo-Shiang Lin," A Vision-based Human Identification Method", International Conference on Consumer Electronics-Taiwan 2016 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW) 2016.

[28] P. Nangtin, P. Kumhom, and K. Chamnongthai," Human Identification by using Human Part Similarity", Proceedings of 2007 International Symposium on Intelligent Signal Processing and Communication Systems Nov.28-Dec.1, 2007 Xiamen, China.

[29] Beatrice de Gelder , Jan Van den Stock , Hanneke K.M. Meeren , Charlotte B.A. Sinke ,Mariska E. Kret , Marco Tamietto," Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions, Neuroscience and Biobehavioral Reviews 34 (2010) 513–527.

[30] Isaac COHEN, Hongxia LI," Inference of Human Postures by Classification of 3D Human Body Shape", Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03).

[31] Markus Bindemann, Christoph Scheepers, Heather J. Ferguson, A. Mike Burton," Face,Body, and Center of Gravity Mediate Person Detection in Natural Scenes, Journal of Experimental Psychology: Human Perception and Performance,2010, Vol. 36, No. 6,1477–1485.

[32] Zhenning Li and Dana Kuli," Local Shape Context Based Real-time Endpoint Body Part Detection and Identification from Depth Images", IEEE Xplore,978-0-7695-4362-8/11.

[33] Cullen D. Jackson, Michael J. Tarr, Athinodoros Georghiades, "Identifying Faces Across Variations in Lighting: Psychophysics and Computation". ACM Transactions on Applied

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Perception Volume 5Issue 2Article No.: 10pp 1–25

[34] A.Mike Burton, Vicki Bruce and J.B.Hancock, From Pixels to People: A Model of Familiar Face Recognition", COGNITIVE SCIENCE Vol23 (1) 1999, pp. l-31.

[35] Jasvinder Pal Singh, Sanjeev Jain," Person Identification Based on Gait using Dynamic Body Parameters", IEEE conference in Trendz in Information Sciences & Computing(TISC2010),2010

[36] Celia Foster and Mintao Zhao," Separated and overlapping neural coding of face and Body Identity" WILEY publications Hum Brain Mapp. 42: 4242–4260, 2021.

[37] Phillip Isola Jun-Yan Zhu Tinghui Zhou Alexei A. Efros," Image-to-Image Translation with Conditional Adversarial Networks", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017.

[38] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015.