International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.139

Deep Neural Network-Based Generative Models to improve Surveillance Techniques

Pranav B Kashyap
Department of Data Science and Business Systems
SRM Institute of Science and Technology,
Kattangulathur, Chennai, India
pk6859@srmist.edu.in

S Suchitra
Department of Data Science and Business Systems
SRM Institute of Science and Technology,
KattangulathurChennai, India
suchitrs@srmist.edu.in

***Abstract*—Video Surveillance is the process which involves a network of cameras, recorders, and monitors to observe a scene or look for something distinct. It is an effective mechanism that helps counter crime, secures the area, helps cut down security costs, and provides fool-proof security coverage. Computer Vision (for Facial Recognition) plays a significant role in these surveillance systems. They help identify and recognise potential threats. However, one of the significant drawbacks of these face recognition systems is that they can only detect the person if the image is precise or at a particular angle. We propose an image modification technique using Generative Adversarial Networks (GANs) to overcome this. GANs are a generative modelling technique that generates new samples (images, videos). GANscan not only help generate very realistic face images but also modify the existing features of an image. The motivation of the proposed method is to identify the subject by modifying images obtained from the surveillance feed.**

***Index Terms*—Generative Adversarial Networks, Deep Learn- ing, Object Detection, Computer Vision**

I. INTRODUCTION

The Crime Rate in India is slowly but steadily increasing. According to the National Crime Records Bureau (NCRB),the IPC crime rate in India has been steadily growing at a rate of 3.4% over the last 15 years [Bureau,].

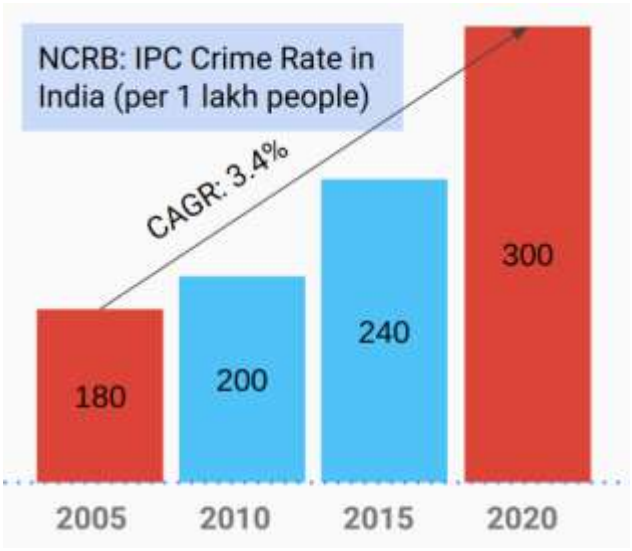

Fig.1.SteadyRiseinCrimeRate

Thisincreaseinthecrimerateisallhappeningamidsttheattemptstocurbit,includinginstallingsecurityandsurveillance systems.VideoSurveillanceisacommontechniquetocounter crime and monitor day-to-day activities. Nevertheless, they have their fair share of drawbacks. The surveillance systems installed in the country are not uniform. According to the report published by Comparitech, almost 92% of all CCTV cameras are installed in only 4 Indian states. The crime rate could be curbed with the larger volume of CCTV cameras. However, that is not the case, as there is no correlation between the number of CCTV camera sand the crime rate. Forinstance, one can see from Fig 2. that Delhi has the 3rd most significant number of CCTV cameras per 1000 persons, but the crimerate is the highest. We propose to enhance the quality of currently existing surveillance techniques with the help of our face modification and generation idea.

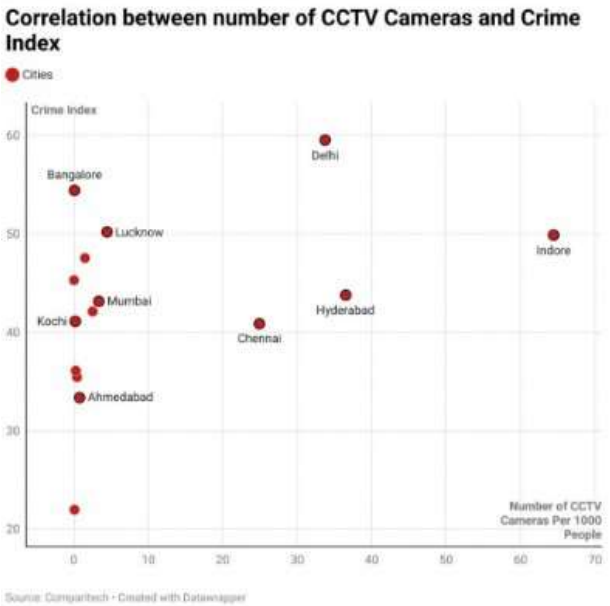

Fig.2.Correlation between crime index and number of CCTV cameras

With the growth of AI, surveillance techniques have only improved with time. AI uses software programs to analysethe audio, video or image files obtained from the surveillance feed.Oneofthemorecommonusecasesisfacedetectionorface recognition. Automating the process of recognising faces helpsimprovesecurityandautomatedidentification.However, it also comes with its fair share of drawbacks. One of the commondrawbacksisthatthemodelcaneasilybethrown off by different angles or poses, leading to wrong predictions. Thesolutionistomodifyimagesobtainedfromthefeedusing GANs. With the help of object detection techniques of the YOLOalgorithm,onecandetectandobtainimagesofpeople's heads. That can then be fed into the GAN network, which will regenerate/modify the face such that the frontal view of the face is obtained. This way, the surveillance system will detect the person, irrespective of their posture.

Tosummarise,thesearethemaincontributionsofthepaper:

- Auto-detecthumanbeingsinavideofeedusingtheobject detection techniques
- Modify images of the face/head to obtain a completeview of the face from the partial profile obtained in the surveillance feed using Generative modelling.

II. RELATEDWORK

GenerativeAdversarialNetworks(GAN)[Goodfellow et al., 2014] They are used for generative modelling using Deep Neural Networks. They help create (generate) new data instances that resemble the training data. A GAN network generally comprises two models: a generator model (used to generate new examples/samples) and a discriminator model (used to classify whether the generated images are real or fake). During the training phase, the generator produces fake data,whichthediscriminatorrecognisesandflagsasfake.This process (training) continues until the discriminator cannot dif- ferentiate between real and simulated data. The generator and discriminator are Deep neural networks. The generator output isconnectedtothediscriminatorinput.Differentvariations of the GAN algorithm include CycleGAN [Zhu et al., 2020], DCGAN [Radford et al., 2016], StyleGAN [Karras et al., 2019], etc. ESRGAN [Wang et al., 2018] enhances the quality of the image obtained from the surveillance feed before the image translation.

*ImageTranslation*Pix2PixGAN[Isolaetal.,2017]is a Conditional GAN (cGAN) [Mirza and Osindero, 2014] known for its image mapping and translational capabilities. The difference is that the input initially fed to the generator is notarandomvectorfromthelatentspace.Inthisscenario,the input image is the original image of the person's face profile that needs modification. The source of randomness for the generator comes from its dropout layers. Unlike the generator in a normal GAN, the generator follows a U-net architecture, with skip connections between layers of the same size, which helps in the image translation process. The traditional GAN model uses a Deep Convolutional Neural Network (DCNN)as a discriminator to classify images as real or fake. Here, a PatchGAN is used. Patch GAN is a DCNN that will classify patches of the image (as fake or real) rather than the whole image.

Object Detection: YOLO [Redmon et al., 2016] is awidelyusedalgorithmforobjectdetectionbysolvingitasaregre ssionproblem.YOLOv5[Jocheretal.,2022]helpsdetect faces/heads in a video or image feed. The detected images are then fed to the GAN model, which will help reconstruct the original face.

A. *ExistingSystems*

Kumar, VD Ambeth, et al. [V D et al., 2018] proposed a facialrecognitionsystemmodel.Herethesurveillancecamera is connected to a central server with the suspect database. The system will be alerted when the cameras have found a person whoexistsinthedatabase.Oneofthemajordrawbacksisthat the model used here is localised. The model will struggle to give accurate resultswhen a new face is givenas input (which is the real-world scenario). We can solve this issue with the help of Generative Learning.

Wang, Xin, et al. [Wang et al., 2023] proposed a face de- tection technique using images synthesised using GANs. The significant drawback is that complete face input is required to generate faces. It does not help us regenerate the whole face from a face profile obtained from the surveillance camera.

III. PROPOSEDMETHOD

A. *OverallArchitecture*

The objective of the paper is to improve and automate current surveillance techniques. We propose an ensemble techniquetoidentifypotentialsuspectsusingComputerVision andGANs.Ourapproachwillidentifyfacialimagesfrom a surveillance recording and modify them accordingly. With the help of GANs, we can regenerate the whole face from whateverisvisibleinthevideofeed.Theprocessforthesame is shown in the diagram below.

Fig.3.ProposedArchitecture

First, we will feed the video, with a potential suspect, as input to our model. With the help of the YOLO algorithm, we will identify all the people in the image. These images will then be cropped out and saved in another folder (potential suspect list). Next, we will pass all these images through an image enhancer. Image enhancer is a basic implementation of ESRGAN. ESRGANs help improve image quality, whose output will eventually feed to the face-generating model. The face-generation model will then input the head angle obtained fromthefeedandusetheGANtogeneratetheentireface for the given person. We have divided the procedure intothree main modules; Pre-Processing (Footage Analysis and Image Extraction), Image Quality Enhancement, and Image Generation/Modification.

B. *Pre-Processing*

YOLOv5helpsintheprocessofobjectdetection.Generally, whenweneedtoidentifyapersonfromavideofeed,weman- ually look through the video until we recognise the subject. We can use the YOLO algorithm's object detection technique to simplify and automate this process. Object detection helps identify and locate a particular object in images or videos.The YOLO algorithm takes a given image/video as input and uses a deep neural network to detect the objects present in the same.

DOI: rp-9788770040723.139

Fig. 4. Detect Human face

Unlike the previous versions of YOLO, YOLOv5 uses EfficientNet architecture (based on the Efficient Net Architecture [Tan and Le, 2020]), allowing the network to generalise better to a broader range of objects. We have trained the YOLO model on a dataset of over 2000 images of a room with people. The model is trained to identify all the heads in the room.

C. Image Quality Enhancement

The image we obtain from the feed might need to be of better quality. It is essential to enhance the quality of the image and improve its resolution before feeding it to the modifying network. SRGANs help achieve the same. They are a generative network used for image super-resolution. The SRGAN uses a perpetual loss function, a weighted sum of a content loss and adversarial loss. The adversarial loss pushes the solution to look similar to the natural image via the discriminator network, which is trained to differentiate between the original and enhanced images. In contrast, the content loss helps identify the pixel-wise error between the original image and the generated image.

D. Image Generation

After pre-processing the video and obtaining a limited image of the potential suspect, we will use a GAN to regenerate the whole face. Pix2pix GAN, used for Image-to-Image translation, helps generate a full-face image.

Pix2Pix is a conditional GAN (cGAN), where the generation of the output image is dependent on an input image. The discriminator, in this case, is provided with a source image and target image, and its role is to determine whether the translation is possible.

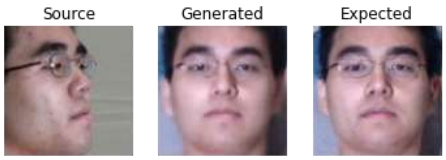

Fig. 5. Sample Model Output

IV. EXPERIMENTS

A. Dataset

Since we are working with different models, we had to train them on different datasets.

1) ***Image Detection****:*For the head detection problem using YOLO,weusedtheSCUT-HEADDataset[Pengetal.,2018]. Thisdatasetconsistsofimagesobtainedfromthemonitorfeed of a school classroom. The YOLO model is trained on this dataset and helps detect or find human heads in any public space.

TABLE IDATASET DETAILS

Task	Dataset	Data InstancesTrained	Epochs Trained	Model/Concept
Object Detection	SCUT-HEAD	2000	200	YOLO Algorithm forObjectDetection
ImageTrans-lation	Multi-PIE	3000	30000	GANsforImage Generation

2) Image Generation: Here, we are trying to generate the complete face profile from whatever view of the face is seen in the feed. So we used the CMU Multi-PIE dataset [Gross et al., 2008]. This dataset consists of face images from 337 subjects, all taken under different poses, expressions and illumination. The GAN is trained on this model. The aim is to generate the complete face of any subject when given a particular posture of that subject.

B. Results

1) Object Detection: We tested the models working with different scenarios. These scenarios were all defined by the number of people in a particular class. As you can see, the model gives us excellent results for all the scenarios. The table below summarises the results.

TABLE IIOBJECT DETECTION RESULTS

Scenario	No. of people	Predicted	Accuracy
Few People	3	3	100
Many People	11	10	90.90

Fig. 6. Object Detection when there are fewer people

Fig. 7. Object Detection when there are a decent number of people

1) Image Translation: Here we tested the model's working with the following different facial profiles:

- People with beards and facial hair
- Different Genders and Age Groups
- Only half the face is visible
- Identifying glasses and frames

DOI: rp-9788770040723.139

V. CONCLUSION

Thus, we could utilise the capabilities of computer vision and generative modelling to solve a real-world face detection problem primarily focused on reducing the crimerate and pro- viding more accurate results. It also help sminimise workforce and human errors. The proposed method uses Pix2Pix GANto generate new images. The images fed into the GAN are obtained using the object detection mechanism of YOLOv5. This method can be extended and implemented with other GANs for various use cases.

Fig. 8. GAN output for facial hair/beards

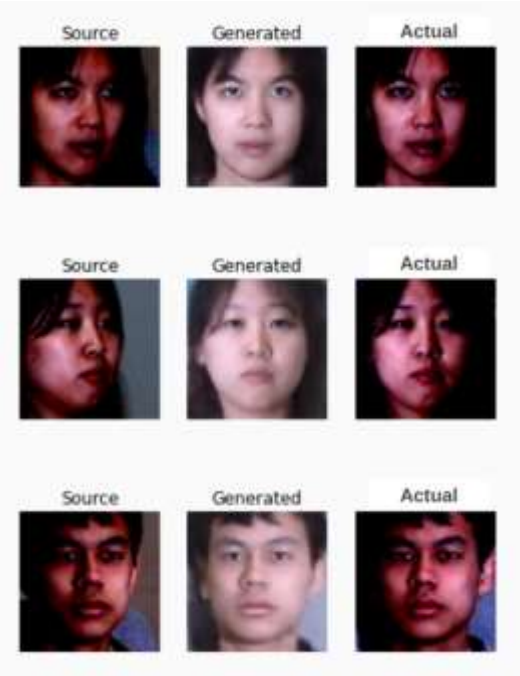

Fig. 9. GAN output for different age groups and genders

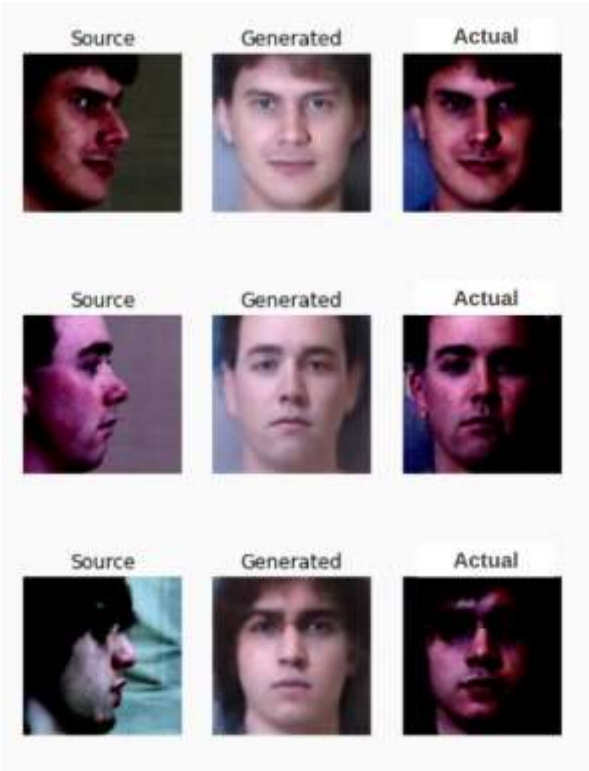

Fig. 10. GAN output when only half the face is visible

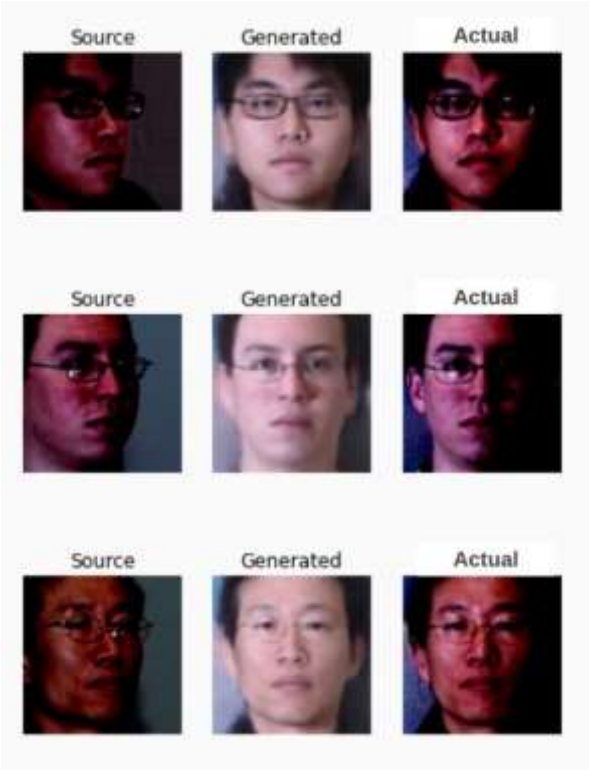

Fig. 11. GAN output for people wearing spectacles

REFERENCES

[1] N.C.R.Bureau,CrimeinIndia, Statistics vol. 1, 2020,https://ncrb.gov.in/sites/default/files/CII%202020%20Volume%201.pdf.

[2] I.J.Goodfellow,J.Pouget-Abadie,M.Mirza,B.Xu,D.Warde-Farley,S.Ozair,A.Courville, and Y.Bengio, "Generative adversarial networks," 2014.

[3] R.Gross,I.Matthews,J.Cohn,T.Kanade,and S.Baker, "Multi-pie", 2008.

[4] P.Isola,J.Y.Zhu,T.Zhou,andA.A.Efros,"Image-to-imagetranslationwithconditionaladversarialnetworks," 2017.

[5] G. Jocher, A.Chaurasia, A.Stoken, J.Borovec, Y.NanoCode012, Kwon, K. Michael, J.TaoXie, Fang, Imyhxy, Lorna,Yifu,C. V. A. Wong,D.Montes,Z.Wang,C.Fati,J.,Nadar,V.Laughing,UnglvKitDe, Sonck, P.tkianai, yxNONG, Skalski, A. Hogan, D. Nair,M.Strobel, and M. Jain, 2022,ultralytics/yolov5: v7.0 - YOLOv5SOTA Realtime Instance Segmentation.

[6] Sitharthan, R., Vimal, S., Verma, A., Karthikeyan, M., Dhanabalan, S. S., Prabaharan, N., ...&Eswaran, T. (2023). Smart microgrid with the internet of things for adequate energy management and analysis.Computers and Electrical Engineering, 106, 108556.

[7] M.Mirza, and S. Osindero, S.,"Conditionalgenerative adversarial nets," 2014.

[8] D. Peng, Z. Sun, Z. Chen, Z.Cai, L.Xie, and L. Jin, "Detecting heads using feature refine net and cascaded multi-scalearchitecture," 2018,arXiv preprint arXiv:1803.09256.

[9] Rajesh, M., &Sitharthan, R. (2022). Introduction to the special section on cyber-physical system for autonomous process control in industry 5.0.Computers and Electrical Engineering, 104, 108481.

[10] J.Redmon,S.Divvala,R.Girshick, and A.Farhadi,"You only look once: Unified, real-time object detection," 2016.

[11] M. Tan, and Q.V. Le, "Efficientnet: Rethinkingmodel scaling for convolutional neural networks," 2020.

[12] VD,A.K.,V. Kumar,M. Subramanian,K. Vengatesan,and M. Ramakrishnan, "Facial recognition system for suspectidentification using a surveillance camera," 2018.

[13] X.Wang,H.Guo,S.Hu,M.C.Chang, and S.Lyu,"Gan-generated faces detection: A survey and new perspectives," 2023.

[14] X.Wang,K.Yu,S.Wu,J.Gu,Y.Liu,C.Dong,C.Loy,Y.Qiao, andX.Tang,"Esrgan: Enhanced super-resolutiongenerative adversarial networks," 2018.

[15] J.Y.Zhu,T.Park,P.Isola, and A. A. Efros,"Unpaired image-to-image translation using cycle-consistent adversarialnetworks".