PATTERN RECOGNITION LABORATORY

Department of Computer Science and Engineering
Indian Institute of Technology (Banaras Hindu University), Varanasi

Lab-in-Charge: Dr. Pratik Chattopadhyay
Assistant Professor, CSE

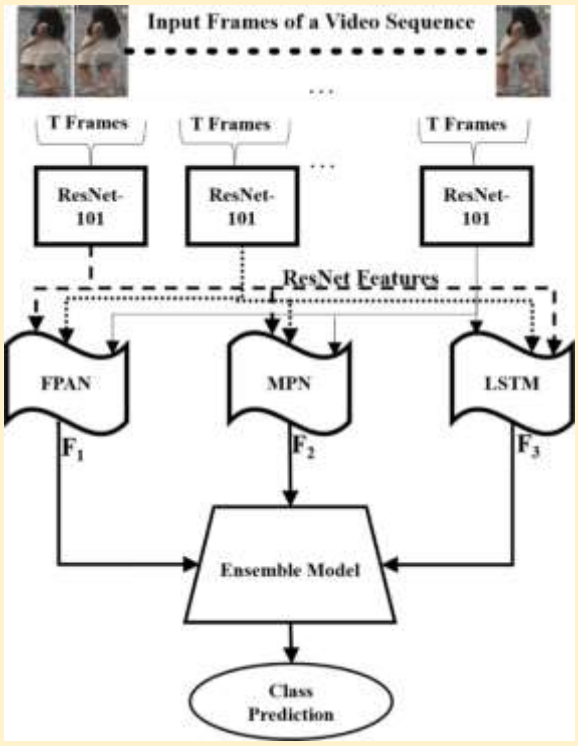
Lab Server: Three graphics processing units (GPUs): Nvidia Titan Xp with 12-GB RAM, total FB memory as 12196 MB and total BAR1 memory as 256 MB, and the other two are Nvidia GeForce GTX 1080 Ti with 11-GB RAM, total FB memory as 11178 MB and total BAR1 memory as 256 MB, OS: CentOS, System RAM 96 GB.

Research Theme: Computer Vision, Pattern Recognition, Machine/Deep Learning Applications

Main Research Activities Include:

Person Re-identification

Person re-identification plays a central role in tracking and monitoring crowd movement in public places, and hence it serves as an important means for providing public security in video surveillance application sites. The problem of person re-identification has received significant attention in the past few years, and with the introduction of deep learning, several interesting approaches have been developed. In one approach, we have developed an ensemble model called Temporal Motion Aware Network (T-MAN) for handling the visual context and spatio-temporal information jointly from the input video sequences. Our methodology makes use of the long-range motion context with recurrent information for establishing correspondences among multiple cameras. The proposed T-MAN approach first extracts explicit frame-level feature descriptors from a given video sequence by using three different sub-networks (FPAN, MPN, and LSTM), and then aggregates these models using an ensemble technique to perform re-identification. The method has been evaluated on three publicly available data sets, namely, the PRID-2011, iLIDS-VID, and MARS, and re-identification accuracy of 83.0%, 73.5%, and 83.3% have been obtained from these three data sets, respectively. The network architecture is shown below.



We are also working on other dimensions of the problem and challenges such as occlusion and open-set re-identification.

***Related Publications**:*

Tagore, Nirbhay Kumar, Pratik Chattopadhyay, and Lipo Wang. "T-MAN: a neural ensemble approach for person re-identification using spatio-temporal information." Multimedia Tools and Applications 79, no. 37 (2020): 28393-28409.

Nirbhay Tagore, Pratik Chattopadhyay SMSNet: A Novel Multi-Scale Siamese Model for Person Re-Identification, 17th International Conference on Signal Processing and Multimedia Applications. (accepted in 2020).

Gait Recognition

Computer vision-based gait recognition has evolved into an active area of research since the past decade, and a number of useful algorithms have been proposed over the years. Among the existing gait recognition techniques, pose-based approaches have gained more popularity due to their inherent capability of capturing the silhouette shape variation during walking at a high resolution. However, a short-coming of the existing pose-based gait recognition approaches is that their effectiveness depends on the accuracy of a pre-defined set of key poses and are, in general, not robust against varying walking speeds. We have proposed an improvement to the existing pose-based approaches by considering a gallery of key pose sets corresponding to varying walking speeds instead of just a single key pose set. This gallery is generic and is constructed from a large set of subjects that may/may not include the subjects present in the gait recognition data set. Comparison between a pair of training and test sequences is done by mapping each of these into the individual key pose sets present in the above gallery set, computing the Active Energy Image for each key pose, and next observing the frequency of matched key poses in all the sets. Our approach has been evaluated on two popular gait data sets, namely the CASIA B data and the TUMGAID data.

We are also working on handling occlusion in gait recognition and view-invariant gait recognition.

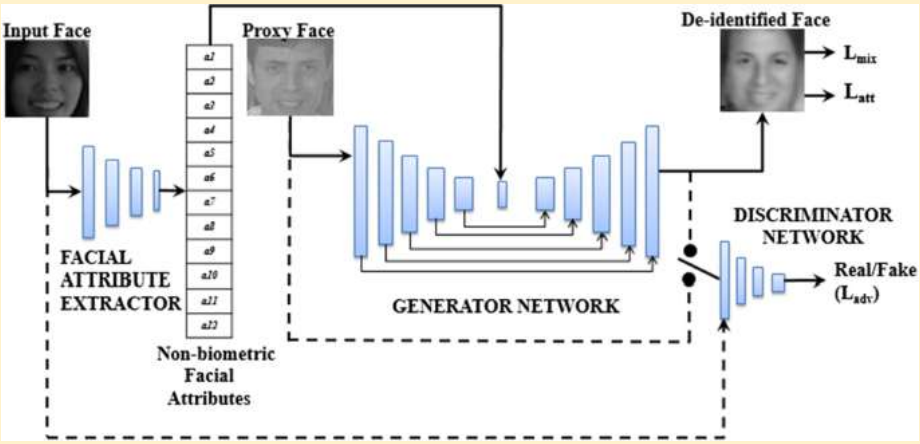
***Related Publications**:*

Gupta, Sanjay Kumar, and Pratik Chattopadhyay. "Exploiting pose dynamics for human recognition from their gait signatures." Multimedia Tools and Applications (2020): 1-19.

Gupta, Sanjay Kumar, Gaurav Mahesh Sultaniya, and Pratik Chattopadhyay. "An Efficient Descriptor for Gait Recognition Using Spatio-Temporal Cues." In Emerging Technology in Modelling and Graphics, pp. 85-97. Springer, Singapore, 2020.

Face De-identification

Due to the availability of low-cost internet and other data transmission media, a high volume of multimedia data gets shared very quickly. Often, the identity of individuals gets revealed through images or videos without their consent, which affects their privacy. Since face is the only biometric feature that reveals the most identifiable characteristics of a person in an image or a video frame, the need for the development of an effective face de-identification algorithm for privacy preservation cannot be over-emphasized. Existing solutions to face de-identification are either non-formal or are unable to obfuscate identifiable features completely. We are working towards developing an automated face de-identification algorithm that takes as input a facial image and generates a new face preserving the emotion and non-biometric facial attributes of a target face. Generative Modelling techniques are being employed to carry out this work. A schematic diagram of our proposed network architecture is as follows:



***Related Publications**:*

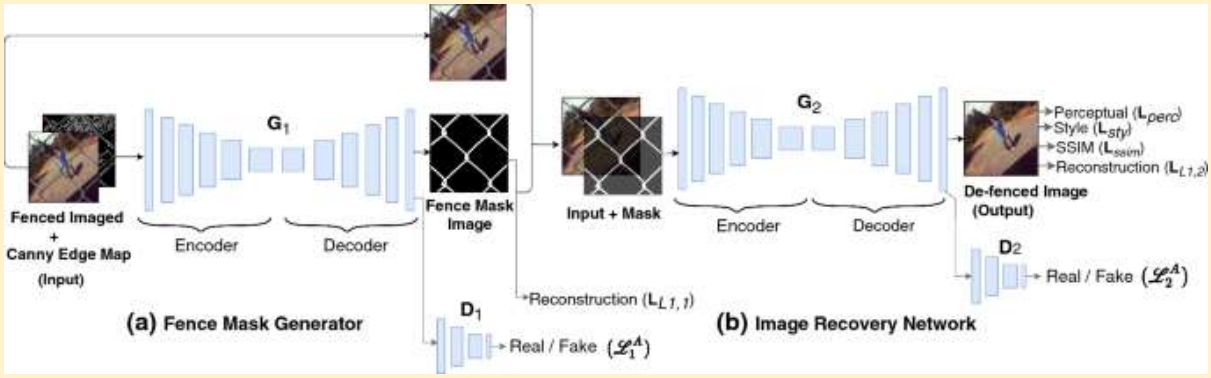
Agarwal, Ayush, Pratik Chattopadhyay, and Lipo Wang. "Privacy preservation through facial de-identification with simultaneous emotion preservation." Signal, Image and Video Processing (2020): 1-8.

Aggarwal, Alakh, Rishika Rathore, Pratik Chattopadhyay, and Lipo Wang. "EPD-Net: A GAN-based Architecture for Face De-identification from Images." In 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), pp. 1-7. IEEE, 2020.

Image Defencing

Image de-fencing is one of the most important aspects of recreational photography in which the objective is to remove the fence texture present in an image and generate an aesthetically pleasing version of the same image without the fence texture as shown in figure next.

We are developing automated and effective techniques for fence removal and image reconstruction using conditional generative adversarial networks (cGANs). These networks have been successfully applied in several other domains of computer vision, focusing on image generation and rendering. One of our approaches is based on a two-stage architecture involving two cGANs in succession, in which the first cGAN generates the fence mask from an input fenced image, and the next one generates the final de-fenced image from the given input and the corresponding fence mask obtained from the previous cGAN as shown in the network architecture next.



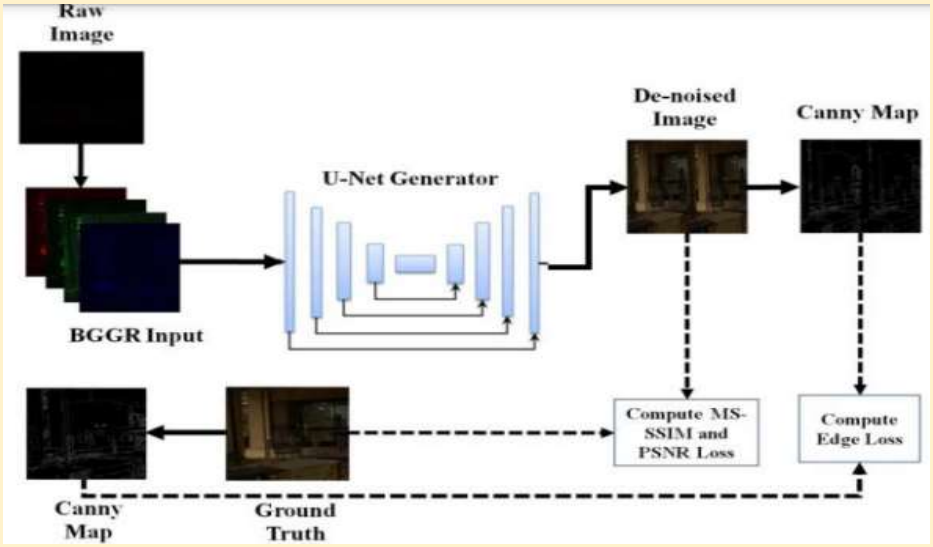
In another approach, we are trying to come up with a single network to perform image defencing efficiently without any intermediate mask generation step.

***Related Publication**:*

Gupta, Divyanshu, Shorya Jain, Utkarsh Tripathi, Pratik Chattopadhyay, and Lipo Wang. "A robust and efficient image de-fencing approach using conditional generative adversarial networks." Signal, Image and Video Processing (2020): 1-9.

Image Denoising

Image noise refers to the specks of false colors or artifacts that diminish the visual quality of the captured image. It has become our daily experience that with affordable smart-phone cameras we can capture high clarity photos in a brightly illuminated scene. But using the same camera in a poorly lit environment with high ISO settings results in images that are noisy with irrelevant specks of colors. Noise removal and contrast enhancement in images have been extensively studied by researchers over the past few decades. But most of these techniques fail to perform satisfactorily if the images are captured in an extremely dark environment. In recent years, computer vision researchers have started developing neural network-based algorithms to perform automated de-noising of images captured in a low-light environment. Although these methods are reasonably successful in providing the desired de-noised image, the transformation operation tends to distort the structure of the image contents to a certain extent. We propose an improved algorithm for image enhancement and de-noising using the camera's raw image data by employing a deep U-Net generator. The network is trained in an end-to-end manner on a large training set with suitable loss functions. To preserve the image content structures at a higher resolution compared to the existing approaches, we make use of an edge loss term in addition to PSNR loss and structural similarity loss during the training phase. The network architecture is shown next.



***Related Publication**:*

Krishnan, Utsav, Ayush Agarwal, Avinash Senthil, and Pratik Chattopadhyay. "Image Enhancement and Denoising in Extreme Low-Light Conditions." ICIAMR, (2019).

Insider Threat Detection

An insider threat scenario refers to the outcome of a set of malicious activities caused by intentional or unintentional misuse of the organization's systems, networks, data, and resources. Prevention of insider threat is difficult, since trusted partners of the organization are involved in it, who have authorized access to these confidential/sensitive resources. The state-of-the-art research on insider threat detection mostly focuses on developing unsupervised behavioral anomaly detection techniques with the objective of finding out anomalousness or abnormal changes in user behavior over time. However, an anomalous activity is not necessarily malicious that can lead to an insider threat scenario. As an improvement to the existing approaches, we propose a technique for insider threat detection from time-series classification of user activities. Initially, a set of single-day features is computed from the user activity logs. A time-series feature vector is next constructed from the statistics of each single-day feature over a period of time. The label of each time-series feature vector (whether malicious or non-malicious) is extracted from the ground truth. To classify the imbalanced ground-truth insider threat data consisting of only a small number of malicious instances, we employ a cost-sensitive data adjustment technique that under-samples the non-malicious class instances randomly. As a classifier, we employ a two-layered deep autoencoder neural network and compare its performance with other popularly used classifiers: random forest and multilayer perceptron. Encouraging results are obtained by evaluating our approach using the CMU Insider Threat Data, which is the only publicly available insider threat data set consisting of about 14-GB web-browsing logs, along with logon, device connection, file transfer, and e-mail log files.

***Related Publication**:*

Chattopadhyay, Pratik, Lipo Wang, and Yap-Peng Tan. "Scenario-based insider threat detection from cyber activities." IEEE Transactions on Computational Social Systems 5, no. 3 (2018): 660-675.