Unpaired Image- to- Image Translation using Cycle Generative Adversarial Networks

Abhinav Dwarkani, Maitri Jain, Jash Thakkar, Kottilingam Kottursamy

Abstract: In this burgeoning age and society where people are tending towards learning the benefits adversarial network we hereby benefiting the society tend to extend our research towards adversarial networks as a general-purpose solution to image-to-image translation problems. Image to image translation comes under the peripheral class of computer sciences extending our branch in the field of neural networks. We apprentice Generative adversarial networks as an optimum solution for generating Image to image translation where our motive is to learn a mapping between an input image(X) and an output image(Y) using a set of predefined pairs[4]. But it is not necessary that the paired dataset is provided to for our use and hence adversarial methods comes into existence. Further, we advance a method that is able to convert and recapture an image from a domain X to another domain Y in the absence of paired datasets. Our objective is to learn a mapping function G: A —B such that the mapping is able to distinguish the images of G(A) within the distribution of B using an adversarial loss.[1] Because this mapping is high biased, we introduce an inverse mapping function F B—A and introduce a cycle consistency loss[7]. Furthermore we wish to extend our research with various domains and involve them with neural style transfer, semantic image synthesis. Our essential commitment is to show that on a wide assortment of issues, conditional GANs produce sensible outcomes. This paper hence calls for the attention to the purpose of converting image X to image Y and we commit to the transfer learning of training dataset and optimising our code.You can find the source code for the same here.

Keywords: Machine learning, General Adversarial Networks, Computer Vision, Convolutional Neural Network, Tensorflow

I. INTRODUCTION

Many problems arising in various domains can be overcome with the advection of translation tasks. Image to image translation comes under the supervised approach for computer vison Cycle GAN helps us to overcome the lack of paired datasets available everytime having to one to one relationship between the source. Domain and target domain. Just as a text can be conveyed in many languages the same way any scene can translated into RGB with the help of translation tasks creating enormous amount of images. We explore Cycle GAN as generative model of data that learn the generative nature of the pictures..

Abhinav Dwarkani*, UG Student, Department of Information Technology, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India. E-mail: takhilabhinav@gmail.com

Maitri Jain, UG Student, Department of Software Engineering, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India. E-mail: jain.maitri7@gmail.com

Jash Thakkar, UG Student, Department of Software Engineering, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India. E-mail: jashthakkar18@gmail.com

Kottilingam Kottursamy, Associate Professor, Department of Information Technology, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India. E-mail: kottilik@srmist.edu.in

This makes cycle GAN suitable to image to image translation where the given particular conditions on an input image it can yield an output image. Our contribution as we propose this paper is that cycle GAN is able to tackle translation tasks with utmost accuracy. We have used Tensor flow framework which alone is sufficient enough to attained desirable results trained on GPU

1.1 Objective:

Our objective is to build a model for automatic translation of images from one domain to another domain without explicitly having the paired datasets of the two domains.
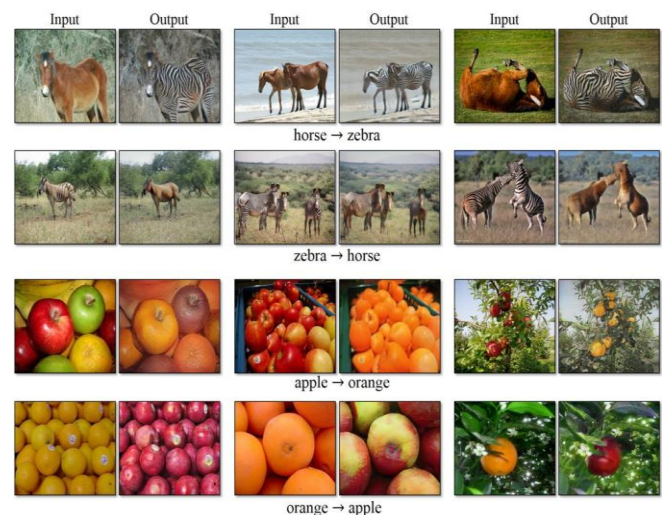

Fig 1-Given any two unpaired datasets our paper proposes translation between two domains

1.2 Motivation

Our project focuses on image to image translation which tends to convert an image representation from one scene to another. The motivation to perform this project comes from the idea of neural style transfer where one tends to convert a realistic scenery image into that of a painting. Our paper overcomes the shortcomings faced in neural style transfer and pix2pix conversion which the explicit one to one relationship between input domain and output domain or in other words paired datasets.
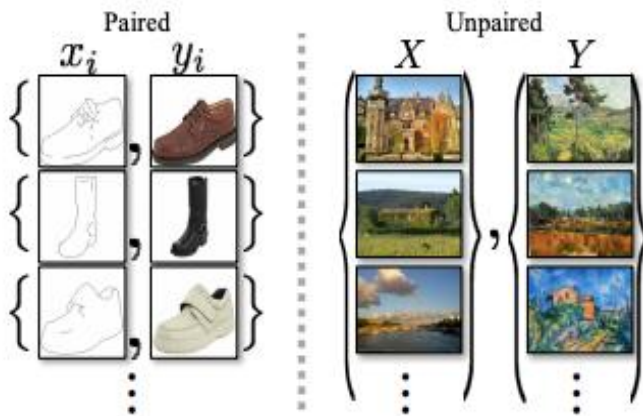
Unpaired Image- to- Image Translation using Cycle Generative Adversarial Networks

Fig -2 (a) Paired Datasets (b) Unpaired Datasets

II. LITERATURE REVIEW

[1]A diverse range of methods apply feedforward networks for translation tasks. They gives us appealing results with the forward propagation completing in a very short span of time. These networks avails in image segmentation and sematic label generation.

[2] Image outpainting is concerned with predicting the visual context of a image beyond its boundary. It uses a three phase training architecture involving DCGAN on a subset of serene365 datasets. Given an image a*b size the model outcasts a*b+2m image such that the images looks realistic and natural.
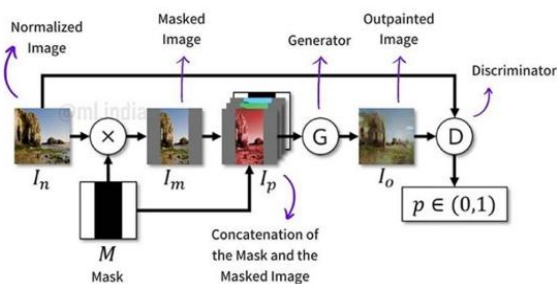

Fig 3-Training pipeline in image outpainting

[3] GAN are class generative models which do not explicitly model the data distribution, but rather provides a sample for it and sampling is performed using deep neural network. This neural network takes input a random noise vector and transforms into the model distribution.

[4] With the upcoming innovation in the GAN ,Neural Style transfer helps to become a artist with a few line of code.It takes an input of two images an actual image and a conversion image and then it blend it and coalesce the contents of the image as if it was real artwork.

[5]The pix2pix cycle generative models were used to perform paired image translation.It is capable of performing edge detection , object detection.It converts the implicit sketch of something into the desirable actual image .

[6]PAN is model based approach that focuses mainly on enhancing the resolution of images. It comes under the class of supervised learning using various CNN architectures and upsampling and downsampling to achieve the desired task.

III. METHODOLOGY

GAN are class generative models which do not explicitly model the data distribution, but rather provides a sample for it and sampling is performed using deep neural network. This neural network takes input a random noise vector and transforms into the model distribution.

$P_{data} = \{x_I\}$ i=1.........N

Our goal is to generate a model that is equivalent or similar to the data provided.

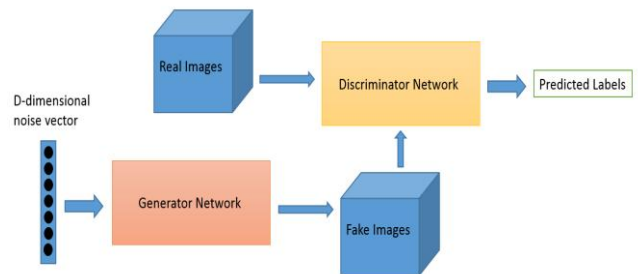$$P_{model}^{(x)} = P_{data}^{(x)} \tag{1}$$

Fig.4- General Methodology for General Adversative Network

The Generative Adversarial framework consists of two neural networks, one is the generator and other is the discriminator. A Generator is a neural network that takes input a random noise and transforms it into a sample from the model distribution. The Discriminator s a neural network that whose function is to determine if an input data is real or fake. Discriminator can be described as analogous to erudite which can detect whether a work is veraciousness or fallacious. [2]

3.1 Cycle GAN

Cycle Gan learns the mapping between two image domains using an unsupervised learning. The method Cycle GAN performs is by training a Generators to understand the relation between a domain A into a image that makes it analogous from domain B or vice versa[7]. Cycle GAN uses cycle consistency loss to train the model without the need for paired datasets. To translate images from one domain to another domain it does not require one to one mapping between the two domains.
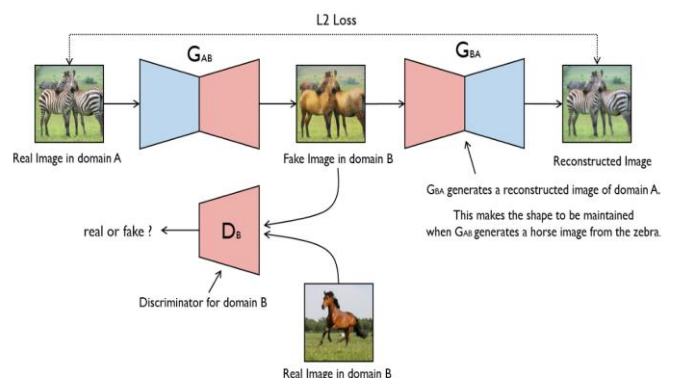

Fig 5 -Cycle GAN Architecture

Given an image generated by the generator as $G_A(z)$ and the discriminator assigns a score $D_B(G_A(z))$ to it. This score will be between 0 and 1 and will tell us the probability of image being real or fake.

3.2 Objective Of Cycle GAN

For a given z, the generator would want to maximize log $(D_B(G_A(z)))$ or minimize $\log(1- D_B(G_A(z)))$[9].This is for single z the generator would like to do this for all possible values of z because every z coming from the normal distribution is a valid input to the generator.[3]

If z was discrete and drawn from uniform distribution (i.e. $p(z)=1/N\forall z$) then the generator objective function would be

$$\min \sum_{i=1}^{N} \frac{1}{N}(\log(1- D_B(G_A(z)))) \qquad (2)$$

However in our case z is continuous and not uniform $(z\sim(0,I)$ so the equivalent objective function would be

$$\int p(z)\ \log(1- D_B(G_A(z))) \qquad (3)$$

Now lets look at the discriminator. The task of the discriminator is to assign high score to real images and low score to fake images. In other words it should try maximize the following objective function

$$\max_{\theta} E_{x\sim p_{data}}[\log D_\theta(x)]+E_{z\sim p(z)}[\log(1-D_\theta(G_\phi(z)))] \qquad (4)$$

3.3 Adversarial Loss

We produce adversarial losses from the two the relating functions and the discriminator associated with it.The first term is computed when X domain is input is Y domain is output and viceversa[11].

$$\mathcal{L}_{adversarial} = \ell_{X\to Y}(G,D_Y) + \ell_{Y\to X}(F,D_X)$$

$$(5)$$

3.4 Cycle Consistency Loss

The cycle consistency loss is computed computing the average root mean square error attained between the translation of an image from one domain to another domain and then again translated to the same domain yields image similar to the actual domain image .Cycle consistency loss is difference in features of the reconstructed original image and the actual original image.[16]

$$forward\ cycle\ consistency\ loss : X-> G(X)-> F(G(X)) \sim \hat{X}$$

$$backward\ cycle\ consistency\ loss : Y-> F(Y)-> G(F(Y)) \sim \hat{Y} \qquad (6)$$

Now we can formulate the objective of the cycle consistency loss by coalescing all the losses associated with the loss by a hyperparameter λ.[8]

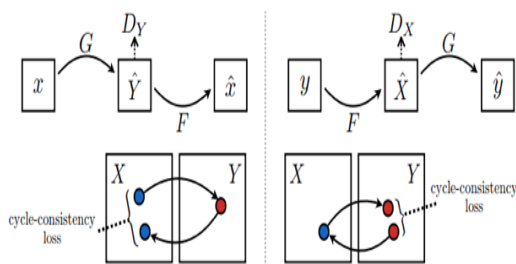$$\text{Loss}_{full}= \text{Loss}_{adv} + \lambda\text{Loss}_{cyc} \qquad (7)$$

Fig 6- Cycle consistency losses when image is translated from one domain to another

IV. IMPLEMENTATION

4.1 Pre-processing Of Images: To avoid overfitting of the model the images are made to go through a no of steps before which the image dataset can be used for modelling. The first basic step crop the images to 256*256 and further on the images are flipped horizontally and randomly shuffled to attain the desired dataset required for modelling.[5]

4.2 Building the Generator:

Each of the cycle GAN generator consists of an encoder , a decoder and a transformer. The input image is feeded to the encoder.[13] The encoder is capable of extracting features from the given input using convolution represenations.The encoder consists of 3 convolutions that reduce the feature after every iteration by one fourth. The first convolution uses 64 filters , the second one uses 128 filters and finally the last one uses 256 filters.[9] The output of the encoder after applying the leaky Relu activation is passed to the transformer. The Transformer performs the upsampling and downsampling based on the size of the input it received. The output of the transformer is then passed to the decoder which uses Transpose Convolution blocks to convert the image back to its original size[4]. The generator loss is computed by calculating the mean error between generated image and the desired target image.
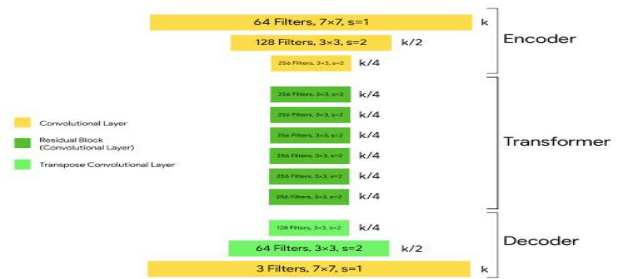
Fig 7-Generator Architecture in GAN

4.3 Build the Discriminator:

The architecture of the discriminator is a patch GAN.Each residual block in the discriminator comprises of 4*4 convolution layer followed by instance normalization with Relu activation with k filters and applying a stride of 2.After the last convolution layer we use a sigmoid operation to attain a 1*1 output. The shape of the output layer attained is (batch_size,30,30,1).The discriminator receives source image and target image which it has objective to distinguish as real and source image and generated image it should distinguish as fake.[9] We coalesce these inputs to attain the desired architecture for the discriminator.
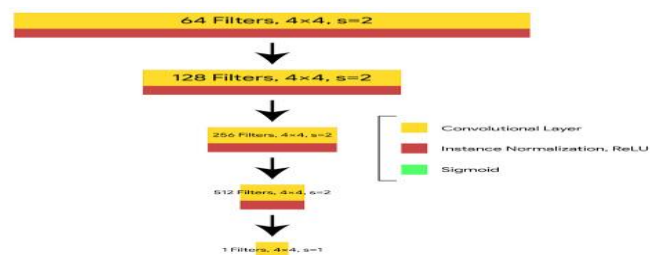
Fig8-Discriminator Architecture

4.4 Network Architecture:

The network architecture comprises of two generators and two discriminators that are to be trained to attain the architecture. The first generator learns to transform the source image to target image. The second generator learns to translate the image back to the original source image. The discriminator is trained to distinguish the real source image and the reconstructed image.[7] Our network objective is to minimize the reconstruction loss so as to attain translation with maximum efficiency.
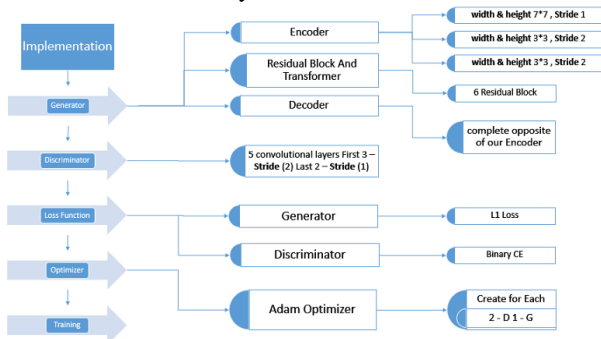

Fig8-Steps of Cycle GAN Architecture Implementation

4.5 Training the model:

The steps involved in training the model basically involves the computation listed below:

- Get the prediction of real vs fake
- Calculate the total loss
- Backpropagate using gradients
- Optimize the gradients

All weights are initialized to a normal distribution centred around zero with a standard deviation of 0.02.For GAN we propose two optimizers one for the generator and other one for the discriminator to help them running simultaneously and thereby minimizing the loss.In this paper we propose two adam optimizers with lamda=10 and beta=0.05 used for training. The generator and discriminator both have opposing loss function fighting against each other to provide the desired result.The training is performed over 100 epochs using a GPU. To increase the efficiency of the model we increase the no of epochs the model through. .Increasing the no of epochs though increases the efficiency but it increases the time consumed to perform the translation.

V.   MODEL VALIDATION

The following is plot of the training losses for the generator and the discriminator after each epoch.

Fig 9-Traning loss for Discriminator and Generator

We can forsee that the generator start with very high errors but with time piecemeal producing germane images thus helps in enervating down the error. This provides us the the behoves of diversity of applications such as neural style transfer and increasing the resolution of given images .The prerequisites to perform the following applications are just the source and target datasets which in this case are just the directory of images. Cycle GAN can be beneficial when need to to domain translation but when the geometric translation comes into focus it fails certainly the reason being it is trained to perform only appearance changes not geometric changes.[8,9,10]

Fig 10- Translation from X to Y after 100 epochs

Fig 11- Translation from Y to X after 100 epochs

Fig 12- Translation from X to Y after 5000 epochs

Fig 13 – Translation from Y to X after 5000 epochs

We are able to recognize that Cycle GAN produce low resolution images which is an ongoing field of research .We observed that our model faced difficulty with matching the colors with high precision which is due to cycle consistency loss being too small for this to overcome this an entire different approach is needed which tends to make our approach supervised.[11]

5.1 Evaluation Metrics

The generator model is trained iteratively over a plethora of epochs as there is no parameter to measure the objective function we do not know when to stop the training process of the model.

AMT Perceptual Studies: For the map->aerial task the authors[12] run real versus the fake perceptual studies on Amazon Mechanical Turk to measure the voraciousness of the results obtained. Observers were given a pair of images and were asked to determine which one of them they thought were real.

French Inception Distance: We adopt FID to measure the distance between the generated image and the real image .We compute FID scores for zebra->horse translation comapare with the other methods.[5]

5.2 Applications:

- **Collection Style Transfer:** Cycle GAN knows to replicate the an entire collection of works instead of just transferring the piece of it.[5]

- **Object Transformation:** Cycle GaN can the transfer the object from one domain to another very beneficial in the health sector for organ transplant.[14]

- **Season Transfer:** Cycle GAN is also capable of inverting the images from one season to another thus help in forecasting.

- **Painting Generation:** Cycle GAN can convert the image into painting and vice versa also thus opening and creating an entire new artwork.

- **Photo Enhancement:** Cycle GAN is also capable of increasing the clarity of a image
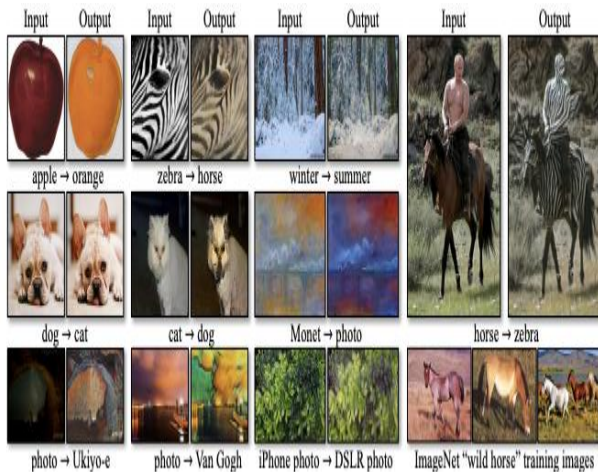thereby increasing dpi associated with every image.

Fig 14- Our failed attempts at image translations

VI. CONCLUSION

In this paper we proposed unsupervised image to image translation using Cycle GAN which produced pleasing and fastidious results with utmost accuracy on various datasets. We have explored datasets that require texture and color transformation. We also applied transformations to geometric aspects but the model still accuracy with these types of transformations. We also observed that computing the cycle consistency loss is behooved in order to attain style changes with maximum accuracy. We depicted the superiority of our method to other models through various models.[7]

This paper depicts how image to image translation synthesis can yield productivity in various domains and enable various forms of interactivity given appropriate output pairs desired. This paper has demonstrated the feasibility of the translation computed using google colab using the Tensorflow

framework of the model cycle GAN posit that these research could be useful in the upcoming time .The

cycle GAN is continuous ongoing field of research that is ameliorating day by day and tends to provide devastating results. This paper enforces the paradoxical boundaries that term the model to be unsupervised. cycle GAN can be used to train various GAN model and ameliorate their performances[14].

Our image translation lacks the dexterity of diversity.We also observe than cycle GAN produce low resolution images the drawback which will be addressed in our future work. In our next future works we would tend to diversify our works into other domains such as Medicare and neural style transfer and image resolution optimization involving multiple generators which shows a dearth of cycle consistency loss thereby ameliorating the resolution of the images[13].

REFERENCES

1. Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz1 Bryan Catanzaro, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks" arXiv preprint arXiv:1711.11585
2. K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan. "Unsupervised pixel-level domain adaptation with generative adversarial networks". In CVPR, 2017.
3. D. Eigen and R. Fergus. "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture". In ICCV, 2015
4. L. A. Gatys, M. Bethge, A. Hertzmann, and E. Shechtman. "Preserving color in neural artistic style transfer". arXiv preprint arXiv:1606.05897, 2016.
5. I. Goodfellow. NIPS 2016 tutorial: "Generative adversarial networks". arXiv preprint arXiv:1701.00160, 2016
6. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. "Image to-image translation with conditional adversarial networks". In CVPR, 2017
7. D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in ICLR, 2015
8. C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cun- ´ ningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. "Photo-realistic single image superresolution using a generative adversarial network". In CVPR, 2017
9. J. Johnson, A. Alahi, and L. Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution". In ECCV, 2016.
10. K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition". In ICLR, 2015.
11. J.-Y. Zhu, P. Krahenb ¨ uhl, E. Shechtman, and A. A. ¨ Efros. "Generative visual manipulation on the natural image manifold". In ECCV, 2016
12. X. Wang and A. Gupta. "Generative image modeling using style and structure adversarial networks". In ECCV, 2016
13. Jarrett, K., Kavukcuoglu, K., Ranzato, M., and LeCun, Y. (2009). "What is the best multi-stage architecture for object recognition?" In Proc. International Conference on Computer Vision (ICCV'09), pages 2146–2153. IEEE.
14. Rezende, D. J., Mohamed, S., and Wierstra, D. (2014). "Stochastic backpropagation and approximate inference in deep generative models". Technical report, arXiv:1401.4082.

AUTHOR PROFILE

Abhinav Dwarkani, is a student of Information Technology, SRM Institute Of Science and Technology, Chennai .He is the one who have a predilection for Deep learning and Data Analysis. He has worked on various projects that involves domains such as sentimental analysis, image segmentation and Graphic designing. His research interests are deep learning , speech recognition.

Unpaired Image- to- Image Translation using Cycle Generative Adversarial Networks

Maitri Jain is a student of Software Engineering ,SRM Institute Of Science and Technology,Chennai.She has a predilection of Deep learning and UI/UX.She has worked on various projects involving domains such as sentimental analysis , web based apps and Image translations, emotion detection. Her research interests are automation , deep learning and web development.

Jash Thakkar, is a student of Software Enigineering,SRM Institute of Science and Technology.He is one of the enthusiast in Artificial Intelligence and Machine Learning.He is one the core member of Software Engineering Association. He has done various projects in similar fields.His interests include Deep Learning, Wed Development and Data Analysis.

Kottilingam Kottursamy, is an Associate Professor in School of Computing, SRMIST, India. He has completed his P.hd from Anna University. He is one of the Co-Principal Investigators of IITK-SPARC Project, in collaboration with UC Davis, USA. A former Member NGNLab, Anna University, Chennai, India. His research interest includes Artificial Intelligence, Next Generation Networks and Bio-Statistics.