Degraded Image Restoration Using an Ensemble of GANs for Improved Classification Performances

A thesis

submitted in partial fulfillment of the requirement for the

Degree of

Master of Computer Science and Engineering

of

Department of Computer Science and Engineering of

Jadavpur University

by

ANIRUDDHA MAITY

Registration No.: 154147 of 2020-2021

Examination Roll No.: M4CSE22023

Under the guidance of

Prof. MAHANTAPAS KUNDU

Department of Computer Science and Engineering

Jadavpur University, Kolkata-700032

India
2022

Faculty of Engineering and Technology

Jadavpur University

Certificate of Recommendation

This is to certify that the dissertation entitled "Degraded Image Restoration Using an Ensemble of GANs for Improved Classification Performances" has been carried out by Aniruddha Maity (University Registration No.: 154147 of 2020-2021, Examination Roll No.: M4CSE22023) under my guidance and supervision and be accepted in partial fulfillment of the requirement for the Degree of Master of Computer Science and Engineering in the Department of Computer Science and Engineering, Jadavpur University. The research results presented in the thesis have not been included in any other paper submitted for the award of any degree in any other University or Institute.

Prof. MAHANTAPAS KUNDU (Thesis Supervisor)
Department of Computer Science and Engineering
Jadavpur University, Kolkata-32

Countersigned

Prof. Anupam Sinha
Professor and Head, Department of Computer Science and Engineering,
Jadavpur University, Kolkata-32

Prof. Chandan Mazumdar
Dean, Faculty of Engineering and Technology,
Jadavpur University, Kolkata-32

Faculty of Engineering and Technology

Jadavpur University

Certificate of Approval*

This is to certify that the thesis entitled “Degraded Image Restoration Using an Ensemble of GANs for Improved Classification Performances” is a bonafide record of work carried out by **Aniruddha Maity** in partial fulfilment of the requirements for the award of the degree of Master of Computer Science and Engineering in the Department of Computer Science and Engineering, Jadavpur University during the period of august 2021 to July 2022. It is understood that by this approval the undersigned does not necessarily endorse, or approves any statement made, opinion expressed or conclusion drawn therein but approve the thesis only for the purpose for which it has been submitted.

____

Signature of Examiner 1

Date:

____

Signature of the Examiner 2

Date:

*Only in the case the project report is approved.

Declaration of originality and compliance of academic ethics

I hereby declare that this thesis entitled "Degraded Image Restoration Using an Ensemble of GANs for Improved Classification Performances" contains a literature survey and original research work by the undersigned candidate, as part of his degree of Master of Computer Science and Engineering.

All information has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have fully referenced all materials and results that are not original to this work.

Name: Aniruddha Maity
Registration No.: 154147 of 2020-2021
Exam Roll No.: M4CSE22023
Thesis Title: Degraded Image Restoration Using Ensemble of GANs for Improved Classification Performances.

____

Signature with date

ACKNOWLEDGEMENTS

The writing of the thesis as well as the related work has been a long journey with input from many individuals, right from the first day till the development of the final project. I would like to express my deepest gratitude to my supervisor, **Prof. Mahantapas Kundu**, Professor, Department of Computer Science and Engineering, Jadavpur University for giving me the opportunity to do research and providing invaluable guidance throughout this work. His dynamism, vision, sincerity, and motivation have deeply inspired me. He has taught me the methodology to carry out the work and to present the works as clearly as possible. It was a great privilege and honour to work and study under his guidance.
I would also thank **Prof. Mita Nasipuri**, Professor, Department of Computer Science and Engineering, Jadavpur University for her patience, guidance, suggestions and moral support in times of need.
I would like to thank **prof. Ram Sarkar** and **Prof. Nibaran Das**, professor, Department of Computer Science and Engineering, Jadavpur University for helping me in every possible way to complete this thesis.
I am also particularly thankful to Dr. Anupam Sinha, Professor and Head of the Department of Computer Science and Engineering, Jadavpur University for allowing us to carry out research in the department.
I would also like to thank all the faculty members of the Department of Computer Science and Engineering of Jadavpur University for their continuous support.
This thesis would not have been completed without the inspiration and support of my family and friends and a number of wonderful individuals including my batchmates of Master of Computer Science and Engineering at Jadavpur University — my thanks and appreciation to all of them for being part of this journey and making this project possible.

____

Aniruddha Maity
Examination Roll No.: - M4CSE22023
Registration No.: - 154147 of 2020-2021
Class Roll No.: - 002010502023

CONTENTS

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Degraded Image Classification

Computer vision and artificial intelligence researchers have been working on image classification for a long time. It's a task to associate an input image with one of the previously defined image classes, such as an item in an image or video. The performances of image classification methods have recently been improved by using supervised deep learning, such as Convolutional Neural Networks (CNNs) [1], [2]. All the experiments and testing of these image classification methods have mostly used clear images. But in practice, images taken from digital cameras are not always clear. They suffer from gaussian-blur, motion blur, salt-paper noise, out of focus blur etc. So, Image degradation needs to be considered for image classification in practice. In Fig, 1.1 Gaussian blurred degraded images have been shown corresponding to its original version. Different types Degradation and each degraded image with different levels of distortion makes the classification task a challenging problem.

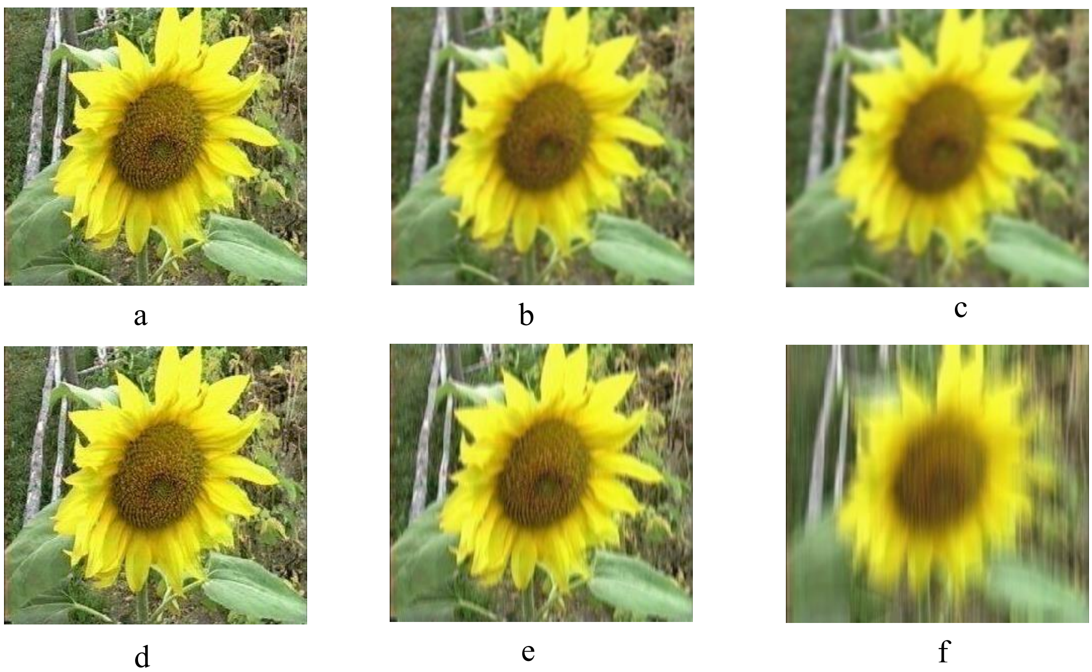


Fig 1.1: a, d clear images, b and c are gaussian blurred images with degradation level 3 and 5 respectively, e and f motion blurred images with blur factor 5 and 15 respectively.

1.2 Importance of Degraded Image Classification

Degraded image classification has great importance as it is required in practice for many computer vision related applications such as,

- Autonomous Driving:

  Autonomous cars need thoroughly surveillance of root during its travel and also need continuous detection of objects. But rainy and hazy weather, car emition etc make surveillance difficult. So there need to be some system that could do object detection and surveillance better even verse situation.

- Underwater Robotics:

  In recent years, the scientific community has become increasingly interested in creating systems for underwater object recognition for underwater world exploration, but underwater images are suffered from several kind of distortion, so common methods of object recognition will not work.

- Video Surveillance:

  In CCTV surveillance, objects face need to recognized but bad weather condition, motion of the object/human makes it difficult.

- Medical Imaging:

  In recent times diagnosis via medical imaging become very interesting area. But Digital and analogue noise, signal distortion make image distorted that why detection of diseases become very complex.

1.3 Problems with Degraded Images

Degraded images are suffered from distortion of the image which also results in perceptual loss. There are many reasons for which images taken by a digital equipment could be degraded like for instrumental problem, low light, unstable object etc. For example, relative motion between object and digital camera results motion blur image, photo taken with highly sensitive digital camera could results salt paper noise etc.

There are several kinds of degradation and each degradation has several levels of distortion which makes it difficult for traditional CNN (Convolution neural network) classifiers to classify the images.

In a study conducted by Yanting Pei et al. [3], they have shown how hidden layers features of the CNN classifiers get distorted with increasing degradation level and how those feature change with changing degradation type.

They have trained a VGGNet-16[4] and extracted features of the first to fifth max-pooling layers labelled as *pool1, pool2, pool3, pool4, pool5* and plot them as image in Figure 1.2.

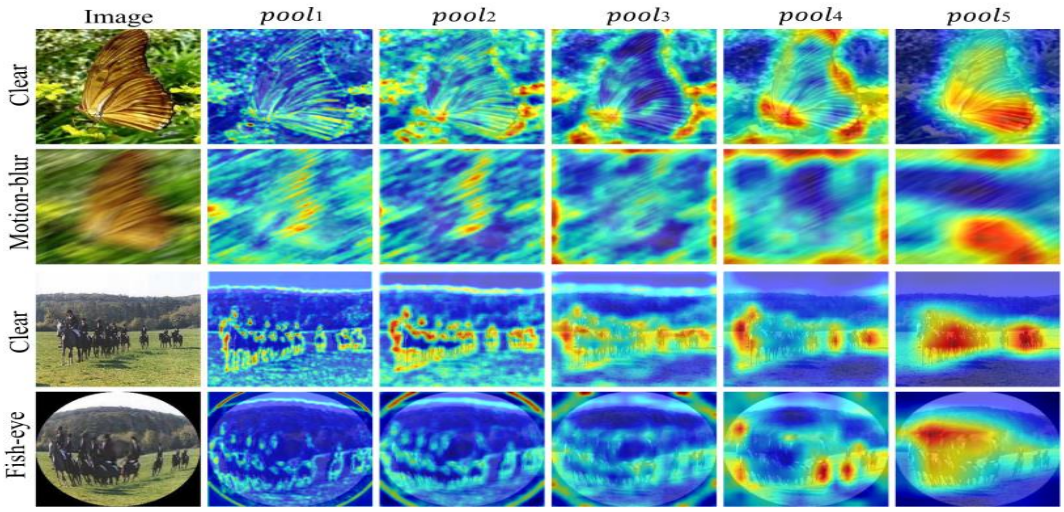
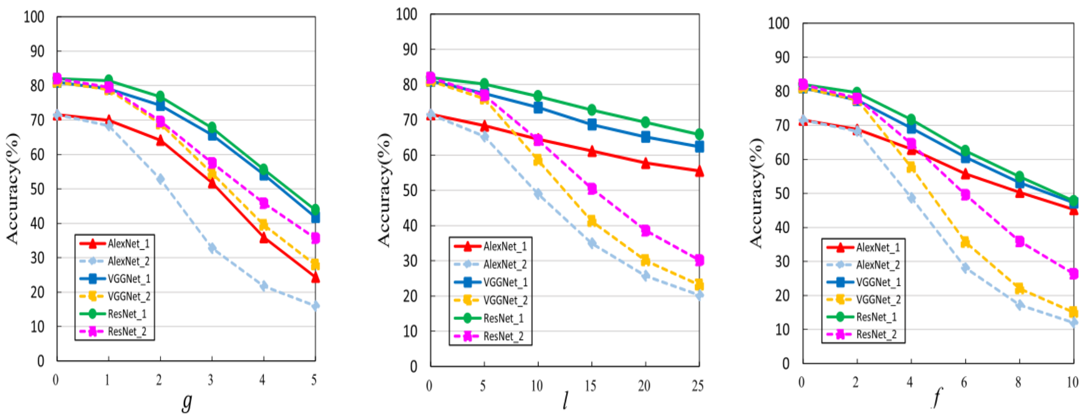


Fig. 1.2. Activations of hidden layers of CNN on image classification. From left to right are input images, and the activations at pool 1 , pool 2 , pool 3,pool 4 and pool 5 layers, respectively.

1.4 Methods for Improving Classification Performances for Degraded Images

Even though the volume of research so far conducted for degraded image classification is quite high, the number of researches attempts in this field is very limited [3][7]. has been done for image classification, but research in the field of degraded image classification is very limited.
Yanting Pei et al. [3] have done an empirical study to show how degradation affects the classification performances of the existing CNN based image classification methods like VGGNet-16, ResNet-50[5] , AlexNet[6] .
They have also shown that existing algorithms for removing haze and motion-blur gradations could not improve the CNN-based classification performance much.
In Figure 1.3, it has been shown how classification accuracy drastically falls with increase of degradation level.



a b c

Fig 1.3: The classification accuracy (%) of various degraded images synthesized from Caltech-256 dataset. From (a) to (c), classification accuracies are computed on images with Gaussian noise, Gaussian-blur images and motion-blurred respectively.

Kazuki Endo et al. came with an ensemble approach to classify degraded images[7]. Their proposed method uses of four distinct networks: a restoration network, two classification networks, an estimation network of degradation levels, and an estimation network of ensemble weights. As shown in figure 1.4. The restoration network begins by restoring degraded images. The restored images are supplied into two classification networks: one trained with clean images and another with restored images. Each classification network calculates its own probability vector. Degraded images are also supplied to the network for estimating degradation levels. The estimated degradation levels are supplied into the ensemble weight estimation network. The ensemble weights estimation network calculates ensemble weights for two probability vectors predicted by two classification networks. The weights have values between [0, 1], and their total is 1. Finally, weighted averaging is used to determine the anticipated probability vector.

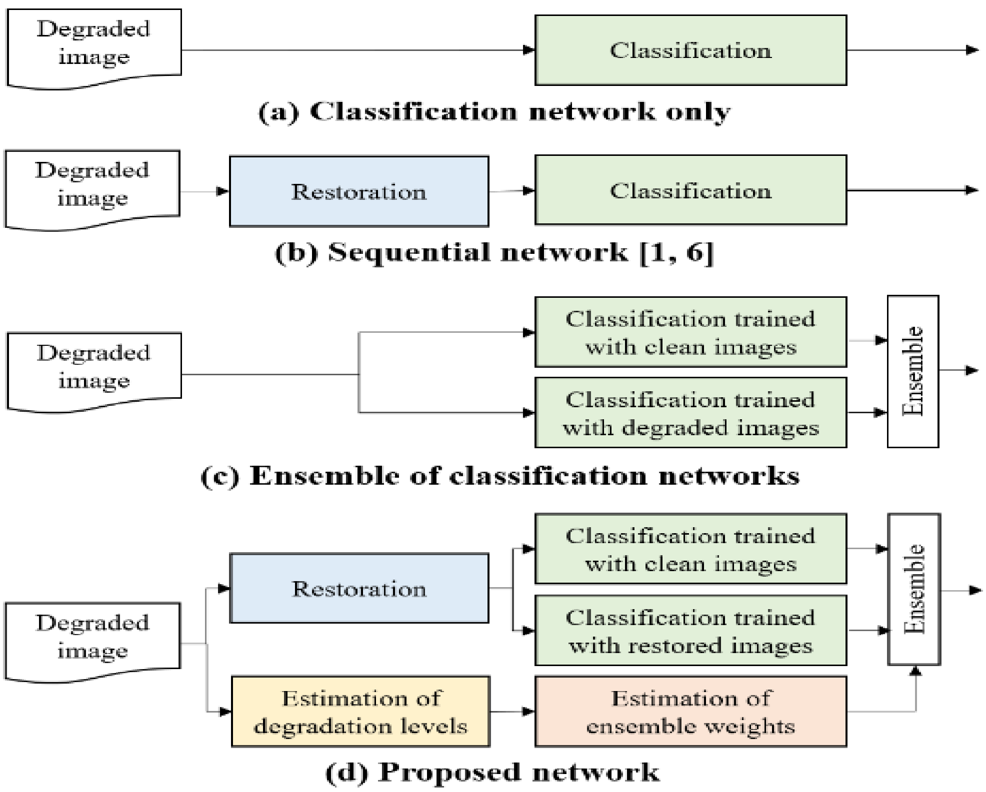


Fig. 1.4 Classification networks of degraded images proposed by Kazuki Endo et. al

1.5 Motivation

It has been observed by various studies that the classification performances of various classifiers trained on clear images are very poor when supplied with degraded images as input. So, it has become a subject of research that how to enhance the performances of the classifiers on degraded images. The methods [3][7] so far investigated have attempted to enhance the classification performances of classifiers by training them exclusively with degraded images. No such attempt has been found effective. So, there is a scope for exploring various alternative approaches to enhance the classification performances of the classifiers on degrade images.

1.6 The Present Work

Considering the latest research efforts for improving classification performances on degraded images, a novel approach is explored here. In this approach, degraded images have been restored first using suitable deep learning models. Then restored images are classified by an ensemble of trainable classifiers. The present work prefers to focus on a specific type of image degradation instead of considering all kind of degradation viz, Motion blur, hazy, white Gaussian noise etc. together.

Specifically speaking two GANs, one pix2pix GAN and the other Deblur-GAN are employed to generate the input image in restored form. The Results obtain from the two GANs combined to form the restored image finally. In the next stage an ensemble of two neural network-based models are trained to classify the restored image separately. ResNet-50 models are employed here as the two classifiers. One model is trained with clear images and other with degraded images. For final classification decision a weighted average of the decisions from the two classifiers is computed.

Chapter 2

Synthesizing Degraded Images

To train and test our proposed method, we need labelled with various degradation level and their corresponding ground truth image. Due to non-availability of such data set, Degraded image has been simulated from clear image by respective degradation model.
Paper [3] has been followed to generate degraded image.
Even though there are several types of degradations, only gaussian blur has been considered to test the proposed methodology.

2.1 Gaussian Blur

To simulate Gaussian blurred images, paper [9] has been followed to add Gaussian blur by varying the standard deviation of Gaussian kernel g. *skimage ver 0.19.2* in *python version 3.8 has been used to* generate Gaussian blurred image. The function takes image and standard deviation as a parameter and generate respective Gaussian blurred image. For standard deviation 1, 2, 3, 4 and 5 the distorted images are referred as degradation level-1, 2, 3, 4 and 5. Some example of degraded image with degradation level of 1,2,3,4 and 5 has shown in figure 2.1.

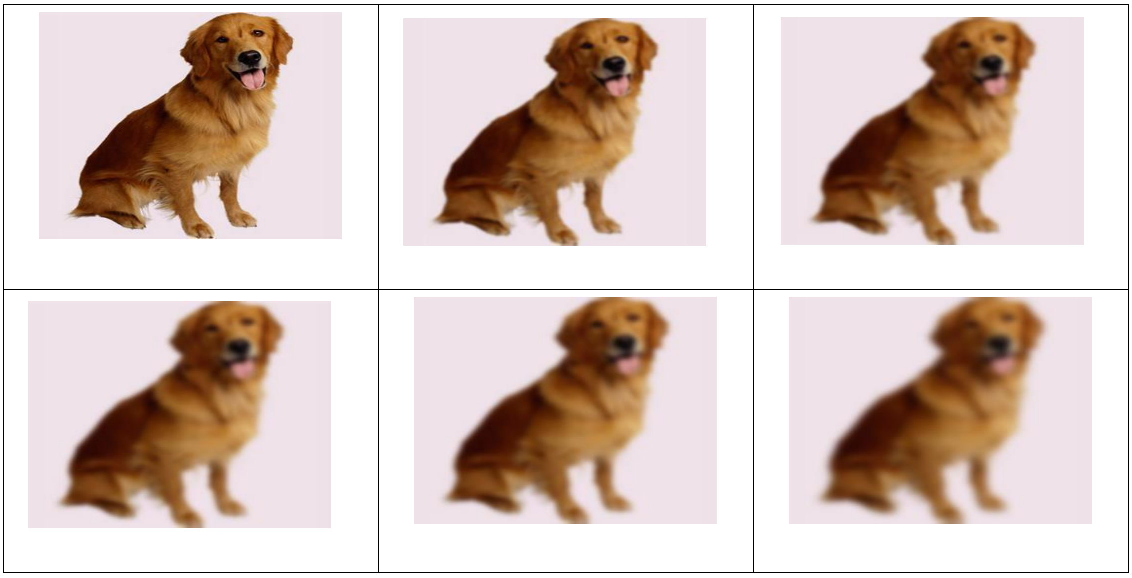


Fig 2.1: Gaussian blurred images (b-f) with standard deviation 1, 2, 3, 4, 5 respectively and a is corresponding clear image.

Chapter 3

Deep Neural Network for Image Classification and Restoration

3.1 Deep Neural Network

Deep neural networks (DNN) are a representation of a networks of linked "neurons" that can calculate the desired output by a collection of values as inputs through an objective function.
In Figure 3.1.

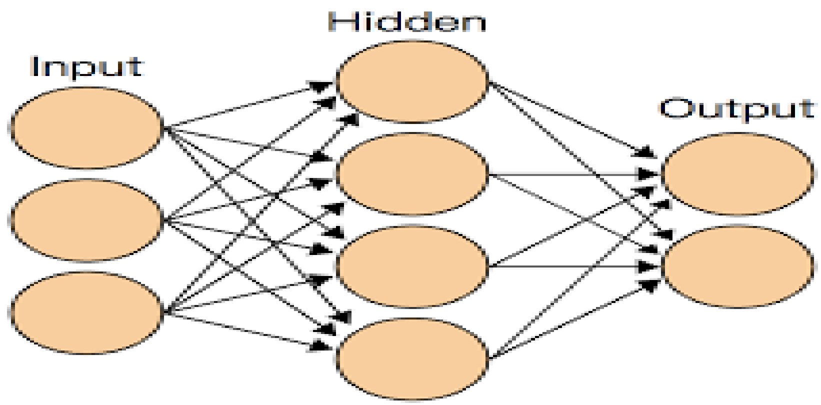


Fig 3.1: one layered neural network

After invention of backpropagation algorithm (BP) [10], DNN become very popular among the researchers in the field of pattern recognition for achieving state-of-the-art performance.
The architecture of DNN is very simple, there will be many nodes representing 'neuron' in every layer as depicts in figure-x. There will be a input layer and output layer and one or more hidden layer/layers.
Number of neurons in the input layer will be as same as size of input vector and number of neurons in output layer will be as same as desired output shape.
Number of hidden layer and each hidden layer's neuron is manually tuned.
These DNN works in two passes, in first pass input is forward to output layer which is called forward pass and during backward pass the error gradient is pass from output to input layer.

During forward pass every neuron in DNN does two operation,

I. **Vector multiplication: -** Take input from previous layer and multiply $\vec{W}$(weights vector) with it and pass the multiplication for next operation.

II. **Activation function: -** The multiplication is then going through suitable activation function like sigmoid, relu etc. This activation function map the multiplication result to a suitable format, like sigmoid usually map the input multiplication to probabilistic value (0 to 1).

During backward pass the weights $\vec{W}$ of the neuron is updated by bp rule.

$$\vec{W} = \vec{W} - l_{rate} * \left(\frac{\partial(error_w)}{\partial(\vec{W})}\right)$$

l_{rate}= Learning rate

3.2 Convolution Neural Network

Even though achieving bench mark results in many problems, DNN could not perform well in case of image classification.
Thus researchers find out a new solution Convolution neural network (CNN) which can classify image better.
Through the use of suitable kernels, a CNN may successfully capture the Spatial and Temporal relationships in a image.
Because to the reduced number of parameters involved and the reuse of weights, the architecture performs superior fitting to the image dataset.

Every CNN is combination four layers, convolution layer, batch normalization, pooling layer and last fully connected layer.

I. **Convolution Layer: -** Convolutional layers perform a convolution on the input with the kernels before forwarding the output to the next layer. The pixels in a convolution's receptive area are all converted to a single value. The outputs of the convolution layers are the new feature maps of the input image or input feature map that map the inputs to a new dimension where image is more separable. The outputs of convolution layer pass through some activation function (usually Relu). In figure 3.2, convolution operation has been shown.

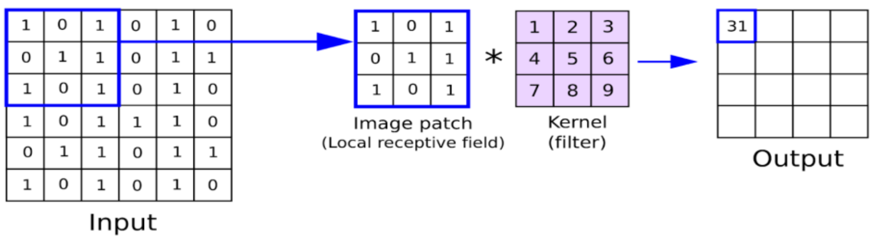


Fig 3.2: Convolution operation

II. **Batch Normalization: -** Batch normalization is a network layer that allows each layer to learn more independently. It's normalizes the outputs of previous layer. In normalization, the activations scale the input layer. Learning becomes more efficient when batch normalization is utilized, and it can also be used as a regularization to avoid model overfitting.

III. **Pooling Layer: -** The pooling operation entails sliding a two-dimensional filter across each channel of the feature map and aggregating the features that fall inside the filter's coverage zone. It consolidates the features learned by the convolutional layer feature map. Pooling layers are very simple because they often use the maximum or average values of input to down sample the data. A 2x2 max pooling operation has been shown in figure 3.3.

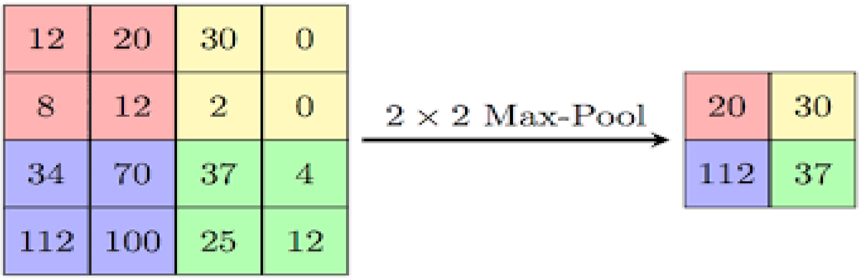


Fig 3.3: Max pooling

IV. **Fully Connected Layer: -** Every CNN classifiers have a fully connected layer at the end which take input learned feature maps from the above-mentioned layers and output the respective class. It is basically a DNN.

Convolution layer, batch normalization layer, pooling layer can be stack together in different combination to form a CNN architecture.
In subsequent time, many CNN architectures have been proposed which achieve state-of-the-art performance of many image classification task. Among these architectures most popular architectures are:

3.2.1 AlexNet

AlexNet was one of the first CNN architecture which won many image recognition competitions. AlexNet has eight layers, the first five were convolution layers, some of them followed by max-pooling layers, and the last three were fully connected layers. It used the non-saturating Relu activation function, which showed improved training performance over tanh and sigmoid. The architecture of AlexNet has shown in figure 3.4.

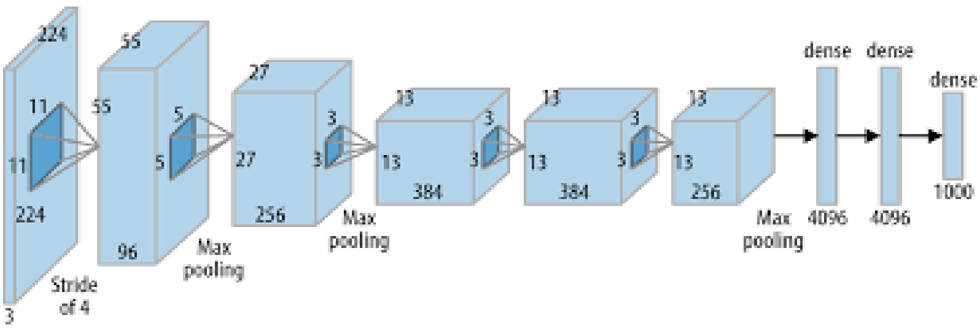


Fig 3.4: AlexNet

3.2.2 VGG-16

The VGG model was proposed by A. Zisserman and K. Simonyan from the University of Oxford. The number 16 in the VGG name refers to the fact that it is a deep neural network with 16 layers (VGGNet). This suggests that VGG16 is a large network with over 138 million parameters. Even by today's standards, it is a massive network. The simplicity of the VGGNet16 architecture, on the other hand, is what makes the network appealing. Its architecture alone can be described as uniform. Following a few convolution layers, a pooling layer reduces the height and width. In figure 3.5, the architecture of VGG-16 has been shown.

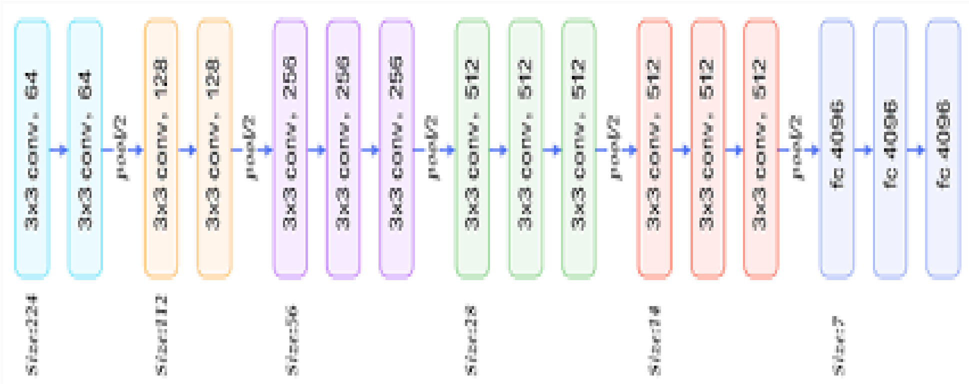


Fig 3.5: VGG16

3.2.3 ResNet-50

ResNet stands for Residual Network. It is an innovative neural network that was first introduced by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun in their 2015. ResNet has many variants that run on the same concept but have different numbers of layers. Resnet50 is used to denote the variant that can work with 50 neural network layers.

Machine learning specialists add extra layers in deep convolutional neural networks to address a computer vision challenge. As deeper layer learns more separable features, thus these extra layers make it easier to solve challenging task in computer vision. While the number of stacked layers might enhance the model's features, a deeper network can reveal the degradation problem. The model's performance consequently declines on both the training and testing sets of data as a result of vanishing gradient problem.

In ResNet researchers solve this issue by adding residual block in CNN .

A residual block is a stack of layers set in such a way that the output of a layer is taken and added to another layer deeper in the block. The non-linearity is then applied after adding it together with the output of the corresponding layer in the main path. This by-pass connection is known as the shortcut or the skip-connection. the presence of the residual blocks prevents the loss of performance whenever, the activations tend to vanish or explode.

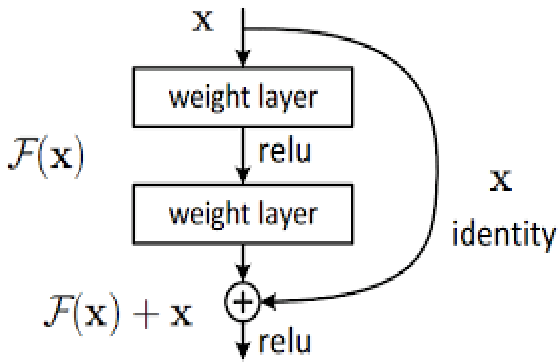


Fig 3.6: Residual Block

The architecture of ResNet50 is given below,

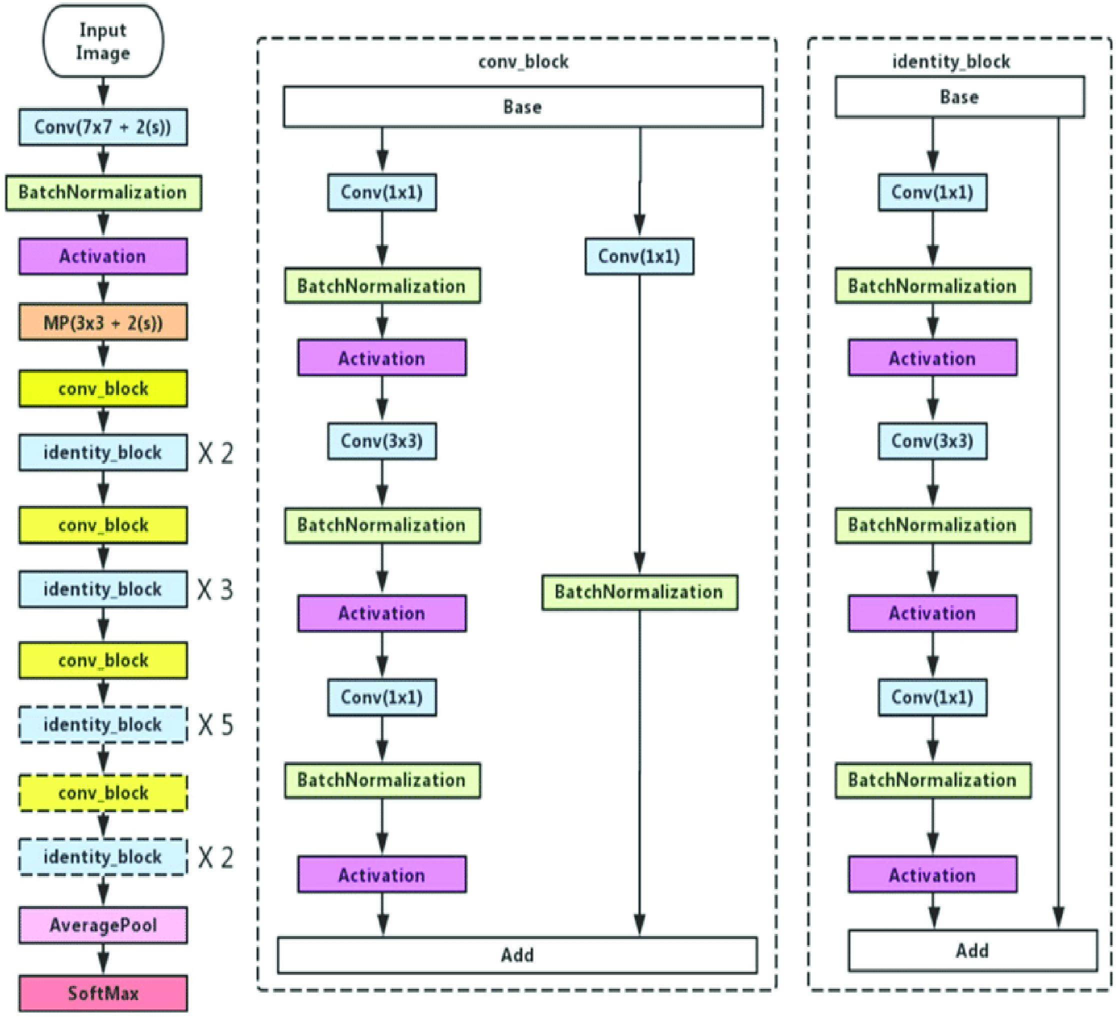


Fig 3.7: ResNet50

3.3 Generative Adversarial Network

GANs (generative adversarial networks) [11] are an interesting new machine learning technique. GANs are generative models, which means they generate new data instances that are similar to training data. GANs, for example, may generate images that resemble images of human faces, despite the fact that the faces do not belong to any actual person.
A generative adversarial network (GAN), figure 3.8, has two parts:

- The **generator** learns to generate plausible data. The generated instances become negative training examples for the discriminator.
- The **discriminator** learns to distinguish the generator's fake data from real data. The discriminator penalizes the generator for producing implausible results.

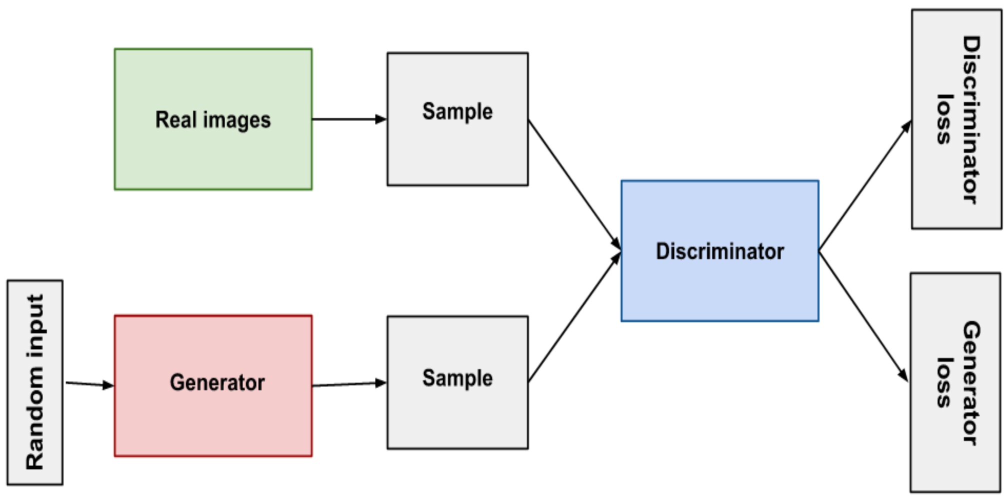


Fig 3.8: GAN architecture

Both generator and discriminator are actually CNN architecture which can be train using back propagation.

Basic GAN uses minimax loss to train both generator and discriminator. Minimax loss is a extension of binary cross entropy loss given belowe,

$$E_x[log(D(x))] + E_z[log(1 - D(G(z)))]$$

where

$D(x)$ is the discriminator's estimate of the probability that real data instance x is real.

E_x is the expected value over all real data instances.

$G(z)$ is the generator's output when given noise z.

$D(G(z))$ is the discriminator's estimate of the probability that a fake instance is real.

E_z is the expected value over all random inputs to the generator (in effect, the expected value over all generated fake instances $G(z)$.

Even though GANs were invented for data creation but later many GAN were developed for image reconstruction and translation. These variant of GANs can regenerate distorted image to clear image and transform one to image to another form. There are many GANs available for these tasks, there are two most used GAN architecture is discussed.

3.3.1 Pix2Pix GAN

Pix2Pix[12] is a Generative Adversarial Network, or GAN, model designed for general purpose image-to-image translation. The Pix2Pix GAN architecture involves the careful specification of a generator model, discriminator model, and model optimization procedure.
Both the generator and discriminator models use standard Convolution-Batch Normalization-Relu blocks of layers as is common for deep convolutional neural networks.

- **Generator Network**

A U-Net [18] model architecture is used for the generator, instead of the common encoder-decoder model. The U-Net model architecture is very similar in that it involves down sampling to a bottleneck and up sampling again to an output image, but links or skip-connections are made between layers of the same size in the encoder and the decoder, allowing the bottleneck to be circumvented.

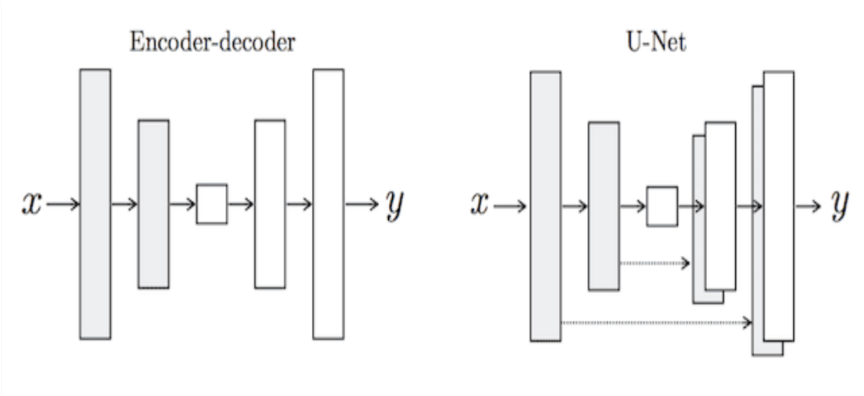


Fig 3.9: Encoder-Decoder and U-Net

- **Discriminator Network**

In contrast to the classic GAN model, which use a deep convolutional neural network to categorise images, the Pix2Pix model employs a PatchGAN. This is a deep convolutional neural network that classifies patches of an input images as real or fraudulent rather than the complete image.

- **Losses**

Generator learns a loss that adapt to the data. There are two loss that generator tries to minimizes

I. Sigmoid cross-entropy loss of the generated images and an array of ones.

II. The L1 loss, which is a MAE (mean absolute error) between the generated image and the target image.

On the other hand, discriminator takes binary cross entropy loss where real image is labelled as ones and generated image is labelled as zeroes.

3.3.2 Deblur-GAN

Deblur-GAN [13], an image to image translation conditional GAN, mainly proposed for motion deblurring of images.

Deblur-GAN contains a generator network which generate restore images and a critic network instead of discriminator.

- **Generator Network**

The generator contains two down sampling blocks, each of which contain a convolution layer, batch normalization layer and relu activation layer. Then there are nine residual block which create skip connections.

Finally there are two up sampling layer which are basically transpose convolution layer.

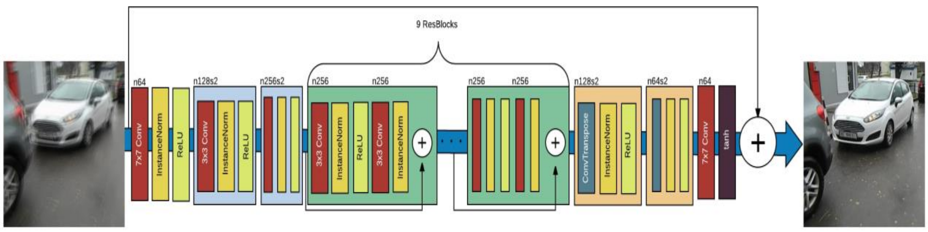


Fig 3.10: Deblur-GAN

- **Critic Network**

Unlike other GAN which contains a discriminator which classify the generated images and real images, in Deblur-GAN there is no discriminator. Instead Deblur-GAN contain a critic network which compute Wasserstein loss between generated image and real image.

- **Losses**

$$\mathcal{L} = \underbrace{\underbrace{\mathcal{L}_{GAN}}_{adv\ loss} + \underbrace{\lambda \cdot \mathcal{L}_X}_{content\ loss}}_{total\ loss}$$

The loss is actually addtion of two loss function

I. Adversarial loss :-

$$\mathcal{L}_{GAN} = \sum_{n=1}^{N} -D_{\theta_D}(G_{\theta_G}(I^B))$$

D_{θ_D} = critic output

$G_{\theta G}(I^B)$ = generator output on input image I^B

II. Content loss:-

L_x= L2 loss between the feature maps obtained from some convolution (after activation) before maxpooling layer within the VGG19 network by generated image and ground truth image. This loss is also called VGG-loss.

Λ = 100

Chapter 4

Restoration of Degraded Images for Improved Classification

4.1 Restoration network

As discussed in the section-1.3, Image classification mainly governed by the hidden feature and distortion of image results distortion in hidden features, thus reduces the classification performances.
To classify degraded images better, there is a need to keep those hidden feature as same as non-degraded images.
Here, a ensemble method has been proposed where the degraded images are restored by two GANs.

- **Selection of GANs**

Basically, two images to image translator conditional GAN has been choose for image restoration. First is pix2pix GAN and another is Deblur-GAN which are discussed is previous section.

- **Loss Function**

Even though Both GAN's architecture remain same but loss function for training of GANs more precisely generator loss has been slightly changed.
As the main goal of the restoration not only gain the perceptual view of the images but also improve the classification performances.
To do that a class loss information is added with loss function of respective GANs.
Class loss is basically the categorical cross entropy loss between outputs of ResNet pre-trained with clear ground truth images, for ground truth image and generated image.

$$L_{class} = catagoricalcrossentropy\left(R(I_{tar}), R\left(I_{pred}\right)\right)$$

where R is Resnet output I_{tar} and I_{pred} are ground truth and generated image respectivly.
In case of discriminator loss of Deblur-GAN, binary cross entropy loss has been used instead of using Wasserstein loss.

- **Ensemble Procedure**

For each pixel position at the outputs, the two GANs produce two-pixel values for each of R, G, B channels. To measure which of these two values is more accurate, the pixel value given for that position in the target image, taken from validation set, is considered. The more frequent a GAN introduces pixel values for some pixel position closer to its target value, the higher becomes authenticity of result for that position. Depending on that an image mask is introduced for containing the result of the two. If pixel position (i, j) in the image mask has value k then pixel value produce by the k^{th} GAN is to be considered to produce the final predicted image. In the present case, since two GANs are trained for image restoration, the value of k = 2. There such mask to be formed on the validation set images for R, G and B channel.

- **Architecture**

The architecture of ensemble of GANs given below in Fig-4.1.

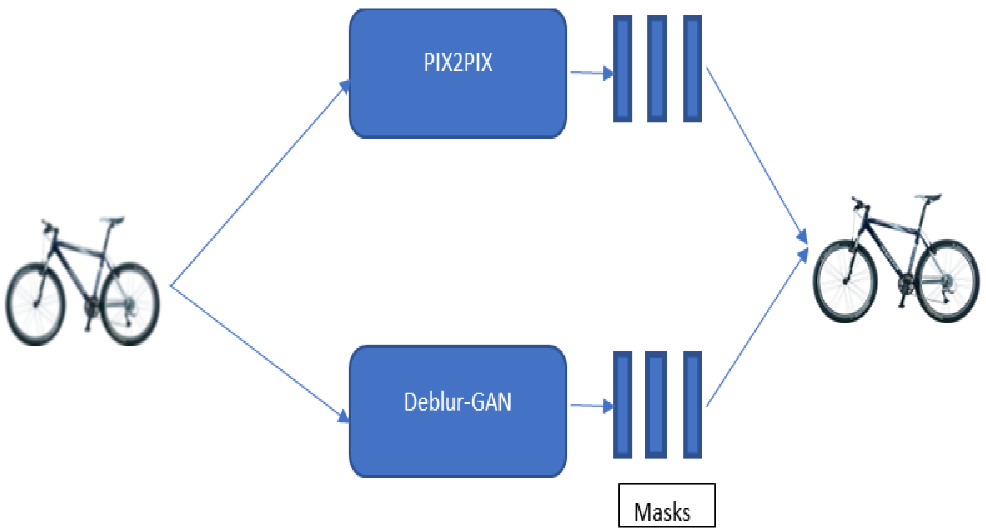


Fig 4.1: Ensemble GAN architecture

Both GANs take input the degraded image and the output of the GANs passes through respective masks and produce the restored image.

4.2 Classification Network

Main aim of the propose work is to classify images. To do so a ensemble of two classification network has been used. Both the networks are actually resnet-50 but trained on different type of training set.

- **Selection Classification Network**

There are many classification networks available for image classification like vgg-16, resnet-50, densenet-169[16] etc. So, selection of classifier from varieties of such classifier is not easy. To select the best classifier, a thorough study has been conducted. Training and test of clear images over the classification networks vgg-16, DenseNet-169 and ResNet-50 has been done. Among them ResNet and Vgg-16 fits the data well but ResNet-50 was choose as it generalizes better.

- **Architecture**

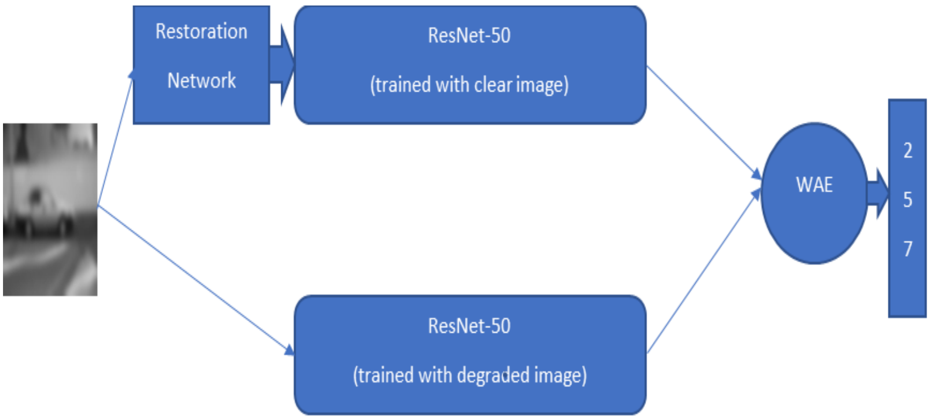


Fig 4.2: Ensemble resnet50 architecture

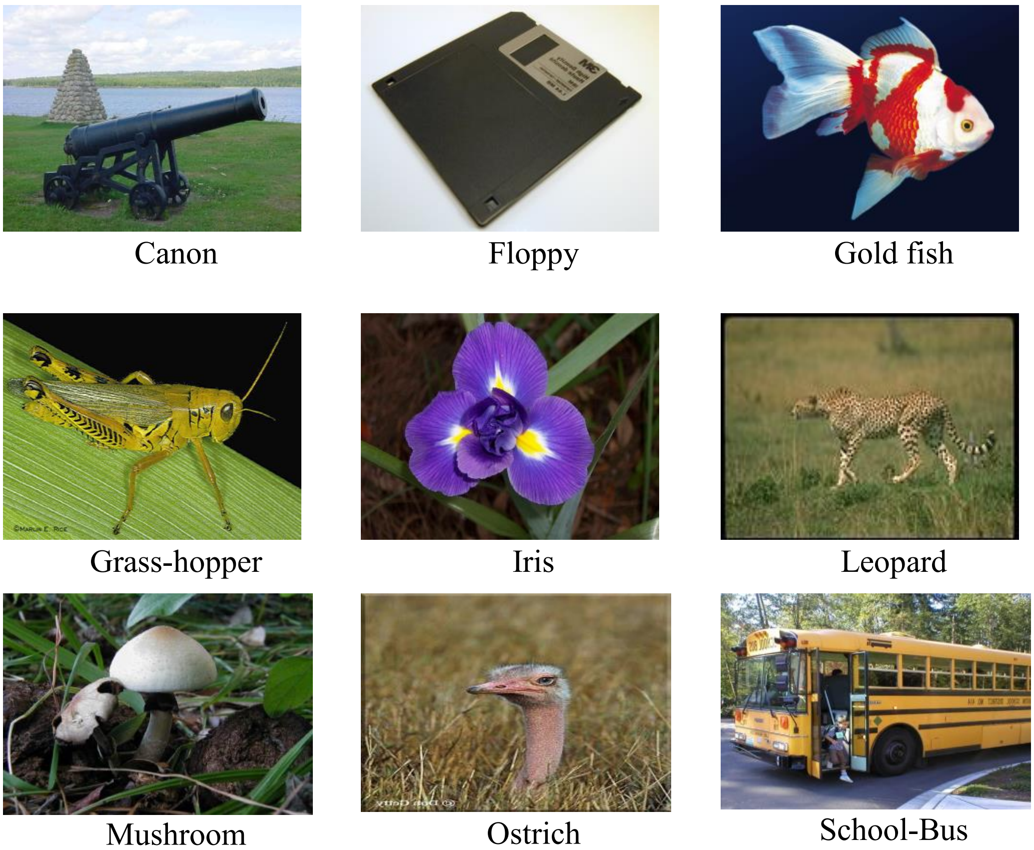
The architecture of ensemble procedure given in figure 4.2 where a degraded image fit into two different networks Simultaneously. First images are fitted into restoration network[section-] the output of restoration network the fitted into ResNet-50 which is trained on clear image (ResNet1).
Second degraded image directly fitted into ResNet-50 which is trained on all level of degradation (ResNet2). The output of both resnet-50 is ensembled using WAE (Weighted average ensemble) approach [14] .

Chapter 5

Experiments and Results

5.1 Data Set

Caltech-256 data set has been used in these experiments. This data set has been widely used for image classification. It contains 30,607 RGB images from 257 classes. 256 classes consist of images of specific objects like unicorn, sunflower, skyscraper etc. and one class called clutter class contains random images. The data set is heavily imbalanced. For example, the class with the maximum number of images contains 527 images where the class with minimum number of images contains only 80 samples. This makes the task of image classification difficult for any classifier in this data set. Images of few samples taken from various classes of Caltech-256 data set are shown in figure-5.1**.**



Canon | Floppy | Gold fish

Grass-hopper | Iris | Leopard

Mushroom | Ostrich | School-Bus

Fig 5.1: Some class of images from Caltech-256

5.2 Formation of Training and Test Set

As specified in previous section, the data set is heavily imbalanced. To maintain a uniformity in the numbers of samples of all classes in the training set, paper[3][4] has been followed to split the data set among train set, validation set and test set. The number of training samples chosen for each image class is kept same as the 60% of the total number of samples present in the smallest class in Caltech-256. The number is 48 and the number of samples for validation and test sets for each image class is set to 32. Since the validation set and the test set consists of 15% and 25% respectively of the all samples, i.e., 80 here, validation and test sets consist of 12 and 20 samples respectively.
After fixing up the numbers of training, validation and test samples for each image class, the samples are randomly chosen from Caltech-256 data set. The images so chosen are degraded with levels-1,2,3,4 of Gaussian blur to complete preparation of training, validation and test sets for two GANs. All these data sets are named as *training-set1, validation-set1* and *test-set1*.
The training, validation and test set for ResNet1(trained with clear images) are all formed with images taken from Caltech-256. In choosing sizes of these data sets, the same procedure has been followed as described above.
The data sets, named *train-set1*^{1}, *validation-set1*^{1} and *test-set1*^{1} used for training, validation and test the ResNet2(trained with degraded images) respectively are as same as those in the corresponding data sets for the GANs with the only difference that image samples are all labeled here with the corresponding image classes.
The validation set used for computing the said weight scores, for WAE method, the union of *validation-set1*^{1,} and validation-set of clear image with class label has been taken.

5.3 Training procedure

For this experimentation, all images have been resized to 256 x 256. Training is performed in phases. In first phase, two GANs, viz, Deblur-GAN and pix2pix GAN are separately trained with *train-set1*. *Adam optimizer* with initial *learning rate 0.00005* has been used to train both the GANs for *10 epochs*.
In second phase, to train the ResNets for classification, Adam optimizers have been used with initial *learning rate* 0.00001 for 120 *epochs*. Both

ResNet were pre-trained on ImageNet dataset [16] and they have been fined tuned using respective data set as specified in previous section.
All these models and experiments have been implemented using *Tensorflow ver 2.9.1* on *google colab.*

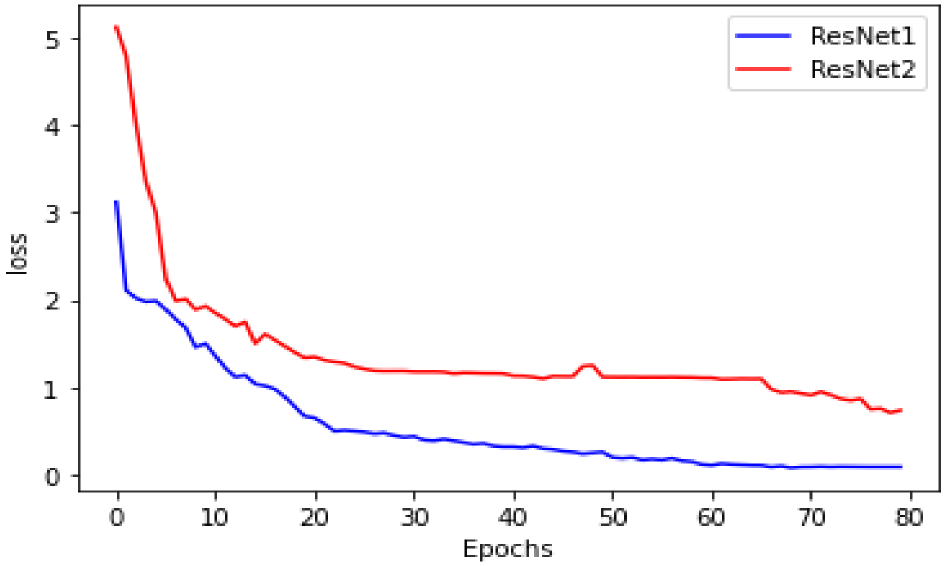


Fig 5.2: Epochs vs Training loss for last 80 epochs of ResNet1 and ResNet2

5.2 Results

5.2.1 Performances of GANs Employed for Image Restoration

In this section, the performances of GANs used for image restoration are presented on *test-set1* .
In figure-5.1, some restored images using proposed method with their corresponding degraded and ground truth images are shown.

Degraded Image	Clear Image	Restored Image

Fig – 5.3: Degraded Image, their corresponding ground truth and restored version.

5.2.1.1 Results on test-set1

The FSIM (feature similarity index measure) [17] has been measured between restored images and ground truth images. It has been also shown that how ensemble method of GANs increase the FSIM compare to pix2pix and deblur-GAN alone together.

Image Quality Measure Metric (FSIM)			
Degradation Level of Gaussian blurred images	**pix2pix GAN**	**Deblur-GAN**	**Ensemble**
1	0.924	0.925	0.939
2	0.8904	0.910	0.911
3	0.8761	0.901	0.905
4	0.8553	0.8701	0.879

Table 5.3

5.2.2 Performances of ResNets as Image Classifiers

In this section, performances of two ResNet-50 and their ensemble are given on *test-set1*.

Classification Rate (Accuracy)			
Degradation Level of Gaussian blurred images	**ResNet-1**	**ResNet-2**	**Ensemble**
0	80.00	78.80	78.00
1	67.24	77.91	78.00
2	61.48	73.10	75.91
3	54.00	72.85	73.12
4	45.12	70.01	71.12

Table 5.5: Accuracy measure of ResNet1, ResNet2 and Proposed Ensemble Model

5.2.3 A comparative Assessment

A comparison between the results of proposed method and previous method done Yanting Pei et al. [3] have been shown in below table:

Gaussian blur images degradation level	**VGG-16 [yanting Pei et. al.]**	**Proposed Method**
0	78.80	78.00
1	78.00	78.00
2	75.50	75.91
3	73.00	73.12
4	70.00	71.12

Table 5.6: Accuracy of previous method and proposed method.

Chapter 6

Conclusion

One major objective of this work is to establish a methodology for classification of degraded images. Previously, some attempts were made to improve classification performances as discussed in section 1.4. Under the present work, an attempt has been made firstly to restore degraded images, specifically images with degradation level-1,2,3,4 of Gaussian blur, by training an ensemble of GANs and then classify the restored images using ensemble of ResNet-50 classifiers.
In this approach, training, validation and test sets are formed as Yanting Pei et al. [3] for their experiments.
The experiment conducted in this approach, provide marginal improvement in classification performances as stated in section 5.2.3. Specifically, the classification performance at 4^{th} level of Gaussian blur degradation improved from 70 percent to 71.12 percent.

Future Work

In this present work, ResNet-50 models are used for classification of restored images. In place of this, VGG-16 network may be tested as classifiers as VGG-16. ResNet-50 need very large training set as it is very complex model. To have more training data, large dataset like ImageNet may be considered for further investigation. Finally, The present methodology needs to be tested for several other kind of degradation like motion blurred, fish eye out-of-focus blurred etc. images.

References

(1) J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2009, pp. 1794–1801.

(2) J. Sanchez and F. Perronnin, "High-dimensional signature compression for large-scale image classification," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2011, pp. 1665–1672.

(3) Yanting Pei , Yaping Huang , Qi Zou , Xingyuan Zhang , and Song Wang "Effects of Image Degradation and Degradation Removal to CNN-Based Image Classification " in IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 43, No. 4, April 2021.

(4) K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. Int. Conf. Learn. Representations, 2015, pp. 730–734.

(5) K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 730–734.

(6) A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. 25th Int. Conf. Neural Inf. Process. Syst., 2012, pp. 1097–1105.

(7) Kazuki Endo, Masayuki Tanaka, Masatoshi Okutomi "Classifying Degraded Images Over Various Levels Of Degradation" in the 27th IEEE International Conference on Image Processing .

(8) Hang zhao , orazio gallo , iuri frosio , and jan kautz, "loss functions for image restoration with neural networks" in IEEE transactions on computational imaging, vol. 3, no. 1, january 2017.

(9) J. Flusser, T. Suk, S. Farokhi, and C. Hschl, "Recognition of images degraded by Gaussian blur," in Proc. Int. Conf. Comput. Anal.Images Patterns, 2015, pp. 88–99.

(10) David E. Rumelhart, G. E. Hinton and Ronald J. Williams , "Learning representations by back-propagating errors" in Nature vol 323, 9 oct. 1986.

(11) Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair† , Aaron Courville and Yoshua

Bengio “Generative Adversarial Networks” in Advances in Neural Information Processing Systems 27 (NIPS 2014).
(12) Phillip Isola, Jun-Yan Zhu, Tinghui Zhou and Alexei A. Efros “Image-to-Image Translation with Conditional Adversarial Networks” in 2017 IEEE Conference on Computer Vision and Pattern Recognition.
(13) Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin and Jiři Matas “DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks” in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.
(14) Shanjiang Tang , Chunjiang Wang , Jiangtian Nie , Neeraj Kumar, Yang Zhang, Zehui Xiong and Ahmed Barnawi “EDL-COVID: Ensemble Deep Learning for COVID-19 Case Detection From Chest X-Ray Images” in IEEE Transactions On Industrial Informatics, Vol. 17, No. 9, September 2021.
(15) Gao Huang, Zhuang Liu, Laurens van der Maaten and Kilian Q. Weinberger “Densely Connected Convolutional Networks” in IEEE Conference on Computer Vision and Pattern Recognition, 2017.
(16) J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in Proc.IEEE Conf. Comput. Vis. Pattern Recognit., 2009, pp. 248–255.
(17) Lin zhang, lei zhang, xuanqin mou and david zhang “fsim: a feature similarity index for image quality assessment” in IEEE transactions on image processing, vol. 20, no. 8, august 2011.
(18) Olaf Ronneberger, Philipp Fischer, and Thomas Brox “U Net: Convolutional Networks for Biomedical Image Segmentation” in conditionally accepted at MICCAI 2015.