
E. Poornima1 *, Suryadevara Mohit1, Kunduru Cheresh Reddy1, Vallepu Hemchandra1, Awadhesh Chandramauli2, Peram Kondal Rao3

1Department of CSE (AIML), GRIET, Hyderabad, Telangana State, India
2Uttaranchal Institute of Technology, Uttaranchal University, Dehradun, 248007, India
3KG Reddy College of Engineering & Technology, Hyderabad, India

Abstract. Remote Sensing (RS) is the process of observing and measuring the physical features of an area from a distance by monitoring its reflected and emitted radiation, usually from a satellite or aircraft. The application of RS spans a wide range of fields, including precision agriculture, disaster management, military operations, environmental monitoring, and weather assessment, among others. Haze or pollution in the satellite images, makes satellite images unsightly and makes valuable information useless. Sometimes satellites must capture images in haze-filled atmospheres, rendering them unusable for study. This proposed work is implemented using the Modern Deep Learning techniques to dehaze the satellite images. We have proposed two GAN architectures, INC-Pix2Pix and RNX-Pix2Pix. A publicly available dataset was used for training our proposed approaches. To eliminate haze from images, we have suggested Deep Generative models by employing the best developments in the field of image processing. By using generative models, images can be dehazed without information loss, supporting the paper's objective. It has the capacity to learn any kind of underlying data distribution using its learning mechanism. Therefore, it can dehaze satellite images that have been corrupted by haze using the approach automated dehazing remote sensing satellite images using deep learning models. Existing systems can be made more efficient by integrating this approach.

1 Introduction

Due to the advancements in technology and knowledge, man is constantly monitoring Earth for managing natural resources, studying climate changes, disaster management, supporting agriculture, maintaining defence, and urban planning to name a few. To study the earth, we can either do on-site observation known as in situ or study the earth remotely, known as RS. RS, in contrast to in situ or on-site observation, is the process of gathering data about a phenomenon or object without in reality contacting it physically.

A body that travels in space and circles earth is called a satellite. Satellites come in two different varieties: natural and man-made. The various shapes, sizes, and instrumentation
that human-made satellites have on-board and allow them to carry out a variety of tasks in orbit. The term 'satellite image' shall now be used to refer to all man-made captured satellite images. Our focus in this study is on data collected from the satellites by various sensors that are equipped with them. We are interested in earth imagery data. Satellite images continuously monitor the earth, collect and transmit data. As a result, huge amounts of data are generated. Poornima and team [24] presented automate the study of satellite imagery humans cannot perform the study efficiently and effectively.

The problem we are focusing on is clearing the haze present in the satellite images. The haze is caused by a combination of agricultural fire smoke, urban air pollution, and climatic conditions. To generalize, haze is a suspension of fine solid or liquid particles. Dehazing is the process of removing layers of haze present in satellite imagery. RS satellite images provide good information about the earth's features suitable for scientific research. The data is very valuable and can be used to gain a lot of insights regarding the study. But haze present in the atmosphere makes the captured data unfit for evaluation and analysis for scientific research.

Modern satellites might be equipped with specialized sensors to resolve this problem, but there are some satellites in service that do not have this specialized equipment and they suffer from haze present in the atmosphere. Also, there might be scenarios where it is inevitable to capture data in haze conditions. For example, during a forest fire, or capturing pollution-prone areas data, etc. These reasons mainly contribute to the collection of hazy satellite imagery data. Fortunately, there are various methodologies to solve this problem. This paper mainly focuses on solving this problem via a learning based methodology i.e. Deep Learning (DL) using Neural Networks (NN) in particular using GANs.

As discussed above, regarding the huge amount of data that is being collected by satellite images and their applications in many fields. There must be some robust techniques and technology acting to generate good analysis. DL with deep NN shines in this area. They facilitate efficient, fast, and required analysis for the task at hand. Also, the generated big data must be stored and mainly should be efficiently processed. NN helps in this regard. Previously, this task of analyzing satellite imagery information was done by humans. Now machines possessing Artificial Intelligence (AI) and using Artificial NN are capable of analyzing data. Authors [23] highlighted the significance of ML in prediction, pattern recognition and error reduction across diverse fields, emphasizing the impact of AI in broad domain. Thus, by enabling machines to learn from vast quantities of data and make judgments with high accuracy, AI and DL are revolutionizing the way we tackle complicated problems.

Artificial NN or ANN’s are made up of nodes called neurons inspired by the human brain and there are a series of layers connected in a sequential fashion directing, the information flow. ANN’s rely on the training data used for their learning or training purposes. They are scientific research backing the fact that they are used in Image Processing, Natural Language processing to name a few. It consists of a layer for taking input, known as the input layer, subsequent layers for processing and identifying patterns, known as hidden layers, and the final layer which holds results, known as the output layer. A popular version of ANNs known as Convolutional NN (CNN) are popular choice for image processing tasks due to its efficiency and effectiveness.

Finally, this solution can be integrated into existing RS application systems which can make use of the precious satellite imagery for exploration and exploitation. This approach can eliminate the extra cost and effort in solving the haze problem using sensors that use traditional methods of estimation of climatic parameters to dehaze images. This learning-based approach based on neural networks, specifically GAN [7] and [21] when applied to satellite imagery can capture quality content in the haze environment. The problem is famously known as Single Image Dehazing (SID) because there is only one ground truth
image available for the approach to dehaze the image. Due to this limitation and for better generalization, learning-based methods shine better and are thus incorporated. For a learning-based algorithm, we need the dataset for automated dehazing images presented in either a paired or unpaired manner. This study can also be effectively used in Unmanned Aerial Vehicles (UAV) domain as well. Authors [22] discussed enhancing farming practices and crop yields through Smart Farming through drones for tasks like soil monitoring and livestock management, revolutionizing traditional agriculture.

2 Literature survey

Huang and his team [1] proposed a model that performed Dehazing multi-sensor Optical Satellite Imagery using a cGAN. This work has implemented a cGAN in which the generator uses the encoder-decoder network with dilated blocks and skip connections and the discriminator was built using 70x70 PatchGAN. In this generator setup, SAR image information was passed. A new synthetic dataset SatHaze1k compromising 1200 images of data was proposed, due to the difficulty in getting real-time data. X. He and his team [2] proposed a work that illustrated Contrastive Learning based cycleGAN to dehaze optical satellite images using unpaired datasets input and target data. In this study, asymmetric contrastive cycle GAN (ACC-GAN) was used to dehaze optical satellite images along with a transfer network to form an asymmetric structure.

A. Hu and his cohorts [3] framed an unsupervised haze removal method for Optical RS Images by improvising existing GAN architecture. The state of art cycleGAN was improvised and used to achieve the motive. An additional Edge-Shaping loss function was introduced apart from the existing ones in a cycleGAN network. X. Chen and Y. Huang [4] proposed Dehazing Single RS imagery using Memory-Oriented (MO) GAN. The objective of this work was to use MO unpaired learning to dehaze Optical Remote Sensing Imagery using GAN. The proposed model has a generator that is made up of a memory module (LSTM) and a U-Net auto-encoder with a discriminator. A dual region discriminator has been constructed to perform better on variable haze levels in photographs for greater differentiation.

X. Sun and J. Xu [5] suggested a method to dehaze RS using Cascade GANs. This paper aimed to create synthetic data similar to real-world data by using a GAN, learning to dehaze the former using another GAN, and finally cascading both. Two GANs were cascaded together at end. The first GAN is built upon convolutional layers and was used for learning to haze, and the second one has a self-attention module built on top for learning to dehaze. Y. Zheng and his colleagues [6] proposed an enhanced attention guide (AG) GAN using unpaired remote sensing data for dehazing. The goal of this paper was to dehaze remote sensing photographs. Using the AG-GAN network with unpaired remote sensing data. The state of art cycleGAN was used to achieve this task with the attention guide mechanism. An additional loss function apart from the traditional CC loss, total variation loss was adopted.

Xianhong Zhang [7] proposed a GAN based on texture attention for dehazing RS images. The objective of this paper was to optimize the existing architecture by adding the attention framework to dehaze RS images. A texture attention generator was built using Cellular Neural Networks (CNN). To increase the quality of the provided results, both global and local discriminators were utilized, combined with the power of the texture attention mechanism. Darbaghshahi and his companions [8] proposed a framework for cloud removal in RS images using GAN and SAR-to-Optical translation. This paper's goal was to remove clouds from RS photos firstly by converting the SAR image to Optical Image and then using a GAN to reconstruct the corresponding cloud image. The work used
two GANS, one for SAR-To-Optical and the other for cloud removal. Both the GANs follow encoder-decoder architecture with Dilated, Residual, and Inception modules.

Zhao and his group [9] have proposed a model for RS image dehazing. The work's objective was to design a GAN based on Attention Encoder-Decoder (AED) for RS image dehazing. A GAN was developed in which the generator was based on AED with skip connections, enhance module, distillation module and local skip connection were used in the generative network for extraction features. The discriminator was the Markovian discriminator network. Wang and his crew [10] have proposed a SI Dehazing. A methodology using twofold GAN. The proposed work has two sub-models. One for generating haze similar to the real world and the other for removing haze trained up on the former model. Dong and his group [11] proposed a fusion-based discriminator GAN for SI dehazing. The objective of this work was to dehaze the images using GAN-based architecture by employing a fusion discriminator that takes into account the image's frequency information. Wang [12] proposed an image dehazing methodology using modified cycle GAN via Spectral Normalized (SN) Soft Likelihood Estimation.

Engin along with his cohorts [13] proposed an improvised version of cycle GAN for SI dehazing. Cycle GAN architecture was used and a new loss function was added named perpetual loss function alongside the existing cycle consistency loss. Qu along with his crew [14] proposed an enhanced Pix2pix model for dehazing images. The goal of this effort was to address the dehazing problem through ITI translation. The generator was based on the pix2pix model with enhancing blocks and a multi-scale discriminator was used. Chaitanya and S. Mukherjee [15] proposed SI Dehazing by improvising existing cycle GAN architecture. The generator network was an encoder-decoder-based architecture with AOD-net and the discriminator. Authors [25] emphasized the significance of feature selection in classification for accuracy and efficiency. It investigates combining features from different methods, demonstrating improved precision, contingent on dataset, algorithm, and metrics used. Image restoration is to enhance images by removing noise and restoring them to their original quality. The present approach explored various methods in both frequency and spatial domains, followed by analysing their performance using simulations [26].

3 Proposed method

As mentioned in the previous chapters this work focuses on automated dehazing the satellite images using generative models. Out of them, GANs were selected for this work due to their high usage and acceptability due to their performance. Basically dehazing is a critical task in remote sensing, as haze degrades the quality of satellite imagery, leading to a loss of information thus making the most valuable RS data useless. But employing traditional methods does not scale and needs a lot of human effort to eliminate the haze from the satellite imagery. Worst case scenario, they may even try to miss our crucial information or reconstruct the images with missing and irregular information.

After the work of Ian Goodfellow along with his cohorts [16] GANs came into the picture and further advancements in them gave a new rise to new generative models which gave convincing results in many applications. Considering this study approach GANs performed well in the image-denoising applications which can be correlated with this approach of automated dehaze the image. Haze can be substituted as image noise. GANs use adversarial training where the two networks of this architecture are trained to outperform one another. In the context of this study after the GAN is trained the generator network will be able to dehaze the haze-filled satellite images. The generative power of a GAN comes from its generator reaching its optimal state due to the discriminator.
GANs are capable of learning the patterns of the underlying imagery data. Here satellite imagery data distribution is learned. Thus GAN is a good option to use for automated dehazing satellite images due to this property. The GAN networks used are based on CNN as the networks work on image data. CNN, is tuned for capturing relevant information from the image this is mainly achieved due to the convolutional layers present in the images. Also, CNNs are capable of recognizing the details in the image regardless of spatial position. That is the reason CNNs have been a popular choice for image processing tasks. Since the final outcome of this work is based on real-world data and CNN are good at adapting and learning to the real-world data, they can be collectively used for our task.

Both GANs and CNNs are effective image processing tools, but they have distinct advantages and disadvantages. Our task is ITI translation task and it entails translating the haze image to a clear image. Due to the adversarial training GANs will perform better than CNNs. GANS are specifically tuned for image generation here generating the dehazed version of the haze satellite image. The generator network in GAN is made in such a way to generate an image capturing crucial information guided by adversarial training using another network known as discriminator. But in CNN, the learning approach utterly depends on the layers and the optimization framework it has adopted.

Finally in the Single Image (SI) dehazing task, for a particular point of observation there is only one image available. Any learning approach needs the following data. It needs the haze images as well as the clearer images, either in paired fashion or unpaired fashion, discussed later. Then the concept of image-to-image translation mentioned comes into the picture. The process is translating the haze image into a clear image. Our proposed architecture can solve this problem using the above mentioned features effectively. This work exploits the concept of image-to-image (ITI) translation. The objectives of the paper are as follows:

• To Reconstruct Haze images ITI translation methods using GANs.
• To construct a Pix2Pix variant using the inception blocks called INC-Pix2Pix.
• To construct a Pix2Pix variant using the ResNeXt blocks, called RNX-Pix2Pix.

3.1 Architecture diagram

In the case of this work, the architecture diagram figure 1 shows how different components and modules of the proposed work would fit together to accomplish the task of dehazing the satellite images. Initially, the diagram starts with the data pre-processing after the suitable dataset has been identified, which is the collection of clear satellite images and haze images from publicly available data sources. A publicly available dataset consisting of clear satellite images and haze images is used, which will be discussed later. The data source is open for research use and is thus incorporated into this work. Data has been normalized so that no part of the data dominates one another and the training process completes smoothly with good generalization.

3.1.1 Data pre-processing

Publicly available dataset for research is used for this proposed work. The dataset consists of haze satellite images and clear satellite images. Firstly, corrupted files are identified and removed from the dataset for further processing. Later, the images are resized to 256x256 dimension scale. Then, after the images are normalized, which ensures all the images and all the pixels in the images have similar magnitude, which prevents certain data from dominating the learning process and overall results. Apart from, it also enables the proposed GAN architectures to train at a faster pace and avoiding overfitting problem. Then the further processing is done.
3.1.2 Data preparation

The dataset used in this study contains paired haze and its corresponding ground truths. This is alright for the conditional GAN based approaches but for the cycle GAN based method the data has to be shuffled, which is done in this stage. Tensorflow Dataset pipeline is used for data preparation. It has been utilized since it provides high-performance boost and has many convenient options to manipulate and use according to the application in hand. It has the capacity to work with the data in batches and seamlessly integrate with the high end computing infrastructure like Graphical Processing Units (GPUs) or Tensor Processing Units (TPUs). The dataset was configured to pre-fetch as per the system requirements and available resources to make the training speed faster.
3.1.3 Model construction and training

Generative models are known for their ability to learn from the underlying data present in the images. They figure out the underlying image's probability distribution. Consequently, data generation is facilitated. Due to this robust feature, they are useful for various applications in generative AI. They are used in many applications, starting from image processing to natural language processing. Some common and most useful features are discussed below. They are now being extensively used in image and video processing and synthesis. They are used for generating multimedia files based on the requirements. The generative models have the capacity to generate new videos and images based on the requirements.

This study is using this application where the satellite image is dehazed based on the ground image on which it is trained and tested. The model learns from the underlying clear images and learns how to generate new clear images. There are various types of generative models, but due to the advent of generative models [16] are being extensively used in many applications discussed so far. As discussed, GAN is made up of two primary networks: generator & discriminator network trained adversarially using adversarial training. The discussion further is based on the [17] network on which the proposed methods have been based upon.

Pix2Pix is a Conditional GANs (cGANs) allows to condition the generated outputs on specific inputs. In cGANs, the generator is conditioned on some data, which might be a class label, a text description, or an image. Because of this conditioning, the generator generates more realistic and diversified outputs that are consistent with the input. The generator network takes two inputs: one noise vector and a conditional input and produces a synthetic output. The discriminator also takes two inputs: the synthetic output and the conditional input and determines whether the output is real or fake. A conditional GAN based on the Pix2Pix network model is developed in this work known as INC-Pix2Pix and RNX-Pix2Pix. The changes or improvements were made in the generator of the generator network of the original pix2pix.

Firstly, INC-Pix2Pix in the generator the U-Net variant to which in the contrasting path inception blocks have been added in the parallel. We have added Inception-blocks in the down-sampling path, which helps the network to learn more features. Networks with inception blocks can efficiently extract multi-scale features. To this we have added inception blocks. Inception blocks are known for capturing more information at a given level. Basically, U-Net has already explored the depth which is responsible for learning hierarchies and other complex features. Now, due to the addition of the inception block, another aspect width will also be explored which is responsible for leading to learning more information and providing higher quality information or representation of input at any given level than usual.

The network may gather data at various spatial scales by employing numerous filter sizes at a level without considerably raising the computational cost. Also, an Inception block's concurrent application of filters of various sizes enables the network to learn a wide variety of characteristics. This gives the model more expressive capacity and enables it to recognise intricate patterns and data structures. Inception blocks make it easier to move information between network layers by concatenating multi-level feature maps from various filter sizes. The architecture of the inception block that we have used is taken from the original work P. Isola, along with his team [18] as shown in figure 2.

Our next proposed variant RNX-Pix2Pix is constructed by adding res-next blocks instead of vanilla residual blocks in the generator of the original work. ResNeXt blocks proposed by S. Xie and his cohorts [19]. The network in ResNeXt is segmented into several parallel routes, a phenomenon known as "cardinality." Each route is made up of a number of transformation layers that take the input data's characteristics and refine them.
Splitting the input data into distinct branches enables each branch to develop a unique representation. Next, the convolutional layers found within each branch are referred to by the transform operation. These layers alter the input data using a number of convolutional filters to capture and change the features at various spatial scales and levels of abstraction. The outputs of various branches can be concatenated or added together to execute the merge procedure.

Through the process of merging, the network is given access to the different perspectives and collective knowledge of the parallel pathways, improving its capacity for representation. We have utilized one of the equivalents ResNeXt blocks that is mentioned as shown in Figure 3.

**3.1.4 Model evaluation**

Model evaluation is the process of assessing a NN model's performance on a certain dataset. The purpose is to find any flaws or limits in the model as well as to assess how well the model generalises to fresh, untested data. To make sure that the model is not over-fitting
the training data and performing well, it is necessary to assess the model's performance on several subsets of data. As a result, the model will become more durable and facilitate a seamless interaction with the RS application pipeline for Geo AI applications. For evaluating the quality of reconstructed or compressed pictures or multimedia, the PSNR (Peak Signal to Noise Ratio) statistic is frequently used. It determines the best possible signal value to the noise added during the reconstruction process in relation to that signal. PSNR is measured in decibels (dB) and provides a quantitative measure of the distortion introduced during the reconstruction or compression process. A lower level of distortion is indicated by a greater PSNR number, which results in a higher-quality image or video. PSNR is calculated as follows:

$$PSNR = 10 \log_{10} \left( \frac{MAX^2}{MSE} \right)$$

where $MAX$ is the maximum possible pixel value and $MSE$ is the mean squared error.

**Fig. 4.** Dehazing images using INC-Pix2Pix.

The SSIM statistic is frequently employed to assess how similar two images are. It is meant to assess the structural information similarity between the reference picture and the distorted image rather than relying just on pixel-level alterations Wang and his crew [20]. Luminance, contrast, and structure are used to determine the three components that make up the SSIM index. The texture or pattern difference between the two photos is measured by the structural component. Thus, basically SSIM is unique in that it is based on how the
human visual system perceives images and takes into consideration the fact that changes in structure information are more perceptible to humans than changes in pixel values.

Fig. 5. Dehazing images using RNX-Pix2Pix.

4 Results and discussions

A dataset known as Haze1k [1] is used in this work. It contains the images with different haze levels, namely Thick, Medium and Thin. Performing the Model training and testing on non-uniform haze conditions will help the model to generalize well to the real-world conditions. Researchers and practitioners may fully manage the data production process with synthetic datasets, allowing them to adjust things like object location, lighting, backdrops, and item variants. Compared to the traditional method of using statistical and mathematical models to dehaze the images, this method of using GAN is easy and adapts easy to the real-world data. Wherein the traditional methods involve in finding the atmospheric metrics and needs lot of parameters and tuning to dehaze the satellite imagery. We know that satellites collect lots of data and are known to be Big Data, employing the DL based models is more advisable due to its adaptability.
GAN can be used to improve the generalization ability of computer vision applications. This can help to improve the accuracy and reliability of computer vision models when they are deployed in the real world. GANs have a lot of advantages, including the capacity to produce realistic and varied data, the capacity for unsupervised learning, the possibility for creative uses, and the capacity to be utilized for transfer learning. GANs are advantageous tools for a variety of such applications in numerous industries. Both the proposed works, INC-Pix2Pix and RNX-Pix2Pix are constructed as specified in the previous section trained on the Haze1k dataset. The results of the two approaches are illustrated in Figure 5.

One such application is image denoising and automated dehazing that comes under the image-to-image translation task. This fact is used for this study. The main technique that is used here is Image-to-Image translation. The proposed work is fully relied and, based on this fact, ITI mappings here mapping from haze images to clear image. Table 1 shows the performance analysis of various approaches along with ours, since all the other models do not release the code, we have taken the data for results from [4].

Table 1. Comparison of our models with the existing approaches.

<table>
<thead>
<tr>
<th>Model</th>
<th>Thin Haze</th>
<th>Moderate Haze</th>
<th>Thick Haze</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SSIM</td>
<td>PSNR</td>
<td>SSIM</td>
</tr>
<tr>
<td>DCP</td>
<td>0.724</td>
<td>12.77</td>
<td>0.573</td>
</tr>
<tr>
<td>DehazeNet</td>
<td>0.895</td>
<td>19.75</td>
<td>0.855</td>
</tr>
<tr>
<td>SAR-OPT-CGAN</td>
<td>0.841</td>
<td>20.19</td>
<td>0.794</td>
</tr>
<tr>
<td>B.Huang</td>
<td>0.906</td>
<td>24.16</td>
<td>0.926</td>
</tr>
<tr>
<td>SkyGAN</td>
<td>0.924</td>
<td>25.38</td>
<td>0.903</td>
</tr>
<tr>
<td>H2RL-Net</td>
<td>0.929</td>
<td>25.84</td>
<td>0.928</td>
</tr>
<tr>
<td>Mo-GAN</td>
<td>0.926</td>
<td>25.47</td>
<td>0.9176</td>
</tr>
<tr>
<td>RNX-Pix2Pix</td>
<td>0.921</td>
<td>22.27</td>
<td>0.928</td>
</tr>
<tr>
<td>INC-Pix2Pix</td>
<td>0.954</td>
<td>25.55</td>
<td>0.949</td>
</tr>
</tbody>
</table>

5 Conclusion and future enhancement

After discussing the use cases of remote sensing and the importance of satellite imagery, it is understood that the information they carry plays a huge role in the application that is being observed and solved. Satellites revolve around the earth continuously generating lots of data. Therefore, to manage, maintain and process, there is a need for automation. For that automation to work smoothly, there is a need for the good quality of data. This study tries to dehaze the satellite imagery that holds crucial information about a place of study for further analysis needed by the application. Solving the dehazing using DL using NN is a very good approach. This study is performed using the GANs. With the introduction of GANs, numerous applications were quickly handled.

The implementation of the networks are extended on the original based paper [17]. They are built using the same architecture with the addition of new blocks. For Pix2Pix first, we have added ResNeXt based blocks instead of the plain residual blocks in the Generator and for other variants, we have added Inception blocks in parallel, similar to the convolution blocks of the U-Net. The performance metrics have clearly shown the reliability of using methods mentioned. The proposed methods which have performed better compared to the existing works. These days, satellites are equipped with multiple sensors capturing multi-spectral and hyper-spectral images. Working with that data gives a lot more challenges and opens the door to other untapped opportunities. This work is basically based on single image dehazing. If there are multiple images of the same place at multiple instants of time, many other new insights can be drawn. This will significantly impact the data volume and data variety, thus leading to more generalized results, which
may subsequently improve the image reconstruction quality considerably for enhancement. Therefore, in the future, we intend to work with temporal based models and multi-sensor data.

References

2. X. He, W. Ji and J. Xie, IEEE Access 10, 67316 (2019)


24. E. Poornima, M. Bala Anand, Agrawal, R. Kumar, Dhingra, Mallika, Asaad, Renas, Jumani, Awais, Multi. Tools Appl 6, 15086 (2023)
