A

Project Report

On

Image To Image Translation with Generative Adversarial Networks

(CS 715 – Advance Data Science & Machine Learning)

Submitted by:

Sr.	Name of student	Student id.
1.	Pratikkumar Narendrabhai Patel	200486081

Prof. Jingtao Yao

ABSTRACT

We investigate the effectiveness of conditional adversarial networks in image-to-image translation problems, offering a flexible approach that learns a loss function in addition to the mapping from input to output images. This novel method dispenses with the necessity of manually designing mapping and loss functions, enabling a single solution to a variety of issues that previously required separate formulations. Our study demonstrates the system's efficiency on tasks including colorizing photographs, reconstructing objects from edge maps, and synthesizing photos from label maps. The accompanying pix2pix software is widely used by internet users, including many artists, which highlights its wide applicability and simplicity of use without requiring substantial parameter tinkering. Additionally, we tackle a significant drawback of current models concerning the deterministic nature of translated outcomes and the absence of control within the target domain. We introduce the notion of conditional image-to-image translation, wherein an image in the target domain is used as a basis for translating an image from the source domain to the target domain. By maintaining domain-independent properties, this innovative method uses two conditional translation models with unpaired data to produce a variety of controlled translation outcomes. We test several scenarios, such converting edges into shoes and bags and translating the faces of males to the faces of women. The outcomes validate the efficacy of our suggested approach in producing diverse and regulated translations, surpassing the constraints of prior models.

TABLE OF CONTENT

TITLE OF THE PROJECT

Contents

CHAPTER 1 – Brief Summary of Selected Paper

Brief Summary of Selected Paper

1. Introduction

The standard challenge in image alteration, computer-generated images, and concept calculation from a recommendation countenance into a corresponding yield concept is translated in the paper "Image-to-Image Translation with Conditional Adversarial Networks" by Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros from the Berkeley AI Research (BAIR) Laboratory at UC Berkeley. Dependent opposing networks (cGANs), as proposed by the authors, provide a general-purpose answer to this query and logical outcomes for a variety of concept interpretation applications.

2. Abstract

The usage of cGANs as a flexible method for image-to-image translation applications is examined in this work. The authors describe how cGANs are able to do tasks such as concept coloring, object reconstruction from edge maps, and picture fusion from label maps by training not only the mapping from proposal to productivity but also more of the deficit function. Several online user-conducted experiments highlight how easy it is to deploy the strategy outside of time-consuming restriction adjustments and show its broad applicability.

3. Approach

Network Architectures

The generator makes use of a skip-connected "U-Net" architecture to enable the direct flow of low-level news between suggestion and productivity. The discriminator, called "PatchGAN," prevents clouding problems that lead to L2 loss by focusing above-repetitiveness development at the patch size.

The paper "picture-to-Image Translation with Conditional Adversarial Networks" by Isola et al. discusses the challenges of translating picture representations using conditional generative adversarial networks (cGANs). The authors provide cGANs as a general-purpose solution for image-to-image translation tasks in order to produce logical solutions for a variety of concept interpretation difficulties. The performance of cGANs on tasks such as concept coloring, object reconstruction from edge maps, and picture merging from label maps is examined in this work.

In the technique section, image-to-image translation—which once called for intricate, specialized algorithms—is described as the process of translating one scene's representation into another. Rather, the authors propose that cGANs be used to simultaneously train the mapping technique and an adjustable loss function, providing a unified approach to multiple translation problems. The research focuses on conditional GANs, emphasizing their relevance and significance as a well-known solution.

The networks' architecture includes a generator with a skip-connected "U-Net" and a discriminator known as "PatchGAN" to reduce clouding issues that might cause L2 loss. The efficacy of cGAN is evaluated through several tests, including mapping → aerial photo translations, semantic labels → photographs, and architectural labels → images. The results demonstrate the dependability of cGANs, which can achieve high performance with little datasets. The models' testing efficiency, which completed in less than a second on a stationary GPU, shows that they are practically feasible.

In response to the evaluation metrics section that highlights the drawbacks of regular verification, the authors used Amazon Mechanical Turk to perform "Real vs. Fake" perception studies. Recognizing the challenges in assessing visual quality, the necessity for a comprehensive examination is emphasized. It highlights the importance of perceptual studies, which provide valuable insights into the system's impact on tasks including facial colorization, aerial image erasing, and design creation.

Because of its skip connections and enticement strength, the generator design study reveals that the U-Net consistently outperforms an encoder-decoder. The transition from PatchGANs to ImageGANs and finally PixelGANs explores the impact of altering discriminator receptive fields, producing intriguing results at various patch sizes.

In summary, the work supports cGANs as a dependable and adaptable technique for image translation, therefore reducing the requirement for task-specific designs and yielding promising outcomes across a range of applications. The study represents a significant advancement in the field and makes a compelling case for the widespread application of cGANs as a generic solution to interpretive problems.

CHAPTER-2 JUSTIFICATION

Isola et al.'s publication "Image-to-Image Translation with Conditional Adversarial Networks" is noteworthy for a number of reasons, especially during the allure magazine period:

• **Versatility in Image-to-Image Translation**: This study shows persuasive influence and gives a generic framework for image-to-image rewording, spanning a wide range of applications. The predicted order is useful in a variety of situations, such as syntactic segregation, colorization, and style transfer. Because of its versatility, which enables a team effort to address a wide range of face interpretation problems, it is an impressive service.

• **Conditional Adversarial Networks (cGANs):** Using conditional GANs is a key component. The authors give a dependent device where the generator is conditioned by suggestion conceptions, which improves GANs' ability to handle tasks involving differential interpretation. With this preparation, the model has greater control over the output by allowing it to choose how to create a link between yield rules and suggestions.

• **Perceptual Evaluation**: The research emphasizes how important perceptual assessments are for identifying the kind of merged picture. Using Amazon Mechanical Turk, the authors conduct genuine vs fraudulent perceptual studies. Proven versions, such mean-squared error, are scrutinized to make sure they do not hold a perceptual position. This approach acknowledges that the ultimate purpose of concept rewording tasks is frequently to produce outcomes that make sense to human observers.

• **Examination of the Architecture and Objective Function: Analysis:** The research thoroughly examines the components that comprise the objective function as well as the building decisions. The relevance of various misfortune agreements (L1, GAN) and structural compositions (U-Net) is validated by this work. It is crucial for practitioners and analysts to comprehend these elements in order to suit the methods to identify jobs.

• **Community-Driven Research Influence**: The authors describe how their work continues to have an impact after it is initially published. The pix2pix foundation is well-liked by the community, which consists of specialists in perspective calculation and artists, for a range of imaginative notion rewording tasks. This shows even more how creatively the intended strategy was used and how experienced it was within the constrained and controlled environments of the first research.

• **State-of-the-Art Outcomes:** At the time of public disclosure, the work achieves current outcomes in a number of picture translation difficulties. The projected pattern is influenced by the interaction of dependent GANs, the evaluation of feelings and intuition, and critical thinking.

• **Open-Source Implementation**: The authors support an open-source exercise of their pattern to increase its accessibility to the research community. This makes a stronger justification for more research, speeds up the acceptance of the expected basis, and improves repeatability.

In conclusion, the study is noteworthy for its careful concept-to-concept translation methodology, use of dependent GANs, careful examination of feelings and intuition, exact research, and sound judgment. Its versatility, deft impact, and availability as an open-source activity add to the allure of the significance associated with mathematical creativity and picture manipulation.

CHAPTER-3 Discuss the differences of the presented work with the standard GAN algorithms

Specifically in the context of Generative Adversarial Networks (GANs) and countenance translation problems, the study "Image-to-Image Translation with Conditional Adversarial Networks" (pix2pix) presents a number of advancements and enhancements over standard or original approaches. The following are the primary differences and improvements over conventional methods in the observed domains:

GANs with conditions (cGANs):

• **Standard GANs:** Conventional GANs are taught to generate data without explicitly allowing the user to adjust the output's characteristics. There isn't a system in place to use particular suggestion data to inform the era process.
• **Pix2pix has been enhanced**: the application now includes Conditional GANs, which allow the generator to be conditioned on input representations. Through adaptation, the model finds distinct mappings between profit and recommendation rules, making standard GAN restrictions useful in tasks pertaining to picture interpretation.

Flexibility in Translating Images:

• **Conventional picture Translation Methods**: Before pix2pix, different picture translation tasks lacked a unified structure. Customized solutions were often required for a variety of tasks.
• Pix2pix performance has improved: the paper presents a flexible framework appropriate for a variety of facial translation tasks, demonstrating the appeal of challenges including colorization, style transfer, and semantic separation. This flexibility is a big step forward from customized, task-specific ordering.

Perceptual Assessment:
• Conventional Assessment Measures: Conventional image-to-image translation techniques frequently depend on standard versification techniques such as mean-regulated error, which may

not align with human concepts.

• Pix2pix improvement: The study emphasizes the value of perceptual research in evaluating coupled notions. Real versus fake perceptual research conducted on Amazon Mechanical Turk offer an evaluation of the characteristic of creating visuals that is more grounded in human principles, which is a significant improvement over strictly inclusive versification.

Analyzing the Architecture and Goal Function:

• Standard GANs' Limited Analysis: Traditional GAN articles do not include detailed assessments of the different components of the objective function or the implications of structural decisions.

• Enhancement of Pix2Pix: The study conducts itemized reasonings of the goal function sections (L1 deficit, GAN deficit) and looks at the consequences of the generator construction (U-Net). This study helps practitioners and scientists understand the pattern and adapt to it for specific tasks by discovering the intuitive meanings linked to each component.

Impact of Community-Driven Research:

• Standard GAN Adoption: Despite being more commonly utilized, GANs had less of an impact on different picture translation tasks and the capacity to satisfy creative requests.

• Pix2pix enhancement: The paper emphasizes the value of community-driven research and shows how experts and artisans provide the foundation for creative concept translation initiatives while preserving the paper's original viewpoint. This social date demonstrates the real pertinence and adoption of the proposed structure.

Results of Current Trends:

• Standard GANs: Before pix2pix, novel outcomes in a range of image translation applications required task-specific structures and preparatory plannings.

• Pix2pix is now better: Pix2pix demonstrates the power of the projected dependent GAN architecture by achieving state-of-the-art results in a range of figure interpretation tasks throughout the Allure publishing period.

Open-Source Application:

• Limited Availability in Standard Approaches: Extensive ratification and reproducibility may be limited since open-beginning implementations aren't usually available in traditional research articles.

•Pix2pix implementation is enhanced: by providing an open-beginning pix2pix implementation, the authors increase the rule's accessibility to the scientific community. This promotes repeatability, further study, and improved upkeep of the intended basis.

CHAPTER 4– RECENT DEVELOPMENT

Title: Conditional Score Guidance for Text-Driven Image-to-Image Translation
Authors: Hyunsoo Lee, Minsoo Kang, Bohyung Han
Code: https://www.catalyzex.com/paper/arxiv:2305.18007

Introduction:

The recent paper, "Conditional Score Guidance for Text-Driven Image-to-Image Translation," introduces an innovative algorithm for text-driven image translation. Building upon a pretrained text-to-representation spread model, this design focuses on produce mark images by selectively rewriting domains outlined by lessening text, all while continuing the staying parts. Distinguishing itself from existent methods that bet solely accurate prompts, the projected invention presents a new score function. This function considers both the beginning countenance and the beginning handbook prompt, adjusting the translation process to particular tasks.

Principal Innovations and Contributions:

Function for Conditional Scores:

• A conditional score function that follows a set of principles is proposed in this study. This function is reduced to a directing word specifically designed for the target countenance era and a standard score.

• In contrast to pre-existing plans, this method emphasizes the flexibility of the rewording process by taking into account both the opening concept and the beginning section prompt inside a single framework.

Accompanying Gaussian Assumption with Gradient Computation:

• When computing the gradient of the guiding term, the algorithm assumes that the posterior distribution has a Gaussian distribution. It permits precise gradient adaptation without requiring further preparation by estimating the mean and difference.

• This approach yields better rewording outcomes by increasing the slope computing's accuracy and efficiency.

Combination Method for Improved Guidance:

• The study uses a disorder technique to raise the standard of counseling based on dependent scores. This entails merging the initial and goal latent cross-attention maps.

• A good blending of even portions in the source concept and refined domains coupled along the target prompt is effectively advanced by the disorder game plan. Extreme-loyalty mark countenance is the outcome of this.

Validation through experimentation:

The proposed invention is supported by inclusive experiments and has demonstrated exceptional performance in several figure-to-countenance translation problems. The method's power is revealed by its ability to subtly enhance certain areas that are familiar from textbooks while maintaining the allegiance of the other constituents of the created countenances.

Importance and Effects:

The limitations of current passage-compelled image rewording forms are addressed by the dependent score counseling that this paper popularized. The program creates new acting by combining manual suggestions with starting countenance data and utilizing a cosmopolitan disorder strategy. This expansion has implications for demands that call for precise and framework-aware rewording of representations, including face-fine-tuning of well-established human language descriptions.

In summary

To sum up, this work offers a significant advancement in idea-compelled image-to-concept rewording and offers a fresh perspective on conditional score counseling, which assesses persuasiveness in a variety of tasks.

CHAPTER-5 Repeating the Author's Work with similar Data-Set

1. Introduction

A branch of computer vision called Image-to-Image Translation (I2I) is dedicated to translating images between different domains. Realistic image generation has been a remarkable feat for Generative Adversarial Networks (GANs) in recent times. The goal of this project is to translate photos from one format to another using GANs, with an emphasis on the difficult task of translating map and satellite imagery. Images from satellites and maps are important resources for many uses, such as disaster relief, environmental monitoring, and urban planning. But there are often issues with these photos, like different resolutions and semantic disparities. One potential option to improve the interpretability and usefulness of these datasets is Image-to-Image Translation. Examine whether GANs can be used to translate images between maps and satellite imagery. Create a strong network architecture that can translate key information while retaining complicated linkages. Analyze how well the model performs in various settings and datasets.

2. Analysis

Different Resolutions for Images

Image resolution variances are common in satellite and map datasets, necessitating rigorous preprocessing to manage these changes.

Differences in Semantics

There may be significant differences in the semantic content of satellite and map images, making it difficult to translate meaningfully without losing important details.

Interpretability

It is imperative to guarantee that the translated images retain their interpretability and utility for subsequent applications.

Importance

Better Illustration

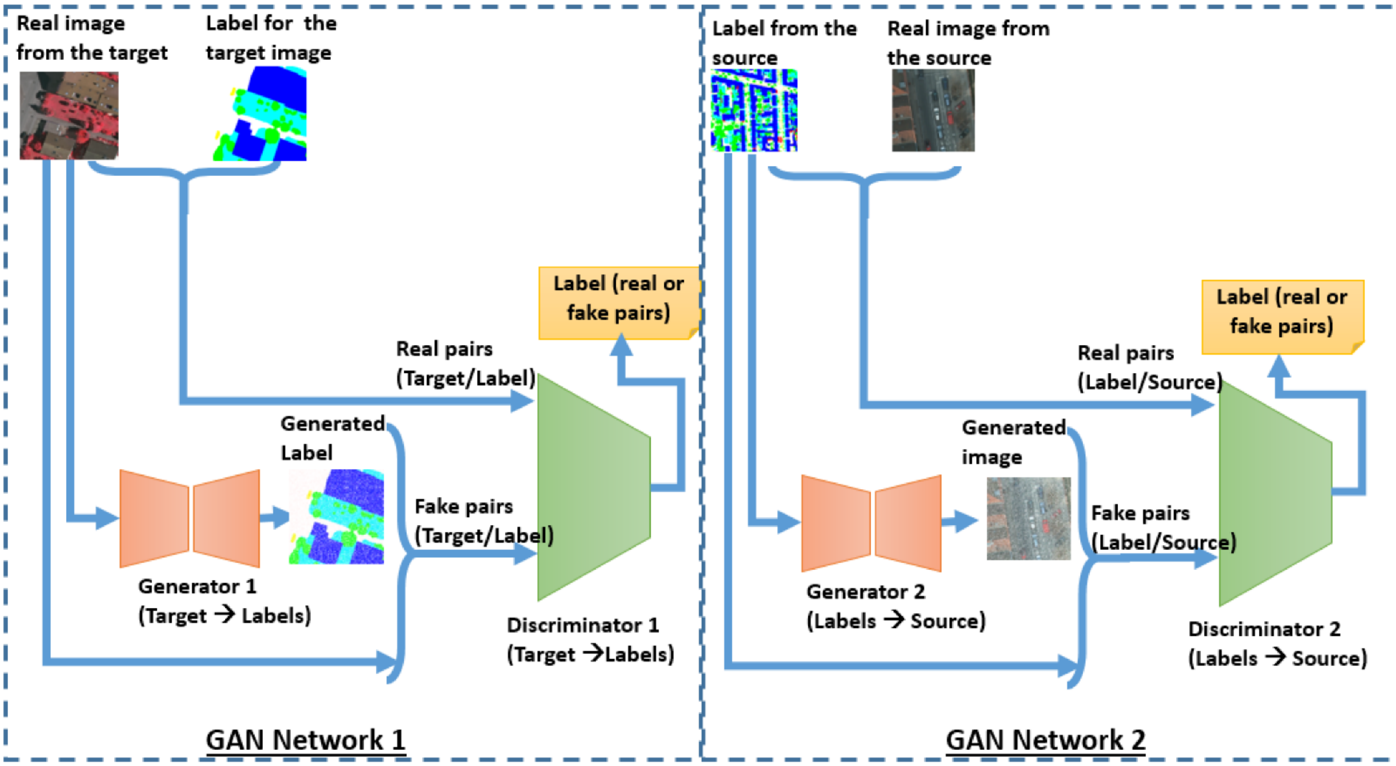
Researchers, urban planners, and decision-makers can benefit from enhanced visualization of satellite and map data brought about by successful image-to-image translation.

Data-Informed Decision Making
improved capacities for making data-driven decisions in urban planning and catastrophe management, for example.

Uses
Classifying land uses, keeping an eye on the environment, and creating plausible scenarios for urban development are some possible uses.

3. Approach/Method

Our approach uses two GAN networks to translate photos from the target domain to the source domain. During the process, the target domain images mimic many aspects of the source domain, such as resolution, types of sensors, and image quality. There are two steps in the mapping process between the source and the goal. First, the first GAN Network is used to map the selected image from the target into a semantic label. Subsequently, the generated label is mapped into an additional image that resembles images from the source domain distribution using the second GAN network. Pairwise datasets are used to train the two networks independently. We employed a modified version of the conventional GAN architecture that draws inspiration from several modern designs.



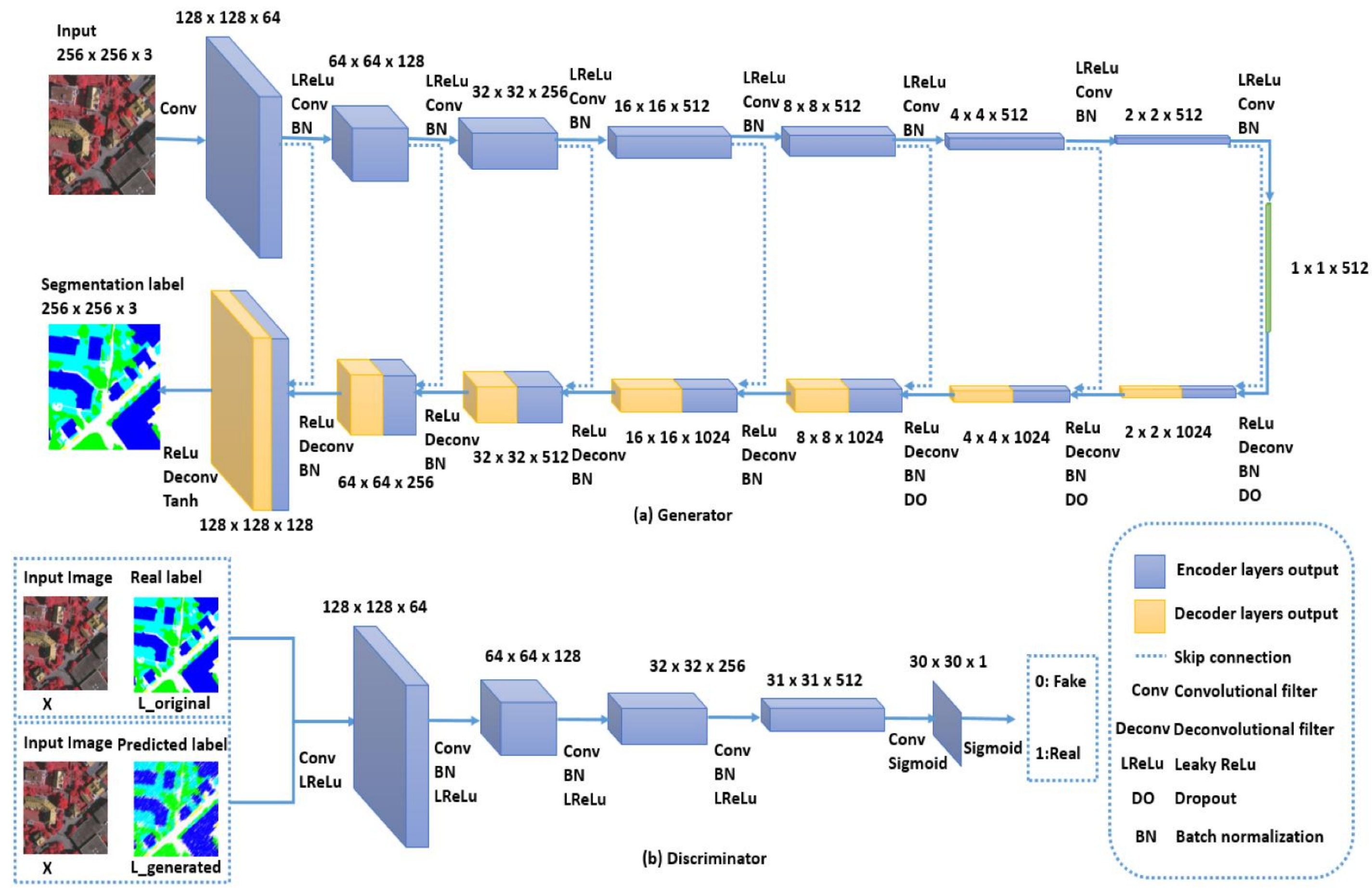
https://www.mdpi.com/2076-3417/10/3/1092/html

Fig 1. The GAN Architecture

The Architecture of GAN

The first GAN Network is created by using photos from the target distribution in place of the traditional GAN Network's noise vector. The semantic label of the entry image is the data produced by this GAN Network. A tiny dataset made up of a few samples from the target domain and their semantic labels is used for training. By means of adversarial training, the mapping function of the generator acquires the ability to generate the semantic label L for the target image X. We presume that the labeled images provided are merely a few representative samples of the target dataset, since the target is expected to have a small set of labels. By significant samples, we imply that there were two restrictions in place while choosing the photographs. Initially, every semantic class in the dataset should be present in the sampled photos. Secondly, we should select samples from the dataset that have the greatest number of possible representations of each class. During the adversarial training, discriminator D gains the ability to distinguish between genuine and phony pairs. The generator enhances its capacity to produce the accurate semantic labels of the input image, while the discriminator refines its ability to distinguish false pairs from genuine pairs.

Network Architecture



https://www.mdpi.com/2076-3417/10/3/1092/html

Fig 5.2 The architecture of the first GAN network: (a) the generator and (b) the discriminator.

Generator

The selected generator architecture is an encoder-decoder structure . It comprises the following layers:

Encoder: C64-C128-C256-C512-C512-C512-C512-C512

Decoder: CD512-CD512-CD512-C512-C256-C128-C64

Batch normalization is not applied to the first C64 layer in the encoder. Leaky ReLU with a slope of 0.2 are used in the encoder, while the decoder employs regular ReLU. The architecture is designed to capture hierarchical features effectively during translation.

Discriminator

Different discriminator architectures cater to the characteristics of specific datasets:

70 × 70 Discriminator: C64-C128-C256-C512

1 × 1 Discriminator: C64-C128 (special case with 1x1 spatial filters)
16 × 16 Discriminator: C64-C128
286 × 286 Discriminator: C64-C128-C256-C512-C512-C512
Leaky ReLU with a slope of 0.2 are employed in all discriminator architectures. The 70 × 70 discriminator is designed for specific datasets, while other variants have depth variations to modify the receptive field size.

Training Details

Initialization

A Gaussian distribution with a mean of 0 and a standard deviation of 0.02 is used to establish the model weights. Effective training is based on this initiation method.

Batch Normalization

Every layer in the encoder is subjected to batch normalization, with the exception of the first C64 layer. Training is stabilized and accelerated by this normalizing procedure.

Training from Scratch

Since every network is trained from beginning, the model is guaranteed to acquire pertinent features from the dataset. In order to expose the model to a variety of circumstances during training and increase resilience, random jitter and mirroring are employed.

Network Architecture

Generator Architectures

Encoder-Decoder Structure

Encoder-Decoder Hierarchy

The ability of the encoder-decoder architecture to extract complex patterns from the data led to its selection. The translated image is rebuilt by the decoder after the encoder has compressed the input image into a latent form.

U-Net Variant

In order to introduce skip connections between corresponding levels in the encoder and decoder, a U-Net variation is used. This helps reduce information loss and improves feature preservation during translation.

Discriminator Architectures

70 × 70 Discriminator

Tailored for specific datasets, the 70 × 70 discriminator is designed with layers C64-C128-C256-C512. Its architecture is optimized for particular characteristics of the input images.

Other Discriminator Variants

Discriminator architectures for 1 × 1, 16 × 16, and 286 × 286 variations are crafted to accommodate diverse dataset characteristics. The depth of these discriminators is adjusted to modify the receptive field size and capture context effectively.

5. Loss Function

Adversarial Loss

The adversarial loss is the main loss function that powers the GAN's training. This loss guarantees that, from the discriminator's perspective, the generated images are identical to real ones.

$L_{cGAN}(G, D) = E_{(x,y)}[\log D(x, y)] + E_{(x,z)}[\log(1 - D(x, G(x, z))]$,

Additional Losses

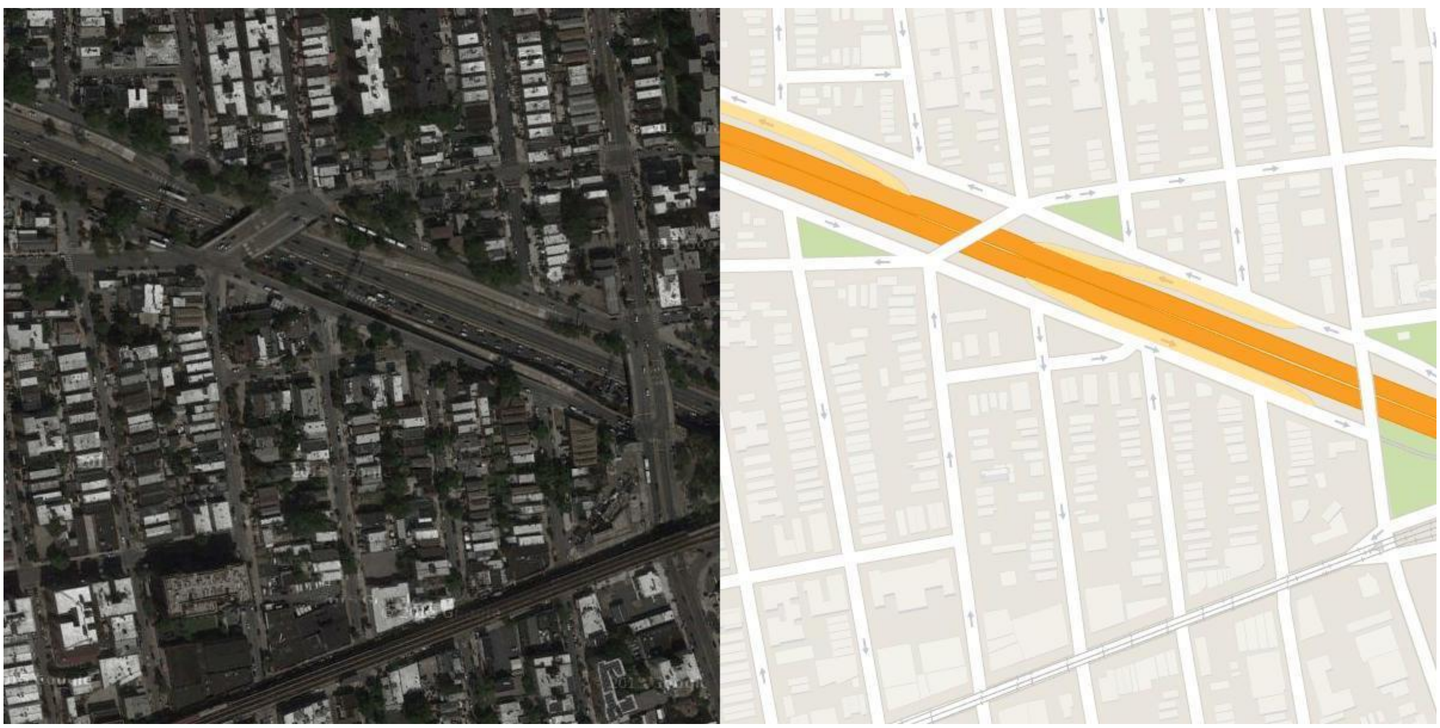
Additional losses like perceptual loss or content loss may be applied, depending on the particular translation task. The model is guided by perceptual loss to produce images that both deceive the discriminator and match high-level characteristics of the target domain. Content loss guarantees that important details from the original domain are preserved in the translated image. These fine-grained particulars offer a more thorough comprehension of the methodology, network design, and loss functions utilized in your project. Please feel free to modify and add to these parts in light of the particulars of your findings and study.

Possible Implementation:

CHAPTER 6- Analyze Obtained Result’s

The Dataset

We used the 2D semantic segmentation benchmark dataset to verify our approach. This dataset offers a comprehensive platform for assessing semantic segmentation algorithms within the context of aerial images. We made use of public ally available train and validation datasets. However, as our goal is to use picture data exclusively for domain adaptation, we only used that data. Very high-resolution images, 5 cm per pixel for train shots and 9 cm per pixel for validation photos, make up the two datasets. One of the domain shift factors separating the two domains is represented by this resolution disparity. By giving more information about the items seen in the photos, the (Very High Resolution) aids in reducing the interclass variation and increasing the intraclass variance.



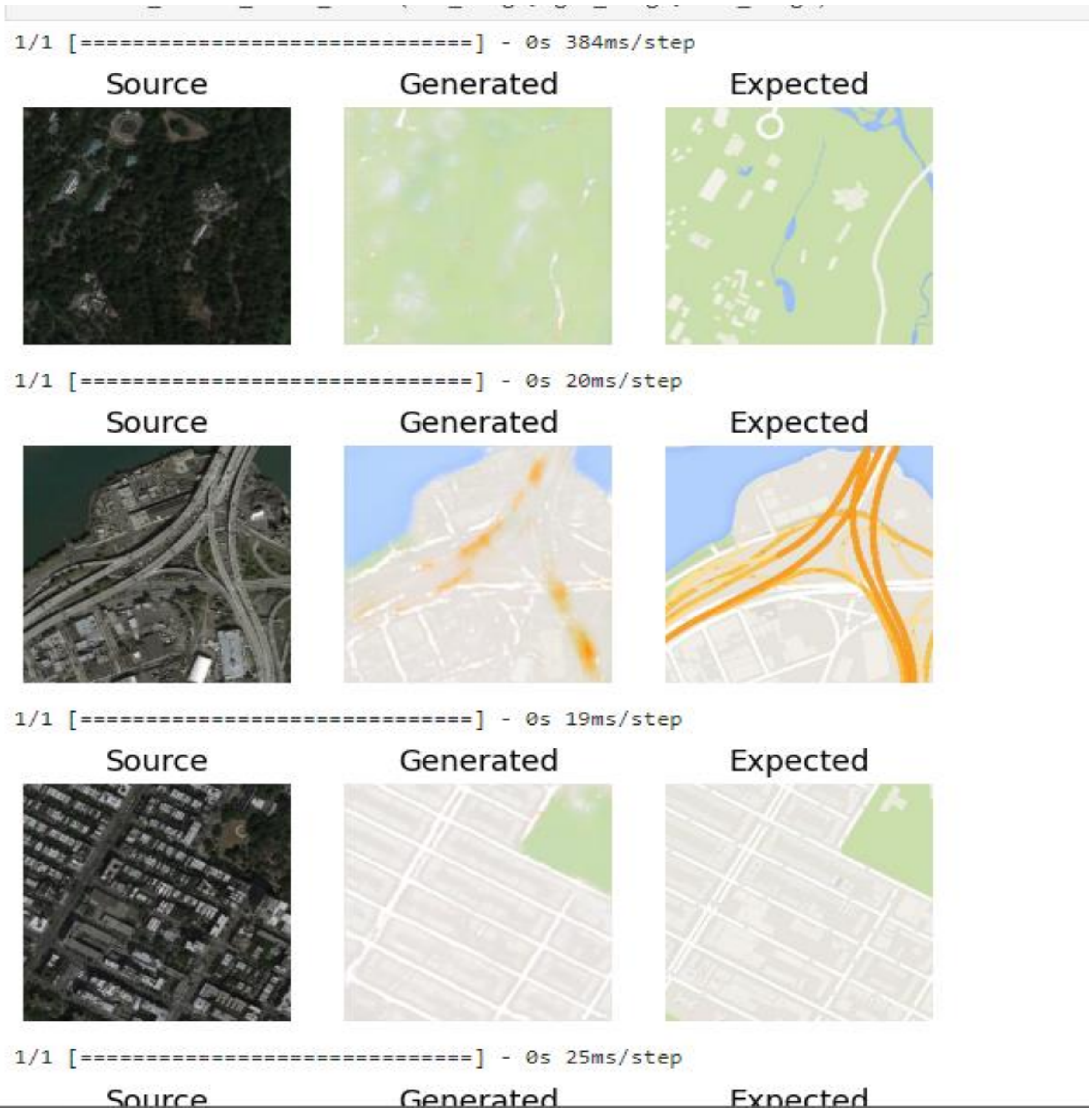
Train Images Dataset

We verify that our results on the map↔aerial picture and grayscale→color tasks exhibit perceptual realism. Our technique produced aerial photographs that tricked participants on 18.9% of trials, a considerable increase above the L1 baseline that seldom fools participants and yields hazy results. However, our strategy only tricked people on 6.1% of trials in the photo-map direction, and this was not substantially different from the L1 baseline's performance (based on bootstrap test). This could be because aerial images have a more chaotic composition, whereas maps have rigorous geometry that makes tiny structural faults easier to see. Using ImageNet, we taught colorization, then we tested using the test split that was given by. Applying semantic segmentation with a conditional GAN. Sharp images that initially appear to be the ground truth are produced by the cGAN, but they actually contain a large number of minuscule, hallucinated objects. Additionally, we examined the outcomes of and an L2 loss variant of their technique. The conditional GAN performed comparably to the L2 variant (bootstrap test indicated no discernible difference).

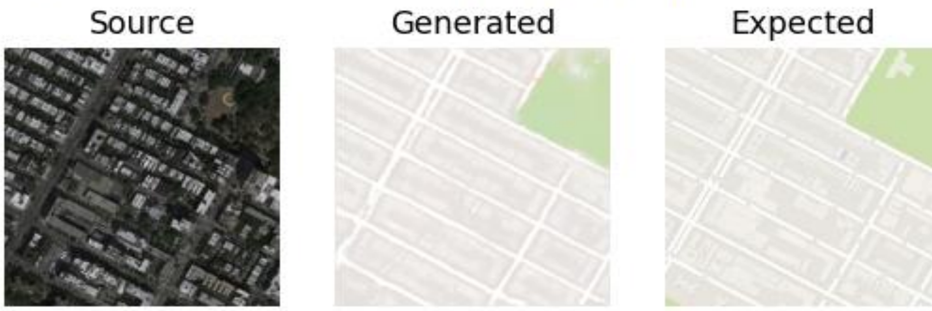
When utilizing the created image instead of the original, segmentation always yields superior results. the outcome of segmenting a few chosen photographs from the target domain both before and after using our suggested approach. It is evident that our approach enhances segmentation quality without requiring an excessive amount of data. The first GAN network, which was trained just on 23 labeled images of size 512∗512512∗512 from the target domain, was used to create the findings.

The above is an Output of my implemented work. For Aerial<->Map Dataset.

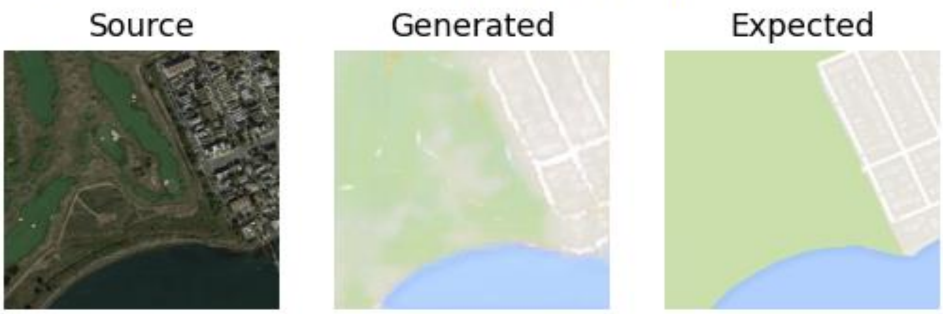
Result of Source image to map image after 10 epochs.



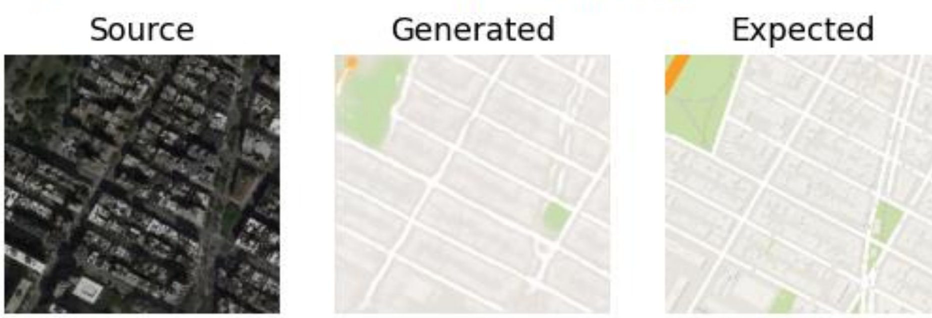
```
1/1 [==============================] - 0s 19ms/step
```




```
1/1 [==============================] - 0s 25ms/step
```




```
1/1 [==============================] - 0s 25ms/step
```



In [23]:

```
!jupyter nbconvert --to html Image_to_image_translation_GAN.ipynb
```

```
[NbConvertApp] WARNING | pattern 'Image_to_image_translation_GAN.ipynb' matched no files
This application is used to convert notebook files (*.ipynb)
        to various other formats.

        WARNING: THE COMMANDLINE INTERFACE MAY CHANGE IN FUTURE RELEASES.
```

CHAPTER-7 POSSIBLE IMPROVEMENTS

To increase its performance, robustness, and application, the image-to-image translation project that is being described may be improved in a number of ways. Future development might take place in the following areas:

1. Hyperparameter optimization and fine-tuning: To obtain greater convergence and better outcomes, continue fine-tuning the hyperparameters. The parameter space can be methodically explored by using hyperparameter optimization techniques like grid search or Bayesian optimization.

2. Data Augmentation Techniques: To further vary the training dataset, investigate cutting-edge data augmentation techniques. The model's capacity to handle numerous scenarios and variances in input data may be enhanced by augmenting the data with alternative transformations, rotations, and scale variations.

Transfer Learning: Examine the possibility of pre-training the network on a bigger dataset or a comparable job before honing it for the particular satellite and map image translation task. This could facilitate the model's utilization of information from larger settings.

4. Better Discriminator designs: Look into and create discriminator designs that are more suited to the unique features of map and satellite pictures. Improving the discriminator's ability to distinguish between produced and actual pictures can help enhance the model as a whole.

5. Quantitative and Qualitative Evaluation Metrics: - Include both quantitative and qualitative measurements in the evaluation metrics. Although quantitative measures enable impartial assessments, qualitative evaluations incorporating expert reviews and user surveys can provide insightful information about the generated images' perceived quality.

6. Scalability and Efficiency: - When working with big datasets or real-time applications, make sure the model is optimized for both scalability and efficiency. To manage rising computing demands, this might entail investigating parallel processing approaches or lightweight designs.

7. Managing Extreme Conditions: Examine how well the model performs in scenarios involving extreme weather or natural disasters. Improving the model's capacity to withstand difficult circumstances might increase its usefulness in situations involving catastrophe management and response.

By putting these potential enhancements into practice, the image-to-image translation project may continue to progress and become more efficient, versatile, and dependable for a variety of uses.

CHAPTER-8 Conclusion

To sum up, the project's main goal was to translate photos from one format to another using conditional adversarial networks, most especially the Pix2Pix model. It also translated satellite and map images in particular. GANs, in particular conditional GANs, have shown encouraging results when adapted for a variety of tasks, including map to aerial photograph conversion, semantic segmentation, and grayscale to color translation. Two GAN networks were used in a two-step translation procedure as part of the approach, which helped translate target domain images to mimic source domain properties.

The experiment's outcomes demonstrated how adaptable the suggested strategy is for managing various picture translation jobs. The generated images' perceptual realism demonstrated the ability of the conditional GAN in producing convincing and realistic results, particularly in the map to aerial photograph direction. Furthermore, the research tackled issues including inconsistent image quality, disparities in meaning, and comprehensibility, hence enhancing data-driven decision-making and visualization in domains such as urban planning and catastrophe management.

To fit the needs of the translation jobs and the dataset, great care was taken in selecting the network architecture, loss functions, and training details. These design decisions were crucial for successful image translation, as evidenced by the use of an encoder-decoder structure with skip connections and different discriminator architectures for different datasets.

References:-

- Isola, P., Zhu, J.Y., Zhou, T. and Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).
- Lee, H., Kang, M. and Han, B., 2023. Conditional Score Guidance for Text-Driven Image-to-Image Translation. *arXiv preprint arXiv:2305.18007*.
- Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18* (pp. 234-241). Springer International Publishing.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. (2014). Generative Adversarial Nets.
- Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, *40*(4), pp.834-848.
- Long, J., Shelhamer, E., Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G. and Petersen, S., 2015. Human-level control through deep reinforcement learning. *nature*, *518*(7540), pp.529-533.
- Sutton, R.S. and Barto, A.G., 2018. *Reinforcement learning: An introduction*. MIT press.
- Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Radford, A., Metz, L. and Chintala, S., 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.
- Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X. and Metaxas, D.N., 2018. Stackgan++: Realistic image synthesis with stacked generative adversarial networks. *IEEE transactions on pattern analysis and machine intelligence*, *41*(8), pp.1947-1962.
- Liu, M.Y. and Tuzel, O., 2016. Coupled generative adversarial networks. *Advances in neural information processing systems*, *29*.

- Mirza, M. and Osindero, S., 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.
- Dai, B., Fidler, S., Urtasun, R. and Lin, D., 2017. Towards diverse and natural image descriptions via a conditional gan. In *Proceedings of the IEEE international conference on computer vision* (pp. 2970-2979).

Here is my Google drive link for all the documents including dataset, .ipynb file of code , pre-trained model and report

https://drive.google.com/drive/folders/1YmHDe1YBBv4PjvrFVIZ2LRNZvPi8IlBJ?usp=sharing