AI AND ML MINI PROJECT REPORT

HOUSE FLOOR PLAN GENERATION

DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY

COLLEGE OF ENGINEERING, GUINDY

ANNA UNIVERSITY, CHENNAI

SUBMITTED BY:

A AZARUDEEN(2022179017)

M NANDHINI DEVI(2022179018)

P SHAKTHI(2022179055)

SUBMITTED TO:

DR.M.DEIVAMANI

INTRODUCTION

The proliferation of Generative Adversarial Networks (GANs) has revolutionized the field of artificial intelligence, particularly in the domain of image synthesis. In this project, we embark on a journey to harness the power of GANs for the generation of house plots. The fusion of cutting-edge deep learning techniques with architectural visualization holds the promise of transforming how we conceptualize and create digital representations of residential spaces.

BACKGROUND

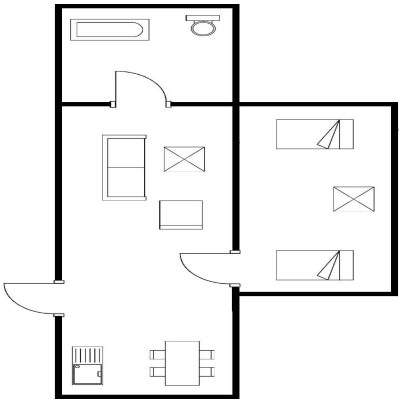
Our project focuses on generating floor plans using a Generative Adversarial Network (GAN) that relies heavily on the quality and coherence of the input data. Data cleaning, as a crucial pre-processing step, plays a pivotal role in ensuring that the training dataset is free from inconsistencies and irrelevant information. This abstraction outlines the key aspects of data cleaning tailored for a GAN-based house plot generation project. The process involves handling missing values, correcting outliers, standardizing formats, and addressing noise in the dataset. By implementing effective data cleaning techniques, the GAN can learn more effectively, leading to the generation of coherent and visually appealing houseplots. Specifically implementing an autoencoder for denoising floor plans. The autoencoder aims to remove noise from the input data, enhancing the quality of the generated floor plans.

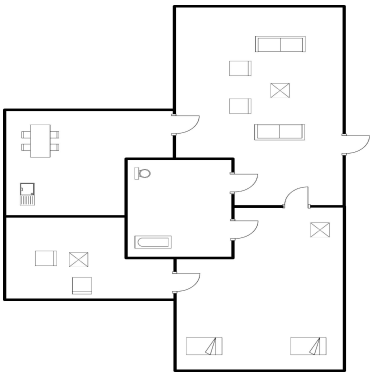
DATA

DATA COLLECTION

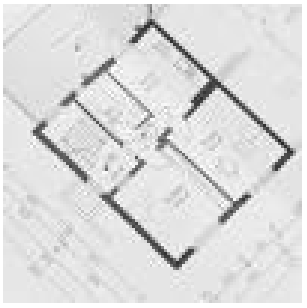
In the initial stage we were given a data consisting images around 6000 which were download in bulk using webscraping method. Those 6000 images are splited among the 6 teams for our team we got around 1000 images and in those images many images were irrelavent and most of images were so blury and uneditable. The number of images which were good to edit (or) clean were around 400. Other than that we got a synthatic dataset of images about 70. The synthatic images were called Robin dataset.

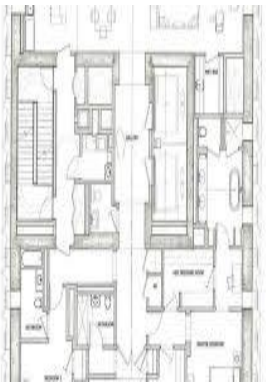
The images from Robin dataset



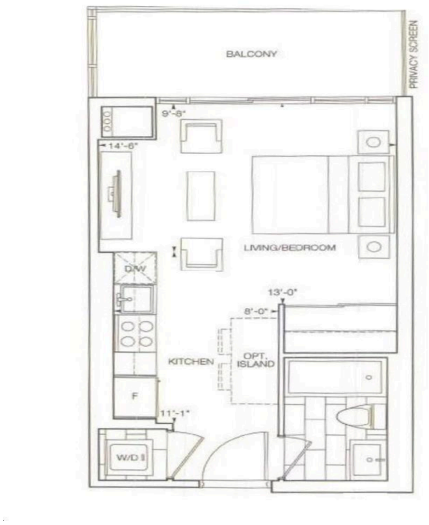


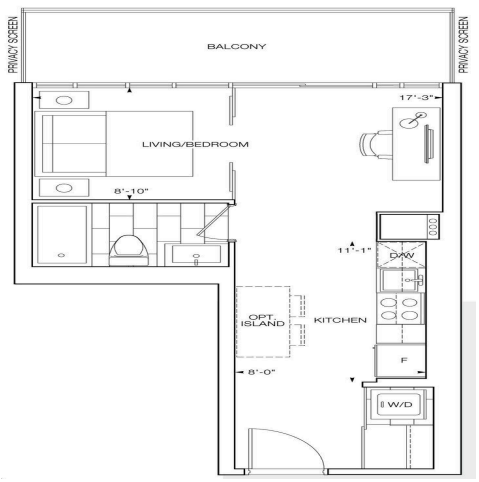
The images which are not able to used





The images which are able to use





DATA CLEANING

From the collected 6000 images 1000 images where splited with in the teams. For our team we got 1000 images and from the 1000 images some images were courpted and some images are irrelavent to the domin since it was a bulk download and webscraping. And there were many duplication within our Team.

We user a open source software called AntiDupl. This problem was with all the teams. So make the proess quick a python script was written to remove the duplicates.

APPROACH USED

Tools used

1. Photopea
2. Python(For removing duplicates)
3. AntiDupl (For removing duplicates)

Removing duplicates

First we used a python script to remove the images which are duplicates from all the groups. we manually deleted the imaegs are courpted and which were irrelavent.

The irrelavent images :

Cleaning the raw images

We used a photo edting tool named Photopea which is a free alternative for photoshop. instead earseing the unwanted text from the images we added lines to over the images layer by layer. From one raw image we need to output one is Layout image and another one is outline image.

Rules to be followed while cleaning a image

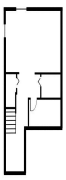
- Removing text from the image
- Making the background as white
- All walls should be black
- Outer walls of the buliding should be 8px rectangle with fill
- Inner walls of the buliding should be 4px rectangle with fill

- Doors and windows of the buliding should be 1px rectangle with fill
- Balcony of the buliding should be 8px rectangle without fill
- Sliding door of the buliding should be 4px rectangle without fill 1px line in the middle

RESULTS

Raw image

Layout image



Outline image

DISCUSSION AND CHALLENGES

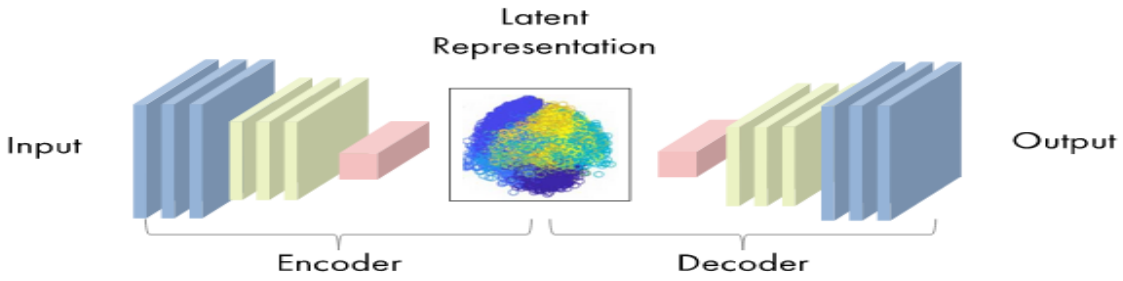
The main discussion in the project was all about the data cleaning and auto encoder which will be useful in creating GANs.

GENERATIVE ADVERSARIAL NETWORK

GAN stands for Generative Adversarial Network. It is a type of neural network architecture that consists of two main components a **generator** and a **discriminator**. The generator creates synthetic data, and the discriminator evaluates the generated data against real data. Both components are trained simultaneously, leading to the improvement of the generator's ability to create realistic data.

AUTO ENCODER

An autoencoder is a type of neural network designed for **unsupervised learning**. It consists of an **encoder** and a **decoder**. The encoder compresses the input data into a latent representation, and the decoder reconstructs the original data from this representation. Autoencoders are often used for tasks such as **data denoising**, feature learning, and dimensionality reduction.



During cleaning the image many teams faced issues in multiple ways like using the photopea tool and regularizing the output for input since there was a variability in the data it was hard to clean image in different scenario and another problem in this project was debuging the Auto encoder code for denoising the image was one of the biggest challenges in the project since theoretically, auto encoders are very new to our team It was hard to implement it.

CONCLUSION

As a part of the data cleaning process our team was able to clean up to 180 images including the Robin dataset. Our team was able to learn what are the difficulties in cleaning a dataset for training a GAN model. Came to know about what are autoencoders and how they are used in the development of GAN.

REFERENCES:

- aakgna/Estate_plot_groups (github.com)
- Photopea tutorial - YouTube
- Deep CNN Autoencoder - Denoising Image | Deep Learning | Python
- Autoencoder Explained - Deep Neural Networks