

## Chapter-1

### **Problem Statement**

Pictorial representation conveys more information than mere words. The human brain can process images up to 60,000 times faster than words. Materialistic realization of our imagination is quite difficult. This is because the media often used for this purpose is speech, which can be interpreted in different ways by different people. This always leads to some ambiguity while expressing our ideas.

This problem can be solved by having a way to convey our ideas through a pictorial medium. For this purpose, Generative Adversarial Network (GAN), deep neural network is required.

Despite the many recent successes of supervised neural networks, it remains challenging to scale such models to accommodate an extremely large number of predicted output categories.

Previous methods like pix2pix, directly feed the semantic layout as input to the deep network, which is then processed through stacks of convolution, normalization, and nonlinearity layers. This is suboptimal as the normalization layers tend to “wash away” semantic information.

Therefore, GAN is used in combination with SPADE, so as to overcome the above mentioned problems.

## Chapter-2

# Objectives and Scope

## 2.1 Objectives

- Semantic Image Synthesis based on user needs.
- Accurately modelling images using Segmentation Map.
- Spatially Adaptive Interaction among segments of Segmentation Map.

## 2.2 Scope

- Designers and Artists will find this useful for immediately recreating their ideas.
- Architectural design can easily be conveyed by the clients without ambiguity.
- Virtual worlds can be simulated with less development cost.

## Chapter-3

# Project Description

The task of generating photorealistic images from the Segmentation Map includes a GAN (Generative Adversarial Network). A specific modification of GAN with SPADE (Semantic Image Synthesis with Spatially-Adaptive Normalization) synthesizes visually effective images over suboptimal neural networks such as pix2pixHD, CRN, or SIMS <sup>[1]</sup>.

### 1. Dataset:

To train a GAN we require a large number of images with Segmentation Map. Segmentation Map is a manipulated result obtained by applying the function  $m \in L^{(H \times W)}$  where,  $m$  - Segmentation Mask,  $H$  - Height,  $W$  - Width of the Image,  $L$  - Semantic Labels <sup>[1]</sup>. Data is obtained from the COCO Stuff dataset <sup>[4]</sup>.

### 2. GAN:

Generative Adversarial Network consists of Generator and Discriminator. The goal of the generator is to produce realistic images so that the discriminator cannot tell the synthesized images apart from the real ones <sup>[2]</sup>.

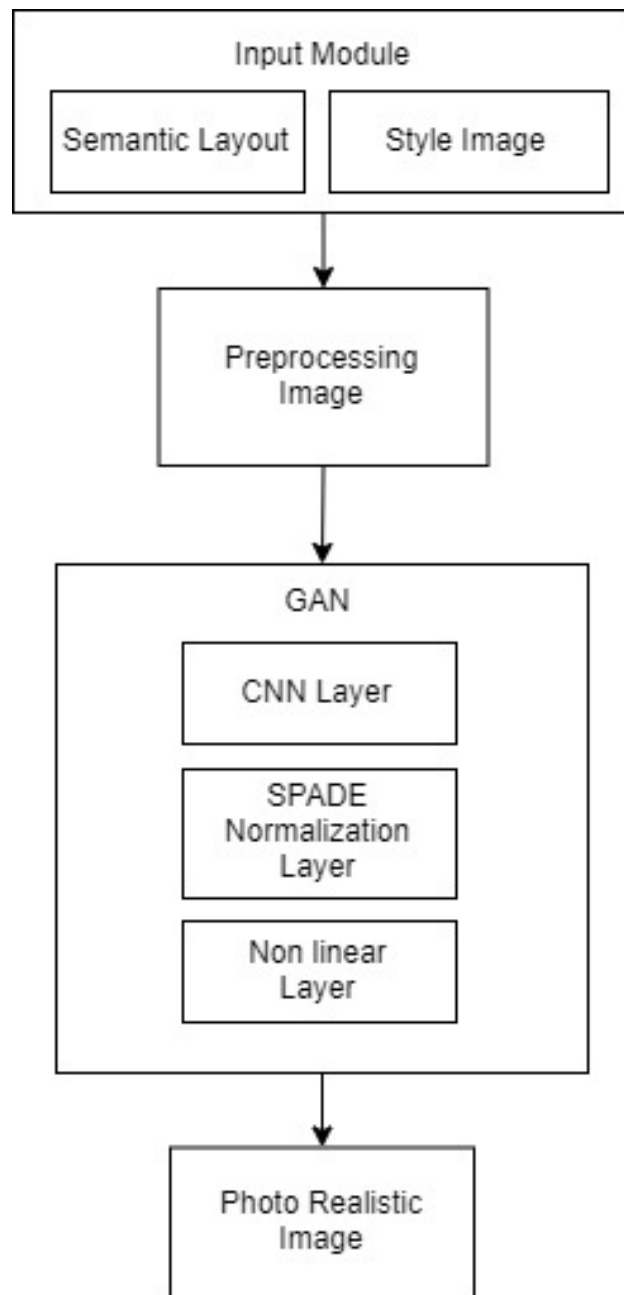
### 3. SPADE:

Spatially Adaptive (De)Normalization is a layer that is used to obtain the normalization of the tensors based on Batch Norm. This provides visual fidelity and alignment with input layouts <sup>[1]</sup>.

### 4. CNN:

Convolutional Neural Network, class of deep neural networks, most commonly applied to analysing visual imagery. Major operations of CNN Convolution, Non-Linearity (ReLU), Pooling or Sub Sampling, Classification <sup>[3]</sup>.

## Chapter-4

**Methodology***Figure 4.1: Flowchart for synthesis of Photo Realistic Image*

## Chapter-5

# Hardware and Software requirements

## 5.1 Hardware Requirements

The minimum requirements are

- Processor with Intel Core i7-7700K (4GHz) or AMD Thread Ripper (32 Core)
- 16 GB RAM
- nVidia Titan X (12 GB) x 2 with SLI Bridge
- 1024 X 720 or higher resolution display with 16bit color space

## 5.2 Software Requirements

The minimum requirements are

- OpenGL ES / WebGL for interfacing
- Tensorflow (Keras) / PyTorch Libraries
- COCO Stuff, Cityscape Dataset

## Chapter-6

### **Possible Outcomes**

- Generation of photorealistic images.
- Improved communication by better representation of ideas.
- User friendly interface to give semantic Input.
- Improved visual fidelity and alignment with input layouts.

## Chapter-7

### References

- [1]. Semantic Image Synthesis with Spatially-Adaptive Normalization - Taesung Park,  
Ming-Yu Liu, Ting-Chun Wang, Jun-Yan Zhu  
<https://arxiv.org/abs/1903.07291>
- [2]. Conditional Generative Adversarial Nets - Mehdi Mirza, Simon Osindero  
<https://arxiv.org/pdf/1411.1784.pdf>
- [3]. Understanding of a Convolutional Neural Network  
<https://ieeexplore.ieee.org/document/8308186>
- [4]. Microsoft COCO: Common Objects in Context  
<https://arxiv.org/pdf/1405.0312.pdf>