HUJBERT PATRIK

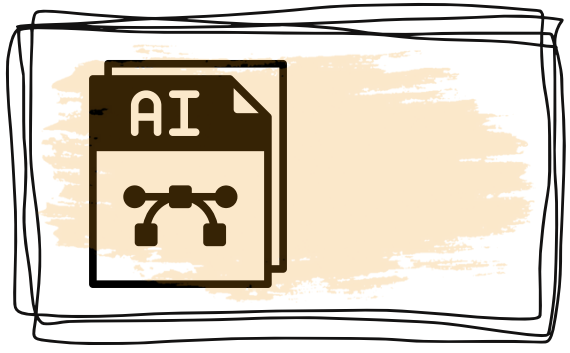
2022

SCENE GEN

AI IMAGE GENERATOR

IMAGE GENERATION

FROM SCENE GRAPSH



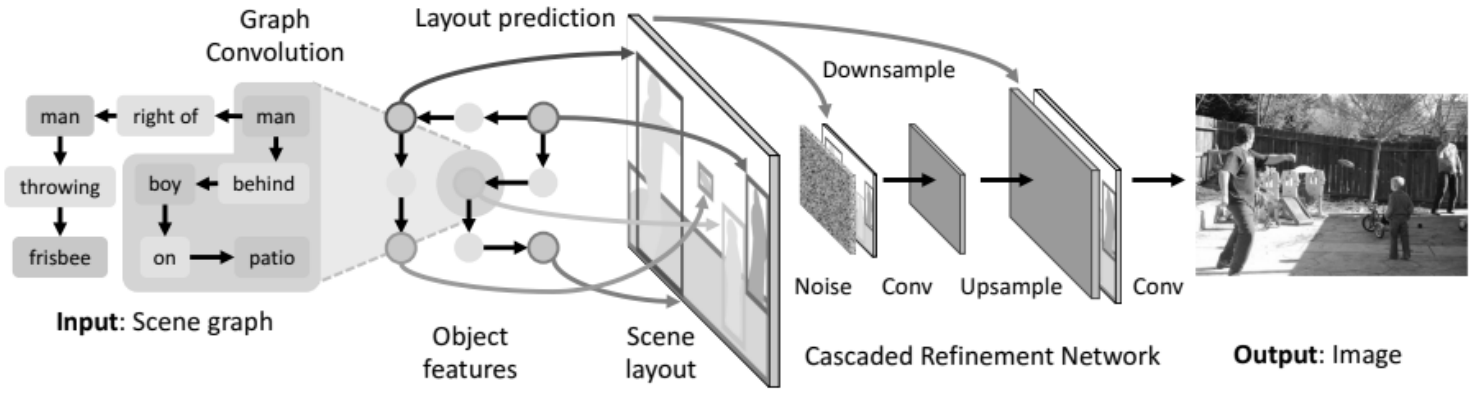
MOTIVATION

- The most popular image generation networks: text2image
- More structured input: scene graph
- Improve upon GAN solution with diffusion

SG2IM MODEL

GAN

- Generative Adversarial Network
- Trained on COCO-Stuff, Visual Genome
- Image, object discriminator
- Generator network



DATASET

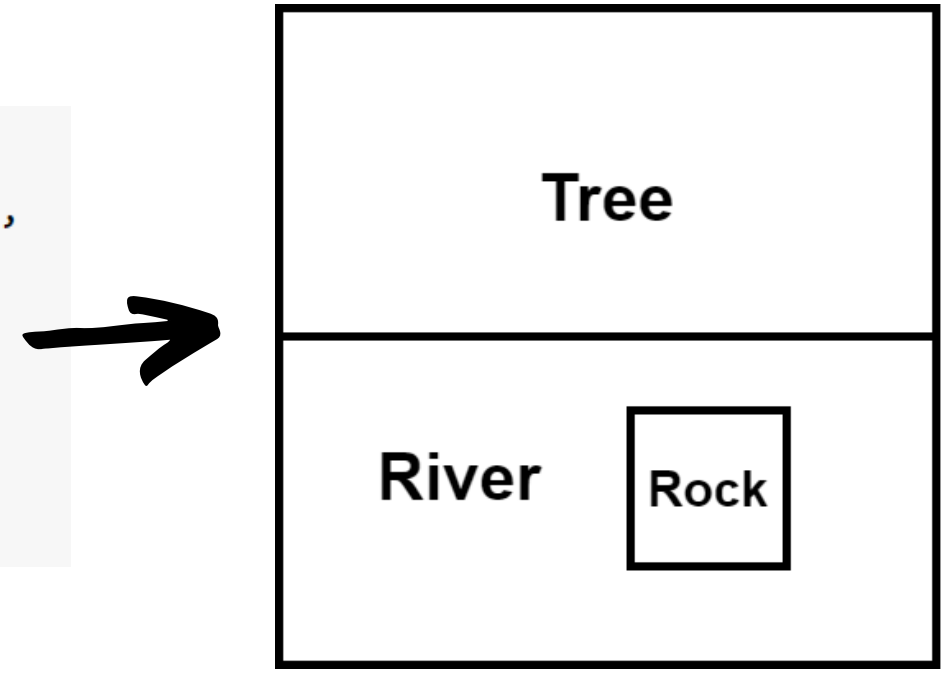
COCO-STUFF

- 164K complex images from COCO
- Dense pixel-level annotations
- 80 thing classes, 91 stuff classes
- Complex spatial context between stuff and things

SCENE GRAPH

Represent scenes as directed graphs, where nodes are objects and edges give relationships between objects

```
{
    "objs": ["river", "rock", "tree"],
    "triples": [
      ["rock","inside","river"],
      ["river","below","tree"],
      ["rock","below","tree"]
    ]
  }
```



LET'S ANALYZE THE ARCHITEC-TURE

QUESTION 01

What type of network to process the input graphs?

QUESTION 02

How to guide the diffusion network to generate images based on the given scene graph?

QUESTION 03

How to connect the two network?

GRAPH CONVO-LUTION

To process the input graphs I used graph convolution

1 PROBLEM

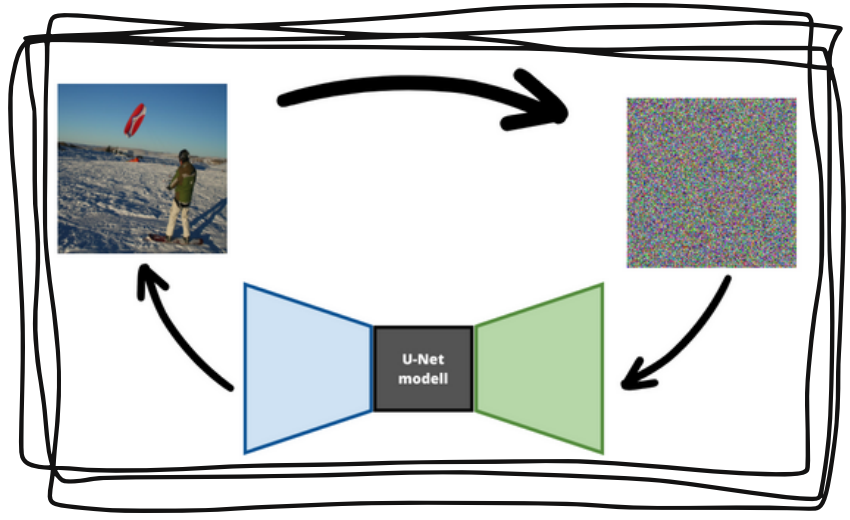
What type of GCN architecture?

How to make the training process easier?

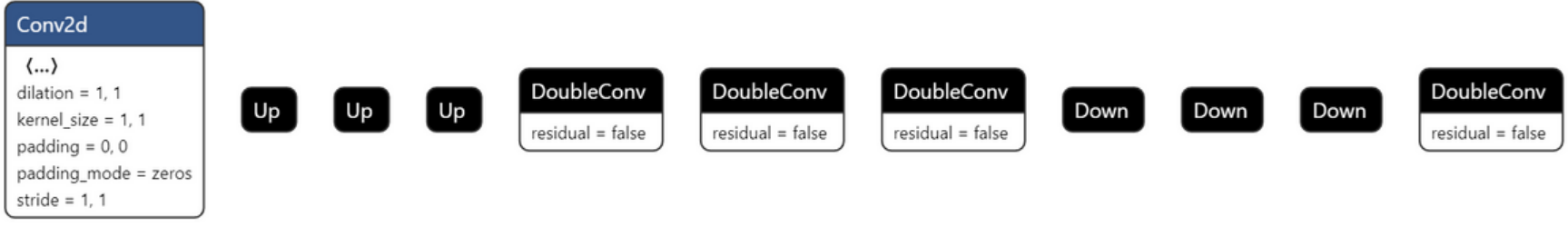
2 SOLUTION

Transfer learning

DIFFUSION MODEL



To generate the images I implemented a Denoising Diffusion Probabilistic Model using Classifier-free Diffusion Guidance



During training the model was sampled several times
To validate it's performance I used the validation dataset

LOOK AT THESE TEST IMAGES

Scene

A river with a rock in it and above trees

Objects: river, rock, tree
Triples:

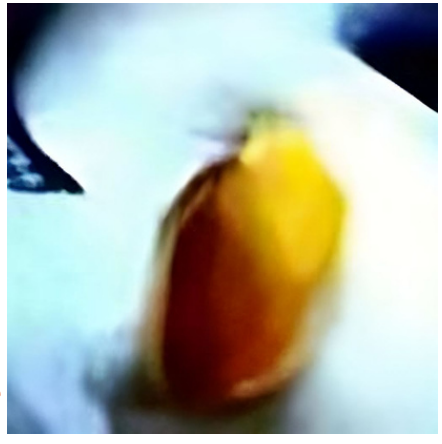
- **rock inside river**
- **river below tree**
- **rock below tree**

Scene

Orange on a table

Objects: orange, table
Triples:

- **table surrounding orange**
- **orange inside table**



Scene

Beach: sea at the top and sand below

Objects: sea, sand
Triples:

- **sand below sea**
- **sea above sand**

PERFORMANCE

Analyzing the model's performance and it's lacking features as well

2-4 OBJECTS

The model was trained on images containing 2-4 objects.
It performes best with two objects in a scene.

IMAGE QUALITY

The main goal of the project wasn't the quality of the images, rather the scene representation ability of the model.

DIVERSITY

It performes poorly with scenarios that aren't likely in the real world

FURTHER IMPROVEMENTS

01 ARCHITECTURAL CHANGES →

02 HIGHER RESOLUTION →

03 FASTER SAMPLING TIME →

04 USER INTERFACE →

SceneGen
Hujbert Patrik

THANK YOU