National College of Ireland

Configuration Manual

MSc Research Project
MSc in Artificial Intelligence

Sameera Sudath Tennakoon Wanninayaka Tennakoon Mudiyanselage
Student ID:23399031

School of Computing
National College of Ireland

Supervisor: Vikas Sahni

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Sameera Sudath Tennakoon Wanninayaka Tennakoon Mudiyanselage
Student ID:	23399031
Programme:	MSc in Artificial Intelligence
Year:	2025
Module:	MSc Research Project
Supervisor:	Vikas Sahni
Submission Due Date:	11/12/2025
Project Title:	Configuration Manual
Word Count:	883
Page Count:	7

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Sudath Tennakoon
Date:	11th December 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Sameera Sudath Tennakoon Wanninayaka Tennakoon Mudiyanselage
23399031

1 Introduction

This configuration manual explains how to configure and how to reproduce the research content and run the Grad-CAM,VQ-GAN and MAE Python codes used in this research without any issue. It documents system requirements, dependency installation, repository layout, configuration variables and notes for reproducing results in a clean environment. The purpose of this manual is to ensure that the entire process from dataset preparation to model training, evaluation, and visualization can be easily reproduced on any compatible system.

2 System requirements

2.1 Hardware

- **GPU**: NVIDIA RTX-4050 6GB with CUDA.
- **CPU**: AMD Ryzen 5 8645HZ
- **Memory**: 16GB
- **Disk space**: Ensure sufficient storage for intermediate and final outputs (This requirements depend on dataset size and number of mask ratios,recommend at least 50GB - 75GB free space).

2.2 Software

All research and development activities were done on a Windows 11 Home system using Python 3.10. All the required python libraries are listed below and also in the root directory you can find it in requirements.txt file too. You can simply install all the dependencies in your virtual environment using this file. Use *pip install -r requirements.txt* command in the terminal after setting the virtual environment. The full project code is available at github(`https://github.com/SudathTennakoon/practicum`). Clone the repository and then donwnload the *mae_visualize_vit_large_ganloss.pth*[1] from official MAE Repository and paste it in the root directory.

- anyio==4.11.0
- argon2-cffi==25.1.0

[1] `https://dl.fbaipublicfiles.com/mae/visualize/mae_visualize_vit_large_ganloss.pth`

- argon2-cffi-bindings==25.1.0
- arrow==1.4.0
- asttokens==3.0.0
- async-lru==2.0.5
- attrs==25.4.0
- babel==2.17.0
- beautifulsoup4==4.14.2
- bleach==6.3.0
- blobfile==3.1.0
- certifi==2025.10.5
- cffi==2.0.0
- charset-normalizer==3.4.4
- colorama==0.4.6
- comm==0.2.3
- contourpy==1.3.2
- cycler==0.12.1
- debugpy==1.8.17
- decorator==5.2.1
- defusedxml==0.7.1
- einops==0.8.1
- einx==0.3.0
- exceptiongroup==1.3.0
- executing==2.2.1
- fastjsonschema==2.21.2
- filelock==3.20.0
- fonttools==4.60.1
- fqdn==1.5.1
- frozendict==2.4.7
- fsspec==2025.9.0

- h11==0.16.0
- httpcore==1.0.9
- httpx==0.28.1
- huggingface-hub==0.35.3
- idna==3.11
- ipykernel==7.1.0
- ipython==8.37.0
- isoduration==20.11.0
- jedi==0.19.2
- Jinja2==3.1.6
- json5==0.12.1
- jsonpointer==3.0.0
- jsonschema==4.25.1
- jsonschema-specifications==2025.9.1
- jupyter-events==0.12.0
- jupyter-lsp==2.3.0
- jupyter_client==8.6.3
- jupyter_core==5.9.1
- jupyter_server==2.17.0
- jupyter_server_terminals==0.5.3
- jupyterlab==4.4.10
- jupyterlab_pygments==0.3.0
- jupyterlab_server==2.28.0
- kiwisolver==1.4.9
- lark==1.3.1
- lxml==6.0.2
- MarkupSafe==3.0.3
- matplotlib==3.10.7
- matplotlib-inline==0.2.1

- mistune==3.1.4
- mpmath==1.3.0
- nbclient==0.10.2
- nbconvert==7.16.6
- nbformat==5.10.4
- nest-asyncio==1.6.0
- networkx==3.4.2
- ninja==1.13.0
- nn==0.1.1
- notebook==7.4.7
- notebook_shim==0.2.4
- numpy==1.23.5
- opencv-python==4.12.0.88
- overrides==7.7.0
- packaging==25.0
- pandas==2.3.3
- pandocfilters==1.5.1
- parso==0.8.5
- pillow==12.0.0
- platformdirs==4.5.0
- prometheus_client==0.23.1
- prompt_toolkit==3.0.52
- protobuf==6.33.1
- psutil==7.1.3
- pure_eval==0.2.3
- pycparser==2.23
- pycryptodomex==3.23.0
- Pygments==2.19.2
- pyparsing==3.2.5

- python-dateutil==2.9.0.post0
- python-json-logger==4.0.0
- pytz==2025.2
- pywinpty==3.0.2
- PyYAML==6.0.3
- pyzmq==27.1.0
- referencing==0.37.0
- requests==2.32.5
- rfc3339-validator==0.1.4
- rfc3986-validator==0.1.1
- rfc3987-syntax==1.1.0
- rpds-py==0.28.0
- safetensors==0.6.2
- scipy==1.15.3
- Send2Trash==1.8.3
- six==1.17.0
- sniffio==1.3.1
- soupsieve==2.8
- stack-data==0.6.3
- sympy==1.14.0
- tensorboardX==2.6.4
- terminado==0.18.1
- timm==1.0.20
- tinycss2==1.4.0
- tomli==2.3.0
- torch==1.13.0
- torchtools: git+https://github.com/pabloppp/pytorch-tools@6472bd5e223142bfbe105e2006432252f1ce3709
- torchvision==0.14.0
- tornado==6.5.2

- tqdm==4.67.1
- traitlets==5.14.3
- typing_extensions==4.15.0
- tzdata==2025.2
- uri-template==1.3.0
- urllib3==2.5.0
- warmup_scheduler==0.3
- wcwidth==0.2.14
- webcolors==25.10.0
- webencodings==0.5.1
- websocket-client==1.9.0

3 Dataset

The dataset used in this study consists of a single semantic class(identifier=n02690373) extracted from ImageNetDeng et al. (2009). The dataset contains approximately 1300 images. To have an uniform resolution though out the study following preprocessing steps were done on each image.

1. All the Images were center-cropped.
2. All the Images were resized to 512×512.
3. Then converted into python tensors and normalized using the standard ImageNet mean and standard deviation values.

These initial preprocessing steps make sure that it works properly with Grad-CAM+ and the same preprocessing also makes sure compatibility with the Vision Transformer-based Masked Autoencoder used later for reconstruction tasks.

4 Repository layout (expected by the script)

The scripts expect below folder layout and files to run those without any issue.

```
./ # working dir where you run the script
 mae_visualize_vit_large_ganloss.pth   # MAE checkpoint
 models/
    vqgan.pt           # models directory
 models_mae.py           # contains MAE model factory functions
 src/
    vqgan.py           # contains `VQModel` class
 utils/                  # utility files
```

```
   lr_decay.py          # File from official MAE Repo
   lr_sched.py          # File from official MAE Repo
   misc.py              # File from official MAE Repo
   pos_embed.py         # File from official MAE Repo
Data/val                 # dataset folder
```

5 Key configuration variables

1. **masking_ratios** - List of mask ratios used (floating fractions, e.g., 0.25 for 25%). Edit to change which mask percentages run.
2. **num_samples** - *num_samples=None* will process the entire dataset, or you can set a value you want to process.
3. **dataset_path** - Set the dataset path using this variable.
4. **chkpt_dir** - Path to MAE checkpoint loaded by `prepare_model()`.
5. **base_dir** = "vqvae results" - Base output directory. This contains `original/`, `gradcam/`, `eval/`, and `recon/` subfolders.
6. **OUTPUT_SIZE** = (512, 512) - Final saved output image size (width, height).
7. **gradcam_fraction** = default(0.5) - Using this variable you can change between random, saliency, and hybrid masking.
   - For random masking, set this to 0.
   - For saliency, set this to 1.
   - Anything between 0 and 1 will be hybrid masking, representing the fraction of patches selected using the saliency map.

References

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database, *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255.