GEN AI - IMAGE RESIZERS

Team

1. College Professor(s):
   Dr. Ranjeet Kumar - ranjeet.k@vit.ac.in
   Dr. Jayanthi R - jayanthi.r@vit.ac.in
1. Students:
   1. Pojesh Kumar R - pojeshkumar.r2022@vitstudent.ac.in
   2. C R Lavanya - lavanya.2022a@vitstudent.ac.in
2. Department:
   SCOPE & SENSE

Date: 29 Aug 2025

GenAI | Image Resizers

SAMSUNG

Problem Statement

Context

GenAI image generation models are good at generating images in trained resolution (For example 1024*1024). However, this limitation is not desirable in real life situation where we need to have images in different resolutions.
Another aspect is generally images generated are in square format which is not suitable for usage on mobile or laptop having rectangle format predominant.

Statement

Resizers and auto-upscalers (2X, 4X) for the images. Change to landscape and portrait.

Worklet Details

6	4	Gokul Nair +91-9496059290 gokul.gn@samsung.com
Duration (Months)	**Members Count**	Mentors

Pre-Requisite

- https://openmodeldb.info/
- https://paperswithcode.com/task/image-super-resolution

Expectations

Undertaken Tasks

- Conduct Literature survey
- Identify the suitable framework
- Build a framework for image resizing and upscaling

KPI

- Web application with simple UI. ComfyUI is preferred.
- It should seamlessly integrate with the backend GenAI models. SDXL etc.
- Latency should be <10 seconds
- Original image contents should remain constant.
- No visible drop in image quality.

Timeline

Kick Off
< 1st Month >

- Problem Briefing
- Check Feasibility
- Technical survey

Milestone 1
< 2nd Month >

- Finalize the model/pipeline
- Architecture diagram

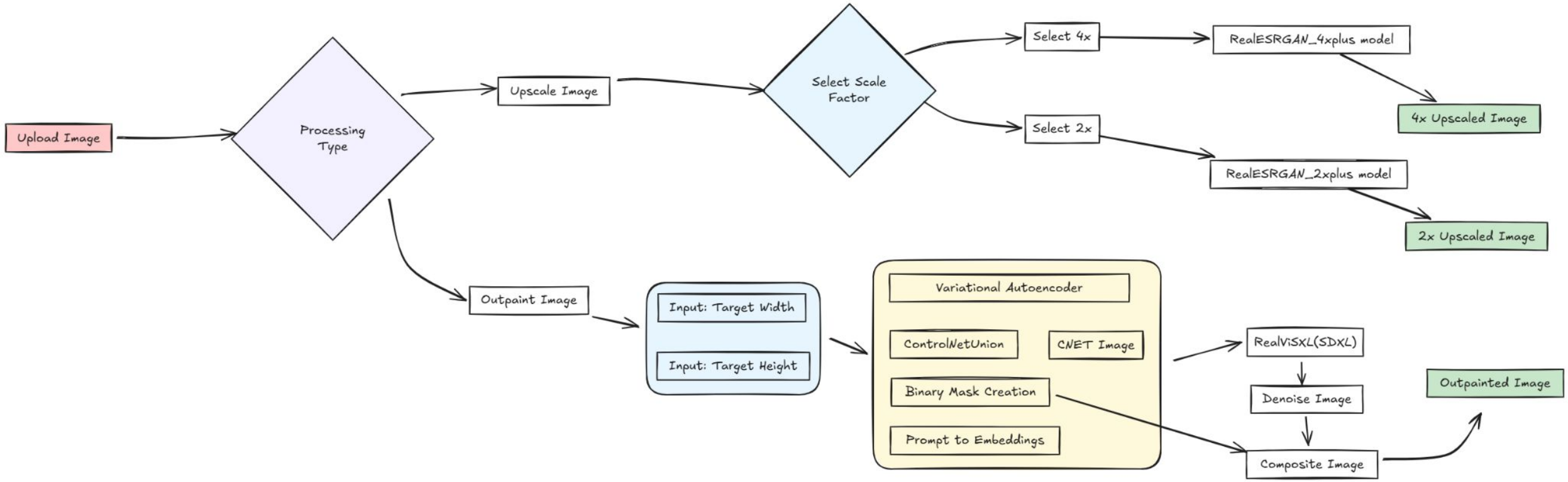
Milestone 2
< 6th Month>

- Final product

Complexity

Approach

- **Architecture Diagram** :



Dataset(s) Analysis / Description

- **Dataset Capture / Preparation / Generation** :

Flux generated images in huggingface spaces - black-forest-labs/FLUX.1-schnell

- Generated synthetic images using **Flux (FLUX.1-schnell)** model on **HuggingFace Spaces**.
- Contains **images of different sizes and resolutions**, ensuring diversity for model training and testing.
- Both **portrait and landscape orientations** are included, allowing experiments in orientation conversion.

- **Dataset Understanding / Analysis** :

Fictional scenario images of various width and height

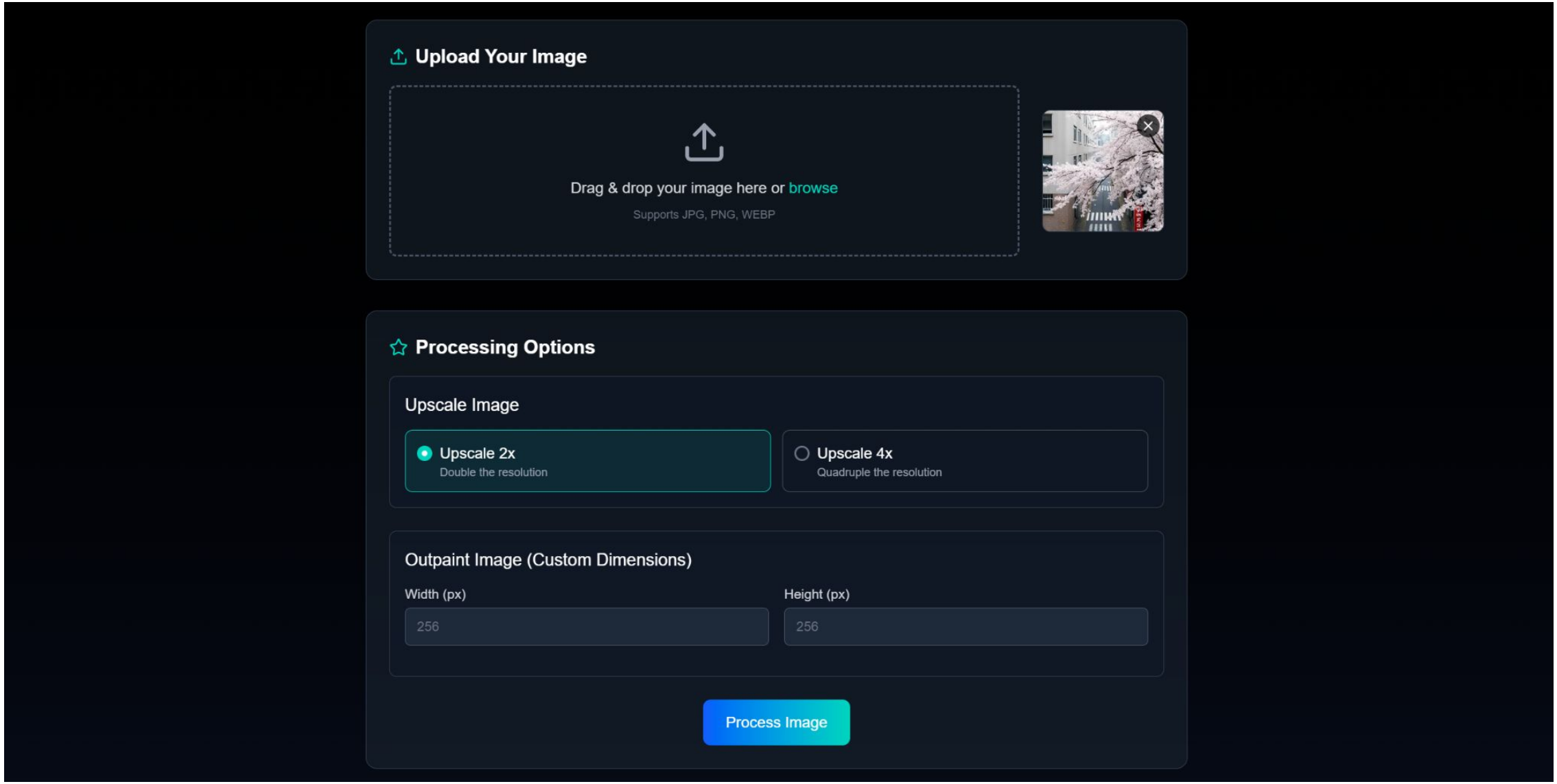
- The dataset consists of **synthetically generated images** with varying resolutions.
- The dataset shows **good diversity in content**, covering different backgrounds, objects, and contexts, which reduces model overfitting.
- Image resolution varies across samples, which makes it suitable for **outpainting and super-resolution experiments**.

- **Data Preprocessing/Related Challenges (if any)** :

No preprocessing or challenges were involved as the dataset images were already perfect and ready to use.

Experimental Results

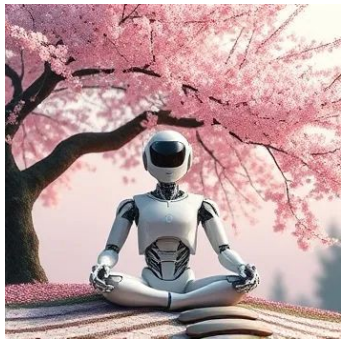
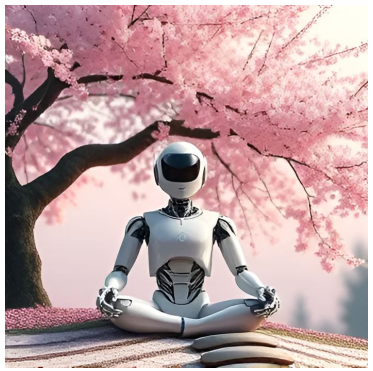
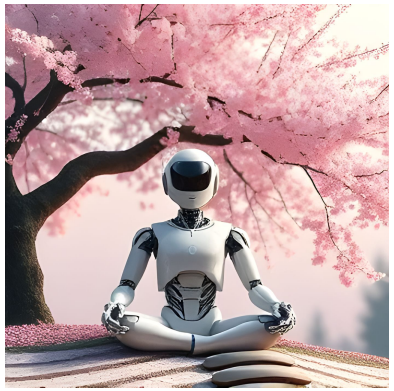
- [Frontend Interface]



Experimental Results

- **Super Resolution Results :**

Super Resolution Results :

Original Image	2X	4X
 352 X 352	 704 X 704	 1408 X 1408
 1024 X 683	 2048 X 1366	 4096 X 2732

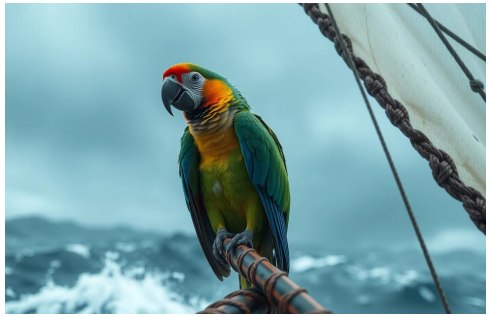
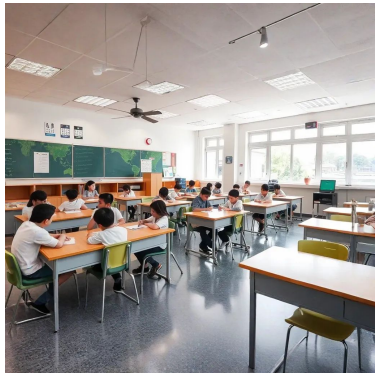
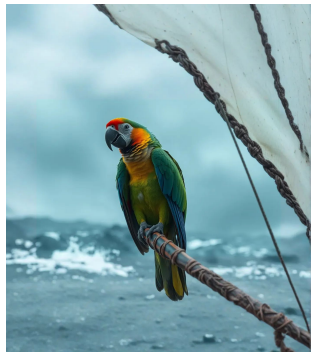
Experimental Results

- Super Resolution Analysis

Original	Original cropped
	
2x Cropped	4x Cropped
	

Experimental Results

- **Outpainting Results :**

Original Image	 992 X 608	 1024 X 1024	 1392 X 896	 992 X 736
Resized Image	 1024 X 1024	 1800 X 1024	 1400 X 1600	 1920 X 1080

Experimental Results / Observations

- **Results** :

Model	Memory Usage(Idle)	Inference Time
Real-ESRGAN	2.3GB	2 - 4 sec
SDXL Pipeline	9.6GB	20 - 50 sec

- **Major Observations / Conclusions & Challenges :**

- Upscaling with **RealESRGAN** successfully upscaled the resolution of the image to 2x and 4x.
- Outpainting with **SDXL** successfully expanded images, allowing **portrait** → **landscape** and **landscape** → **portrait** transformations.
- **Flux-generated images** provided a **clean, high-quality** input dataset.
- Processing of images of resolution more than **2k** of dimension had considerable **increase** in **GPU memory**.

Deliverable

- **Final Deliverables** :
  - Created a **synthetic dataset** using Flux for input images.
  - Implemented **Real-ESRGAN** for image scaling (super-resolution).
  - Applied **SDXL outpainting** for **portrait ↔ landscape conversion**.
  - Explore integration of **real-world data** alongside **synthetic data**.
- **IP / Paper Publication Plan** :
- **KPIs delivered/Expectations Met**:
  - Web application with simple UI built with Next.js.
  - Backend server built with Fast framework.
  - Latency obtained <10 seconds.
  - No visible drop in image quality.

Work-let Closure Details

- **Code Upload details:**

Items	Details
KLOC (Number OF Lines of codes in 000's)	2.3 KLOC
Model and Algorithm details	Models - Real-ESRGAN, SDXL
Is Mid review, end review report uploaded on Git ?	Yes
Link for Git	https://github.ecodesamsung.com/SRIB-PRISM/VITC_24GAI15VITC_GenAI_Image_Resizers.git

- **Data details:**

Items	Data folder 1	Data folder 2	Data Folder 3……..
Name & Type of Data (Audio/Image/Video)	Image		
Number of data points	-		
Source of Data (self collected, Scrapped, available on open source)	HuggingFace space - black-forest-labs/FLUX.1-schnell		
Google drive link/ git link to access data	Worklet repository		

Note: If data uploaded on google drive, access to be shared to prism.srib@gmail.com

Thank you