AI-Powered Food Recipe & Dish Image Generator

Mini Project Report

Prepared by

Tejaswini K

2022510010

B. Tech AI&DS (6/8)

AZ5012 - Applied AI

Introduction:

In recent years, Artificial Intelligence (AI) has made remarkable strides in computer vision, natural language processing, and generative modeling. This project, **AI-Powered Food Recipe & Dish Image Generator**, brings together these capabilities to create an intelligent assistant that can convert a simple food image into a complete culinary experience. With the increasing popularity of food photography and cooking content on social media, there's a growing demand for AI systems that can understand, interpret, and even imagine food in creative ways.

This project is extended to help amateur cooks, inspire professional chefs, assist visually impaired users, or even enhance smart kitchen appliances. By combining object detection (YOLOv3), a transformer-based language model (LaMini-GPT), and two types of generative models (GAN and Stable Diffusion), this project showcases the power of integrating multiple AI domains into a single cohesive application.

Objectives:

The main objective of this project is to build an AI-based system that can generate a complete recipe and visual representation of a dish from a single food image. The specific goals are:

- **Ingredient Detection:** Use object detection (YOLOv3) to accurately identify food items or ingredients present in a given image.
- **Recipe Generation:** Utilize a pre-trained language model (LaMini-GPT) to generate a well-structured and human-readable recipe based on the detected ingredients.
- **Dish Image Generation:** To use a pretrained GAN model trained on the Food-101 dataset to generate a synthetic image of a dish. To use Stable Diffusion to generate a high-quality image based on the recipe title or dish name.
- **User Interface:** Provide an interactive interface using Gradio for users to upload images and view AI-generated outputs.
- **Societal Application:** Support accessibility by assisting visually impaired users in understanding food through descriptions and visuals, and promote food education and creativity in home kitchens.

Technologies used:

This project integrates several AI and software development tools across three major domains: computer vision, natural language processing, and image generation. Below are the key technologies and frameworks used:

Object Detection

- **YOLOv3 (You Only Look Once)**: Used to detect food items from input images.
- **OpenCV**: Utilized for image loading, preprocessing, and interfacing with the YOLO model.

Natural Language Processing

- **Transformers (Hugging Face) - LaMini-GPT-774M**: A causal language model used to generate recipes based on detected ingredients.
- **Tokenizers**: AutoTokenizer from Hugging Face for processing input prompts.

Image Generation

- **Generative Adversarial Network (GAN)**: A custom-trained GAN using, trained on the Food-101 dataset to generate dish images from random noise.
- **Stable Diffusion (CompVis)**: A powerful text-to-image model used to generate realistic dish images based on recipe titles.

Core Frameworks & Libraries

- **PyTorch**: Used for building and training the GAN and for running inference.
- **Torchvision**: For dataset handling and image transformations.
- **Diffusers**: Hugging Face library for deploying and using the Stable Diffusion pipeline.

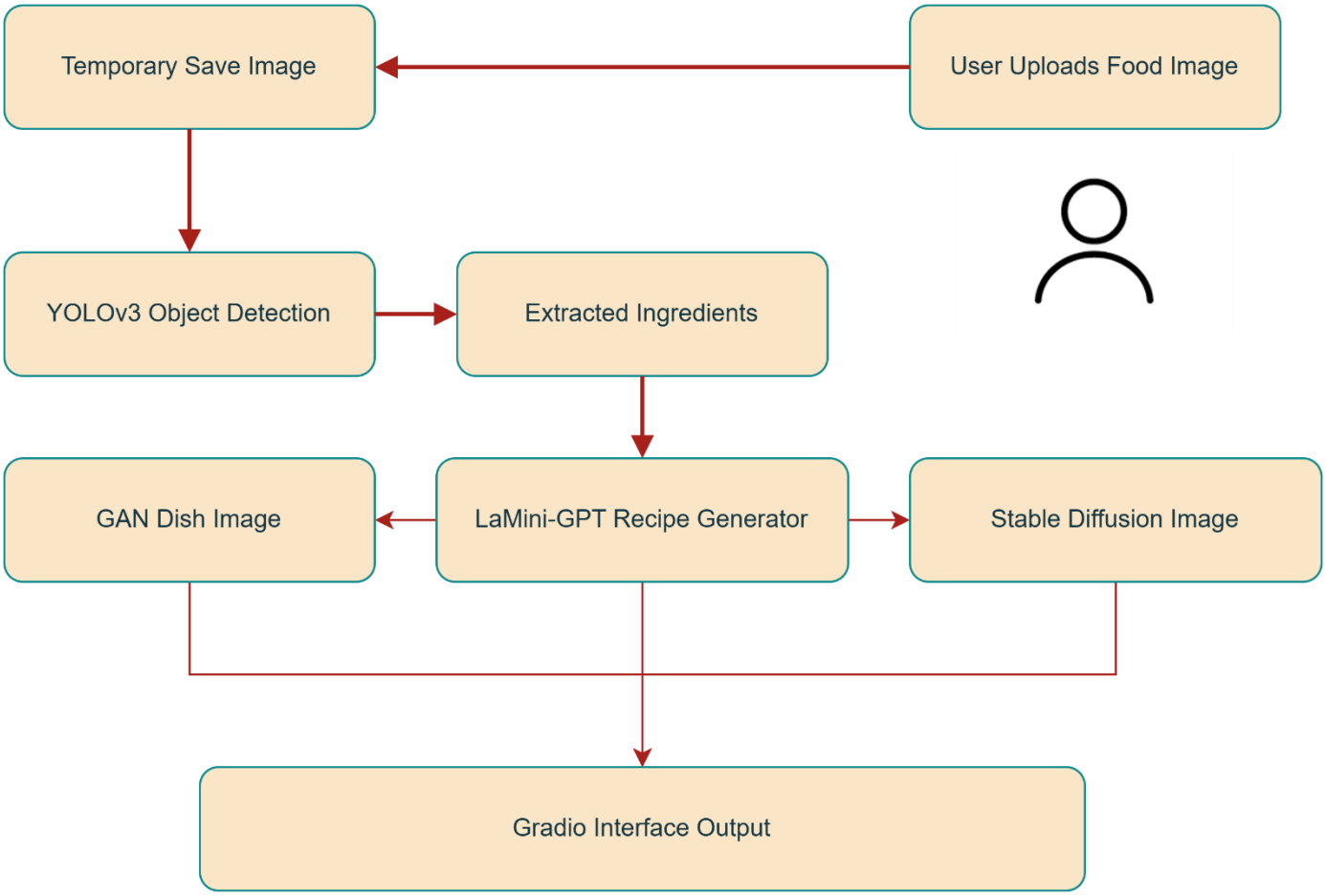
Interface and Visualization

- **Gradio**: A Python library used to build a simple, interactive web-based UI for image upload and displaying outputs.

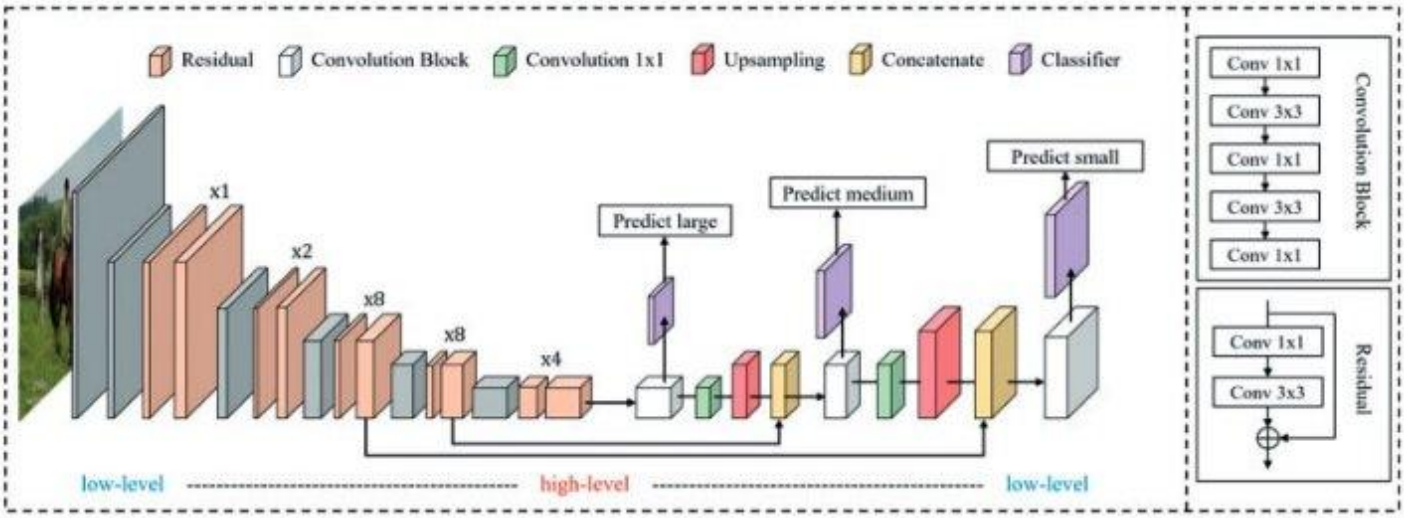
Dataset

- **Food-101**: A large-scale dataset containing images of 101 types of food, used to train the GAN model.

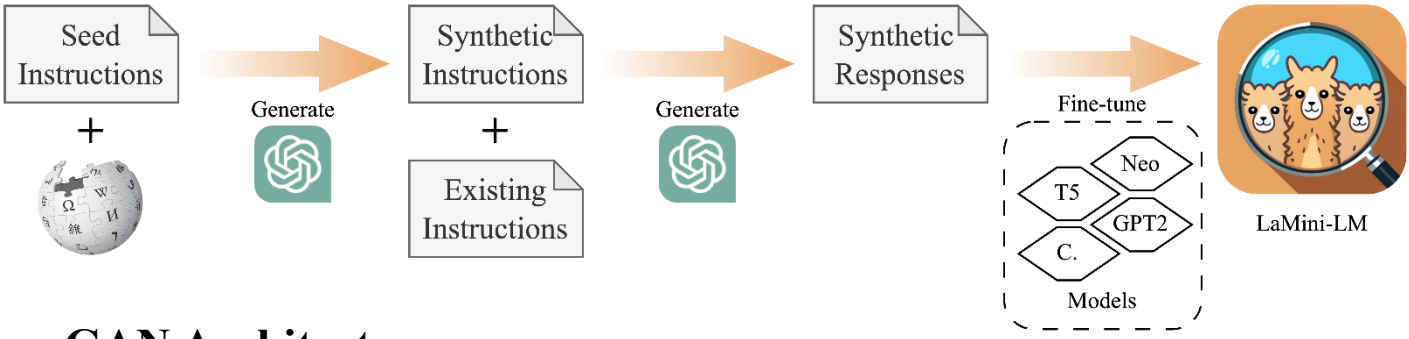
System Architecture:



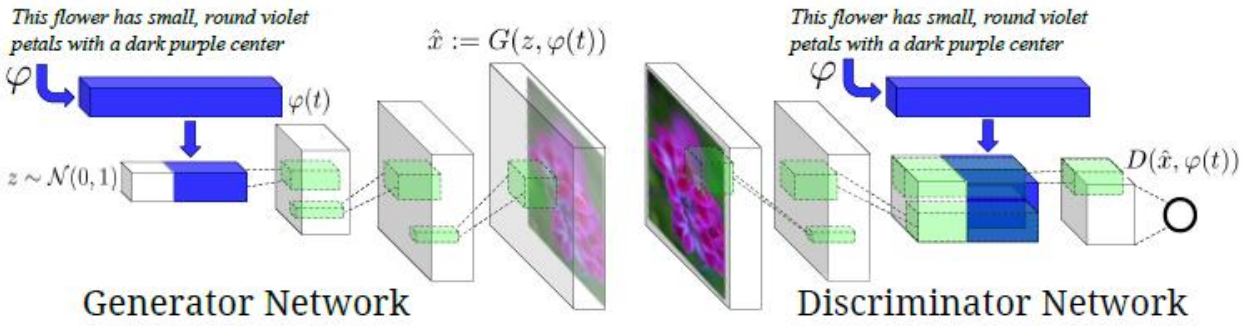
Yolo V3 Architecture:



LaMini-GPT Architecture:



GAN Architecture:



Working:

1. **Image Input & Ingredient Detection (Object Detection)**
   The user uploads a food image via the Gradio interface. The image is processed using the YOLOv3 object detection model. YOLOv3 identifies objects in the image and returns labels like "banana", "carrot", or "pizza" — filtered to include only food-related classes. These are used as the key ingredients for the recipe.
2. **Recipe Generation (NLP)**
   Once ingredients are identified, a prompt is dynamically created and sent to the **LaMini-GPT** language model. This prompt asks the model to generate a full recipe using the detected ingredients. The output includes:
   - Dish title
   - Preparation & cooking time
   - Ingredients with quantities
   - Step-by-step instructions

- Serving suggestions and nutrition tips
  This allows for automatic recipe generation using NLP.

3. **Dish Image Generation (GAN & Diffusion)**
   Two techniques are used for generating visual representations:
   - A **custom GAN (DCGAN)** trained on the Food-101 dataset takes random noise as input and produces a 32×32 dish-like image. It visually captures food-like textures learned from the training data.
   - A **Stable Diffusion model** takes the generated dish title as a text prompt and produces a high-quality, realistic image of the dish. This helps visualize the expected output of the recipe in detail.
4. **Output Display**
   The final output consists of:
   - The full AI-generated recipe (text)
   - A dish image generated by GAN (image learned from dataset patterns)
   - A dish image generated by Stable Diffusion (text-to-image based)
     All of these are shown on a user-friendly Gradio interface.

Results:

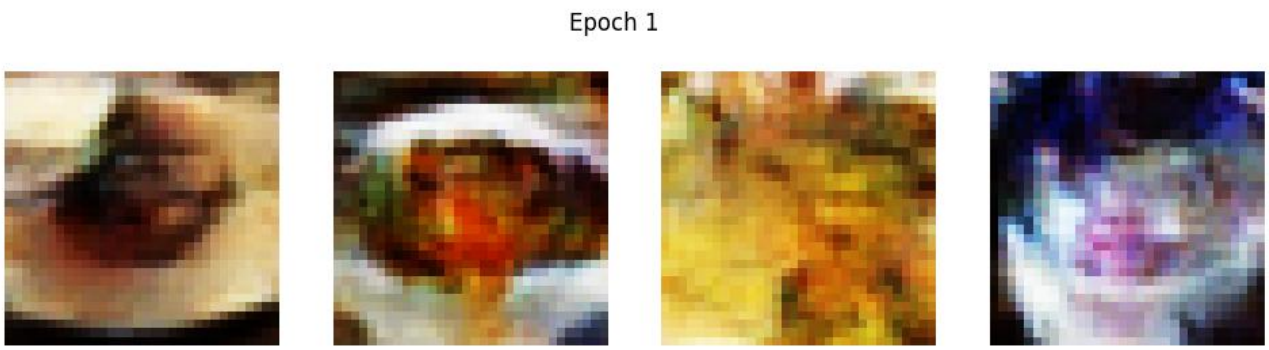
Epoch [1/5] Batch 0/6313 Loss D: 1.3336 G: 1.7430

Epoch [1/5] Batch 100/6313 Loss D: 0.0410 G: 5.7622

...

Epoch [1/5] Batch 6200/6313 Loss D: 1.1141 G: 1.1968

Epoch [1/5] Batch 6300/6313 Loss D: 0.3133 G: 3.2542



Epoch [2/5] Batch 0/6313 Loss D: 0.2762 G: 2.4181

Epoch [2/5] Batch 100/6313 Loss D: 0.5659 G: 3.4793

...

Epoch [2/5] Batch 6200/6313 Loss D: 0.8840 G: 1.7345

Epoch [2/5] Batch 6300/6313 Loss D: 0.3506 G: 4.4166

Epoch 2

Epoch [3/5] Batch 0/6313 Loss D: 0.2886 G: 2.7361

Epoch [3/5] Batch 100/6313 Loss D: 0.4703 G: 2.8258

...

Epoch [3/5] Batch 6200/6313 Loss D: 0.3054 G: 3.5162

Epoch [3/5] Batch 6300/6313 Loss D: 0.3799 G: 2.4282

Epoch 3

Epoch [4/5] Batch 0/6313 Loss D: 0.2782 G: 3.3541

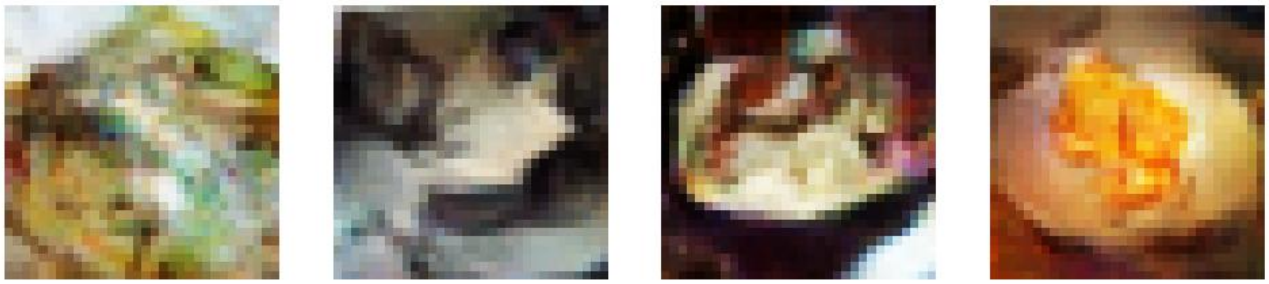
Epoch [4/5] Batch 100/6313 Loss D: 0.2235 G: 3.6304

...

Epoch [4/5] Batch 6200/6313 Loss D: 0.3614 G: 2.2735

Epoch [4/5] Batch 6300/6313 Loss D: 1.0611 G: 1.8947

Epoch 4



Epoch [5/5] Batch 0/6313 Loss D: 0.3318 G: 3.3671

Epoch [5/5] Batch 100/6313 Loss D: 0.1603 G: 4.0658

...

Epoch [5/5] Batch 6200/6313 Loss D: 0.3349 G: 3.7755

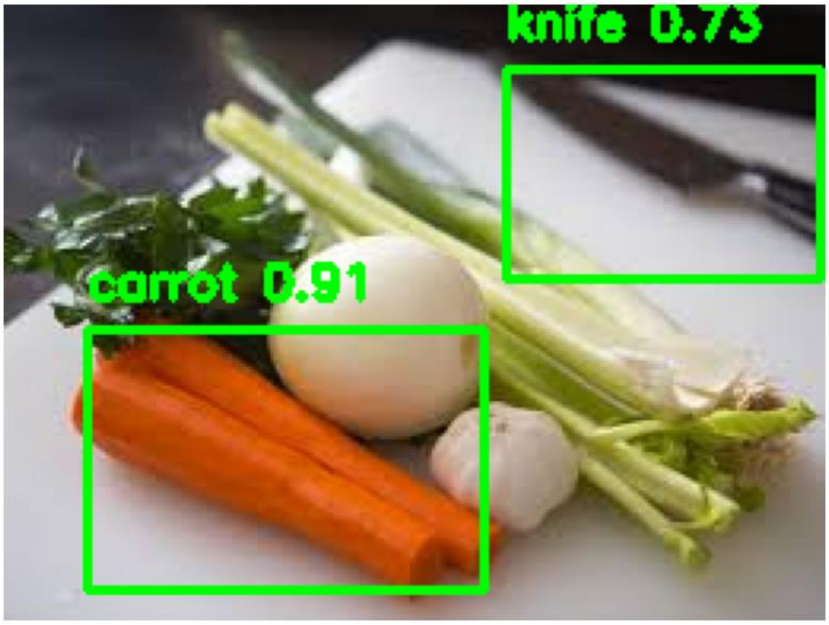
Epoch [5/5] Batch 6300/6313 Loss D: 0.3861 G: 3.7460

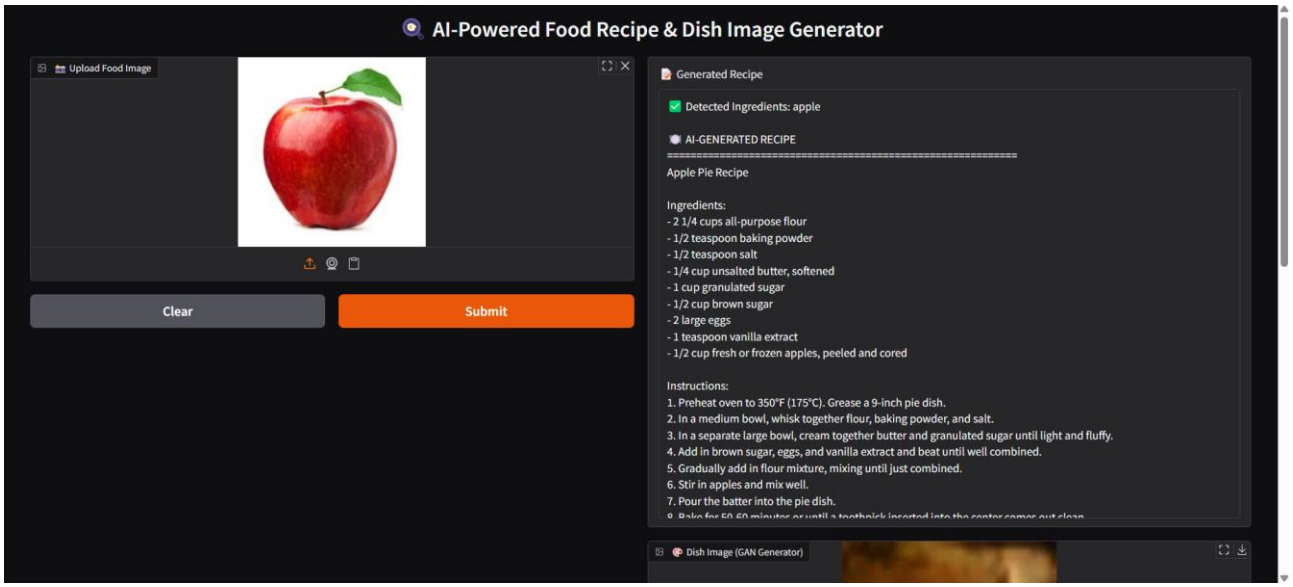
Epoch 5

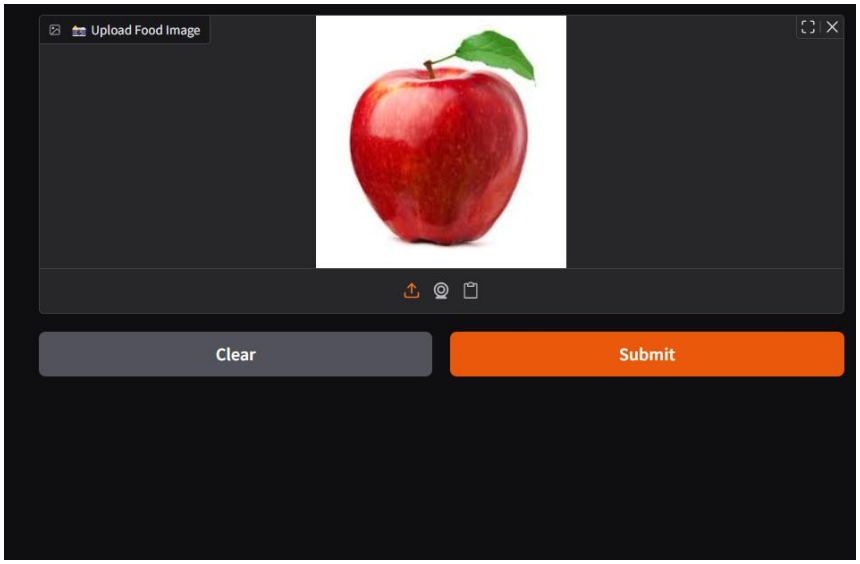
The project successfully achieves its goal of converting a food image into a complete cooking experience using AI:

- YOLOv3 detects key food ingredients from the uploaded image with good accuracy.
- LaMini-GPT generates a detailed and readable recipe using the detected ingredients.
- GAN creates low-resolution but realistic food-like images learned from the Food-101 dataset.
- Stable Diffusion produces high-quality, photorealistic dish images based on the recipe title.
- All outputs are presented in a clean, user-friendly Gradio interface.

Sample Outputs:







Generated Recipe

Detected Ingredients: apple

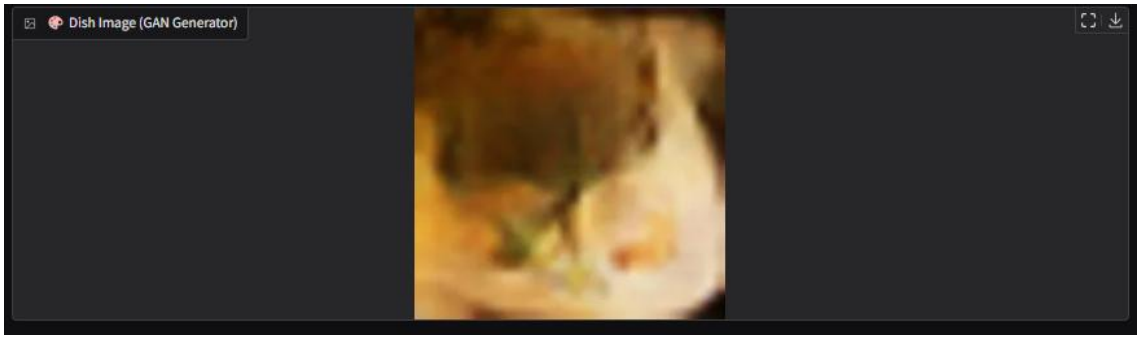
AI-GENERATED RECIPE

==========

Apple Pie Recipe

Ingredients:
- 2 1/4 cups all-purpose flour
- 1/2 teaspoon baking powder
- 1/2 teaspoon salt
- 1/4 cup unsalted butter, softened
- 1 cup granulated sugar
- 1/2 cup brown sugar
- 2 large eggs
- 1 teaspoon vanilla extract
- 1/2 cup fresh or frozen apples, peeled and cored

Instructions:
1. Preheat oven to 350°F (175°C). Grease a 9-inch pie dish.
2. In a medium bowl, whisk together flour, baking powder, and salt.
3. In a separate large bowl, cream together butter and granulated sugar until light and fluffy.
4. Add in brown sugar, eggs, and vanilla extract and beat until well combined.
5. Gradually add in flour mixture, mixing until just combined.
6. Stir in apples and mix well.
7. Pour the batter into the pie dish.
8. Bake for 50-60 minutes or until a toothpick inserted into the center comes out clean



Conclusion

This project showcases the power of combining multiple AI techniques—Object Detection, Natural Language Processing, and Generative Adversarial Networks—to deliver a complete food-related AI solution. From detecting ingredients in a food image to generating a personalized recipe and visualizing the final dish, the system offers a seamless and interactive experience. The use of GANs and Stable Diffusion demonstrates the creative capabilities of AI, while NLP ensures the output is both useful and human-readable. This work not only highlights the technical potential of AI but also shows how it can be applied in practical, everyday use cases such as cooking assistance and smart kitchen systems.