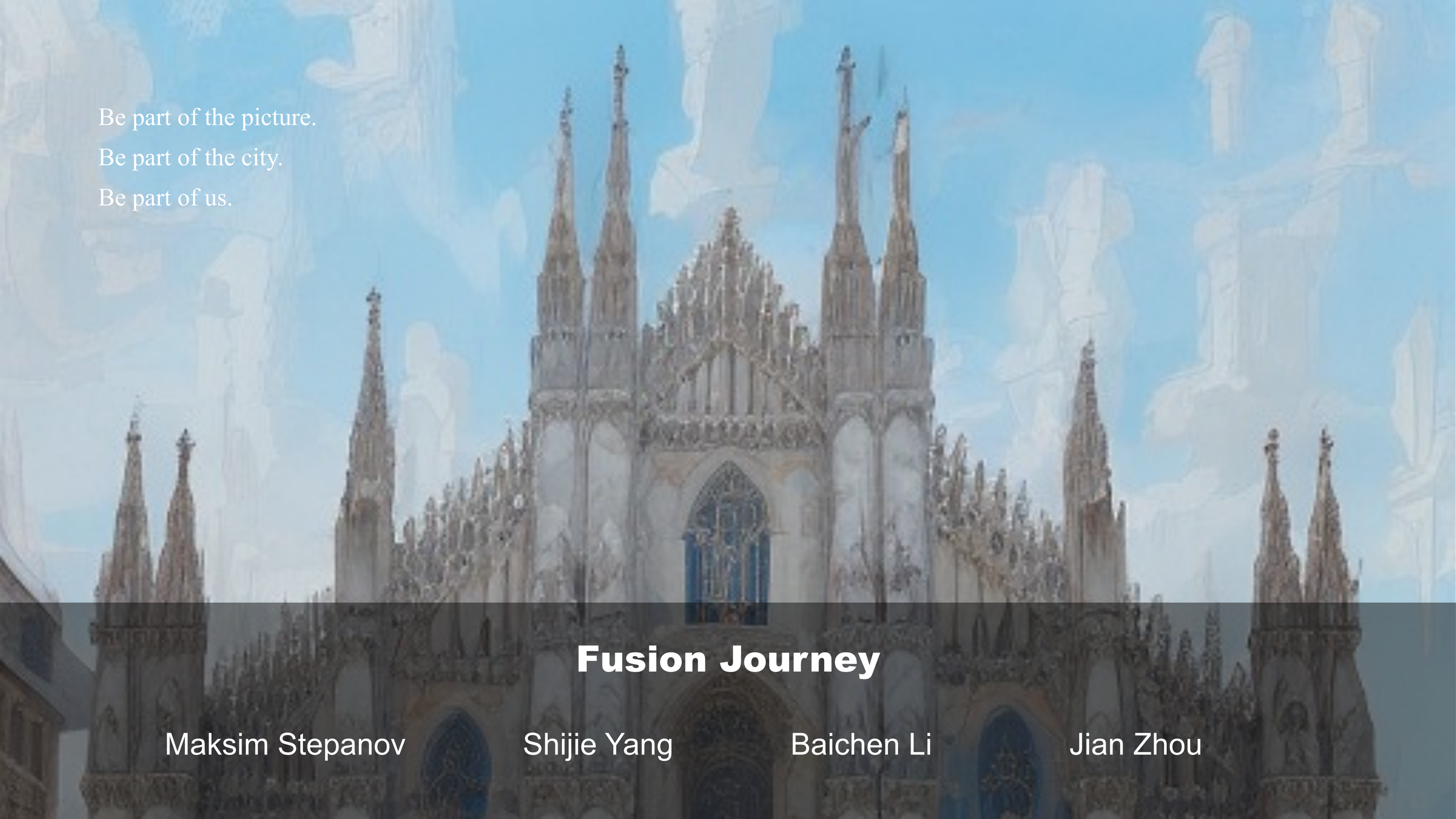


Fusion Journey

Maksim Stepanov Shijie Yang Baichen Li Jian Zhou

Introduction

Goal: Offer travelers a pleasant and engaging introduction to the city, by showcasing its scenery, landmarks, and culture.

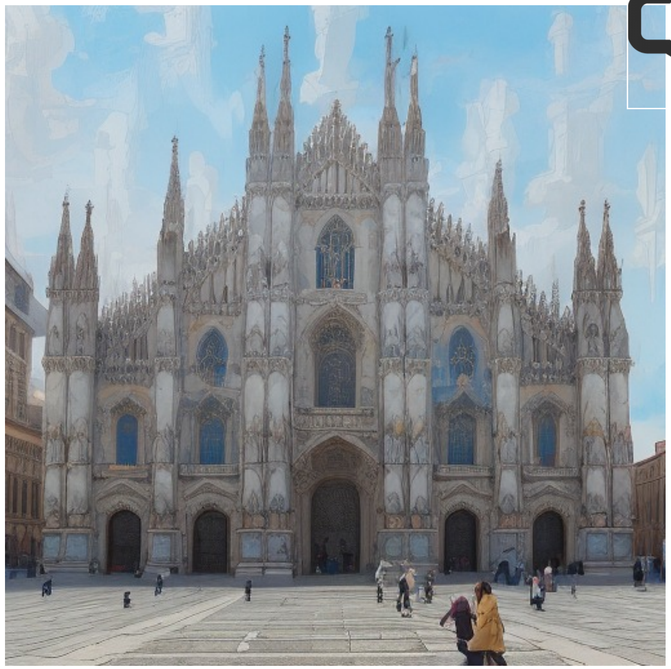
Where: Installed in the public facilities(train station, airport, etc.)

What: An art installation with a big screen, speaker, humidity and temperature sensor and a camera.

“Let everyone be part of the city”

User Experience

People detected in the camera



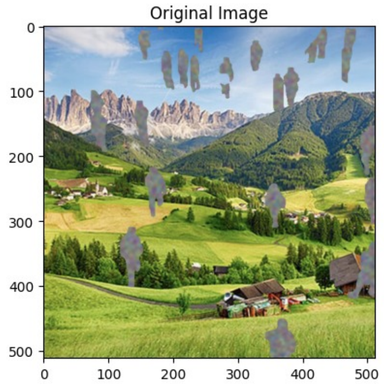
Transform them into artistic parameters of the picture

(clouds of the sky, trees or animals of the countryside.)

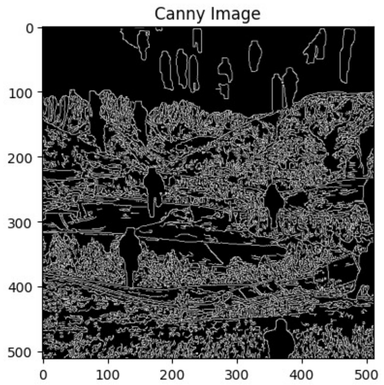
Technical Analysis

Technical Route

Object Detection



Overlap with base image



edge detection

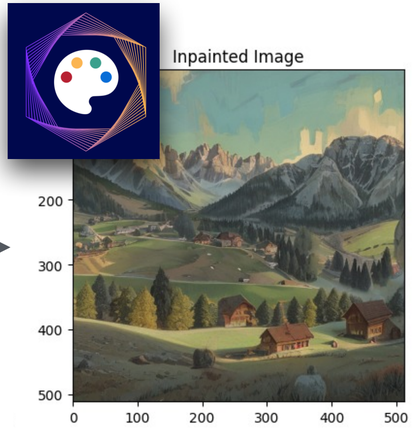
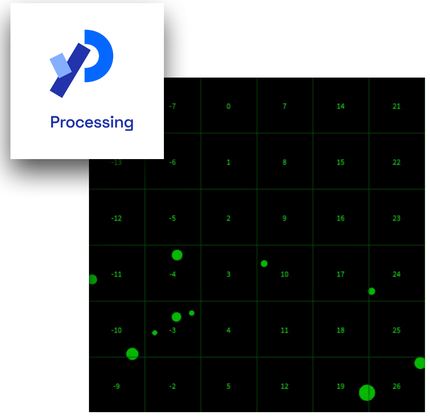


Image Generation

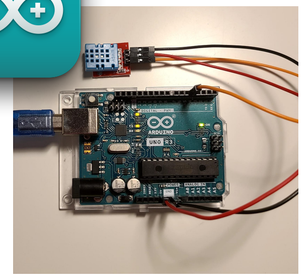
Prompts Generator

Realtime weather info



Coordinates Mapping

Audio Generation



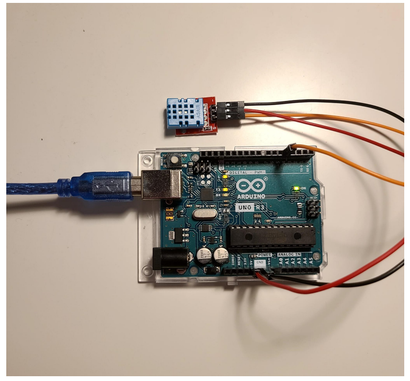
Temperature and Humidity Sensor

Image Generation

Environment perception

- **Real-Time Weather Monitoring:**

We use WeatherAPI to fetch real-time weather data from the most iconic locations in the city, then the information will be used to generate tailored prompts for image creation.



- **Dynamic Adjustment of Image Gen:**

Additionally, an Arduino system monitors the local temperature and humidity at the airport. If any significant change is detected locally, the system will trigger a new call to WeatherAPI for an updated weather report, ensuring that the image dynamically reflects both the cultural essence and current environmental state of the city.

Image Generation

Object segmentation & image generation

- **Camera Capture:**

Our system starts with a high-resolution camera installed at the top of large screens in the airport. These cameras continuously capture the flow of people moving through the area.

Fig. people become sheep in Dolomiti

- **Real-Time Object Segmentation with YOLO:**

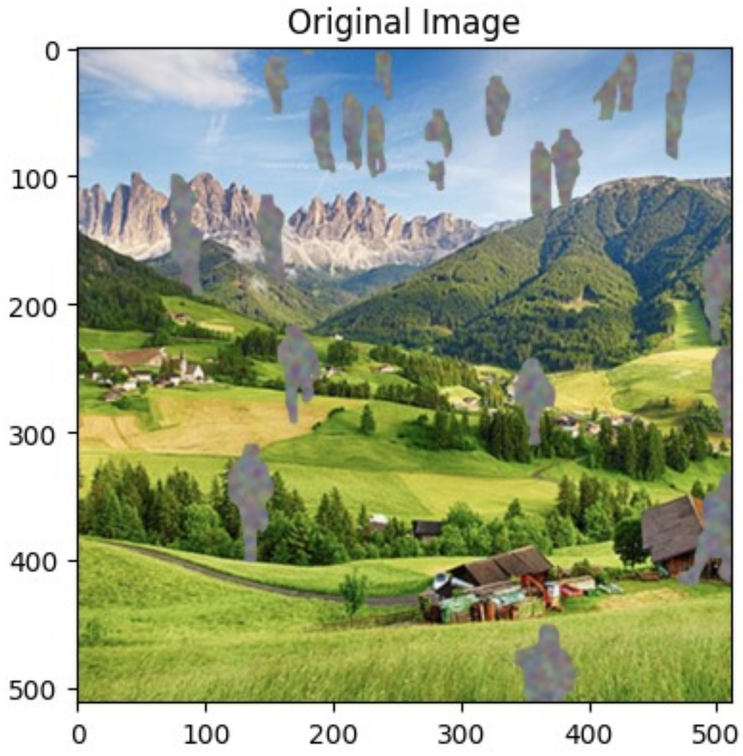
The video stream is processed in real-time using YOLO (You Only Look Once), an advanced object detection model. YOLO identifies and cuts out people's edge detected in the frame, then their centroid is calculated and stored for audio processing.

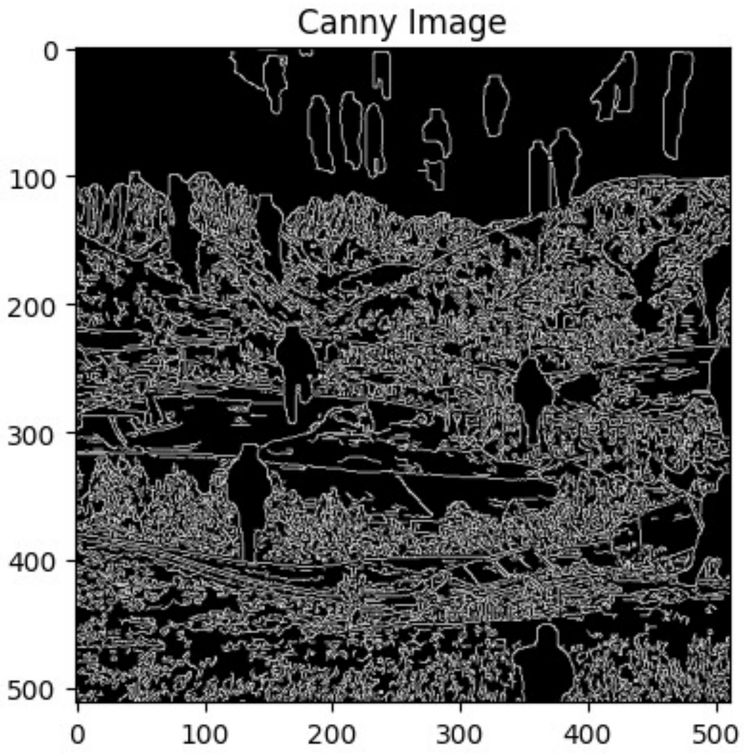
- **Image Transformation with Stable Diffusion:**

We then combine the segmentation mask with a base picture and send it to our diffusion model, and use with inpainting. Meanwhile we call gpt to generate a prompt based on time, weather, and city to create personalized and context-aware imagery where travelers can appear as elements like plants, objects, or animals.

Image Generation

An example





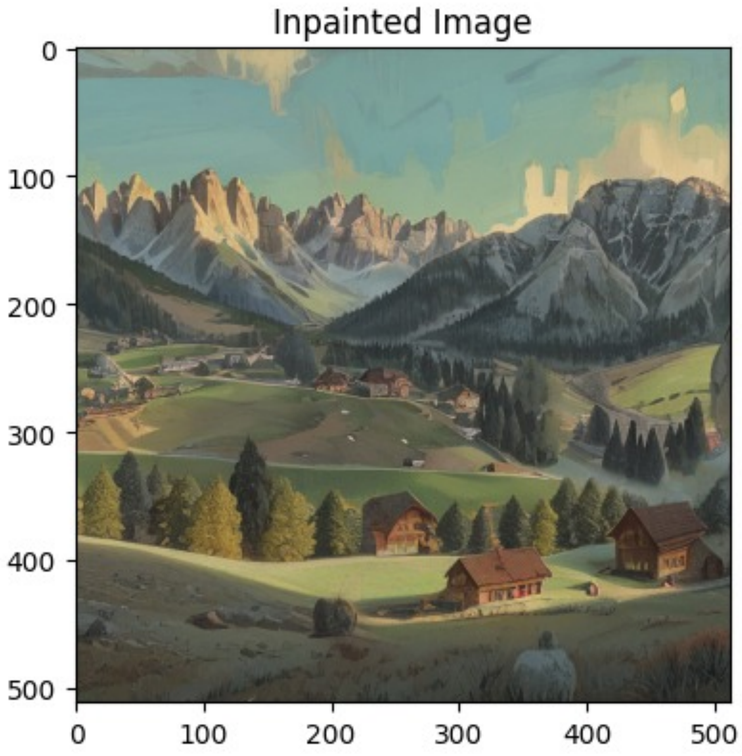


Image Generation

Prompt generation

- **ChatGPT API:**

Use the ChatGPT API to generate prompts for an image generation model. The system will provide GPT-4o with suitable system instructions, user input containing instructions, real-time weathers get from weather API and labels, plus an example. Then receive a prompt in return.

- **Example output:**

'Duomo di Milano under the early morning sky with broken clouds, the iconic cathedral subtly illuminated by streetlights and the quiet ambiance of the city at predawn hour with a few nocturnal birds in flight. cinematic, serene, atmospheric, 8k.'

Image Generation

Call ChatGPT API

- system_message = (

f"""You are a creative assistant specialized in generating informative prompts for image generative models like Stable Diffusion, DALL-E, etc. You are tasked with creating a prompt for a tourism scene showing the most famous landmark with the current weather conditions."""

- user_message = (

f"""Write a one sentence prompt showing the most famous tourism scene in {landmark} with the weather being {weather_description} at {temp_celsius:.1f}°C, observed at time {current_time}. Please accurately reflect the scene within the specified time range, given that the sun rises at {sunrise_time} and sun sets at {sunset_time}, but don't show the exact time and temperature in the prompt. Append suitable descriptive labels you think that may appear in the scene at the end, e.g. 'animals, sheep, birds, cinematic, scenery, 8k', etc. Here is an example of expected prompt: 'Duomo di Milano under the early morning sky with broken clouds, the iconic cathedral subtly illuminated by streetlights and the quiet ambiance of the city at predawn hour with a few nocturnal birds in flight. cinematic, serene, atmospheric, 8k."""

Image Generation

Call ChatGPT API

- **Ask GPT:**

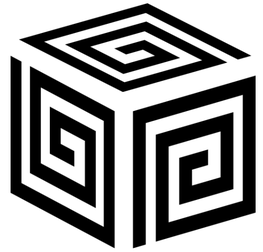
Write a one sentence prompt showing the most famous tourism scene in Duomo with the weather being few clouds at 21.6°C, observed at time 22:47:38. Please accurately reflect the scene within the specified time range, given that the sun rises at 07:34:40 and sun sets at 23:11:03, but don't show the exact time and temperature in the prompt. Append suitable descriptive labels you think that may appear in the scene at the end, e.g. 'animals, sheep, birds, cinematic, scenery, 8k', etc. Here is an example of expected prompt: 'Duomo di Milano under the early morning sky with broken clouds, the iconic cathedral subtly illuminated by streetlights and the quiet ambiance of the city at predawn hour with a few nocturnal birds in flight. cinematic, serene, atmospheric, 8k.'

- **GPT Response:**

The majestic Duomo di Milano bathed in the cool evening glow under a slightly clouded sky, the grand cathedral softly highlighted by city lights and the tranquil atmosphere of a late-night scene with a few people strolling nearby. cinematic, serene, architectural, nightscape, 8k.

Sound Generation

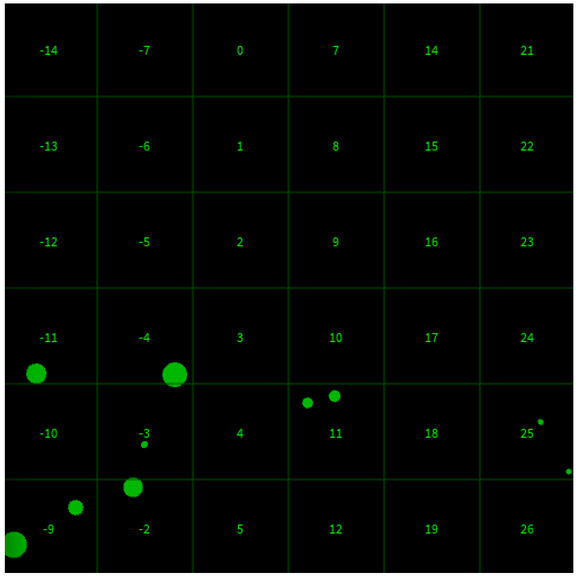
YOLOv8.2



Yolo segmentation

Coordinates

OSC



Tone Grid

Tone value

Amplitude

OSC

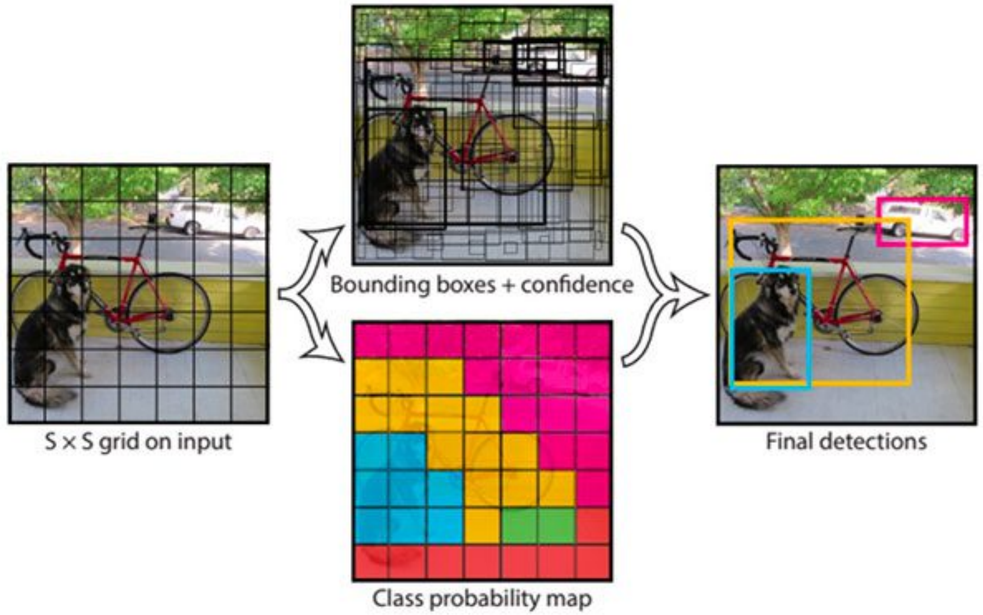
SynthDef

```
(
SynthDef(\bpfsaw, {
    arg atk=1, sus=0, rel=0, c1=1, c2=(-4),
    freq=500, cf=700, rq=0.2, amp=1, out=0;
    var sig, env;
    env = EnvGen.kr(Env([0,1,1,0],[atk,sus,rel],[c1,0,c2]),doneAction:2);
    sig = Saw.ar(freq);
    sig = BPF.ar(sig,cf,rq);
    sig = sig * env * amp;
    Out.ar(out, sig);
 }).add
)
```

You Only Look Once

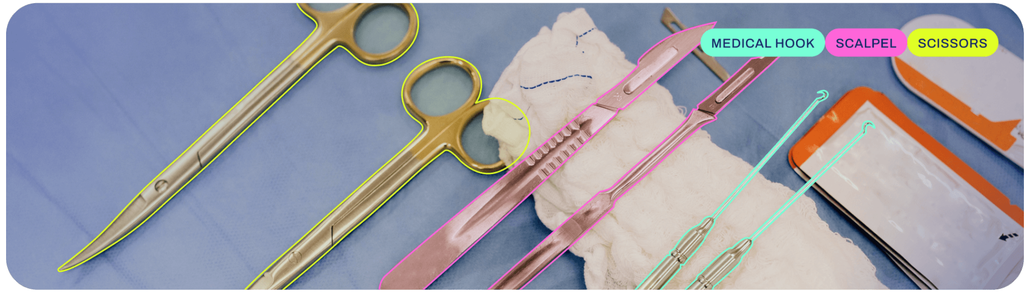
is a real-time object detection system.

process the entire image in a single shot



`yolov8v.pt`

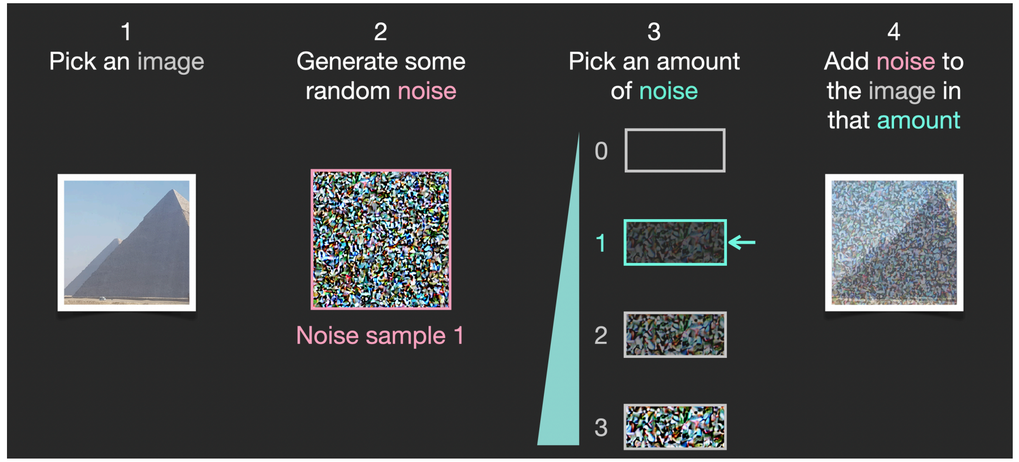
Classic YOLO design



`yolov8n-seg.pt`

Instant segmentatinon

Stable Diffution



Classic

Our way

Demo

https://www.youtube.com/watch?v=QsykiUDhq9U

Future Improvements

User Interaction & Technology

- **Interactive Learning Modules**

We plan to incorporate educational modules into the displays, providing travelers with engaging learning opportunities about local history and culture directly on their personal devices, e.g. adding QR codes.

- **Improved Real-Time Processing**

If we have a faster graphic card, it is possible to enhance the real-time processing capability, reduce latency and increase the responsiveness of our installations to real-time inputs.

Thank you!

Shijie Yang, Maksim Stepanov, Jian Zhou, Baichen Li