Crafting Landscape Videos using Artificial Intelligence Models

Felipe Dias • Zanoni Dias • Helio Pedrini

CHAPTER I

INTRODUCTION

Terravision
ART+COM
Berlin, 1993

Chat GPT

November 30th, 2022
Generative Models

Chat GPT

November 30th, 2022
Generative Models

Apple Vision

February 2nd, 2024
Augmented Reality

Chat GPT

November 30th, 2022
Generative Models

Apple Vision

February 2nd, 2024
Augmented Reality

CHAPTER II

BACKGROUND

Cinema
Elementals from Pixar

Procedural Techniques for Large, Dynamic Sets in *Elemental*

Mike Ravella
mvr@pixar.com
Pixar Animation Studios

Aylwin Villanueva
aylwin@pixar.com
Pixar Animation Studios

Brandon Montell
montell@pixar.com
Pixar Animation Studios

Ting Zhang
tingzhang@pixar.com
Pixar Animation Studios

Hosuk Chang
hosuk@pixar.com
Pixar Animation Studios

ABSTRACT

The world of Pixar's film *Elemental* is inhabited by characters made of fire, water, air and earth, and we needed to give these characters a home that was just as dynamic as they were. Specifically, we needed to build a city with new and distinct forms of architecture for each element, fill this city with fire, water, smoke and vegetation, and add animation to make everything feel alive in the way a bustling city should. In this talk, we present our techniques for handling problems of scale, such as parameterized building generation and dressing, application of a variety of fx elements within large sets, as well as some novel approaches for automated color palette generation, both in an asset and shot context.

KEYWORDS

environments, lookdev, production design, houdini, proceduralism

ACM Reference Format:
Mike Ravella, Aylwin Villanueva, Brandon Montell, Ting Zhang, and Hosuk Chang. 2023. Procedural Techniques for Large, Dynamic Sets in *Elemental*. In *Proceedings of SIGGRAPH Talks*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

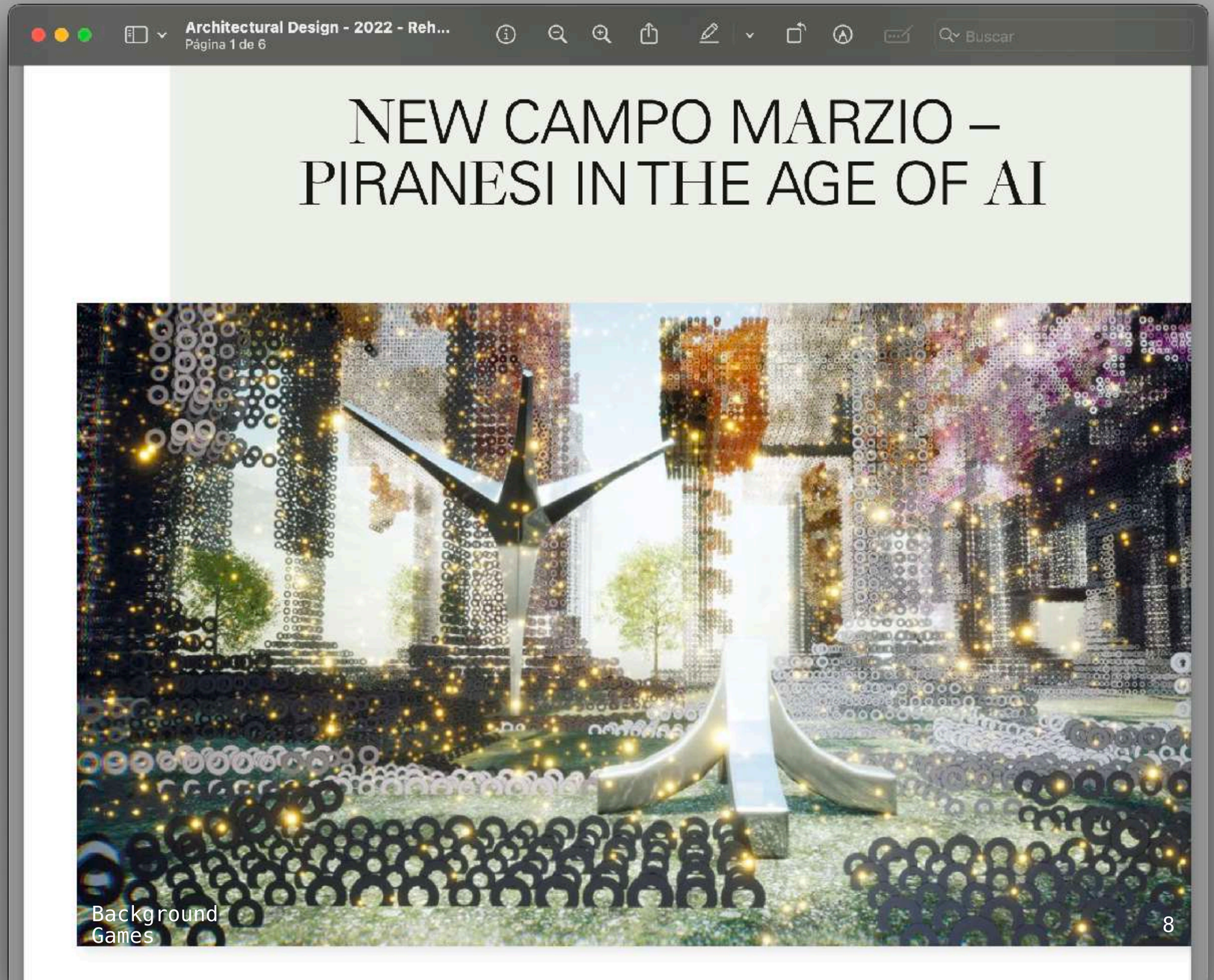
1 PALETTE BASED SHADING

Fire Town is a large city in Elemental consisting of stylized buildings for fire characters. The building shapes are very distinct from past Pixar films, and the number of Fire Town building designs was relatively limited. One way that we added variation was by anced buildings using an image-based color picking e the palette of the city was generated from one or more input images. One of three materials is assigned to each

Figure 1: A section of procedural city with a palette based on the images of the three people below. ©Pixar.

Figure 2: The left image represents how this set was shaded

Games
New Campo Marzio



Technology
Movie Gen by Meta

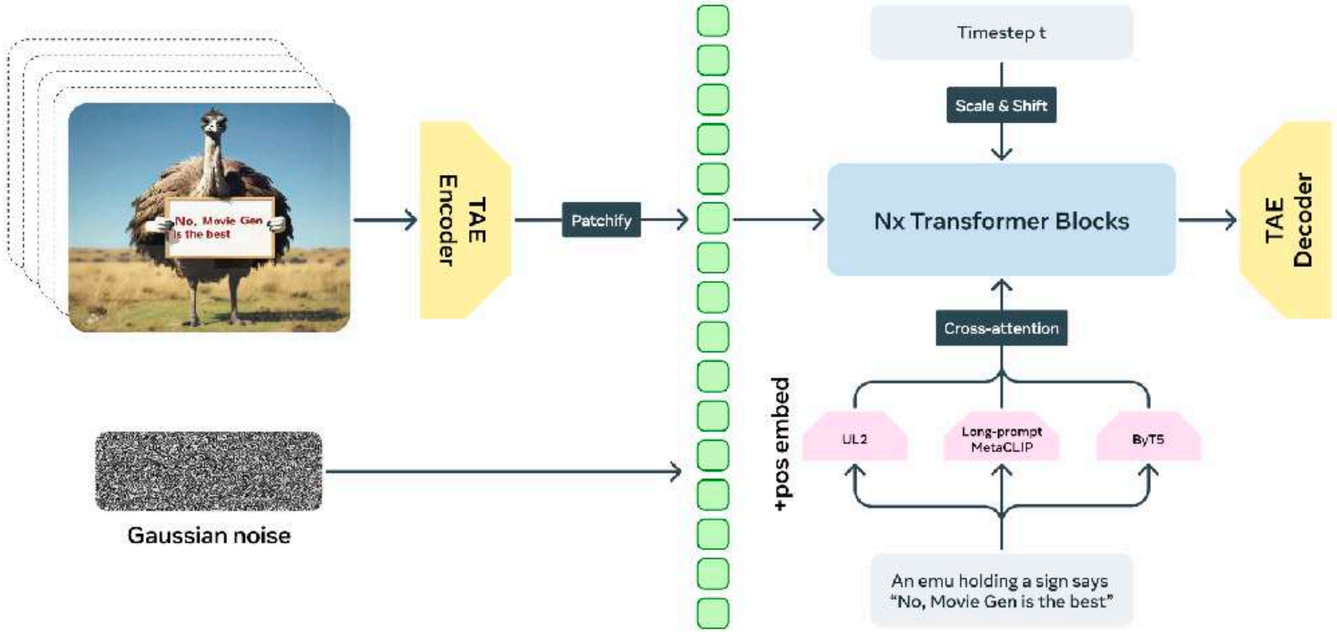


Figure 3 Overview of the joint image and video generation pipeline. We train our generative model on a spatio-temporally compressed latent space, which is learned via a temporal autoencoder model (TAE). User-provided text prompts are encoded using pre-trained text-encoders, and used as conditioning. Our generative model takes sampled Gaussian noise and all provided conditioning as input, and generates an output latent, which is decoded to an output image or video using the TAE decoder.

3.1 Image and Video Foundation Model

We describe the key components of the MOVIE GEN VIDEO model—the spatio-temporal autoencoder (TAE), the training objective for image and video generation, model architecture, and the model scaling techniques we use in our work.

Background
Technology

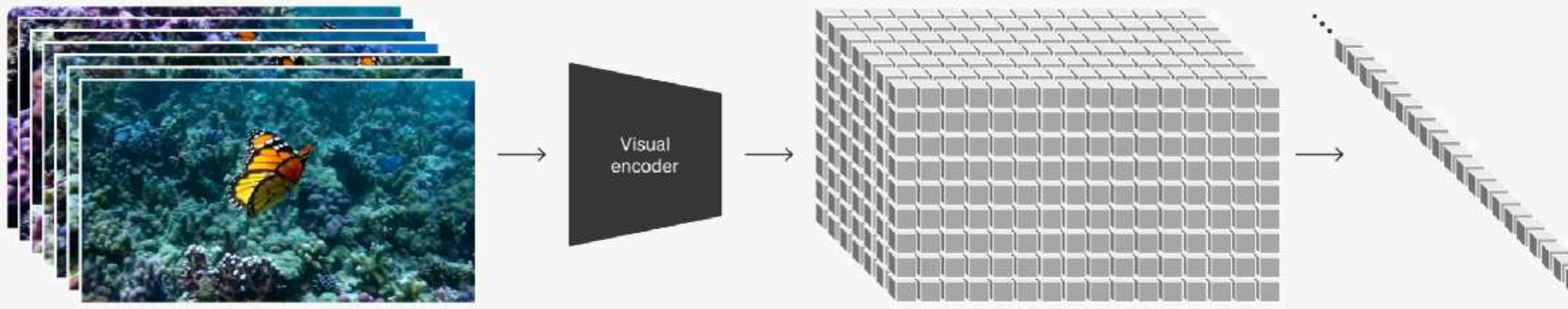
l Autoencoder (TAE)

Technology
OpenAI Sora

openai.com

Turning visual data into patches

We take inspiration from large language models which acquire generalist capabilities by training on internet-scale data.[13, 14] The success of the LLM paradigm is enabled in part by the use of tokens that elegantly unify diverse modalities of text—code, math and various natural languages. In this work, we consider how generative models of visual data can inherit such benefits. Whereas LLMs have text tokens, Sora has visual *patches*. Patches have previously been shown to be an effective representation for models of visual data.[15, 16, 17, 18] We find that patches are a highly-scalable and effective representation for training generative models on diverse types of videos and images.



At a high level, we turn videos into patches by first compressing videos into a lower-dimensional latent space,[19] and subsequently decomposing the representation into spacetime patches.

Rio de Janeiro landscape clear sky, lush vegetation, few boats in the water, calm sea.

Rio de Janeiro landscape clear sky, vegetation covered in snow, frozen water, sunset lights, calm sea.

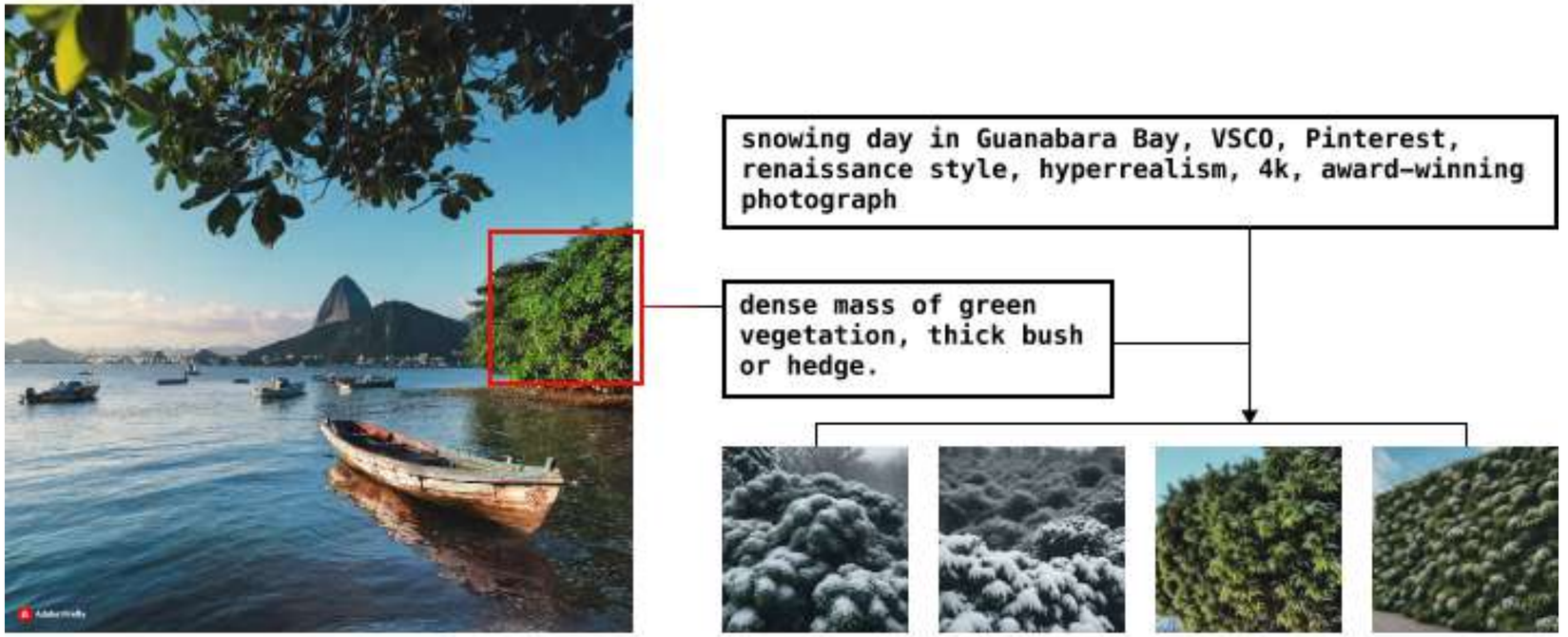
Research Questions

- What methodologies and techniques can be employed to exert greater control over the specific visual elements that AI hallucinates and reproduces?

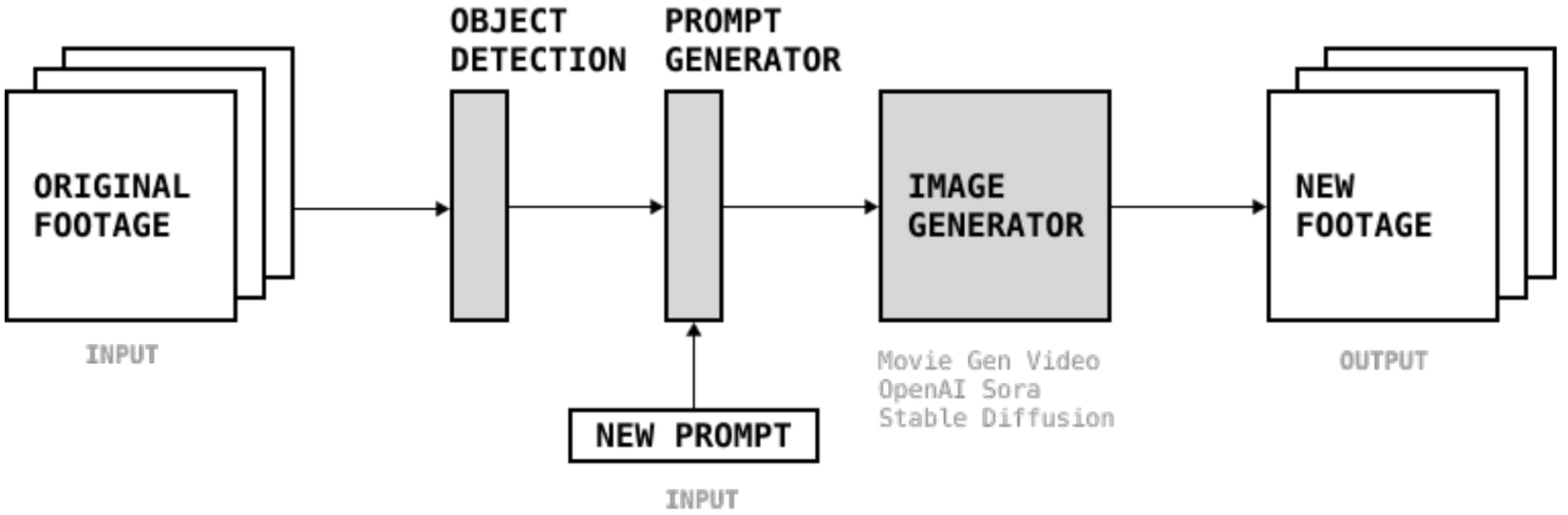
- How does leveraging additional data sources beyond the prompt contribute to enhancing the precision and accuracy of AI-generated visual content by providing contextual and semantic information?

CHAPTER III

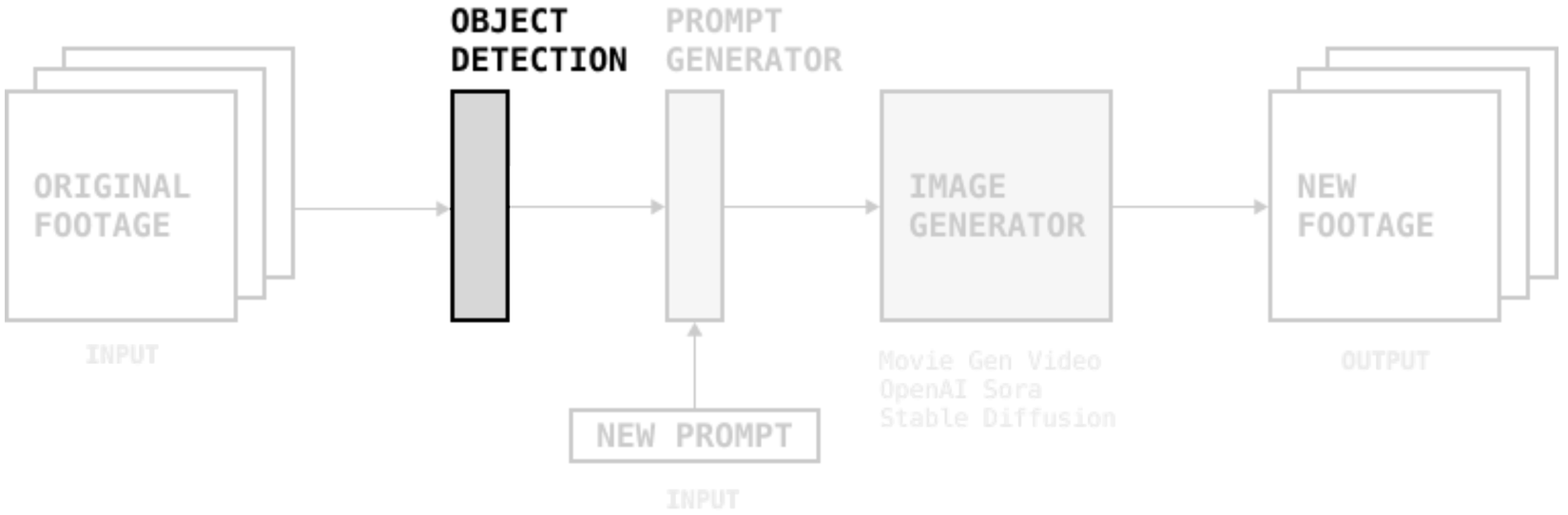
METHODOLOGY

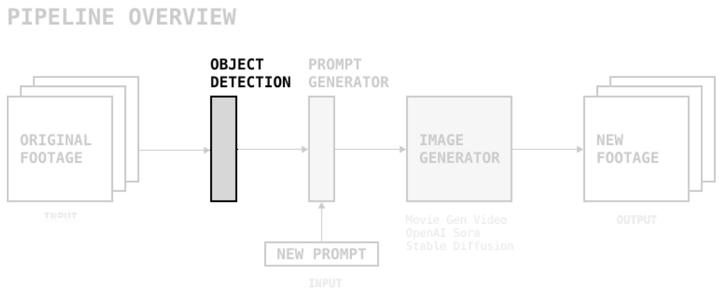

snowing day in Guanabara Bay, VSCO, Pinterest, renaissance style, hyperrealism, 4k, award-winning photograph
dense mass of green vegetation, thick bush or hedge.

PIPELINE OVERVIEW

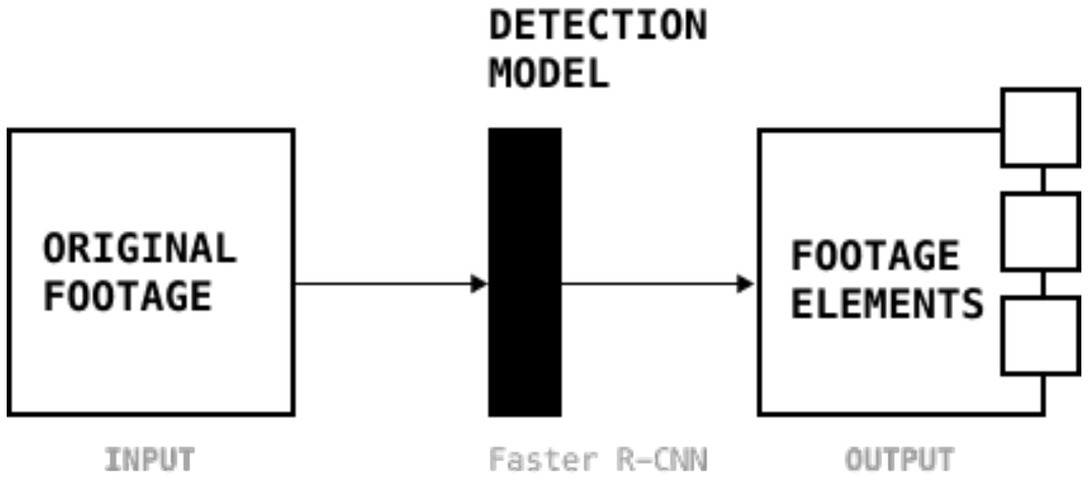


PIPELINE OVERVIEW

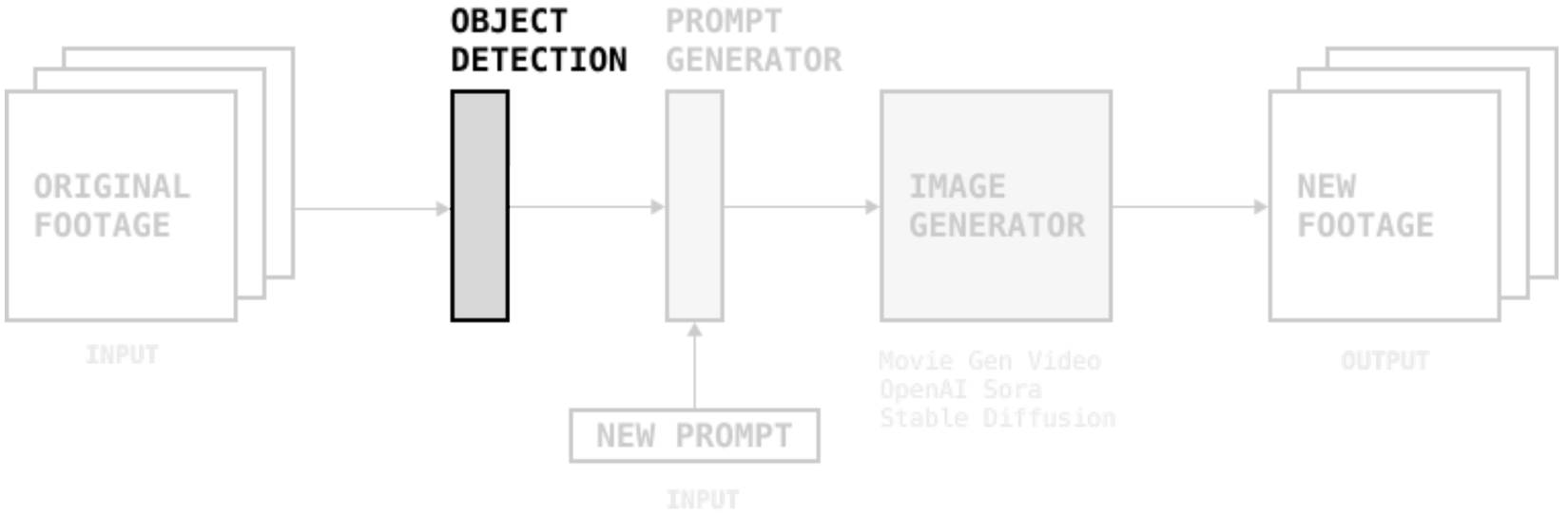




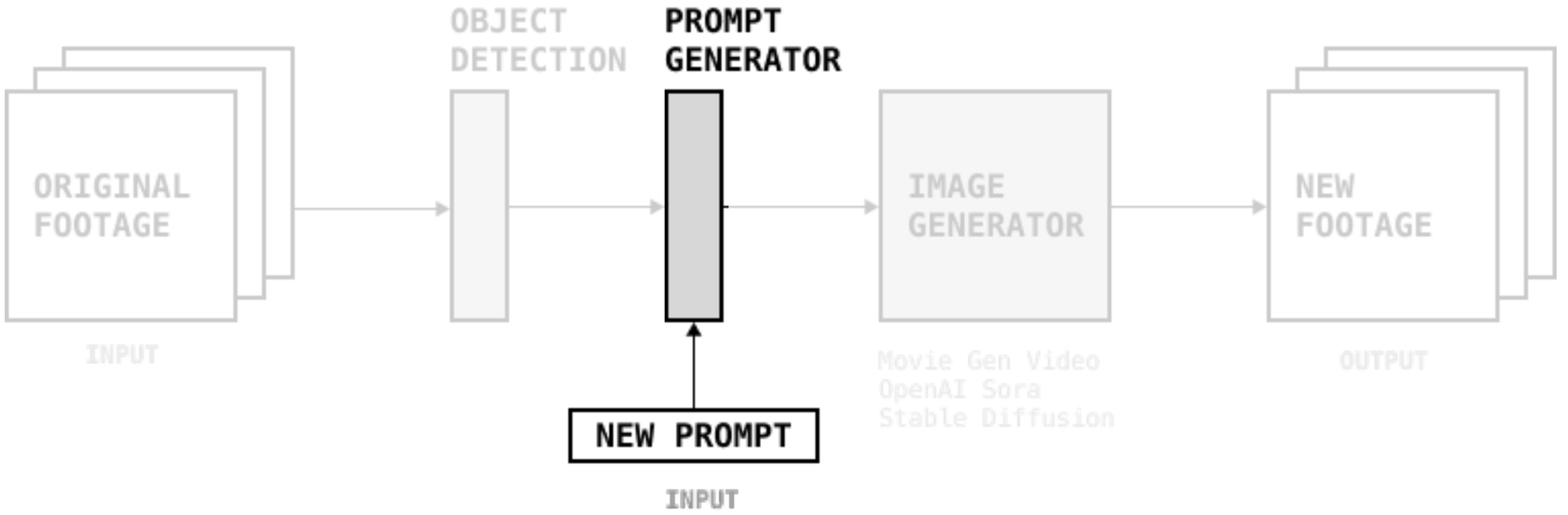
OBJECT DETECTION

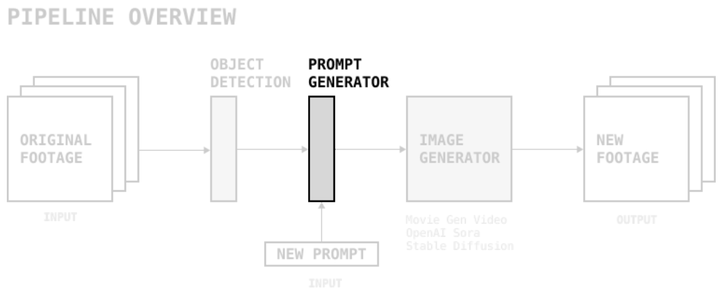


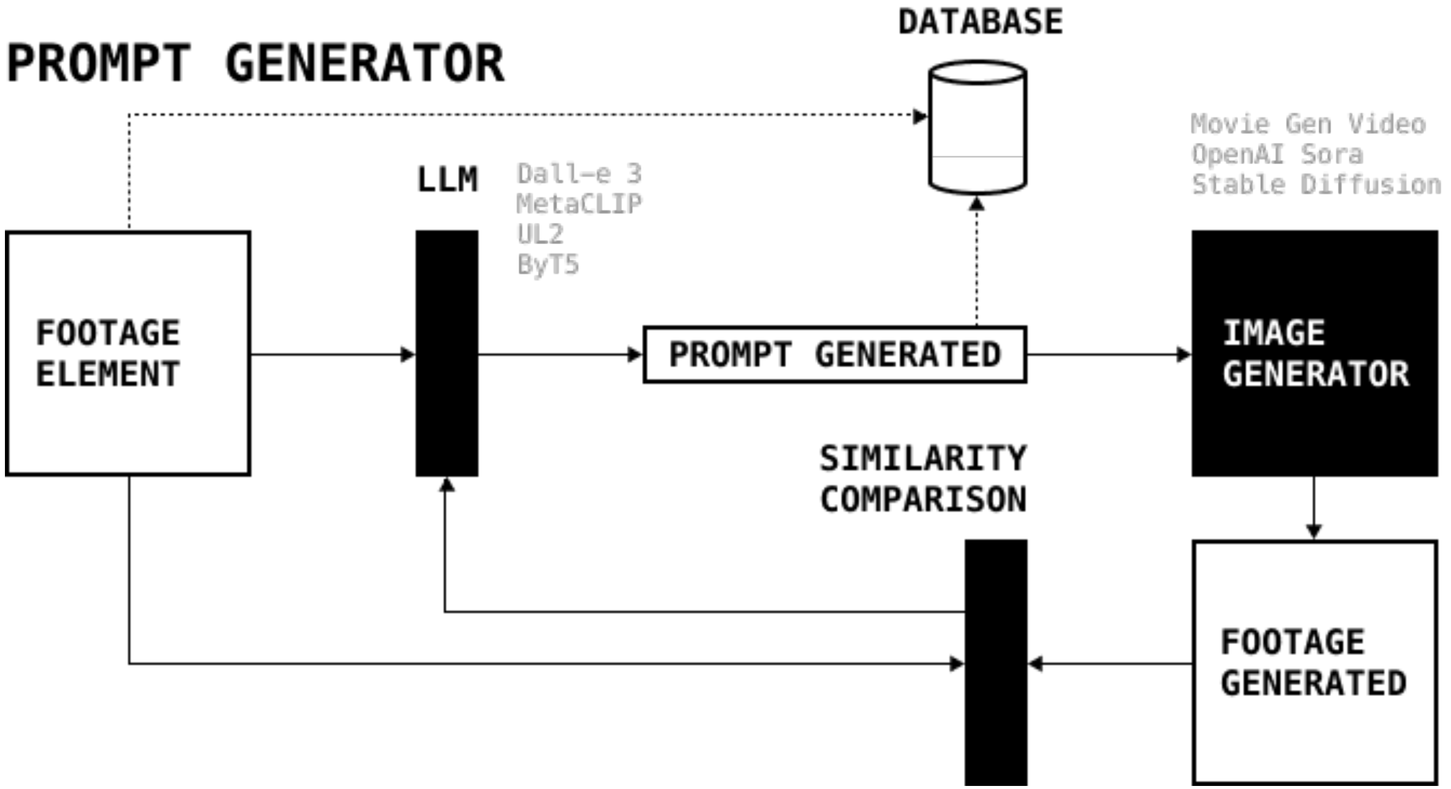
PIPELINE OVERVIEW

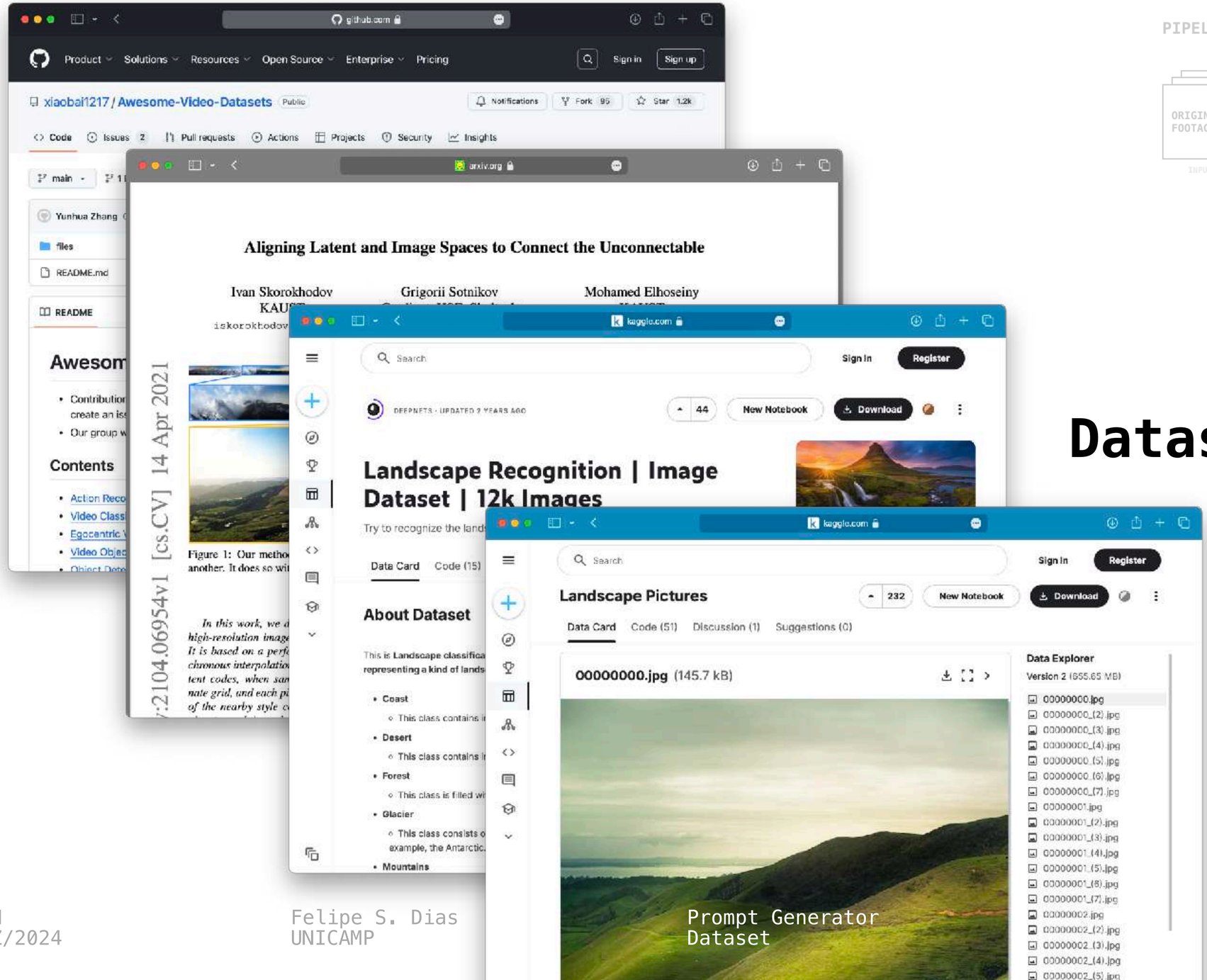


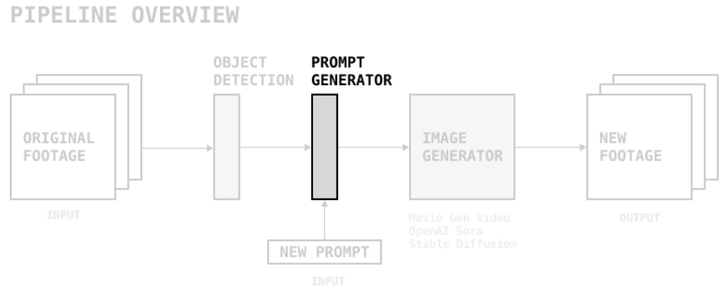
PIPELINE OVERVIEW




PIPELINE OVERVIEW
ORIGINAL FOOTAGE
INPUT
OBJECT DETECTION
PROMPT GENERATOR
NEW PROMPT
INPUT
IMAGE GENERATOR
Movie Gen Video
OpenAI Sora
Stable Diffusion
NEW FOOTAGE
OUTPUT

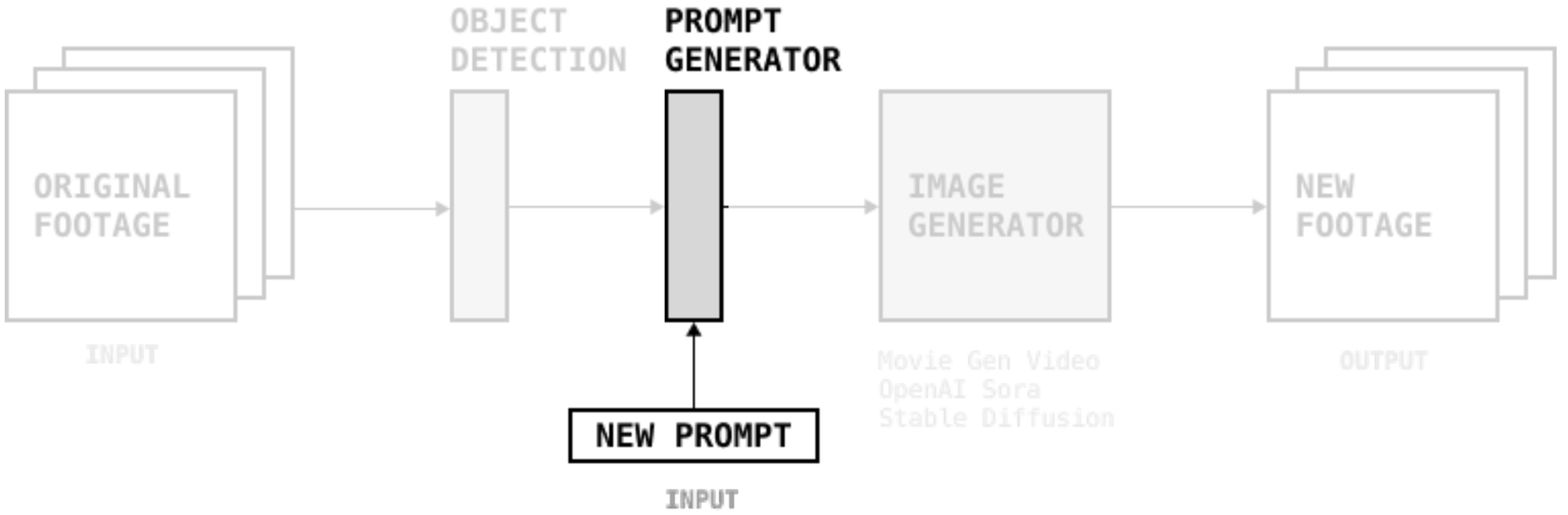

PROMPT GENERATOR
DATABASE
LLM
Dall-e 3
MetaCLIP
UL2
ByT5
Movie Gen Video
OpenAI Sora
Stable Diffusion
FOOTAGE ELEMENT
PROMPT GENERATED
IMAGE GENERATOR
SIMILARITY COMPARISON
FOOTAGE GENERATED


github.com
Product
Solutions
Resources
Open Source
Enterprise
Pricing
Sign in
Sign up
xiaobai1217 / Awesome-Video-Datasets
Public
Notifications
Fork 95
Star 1.2k
Code
Issues 2
Pull requests
Actions
Projects
Security
Insights
main
Yunhua Zhang
files
README.md
README
Awesom
Contents
arxiv.org
Aligning Latent and Image Spaces to Connect the Unconnectable
Ivan Skorokhodov
Grigorii Sotnikov
Mohamed Elhoseiny
iskorokhodov
Figure 1: Our metho
another. It does so wit
In this work, we
high-resolution image
It is based on a perf
chronous interpolatio
tent codes, when sa
nate grid, and each pi
of the nearby style c
v:2104.06954v1 [cs.CV] 14 Apr 2021
kaggle.com
Search
Sign In
Register
DEEPNETS · UPDATED 2 YEARS AGO
44
New Notebook
Download
Landscape Recognition | Image Dataset | 12k Images
Try to recognize the land
Data Card
Code (15)
About Dataset
This is Landscape classifica
representing a kind of lands
Coast
This class contains
Desert
This class contains
Forest
This class is filled wi
Glacier
This class consists o
example, the Antarctic
Mountains
kaggle.com
Search
Sign In
Register
Landscape Pictures
232
New Notebook
Download
Data Card
Code (51)
Discussion (1)
Suggestions (0)
00000000.jpg (145.7 kB)
Data Explorer
Version 2 (655.65 MB)
00000000.jpg
00000000_(2).jpg
00000000_(3).jpg
00000000_(4).jpg
00000000_(5).jpg
00000000_(6).jpg
00000000_(7).jpg
00000001.jpg
00000001_(2).jpg
00000001_(3).jpg
00000001_(4).jpg
00000001_(5).jpg
00000001_(6).jpg
00000001_(7).jpg
00000002.jpg
00000002_(2).jpg
00000002_(3).jpg
00000002_(4).jpg
00000002_(5).jpg
Prompt Generator
Dataset

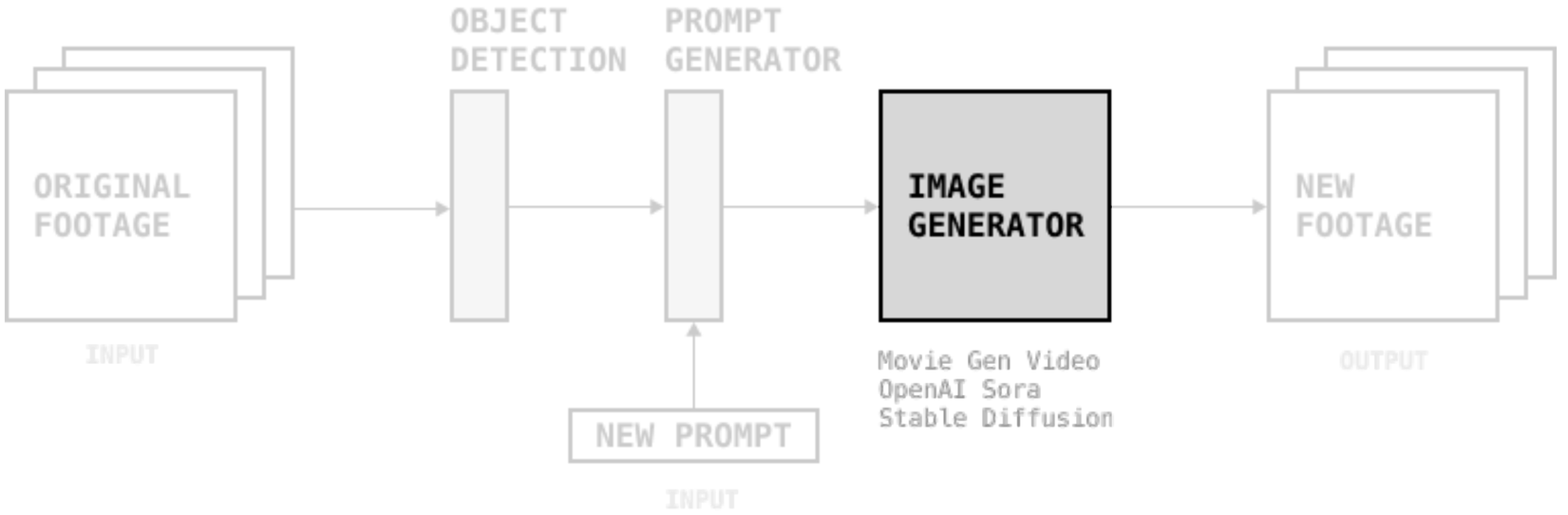

PIPELINE OVERVIEW
ORIGINAL FOOTAGE
INPUT
OBJECT DETECTION
PROMPT GENERATOR
NEW PROMPT
INPUT
IMAGE GENERATOR
Movie Gen Video
OpenAI Sora
Stable Diffusion
NEW FOOTAGE
OUTPUT

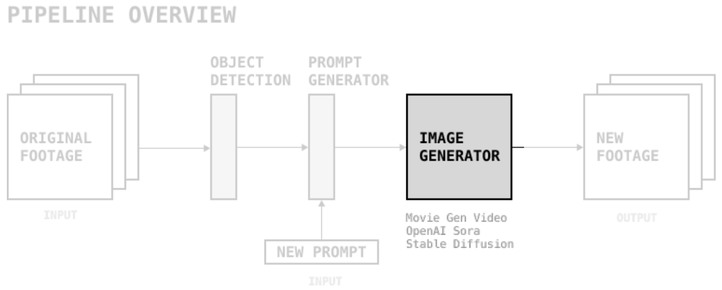
Dataset

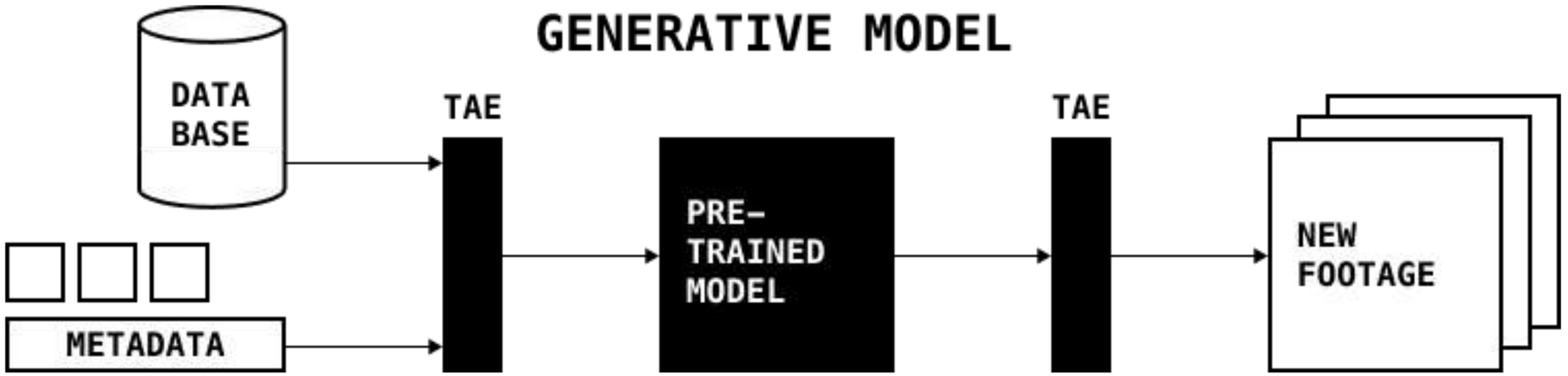
PIPELINE OVERVIEW

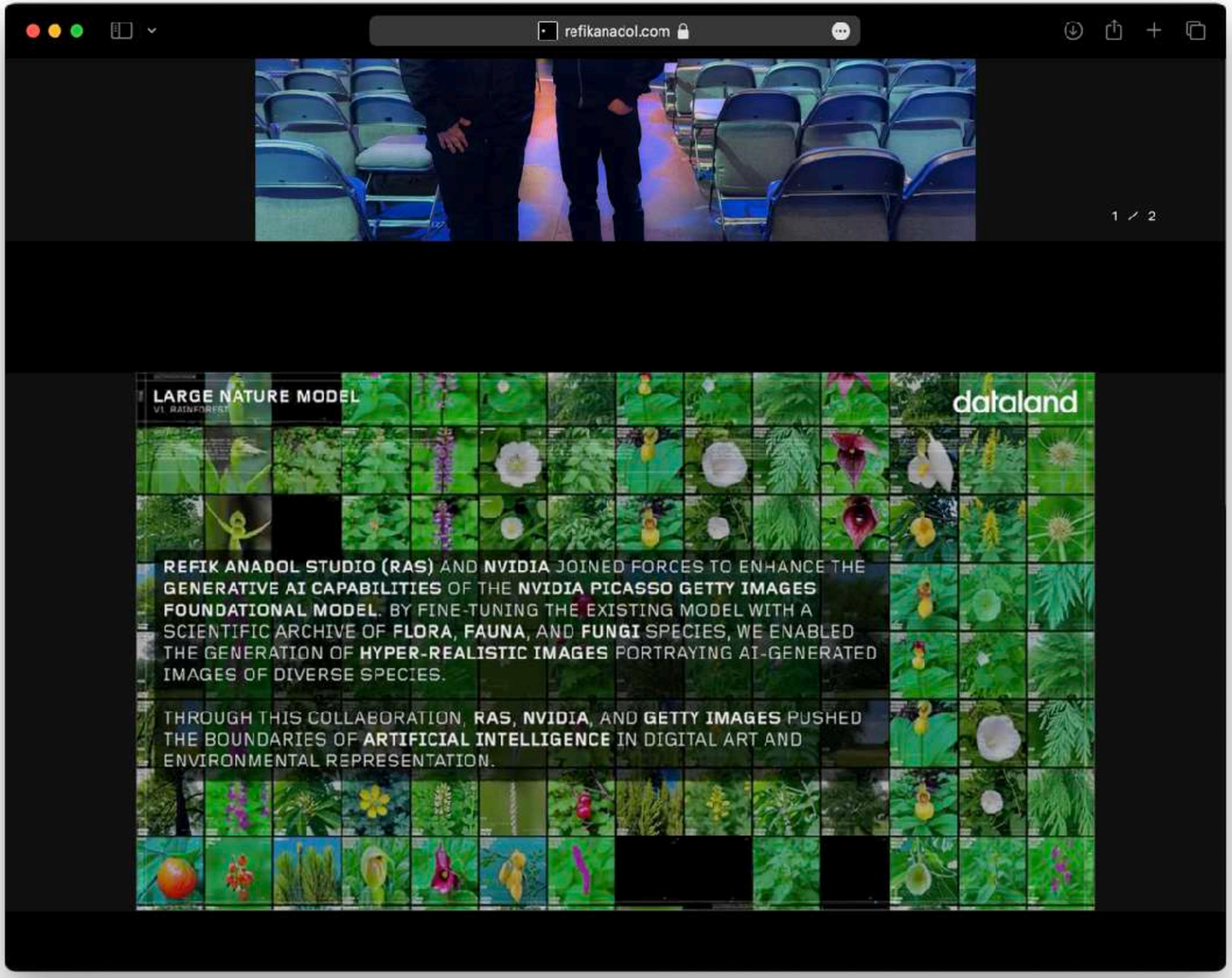


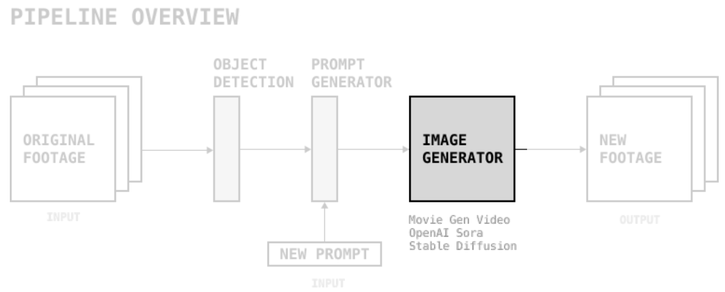
PIPELINE OVERVIEW




PIPELINE OVERVIEW
ORIGINAL FOOTAGE
INPUT
OBJECT DETECTION
PROMPT GENERATOR
IMAGE GENERATOR
Movie Gen Video
OpenAI Sora
Stable Diffusion
NEW PROMPT
INPUT
NEW FOOTAGE
OUTPUT

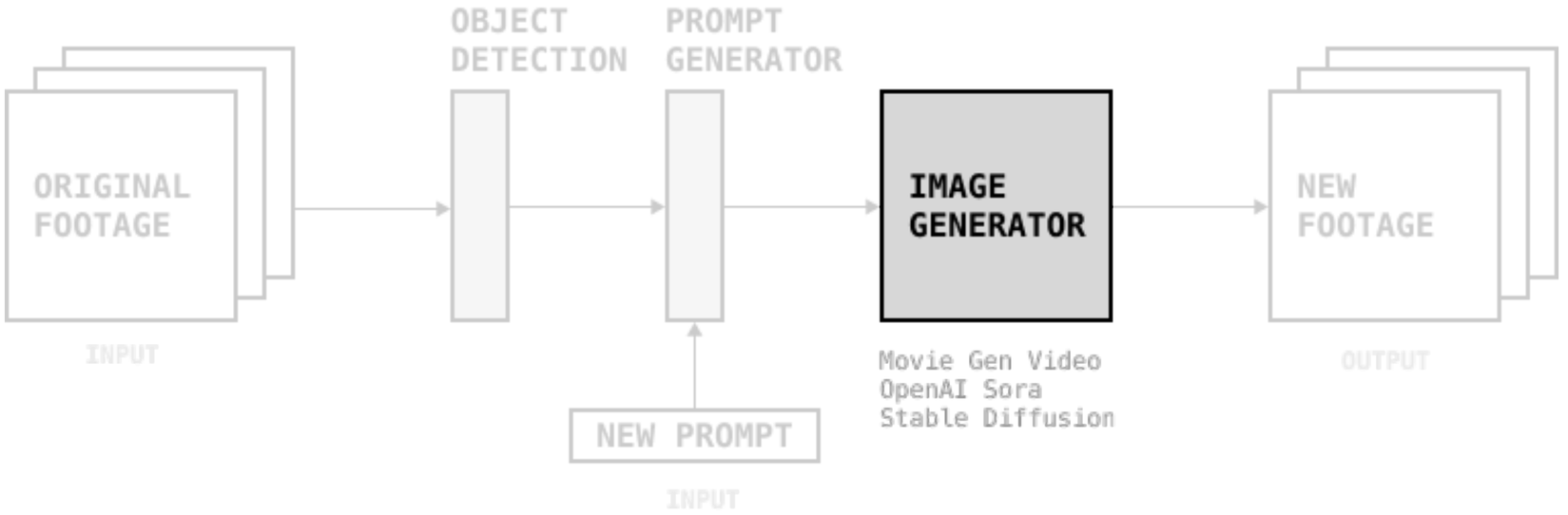

GENERATIVE MODEL
DATA BASE
METADATA
TAE
PRE-TRAINED MODEL
TAE
NEW FOOTAGE


refikanacol.com
1 / 2
LARGE NATURE MODEL
V1 RAINFOREST
dataland
REFIK ANADOL STUDIO (RAS) AND NVIDIA JOINED FORCES TO ENHANCE THE GENERATIVE AI CAPABILITIES OF THE NVIDIA PICASSO GETTY IMAGES FOUNDATIONAL MODEL. BY FINE-TUNING THE EXISTING MODEL WITH A SCIENTIFIC ARCHIVE OF FLORA, FAUNA, AND FUNGI SPECIES, WE ENABLED THE GENERATION OF HYPER-REALISTIC IMAGES PORTRAYING AI-GENERATED IMAGES OF DIVERSE SPECIES.
THROUGH THIS COLLABORATION, RAS, NVIDIA, AND GETTY IMAGES PUSHED THE BOUNDARIES OF ARTIFICIAL INTELLIGENCE IN DIGITAL ART AND ENVIRONMENTAL REPRESENTATION.

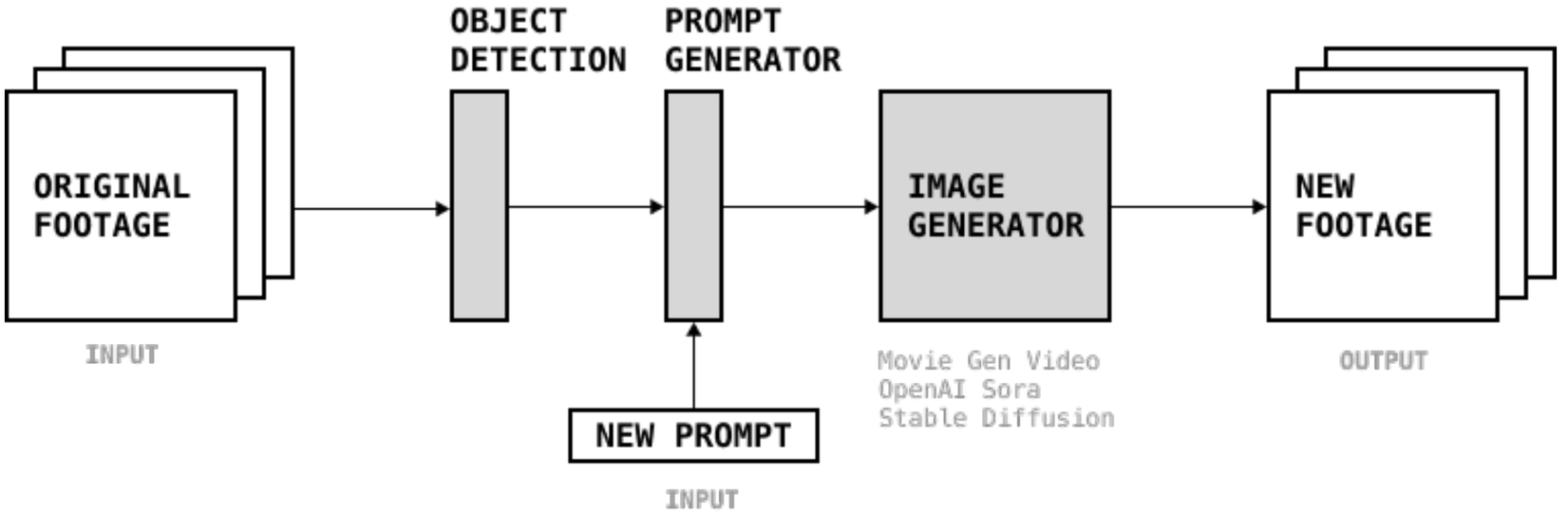

PIPELINE OVERVIEW
ORIGINAL FOOTAGE
INPUT
OBJECT DETECTION
PROMPT GENERATOR
IMAGE GENERATOR
Movie Gen Video
OpenAI Sora
Stable Diffusion
NEW PROMPT
INPUT
NEW FOOTAGE
OUTPUT

Dataset

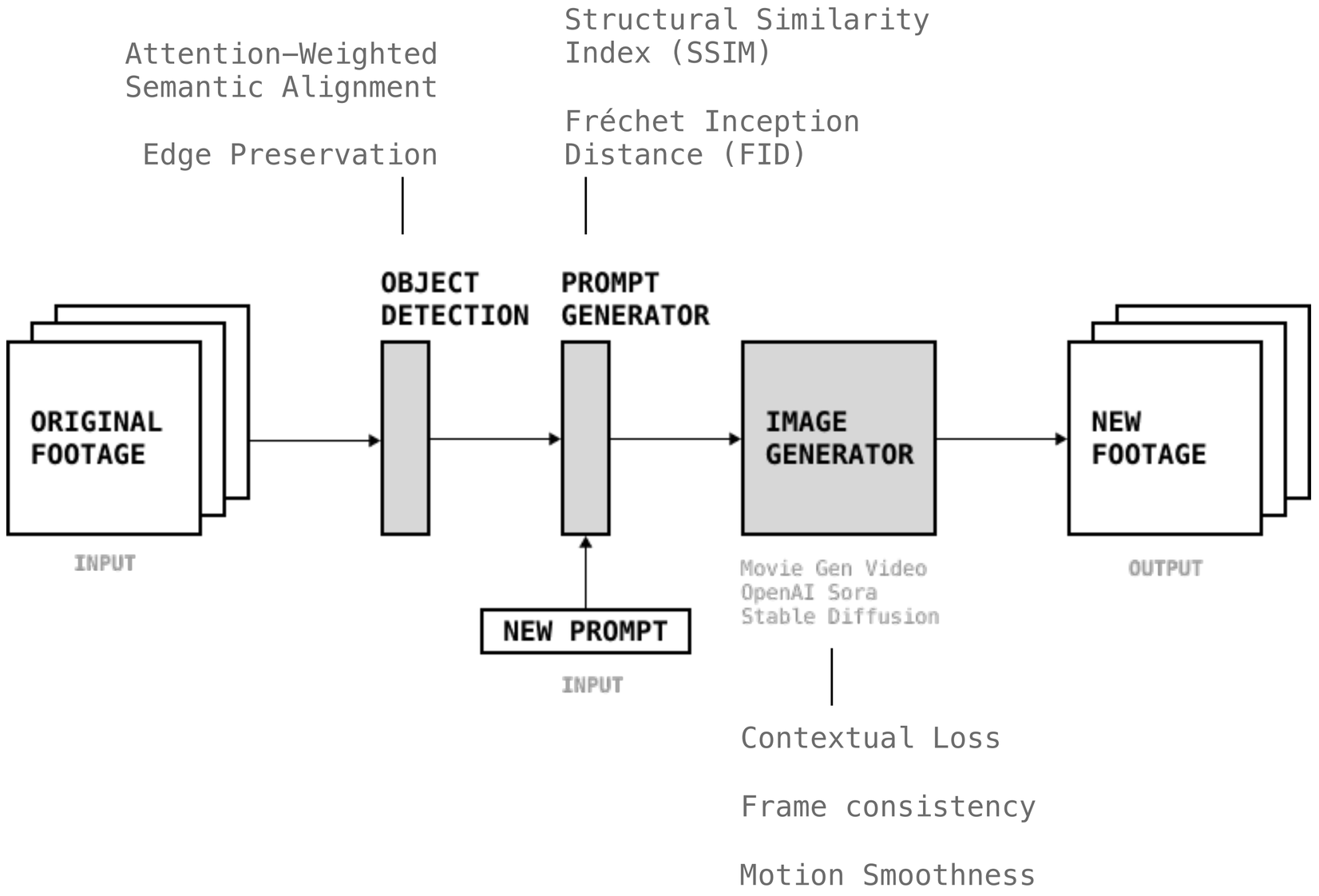
PIPELINE OVERVIEW



PIPELINE OVERVIEW



Metrics


Attention-Weighted Semantic Alignment
Edge Preservation
Structural Similarity Index (SSIM)
Fréchet Inception Distance (FID)
OBJECT DETECTION
PROMPT GENERATOR
ORIGINAL FOOTAGE
INPUT
NEW PROMPT
INPUT
IMAGE GENERATOR
Movie Gen Video
OpenAI Sora
Stable Diffusion
NEW FOOTAGE
OUTPUT
Contextual Loss
Frame consistency
Motion Smoothness

CHAPTER IV

TIMELINE

ACTIVITIES	1st Year						2nd Year					
	1	2	3	4	5	6	1	2	3	4	5	6
PREPARATION	●	●	●	●	●	●						
Mandatory credits in subjects.					●	●						
Bibliographic review	●	●										
EQM			●	●								
PHASE I: Static Images							●	●				
Object Detection Model							●					
Prompt Generation							●					
Image Generation							●	●				
Inpainting Model								●				
PHASE II: Animated Images								●	●			
Temporal Generation								●	●			
PHASE III: 3D Environments										●	●	●
Spatial Generation										●	●	●
CONCLUSION											●	●
Document and publish results											●	
Writing the dissertation document											●	●
Defense of Master's Thesis												●

THANK YOU.

f215775@dac.unicamp.br
linkedin.com/in/fesdias