Computer Vision and Generative Visual AI Workloads

Algorithms and Deployment Challenges

Activities Report
December 2025

Ruben Tous

Associate Professor. Computer Architecture Department

Universitat Politecnica de Catalunya (UPC)

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

UPC

CROMAI
Computational Resources Orchestration
and Management for AI

Introduction

- BIMAI (Barcelona Image and Artificial Intelligence Lab)
- AI, Computer Vision and image/video synthesis
  - Focus on visual arts
- Algorithms
  - Emphasis on non-mainstream creative flows (e.g. traditional 2D animation, educational materials, indie videogames, etc.)
- Computational performance
  - AI inference workflows and deployment on commodity hardware

#video-to-animation
#video-to-video
#rotoscoping

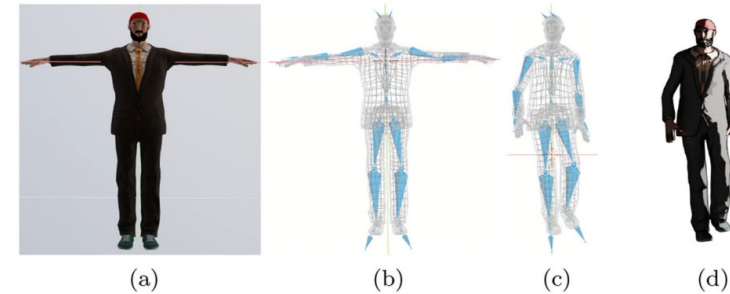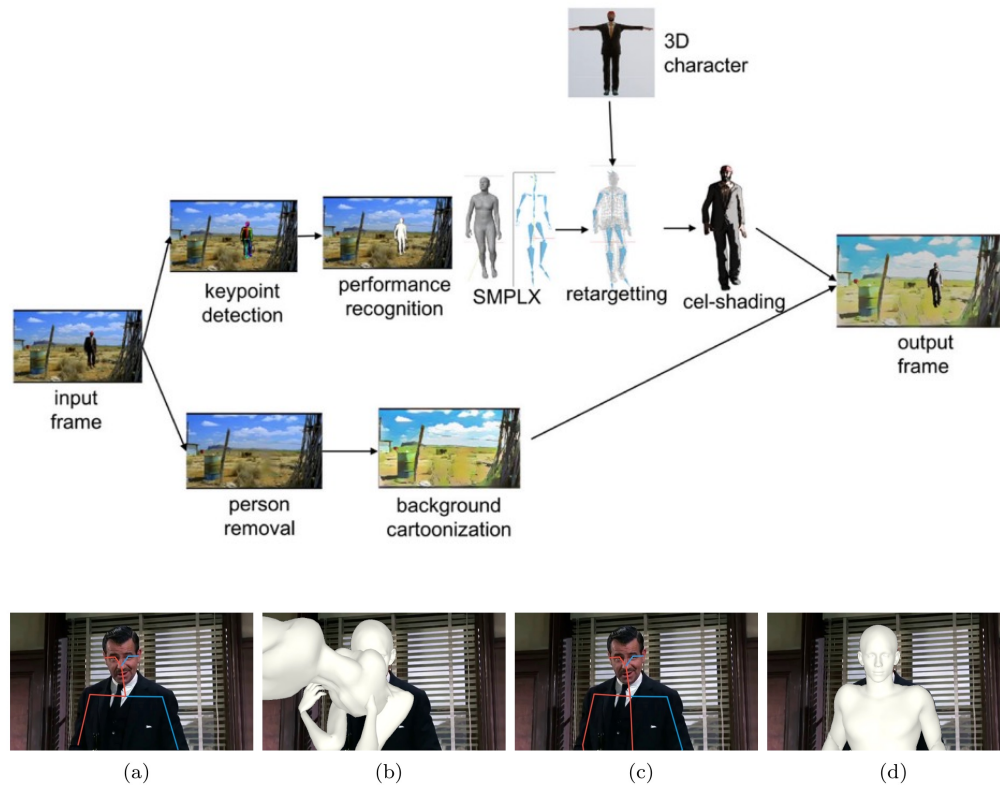# Rotoscoping through 3D Human Pose Estimation

Fig. 5 Pose retargeting. (a) 3D Mixamo character, (b) Rigged with the SMPL-X skeleton, (c) posed, (d) cel shaded

[1] Ruben Tous. Pictonaut: Movie Cartoonization Using 3D Human Pose Estimation and GANs. Multimedia Tools and Applications volume 82, pag. 21101–21115 (2023). **2023**

[2] Ruben Tous, Jordi Nin and Laura Igual. Human Pose Completion in Partial Body Camera Shots. Journal Of Experimental & Theoretical Artificial Intelligence. July **2023**.
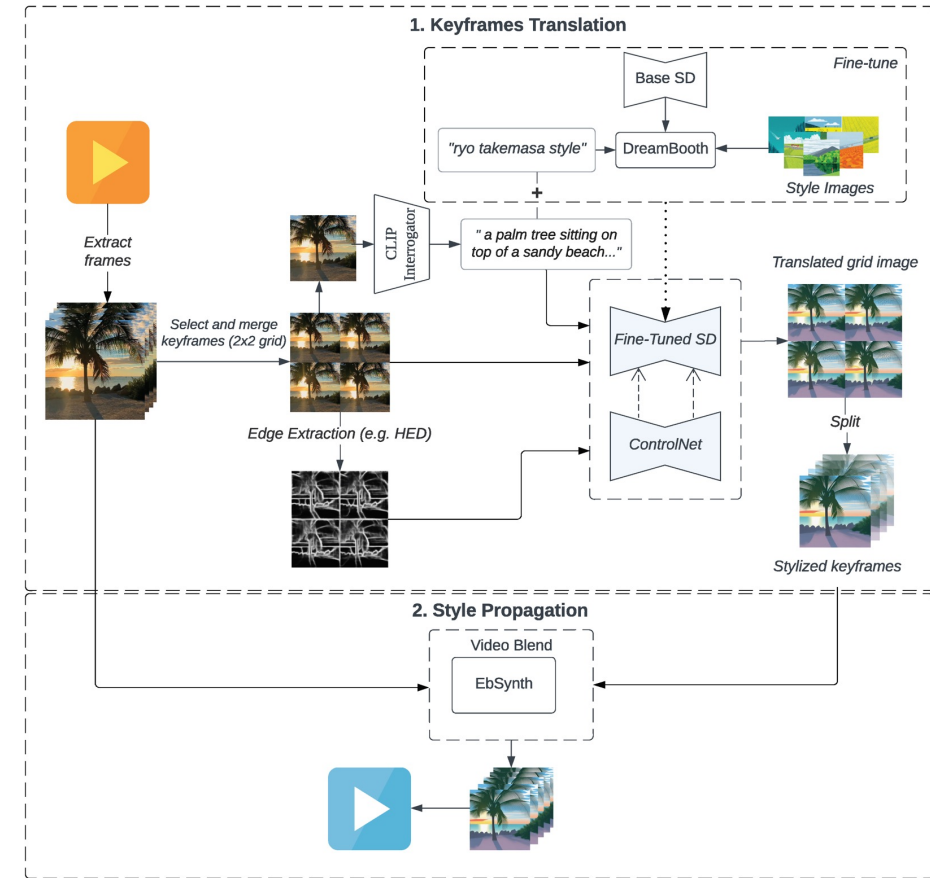
[3] Beatriz Gomes, Ruben Tous. A 3D Pipeline for 2D Pixel Art Animation. **2023**

# Temporally Coherent Video Cartoonization for Animation Scenery Generation

Sunset Over the Rice
Fields
Ryo Takemasa

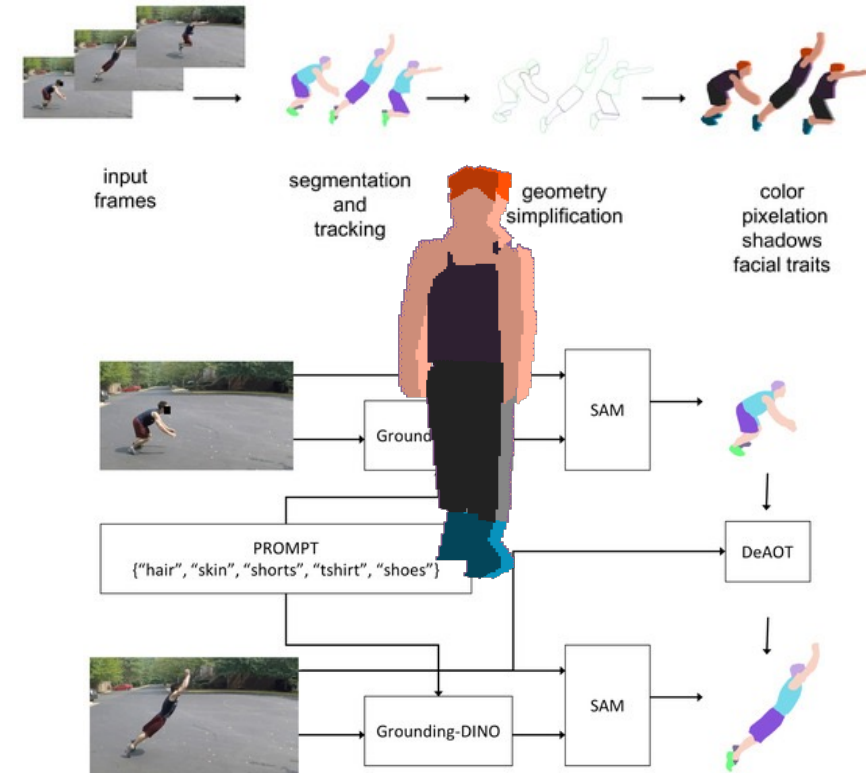Demos: https://github.com/gustavorayo/video-to-cartoon

[4] Gustavo Rayo and Ruben Tous. Temporally Coherent Video Cartoonization for Animation Scenery Generation. Electronics, 13(17), 3462. **2024**

Lester: Rotoscope Animation Through Video Object Segmentation and Tracking

Project page: https://github.com/rtous/lester
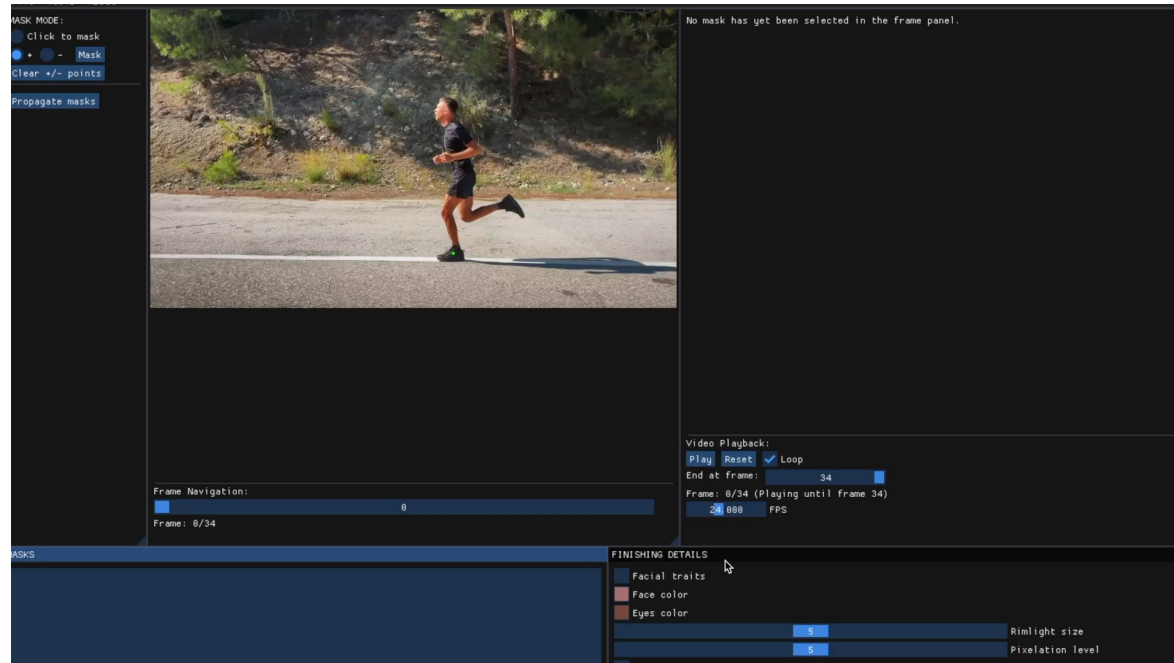
[5] Ruben Tous. Lester: Rotoscope Animation through Video Object Segmentation and Tracking. Algorithms 17, no. 8: 330. **2024**
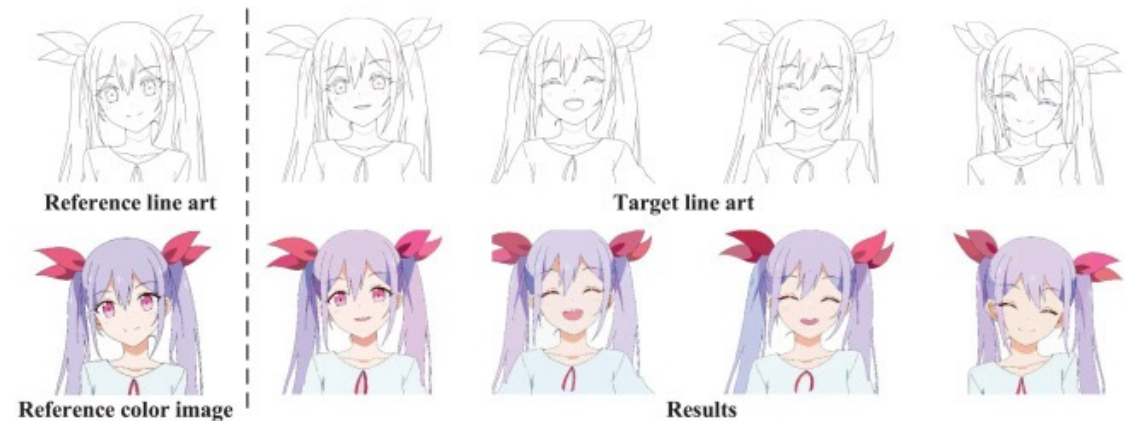
#line-art
#colorization
#interpolation
#superresolution
#2D

2D Animation In-betweening

- Reference-based animation line-art colirization

- Lack of groundtruth data
  - Typically projections of 3D clips

- Novel line-art extraction method combining:
  - Super-resolution
  - Video segmentation

Reference line art | Target line art

Reference color image | Results

[6] Judith Caldés, Ruben Tous, Beatriz Otero. Quality Assessment of State-of-the-Art 2D Animation Interpolation. Technical report. **2024**
[7] Maria Bertolin, Ruben Tous. Reference-based Animation Line-Art Colorization. Technical report. **2025**
[8] Rosa Ceano , Ruben Tous. Automatic Line Art Extraction from 2D Color Cartoons and Animations. Technical report. **2025**

#efficiency
#edge
#interpretability

Deploying Vision Foundation AI Models on the Edge. The SAM2 Experience

- Segmentation Anything Model 2 (SAM2)
  - Released by Meta on summer 2024
  - Promptable video segmentation
- SAM2 Migration to C++
- 6 ONNX models
- 3000 lines of pipeline Python code
- Performance evaluation
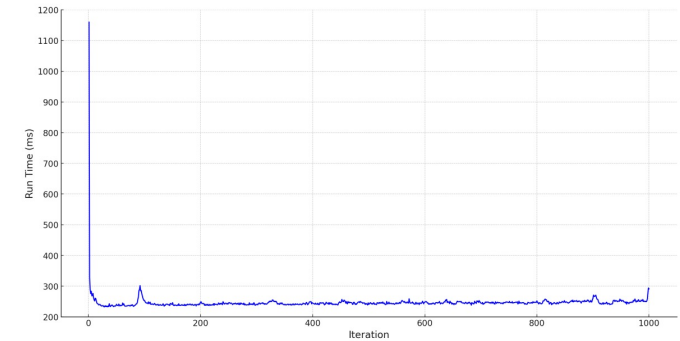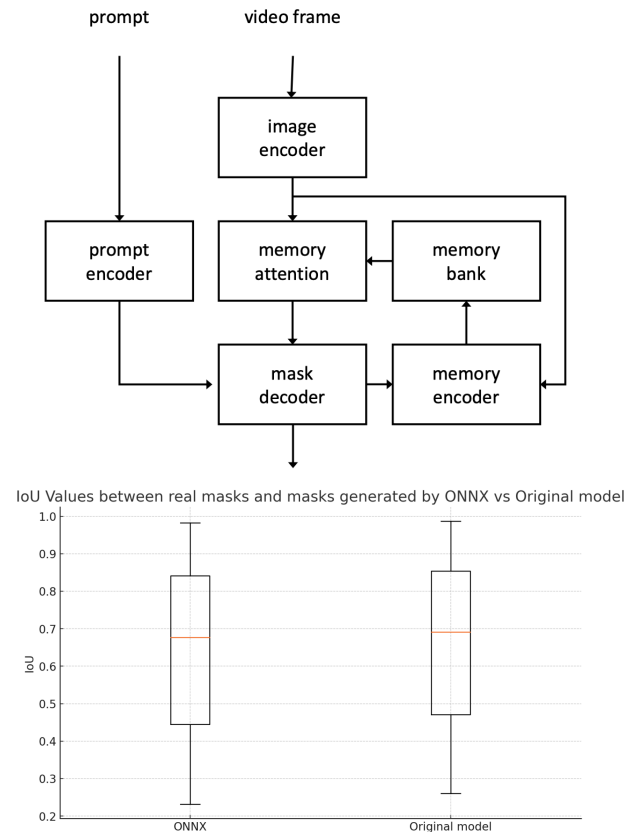- Comparison with cloud depl.
- Accuracy evaluation

Figure 3: Execution Time on a Local Machine

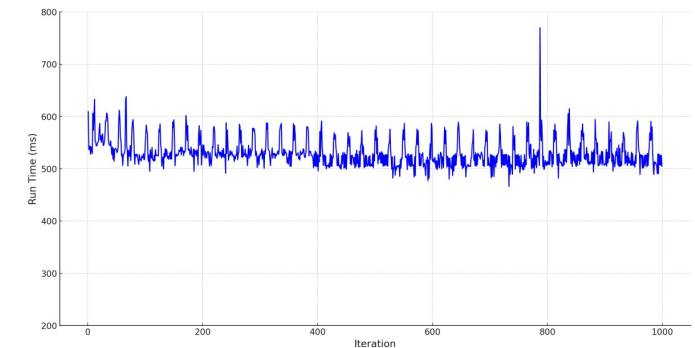IoU Values between real masks and masks generated by ONNX vs Original model

Figure 4: Execution Time in a Cloud Environment

[9] Zheshuo Lin, Ruben Tous and Beatriz Otero. Deploying Vision Foundation AI Models on the Edge. The SAM2 Experience. Proceedings of the 18th International Work-Conference on Artificial Neural Networks (IWANN 2025). A Coruna, Spain, June 16-18, **2025**.

Efficient Vision Transformer Inference on Edge Devices

- ViT + Quantization, knowledge distillation, token merging, and pruning

- Individual techniques and their combinations

- latency-accuracy and memory-accuracy trade-offs
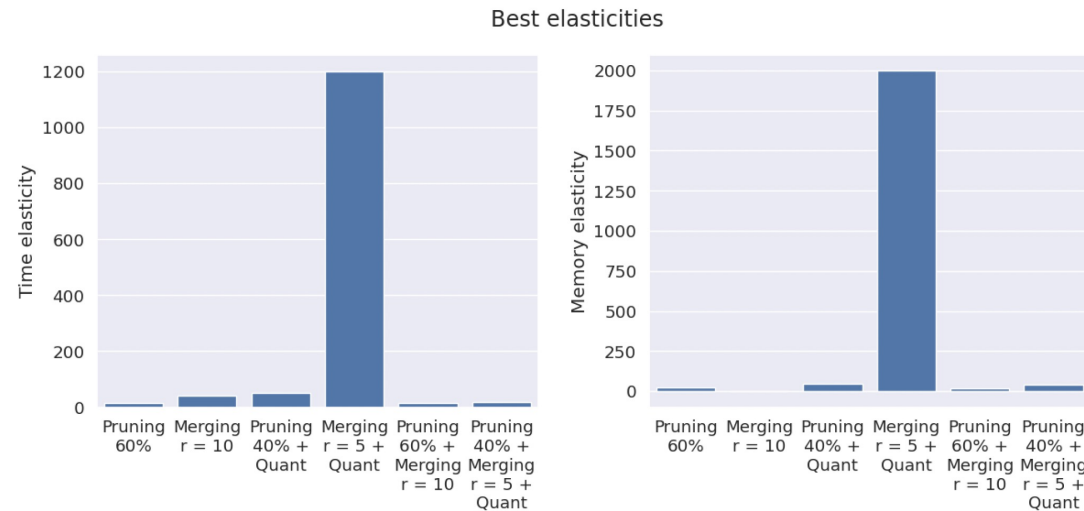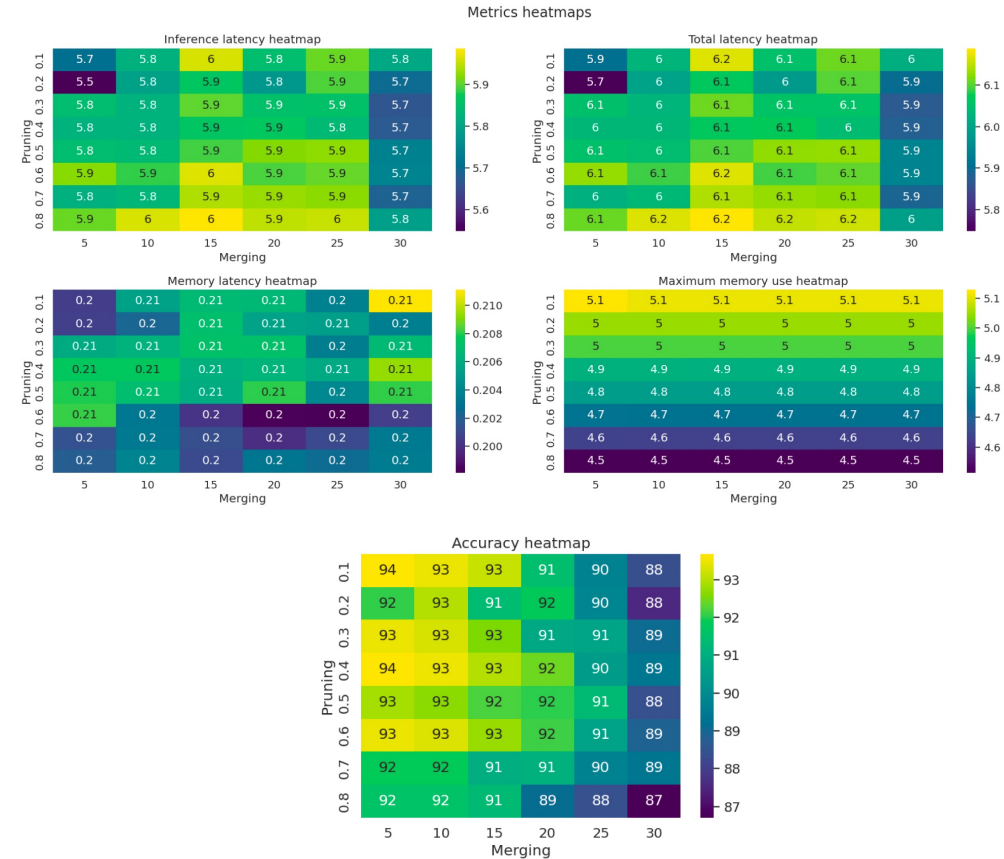
[10] Aleix Armero, Efficient Vision Transformer Inference on Edge Devices. Technical report. **2025**

Interpretability Techniques in BEV Autonomous Driving Models

- ViT, 6 cameras, time, 3D detection+segmentation

- Vanilla gradients and integrated gradients

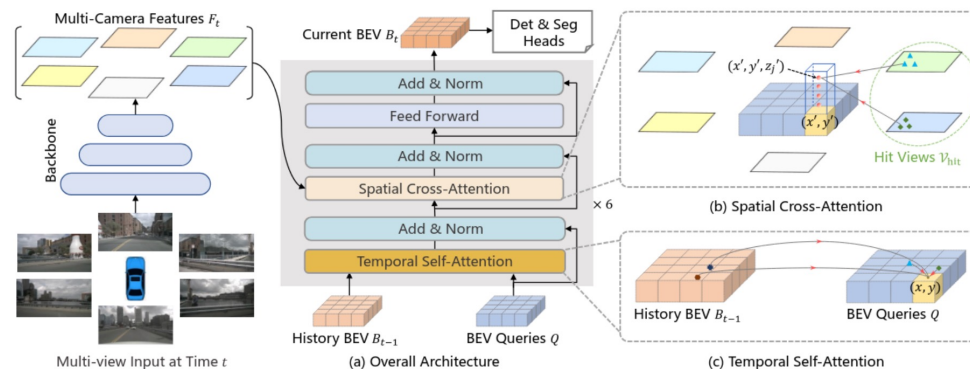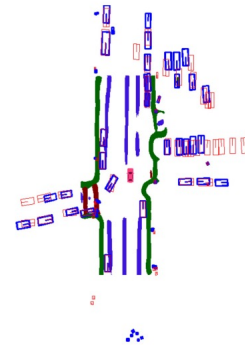- Differentiable NDS

- Perturbation tests
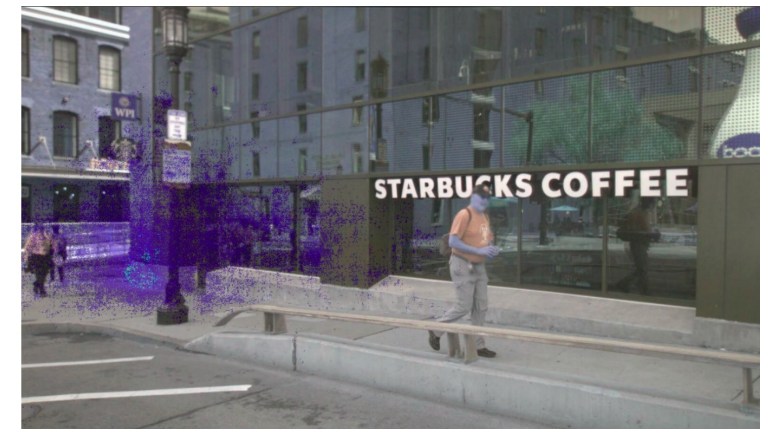

Figure 7: BEVFormer Architecture. [2]

Figure 16: VG Saliency for Bicycle Class, Front-Right Camera, frame 3. *Own generation.*

[11] Adria Ruiz, Ruben Tous and Josep Ramon Casas. Interpretability Techniques in BEV Autonomous Driving Models. **2025** (to be submitted)

BIMAI

Barcelona Image and Artificial Intelligence Lab

CONTRIBUTORS:

Ruben Tous   Camila Pérez
Beatriz Otero   Héctor Antona
Aleix Armero   Joffre Alcivar
Adrià Ruiz   Roger García-Alzórriz
Eduardo Cid   Silvia Arellano
Rosa Ceano   Carlos Valverde
Maria Bertolin   Beatriz Gomes
Judith Caldés   Roc Masana
Albert Ruiz   Sergi Espinosa
Gustavo Rayo   Sebastià Nicolau

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH
UPC

CROMAI
Computational Resources Orchestration
and Management for AI