

# Kinaema: A recurrent sequence model for memory and pose in motion

NAVER LABS Europe

https://europe.naverlabs.com/kinaema

bulent.sariyildiz@naverlabs.com

Mert Bülent Sarıyıldız Philippe Weinzaepfel Guillaume Bono Gianluca Monaci

# Introduction

Key aspect of spatially aware robots is the ability to situate themselves in previously seen spaces

Kinaema is a new memory model that can

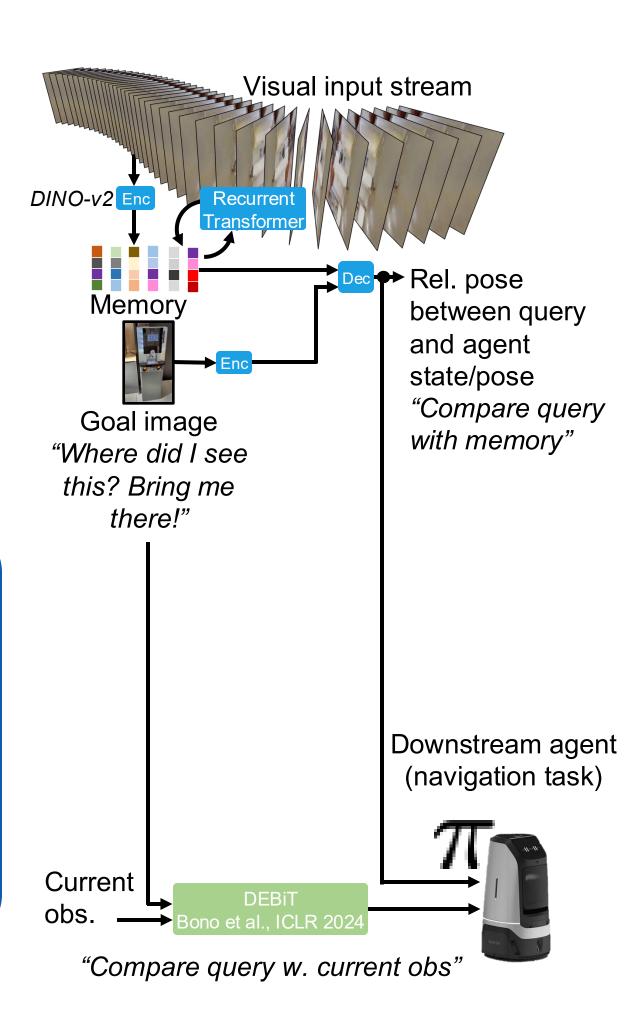
- → integrate a stream of visual observations while moving in a potentially large scene,
- → predict relative pose of query images with respect to any memory location.

## **Design objectives**

Goal 1: Fast, robotics friendly – O(1) updates!

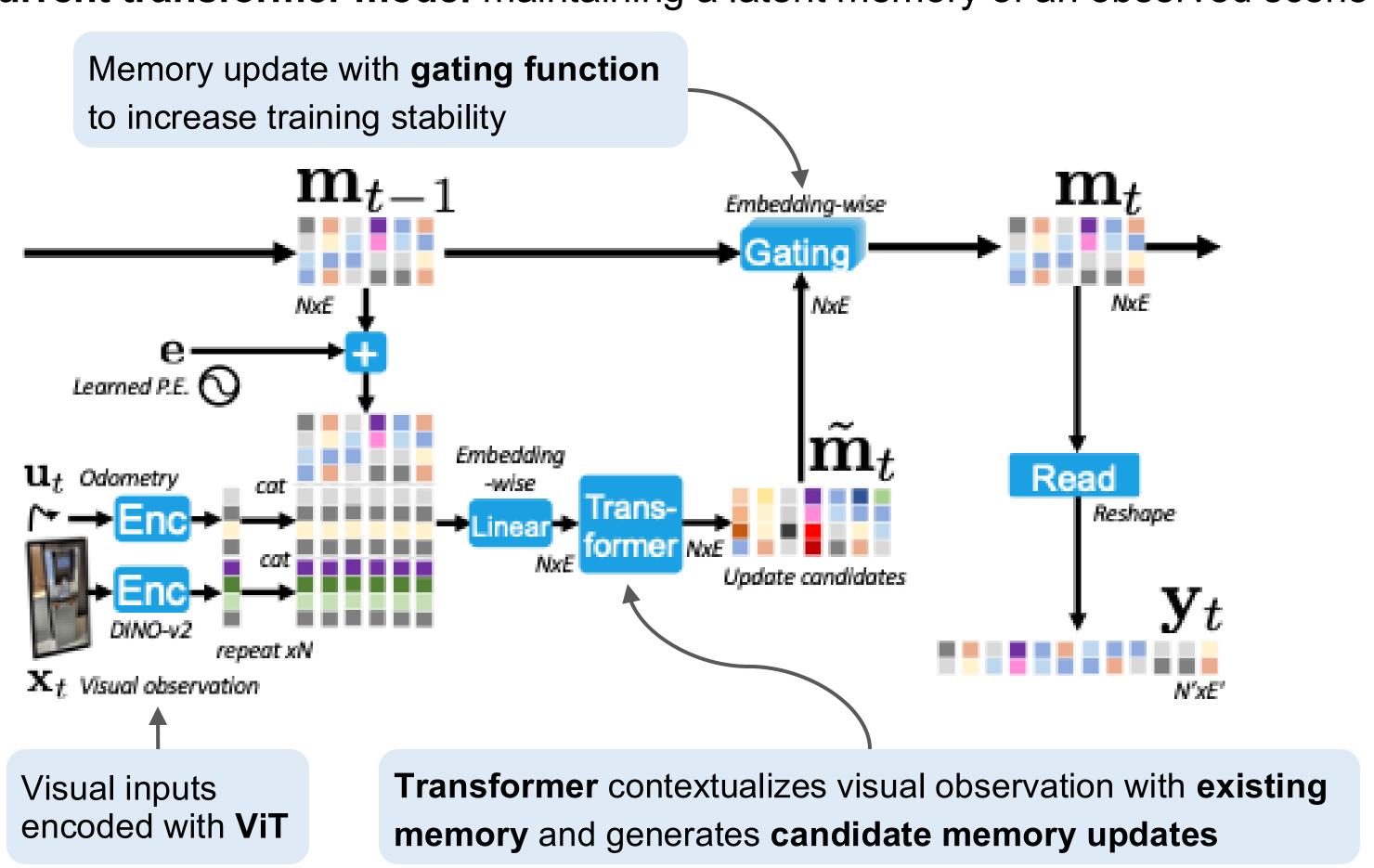
Goal 2: Scale memory – without affecting network capacity, in contrast to GRU, Mamba etc.

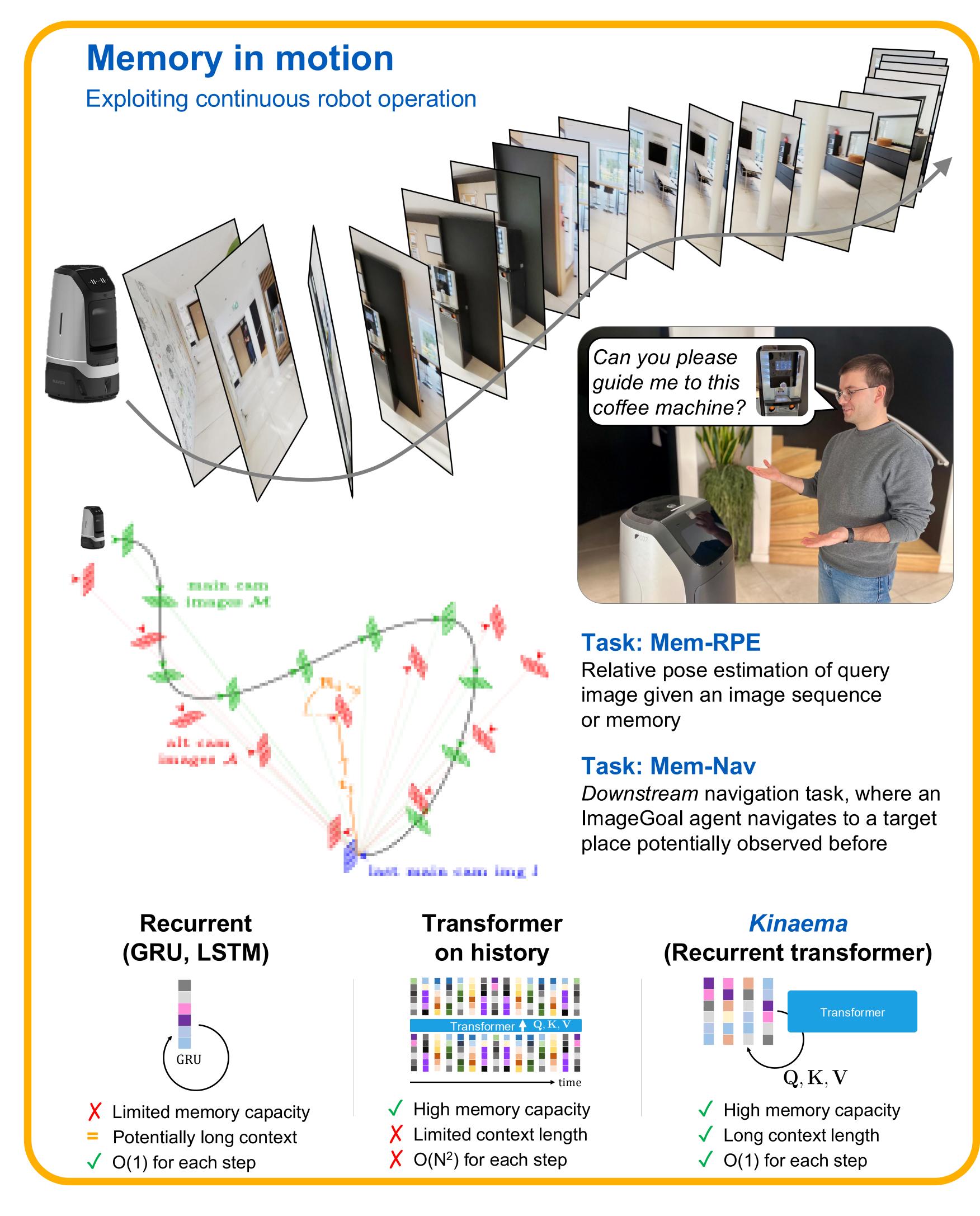
Goal 3: Stability – integrate gating functions into recurrent transformer updates



## Kinaema model

Recurrent transformer model maintaining a latent memory of an observed scene

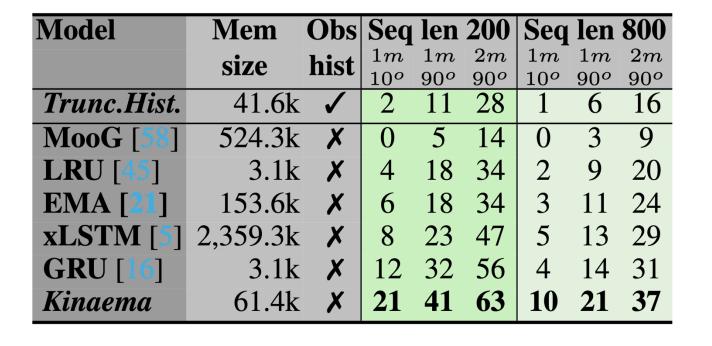


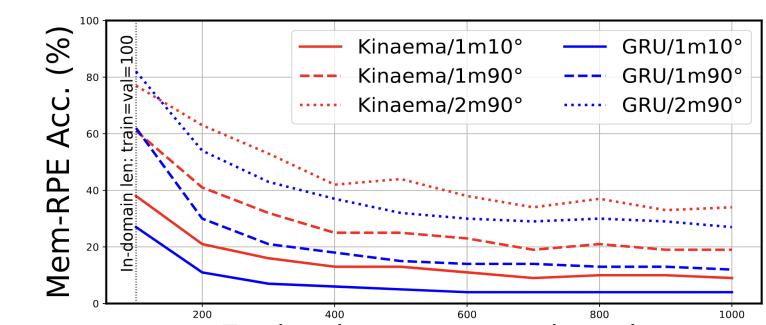


# **Experimental results**

All models trained with randomized sequence lengths between T=50 and T=100 Evaluated in significant out-of-distribution settings

### **Mem-RPE – Pose estimation accuracy**



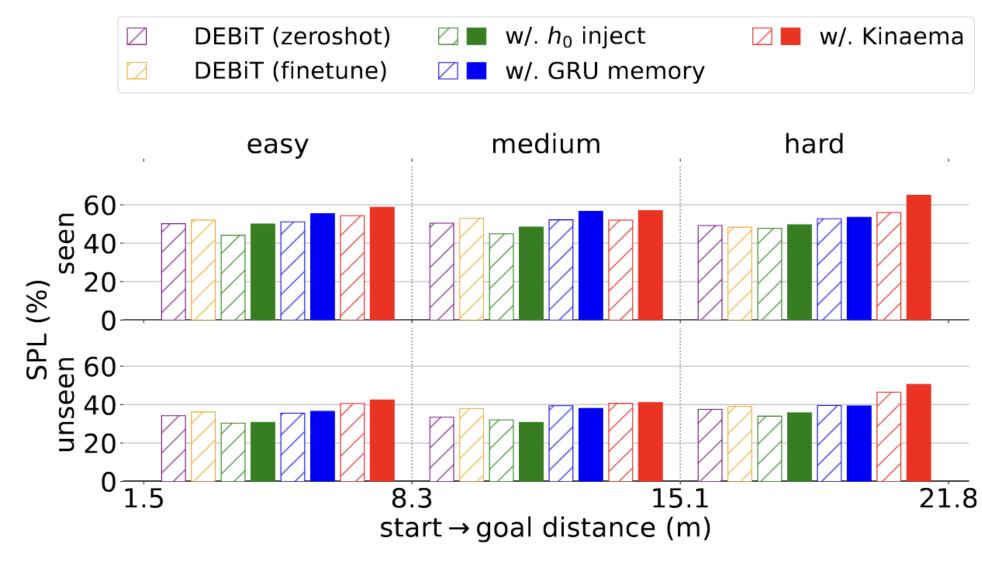


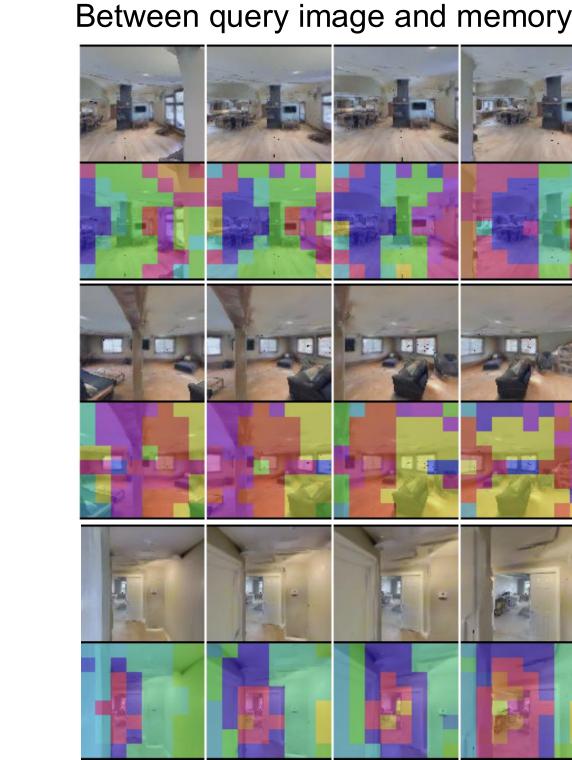
**Cross Attention** 

#### Mem-Nav – Navigation efficiency (SPL)

GRU and Kinaema are integrated into the baseline DEBiT agent, with (■) and without (□) priming sequence.

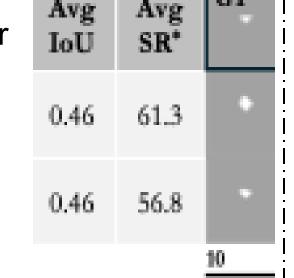
Kinaema is particularly useful to solve hard nav episodes, where start → goal distance > 15m

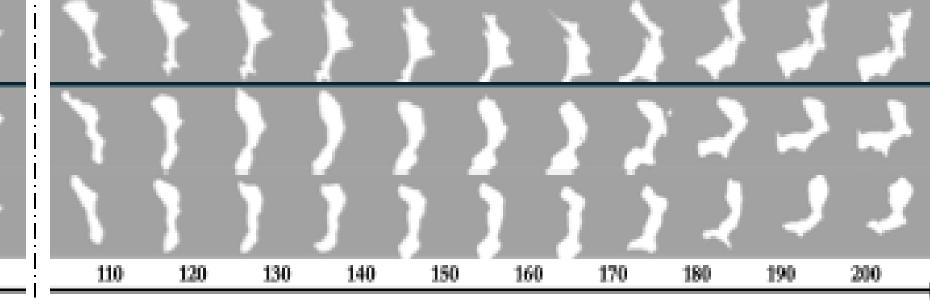




#### Occupancy probing Probing network can infer

scene structures from Mem-RPE pre-training





## **NAVER LABS** is hiring!

We have permanent and internship positions open. Consider applying if you are interested in Robotics, CV, ML and NLP.

Join us and work at the gateway to the French Alps!

