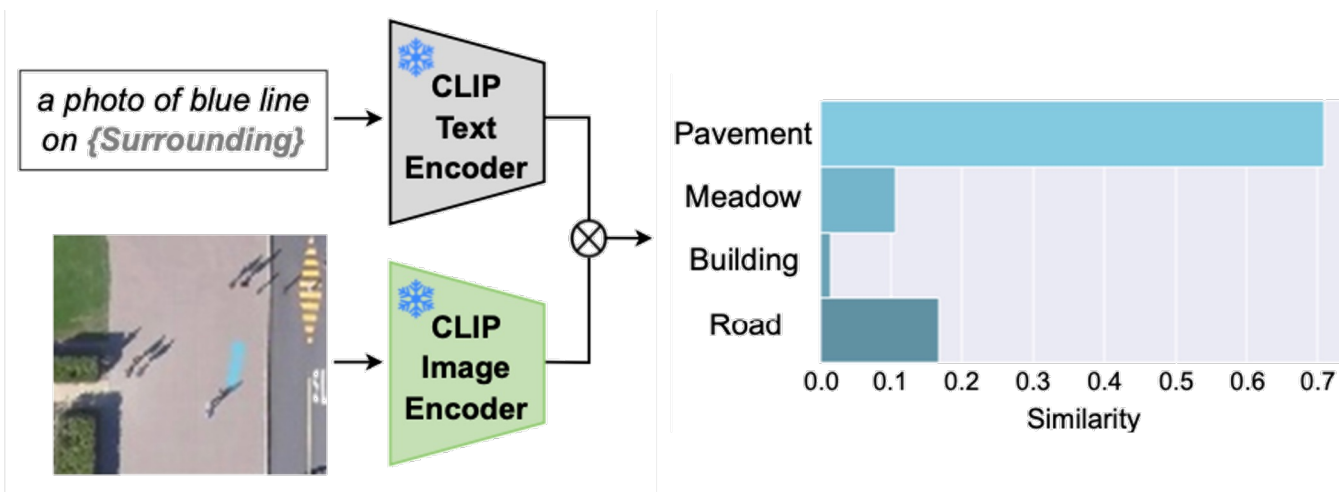


# TrajPrompt: Aligning Color Trajectory with Vision-Language Representations

Li-Wu Tsao, Hao-Tang Tsui, Yu-Rou Tuan, Pei-Chi Chen,

Kuan-Lin Wang, Jhih-Ciang Wu, Hong-Han Shuai, and Wen-Huang Cheng


National Yang Ming Chiao Tung University, National Taiwan University




# How do trajectory realize the map?

---

Past Trajectory + BEV Map → Future Trajectory

  
 $(x_1, y_1) \dots (x_i, y_i)$



  
 $(x_{i+1}, y_{i+1}) \dots (x_j, y_j)$

- Localization is the key!

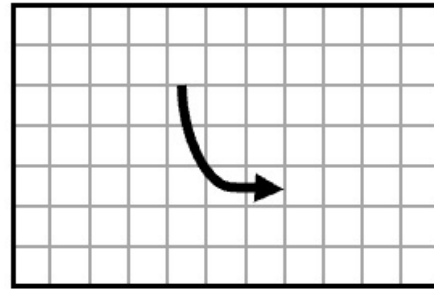
# Which representation better fits trajectory?

**Cartesian Coordinate**

**Occupancy Grid**

**Trajectory on Map**

$(x_1, y_1) \dots (x_i, y_i)$



- While considering the embedding cost:

**Cartesian Coordinate** < **Occupancy Grid** ≤ **Trajectory on Map**  
 (Less) (More)

# Our Goal

- We provide brand new perspectives for aligning vision and trajectory representation using [color trajectory prompts](#) and the properties below:

**X Human Annotation (No Pixel-wise Label)**

**V Text Guidance (Token-wise Dictionary)**

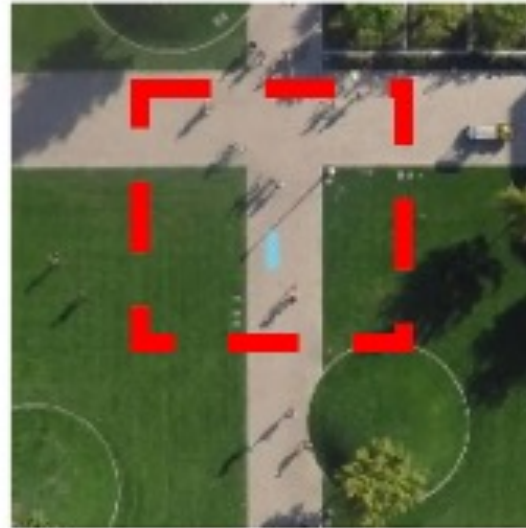
**V Lightweight Structure (Learnable Prompt)**

# How important are color trajectory prompts?



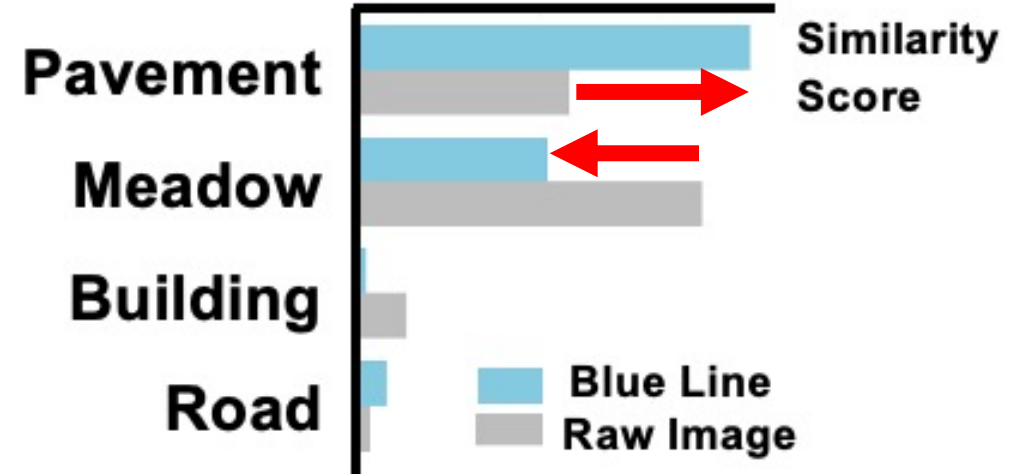
**Raw Image**

*A photo of ...*



**Blue Line**

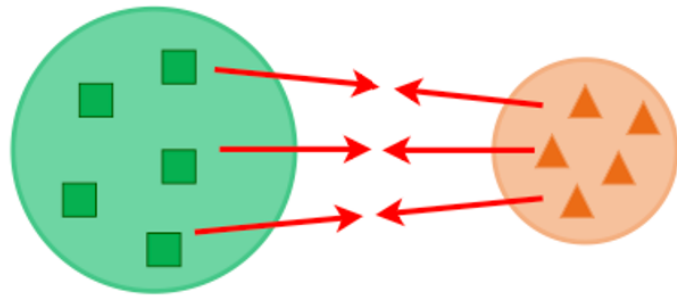
*A photo of blue line on ...*



*Capture the local surroundings effectively.*

# How do we combine BEV scene and trajectory?

## Contrastive Learning



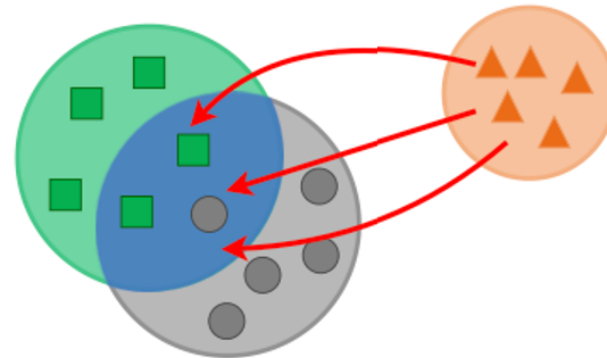
- Image Space
- Trajectory Space
- Text Space

ADE (↓) / FDE (↓): 7.90 / 15.37

↓ over 40%



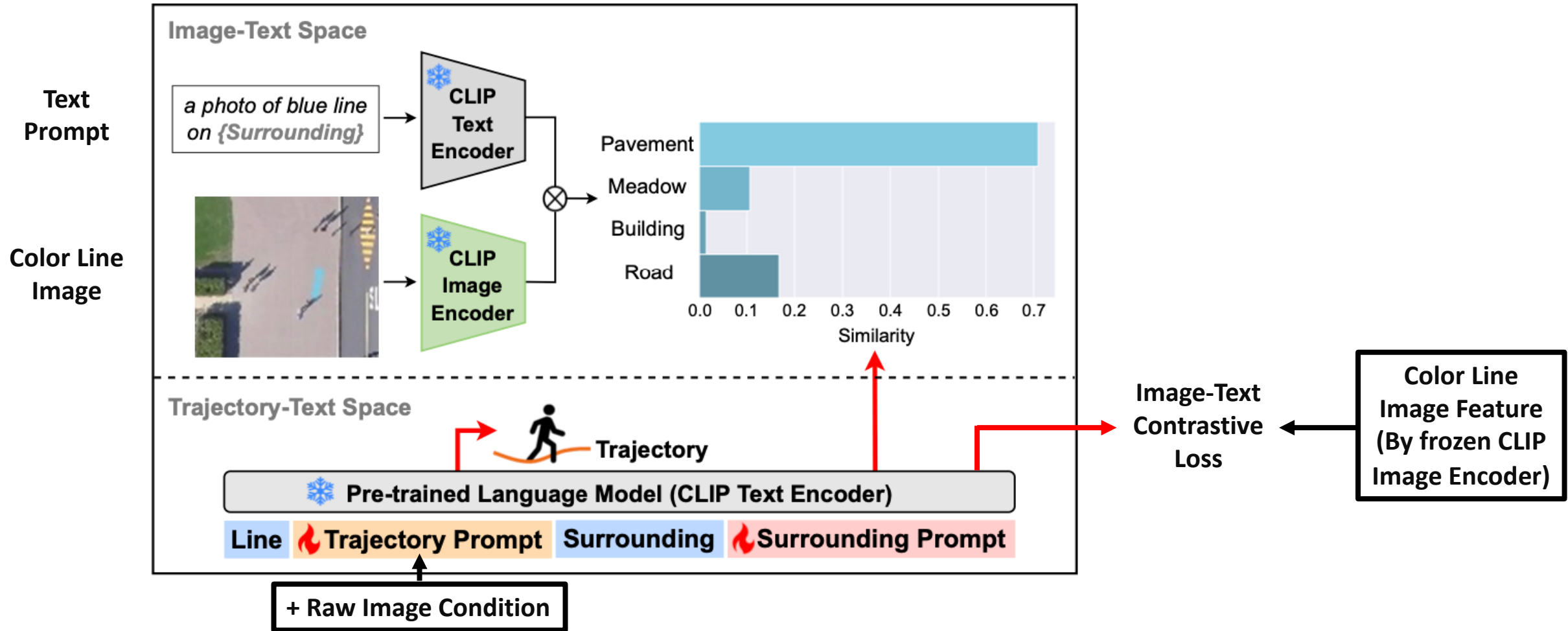
## Our Contrastive Learning



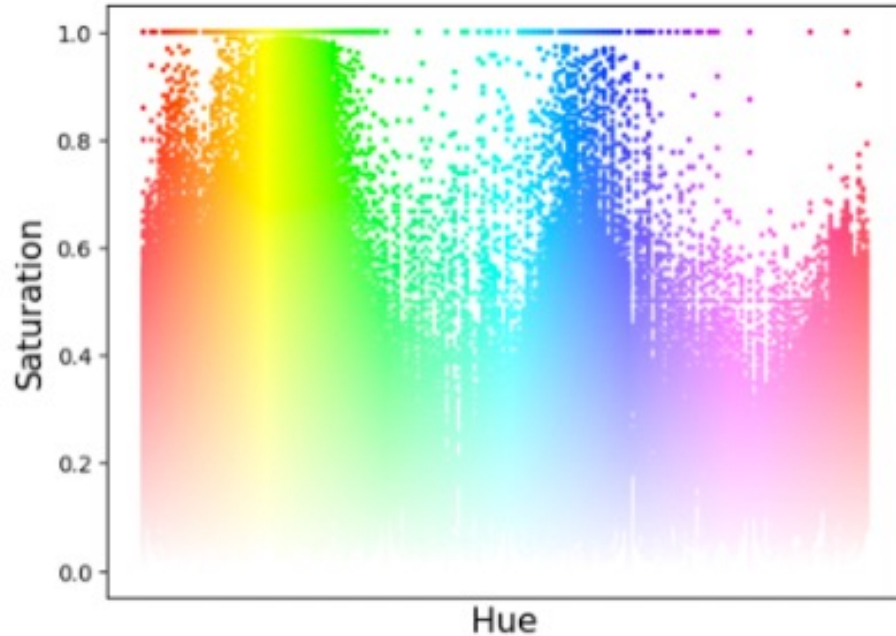
ADE (↓) / FDE (↓): 4.69 / 8.29

*Maintain the image semantics with proper text guidance.*

# How to ensure the learning of correct localization?



# Meaning of color in frozen CLIP



**Color Distribution on SDD dataset**



**Comparing Different Colors**

*Less common colors in the dataset may attract more attention.*



# THANK YOU FOR LISTENING

---

## Project Page

