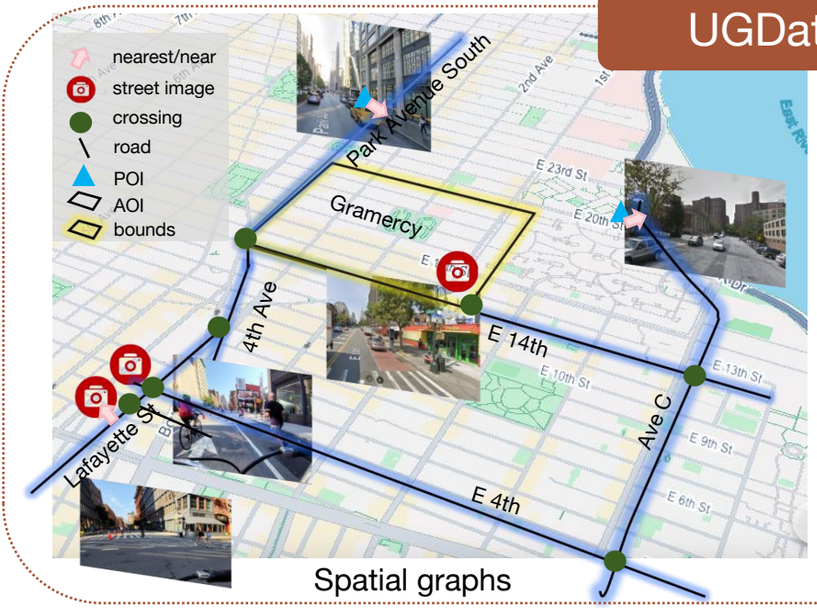


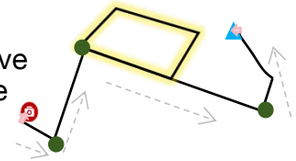
Spatial graphs

UGData

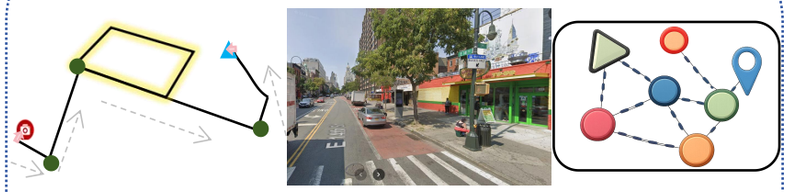
Stage 1 data

"<image>The image captures a sunlit urban street scene in Kips Bay neighborhood of Manhattan, near the intersection of E 25 St and Ave C..."

"<image>SRP: (image, nearest, E 4th)...-> (160m, 135°(SE)) -> ... (Asser Levy Playground, near, Ave C) (29m, 270°(W))...Based on the spatial resoning path, you can reach Asser Levy Playground..."



Stage 2 data



"**<graph>**<image>Based on the spatial context represented in the spatial reasoning path, ...you can reach Asser Levy Playground..."

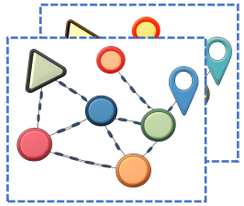
"**<graph>**<image>This image aligns closely with the spatial context of Kips Bay neighborhood of Manhattan, near Asser Levy Recreation Center ..."

"<image>The image captures

"<image>SRP: image,nearest, E 4th) -> (160m, 135°(SE)) ..."

Text Encoder *LoRa*

Image Encoder



NodeAttri Encoder *LoRa* + Position Encoder

Edge Encoder

"<graph><image>This

"**<graph>**<image>Given the spaital graph and the SRP, you can reach...

Text Encoder *LoRa*

UGE

Stage 1

Z_i

InfoNCE

Z_t

Z_g

Z_i

InfoNCE

Z_t

Stage 2

UGBench

Geolocation Ranking

"**<image>**Where is this street view image taken? "

Image Retrieval

"**<graph>**Refer to the graph, find the image which has a path that reaches Greenwood Ave"

"**<graph>**Using the spatial graph provided, find the image whose spatial context matches: This image is situated within a dense urban network centered around Queens Boulevard"

Urban Perception

"**<graph><image>**What is the perception of depressing for this urban location?"

Spatial Grounding

"**<graph><image>**Refer to graph, which location/street is nearest to the current image location?"

"**<graph><image>**Refer to graph, How far is the nearest church from the closest commercial?"

"**<graph><image>** if you walk approximately 200 meters south from here, what landmark will you encounter?"