



I am in front of a restaurant named 'Burger Mania', which features a modern facade with a neon sign. The establishment offers a variety of menu items, including burgers, sandwiches, pancakes, craft beer, coffee, tea, and cake, as indicated by the signage above the entrance. There is a welcoming atmosphere with greenery at the base of the window...

**Retrieval Reason :** The model's text attention focuses on the store sign "Burger Mania" and the food offerings. In the OSM data, the model's attention is primarily on restaurant information, with the retrieved OSM data containing a matching "Burger Mania" location. Therefore, the retrieval reason is identifying the same "Burger Mania" establishment.



I am at an intersection with a clear zebra crossing sign and a clear white gridline on the ground. In front of me is a multi-storey building with a large billboard prominently displayed on it, showing an advertisement. Traffic lights are present to control vehicle flow. To my left, there are several fast food restaurants including 'Black Burger' and a convenience store. The area is bustling with activity, typical of a city center...

**Retrieval Reason :** The model's text attention focuses on the zebra crossing and the mesh-like white lines on the ground. In the satellite data, the model's attention is also directed towards these markings, with the retrieved satellite image containing intersections and mesh-like white lines. Thus, the retrieval reason is recognizing the same intersection and ground mesh patterns.

1<sup>st</sup>



2<sup>nd</sup>



3<sup>rd</sup>

