Spaces as Graphs

Seyran Khademi

About Me

- **PhD @ TUDelft (2011-2015)** Statistical Signal Processing
- **Postdoc @ TUDelft (2015-2017)** Speech Processing Lab
- **Postdoc @ TUDelft & VU Amsterdam (2017-2021)** Computer Vision Lab
- **Research in Residence @ Royal Library of the Netherlands (2021)**
- **Assistant Professor @ TUDelft (2021-Now)** Co-director of AiDAPT Lab

Outline

1. **Spatial Data Representation — the world encoded as visual, geometric, and topological data**
   How spaces transform from pixels into structures we can compute on.

2. **Visual Place Recognition — challenges and why graph reasoning helps**
   Appearance/viewpoint change and aliasing require moving beyond raw image features.

3. **Scene Graphs & Space Understanding — relational representations of 3D environments**
   Using object–object and spatial relations to understand scenes more robustly.

4. **Floorplan Analysis & Synthesis**
   Treating topology as the fixed structure and geometry as the variable to be generated and matching, and retrieving spatial data through graph embeddings and structural similarity.

1. Spatial Data Representation

Spatial Data Representation

- Our world is spatial; to analyze it, we need effective data representations of space in various scales.

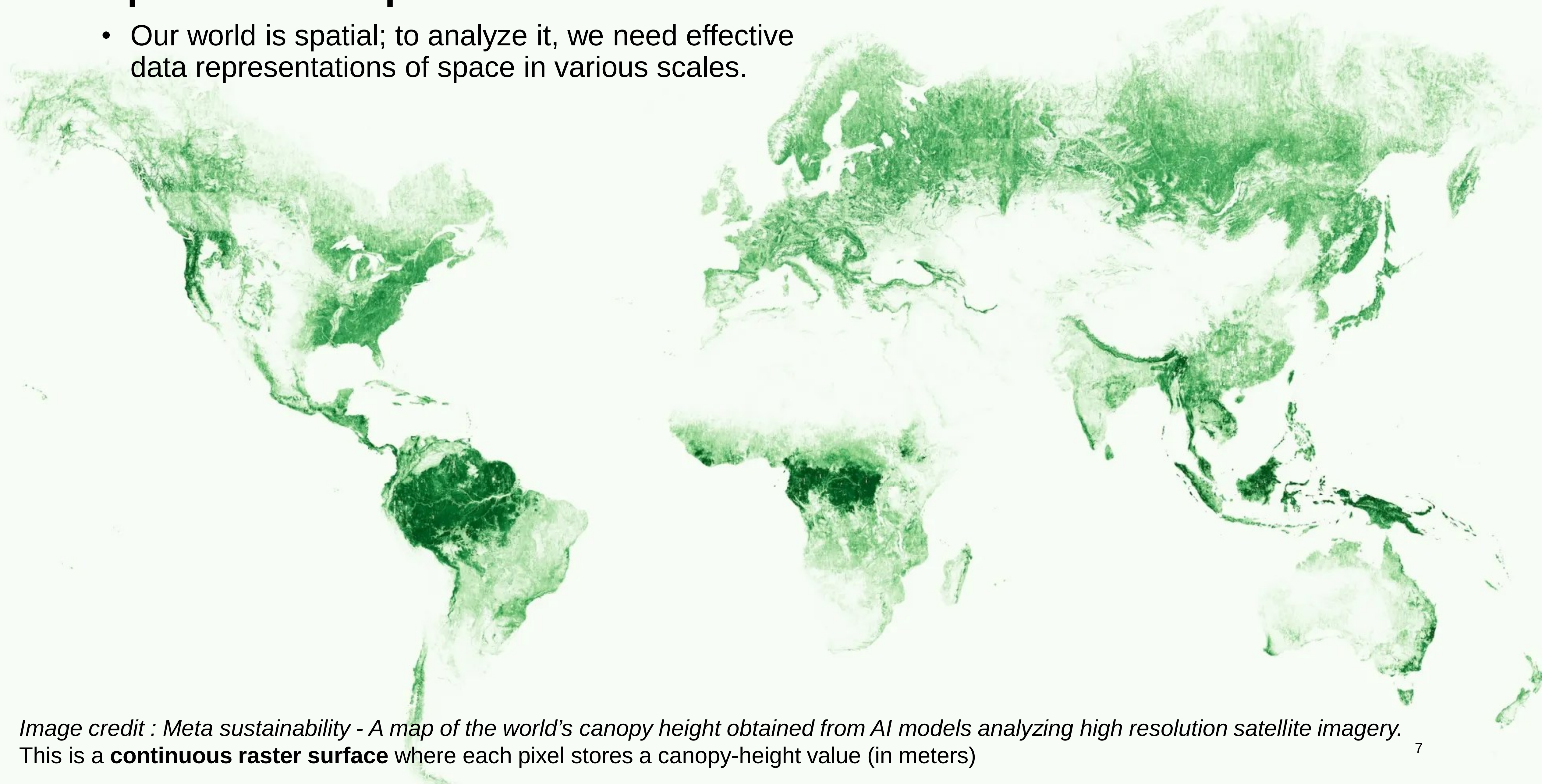
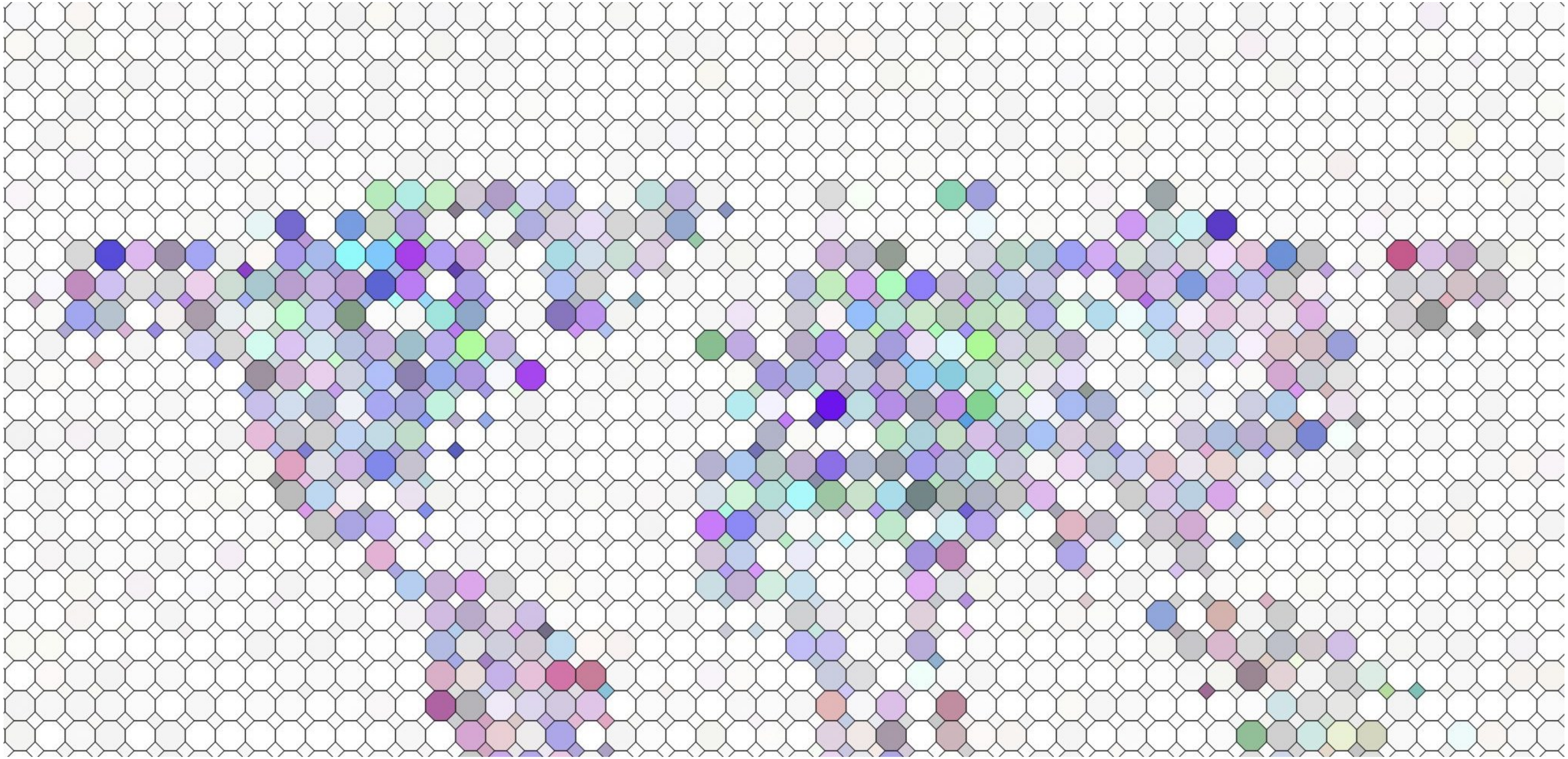


Image credit : Meta sustainability - A map of the world’s canopy height obtained from AI models analyzing high resolution satellite imagery.
This is a **continuous raster surface** where each pixel stores a canopy-height value (in meters)

Maps and Layouts



Urban Network Map

•**Urban maps encode topological relationships** between spatial entities—such as roads, buildings, amenities, and bridges—using geo-referenced coordinates.

•**OpenStreetMap (OSM)** provides a global, crowdsourced geospatial dataset that can be represented as a spatial graph, where nodes and edges correspond to real-world features enriched with metadata (e.g., road type, speed limits, access rules).

•**OSMnx**, a Python library, automatically downloads and converts OSM data into structured graph models, enabling routing, network analysis, topological simplification, and high-quality visualizations at urban to regional scales.

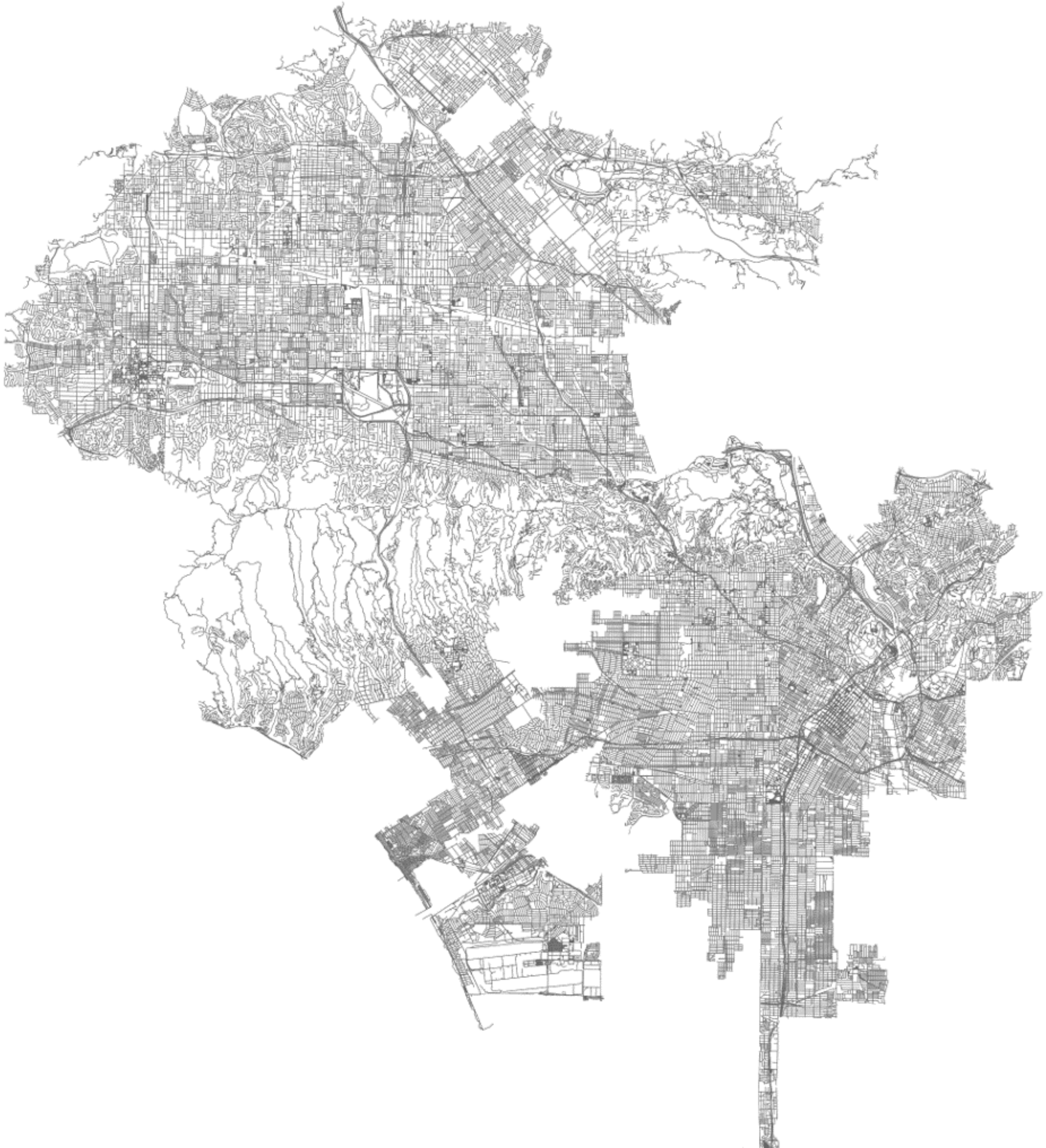


Image Credit: © OSMnx (data from OpenStreetMap) — image from “Modeling and Analyzing Urban Networks and Amenities with OSMnx”, Geographical Analysis

Floorplan Representation

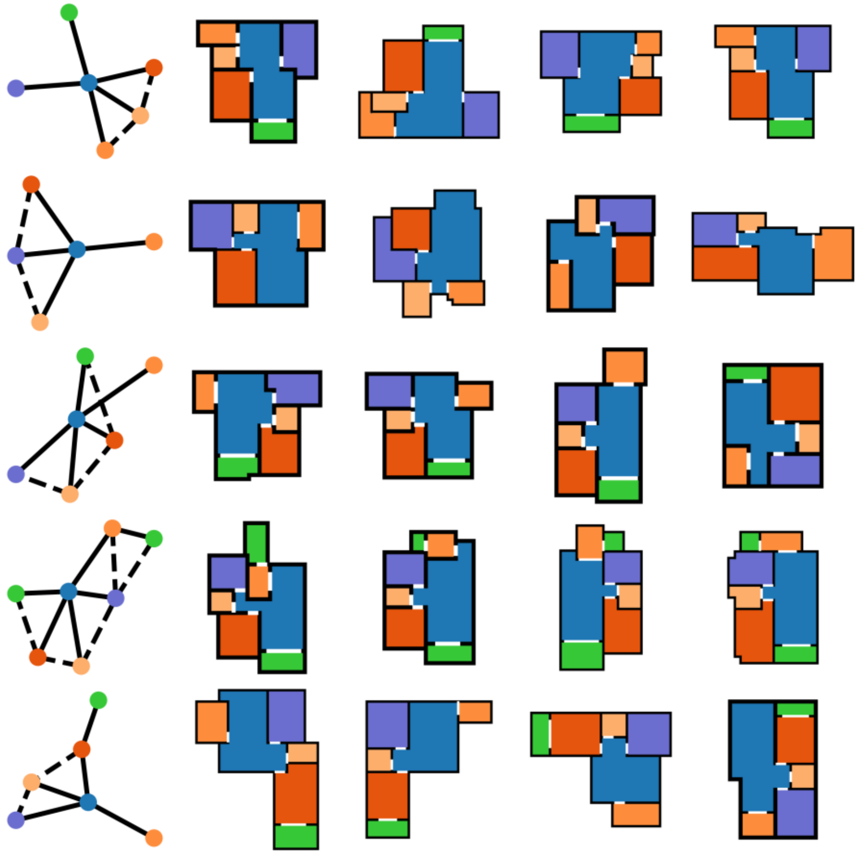
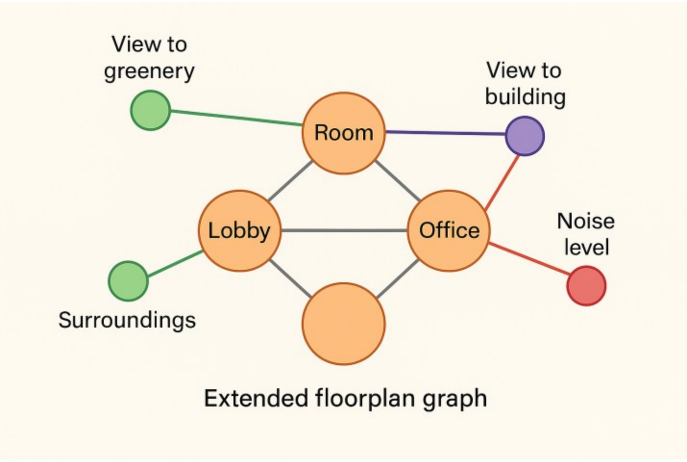


Image credit : from ICCV paper "SSIG: A Visually-Guided Graph Edit Distance for Floor Plan Similarity"

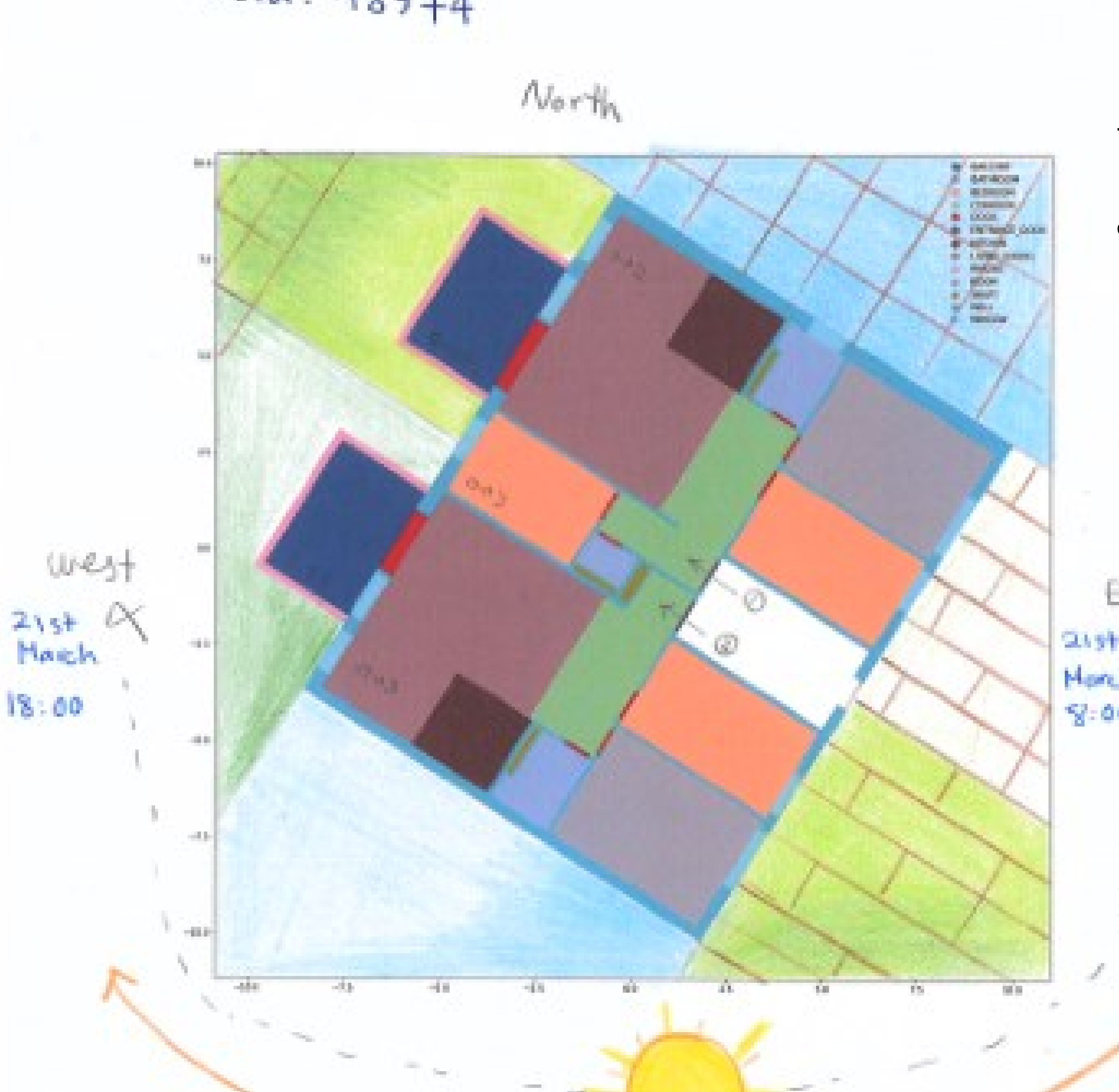
- Floor plans are simple but powerful representations of space, conveying a richness of information about the **compositional structure of buildings.**
- Floor plans are inherently multi-modal—pictorial, geometric, and **topological**—making them suitable for diverse data models and machine-learning frameworks for digital representation and understanding.
- This figure shows graph to image examples. Each row represents an access graph (column 1) and several randomly selected corresponding floor plans semantic images (columns 2 - 5).

Environmental Features

- **Include environmental facto**
  Add elements like *greenery vie*
  and *noise sources* as heteroge
- **Model their influence throug**
  Connect these nodes to indoo
  that encode their impact on en
- **Extend the floorplan graph t building's surroundings**
  Integrate immediate outdoor c
  more complete spatial–enviro

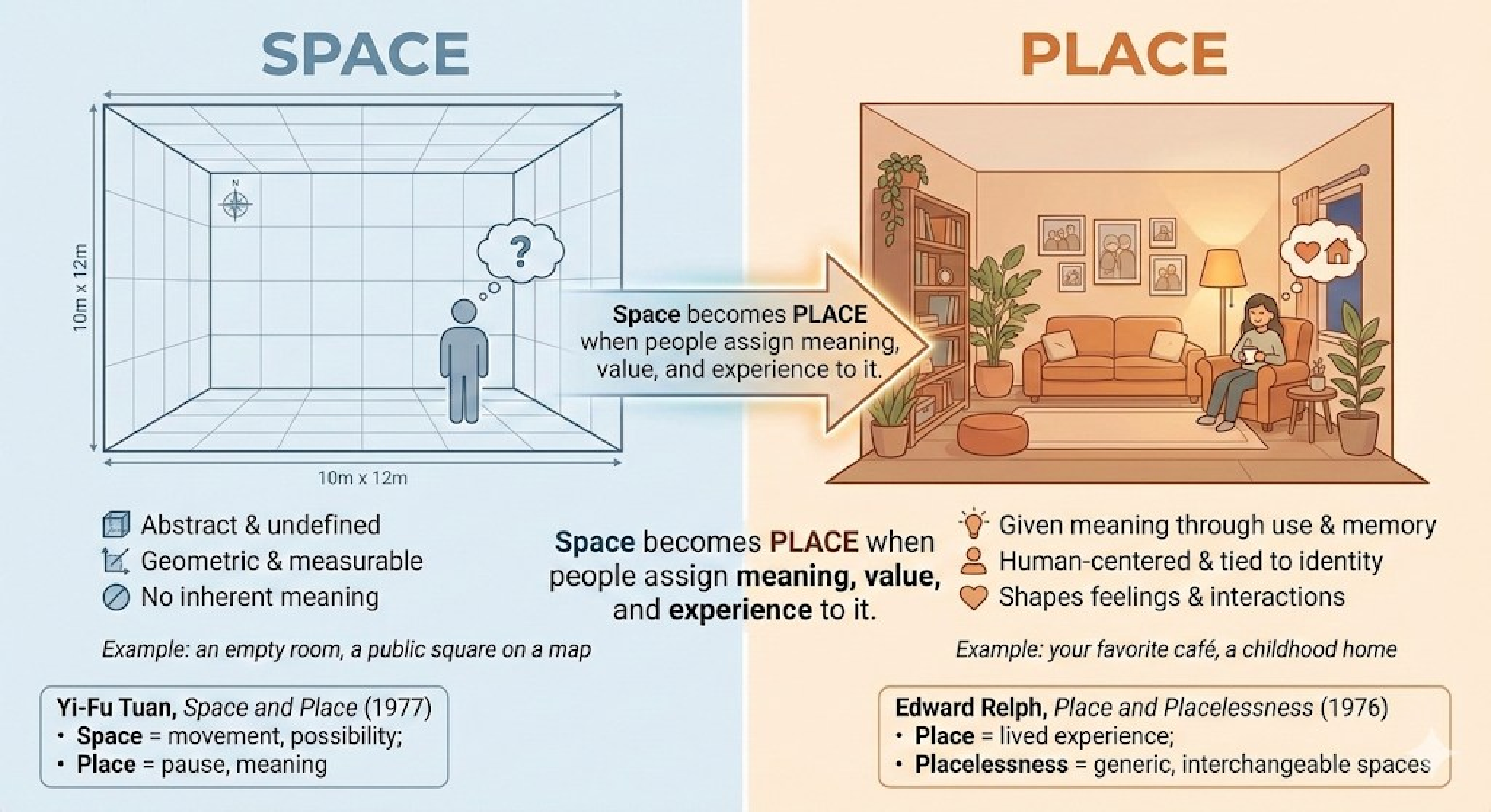


Seyran & Casper



Mostafavi, F. & Khademi, S. (2023). *Micro-Climate Building Context Visualization: A Pipeline for Generating Buildings' Environmental Context Maps Using Numerical Simulation Data.*

From Space to Place



2. Visual Place Recognition

Places in Pixel Space

Visual Place Recognition (Geo-Guesser)

Real-World Applications of Visual Place Recognition (VPR)

Long-Term Mapping : City-scale updates across seasons and years (e.g., Street View)

Indoor Localization: Navigation in airports, malls, museums & assistance for visually impaired users

Image/Video Geo-Localization: Automatically determining where a photo/video was taken that is used in photo organization and social media platforms

Agriculture & Mining Robotics: Localization in fields, orchards, underground tunnels

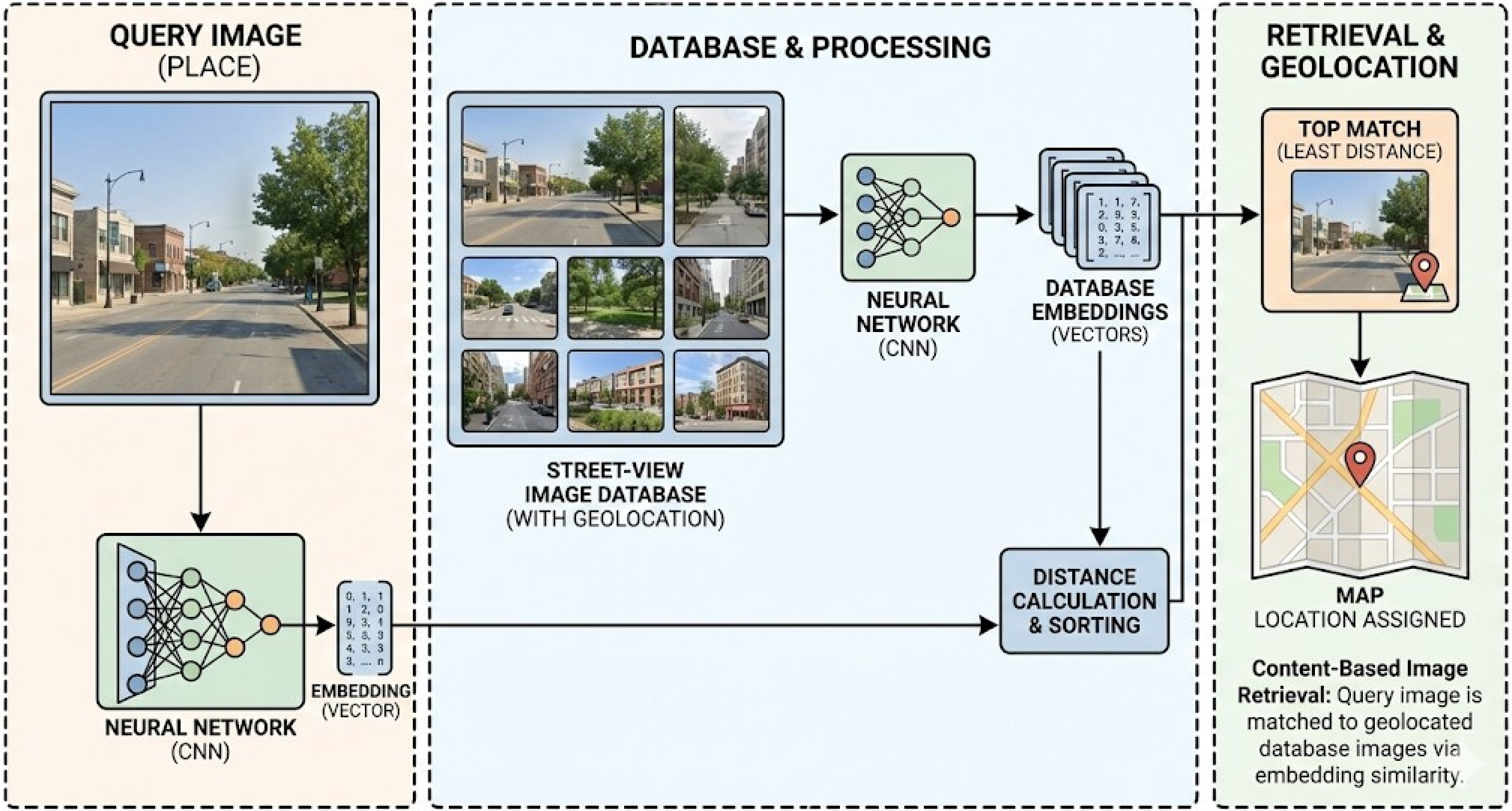
Autonomous Trains & Trams: Station recognition and navigation where scenery are more predictable

Security & Forensics: Matching scenes across cameras or videos, location identification under appearance change

Visual Place Recognition (VPR)

Yildiz, B., Khademi, S., Siebes, R. M., & van Gemert, J. (2022). *AmsterTime: A Visual Place Recognition Benchmark Dataset for Severe Domain Shift.* arXiv:2203.16291.

VPR is Content Based Image Retrieval Problem



Challenges of Image-Based VPR

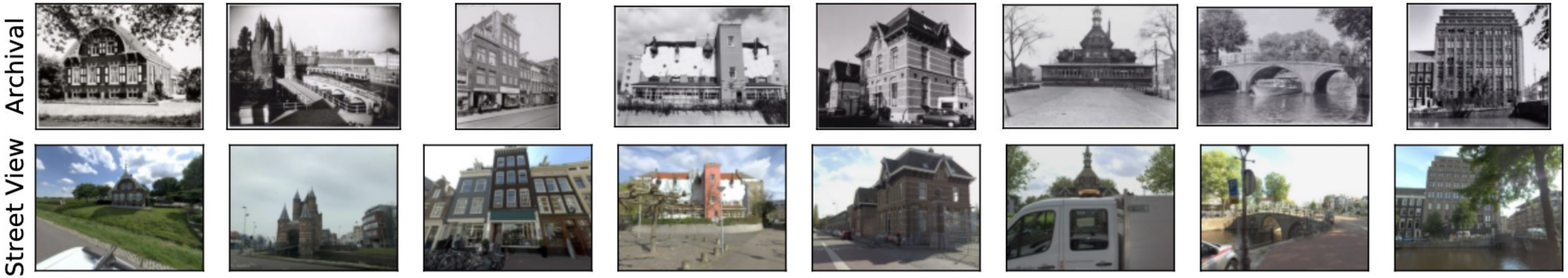


Fig. 2. Sample image pairs from AmsterTime dataset. Challenges are extreme occlusions, view point changes, camera lens distortions, color changes.

- **Appearance Change**
  The system must stay reliable despite major variations in lighting, weather, shadows, seasons, vegetation, or structural changes.
- **Scalability**
  As autonomous systems collect millions of images, recognition must remain efficient and avoid computational growth with database size.
- **Viewpoint Change**
  Changes in camera pose, trajectory, or field of view can drastically alter pixel arrangements, causing retrieval failures.
- **Perceptual Aliasing**
  Visually distinct places may look nearly identical (e.g., corridors, forests, office hallways, similar suburban streets), making confusion likely.

Graph-based Solution for Appearance Change & Scalability

Doan, A.-D., Latif, Y., Chin, T.-J., Liu, Y., Do, T.-T., & Reid, I. *(2019).*
Scalable Place Recognition Under Appearance Change for Autonomous Driving.
In ***Proceedings of the IEEE/CVF Conference on International Conference of Computer Vision (ICCV)****.*

- **HMM (**Hidden Markov Model**)** Based Graph Formulation for VPR
- **Places as Nodes:** Each database place/image is a node $V_1, V_2, \dots, V_K$
- **Motion as Edges:** Transitions modeled by $P(k_2 \mid k_1)$
- **Sequence Matching:** HMM matches query sequence $Q_1, \dots, Q_T$ to the most likely node sequence.
- **Robustness:** Global sequence reasoning reduces errors from weather, lighting, and viewpoint changes.
- **Scalability:** Graph + HMM can be efficiently updated as new routes are added.

Co-Visibility Graphs for Multi-View VPR

Stumm, E., Mei, C., Lacroix, S., Nieto, J., Hutter, M., & Siegwart, R. (2016).
Robust Visual Place Recognition with Graph Kernels.
Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

What is a Co-Visibility Graph? It is a **structural graph** that models the **spatial relationships** in a scene based on shared viewing potential. It transforms a collection of raw images into a **semantic, topological structure** for **spatial reasoning**.

Graph component: **Nodes (Vertices):** Represent individual **viewpoints** (single images) or specific **physical areas/landmarks**. * **Edges (Connections):** Exist between two nodes if they **share a significant, visible surface area** in the 3D scene (i.e., they can "co-see" the same region).

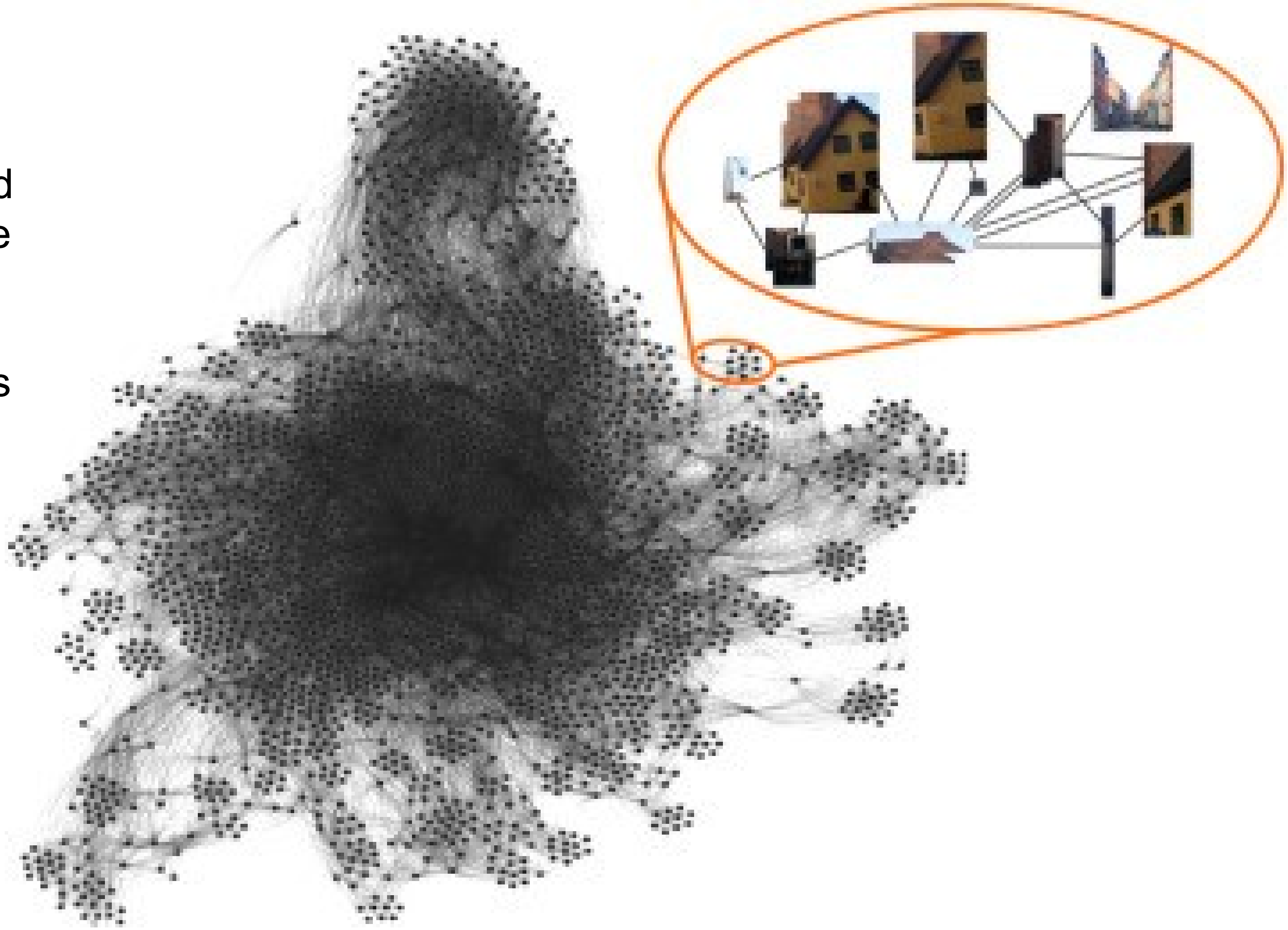
Benefit: It addresses the view-point change challenge and perceptual aliasing problem. Two places may look identical in one image, but their multi-view co-occurrence patterns of landmarks are different covisibility graphs capture those patterns and therefore break the aliasing symmetry.

Co-Visibility Graph Example

•**Global c-ovisibility graph** where nodes are images and edges indicate shared visible 3D structure.

•**Zoomed-in example** shows how neighboring viewpoints connect through overlapping landmarks.

•Supports **robust loop closure** by leveraging multi-view structural relationships beyond single-image appearance.



Cascianelli, S., Costante, G., Bellocchio, E., Valigi, P., Fravolini, M. L., & Ciarfuglia, T. A. (2017). ***Robust visual semi-semantic loop closure detection by a covisibility graph and CNN features.*** *Robotics and Autonomous Systems, 92, 53–65.*

3. Scene Graph

Structured Image Understanding : Scene Graphs

- **Vision-Language (V-L) Reasoning:** Provides an explicit, structured link between vision and language, enabling relational and contextual reasoning for VQA and detailed captioning.
- **Controllable Image Generation:** Acts as a semantic blueprint that allows for precise, compositional control over generated and edited scenes, ensuring accuracy in object placement and interaction.
- **Affordance & Embodied AI:** Forms a knowledge base for Robotics and Embodied AI by encoding affordances (how objects can be used) and functional relationships for effective task planning.

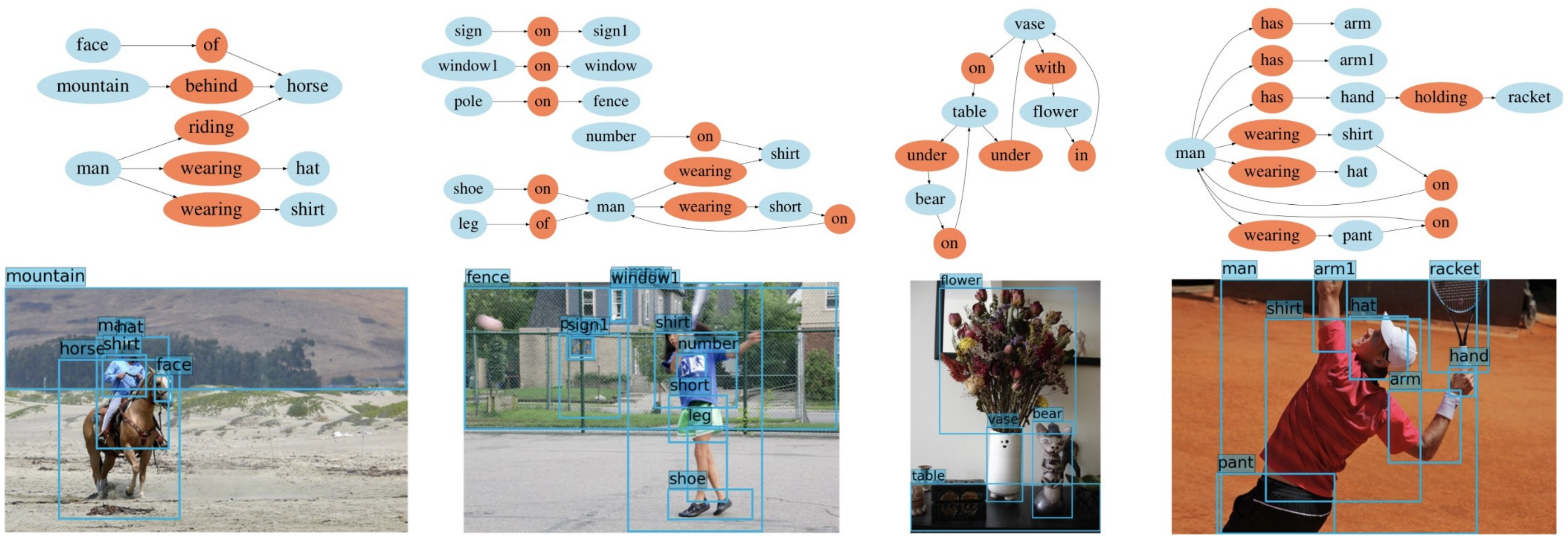
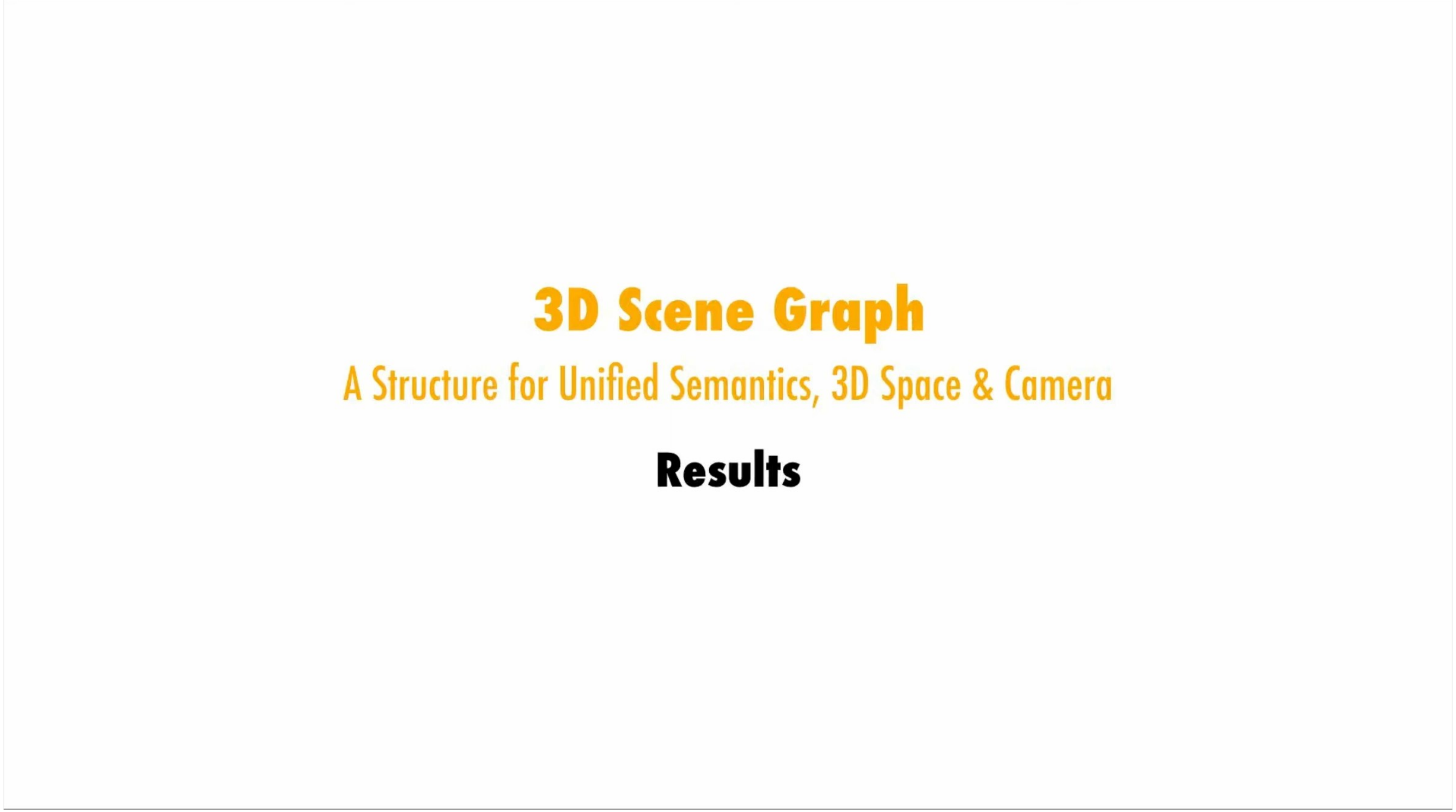


Image credit : Scene Graph Generation by Iterative Message Passing, Xu, Danfei and Zhu, Yuke and Choy, Christopher and Fei-Fei, Li, CVPR 2017

3D Scene Graph



3D Scene Graph: A Structure for Unified Semantics, 3D Space, and Camera

Iro Armeni[1] Jerry Zhi-Yang He[1] JunYoung Gwak[1] Amir R. Zamir[1,2]
Martin Fischer[1] Jitendra Malik[2] Silvio Savarese[1]
[1] Stanford University [2] University of California, Berkeley
http://3dscenegraph.stanford.edu

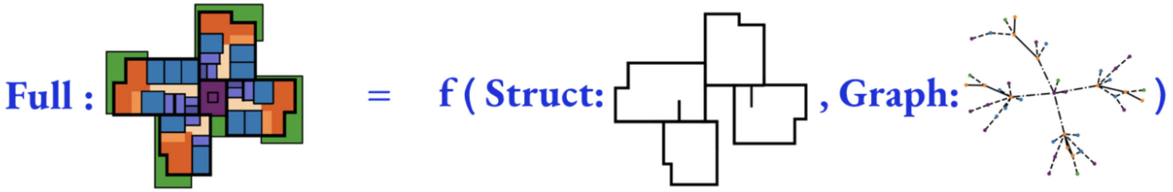
Co-Visibility Graph vs. Scene Graph

Feature	Co-Visibility Graph (VPR/SLAM)	Scene Graph (Reasoning/VQA)
Nodes represent	**Visual Landmarks** (Keypoints, Visual Words) or **Images/Keyframes**.	**Semantic Objects** (Car, Person, Building).
Edges represent	**Co-visibility** (landmarks seen together in the same image) or **Spatial Proximity/Overlap**.	**Semantic Relationships** (on, next to, holding).
Goal	**Localization and Mapping.** Used for loop closure and building a robust geometric map.	**Interpretation and Reasoning.** Used for high-level semantic understanding.

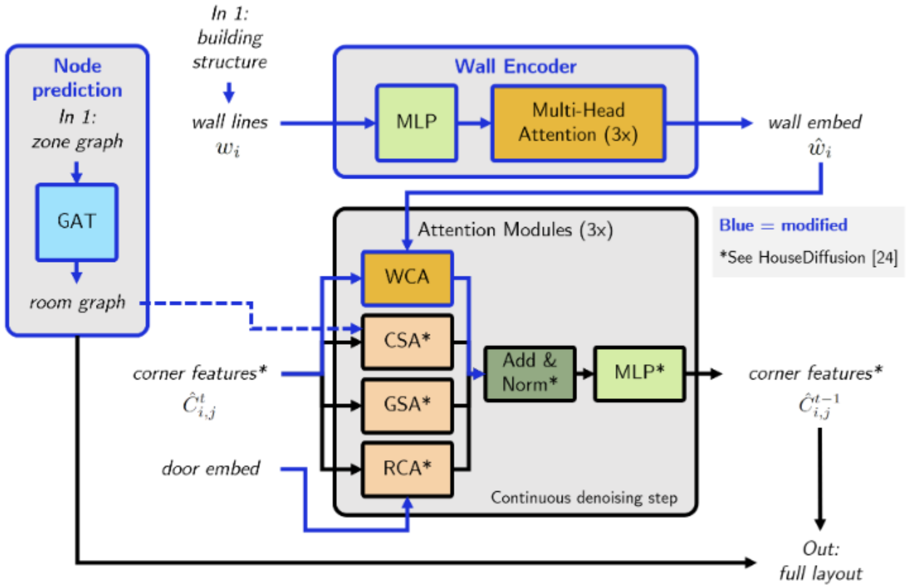
4. Floorplan Generation & Search

Floor plan generation

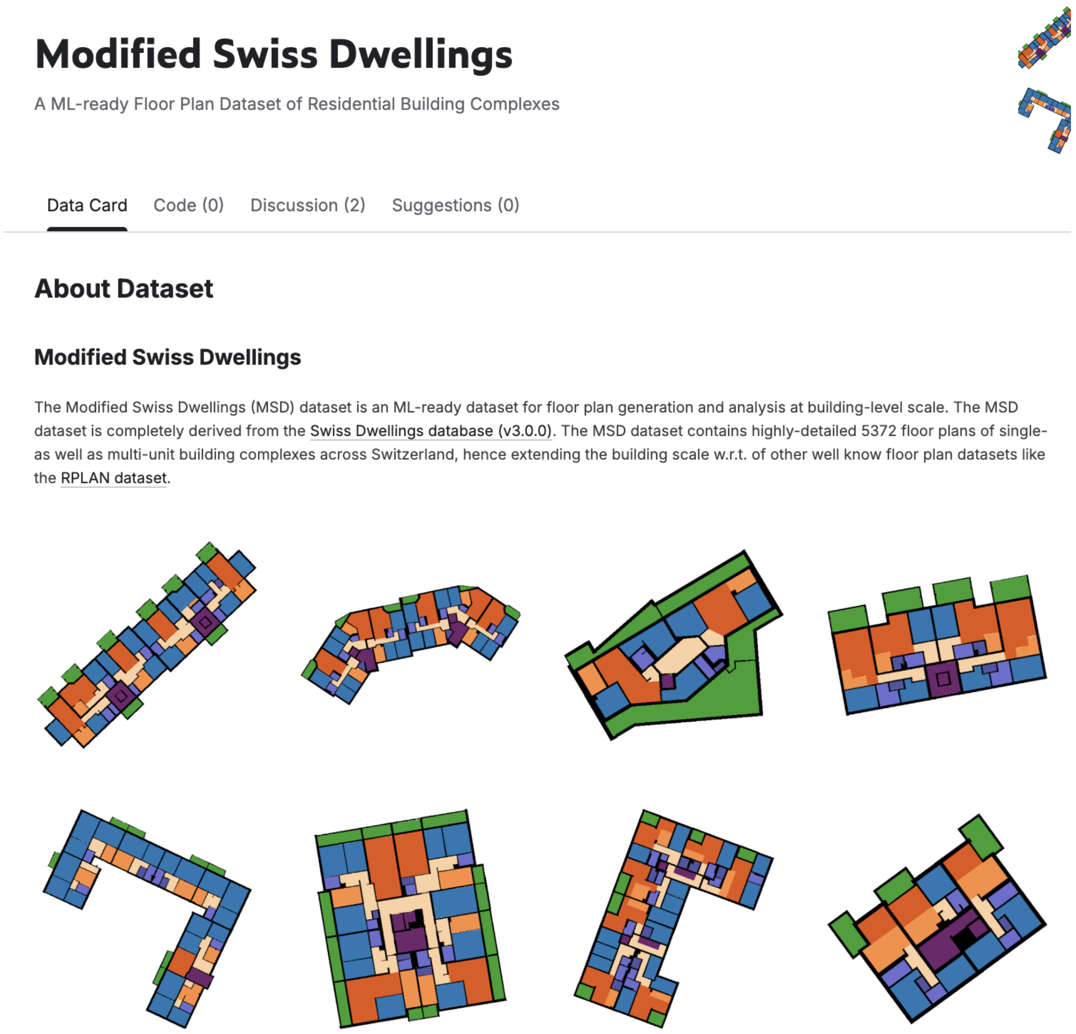
The floor plan generation task takes as input the boundary of a building, the structural elements necessary for the building's structural integrity, and a set of user constraints formalized in a graph structure, with the goal to automatically generate the full floor plan. While previous research on floor plan generation has mainly focused on the scale of individual apartments, our challenge sets the stage for floor plan generation at a larger scale: the scale of the apartment complex. With the help of Archylise, we developed Modified Swiss Dwellings: a ML-ready Dataset for Floor Plan Generation at Scale which was used for the challenge.



MSD Benchmark Dataset

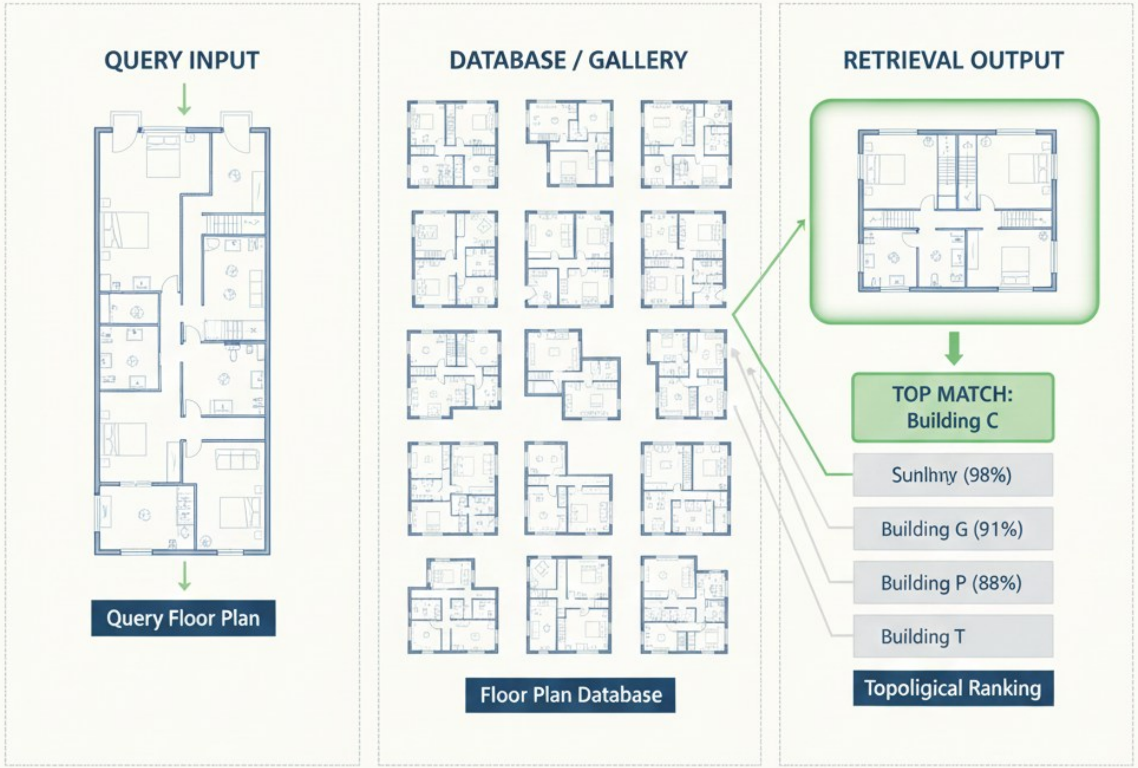


Modified House Diffusion (MHD) extends HouseDiffusion by enriching the representation of room boundaries and relationships:



van Engelenburg, C., Mostafavi, F., Kuhn, E., Jeon, Y., Franzen, M., Standfest, M., van Gemert, J., & Khademi, S. *(2024). MSD: A Benchmark Dataset for Floor Plan Generation of Building Complexes. In* ***European Conference on Computer Vision (ECCV 2024)***

Topological Search Engines



Topological Search Engines are advanced tools designed for experts working with complex, structured knowledge, such as **architects, spatial planners, and urbanists.**

How they work:

•They accept a **graph** (nodes and edges with diverse data) as a query.

•They search the database, **ranking entities** based on their **structural and data similarity** to the query graph.

Why they matter: This type of search is crucial for finding **structurally related information l**ike similar floor plans, urban layouts.

Conclusion

1. Structured spatial data unlocks meaningful analysis.
Turning the continuous physical world into digital representations—maps, images, and graphs—enables efficient computation, reasoning, and explicit modeling of topology.

2. Graphs unify spatial representation across all scales.
From planetary systems to cities and buildings, graphs provide a flexible way to encode connectivity, relationships, and structure.

3. Adding semantics transforms *space* into *place*.
Enriching graph nodes and edges with functional, environmental, or experiential attributes makes spatial data interpretable and actionable.

4. Vision + Graphs enable robust world understanding.
Combining visual features with graph structure supports reliable localization, mapping, and high-level reasoning in complex environments.

5. Efficient graph-based search engines will power future spatial intelligence.
Advancing retrieval and reasoning algorithms will allow us to build powerful knowledge machines capable of supporting rich, data-driven decision-making in spatial and structural domains.

Extra References

- **Scene Graphs / GNNs**

Yang, R., Li, X., Chen, Z., & Zhang, Y. (2018). **Graph R-CNN for Scene Graph Generation**. *Proceedings of the European Conference on Computer Vision (ECCV)*.

Johnson, J., Hariharan, B., Maaten, L. V. D., Fei-Fei, L., Zitnick, C. L., & Girshick, R. B. (2018). **Image Generation from Scene Graphs**. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Hildebrandt, M., Hu, M., Zhu, Y., & Li, Y. (2020). **Scene Graph Reasoning for Visual Question Answering**. *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*.

- **3D Scene Graphs / Embodied AI**

Armeni, I., Gholami, S., Sunkavalli, K., Thies, J., & Nießner, M. (2019). **3D Scene Graph: A Structure for Unified Semantics, 3D Space, and Camera**. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Hughes, J., Pumarola, A., Rosinol, A., & Sola, A. (2022). **Kimera: from SLAM to Spatial Perception with 3D Dynamic Scene Graphs**. *The International Journal of Robotics Research (IJRR)*.

Murali, A., Chen, Y., Schill, B., & Zhang, Q. (2021). **TASKOGRAPHY: Evaluating robot task planning over large 3D scene graphs**. *Proceedings of the Conference on Robot Learning (CoRL)*.

- **Visual Place Recognition (VPR)**

Arandjelovic, R., Gronat, P., Torii, A., Schindler, G., & Sivic, J. (2016). **NetVLAD: CNN Architecture for Large-Scale Place Recognition**. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Hou, K., Chen, R., Zhang, J., Cui, M., Su, S., Lin, B., Zhang, Y., & Liu, X. (2021). **AmsterTime: A Visual Place Recognition Benchmark Dataset for Severe Domain Shift**. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Qin, Y., Chen, Z., Ma, T., & Wang, H. (2024). **A visual place recognition approach using learnable feature map filtering and graph attention networks**. *IEEE Robotics and Automation Letters (RAL)*.