GG-SSMs: Graph-Generating State Space Models

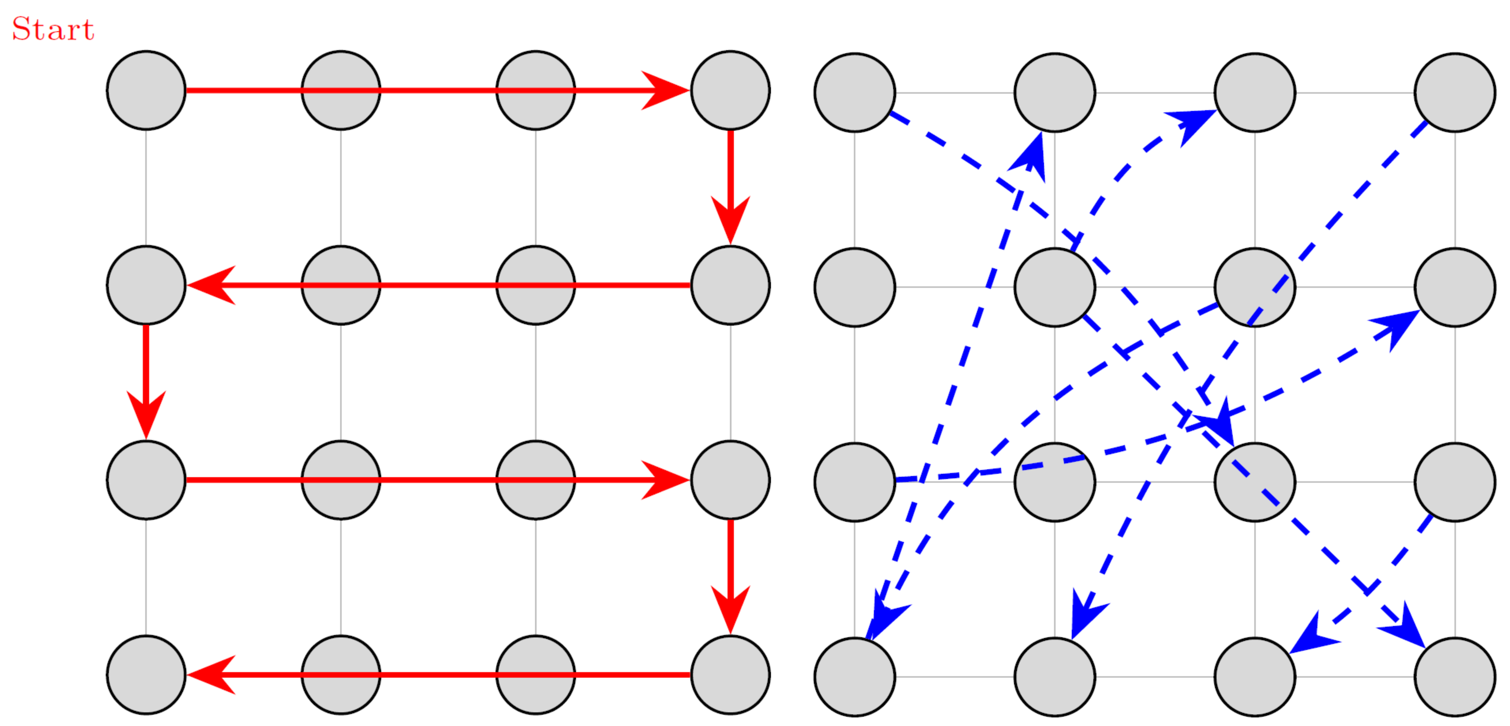
University of Zurich[UZH]

ROBOTICS & PERCEPTION GROUP

Nikola Zubić and Davide Scaramuzza

Motivation:

State Space Models (SSMs) excel at 1D sequential data but struggle with high-dimensional inputs like images. Traditional SSMs use fixed, one-dimensional scanning, limiting their ability to capture non-local interactions. Even recent methods (Mamba, Vim, VMamba) use predetermined scan paths, failing to adapt to complex, data-inherent structures.



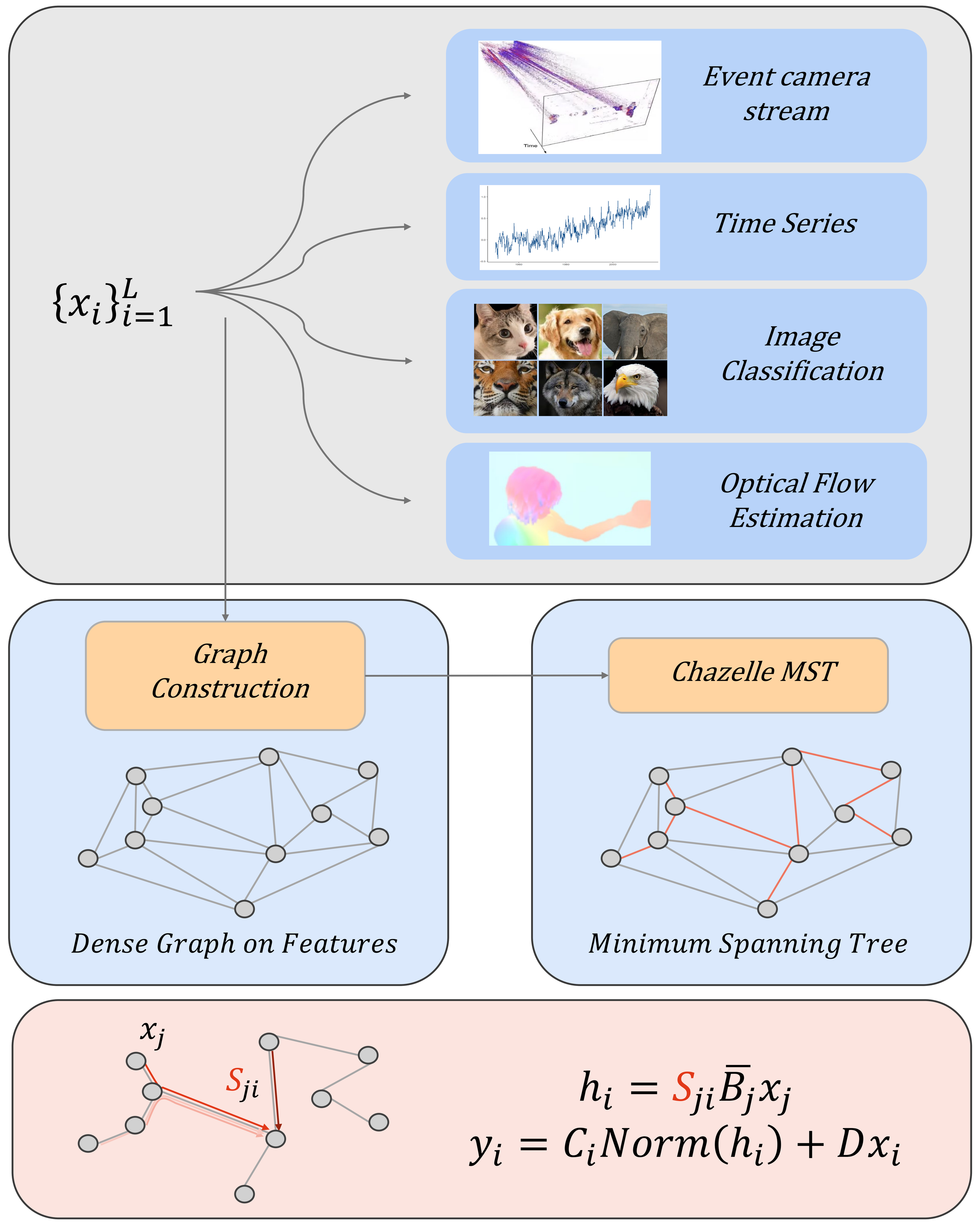
Fixed 1D Scan Path   Desired Complex Interactions

Our Key Insight: Dynamic Graph Generation

- We introduce Graph-Generating State Space Models (GG-SSMs)
- GG-SSMs dynamically construct graphs based on feature relationships within the data
- This allows SSMs to adapt their information propagation pathways to the data's intrinsic structure, capturing complex dependencies efficiently
- Core Contributions:
  - **Dynamic Graph-Based SSMs:** Novel integration of dynamic graph structures into SSMs
  - **Efficient Computation with MSTs:** Leveraging Chazelle's MST algorithm for scalable graph construction
  - **State-of-the-Art Performance:** Outperforming existing methods across 11 diverse vision and time series datasets

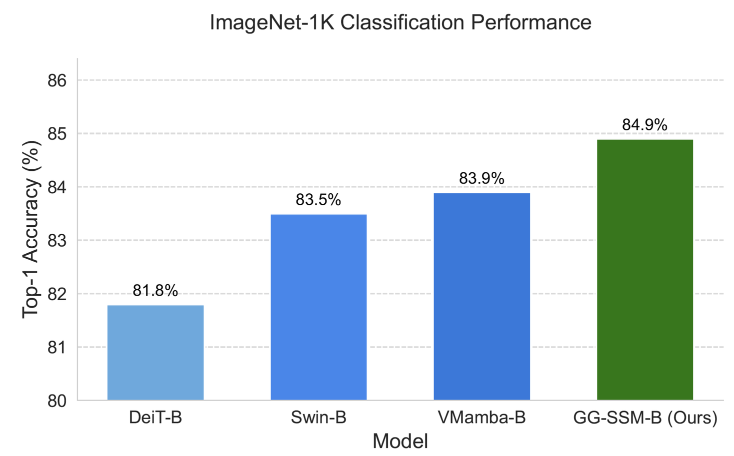
The GG-SSM Framework:

- **Input Features $\{x_i\}$:** Given an input feature set (e.g., image patches, time series tokens)
- **Graph Construction:** A dense graph is implicitly formed based on feature dissimilarities (e.g., cosine distance)
- **Chazelle's MST:** We efficiently compute a Minimum Spanning Tree (MST) from this graph. This tree captures the most significant feature relationships with minimal edges
- **SSM Propagation:** State information is propagated along the paths of this dynamically generated MST, enabling adaptive and robust feature learning (y_i)

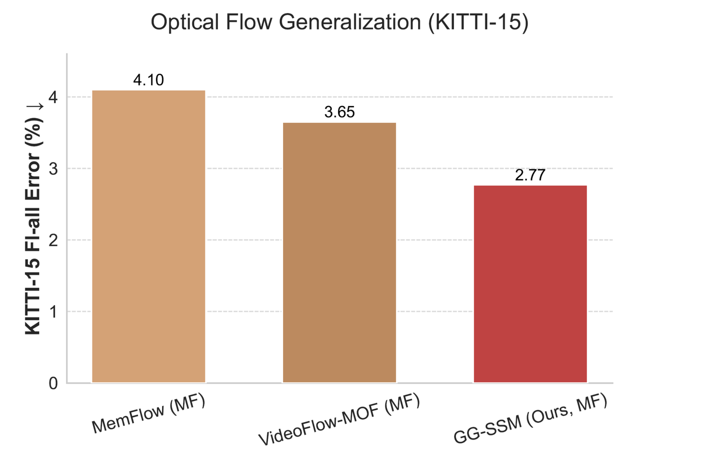


Experimental Results

State-of-the-Art Across 11 Diverse Datasets



GG-SSM-B achieves 84.9% Top-1 accuracy, outperforming prominent Transformer (Swin-B) and prior SSM (VMamba-B) architectures



GG-SSM achieves a SOTA Fl-all error of 2.77%, demonstrating superior generalization in motion estimation (MF denotes multi-frame input)

Dataset (Avg. MSE / MAE)	GG-SSM (Ours) (Lower is Better)	Next Best Competitor (Model Name)
Weather	**0.225 / 0.262**	0.251 / 0.276 (S-Mamba)
Solar-Energy	**0.183 / 0.246**	0.233 / 0.262 (iTransformer)
ETTh2	**0.353 / 0.401**	0.374 / 0.398 (RLinear MAE)
Traffic	**0.374 / 0.263**	0.414 / 0.276 (S-Mamba)

GG-SSM consistently sets new state-of-the-art or achieves highly competitive results (Avg. MSE/MAE) across six diverse time series forecasting benchmarks, outperforming specialized models like S-Mamba and iTransformer

Efficiency and Scalability

Model (ImageNet)	Params (M) ↓	FLOPs (G) ↓
VMamba-T	30	4.9
GG-SSM-T (Ours)	**28**	**4.4**
VMamba-S	50	8.7
GG-SSM-S (Ours)	**49**	**6.6**
VMamba-B	89	15.4
GG-SSM-B (Ours)	**87**	**14.1**

Our GG-SSM variants (highlighted in green) consistently demonstrate lower parameter counts (Params (M)) and computational costs (FLOPs (G)) compared to their VMamba counterparts

Ablation study of MST algorithms on the ETTh2 dataset. Chazelle's algorithm achieves comparable predictive performance (MSE) while offering significantly better runtime efficiency

Algorithm	ETTh2 (Horizon=192) MSE	Time (s/epoch)
Kruskal	0.356	$1.00\times$
Prim	0.358	$0.95\times$
Chazelle	0.353	$0.88\times$

Check out our Paper, Code and more!

Paper link   Code   ML research at RPG   Nikola Zubić's homepage

Sponsors: Horizon2020 European Union Funding for Research & Innovation   erc

Efficient Graph Construction: Chazelle's MST Algorithm

- Chazelle's algorithm finds the MST - a tree connecting all nodes with the minimum total edge weight
- **Why Chazelle's?** It enables near-linear time complexity ($O(E\alpha(E,V))$, where E, V are edges and vertices, and α is the very slow-growing inverse Ackermann function, making it extremely fast even for large, dense graphs
- **In GG-SSMs:** It allows us to dynamically build an optimal sparse graph (the MST) from input features on-the-fly, forming the backbone for efficient state propagation without predefining scan paths

The Power of Dynamic Adaptation

- Dynamic graph construction adapts to inherent data structures, unlike fixed scans
- MSTs provide an optimal, sparse backbone for feature propagation, efficiently modeling complex, long-range dependencies
- Versatile and effective for various computer vision and time series tasks