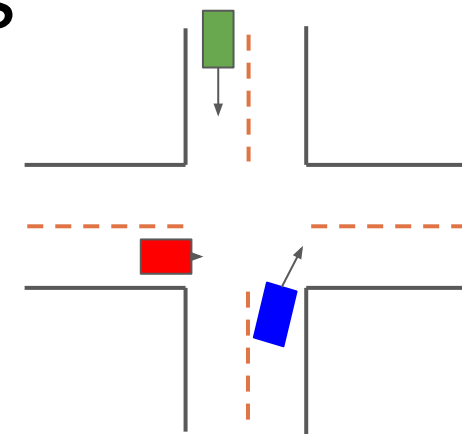


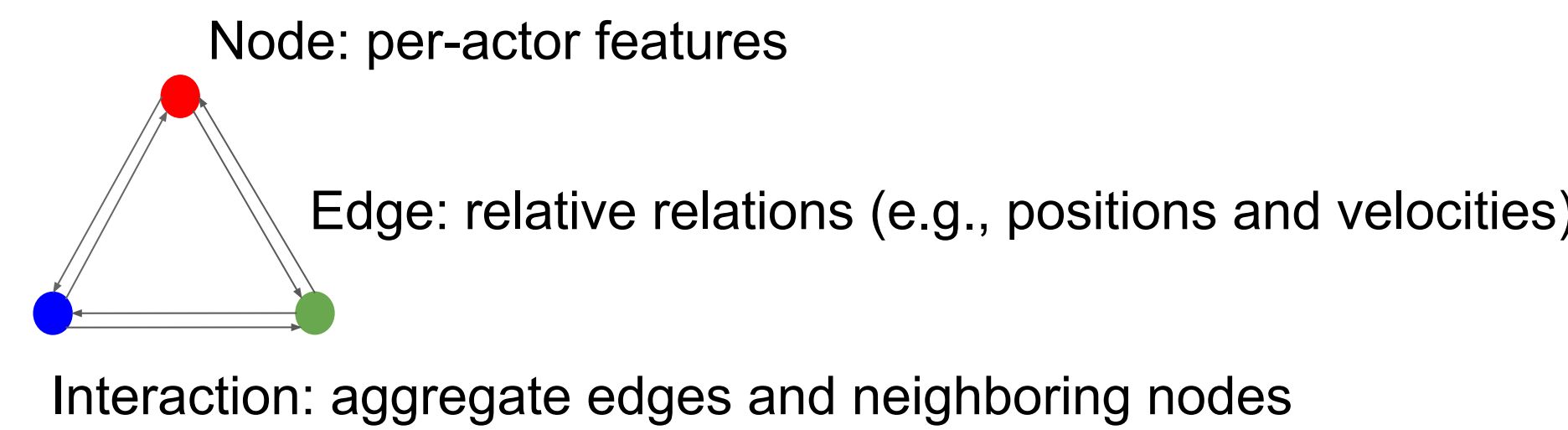
Spatial interaction:

- Spatial interaction: relations in Euclidean space. The relative spatial relation matters
- Common and critical. E.g., forecasting behaviors of traffic actors



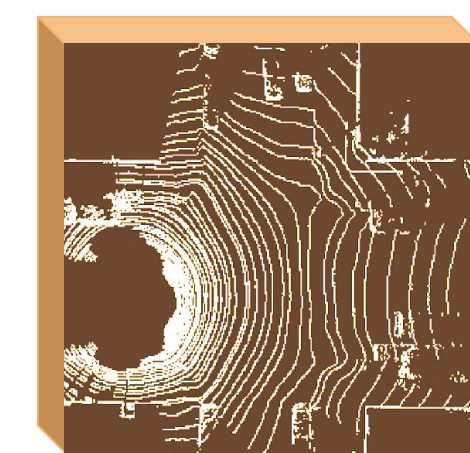
Approaches to model spatial interaction:

- Graph neural networks (GNNs)



1. Have to hand-craft and add relative relations to the edges
2. Slower than Convolutional neural networks (CNN)

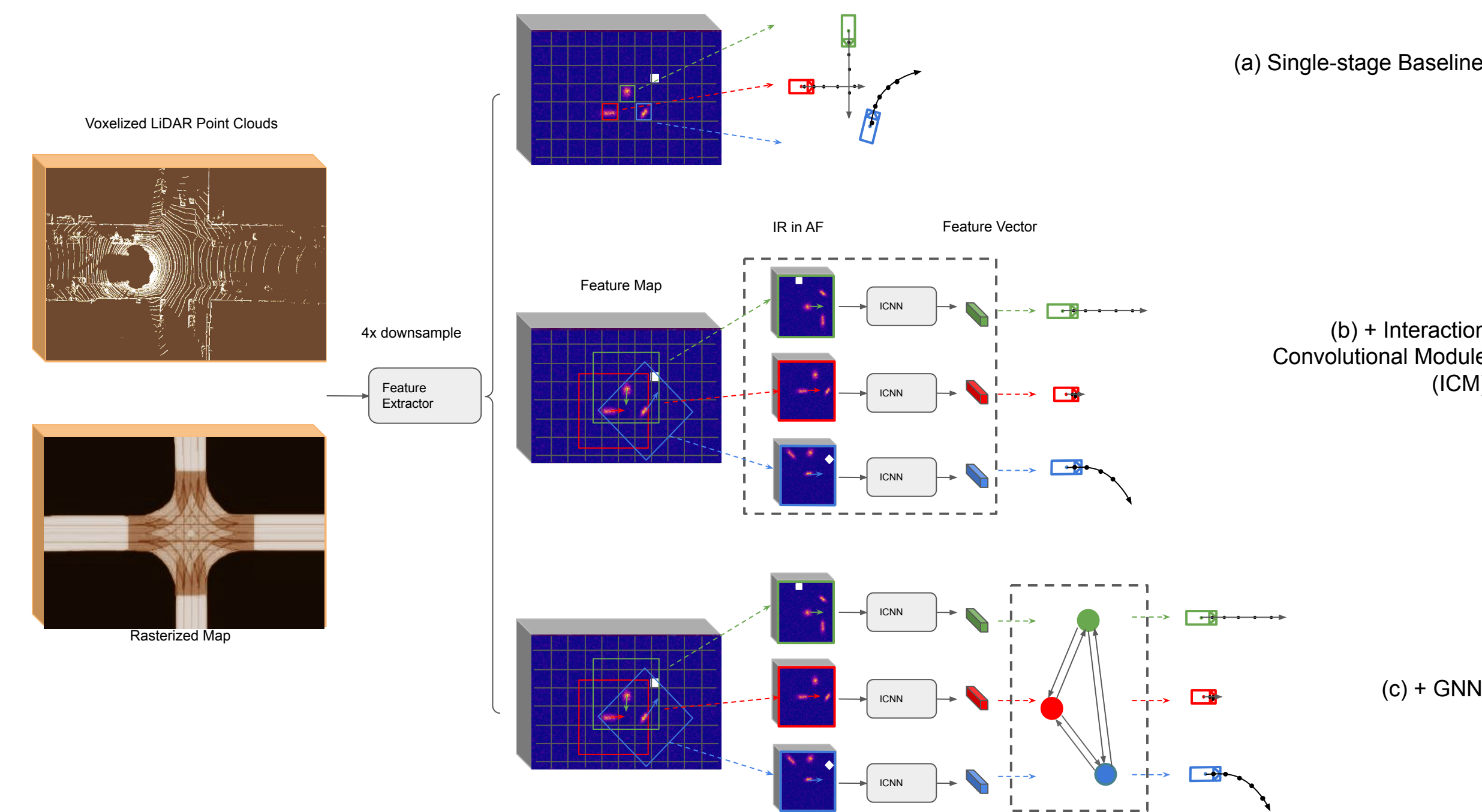
- How about convolutional layers or CNN?
 - Intuitively, convs can model spatial interactions
 - 2D and 3D conv-layers operate on data in grid forms → spatial relations are intrinsically represented
 - Large receptive fields → aggregate non-local information
 - Why interaction is ineffectively modeled, even large CNN backbones are widely used?



Voxelized lidar point-cloud at an intersection

Experiments

- Test field: forecasting trajectories of traffic actors
 - 2D top-down view
- Test data: large autonomous driving data
- Control experiments: models are almost identical, except for
 - additional light-weight conv-layers and GNNs
- Explicite interaction metrics: overlaps (i.e. collisions)

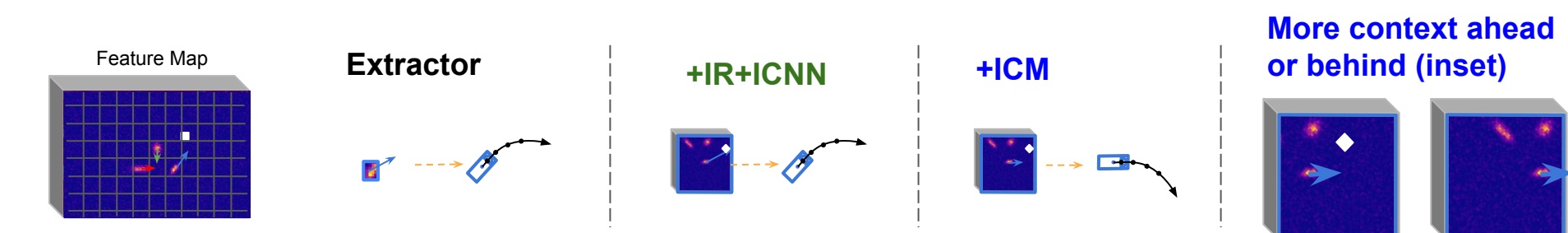
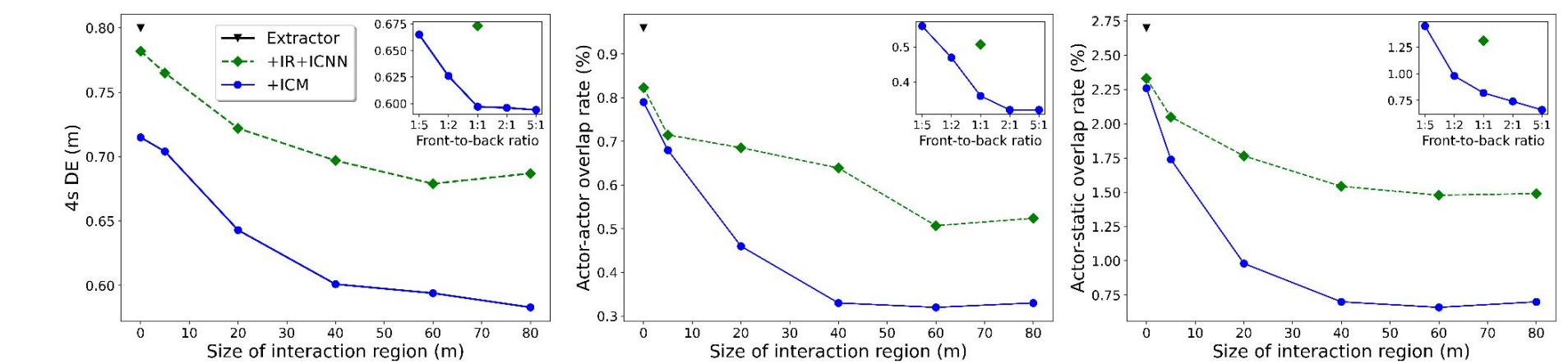


Takeaways of the experiments:

- 3 characteristics to “activate” conv-layers:
 - Large and relevant context as the input to conv-layers
 - Aggregation of per-actor feature maps using a few downsampling conv-layers
 - Overcoming the rotational variance of convolutions
- Conv approach vs. GNN
 - Convs can perform similarly to or better than GNNs
 - Adding the convs considerably improves interaction modeling even when a GNN is used
 - Adding a GNN demonstrates only minor additional gain when the convs is already used

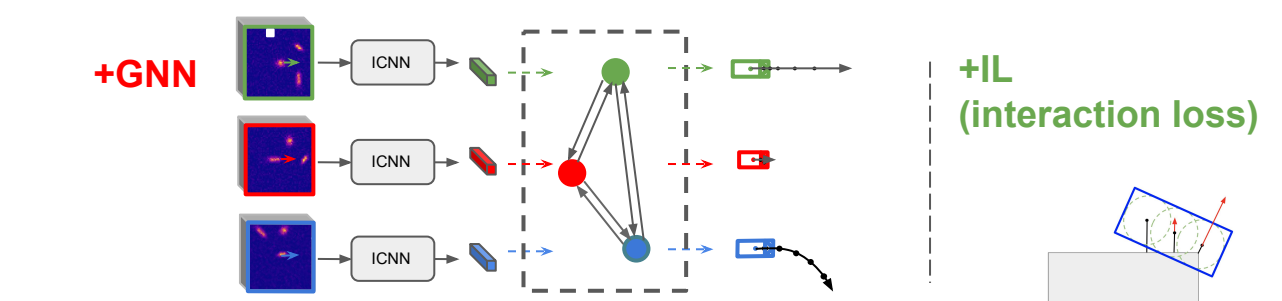
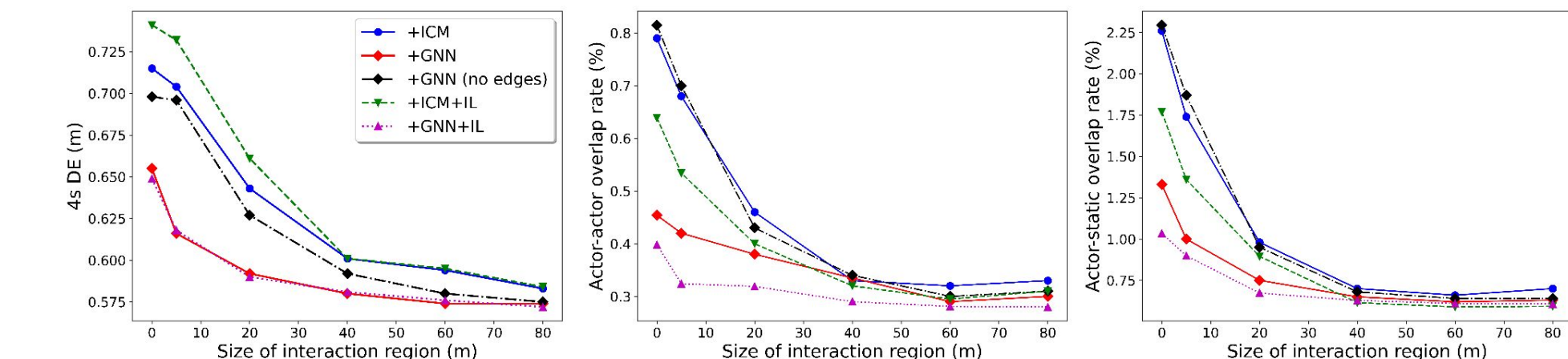
Results:

Effective convs



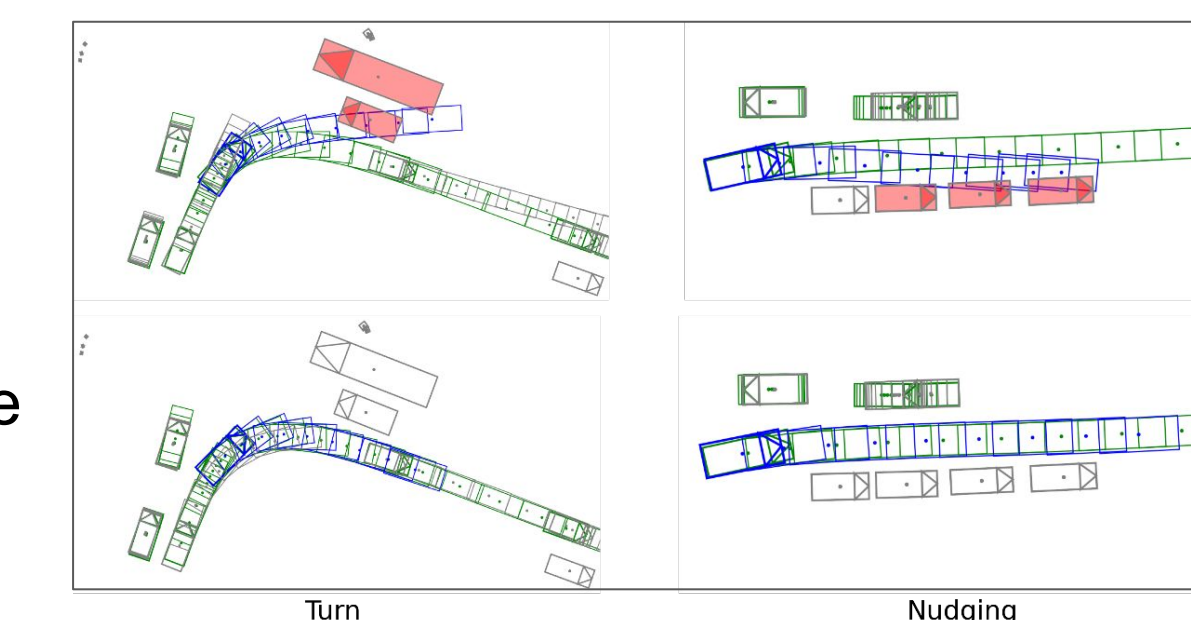
Results:

Convs vs. GNN



Baseline

Convs (crop size of 60m)



Inference times of additional modules (baseline takes 45.6 ms)

Module	Crop size (m)	Inference (ms)
Convs	0	5.2
Convs	80	8.1
GNN	-	46.9

Summary

- Identified 3 characteristics that affects conv-layers in model spatial interaction
- 2D motion forecasting evidences that convs demonstrates comparable or stronger ability than GNNs in modeling interaction
- Future: generalization to other 2D and 3D tasks with interactions

Videos ([link](#)) and supplementary ([link](#))