Convolutions for Spatial Interaction Modeling

Zhaoen Su, Chao Wang, David Bradley, Carlos Vallespi-Gonzalez,
Carl Wellington, Nemanja Djuric

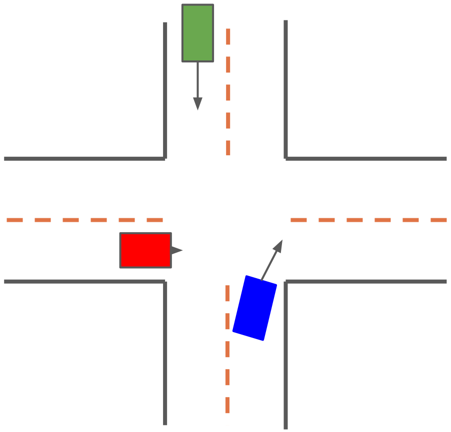
June, 2022.

Outline

- Spatial interaction
- Approaches to spatial interaction modeling
  - Graph neural networks (GNN)
  - Convolutional layers
- Empirical studies
- Summary

Spatial interaction

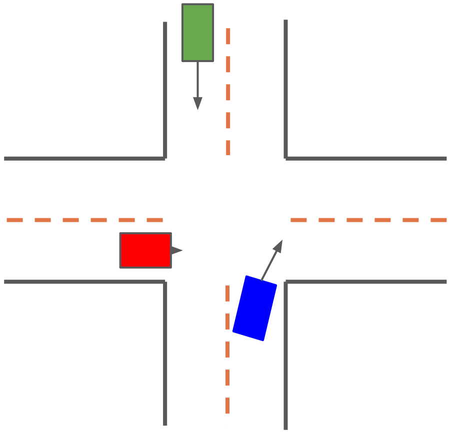
- Interactions between objects – common and critical in many application areas (e.g., robotics, self-driving cars, social networks)
- Spatial interaction – the relative spatial relation between objects matters the most
- E.g, forecasting the behavior of traffic actors which depends on both the history as well as the interactions with other actors and the environment



Popular approach: Graph Neural Network (GNN)

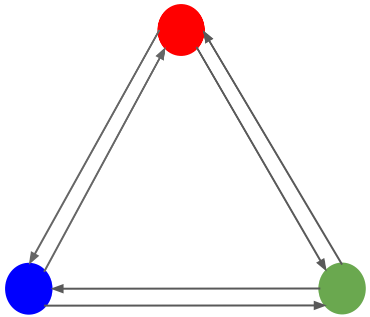
Building a graph:

- *Node*: per-actor features
- *Edge*: relative relations (e.g., positions and velocities)
- *Interaction*: aggregate neighboring edge and node features via message passing



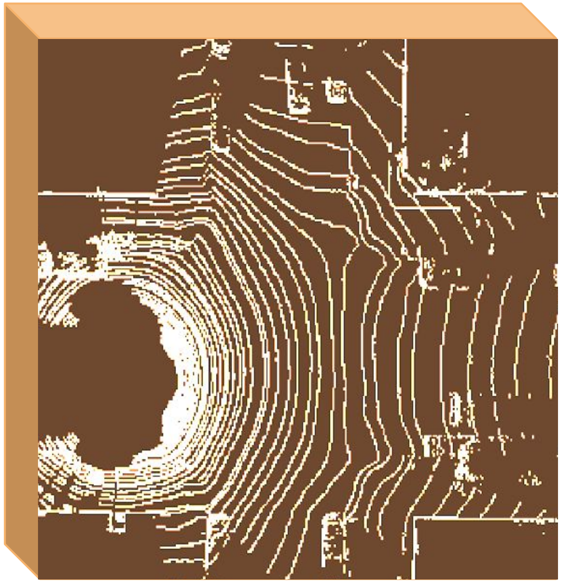
Cons:

- Have to handcraft and add Euclidean relations to the graphs
- Slower than Convolutional Neural Network (CNN)



An alternative: Convolutional Neural Network (CNN)

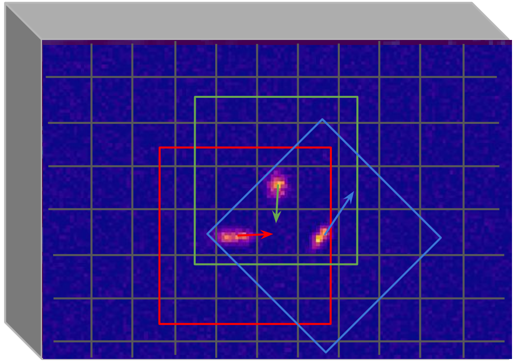
- Intuitively, conv-layers model spatial interactions
  - 2D and 3D conv-layers operate on data in grid forms: spatial relations are intrinsically represented in the Euclidean space
  - Propagation of non-local information between objects by sufficiently large receptive fields
- But why is it modeled ineffectively, when large CNN backbones are already widely used?



Voxelized lidar point-cloud at an intersection

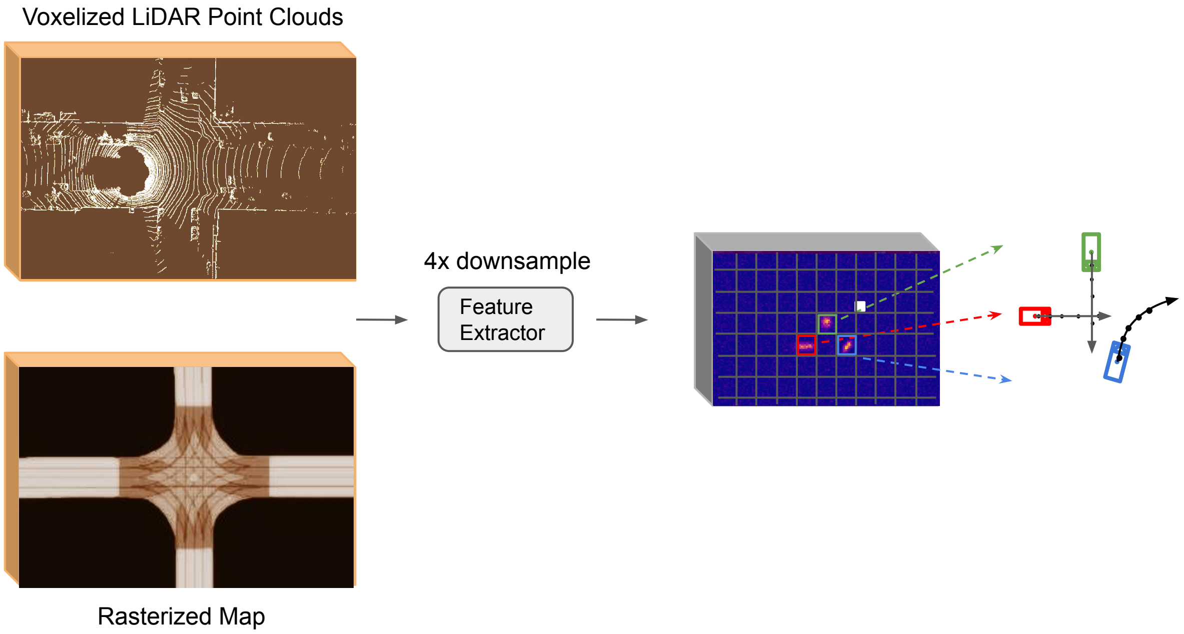
Effective modeling using CNNs

- We focus on per-actor trajectory forecasting, where convolutional layers are used to model spatial interactions with other actors
- We identify three components to improve performance of convolutions for the task of interaction modeling:
  - Large and relevant context as the input to conv-layers
  - Aggregation of per-actor feature maps using downsampling convolutions
  - Overcoming the rotational variance of conv-layers

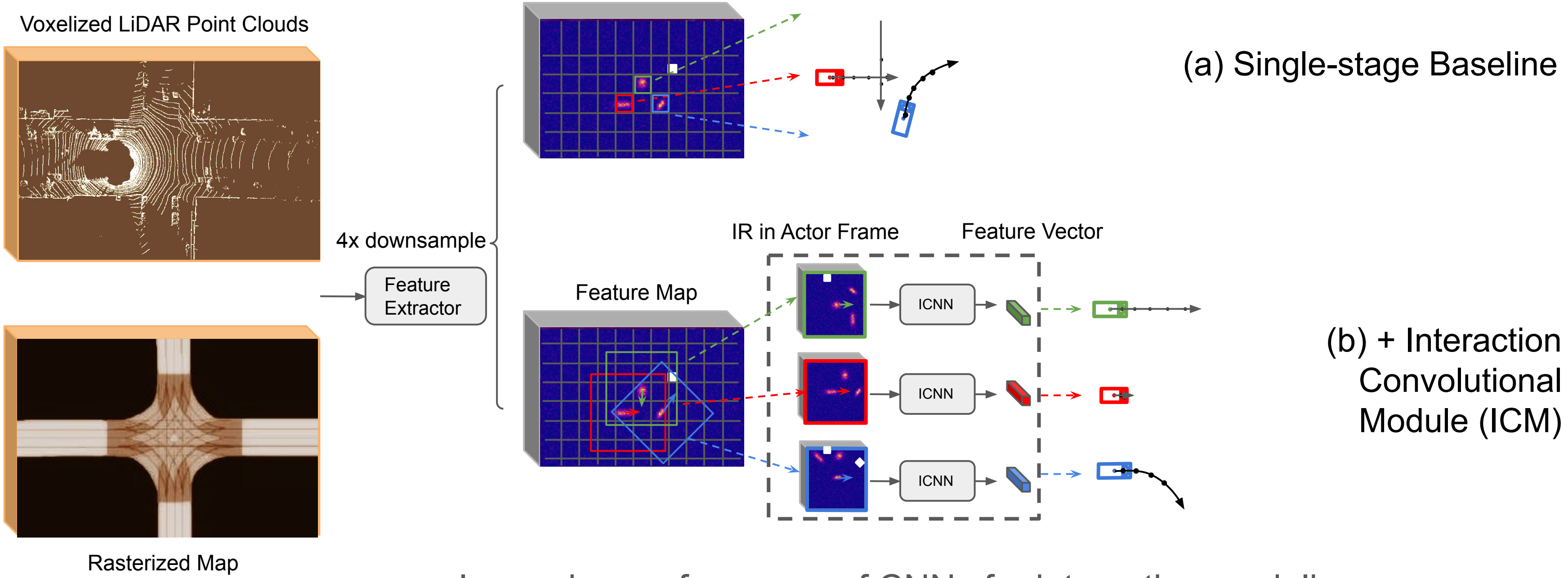


Per-actor modeling: in the 2nd stage each actor is individually processed using a crop of the feature maps around its location

Empirical studies: Baseline model



Empirical studies: Using convolutions (ICM)



Improving performance of CNNs for interaction modeling:

- Large and relevant context as an input to the conv-layers
- Aggregate per-actor feature maps using downsampling convolutions
- Overcoming the rotation variance of conv-layers

Empirical studies: Metrics and data set

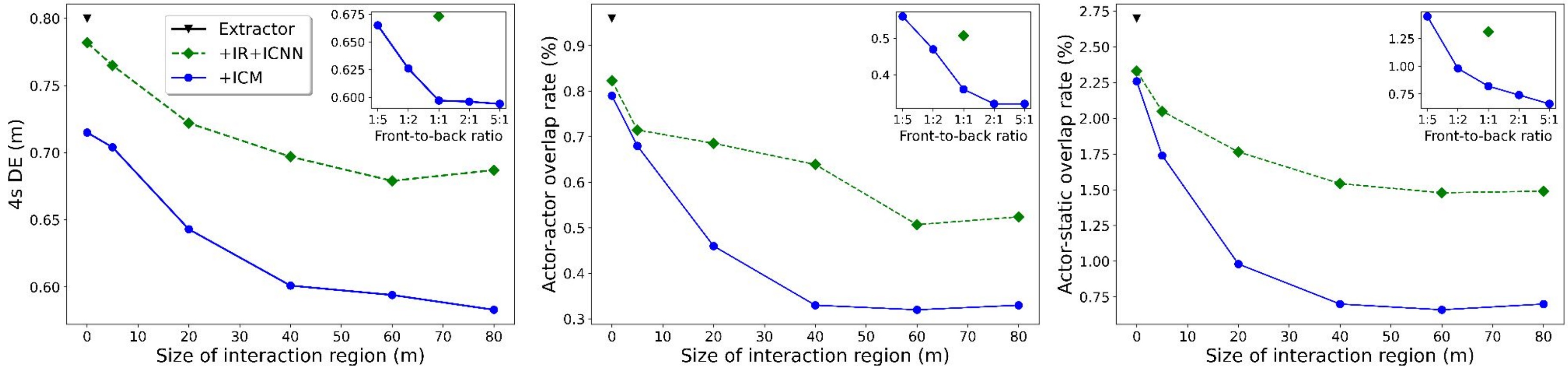
Metrics used:

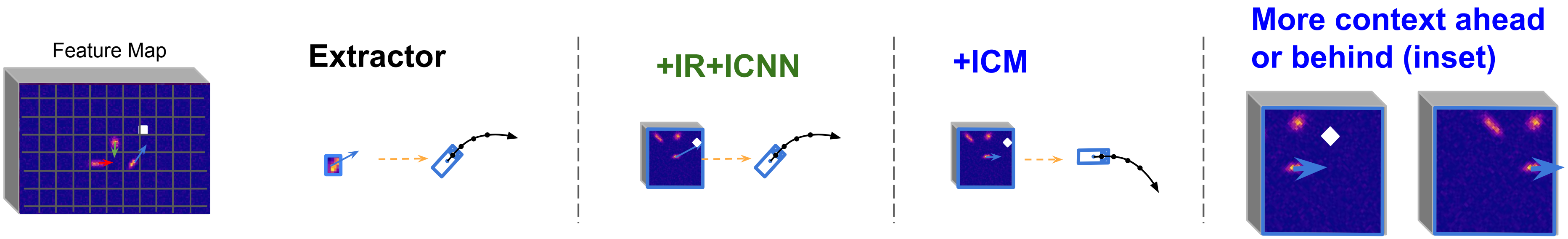
- Motion forecasting displacement errors (at 4s)
- *Actor-actor overlap rate*: percentage of predicted trajectories overlapping with predicted trajectories of other detected actors
- *Actor-static overlap rate*: percentage of predicted trajectories overlapping with ground-truth static traffic objects

Autonomous driving data set

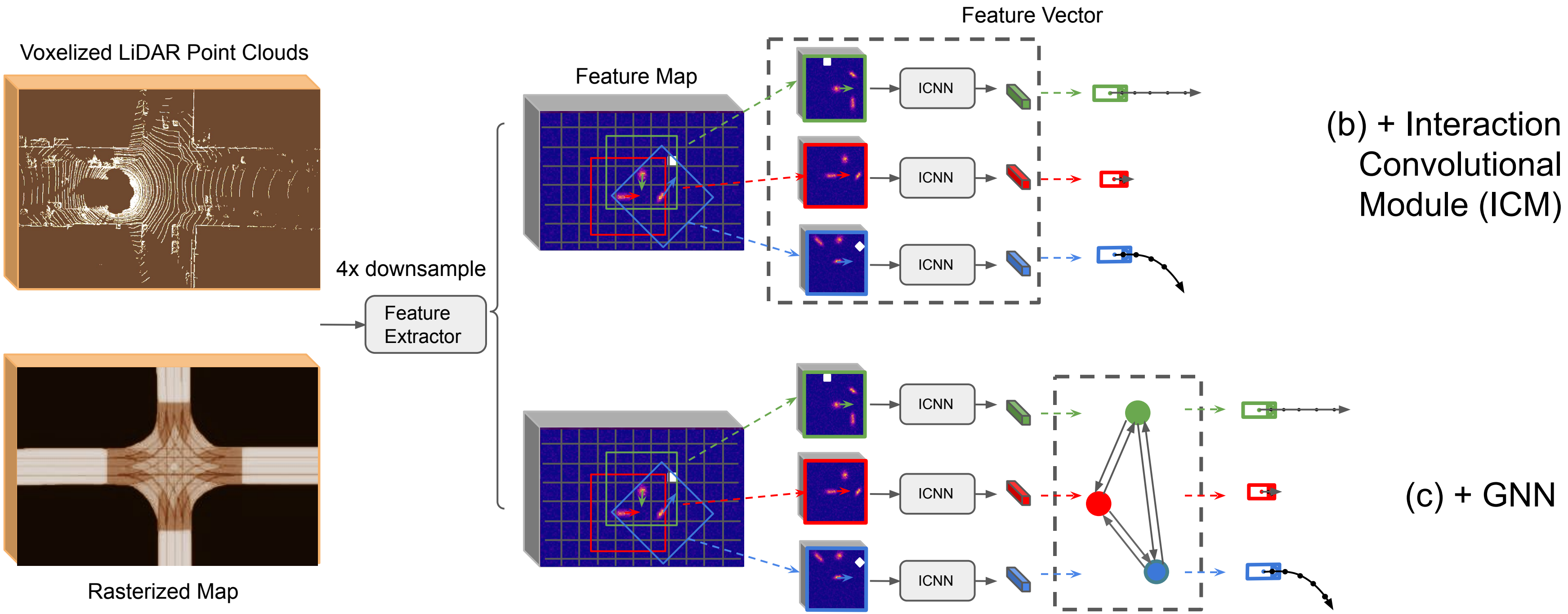
- 19,000 scenes of 25s each; collected across several cities with 10Hz labels
- 5,000 scenes in test set

Empirical studies: Using Convolutions

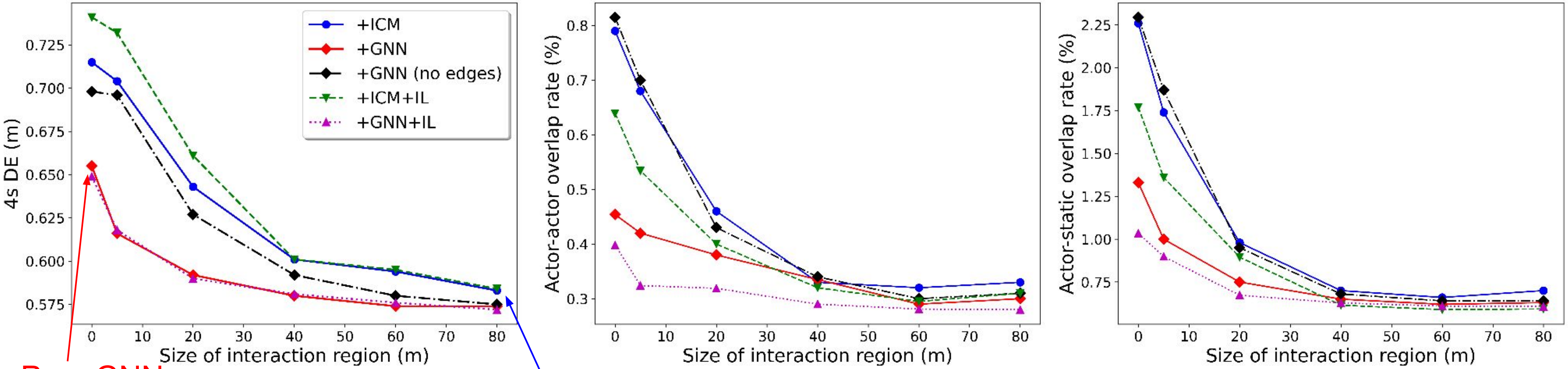




Empirical studies: additional GNN

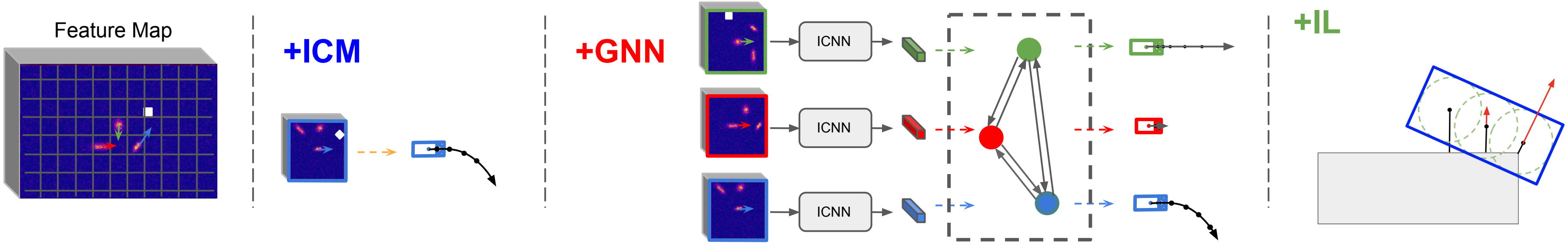


Empirical study: GNNs vs. CNNs



Pure GNN

Pure Convolutional



Qualitative results

Baseline

CNN approach
(crop size of 60m)

Turn Nudging

Red: overlapped obstacles; Blue: forecasts of the actors of interest;
Grey: forecasts of other actors; Green: labels

Summary

- We revisited convolutions for its ability in modeling spatial interaction effectively, and identified three characteristics that affect its performance
- Empirical studies show that convolutions can demonstrate comparable or even stronger ability than GNNs in modeling spatial interaction