Traffic Forecasting on New Roads Unseen in the Training Data Using Spatial Contrastive Pre-Training (SCPT)

Arian Prabowo (RMIT, DATA61/CSIRO), Hao Xue (UNSW), Wei Shao (DATA61/CSIRO), Piotr Koniusz (DATA61/CSIRO), Flora D. Sal-

Challenge: Forecasting on new roads

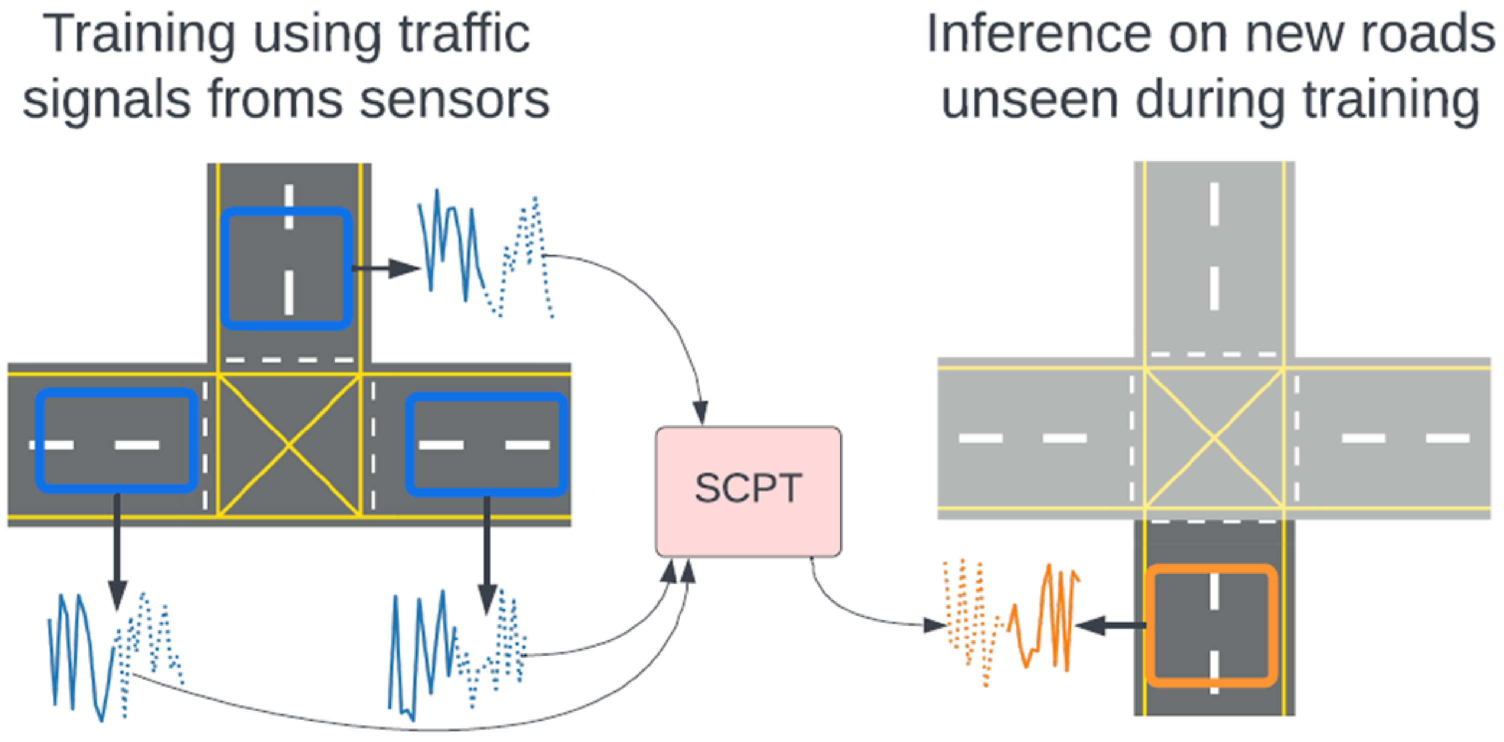


Fig. 1: Our novel traffic forecasting framework, Spatial Contrastive Pre-Training (SCPT), enables accurate forecasts on new roads (orange) that were not seen during training.

- Many models uses learned road embeddings. It is not obvious how to learn new embeddings on new roads.
- We use contrastive learning to learn road embeddings from minimal data.

A new split for a new task

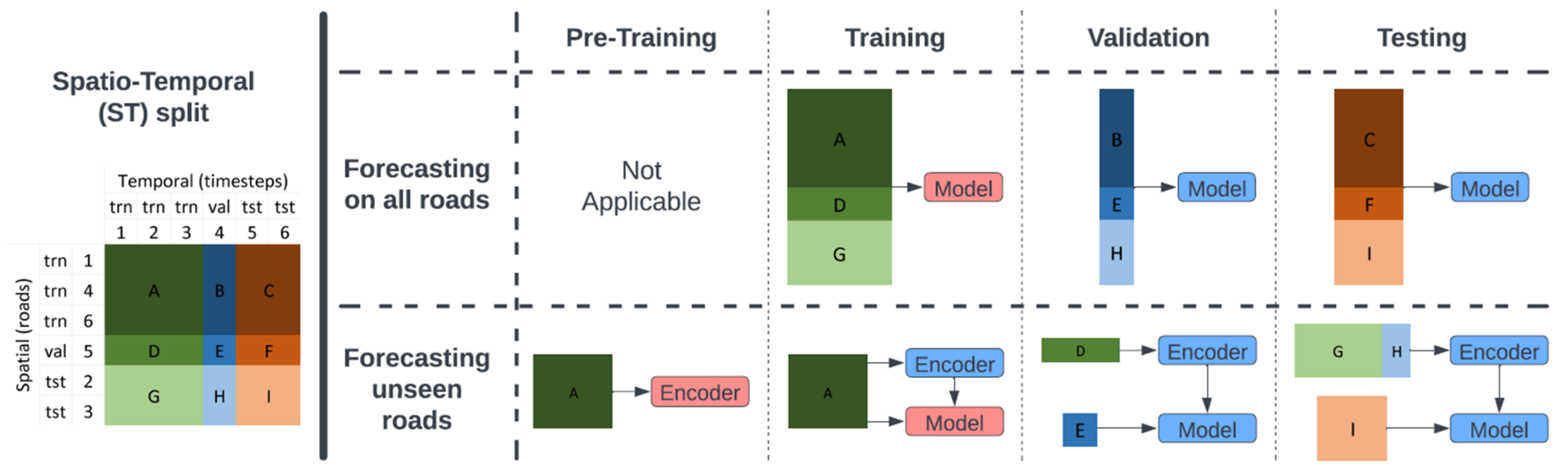
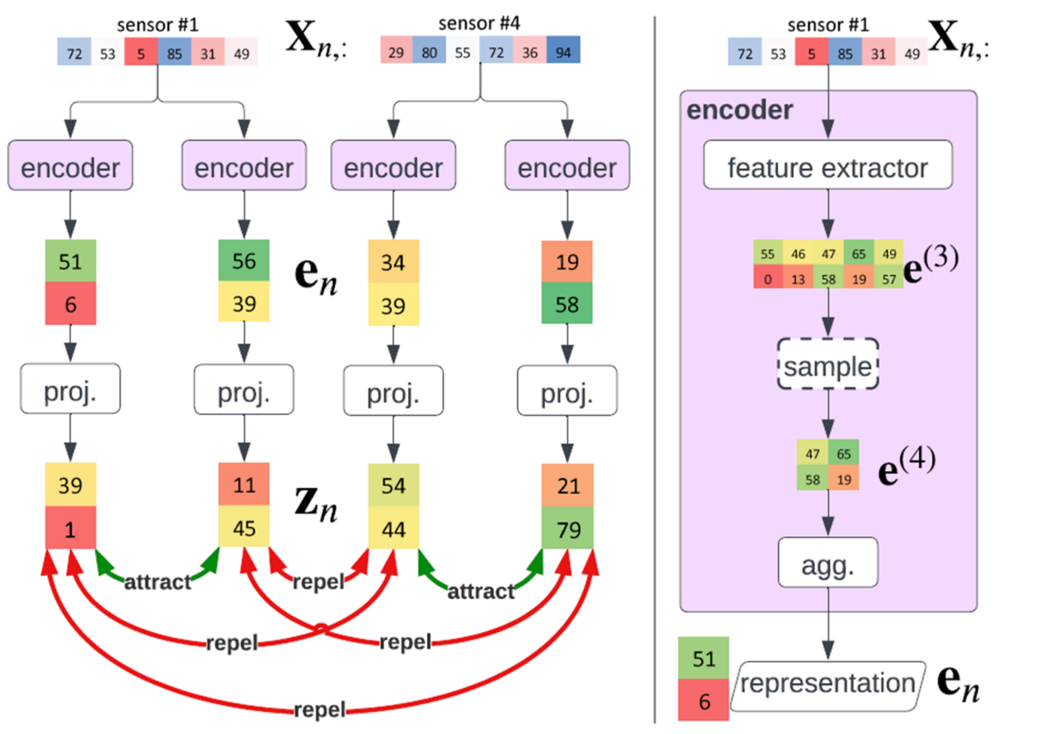


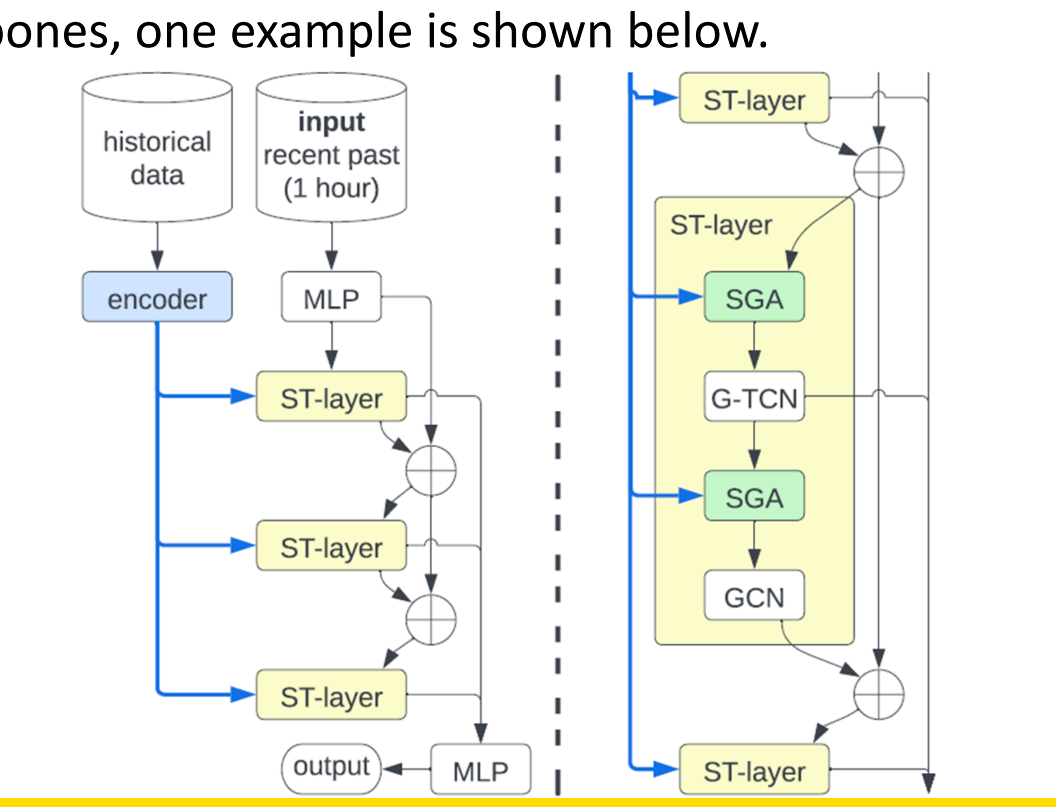
Fig. 2: The ST splitting strategy divides the dataset into nine subsets (left side), while the right side illustrates the usage of different subset combinations at different stages.

- This paper introduce a new traffic forecasting task: forecasting on new roads.
- This require a new strategy to split the dataset such that the roads in the test sets are not seen by the model during training.
- Since we use contrastive pre-training, there is a new pre-training stage here.

Method: SCPT



- We use SimCLR-like method to pre-train the spatial encoder.
- Instead of using augmentation, we use stochastic sampling to generate different views.
- This also allow us to encode timeseries of different lengths.
- The encoder can be integrated to various backbones, one example is shown below.



Results: New roads

Table 1: Performances evaluation of the SCPT framework using ST split. In this setup, the models are trained on only 70%, validated on 10%, and tested on 20% of the roads. This table shows the average performance across 12 timesteps (1 hour) on the 20% of the roads that are unseen during the training. Δ(%) denotes the percentage of error reduction.

Dataset	Methods	RMSE	MAE	MAPE
METR-LA	GWN	10.3405 ± 0.2634	4.7373 ± 0.1618	12.2677 ± 0.8058
	GWN+SCPT	$\mathbf{10.0385} \pm 0.2112$	$\mathbf{4.5645} \pm 0.1556$	$\mathbf{11.5002} \pm 0.8007$
	Δ(%)	3%	4%	6%
PeMS-BAY	GWN	4.5059 ± 0.1613	2.0126 ± 0.1037	4.7779 ± 0.4303
	GWN+SCPT	$\mathbf{3.9658} \pm 0.1266$	$\mathbf{1.8163} \pm 0.0875$	$\mathbf{4.1358} \pm 0.2740$
	Δ(%)	12%	10%	13%
PeMS-D7(m)	GWN	6.4635 ± 0.3103	3.4327 ± 0.1974	8.6896 ± 0.7844
	GWN+SCPT	$\mathbf{5.6893} \pm 0.2552$	$\mathbf{3.0794} \pm 0.1448$	$\mathbf{7.6770} \pm 0.6678$
	Δ(%)	12%	10%	12%

- The result shows that using SCPT on Graph WaveNet (GWN) backbone improves performance across all datasets and metrics.
- More importantly, the improvements become more pronounced at longer forecasting horizon as shown below.

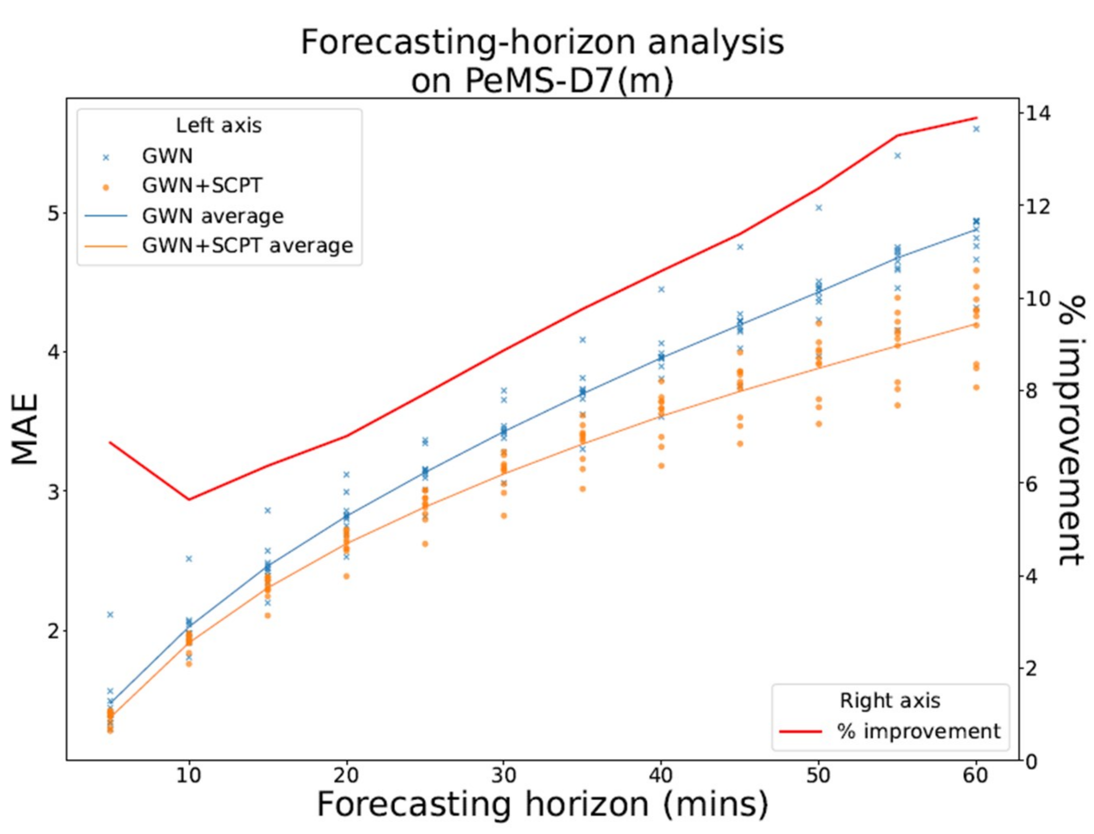


Fig. 5: Performance across forecasting horizons.

Results: Scaling

Table 4: Detailed statistics on the real world datasets.

	Dataset:	METR-LA	PeMS-BAY	PeMS-D7(m)	PeMS-11k(s)
Spatial	Nodes	207	325	228	11,160
	Edges	1,515	2,694	7,304	234,966

- Most papers are benchmarked on artificially small dataset with only few hundred nodes (first three datasets: METR-LA, PeMS-BAY, and PeMS-D7m).
- To scale to large dataset (>10 000 nodes), SCPT can be trained only on 1% of the dataset, while treating the other 99% as new roads. This enable a trade-off between performance and speed.

Method:	GWN	GWN+SCPT	Δ(%)	GP-DCRNN
RMSE	5.6345 ± 0.7469	$\mathbf{4.6741} \pm 0.2089$	17%	
MAE	2.8241 ± 0.2840	$\mathbf{2.4273} \pm 0.2171$	14%	
MAPE	5.6345 ± 0.7469	$\mathbf{4.6741} \pm 0.2089$	17%	
medianMAE12	3.4554 ± 0.2343	3.2442 ± 0.3071	6%	**2.0200**
Training time	**00:16:39**	00:22:28		7 days, 22:34:53
Roads seen in training (count)	111			11160
Roads seen in training (%)	1%			100%

Conclusion

- A new task: forecasting on new roads
- A new method: SCPT
- A new benchmark task: scaling to large network

QR to GitHub: