Investigating Pattern Neurons in Urban Time Series Forecasting

Chengxin Wang, Yiran Zhao, Shaofeng Cai, Gary Tan

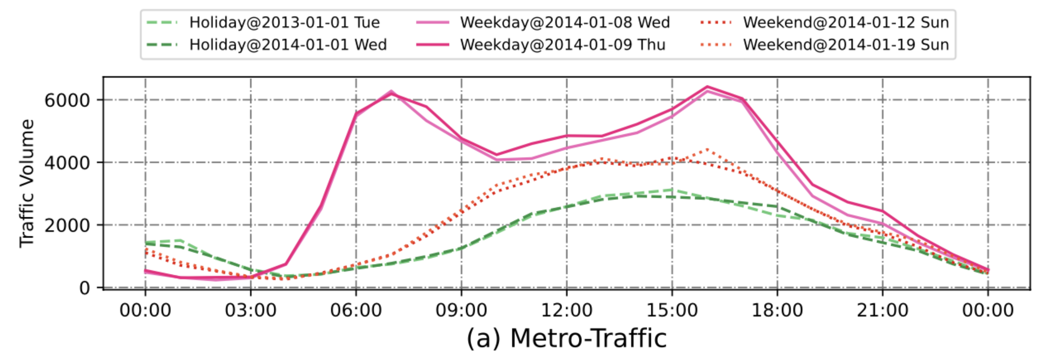
National University of Singapore

ICLR 2025

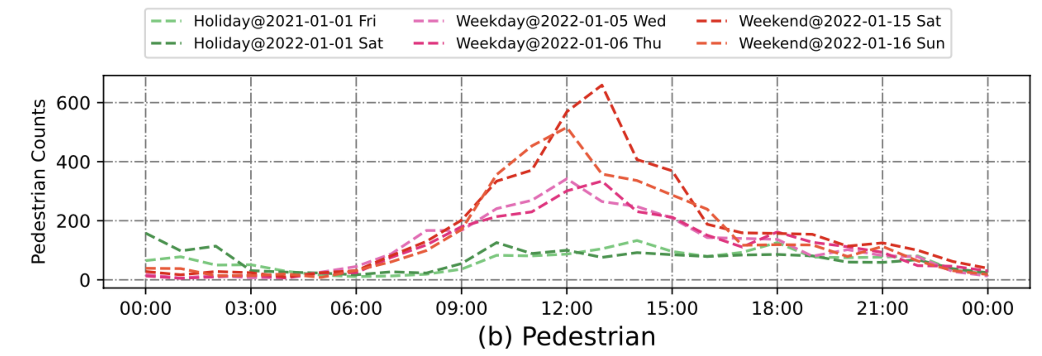
Motivation

Urban time series data exhibits distinct patterns for **high- and low-frequency events:**

- Patterns within each category remain consistent
- Significant differences exist between categories, such as holiday, weekday, weekend patterns



(a) Metro-Traffic



(b) Pedestrian

- **Do neurons associated with low-frequency patterns exist in UTSMs?**
- **If so, how can we enhance the representation learning of these neurons to improve urban time series forecasting**

PN-Train - a pattern neuron-guided training method for urban time series forecasting

- Investigates neurons associated with low-frequency patterns
- Detects them via a perturbation-based detector
- Fine-tunes to enhance their representation

Algorithm

Algorithm 1: Pattern Neuron Guided Training Method

Input: The urban time series model $UTSM$; the training dataset $\mathcal{D}_{\text{train}}$ and validation dataset $\mathcal{D}_{\text{val}}$; the size of the detection sample B, and the size of the fine-tuning sample R; and the learning rates for training, α_1, and fine-tuning, α_2.

Output: The fine-tuned urban time series model $UTSM$

```
// Process fine-tuning samples and training samples
```
$\mathcal{D}_{\text{finetune}} \leftarrow \text{RandomSample}(\{\mathbf{x} \in \mathcal{D}_{\text{train}} \mid \mathbf{x} \text{ is a low-frequency sample}\}, R)$

$\mathcal{D}_{\text{train'}} \leftarrow \mathcal{D}_{\text{train}} \setminus \mathcal{D}_{\text{finetune}}$
```
// Train the urban time series model
```
repeat

  Randomly select a batch of instances $\mathcal{S}$ from $\mathcal{D}_{\text{train'}}$

  Optimize $UTSM$ using AdamW with a learning rate of α_1 on batch $\mathcal{S}$.

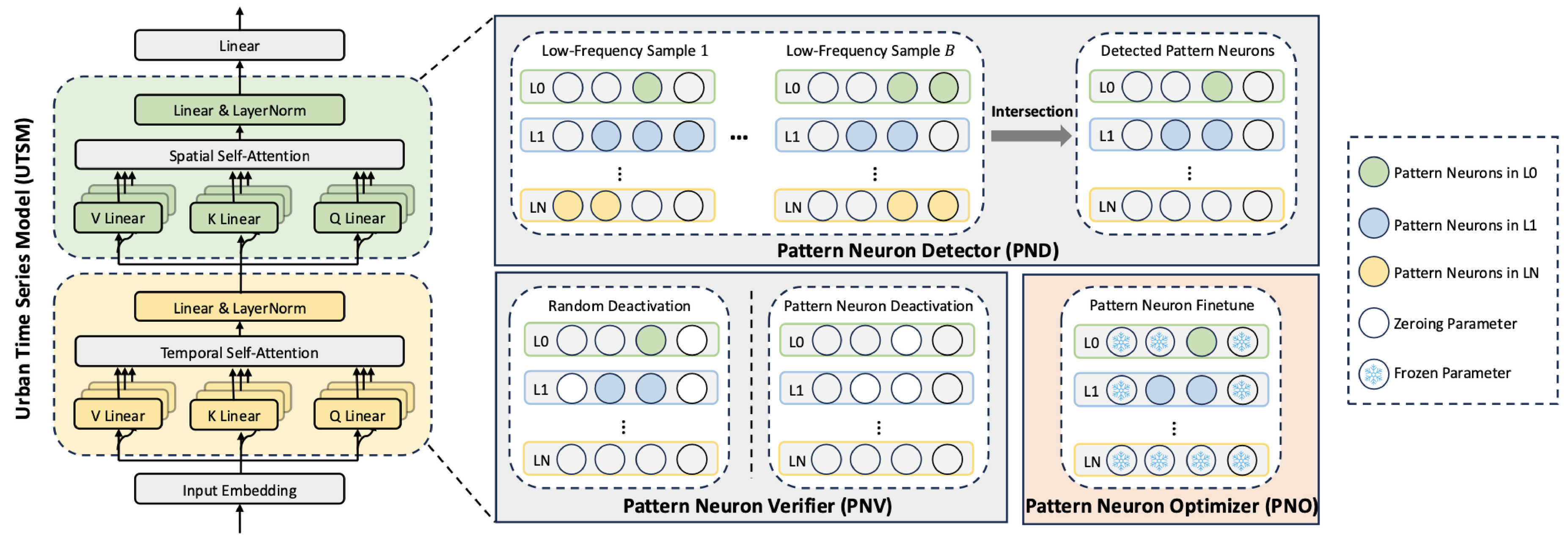
until *met the stopping criteria*;
```
// Select detection samples and detect the pattern neurons
```
$\mathcal{D}_{\text{detect}} \leftarrow \text{RandomSample}(\{\mathbf{x} \in \mathcal{D}_{\text{train}} \mid \mathbf{x} \text{ is a low-frequency sample}\}, B)$

$\mathcal{N}^{p_l} \leftarrow \text{PND}(UTSM, \mathcal{D}_{\text{detect}})$
```
// Fine-tune the detected pattern neurons
```
$\hat{\mathbf{y}} \leftarrow UTSM(\mathcal{D}_{\text{finetune}}, \mathcal{N}^{p})$

$\mathcal{L} \leftarrow \text{MAE}(\hat{\mathbf{y}}, \mathbf{y})$

Optimize pattern neurons $\mathcal{N}^{p}$ using AdamW with a learning rate α_2.
```
// Return the fine-tuned UTSM
```
return $UTSM$

Methodology



Pattern Neuron Detector (PND)

Pattern neurons are those that contribute more significantly to forecasting targets in the perturbation assessment:

$$I(h_i^k|\mathbf{x}^p) = \left\|UTSM(\mathbf{x}^p, \boldsymbol{W}) - UTSM(\mathbf{x}^p, \boldsymbol{W}\backslash \boldsymbol{w}_i^k)\right\|_1$$

We devise an **attribute score** to quantify the influence of the k-th neurons for a specific pattern p given an input x^p with this pattern:

$$\text{Attr}_p(h_i^k \mid \mathbf{x}^p) = \left\|\sum_{s,t} f(\mathbf{x}^p, \boldsymbol{w}_i^k)_{s,t,:}\right\|_1$$

Given a set of samples $\{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_B\}$ corresponding to a specific pattern, the pattern neurons are those whose **attribute scores remain high** across ***all*** B samples:

$$\mathcal{N}^{p_l} = \bigcap_{b=1}^{B} \{n_i^k \mid \text{rank}(\text{Attr}_p(h_i^k \mid \mathbf{x}_b^{p_l})) \leq \epsilon N, \quad \forall i, k\}$$

The detection process can be easily applied to self-attention layers:

$$\text{Attention}(\mathbf{x}) = \text{softmax}\left(Q(\mathbf{x})K(\mathbf{x})^\top/\sqrt{d}\right) V(\mathbf{x}),$$
$$Q(\mathbf{x}) = f(\mathbf{x}, W_Q), \quad K(\mathbf{x}) = f(\mathbf{x}, W_K), \quad V(\mathbf{x}) = f(\mathbf{x}, W_V).$$

Pattern Neuron Verifier (PNV)

Prediction error **increases** significantly without pattern neurons:

$$\sum_{d=1}^{D} \|\mathbf{y}_d - \text{UTSM}(\mathbf{x}_d, \boldsymbol{W}\backslash \boldsymbol{w}_{\text{pattern}})\|_1 \gg \sum_{d=1}^{D} \|\mathbf{y}_d - \text{UTSM}(\mathbf{x}_d, \boldsymbol{W}\backslash \boldsymbol{w}_{\text{random}})\|_1$$

Pattern Neuron Optimizer (PNO)

Fine-tune only the pattern neurons using R samples corresponding to the interested pattern:

$$\mathcal{L}(\hat{\mathbf{y}}, \mathbf{y} \mid \theta_{\boldsymbol{w}_{\text{pattern}}}) = \frac{1}{R}\sum_{r=1}^{R} \|\hat{\mathbf{y}}_r - \mathbf{y}_r\|_1$$

Experiments & Model Analysis

Existence of Pattern neurons in UTSMs:

- Deactivating pattern neurons yields a larger performance drop than random deactivation
- Pattern neurons make up <10% of all neurons

Model	Holiday			Non-Holiday			Overall		
Metro-Traffic (Deactivate ratio 7.76%)	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE
Original	446.04	846.75	16.36%	208.84	339.77	6.19%	220.00	379.14	6.58%
D-Random	492.29	833.99	18.05%	263.34	380.46	7.80%	274.11	413.12	8.19%
D-PN	663.46†	1046.40†	24.33%†	474.02†	586.01†	14.04%†	482.93†	615.44†	14.43%†
Pedestrian (Deactivate ratio 9.77%)	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE
Original	109.01	259.79	29.31%	78.82	196.39	21.70%	80.45	200.33	22.12%
D-Random	116.99	264.03	31.46%	91.75	210.79	25.26%	93.12	214.01	25.60%
D-PN	194.53†	370.80†	52.31%†	174.92†	321.45†	48.15%†	175.98†	324.31†	48.38%†

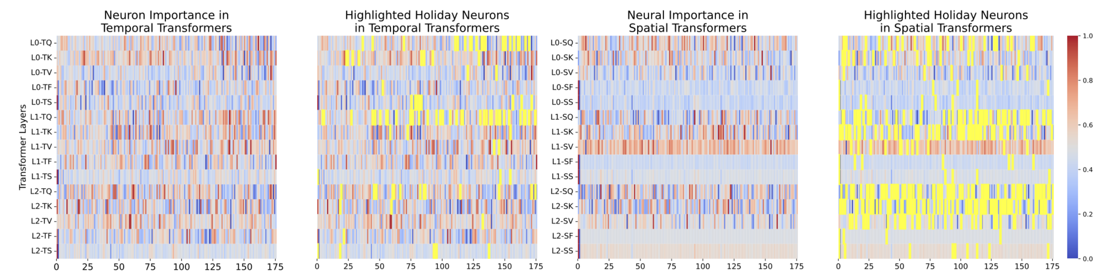
Enhance the representation of pattern neurons associated with low-frequency events:

- Improves performance on low-frequency events
- Also benefits high-frequency events

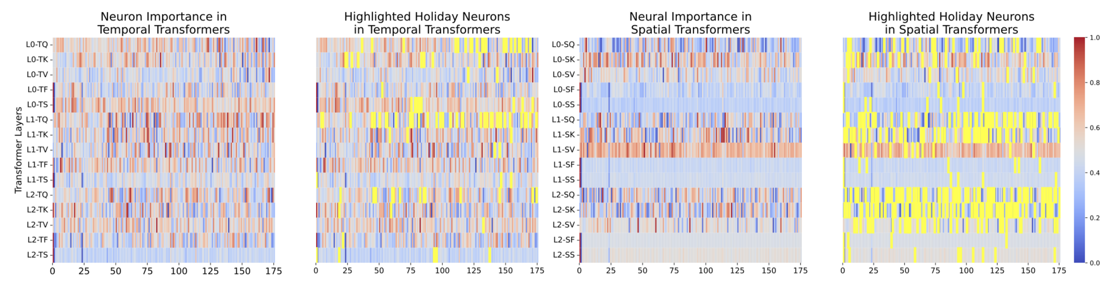
Dataset	Method	Holiday MAE	Holiday RMSE	Holiday WMAPE	Non-Holiday MAE	Non-Holiday RMSE	Non-Holiday WMAPE	Overall MAE	Overall RMSE	Overall WMAPE
Metro-Traffic	HA	156.64	325.58	42.12%	128.28	270.90	35.31%	129.82	274.14	35.69%
	STGCN (Yu et al., 2018)	460.97	739.63	16.91%	289.85	501.33	8.58%	297.90	515.02	8.90%
	GWNet (Wu et al., 2019)	534.76	832.13	19.61%	347.50	582.90	10.29%	356.31	596.97	10.64%
	AGCRN (Bai et al., 2020)	453.23	738.60	16.62%	280.41	496.75	8.31%	288.54	510.71	8.62%
	STID (Shao et al., 2022)	586.90	1031.50	21.52%	216.09	346.81	6.40%	233.54	405.81	6.98%
	PM-MemNet (Lee et al., 2022)	554.33	916.78	20.32%	375.88	666.90	11.13%	384.28	680.71	11.49%
	STWA (Cirstea et al., 2022)	521.02	820.57	19.11%	355.63	619.28	10.53%	364.36	630.61	10.89%
	STAEformer (Liu et al., 2023)	443.23	821.42	16.25%	210.41	343.01	6.23%	221.37	379.29	6.62%
	TESTAM (Lee & Ko, 2024)	486.89	857.99	17.86%	335.05	555.09	9.92%	342.19	572.94	10.22%
	PN-Train *	446.04	846.75	16.35%	208.84	339.77	6.19%	220.00	379.14	6.58%
	PN-Train	**430.40**	**816.50**	**15.78%**	**203.62**	**332.15**	**6.03%**	**214.29**	**369.46**	**6.40%**
Pedestrian	HA	208.49	388.17	64.48%	255.12	471.08	83.46%	253.24	468.01	82.69%
	STGCN (Yu et al., 2018)	120.75	258.53	32.47%	101.61	214.32	27.97%	102.65	216.95	28.22%
	GWNet (Wu et al., 2019)	119.77	267.48	32.21%	113.69	245.87	31.30%	114.02	247.09	31.35%
	AGCRN (Bai et al., 2020)	118.48	267.32	31.86%	108.22	245.55	29.79%	108.78	246.78	29.91%
	STID (Shao et al., 2022)	116.42	263.79	31.31%	85.32	206.36	23.49%	87.00	209.87	23.92%
	PM-MemNet (Lee et al., 2022)	117.44	265.09	31.58%	112.18	246.64	30.88%	112.48	247.69	30.92%
	STWA (Cirstea et al., 2022)	114.18	261.03	30.70%	106.62	234.88	29.35%	106.90	236.13	29.39%
	STAEformer (Liu et al., 2023)	115.24	273.64	30.99%	82.23	202.73	22.64%	84.02	207.19	23.10%
	TESTAM (Lee & Ko, 2024)	**103.79**	**257.10**	**27.91%**	94.04	219.46	25.89%	94.57	221.67	26.00%
	PN-Train *	109.01	259.79	29.31%	78.82	196.39	21.70%	80.45	200.33	22.12%
	PN-Train	106.11	253.86	28.54%	**78.35**	**194.72**	**21.57%**	**79.85**	**198.38**	**21.95%**

Pattern neurons appear in similar positions & more concentrated on **query and key** components

Detected Sample #3

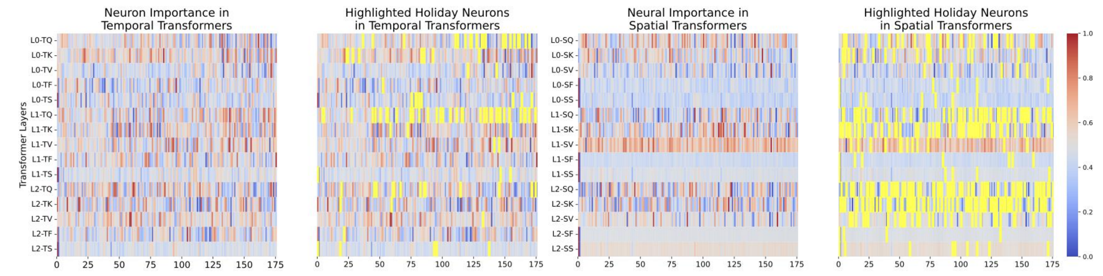


Detected Sample #13

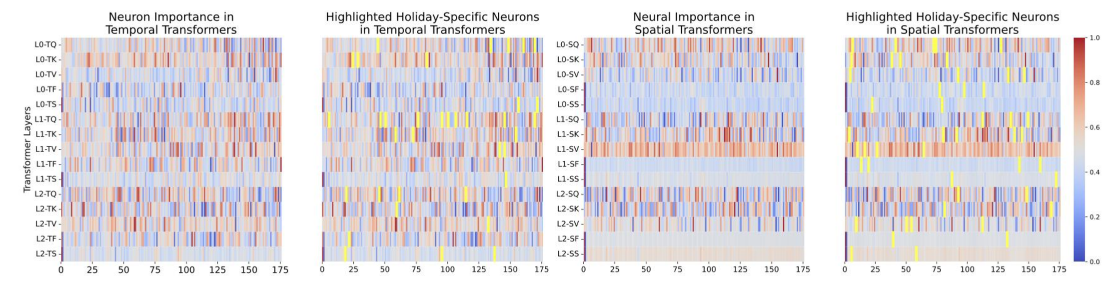


Pattern neurons include both pattern-specific and general neurons

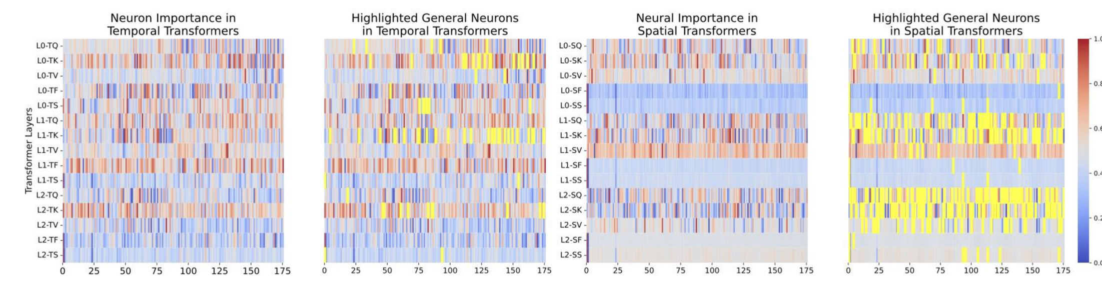
Holiday Neurons



Holiday-Specific Neurons



General Neurons



PN-Train can be generalized to multiple low-frequency patterns

Dataset	Method	Holiday MAE	Holiday RMSE	Holiday WMAPE	Parade MAE	Parade RMSE	Parade WMAPE	Others MAE	Others RMSE	Others WMAPE	Overall MAE	Overall RMSE	Overall WMAPE
GBAP	HA	55.63	76.11	14.88%	48.09	62.53	13.03%	52.13	77.60	16.88%	52.18	77.55	16.83%
	STGCN	38.74	54.67	10.36%	27.52	37.28	7.46%	29.58	45.18	9.57%	29.72	45.33	9.59%
	GWNet	42.75	61.67	11.44%	27.69	38.10	7.50%	31.04	47.88	10.05%	31.22	48.12	10.07%
	AGCRN	38.61	54.26	10.33%	26.18	35.19	7.10%	28.81	44.17	9.33%	28.96	44.34	9.34%
	STID	38.25	54.29	10.23%	26.28	34.84	7.12%	25.43	39.85	8.23%	25.64	40.11	8.27%
	PM-MemNet	34.97	49.23	9.36%	26.41	35.82	7.16%	29.00	43.85	9.39%	29.09	43.93	9.39%
	STWA	40.53	58.04	10.84%	**25.66**	**35.22**	**6.95%**	30.75	46.90	9.96%	30.86	47.01	9.96%
	STAEformer	32.39	45.65	8.67%	27.08	37.20	7.68%	25.33	39.92	8.20%	25.45	40.01	8.21%
	TESTAM	33.86	48.04	9.06%	31.98	45.34	8.66%	28.08	43.52	9.09%	28.18	43.60	9.09%
	PN-Train	**32.25**	**45.44**	**8.62%**	26.73	36.68	7.24%	**25.25**	**39.81**	**8.18%**	**25.37**	**39.90**	**8.18%**