	Method	ETTH2			ECL			Avg
		1	24	48	1	24	48	
Baseline	FSNet	0.466	0.687	0.846	3.143	6.051	7.034	3.038
Ensembling baselines	Average	0.381	0.607	0.595	2.458	2.833	3.309	1.697
	Gating [2]	0.476	0.678	0.782	2.474	2.181	2.301	1.482
	MOE [1]	0.488	0.565	1.238	3.312	3.086	2.497	1.864
	AdaRaker[3]	0.489	0.723	0.934	2.742	2.796	2.842	1.754
	Learn ++. NSE [5]	0.383	0.667	0.672	2.623	2.324	2.412	1.514
Ours	**OneNet**	**0.380**	**0.532**	**0.609**	**2.351**	**2.074**	**2.201**	**1.358**

Table 1: Comparision to advanced ensemble learning baselines.

Params	**ETTH2**			**ECL**		
	1	**24**	**48**	**1**	**24**	**48**
FSNet	1018045	1069726	1123654	1138935	3508878	5981862
OneNet	2037108	2096172	2157804	2157998	4535324	7016012
OneNet-	1105505	1118040	1131120	1105505	1118040	1131120
Inference speed items/s	**ETTH2**			**ECL**		
	1	**24**	**48**	**1**	**24**	**48**
FSNet	7.89	7.85	7.72	8.37	7.71	7.32
OneNet	6.4	5.78	5.27	6.73	6.65	6.3
OneNet-	30.17	25.76	23.24	27.16	23.4	24.2

Table 2: Number of parameters and inference time for different models.

Algorithm 1 Training and inference algorithm of OneNet

Input: Historical multivariate time series $\mathbf{x} \in \mathbb{R}^{M \times L}$ with M variables and length L, the forecast target $\mathbf{y} \in \mathbb{R}^{M \times H}$, where we omit the variable index and time step index for simplicity.
Initialize a cross-time forecaster f_1, a cross-variable forecaster f_2 with corresponding prediction head, long-term weight $\mathbf{w} = [0.5, 0.5]$, short term learning block $f_{rl} : \mathbb{R}^{H \times M \times 3} \rightarrow \mathbb{R}^2$, and step size η for long-term weight updating.
Get prediction results from two forecasters.
$\tilde{\mathbf{y}}_1 \in \mathbb{R}^{M \times H} = f_1(\mathbf{x})$. *// Prediction result from the cross-time forecaster.*
$\tilde{\mathbf{y}}_2 \in \mathbb{R}^{M \times H} = f_2(\mathbf{x})$. *// Prediction result from the cross-variable forecaster.*
Get combination weight from the OCP block.
$\mathbf{b} = f_{rl}\left(w_1 \tilde{\mathbf{y}}_1 \otimes w_2 \tilde{\mathbf{y}}_2 \otimes \mathbf{y}\right)$ *// Calculate the short term weight.*
$\tilde{w}_i = (w_i + b_i) / \left(\sum_{i=1}^{d} (w_i + b_i)\right)$ *// Calculate the normalized weight.*
$\tilde{\mathbf{y}} = w_1 * \tilde{\mathbf{y}}_1 + w_2 * \tilde{\mathbf{y}}_2$. *// The final prediction result.*
Update the long/short term weight.
$w_i = w_i \exp(-\eta \parallel \tilde{\mathbf{y}}_i - \mathbf{y} \parallel^2) / \left(\sum_{i=1}^{2} w_i \exp(-\eta \parallel \tilde{\mathbf{y}}_i - \mathbf{y} \parallel^2)\right)$
$f_{rl} \leftarrow \text{Adam}\left(f_{rl}, \mathcal{L}(w_1 * \tilde{\mathbf{y}_1} + w_2 * \tilde{\mathbf{y}}_2, \mathbf{y})\right)$ *//Parameters such as learning rate are omitted.*
Update the two forecasters.
$f_1 \leftarrow \text{Adam}\left(f_1, \mathcal{L}(\tilde{\mathbf{y}}_1, \mathbf{y})\right)$, $f_2 \leftarrow \text{Adam}\left(f_2, \mathcal{L}(\tilde{\mathbf{y}}_2, \mathbf{y})\right)$

References