Conditional Temporal Neural Processes with Covariance Loss

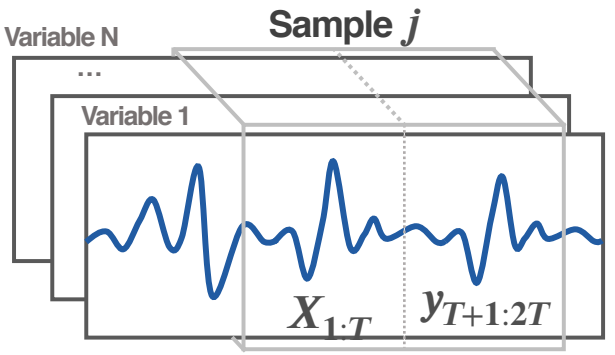
Boseon Yoo[1], Jiwoo Lee[2], Janghoon Joo[2], Seijun Chung[1], Soyeon Kim[1] and Jaesik Choi[1,3]

[1]Graduate School of AI, KAIST, Republic of Korea.

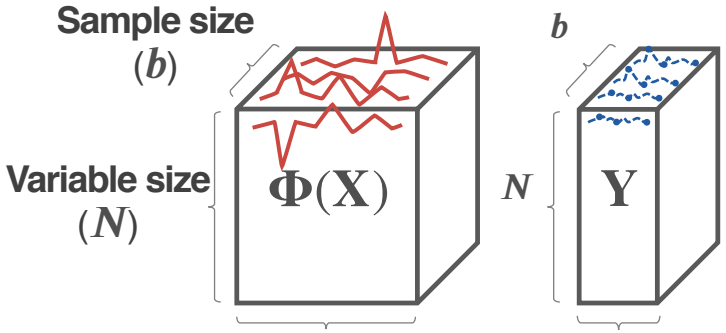
[2]Department of Computer Science and Engineering, UNIST, Republic of Korea.

[3]Ineeji Incorporation, Republic of Korea.

Key Idea



A Temporal DNN

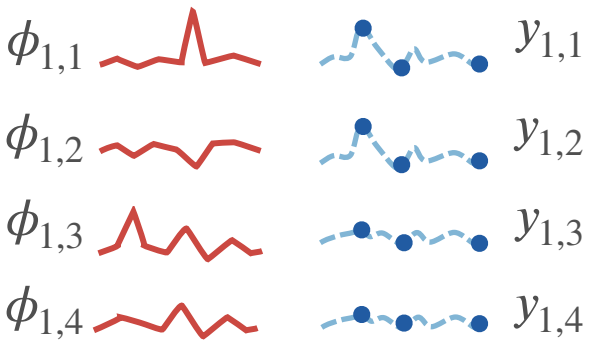


Basis function feature size (F')

Output time step size (T) × Output feature size (S)

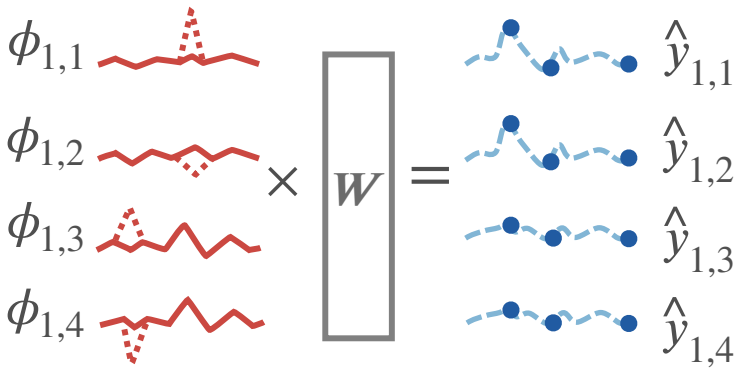
w/o Cov

$\phi_{variable=i,\ sample=j}^{y_{i,j}}$

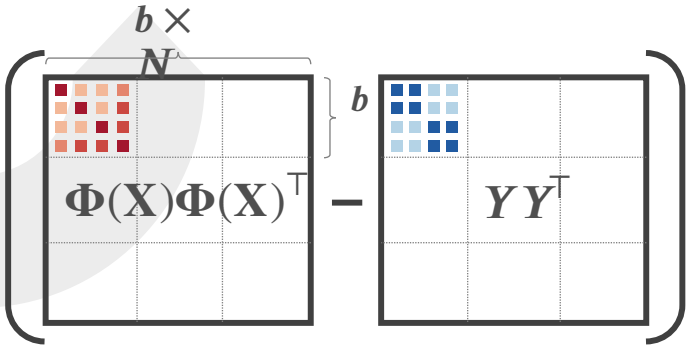


w/ Cov

$\phi_{i,j}$



$$\underset{\Phi}{\operatorname{argmin}}\ \text{(Original loss function)} + \left(\Phi(X)\Phi(X)^{\top} - YY^{\top} \right)$$



CNPs vs Covariance Loss

$$E[\Phi(X)^T\Phi(X)] \rightarrow \quad E[Y^TY]$$

$$-\frac{1}{2}\mathbf{y}^\top\mathbf{K}^{-1}\mathbf{y}-\frac{1}{2}\log|\mathbf{K}|-\frac{n}{2}\log(2\pi)$$

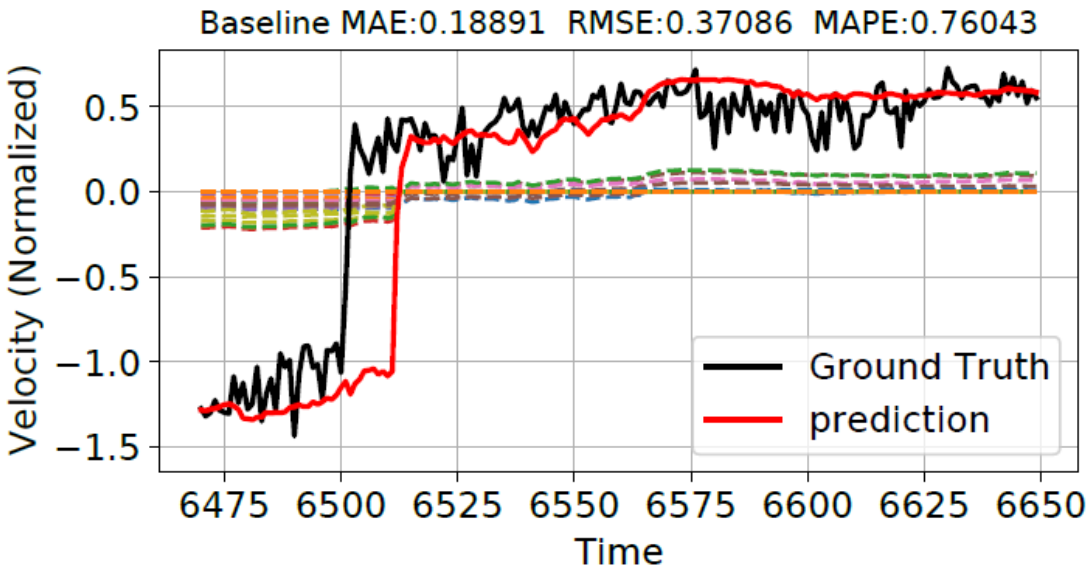
$$\frac{1}{n}(\mathbf{Y}-\hat{\mathbf{Y}})^2+\lambda\left(\frac{1}{n^2}(\widetilde{\Sigma}_{\mathbf{Y}}-\sigma^2\Phi\Phi^\top)^2\right)$$

$$K = \Phi\Phi T$$

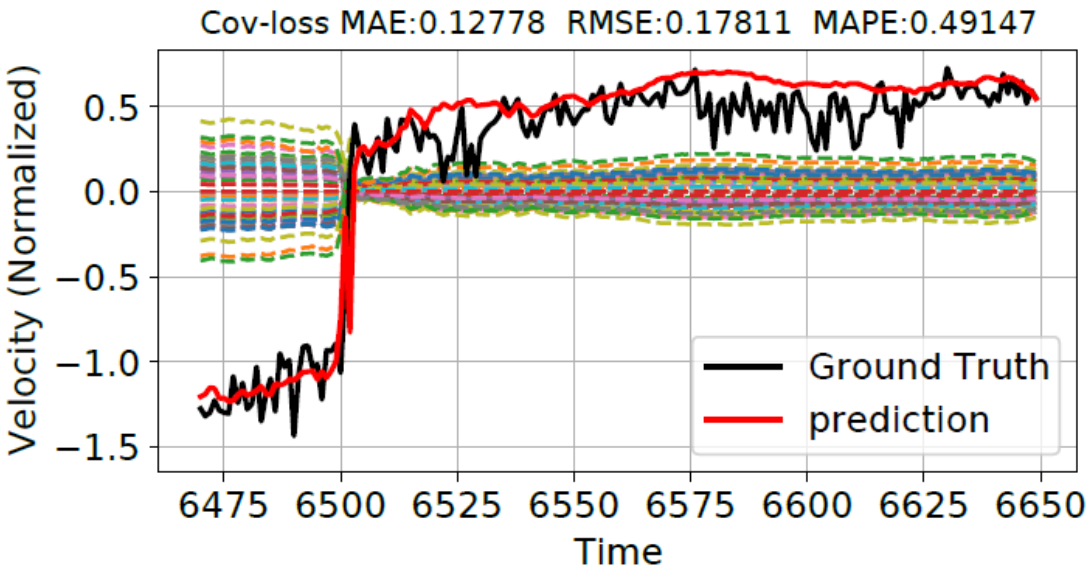
CNPs

Covariance Loss

Result - Regression



STGCN

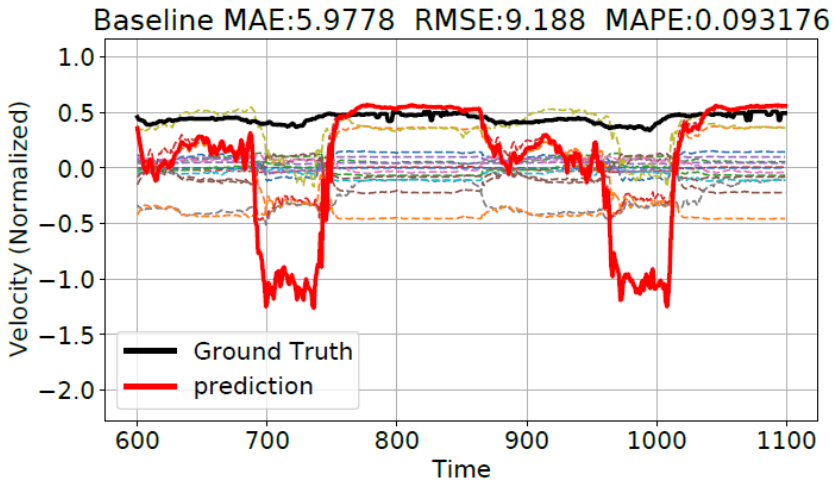


STGCN-Cov

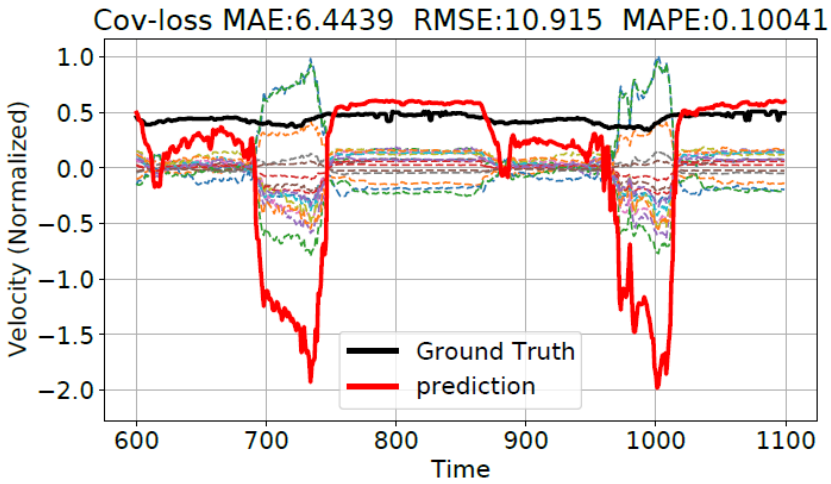
Result - Regression

Table 3. Performance comparison for METR-LA and PeMS-BAY datasets.

Dataset	METR-LA			PeMS-BAY		
Model	MAE (15/30/60 min)	MAPE (15/30/60 min)	RMSE (15/30/60 min)	MAE (15/30/60 min)	MAPE (15/30/60 min)	RMSE (15/30/60 min)
HA	4.16	13.0	7.8	2.88	6.84	5.59
ARIMA	3.99/5.15/6.90	9.6/12.7/17.4	8.1/10.5/13.2	1.62/2.33/3.38	3.5/5.4/8.3	3.30/4.76/6.50
FC-LSTM	3.44/3.77/4.37	9.60/10.9/13.2	6.30/7.23/8.69	2.05/2.20/2.37	4.8/5.2/5.7	4.19/4.55/4.69
WaveNet	2.99/3.59/4.45	8.04/10.3/13.6	5.89/7.28/8.93	1.39/1.83/2.35	2.91/4.16/5.87	3.01/4.21/5.43
DCRNN	2.77/3.15/3.60	7.30/8.80/10.5	5.38/6.45/7.60	1.38/1.74/2.07	2.9/3.9/4.9	2.95/3.97/4.74
STGCN	2.88/3.47/4.59	7.6/9.6/12.7	5.74/7.24/9.40	1.46/2.00/2.67	2.9/4.1/5.4	3.01/4.31/5.73
GWNET	**2.69/3.07/3.53**	6.90/**8.23/9.8**	5.15/6.22/7.37	**1.30**/1.63/1.95	2.7/3.7/4.6	2.74/3.70/4.52
GWNET-Cov	**2.69/3.07/3.53**	**6.83**/8.26/9.85	**5.14/6.17/7.27**	**1.30/1.62/1.91**	**2.69/3.59/4.47**	**2.73/3.67/4.40**



STGCN



STGCN-Cov

Result - Classification

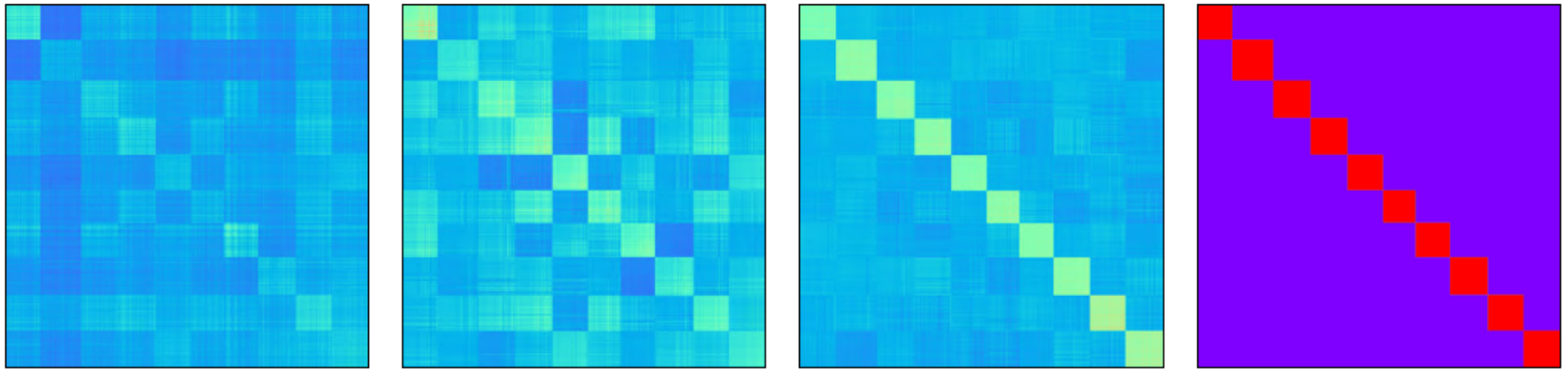


Figure 3. Covariance matrix of **X**, basis function of DNN, basis function of DNN-Cov, and one-hot-encoded label. Brighter color indicates higher covariance value. (b) is similar to (a) and some basis functions have high covariance with basis function that belongs to other classes. In contrast, (c) is similar to (d) and basis functions in each class are mutually exclusive.

Result - Classification

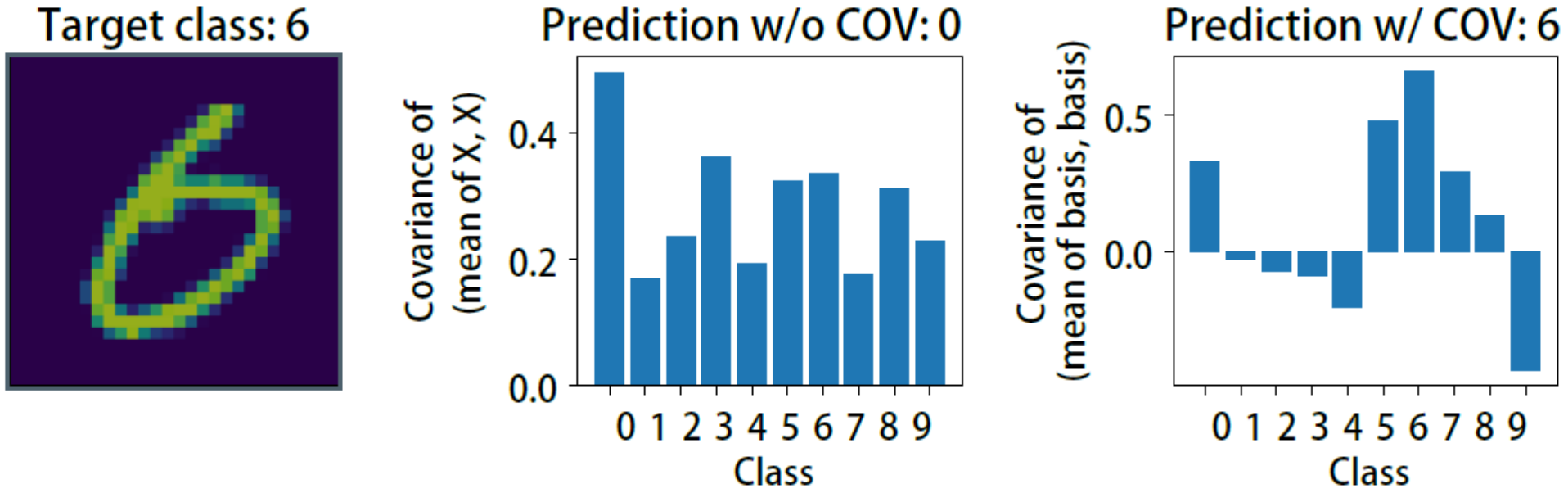


Figure 4. An ambiguous sample in MNIST dataset. The sample (left) is similar to samples of class '0' and the prediction with mean activation of input variable is incorrect (middle). In contrast, the optimization with Covariance Loss shows correct prediction for the ambiguous sample (right).