2025 12 24
발표 자료

광운대학교 로봇학과
FAIR Lab

김한서

KWANGWOON UNIVERSITY

이번 주 진행사항

- ElasTST
  - 논문 리뷰
  - 실험 세팅
  - 실험 결과 및 시각화
  - 결과 정리

ElasTST

ElasTST: Towards Robust Varied-Horizon Forecasting with Elastic Time-Series Transformer

Jiawen Zhang[*]
DSA, HKUST(GZ)
Guangzhou, China
jiawe.zh@gmail.com

Shun Zheng[†]
Microsoft Research Asia
Beijing, China
shun.zheng@microsoft.com

Xumeng Wen
Microsoft Research Asia
Beijing, China
xumengwen@microsoft.com

Xiaofang Zhou
CSE, HKUST
Hong Kong SAR, China
zxf@ust.hk

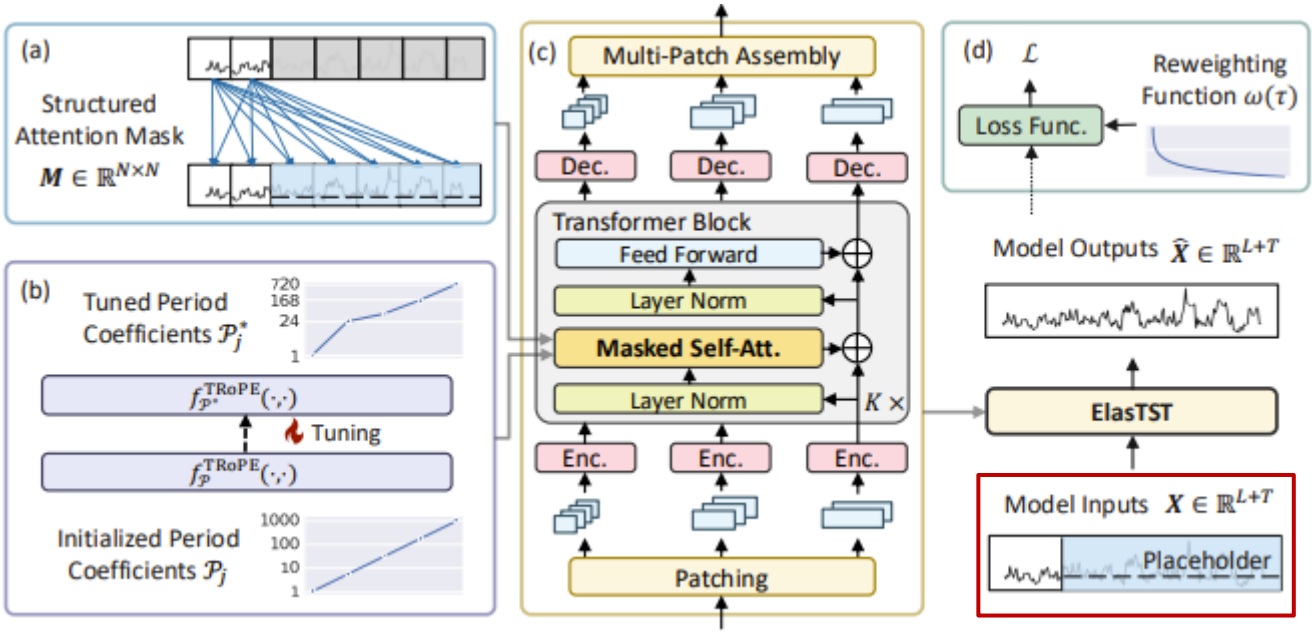
Jiang Bian
Microsoft Research Asia
Beijing, China
jiang.bian@microsoft.com

Jia Li[†]
DSA, HKUST(GZ)
Guangzhou, China
jialee@ust.hk

- arXiv 등록일: 2024-11-04
- 인용 수: 8회(Google Scholar, 2025-12-19)
- Published at NeurIPS 2024
- 기존 시계열 모델들은 특정 예측 길이에 맞춰 학습되는데, 학습 범위를 벗어난 길이를 예측하게 되면 성능이 크게 저하되는 문제가 발생, 이를 해결하기 위해 한 번의 학습으로 다양한 예측 길이에 대해 일관된 성능을 보장하는 ElasTST를 제안

ElasTST

모델 구조

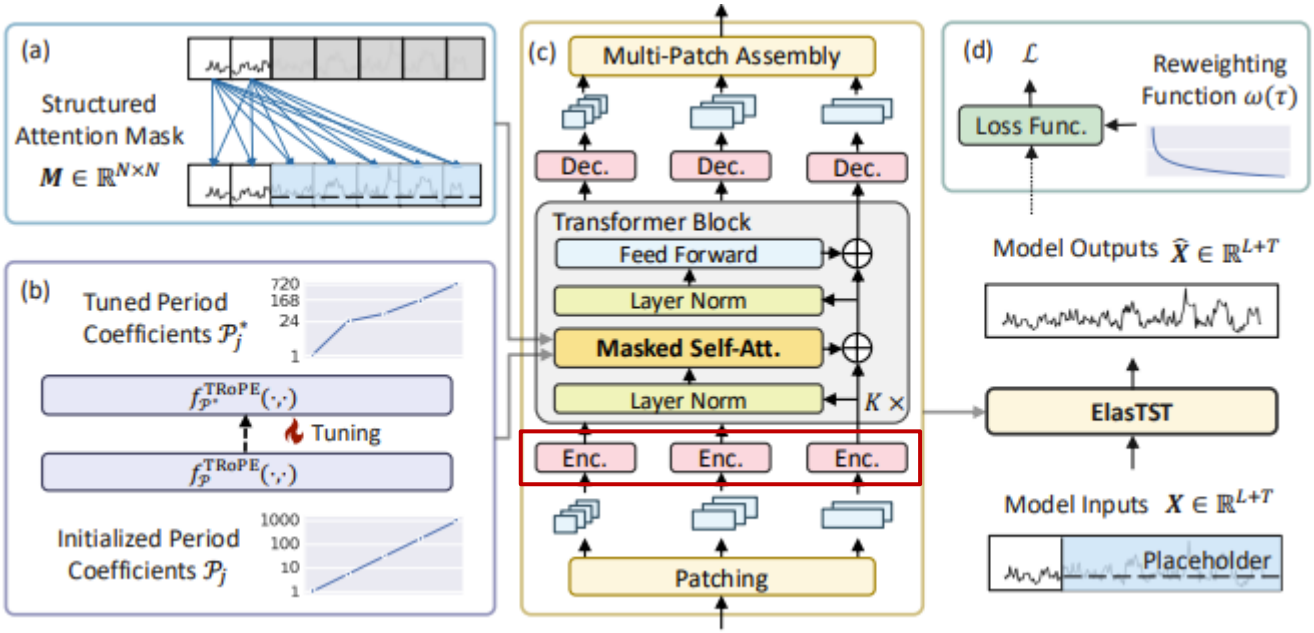


- Model Inputs
  - 과거 시계열 데이터와 예측해야 할 미래 시점인 Placeholder를 Concat하여 입력으로 사용 $X \in \mathbb{R}^{L+T}$
- Multi-Scale Patching
  - 입력 X를 서로 다른 길이의 겹치지 않는 패치 $X^p \in \mathbb{R}^{N+P}$로 분할

L : 시퀀스 길이
T : 예측 길이
N : 패치 개수
P : 패치 길이

ElasTST

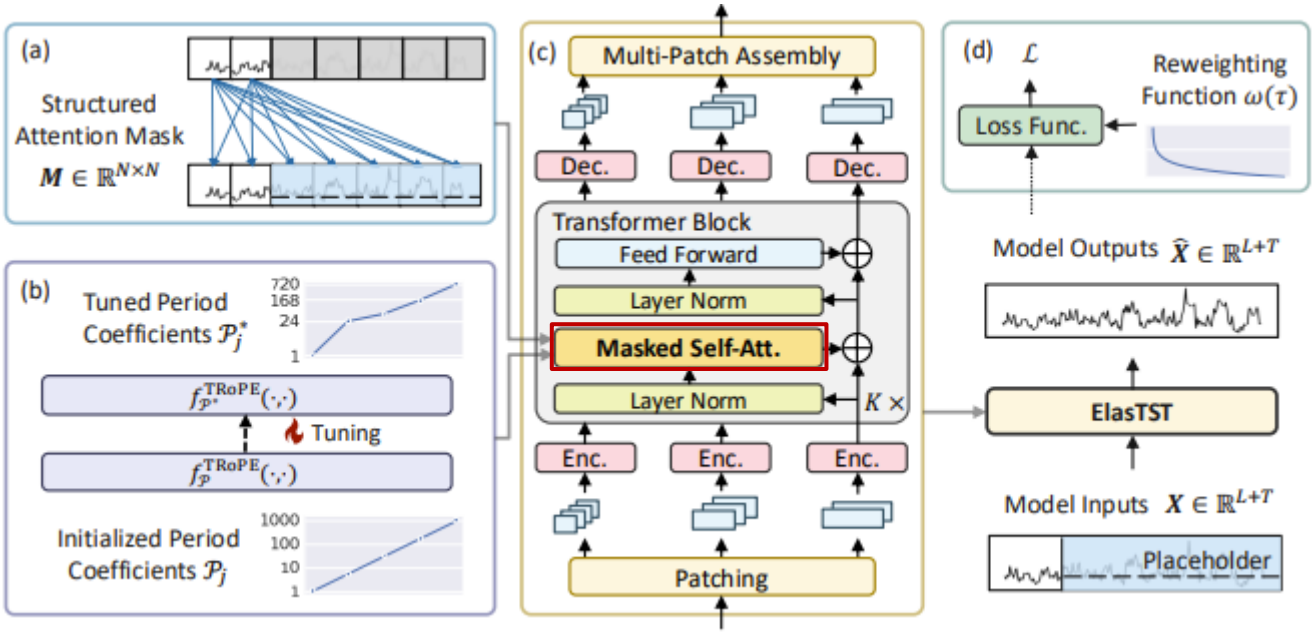
모델 구조



- Encoder
  - 분할된 각 패치는 인코더를 통해 D차원의 임베딩 벡터로 변환 $H = \mathrm{Enc}(X^p),\ H \in \mathbb{R}^{N+D}$
  - 동일한 D차원으로 변환해 Transformer Block에서 한 번에 연산
- Layer Norm
  - 각 패치별로 D개의 특징 값들에 대해 동시에 정규화하여 학습 안정화 및 성능 향상

ElasTST

모델 구조



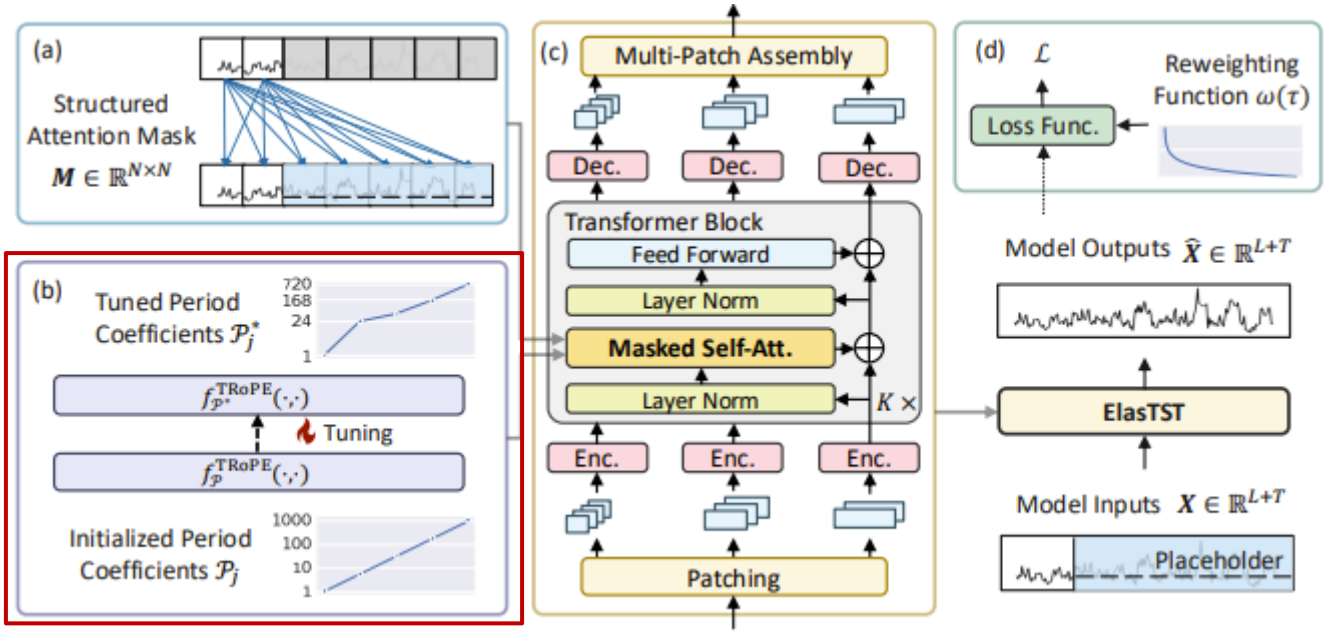
- Masked Self-Attention
  - 정규화된 벡터에 가중치 행렬 W^q, W^k, W^v을 곱해 Q, K, V를 생성하고, TRoPE가 적용된 Q, K의 내적을 통해 Attention Score 계산

ElasTST

FAIR Lab

KWANGWOON UNIVERSITY

모델 구조



- TRoPE (Tunable Rotary Position Embedding)
  - Q와 K 벡터의 D개의 숫자들을 각각 가져와 2개씩 짝지어 $D/2$개의 쌍으로 묶고,
    이후 각각의 쌍은 독립적인 2차원 평면 위의 좌표가 되어 주기를 개별적으로 관리
  - 각 평면마다 데이터의 순서 n과 학습된 주기를 조합한 각도만큼 벡터를 회전 $\theta_j = \frac{2\pi}{P_j} \times n$
  - 회전된 Q, K의 내적을 통해 두 시점 간의 상대적인 각도 차이인 위치 정보를 Attention score에 주입

$$\begin{pmatrix} x'2j-1 \\ x'2j \end{pmatrix} = \begin{pmatrix} \cos\theta_j & -\sin\theta_j \\ \sin\theta_j & \cos\theta_j \end{pmatrix} \begin{pmatrix} x_{2j-1} \\ x_{2j} \end{pmatrix}$$

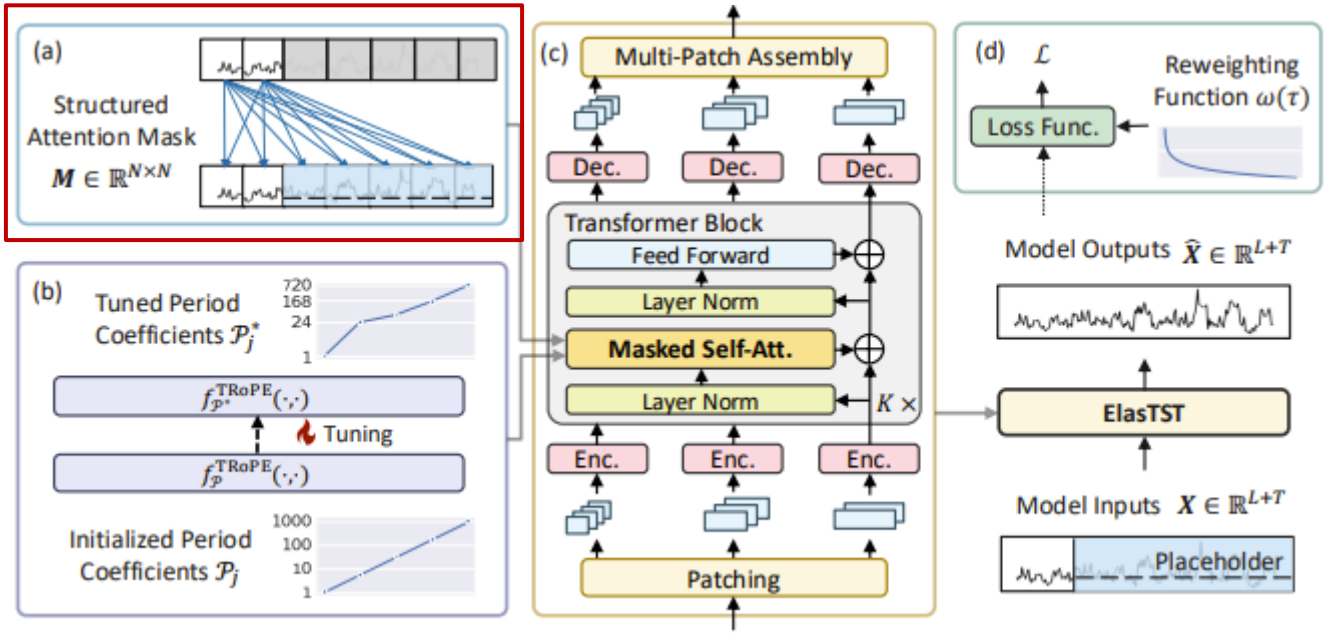
n : 데이터 순서
θ_j: 최종 각도
P_j: 주기
2π: 360도

ElasTST

FAIR Lab

KWANGWOON UNIVERSITY

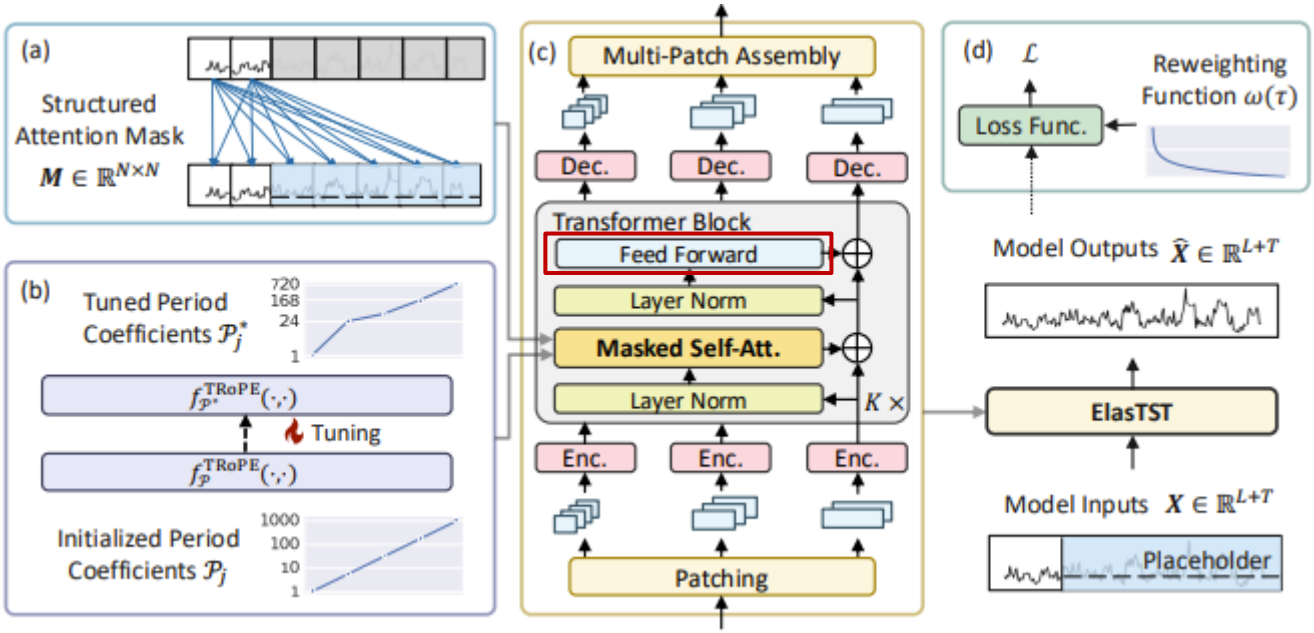
모델 구조



- Structured Attention Mask
  - Attention Score에 Structured Attention Mask 적용
  - 모든 토큰이 위쪽의 과거 데이터 패치만 참조하여 예측 길이가 바뀌더라도 예측 결과가 변하지 않는 예측 기간 불변성(Horizon-invariant property)을 보장

ElasTST

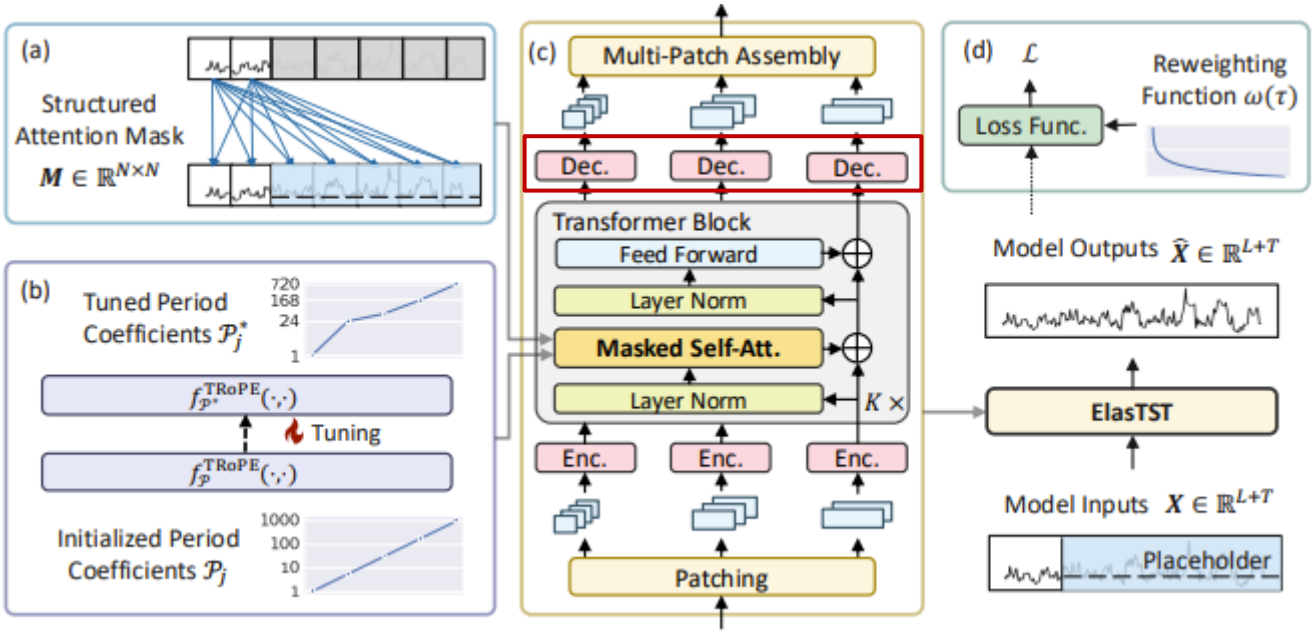
모델 구조



- Feed Forward
  - 비선형성 추가, 또한 각 패치마다 동일한 가중치를 공유하며 독립적으로 적용

ElasTST

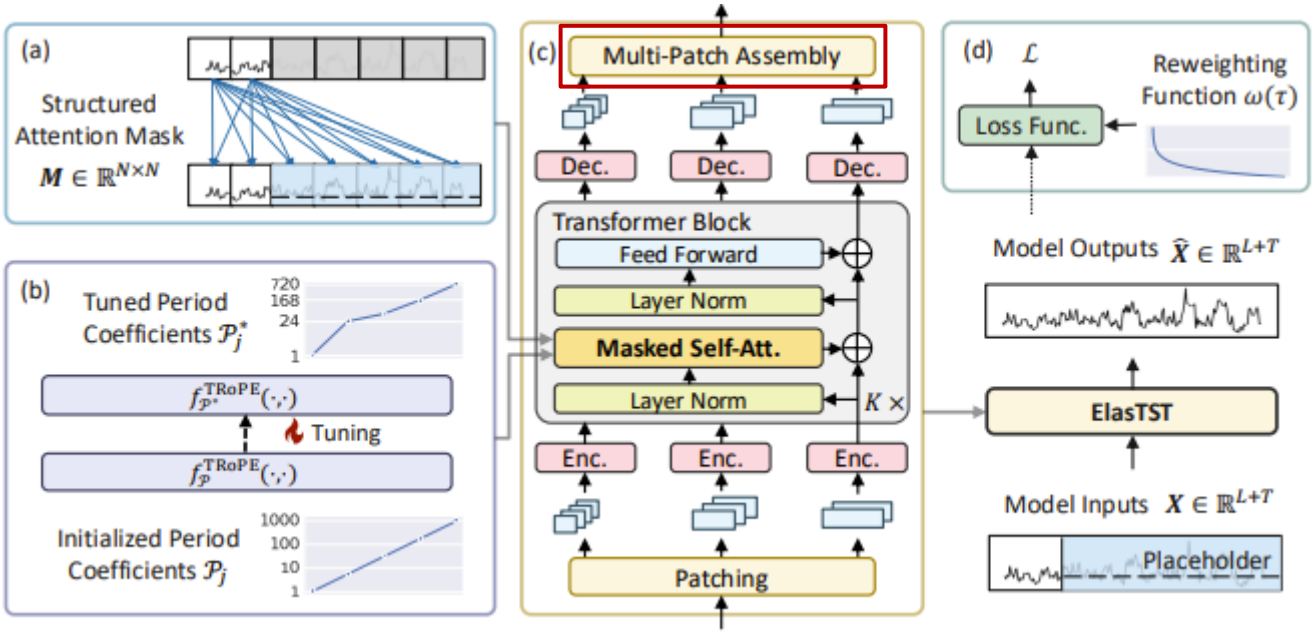
모델 구조



- Decoder
  - Transformer Block에서 나온 결과물인 D차원 특징 벡터를 입력으로 사용하여 원래 패치 길이인 P차원으로 다시 변환

ElasTST

모델 구조



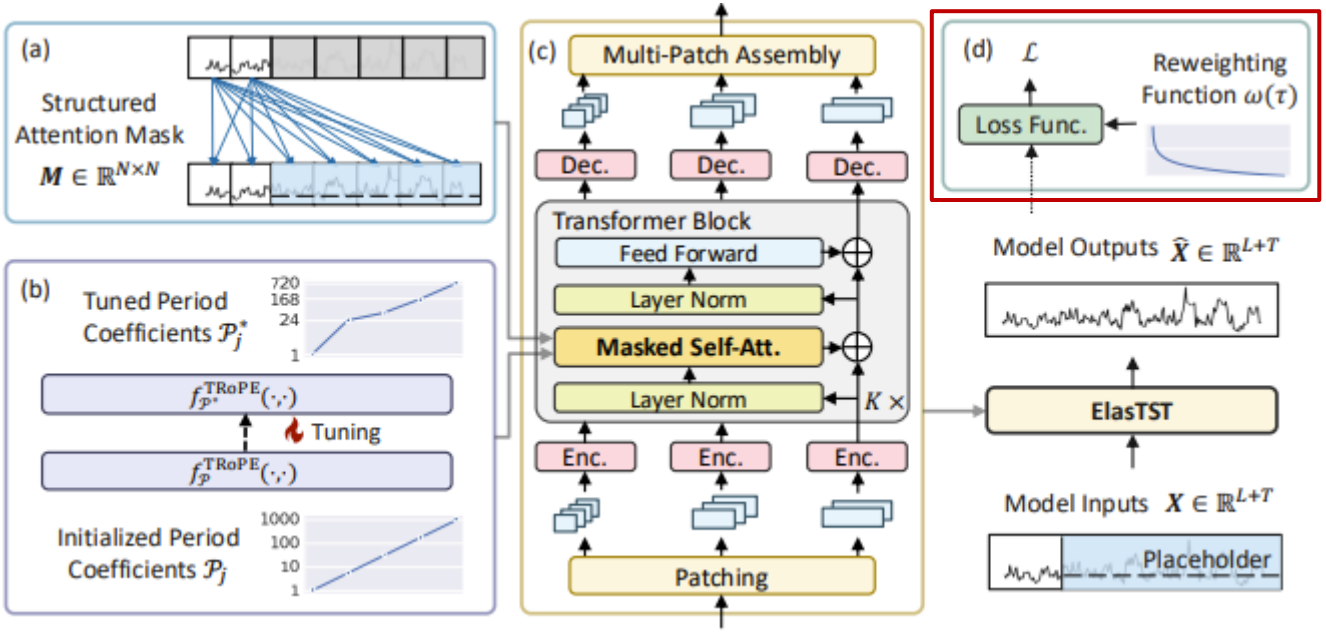
- Multi-Patch Assembly
  - P차원으로 복원된 패치들을 이어 붙여 전체 예측 길이만큼의 시퀀스를 만듦
    이후에 서로 다른 시퀀스를 모두 더한 뒤 시퀀스 개수로 나눠 평균값 출력, 평균값이 최종 예측 결과

ElasTST

FAIR Lab

KWANGWOON UNIVERSITY

모델 구조



- Training Horizon Reweighting
  - 매번 예측 길이를 다르게 샘플링하는 대신, 항상 최대치인 T_{max}까지 학습하도록 하고, 손실 함수를 계산할 때 $[1, T_{max}]$ 범위의 각 시점 τ에 대해 Reweighting을 적용해 매번 길이를 다르게 샘플링하여 학습한 것과 같은 효과를 냄

ElasTST

FAIR Lab

KWANGWOON UNIVERSITY

주요 모델 성능 비교

	pred len	ElasTST NMAE	ElasTST NRMSE	iTransformer NMAE	iTransformer NRMSE	PatchTST NMAE	PatchTST NRMSE	DLinear NMAE	DLinear NRMSE	Autoformer NMAE	Autoformer NRMSE
ETTm1	96	0.273$_{.000}$	**0.488$_{.000}$**	**0.271$_{.000}$**	0.568$_{.000}$	0.272$_{.001}$	0.565$_{.001}$	0.282$_{.002}$	0.573$_{.001}$	0.388$_{.001}$	0.711$_{.003}$
	192	**0.289$_{.000}$**	**0.520$_{.000}$**	0.301$_{.000}$	0.614$_{.000}$	0.295$_{.001}$	0.602$_{.005}$	0.309$_{.004}$	0.617$_{.003}$	0.442$_{.001}$	0.820$_{.003}$
	336	**0.314$_{.000}$**	**0.575$_{.000}$**	0.333$_{.000}$	0.668$_{.000}$	0.323$_{.001}$	0.645$_{.003}$	0.338$_{.008}$	0.654$_{.007}$	0.429$_{.000}$	0.774$_{.001}$
	720	**0.346$_{.000}$**	**0.645$_{.000}$**	0.376$_{.000}$	0.741$_{.000}$	0.353$_{.001}$	0.700$_{.005}$	0.387$_{.006}$	0.737$_{.005}$	0.440$_{.000}$	0.793$_{.000}$
ETTm2	96	0.150$_{.000}$	0.227$_{.000}$	0.137$_{.000}$	0.227$_{.000}$	**0.132$_{.001}$**	**0.220$_{.002}$**	0.138$_{.000}$	0.226$_{.000}$	0.158$_{.000}$	0.254$_{.000}$
	192	0.174$_{.000}$	0.264$_{.000}$	0.161$_{.000}$	0.266$_{.000}$	**0.157$_{.001}$**	**0.259$_{.002}$**	0.163$_{.003}$	0.264$_{.001}$	0.175$_{.000}$	0.283$_{.000}$
	336	0.191$_{.000}$	0.289$_{.000}$	0.180$_{.000}$	0.293$_{.000}$	**0.176$_{.000}$**	**0.286$_{.000}$**	0.188$_{.001}$	0.291$_{.002}$	0.191$_{.000}$	0.307$_{.000}$
	720	0.211$_{.000}$	**0.318$_{.000}$**	0.211$_{.000}$	0.330$_{.000}$	**0.205$_{.001}$**	0.324$_{.002}$	0.219$_{.003}$	0.327$_{.002}$	0.217$_{.000}$	0.338$_{.000}$
ETTh1	96	0.342$_{.000}$	**0.619$_{.000}$**	**0.321$_{.000}$**	0.626$_{.000}$	0.328$_{.003}$	0.640$_{.002}$	0.352$_{.011}$	0.668$_{.012}$	0.367$_{.000}$	0.656$_{.000}$
	192	0.364$_{.000}$	**0.661$_{.000}$**	**0.359$_{.000}$**	0.690$_{.000}$	**0.359$_{.002}$**	0.705$_{.001}$	0.393$_{.001}$	0.745$_{.003}$	0.392$_{.000}$	0.706$_{.000}$
	336	**0.371$_{.000}$**	**0.666$_{.000}$**	0.388$_{.000}$	0.723$_{.000}$	0.384$_{.002}$	0.740$_{.004}$	0.419$_{.007}$	0.778$_{.009}$	0.398$_{.000}$	0.711$_{.000}$
	720	**0.376$_{.000}$**	**0.679$_{.000}$**	0.408$_{.000}$	0.735$_{.000}$	0.397$_{.002}$	0.738$_{.001}$	0.502$_{.029}$	0.860$_{.049}$	0.433$_{.000}$	0.739$_{.000}$
ETTh2	96	**0.158$_{.000}$**	**0.239$_{.000}$**	0.177$_{.000}$	0.279$_{.000}$	0.177$_{.000}$	0.281$_{.001}$	0.211$_{.027}$	0.320$_{.033}$	0.203$_{.000}$	0.317$_{.000}$
	192	**0.170$_{.000}$**	**0.259$_{.000}$**	0.203$_{.000}$	0.314$_{.000}$	0.201$_{.001}$	0.314$_{.001}$	0.238$_{.028}$	0.353$_{.030}$	0.226$_{.000}$	0.346$_{.000}$
	336	**0.188$_{.000}$**	**0.282$_{.000}$**	0.243$_{.000}$	0.372$_{.000}$	0.240$_{.001}$	0.366$_{.001}$	0.284$_{.008}$	0.407$_{.013}$	0.264$_{.000}$	0.398$_{.000}$
	720	**0.215$_{.000}$**	**0.319$_{.000}$**	0.264$_{.000}$	0.386$_{.000}$	0.252$_{.000}$	0.371$_{.000}$	0.307$_{.000}$	0.426$_{.007}$	0.287$_{.000}$	0.416$_{.000}$
Electricity	96	**0.085$_{.000}$**	0.777$_{.000}$	0.098$_{.000}$	**0.772$_{.000}$**	0.086$_{.001}$	0.816$_{.005}$	0.090$_{.001}$	0.863$_{.002}$	0.140$_{.000}$	0.977$_{.016}$
	192	0.093$_{.000}$	0.933$_{.000}$	0.106$_{.000}$	**0.916$_{.000}$**	**0.092$_{.001}$**	0.942$_{.007}$	0.095$_{.001}$	0.974$_{.001}$	0.136$_{.000}$	1.017$_{.000}$
	336	0.101$_{.000}$	1.063$_{.000}$	0.115$_{.000}$	**0.985$_{.001}$**	**0.100$_{.000}$**	1.035$_{.003}$	0.104$_{.000}$	1.066$_{.004}$	0.147$_{.000}$	1.080$_{.006}$
	720	0.117$_{.000}$	1.289$_{.000}$	0.133$_{.000}$	**1.110$_{.001}$**	**0.116$_{.000}$**	1.213$_{.003}$	0.122$_{.001}$	1.259$_{.009}$	0.159$_{.000}$	1.283$_{.005}$
Traffic	96	**0.195$_{.000}$**	**0.461$_{.000}$**	0.246$_{.000}$	0.511$_{.000}$	0.248$_{.001}$	0.527$_{.001}$	0.356$_{.009}$	0.645$_{.017}$	0.293$_{.000}$	0.560$_{.000}$
	192	**0.193$_{.000}$**	**0.459$_{.000}$**	0.259$_{.000}$	0.543$_{.000}$	0.245$_{.001}$	0.528$_{.001}$	0.346$_{.009}$	0.628$_{.009}$	0.318$_{.000}$	0.594$_{.000}$
	336	**0.199$_{.000}$**	**0.468$_{.000}$**	0.283$_{.000}$	0.571$_{.000}$	0.257$_{.002}$	0.550$_{.001}$	0.350$_{.008}$	0.631$_{.008}$	0.332$_{.000}$	0.630$_{.000}$
	720	**0.218$_{.000}$**	**0.497$_{.000}$**	0.275$_{.000}$	0.563$_{.000}$	0.266$_{.001}$	0.559$_{.001}$	0.365$_{.009}$	0.659$_{.009}$	0.341$_{.003}$	0.611$_{.002}$
Weather	96	**0.086$_{.000}$**	**0.287$_{.000}$**	0.089$_{.000}$	0.295$_{.000}$	0.087$_{.002}$	0.294$_{.002}$	0.112$_{.001}$	0.316$_{.000}$	0.239$_{.004}$	0.614$_{.023}$
	192	0.092$_{.000}$	0.312$_{.000}$	0.093$_{.000}$	**0.299$_{.000}$**	**0.090$_{.001}$**	**0.299$_{.001}$**	0.122$_{.001}$	0.331$_{.000}$	0.213$_{.000}$	0.533$_{.002}$
	336	**0.091$_{.000}$**	0.307$_{.000}$	0.096$_{.000}$	**0.297$_{.000}$**	0.092$_{.002}$	**0.297$_{.001}$**	0.130$_{.002}$	0.340$_{.002}$	0.176$_{.000}$	0.413$_{.001}$
	720	**0.093$_{.000}$**	0.308$_{.000}$	0.099$_{.000}$	**0.298$_{.000}$**	0.094$_{.001}$	**0.298$_{.003}$**	0.144$_{.001}$	0.358$_{.002}$	0.170$_{.001}$	0.434$_{.010}$
Exchange	96	0.026$_{.000}$	0.039$_{.000}$	0.025$_{.000}$	0.039$_{.000}$	**0.023$_{.000}$**	**0.036$_{.000}$**	0.024$_{.000}$	0.037$_{.000}$	0.032$_{.000}$	0.049$_{.000}$
	192	**0.033$_{.000}$**	**0.050$_{.000}$**	0.036$_{.000}$	0.056$_{.000}$	0.034$_{.000}$	0.054$_{.000}$	0.035$_{.000}$	0.055$_{.000}$	0.041$_{.000}$	0.065$_{.000}$
	336	**0.041$_{.000}$**	**0.062$_{.000}$**	0.048$_{.000}$	0.072$_{.000}$	0.048$_{.000}$	0.076$_{.000}$	0.048$_{.001}$	0.072$_{.001}$	0.056$_{.000}$	0.091$_{.000}$
	720	**0.059$_{.000}$**	**0.089$_{.000}$**	0.076$_{.000}$	0.114$_{.000}$	0.072$_{.000}$	0.106$_{.001}$	0.075$_{.002}$	0.118$_{.004}$	0.112$_{.002}$	0.164$_{.004}$

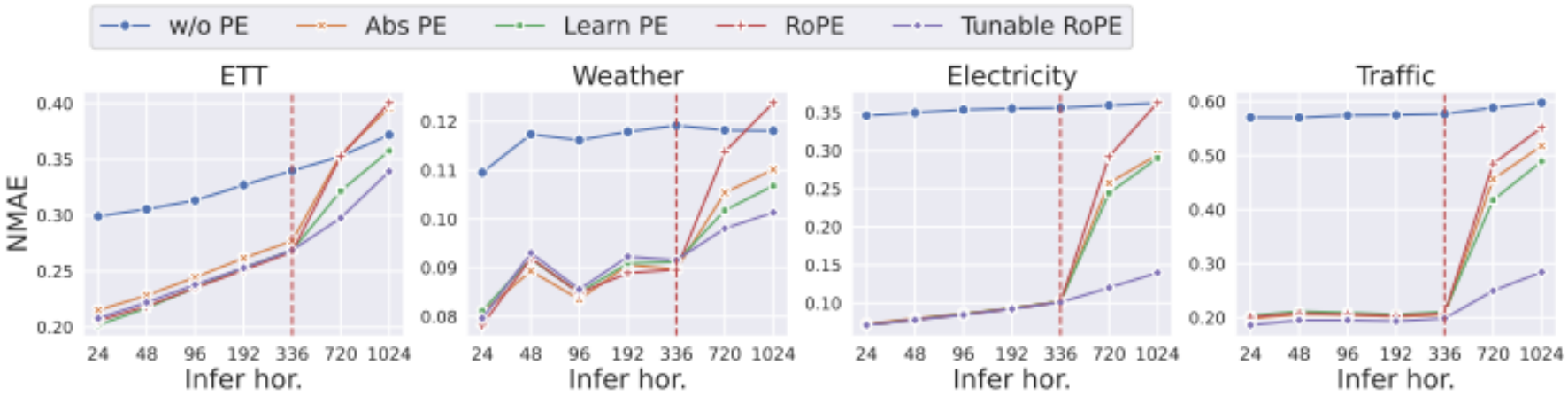
- 기존 Transformer 기반 모델들과 비교하였을 때, ElasTST가 전체적으로 더 뛰어난 성능을 보임

ElasTST

FAIR Lab

KWANGWOON UNIVERSITY

위치 임베딩 비교



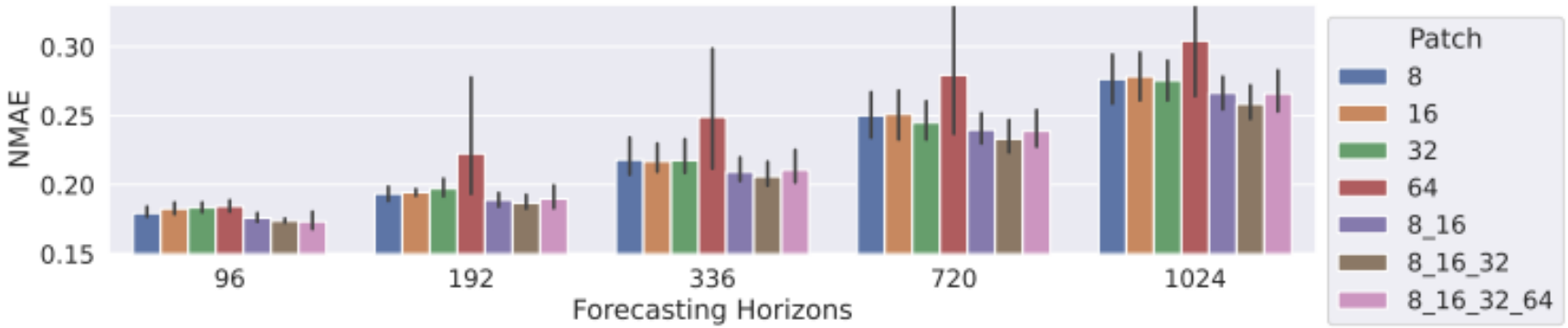
- 데이터의 패턴에 맞춰 학습되는 TRoPE로 인해 학습하지 않은 긴 시계열 데이터가 들어와도 안정적으로 예측이 가능

ElasTST

FAIR Lab

KWANGWOON UNIVERSITY

Patch 크기별 성능 비교



- 단일 패치를 사용하는 것보다 여러 패치를 조합하여 사용하는 것이 모든 예측 구간에서 가장 낮은 오차를 기록
- 세밀한 변동과 전역적인 추세를 동시에 포착하는 Multi-Scale이 예측 정확도를 안정화함

ElasTST

FAIR Lab

KWANGWOON UNIVERSITY

실험 세팅

- 사용한 모델: ElasTST
- 재현 실험 데이터셋: ETTh1

Experiment	ETTh1
Learning rate	10^{-3}
Epoch	20
Batch size	32
Loss function	L1 Loss
Seq_len	96
Pred_len	96/192/336/720
d_model	512
d_ff	512
Patch length	8, 16, 32

ElasTST

ElasTST 재현 실험 (ETTh1)

- 예측 길이가 길어질수록 어느 정도의 오차가 발생하였지만, 전체적으로 논문과 비슷한 수치가 나온 것을 확인

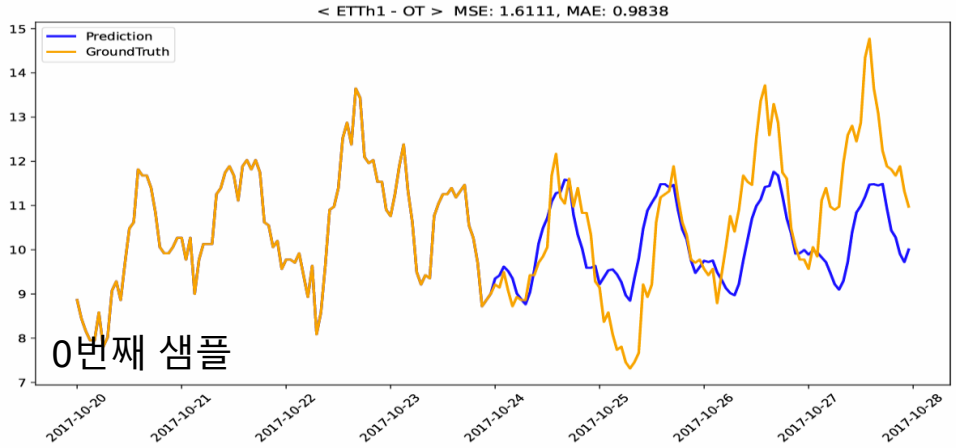
	ETTh1 Paper		ETTh1 Reproduction	
Pred len	NMAE	NRMSE	NMAE	NRMSE
96	0.342	0.619	0.328	0.627
192	0.364	0.661	0.353	0.672
336	0.371	0.666	0.371	0.704
720	0.376	0.679	0.385	0.698

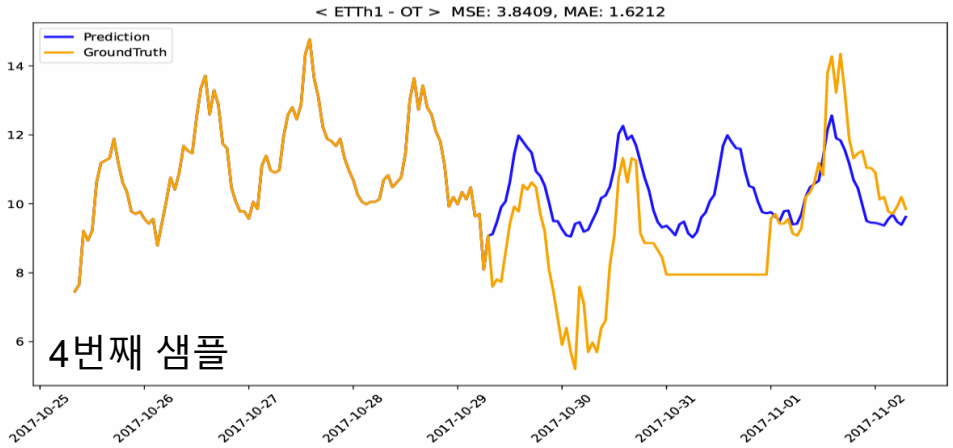
ElasTST

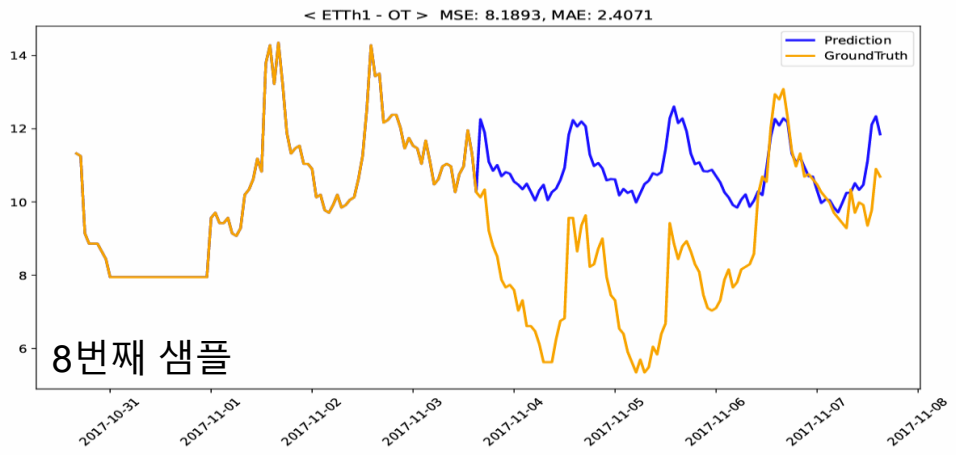
FAIR Lab

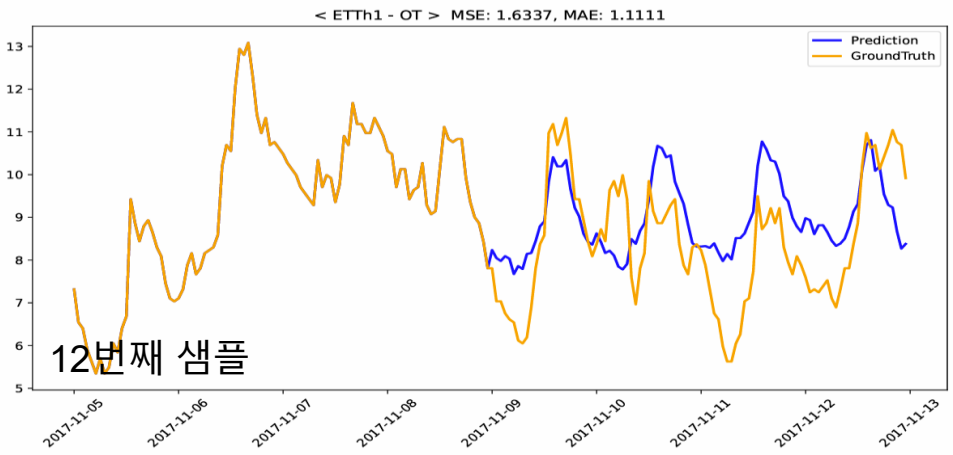
KWANGWOON UNIVERSITY

재현 실험 시각화 (ETTh1)









Seq_len → 96
Pred_len → 96

ElasTST

FAIR Lab

KWANGWOON UNIVERSITY

실험 결과 정리

- 재현 실험
  - 예측 길이가 길어질수록 오차가 발생하였지만 논문과 비슷한 수치가 나옴
- 시각화
  - 전반적으로 예측값이 정답값을 잘 따라가고 있으나, 급격한 변동이나 세밀한 피크 부분은 제대로 반영하지 못함