

2025 12 11
발표 자료

광운대학교 로봇학과
FAIR Lab

김한서

이번 주 진행사항

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

PETFORMER: LONG-TERM TIME SERIES FORECAST- ING VIA PLACEHOLDER-ENHANCED TRANSFORMER

Shengsheng Lin¹, Weiwei Lin^{1,2,*}, Wentai Wu², Songbo Wang¹, Yongxiang Wang¹

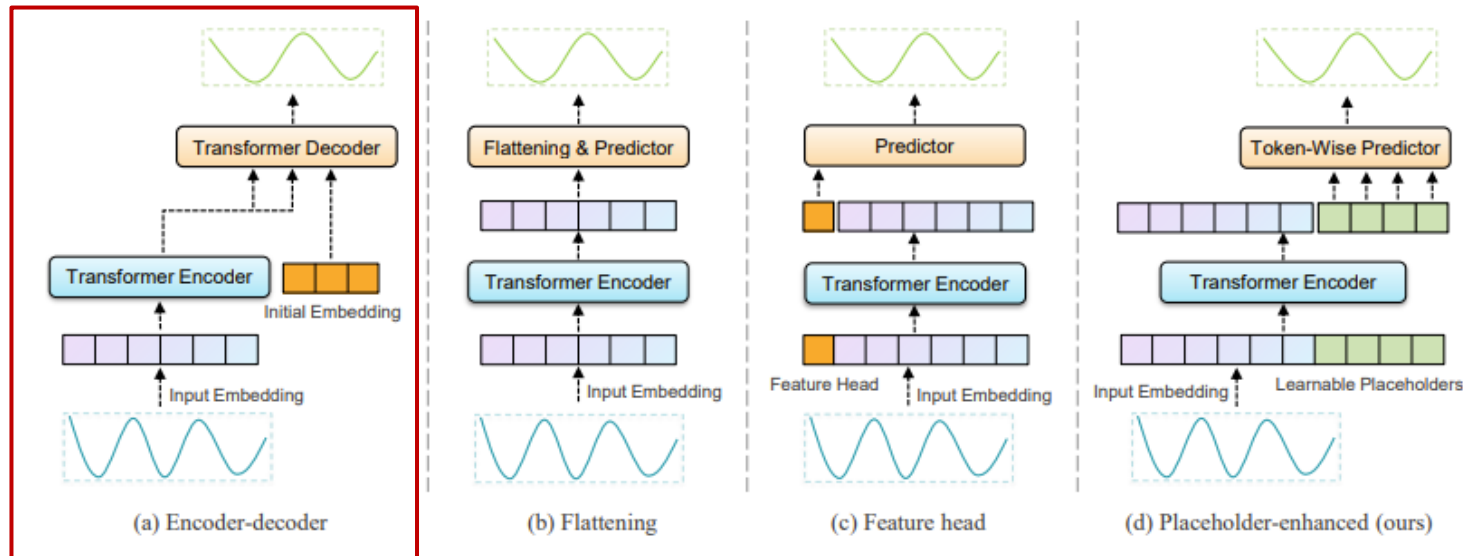
¹South China University of Technology ²Peng Cheng Laboratory

linss2000@foxmail.com, linww@scut.edu.cn, nnwtwu@pcl.ac.cn,
songbo1998@foxmail.com, wangyxv@163.com

- arXiv 등록일: 2023-08-09
- 인용 수: 55회(Google Scholar, 2025-12-07)
- Published at IEEE TETCI 2024
- Cross-Attention의 장기 시계열 예측 한계와 PatchTST의 시간적 특징 손실 문제 해결하기 위해, 미래 예측 시점을 위한 Learnable token으로 변환하여 인코더에 입력하는 Placeholder-enhanced Technique을 제안

PETformer

LTSF 응용 기술 비교

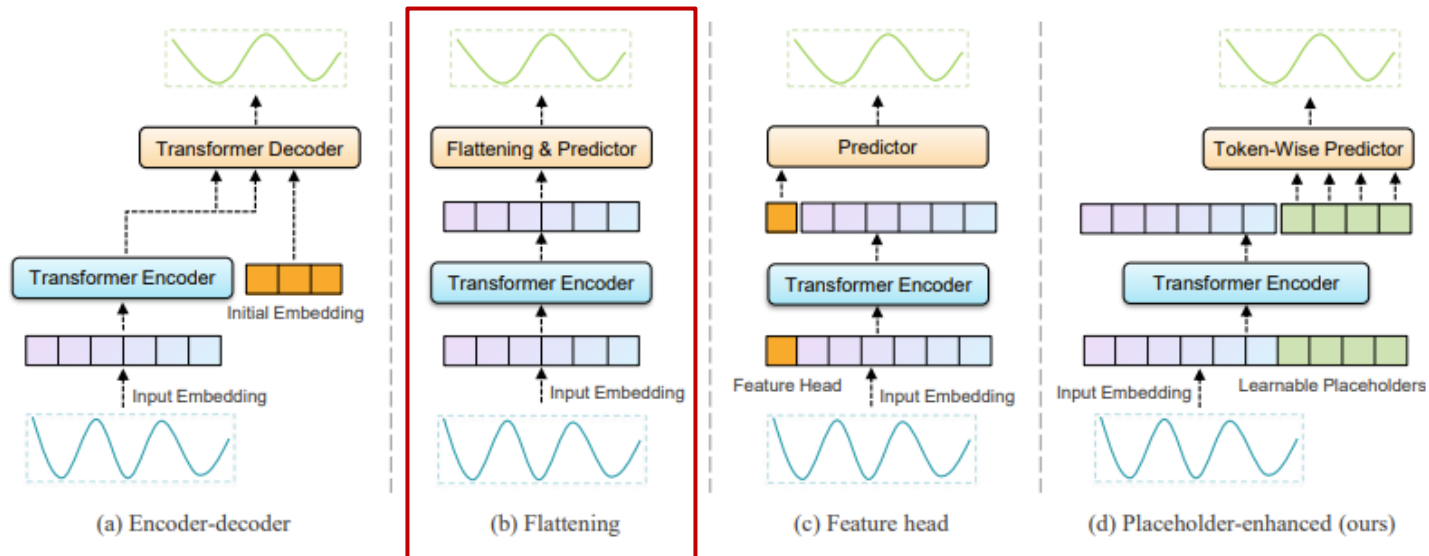


- Encoder-decoder (Transformer 방식)

- 인코더로 입력 시계열을 인코딩하고, 디코더가 초기 embedding을 받아 미래 예측값을 생성
- Step-by-step 생성 방식으로 인해 장기 시계열 예측 시 연산 효율이 낮음

PETformer

LTSF 응용 기술 비교

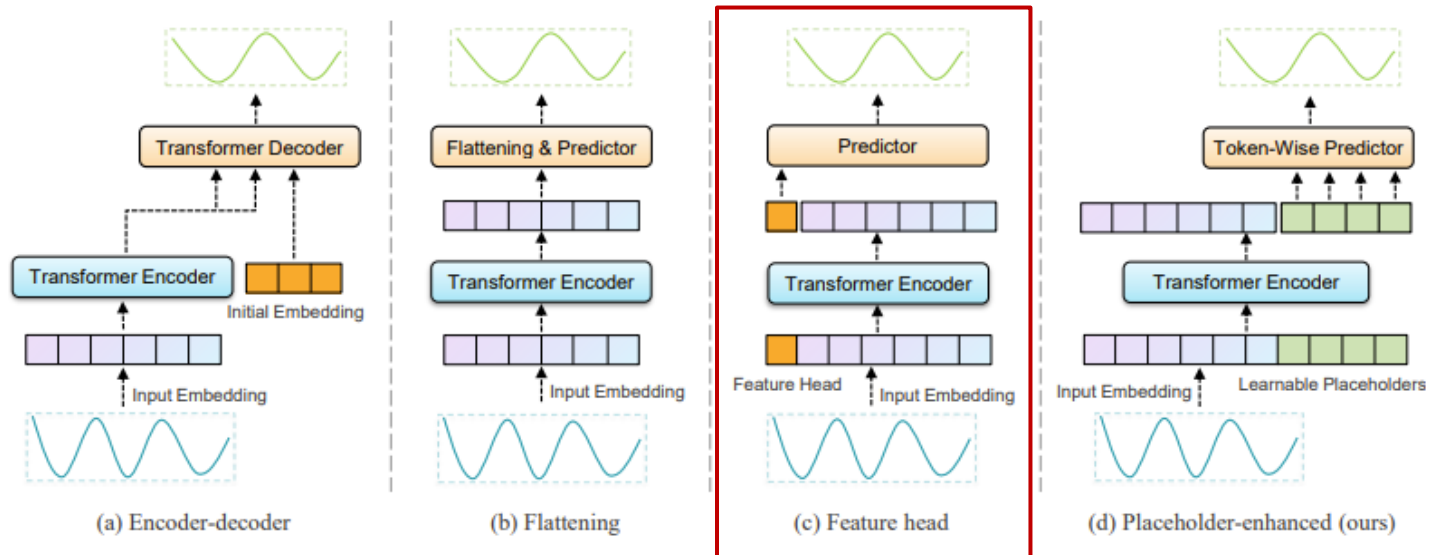


- Flattening (PatchTST 방식)

- 인코더의 모든 출력을 하나의 벡터로 Flatten하고, Predictor에 통과시켜 미래 예측값을 생성
- 연산은 단순해지지만, Flatten 과정에서 시계열의 시간적 상관관계 정보가 손실될 수 있음

PETformer

LTSF 응용 기술 비교

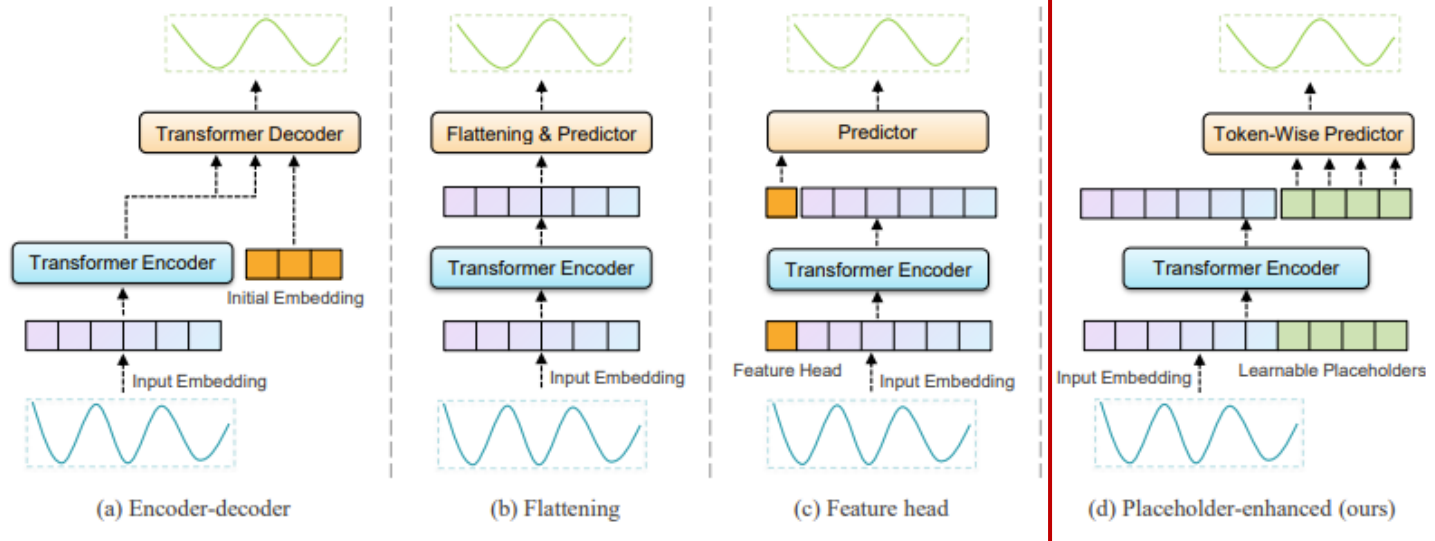


- Feature head

- 인코더 출력 토큰 각각에 Feature Head를 연결해 변수(feature)별로 독립적인 예측을 수행
- 변수 간의 독립성을 고려하지만, 시계열의 시간적 패턴을 충분히 반영하기 어렵다는 한계가 있음

PETformer

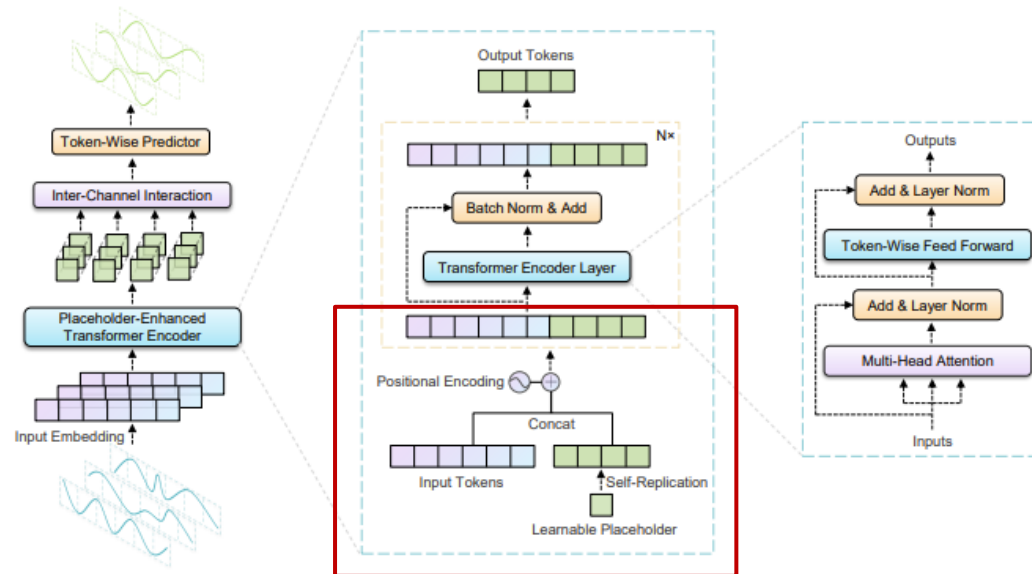
LTSF 응용 기술 비교



- Placeholder-enhanced (PETformer 방식)
 - 임베딩 벡터에 Learnable Placeholder token을 추가하고, Encoder에서 과거 정보인 기존 임베딩 벡터와 Placeholders를 하나로 결합하여 학습
 - Token-wise Predictor로 미래 구간을 직접 예측하며 정보 손실 최소화

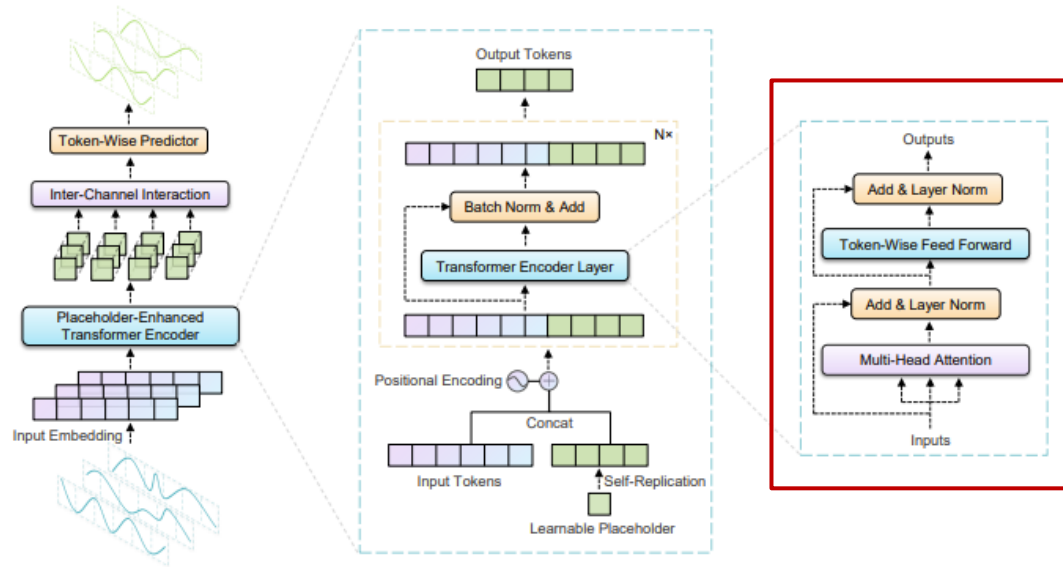
PETformer

모델 구조



- Input Embedding
 - 다변량 시계열 데이터를 부분 시퀀스로 분할 후, Linear projection을 통해 임베딩 벡터로 변환
- Placeholder-Enhanced Transformer Encoder
 - 변환된 임베딩 벡터 뒤에 Self-Replicated Placeholders를 Concat하여 하나의 입력 시퀀스로 결합
 - Positional Encoding을 통해 시계열 순서 정보를 더해줌

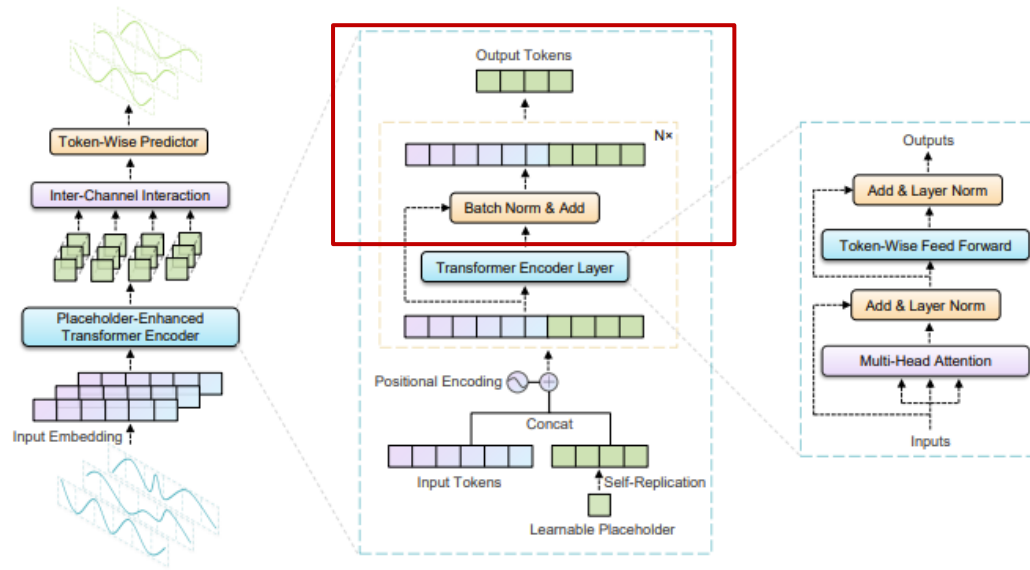
모델 구조



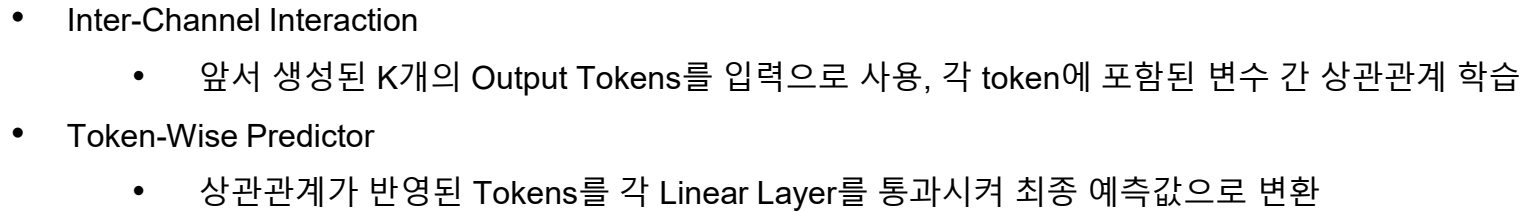
- Transformer Encoder Layer

- Multi-Head Attention을 통해 입력 시퀀스 내 시점 간 관계 학습하며, Placeholder token도 함께 포함되어 있어 과거 정보와 미래 시점의 위치를 나타내는 토큰을 결합하여 학습
- Token-Wise Feed Forward 네트워크로 비선형 변환 및 표현력 강화
- Residual Connection과 Layer Normalization을 통한 안정적 학습

모델 구조

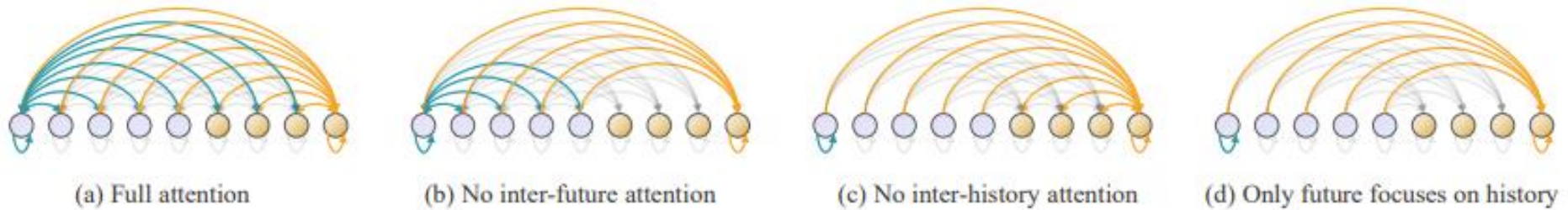


- Batch Norm & Add
 - Encoder를 통과한 시퀀스를 잔차 연결(Residual Connection) 후 정규화(Normalization)를 수행하여 학습 안정성 확보
- Output Tokens
 - 안정화된 시퀀스로부터 미래 예측에 필요한 정보가 압축된 Placeholder를 추출, 이후 K개의 Output Tokens 생성



PETformer

Attention 방식 비교



- Full attention (Transformer 방식)
 - 모든 시점이 서로를 참조하는 모습, 시퀀스 길이에 따라 연산량이 급격히 증가
- Only future focuses on history (PETformer 방식)
 - 미래 시점이 오직 과거 정보만 참조하는 모습, 연산량이 줄어 연산 효율성 증가

채널 간 상호작용 방식

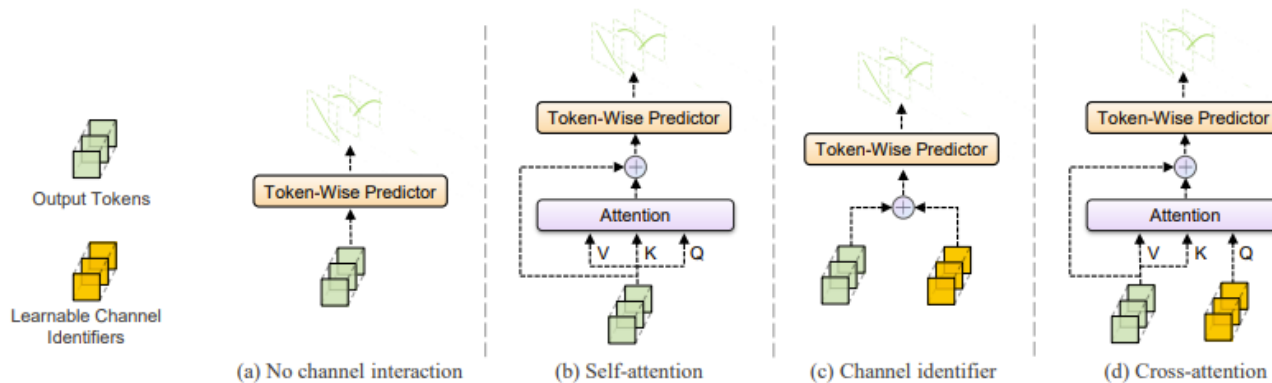


Figure 4: Different inter-channel interactions.

- No channel interaction
 - 변수끼리 참조하는 것이 아닌 개별적으로 예측하는 방식
- Self-attention
 - 모든 변수(Channel)를 토큰으로 취급하여 전체 변수 간의 상관관계를 통합 학습
- Channel identifier
 - 각 변수에 학습 가능한 식별자(Identifier)를 부여하여 변수 고유의 특성을 보존하며 구분
- Cross-attention
 - Query, Key, Value 간의 Cross-attention을 통해 채널 간 상호작용을 효율적으로 학습

주요 모델 성능 비교

Models		PETformer (ours)		PatchTST (2023)		Dlinear (2023)		MICN (2023)		Crossformer (2023)		FEDformer* (2022)		Autoformer* (2021)		Informer* (2021)	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.347	0.377	<u>0.376</u>	0.408	0.378	<u>0.402</u>	0.404	0.429	0.380	0.419	<u>0.376</u>	0.415	0.435	0.446	0.941	0.769
	192	0.390	0.404	<u>0.416</u>	<u>0.423</u>	<u>0.415</u>	<u>0.425</u>	0.475	0.484	0.419	0.445	<u>0.423</u>	0.446	0.456	0.457	1.007	0.786
	336	0.419	0.418	<u>0.425</u>	<u>0.440</u>	0.447	0.448	0.482	0.489	0.438	0.451	0.444	0.462	0.486	0.487	1.038	0.784
	720	0.437	0.449	<u>0.448</u>	<u>0.470</u>	0.480	0.489	0.599	0.576	0.508	0.514	0.469	0.492	0.515	0.517	1.144	0.857
ETTh2	96	0.272	0.329	<u>0.275</u>	<u>0.338</u>	0.282	0.346	0.289	0.354	0.383	0.420	0.332	0.374	0.332	0.368	1.549	0.952
	192	0.338	0.374	<u>0.338</u>	<u>0.378</u>	0.350	0.396	0.408	0.444	0.421	0.450	0.407	0.446	0.426	0.434	3.792	1.542
	336	0.328	0.380	<u>0.329</u>	0.380	0.410	<u>0.437</u>	0.547	0.516	0.449	0.459	0.4	0.447	0.477	0.479	4.215	1.642
	720	<u>0.401</u>	<u>0.439</u>	0.379	0.422	0.587	0.544	0.834	0.688	0.472	0.497	0.412	0.469	0.453	0.490	3.656	1.619
ETTm1	96	0.282	0.325	<u>0.293</u>	<u>0.342</u>	0.306	0.345	0.301	0.352	0.295	0.350	0.326	0.390	0.51	0.492	0.626	0.560
	192	0.318	0.349	<u>0.328</u>	<u>0.365</u>	0.335	<u>0.365</u>	0.344	0.380	0.339	0.381	0.365	0.415	0.514	0.495	0.725	0.619
	336	0.348	0.372	<u>0.362</u>	<u>0.394</u>	0.373	<u>0.391</u>	0.379	0.401	0.419	0.432	0.392	0.425	0.51	0.492	1.005	0.741
	720	0.404	0.403	<u>0.414</u>	<u>0.420</u>	0.422	<u>0.422</u>	0.429	0.429	0.579	0.551	0.446	0.458	0.527	0.493	1.133	0.845
ETTm2	96	0.160	0.248	<u>0.163</u>	<u>0.255</u>	0.164	0.259	0.177	0.274	0.296	0.352	0.18	0.271	0.205	0.293	0.355	0.462
	192	0.217	0.288	<u>0.221</u>	<u>0.292</u>	0.233	0.314	0.236	0.310	0.342	0.385	0.252	0.318	0.278	0.336	0.595	0.586
	336	0.274	0.326	0.270	<u>0.329</u>	0.291	0.355	0.299	0.350	0.410	0.425	0.324	0.364	0.343	0.379	1.27	0.871
	720	0.345	0.376	<u>0.347</u>	<u>0.378</u>	0.407	0.433	0.421	0.434	0.563	0.538	0.41	0.420	0.414	0.419	3.001	1.267
Electricity	96	0.128	0.220	<u>0.130</u>	<u>0.223</u>	0.133	0.230	0.151	0.260	0.198	0.292	0.186	0.302	0.196	0.313	0.304	0.393
	192	0.144	0.236	<u>0.147</u>	<u>0.240</u>	0.147	0.244	0.165	0.276	0.266	0.330	0.197	0.311	0.211	0.324	0.327	0.417
	336	0.159	0.252	0.164	<u>0.257</u>	<u>0.162</u>	0.261	0.183	0.291	0.343	0.377	0.213	0.328	0.214	0.327	0.333	0.422
	720	0.195	0.286	0.203	<u>0.292</u>	<u>0.196</u>	0.294	0.201	0.312	0.398	0.422	0.233	0.344	0.236	0.342	0.351	0.427
ILI	24	1.204	0.687	<u>1.356</u>	<u>0.732</u>	2.000	0.987	2.483	1.058	3.217	1.198	2.624	1.095	2.906	1.182	4.657	1.449
	36	1.246	0.709	1.244	0.705	2.202	1.026	2.370	0.987	3.136	1.199	2.516	1.021	2.585	1.038	4.65	1.463
	48	1.446	0.760	<u>1.604</u>	0.791	2.278	1.059	2.371	1.007	3.331	1.236	2.505	1.041	3.024	1.145	5.004	1.542
	60	1.430	0.774	<u>1.648</u>	<u>0.860</u>	2.478	1.111	2.513	1.055	3.609	1.265	2.742	1.122	2.761	1.114	5.071	1.543
Traffic	96	0.357	0.240	<u>0.367</u>	<u>0.253</u>	0.385	0.269	0.445	0.295	0.487	0.274	0.576	0.359	0.597	0.371	0.733	0.410
	192	0.376	0.248	<u>0.382</u>	<u>0.259</u>	0.395	0.273	0.461	0.302	0.497	0.279	0.61	0.380	0.607	0.382	0.777	0.435
	336	0.392	0.255	<u>0.396</u>	<u>0.267</u>	0.409	0.281	0.483	0.307	0.517	0.285	0.608	0.375	0.623	0.387	0.776	0.434
	720	0.430	0.276	<u>0.433</u>	<u>0.287</u>	0.449	0.305	0.527	0.310	0.584	0.323	0.621	0.375	0.639	0.395	0.827	0.466
Weather	96	0.146	0.186	0.147	<u>0.198</u>	0.169	0.231	0.167	0.231	0.144	0.208	0.238	0.314	0.249	0.329	0.354	0.405
	192	0.190	0.229	0.190	<u>0.241</u>	0.213	0.273	0.212	0.271	0.192	0.263	0.275	0.329	0.325	0.370	0.419	0.434
	336	0.241	0.271	<u>0.243</u>	<u>0.284</u>	0.260	0.314	0.275	0.337	0.246	0.306	0.339	0.377	0.351	0.391	0.583	0.543
	720	0.314	0.323	0.305	<u>0.328</u>	0.315	0.353	<u>0.312</u>	0.349	0.318	0.361	0.389	0.409	0.415	0.426	0.916	0.705
Avg.		0.427	0.369	<u>0.448</u>	<u>0.383</u>	0.565	0.434	0.623	0.455	0.756	0.490	0.651	0.472	0.713	0.497	1.629	0.825

- 대부분의 데이터셋에서 기존 모델인 PatchTST, DLinear 보다 더 뛰어난 성능을 보임

실험 세팅

- 사용한 모델: PETformer
- 재현 실험 데이터셋: ETTh1, Weather

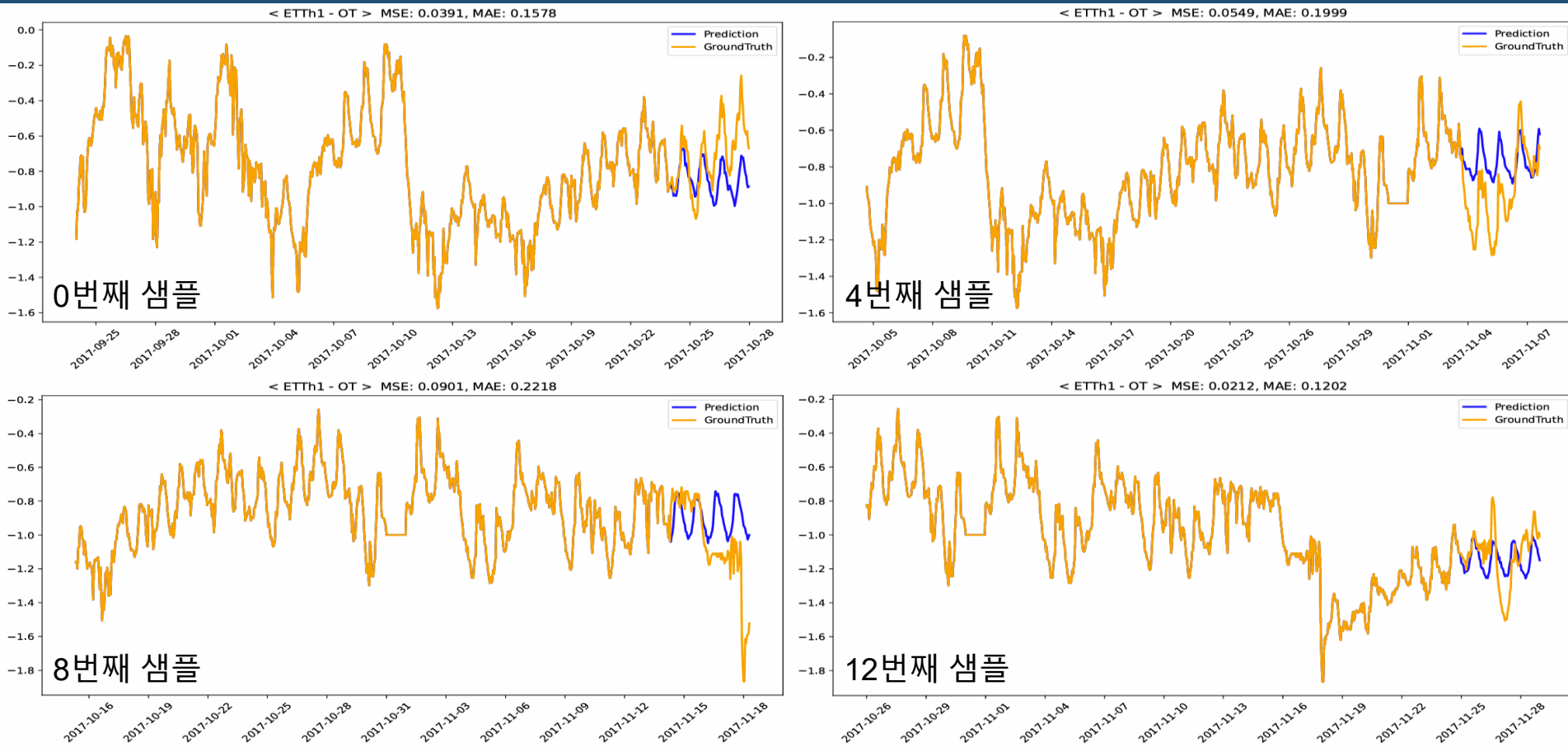
Experiment	ETTh1, Weather
Learning rate	10^{-4}
Epoch	50
Batch size	64
Loss function	Smooth L1 Loss
Seq_len	720
Pred_len	96/192/336/720
d_model	512
d_ff	1024
Patch length	48
Stride	48

PETformer 재현 실험 (ETTh1, Weather)

- 전체적으로 논문 수치와 비슷하게 나오지만 예측 길이가 길어질수록 어느정도 오차가 생기는 것을 확인하였음

	ETTh1 Paper		ETTh1 Reproduction		Weather Paper		Weather Reproduction	
Pred len	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
96	0.347	0.377	0.350	0.381	0.146	0.186	0.145	0.185
192	0.390	0.404	0.392	0.408	0.190	0.229	0.189	0.229
336	0.419	0.418	0.418	0.422	0.241	0.271	0.247	0.276
720	0.437	0.449	0.471	0.469	0.314	0.323	0.321	0.329

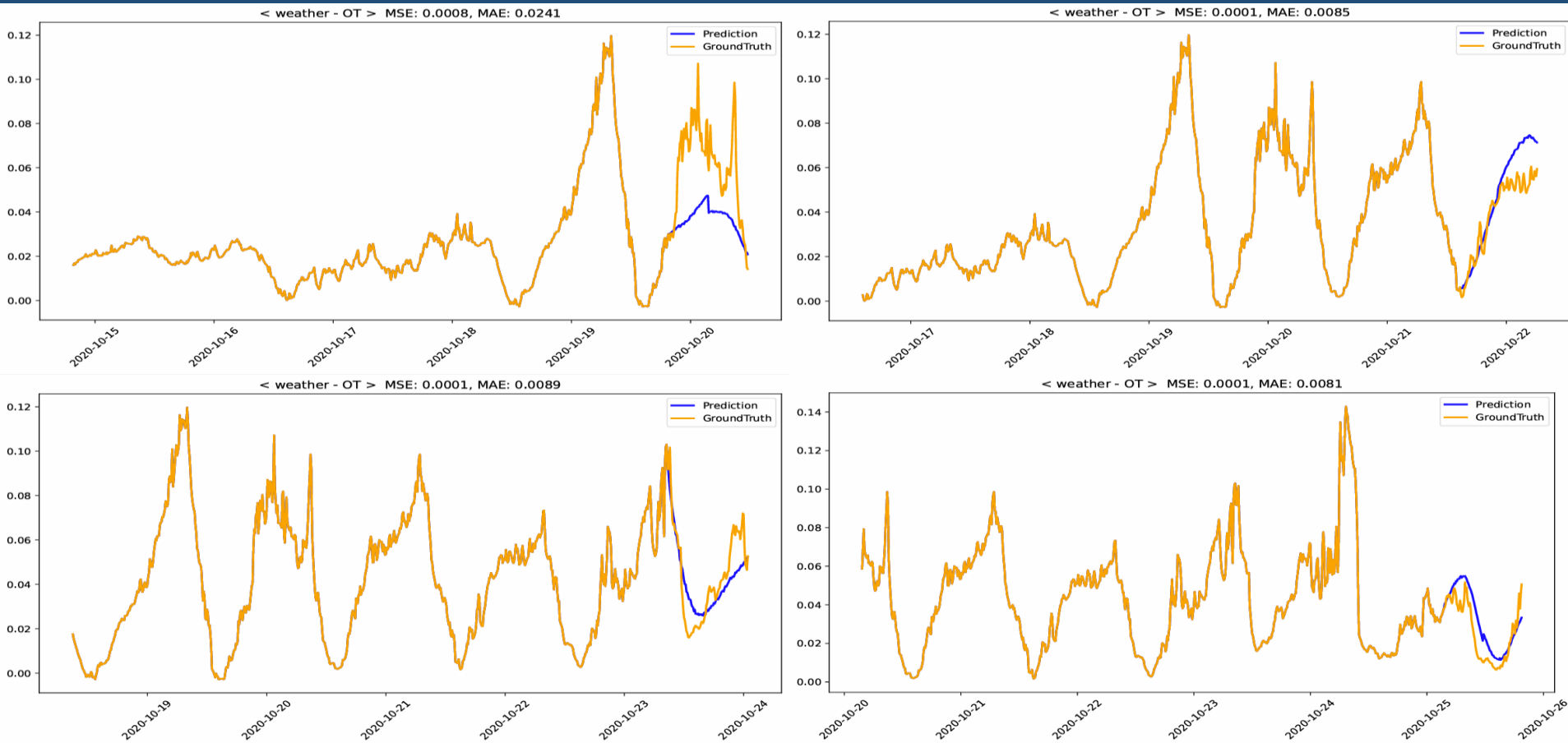
PETformer 재현 실험 시각화 (ETTh1)



Seq_len → 720

Pred_len → 96

PETformer 재현 실험 시각화 (Weather)



Seq_len → 720

Pred_len → 96

PETformer, PatchTST, DLinear 비교 실험 (ETTh1)

- 전체적으로 기존 모델인 PatchTST, DLinear 대비 PETformer가 더 뛰어난 성능을 보이는 것을 확인하였음

	PETformer ETTh1		PatchTST ETTh1		DLinear ETTh1	
Pred len	MSE	MAE	MSE	MAE	MSE	MAE
96	0.350	0.381	<u>0.370</u>	<u>0.400</u>	0.396	0.410
192	0.392	0.408	<u>0.413</u>	<u>0.429</u>	0.445	0.440
336	0.418	0.422	<u>0.421</u>	<u>0.439</u>	0.487	0.465
720	<u>0.471</u>	<u>0.469</u>	0.447	0.468	0.512	0.510

실험 결과 정리

- 재현 실험 및 비교 실험
 - 재현 실험의 경우 전체적으로 논문과 비슷한 수치가 나왔으며, 비교 실험도 마찬가지로 논문의 예측 결과와 동일하게 PETformer의 성능이 가장 좋게 나왔음
- 시각화
 - ETTh1, Weather 모두 전체적으로 예측값이 정답값을 잘 따라가는 모습을 보임