

2026 01 21
발표 자료

광운대학교 로봇학과
FAIR Lab

김한서

이번 주 진행사항

- EMAformer
 - 논문 리뷰
 - 실험 및 결과

EMAformer: Enhancing Transformer through Embedding Armor for Time Series Forecasting

Zhiwei Zhang¹, Xinyi Du², Xuanchi Guo¹, Weihao Wang¹, Wenjuan Han^{1*},

¹School of Computer Science and Technology, Beijing Jiaotong University, Beijing, China

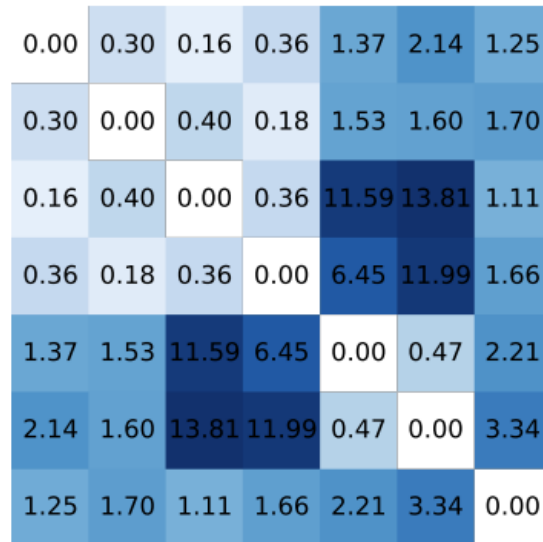
²Beijing Normal University, Beijing, China

{zhiweizhang, xuanchigu, weihaow, wjhan}@bjtu.edu.cn

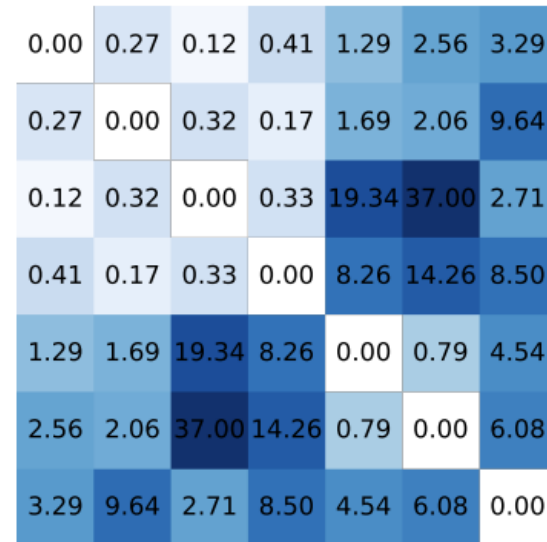
{xinyidu}@mail.bnu.edu.cn

- arXiv 등록일: 2025-11-11
- 인용 수: 0회(Google Scholar, 2026-01-16)
- Published at AAAI 2026
- 링크: <https://arxiv.org/pdf/2511.08396>

Background



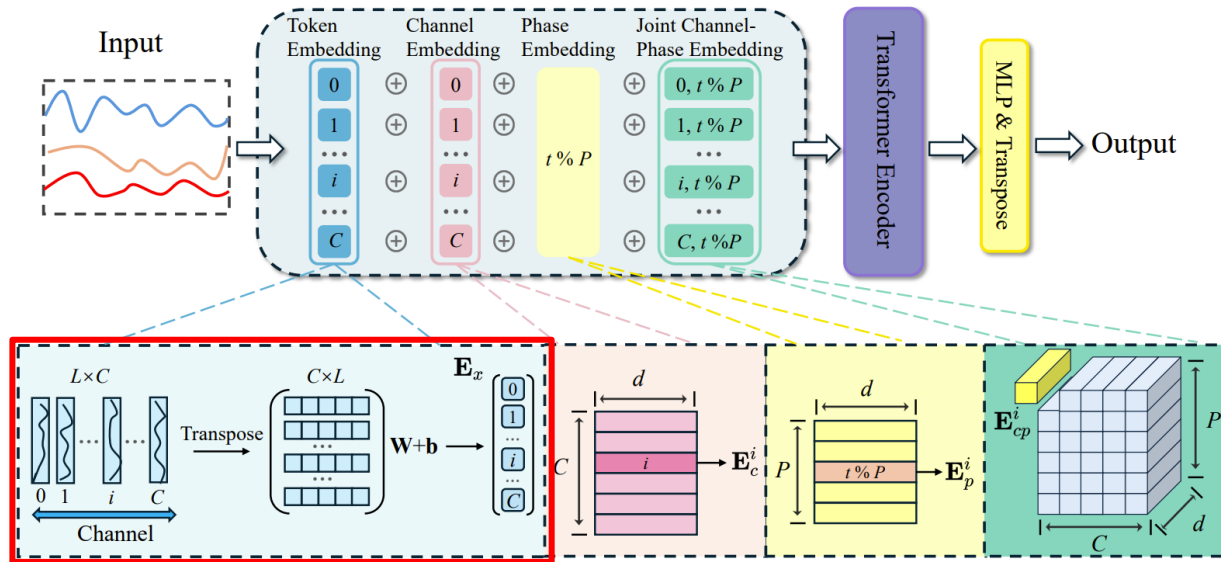
(a) CoV of ETTh2.



(b) CoV of ETTm2.

- 기존 Transformer 모델이 TQNet과 같은 최신 MLP 기반 모델보다 성능이 뒤처지는 현상 발생
- 변동 계수 CoV를 통해 채널 간 상관관계를 분석한 결과, 채널 간 관계가 고정적이지 않고 극심하게 요동침
- 이를 해결하기 위해 EMAformer의 보조 임베딩을 통해 전역적 안정성과 주기적 패턴을 주입

모델 구조



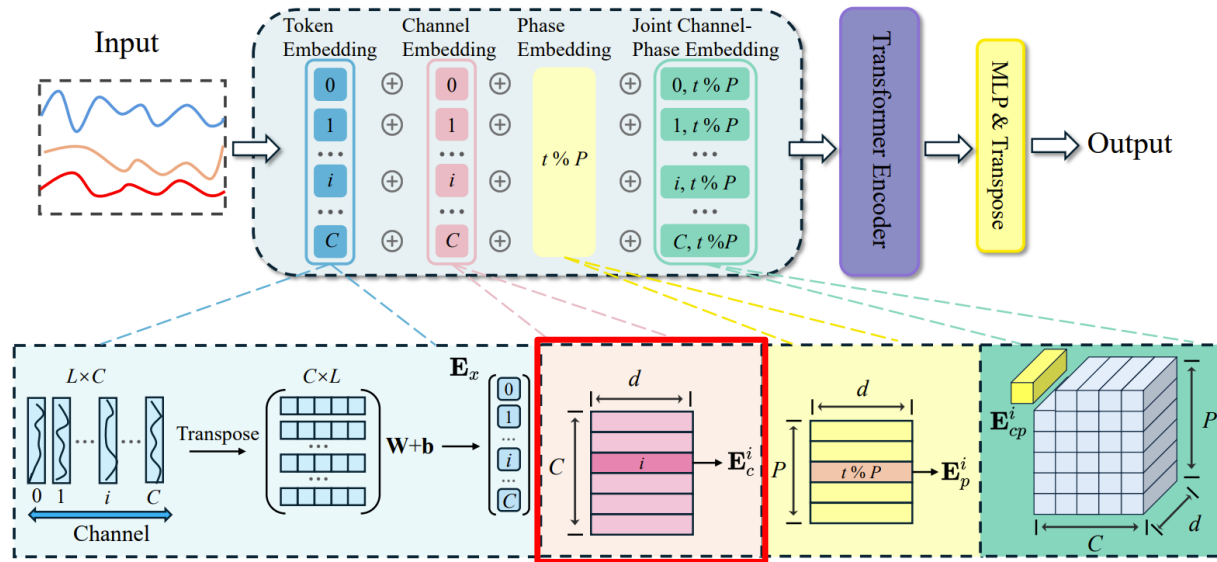
Token Embedding

- 입력받은 행렬 $L \times C$ 를 Transpose하여 $C \times L$ 로 변환 X^T
- 이후 Linear Projection을 통해 d 차원 임베딩 벡터로 변환 $E_x = X^T W + b \in \mathbb{R}^{C \times d}$
- 최종적으로 d 차원의 특징 벡터를 가지는 C 개의 변량 토큰 생성

* C : 변량(채널) 개수
 * L : 과거 타임스텝
 * W : 학습 가능한 가중치 행렬
 * b : 편향 벡터

EMAformer

모델 구조



- Channel Embedding

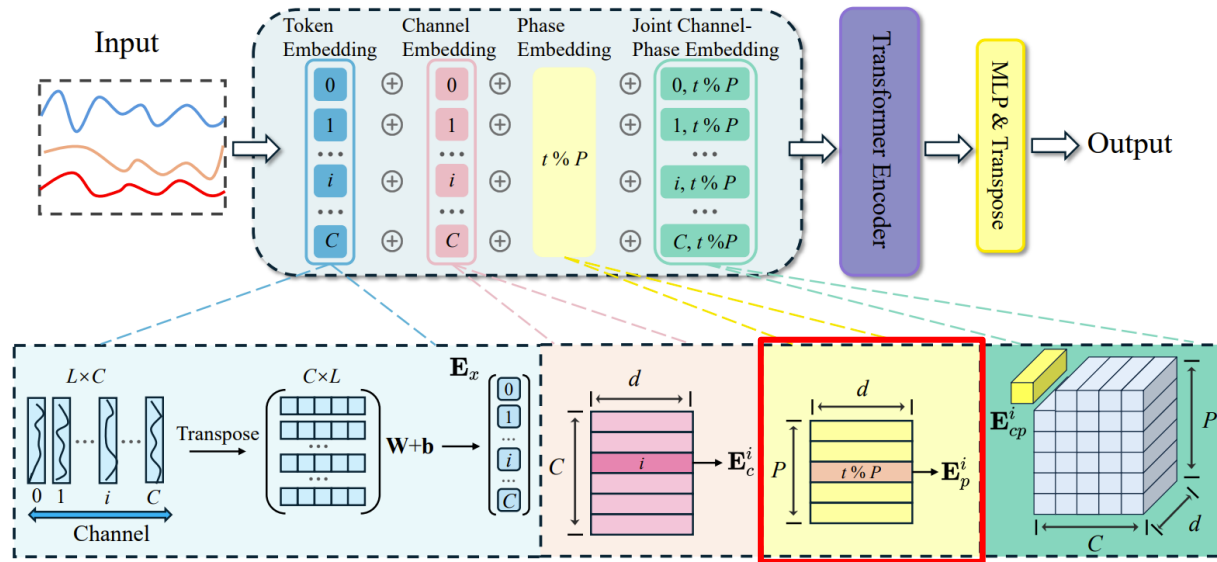
- 변량 토큰에 학습 가능한 행렬 $\Omega_c \in \mathbb{R}^{C \times d}$ 을 Element-wise Sum하여 각 토큰에 채널별 고유 정체성 주입
- Ω_c 은 초기 무작위 숫자 행렬에서 Backpropagation을 통해 업데이트, i 번째 행 벡터를 추출하여 더함

$$E_c^i = \text{Lookup}(\Omega_c, i) \in \mathbb{R}^{1 \times d}$$

* C : 변량(채널) 개수
 * d : 임베딩 차원
 * E_c^i : i 번째 변량 토큰 + i 번째 오메가 행렬 벡터

EMAformer

모델 구조



- Phase Embedding

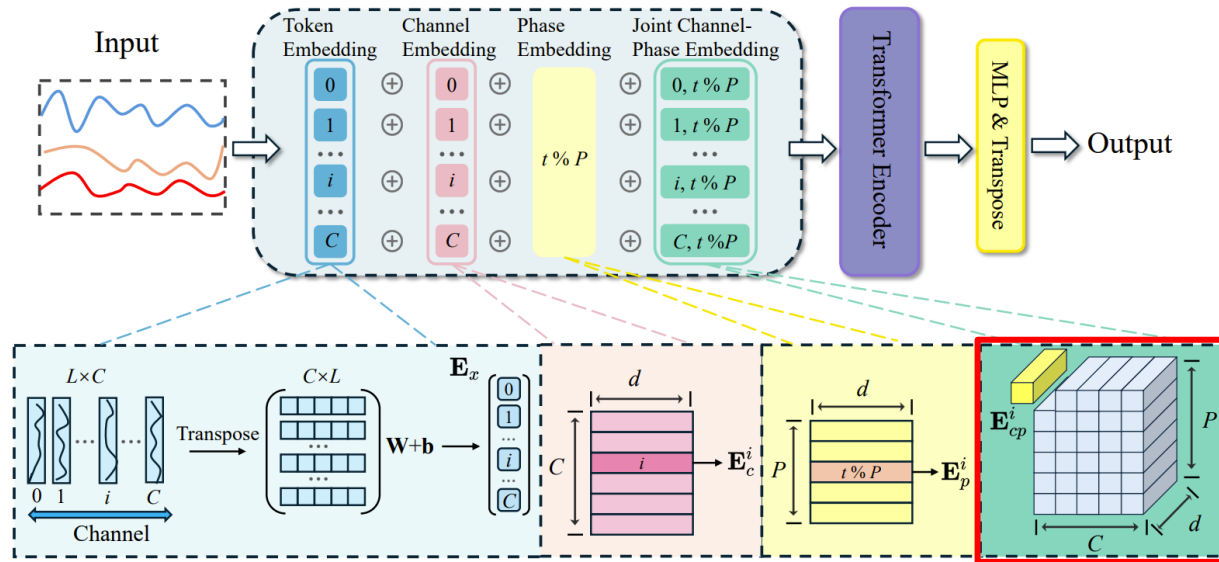
- 변량 토큰에 학습 가능한 행렬 $\Omega_p \in \mathbb{R}^{P \times d}$ 을 Element-wise Sum하여 주기 내의 위치 정보 주입
- 마지막 관측 시점 t 를 주기 길이 P 로 나눈 나머지 ($t \pmod{P}$)를 통해 해당 행에 있는 벡터 추출하여 더함

$$E_p^i = \text{Lookup}(\Omega_p, t \pmod{P}) \in \mathbb{R}^{1 \times d}$$

*P: 주기 길이
 *mod: 나머지 연산

EMAformer

모델 구조



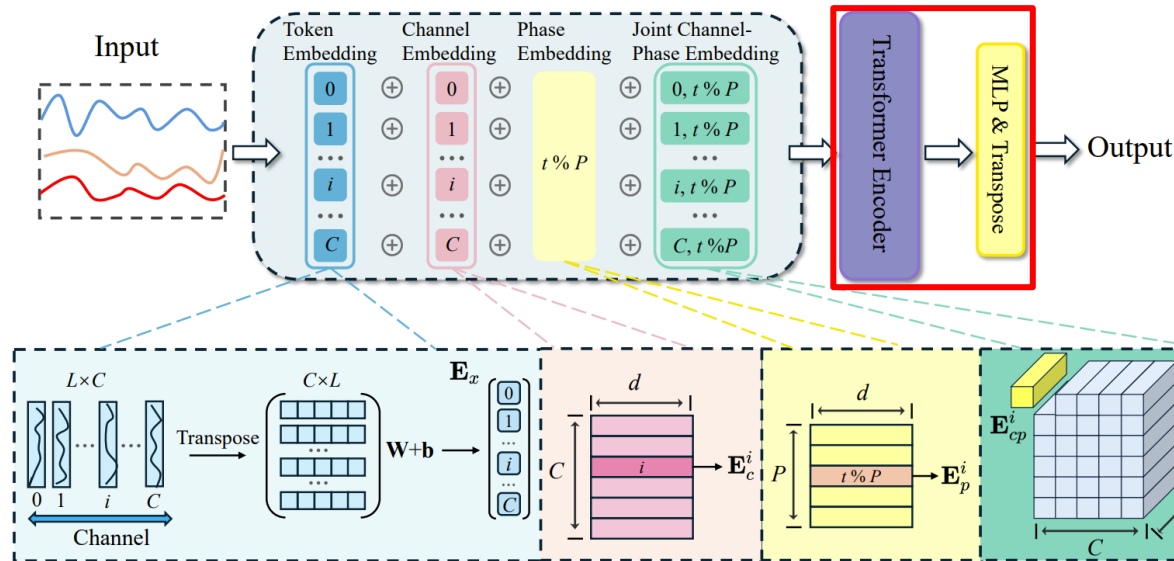
- Joint Channel-Phase Embedding

- 채널과 시간이 복잡하게 얽혀 있어 독립적인 E_c, E_p 만으로는 의존성을 포착하기엔 한계가 있음
- 채널 i 와 위상 t 를 동시에 고려하는 3차원 임베딩 텐서 $\Omega_{cp} \in \mathbb{R}^{C \times P \times d}$ 를 통해 채널과 위상이 교차하는 지점에 있는 고유한 상호작용 패턴을 추출해 더함

$$E_{cp}^i = \text{Lookup}(\Omega_{cp}, i, t \pmod{P}) \in \mathbb{R}^{1 \times d}$$

EMAformer

모델 구조



- Transformer Encoder

- 앞서 생성한 변량 토큰에 보조 임베딩 E_c, E_p, E_{cp} 을 Element-wise Sum으로 더하여 인코더 입력 Z_0 생성

$$Z_0 = E_x + E_c + E_p + E_{cp} \in \mathbb{R}^{C \times d}$$

- 인코더 출력값을 MLP 및 Transpose하여 최종 예측값 생성

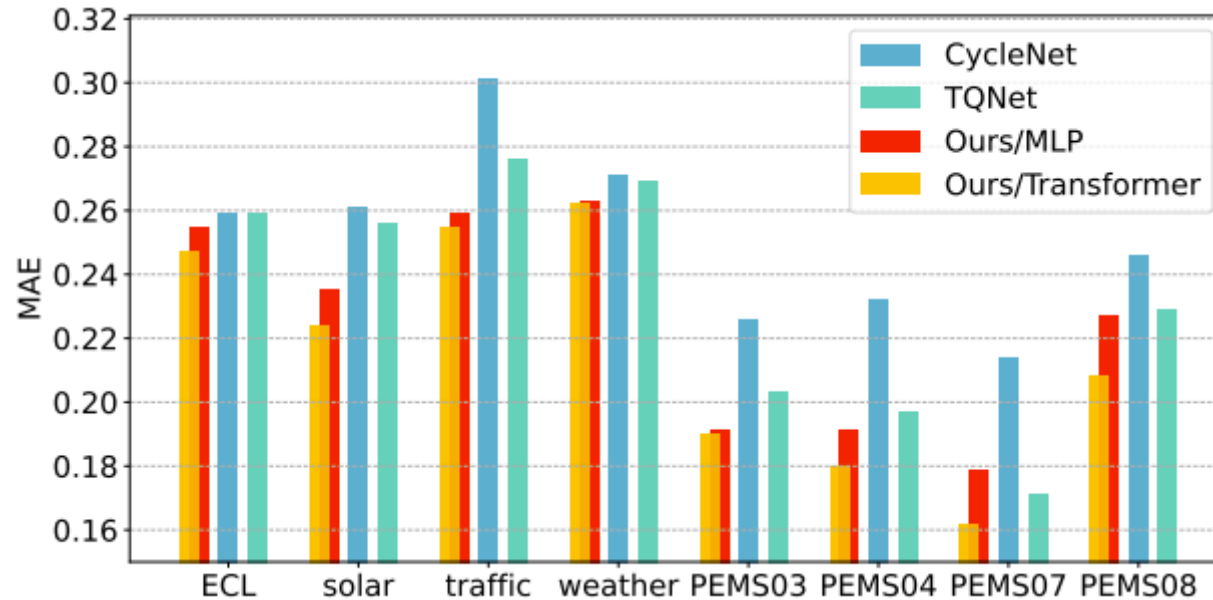
주요 모델 성능 비교

Model	EMAformer (Ours)		TQNet (2025)		TimeXer (2024d)		CycleNet (2024b)		iTransformer (2024c)		TimesNet (2023)		PatchTST (2023)		Crossformer (2023)		DLinear (2023)		SCINet (2022)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	0.432	0.424	0.441	<u>0.434</u>	<u>0.437</u>	0.437	0.457	0.441	0.454	0.448	0.458	0.450	0.469	0.455	0.529	0.522	0.456	0.452	0.747	0.647
ETTh2	<u>0.373</u>	0.395	0.378	0.402	0.368	<u>0.396</u>	0.388	0.409	0.383	0.407	0.414	0.427	0.387	0.407	0.942	0.684	0.559	0.515	0.954	0.723
ETTm1	0.381	0.380	0.377	<u>0.393</u>	0.382	0.397	<u>0.379</u>	0.396	0.407	0.410	0.400	0.406	0.387	0.400	0.513	0.495	0.403	0.407	0.486	0.481
ETTm2	<u>0.271</u>	0.313	0.277	0.323	0.274	0.322	0.266	<u>0.314</u>	0.288	0.332	0.291	0.333	0.281	0.326	0.757	0.611	0.350	0.401	0.571	0.537
ECL	0.158	0.247	<u>0.164</u>	<u>0.259</u>	0.171	0.270	0.168	<u>0.259</u>	0.178	0.270	0.193	0.295	0.205	0.290	0.244	0.334	0.212	0.300	0.571	0.537
Solar	0.197	0.224	<u>0.198</u>	<u>0.256</u>	0.237	0.302	0.210	<u>0.261</u>	0.233	0.262	0.301	0.319	0.270	0.307	0.641	0.639	0.330	0.401	0.282	0.375
Traffic	<u>0.430</u>	0.255	0.445	<u>0.276</u>	0.466	0.287	0.472	0.301	0.428	0.282	0.620	0.336	0.481	0.300	0.550	0.304	0.625	0.383	0.804	0.509
Weather	0.240	0.262	0.242	<u>0.269</u>	<u>0.241</u>	0.271	0.243	0.271	0.258	0.278	0.259	0.287	0.259	0.273	0.259	0.315	0.265	0.317	0.292	0.363
PEMS03	0.089	0.190	<u>0.097</u>	<u>0.203</u>	0.112	0.214	0.118	0.226	0.113	0.222	0.147	0.248	0.180	0.291	0.169	0.282	0.278	0.375	0.114	0.224
PEMS04	0.081	0.180	<u>0.091</u>	<u>0.197</u>	0.105	0.209	0.119	0.232	0.111	0.221	0.129	0.241	0.195	0.307	0.209	0.314	0.295	0.388	0.093	0.202
PEMS07	0.073	0.162	<u>0.075</u>	<u>0.171</u>	0.085	0.182	0.113	0.214	0.101	0.204	0.125	0.226	0.211	0.303	0.235	0.315	0.329	0.396	0.119	0.217
PEMS08	0.128	0.208	<u>0.142</u>	0.229	0.175	0.250	0.150	0.246	0.150	<u>0.226</u>	0.193	0.271	0.280	0.321	0.268	0.307	0.379	0.416	0.159	0.244
Count	20		1		1		1		1		0		0		0		0		0	

- 최신 MLP 모델인 TQNet이나 기존 Transformer 모델들보다 더 뛰어난 성능을 보임

EMAformer

데이터셋별 MAE 성능 비교



- MLP 기반 모델보다 EMAformer의 MAE 성능이 더 뛰어난 것을 확인하였음

핵심 모듈별 성능 비교

		Reformer		Informer		Flowformer	
Models Metric		MSE	MAE	MSE	MAE	MSE	MAE
ECL	Original	0.338	0.422	0.311	0.397	0.267	0.359
	+Inverted	0.208	0.301	0.216	0.311	0.21	0.293
	+Embedding	0.186	0.269	0.186	0.268	0.187	0.269
	Promotion	10.58%	10.63%	13.89%	13.83%	10.95%	8.19%
Traffic	Original	0.741	0.422	0.764	0.416	0.75	0.421
	+Inverted	0.647	0.37	0.662	0.38	0.524	0.355
	+Embedding	0.494	0.311	0.496	0.312	0.49	0.31
	Promotion	23.65%	15.95%	25.08%	17.89%	6.49%	12.68%
Weather	Original	0.803	0.656	0.634	0.548	0.286	0.308
	+Inverted	0.248	0.292	0.271	0.33	0.266	0.285
	+Embedding	0.245	0.265	0.241	0.263	0.244	0.265
	Promotion	1.21%	9.25%	11.07%	20.30%	8.27%	7.02%

- 단순히 Inverted한 것보다 Embedding까지 적용했을 때 모든 모델에서 성능 향상이 이뤄진 것을 확인

실험 세팅

- 사용한 모델: EMAformer
- 재현 실험 데이터셋: ETTh1

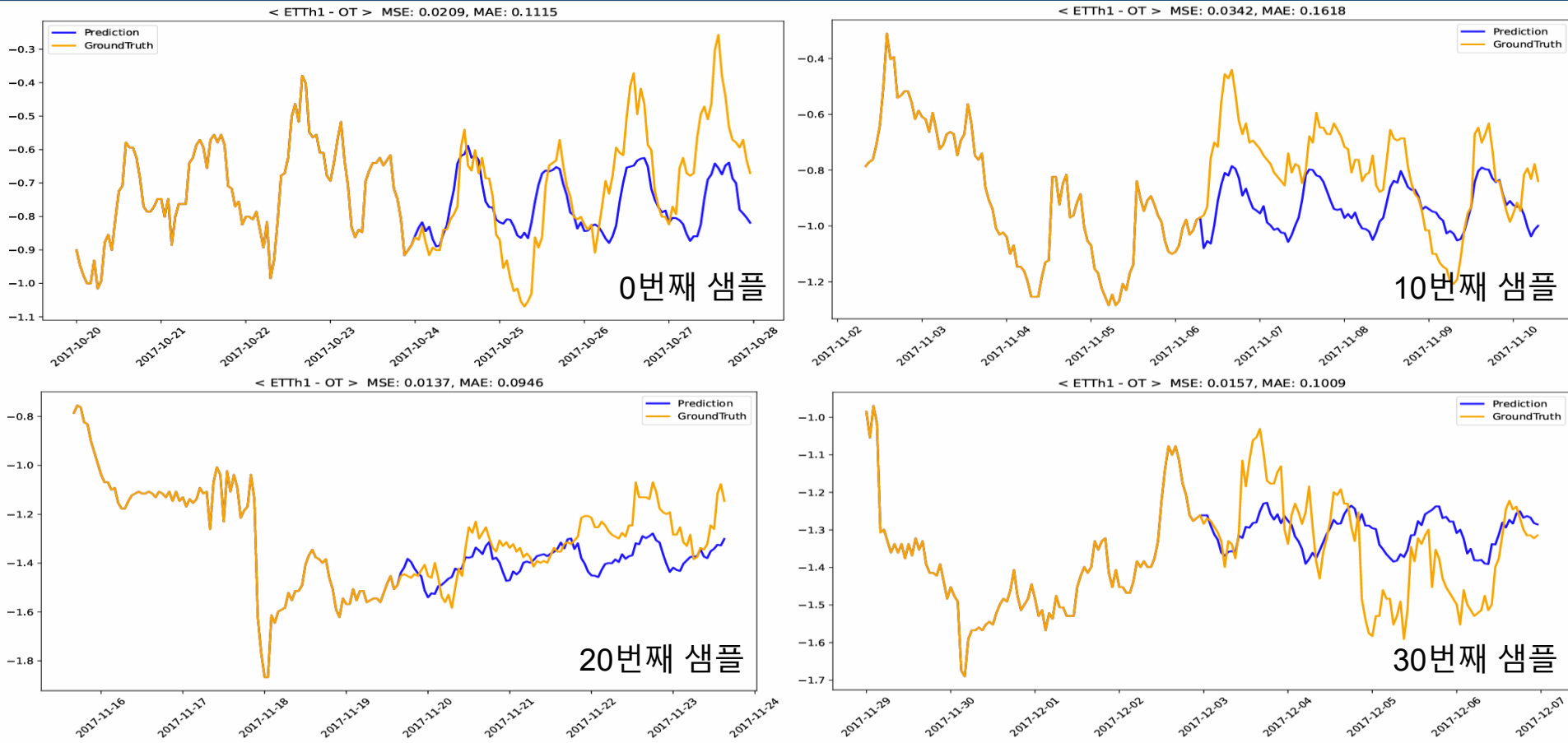
Experiment	ETTh1
Learning rate	10^{-4}
Epoch	10
Batch size	32
Loss function	L1 Loss
Seq_len	96
Pred_len	96
d_model	128
d_ff	128
Cycle	24

재현 실험 (ETTh1)

- 어느 정도의 오차가 있지만 논문 수치와 비슷하게 나온 것을 확인하였음

	ETTh1 Paper		ETTh1 Reproduction	
Pred len	MSE	MAE	MSE	MAE
96	0.374	0.390	0.384	0.401
192	0.428	0.419	0.434	0.429
336	0.469	0.439	0.482	0.451
720	0.456	0.450	0.472	0.464

EMAformer 재현 실험 시각화 (ETTh1)



Seq_len → 96
Pred_len → 96