

2026 01 28
발표 자료

광운대학교 로봇학과
FAIR Lab

김한서

이번 주 진행사항

- EntroPE
 - 논문 리뷰

ENTROPE: ENTROPY-GUIDED DYNAMIC PATCH ENCODER FOR TIME SERIES FORECASTING

Sachith Abeywickrama^{1,2} Emadeldeen Eldele^{3,2} Min Wu² Xiaoli Li^{2,4} Chau Yuen¹

¹School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore.

²Institute for Infocomm Research, A*STAR, Singapore.

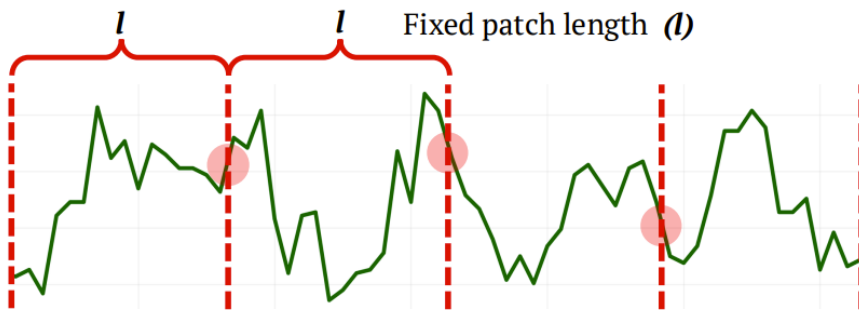
³Department of Computer Science, Khalifa University, UAE.

⁴Information Systems Technology and Design, Singapore University of Technology and Design, Singapore.

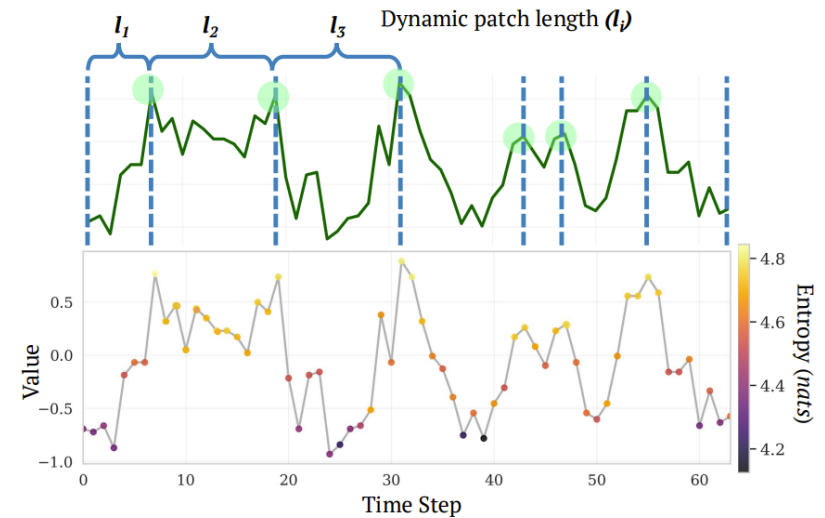
e240203@e.ntu.edu.sg, {emad0002, chau.yuen}@ntu.edu.sg,
wumin@i2r.a-star.edu.sg, xiaoli_li@sutd.edu.sg

- arXiv 등록일: 2025-09-30
- 인용 수: 0회(Google Scholar, 2026-01-16)
- Under review ICLR 2026
- 링크: <https://arxiv.org/pdf/2509.26157>

Background



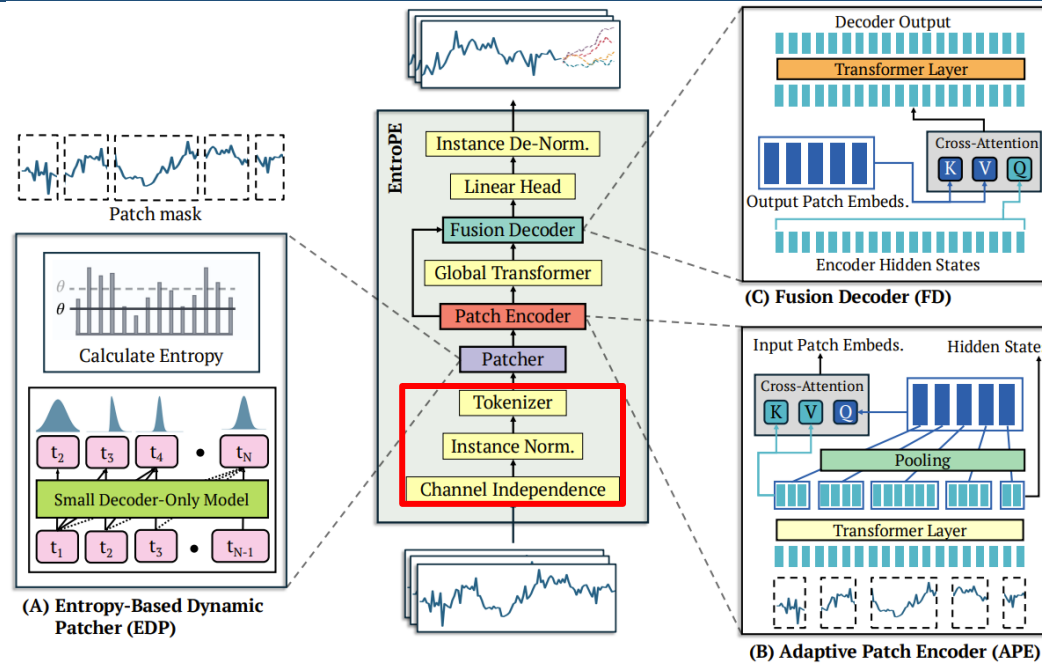
(a) Temporally-agnostic Patching



(b) Temporally-informed Patching

- 기존 방식은 고정 길이 l 로 자르는 방식, 시계열의 중요한 패턴이 잘려 나가는 모습
- EntroPE 방식은 Entropy(불확실성)가 높은 지점인 패턴이 급변하는 전환점을 찾아 Patching하는 모습

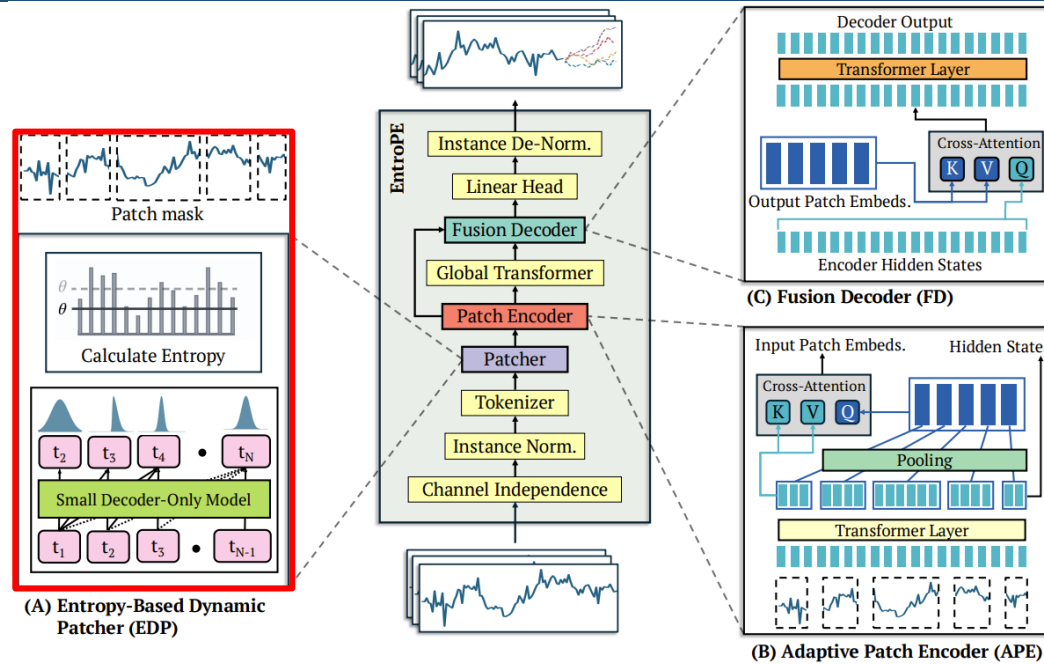
모델 구조



Preprocessing 단계

- Channel Independence로 다변량 시계열의 각 채널을 독립적인 데이터로 취급
- Instance Norm 정규화를 통해 비정상성을 줄여 학습 안정화
- Tokenizer로 시계열 y축을 일정한 간격의 Bin으로 나눈 뒤 각 구간에 고유 번호를 부여해 이산 토큰 x^{tok} 생성

모델 구조



• Patcher(Entropy-Based Dynamic Patcher)

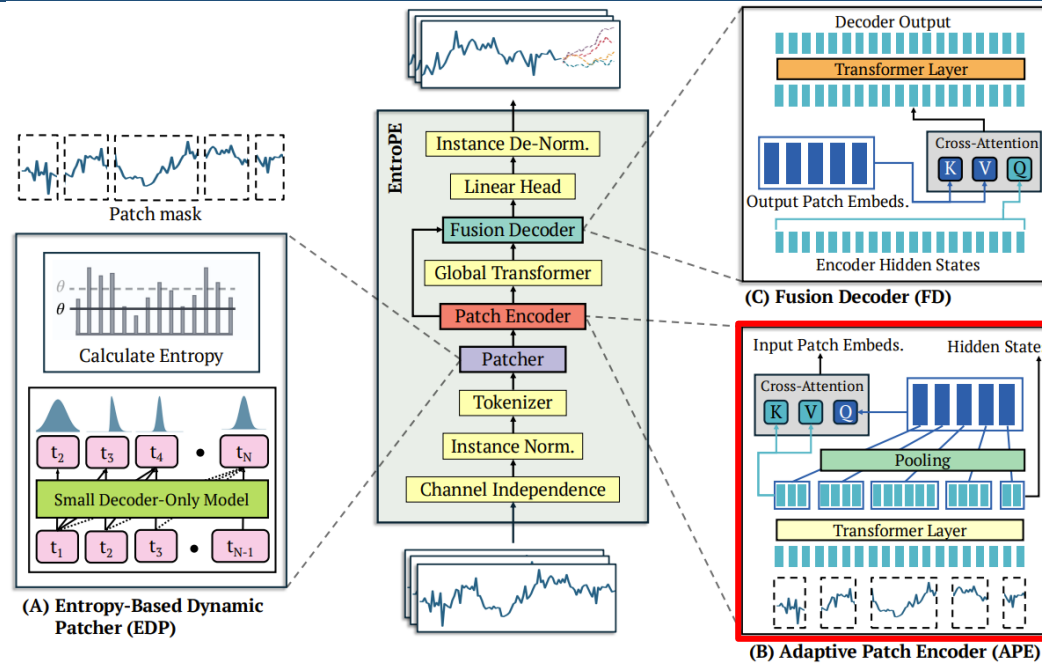
- 먼저 Pre-train 과정, 생성된 이산 토큰을 모델에 입력해 다음 토큰을 예측하는 학습시킴
- 학습된 모델의 가중치를 고정하고 데이터의 각 시점(t)마다 Shannon Entropy를 계산

$$H(x_t) = - \sum_{v \in V} p_{\theta}(x_{t+1} = v | x_{\leq t}) \ln p_{\theta}(x_{t+1} = v | x_{\leq t})$$

- 우측 두 조건을 충족한 경우에만 Patching $H(x_t) > \theta$, $H(x_t) - H(x_{t-1}) > \gamma$

* $H(x_t)$: 시점 t 에서의 Entropy
 * V : 모든 토큰 후보들의 집합
 * p_{θ} : 다음 토큰이 v 일 확률
 * θ : 전역 Entropy 임계값
 * γ : 상대적 Entropy 증가량

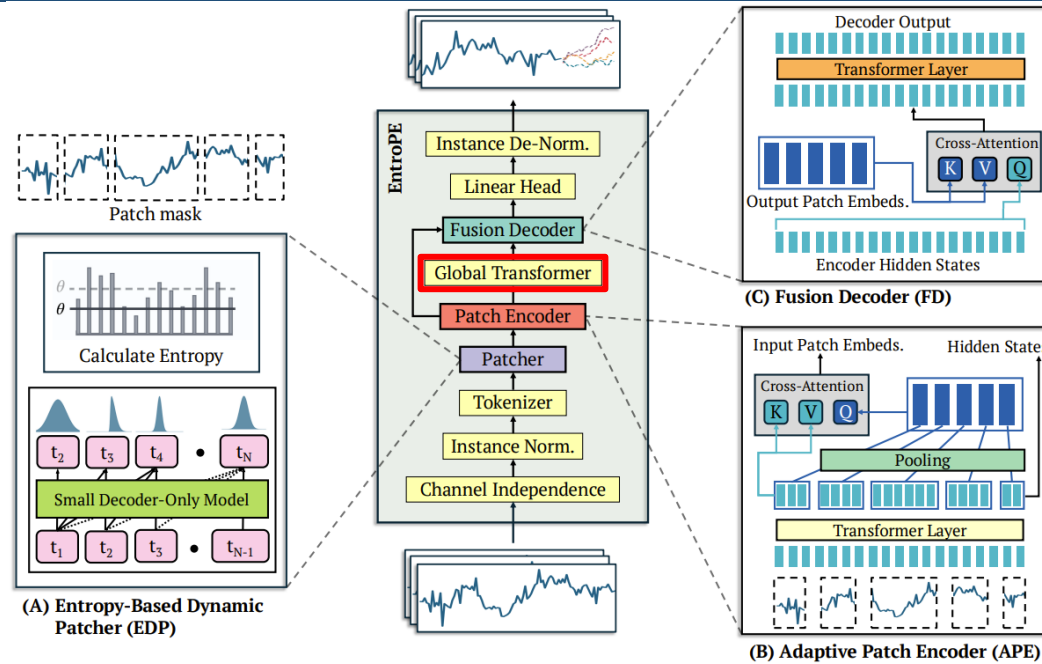
모델 구조



• Patch Encoder(Adaptive Patch Encoder)

- 분할된 각 Patch들(p_j)을 임베딩 벡터로 변환한 뒤 Transformer Layer를 거치게 되는데, 이때 각 시점 임베딩은 자신이 속한 Patch 내부의 시점들하고만 Attention 수행, 나머지는 Masking
- Pooling 전의 임베딩은 Key, Value로, Pooling 후의 임베딩은 Query로 사용

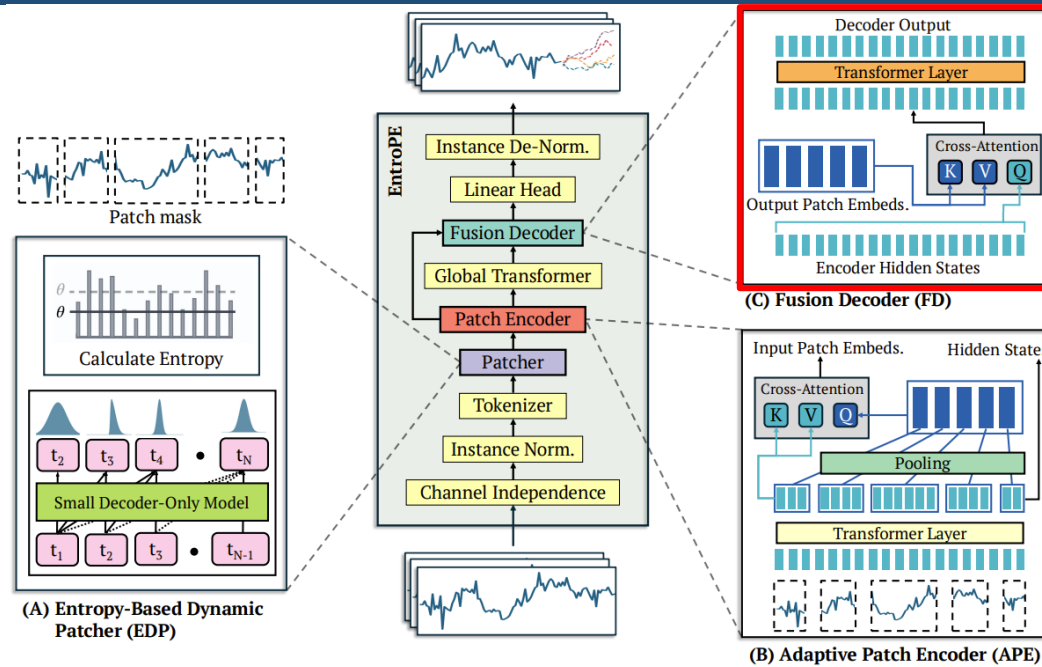
모델 구조



• Global Transformer

- Cross-Attention을 거친 Patch 임베딩을 입력 받아 사용
- APE에서 사용했던 Masking을 없애고, 모든 패치가 서로를 참조하도록 하여 패치 간의 상관관계와 전역적인 맥락을 파악

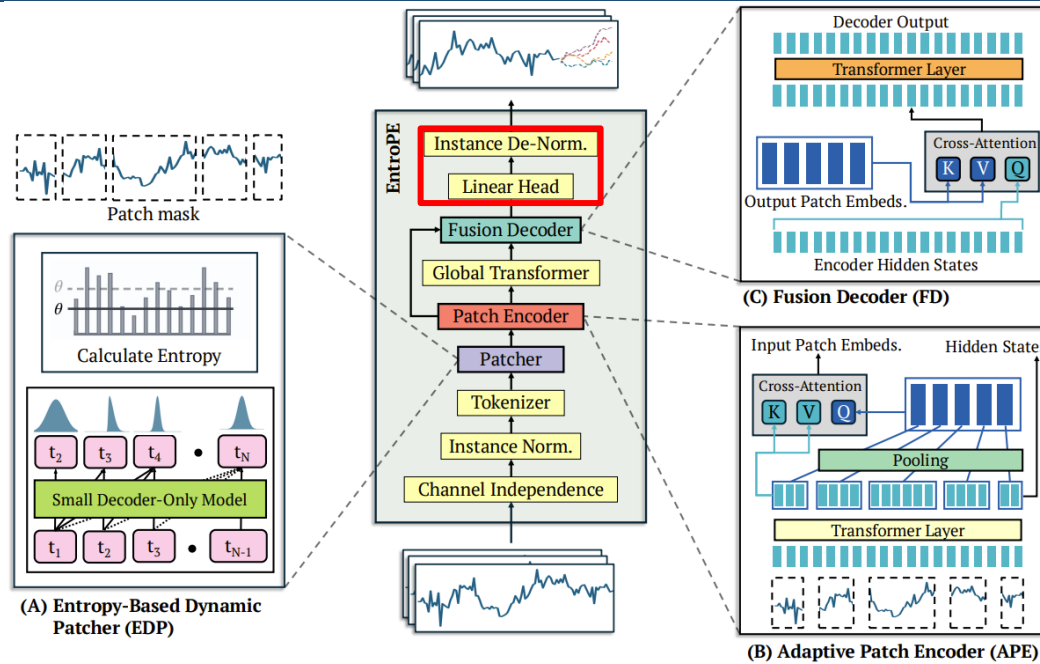
모델 구조



• Fusion Decoder

- Global Transformer의 출력값을 Key, Value로, APE 단계의 Hidden States를 Query로 사용
- Cross-Attention을 통해 각각의 전역적인 추세와 세밀한 지역성을 결합, 이후 Transformer Layer를 거쳐 예측에 최적화된 형태로 변환

모델 구조



• Post-Processing 단계

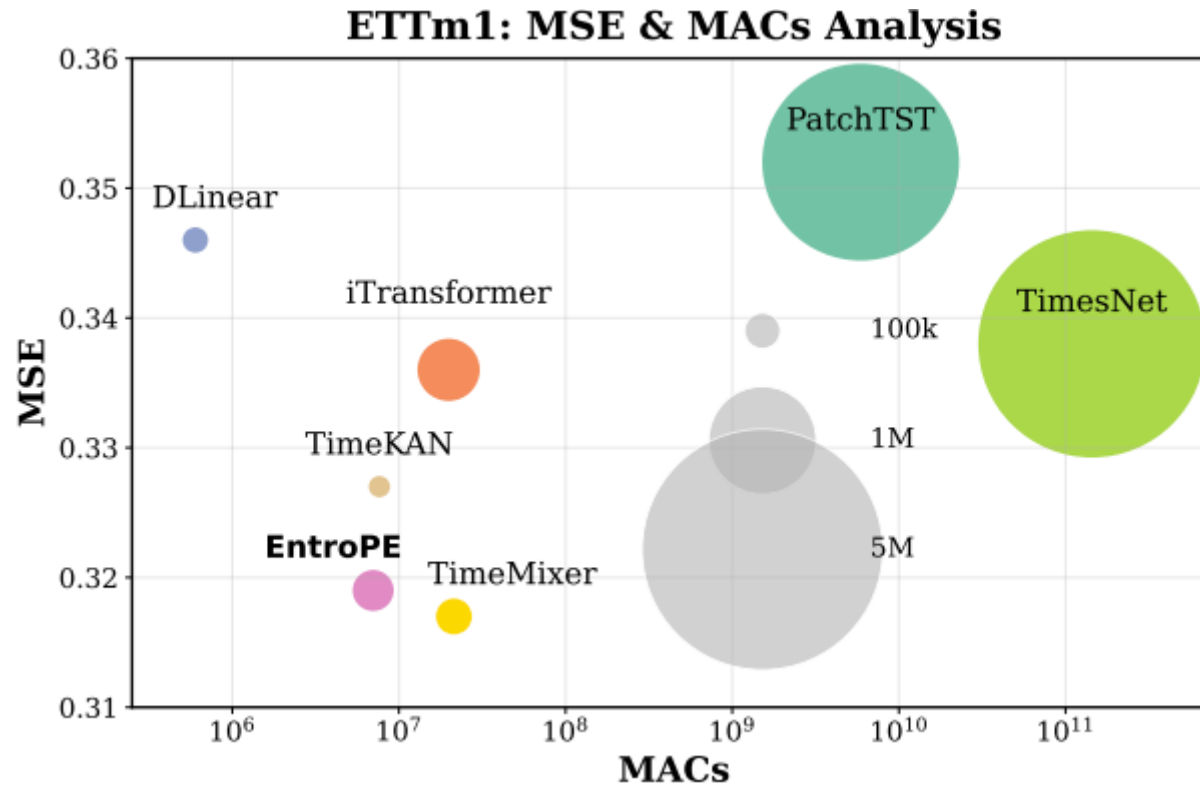
- Decoder의 출력값 $L \times d$ 행렬을 Flatten하여 펼치고, Linear Layer에 통과시켜 예측값 생성
- 생성된 예측값을 Instance De-Norm을 통해 원래 스케일로 복원하여 비정상성 해결

주요 모델 성능 비교

Dataset	T	EntroPE		iTrans.		PatchTST		TimesNet		DLinear		FEDformer		Autoformer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	96	0.317	0.354	0.334	0.368	0.352	0.374	0.338	0.375	0.346	0.374	0.379	0.419	0.505	0.475
	192	0.360	0.379	0.377	0.391	0.390	0.393	0.374	0.387	0.382	0.391	0.426	0.441	0.553	0.496
	336	0.388	0.399	0.426	0.420	0.421	0.414	0.410	0.411	0.415	0.415	0.445	0.459	0.621	0.537
	720	0.446	0.433	0.491	0.459	0.462	0.449	0.478	0.450	0.473	0.451	0.543	0.490	0.671	0.561
	Avg.	0.378	0.391	0.407	0.410	0.406	0.407	0.400	0.406	0.404	0.408	0.448	0.452	0.588	0.517
ETTm2	96	0.180	0.265	0.180	0.264	0.183	0.270	0.187	0.267	0.193	0.293	0.203	0.287	0.255	0.339
	192	0.241	0.311	0.250	0.309	0.255	0.314	0.249	0.309	0.284	0.361	0.269	0.328	0.281	0.340
	336	0.313	0.352	0.311	0.348	0.309	0.347	0.321	0.351	0.382	0.429	0.325	0.366	0.339	0.372
	720	0.411	0.413	0.412	0.407	0.412	0.404	0.408	0.403	0.558	0.525	0.421	0.415	0.433	0.432
	Avg.	0.286	0.335	0.288	0.332	0.290	0.334	0.291	0.333	0.354	0.402	0.305	0.349	0.327	0.371
ETTh1	96	0.375	0.399	0.386	0.405	0.460	0.447	0.384	0.402	0.397	0.412	0.395	0.424	0.449	0.459
	192	0.423	0.425	0.441	0.436	0.512	0.477	0.436	0.429	0.446	0.441	0.469	0.470	0.500	0.482
	336	0.429	0.432	0.487	0.458	0.546	0.496	0.638	0.469	0.489	0.467	0.490	0.477	0.521	0.496
	720	0.439	0.454	0.503	0.491	0.544	0.517	0.521	0.500	0.513	0.510	0.598	0.544	0.514	0.512
	Avg.	0.416	0.425	0.454	0.447	0.516	0.484	0.495	0.450	0.461	0.457	0.498	0.484	0.496	0.487
ETTh2	96	0.281	0.336	0.297	0.349	0.308	0.355	0.340	0.374	0.340	0.394	0.358	0.397	0.346	0.388
	192	0.371	0.393	0.380	0.400	0.393	0.405	0.402	0.414	0.482	0.479	0.429	0.439	0.456	0.452
	336	0.392	0.394	0.428	0.432	0.427	0.436	0.452	0.452	0.591	0.541	0.496	0.487	0.482	0.486
	720	0.421	0.427	0.427	0.445	0.436	0.450	0.462	0.468	0.839	0.661	0.463	0.474	0.515	0.511
	Avg.	0.366	0.387	0.383	0.407	0.391	0.411	0.414	0.427	0.563	0.519	0.437	0.449	0.450	0.459
Weather	96	0.164	0.211	0.174	0.214	0.186	0.227	0.172	0.220	0.195	0.252	0.217	0.296	0.266	0.336
	192	0.210	0.252	0.221	0.254	0.234	0.265	0.219	0.261	0.237	0.295	0.276	0.336	0.307	0.367
	336	0.256	0.290	0.278	0.296	0.284	0.301	0.246	0.337	0.282	0.331	0.339	0.380	0.359	0.395
	720	0.339	0.342	0.358	0.347	0.356	0.349	0.365	0.359	0.345	0.382	0.403	0.428	0.419	0.428
	Avg.	0.242	0.273	0.258	0.278	0.265	0.285	0.251	0.294	0.265	0.315	0.309	0.360	0.338	0.382
Electricity	96	0.163	0.252	0.148	0.240	0.190	0.296	0.168	0.272	0.210	0.302	0.193	0.308	0.201	0.317
	192	0.177	0.268	0.162	0.253	0.199	0.304	0.184	0.322	0.210	0.305	0.201	0.315	0.222	0.334
	336	0.194	0.284	0.178	0.269	0.217	0.319	0.198	0.300	0.223	0.319	0.214	0.329	0.231	0.443
	720	0.235	0.321	0.225	0.317	0.258	0.352	0.220	0.320	0.258	0.350	0.246	0.355	0.254	0.361
	Avg.	0.182	0.271	0.178	0.270	0.216	0.318	0.193	0.304	0.225	0.319	0.214	0.327	0.227	0.338
Exchange	96	0.083	0.140	0.086	0.206	0.088	0.205	0.107	0.234	0.088	0.218	0.148	0.278	0.197	0.323
	192	0.177	0.235	0.177	0.299	0.176	0.299	0.226	0.344	0.176	0.315	0.271	0.315	0.300	0.369
	336	0.299	0.371	0.331	0.417	0.301	0.397	0.367	0.448	0.313	0.427	0.460	0.427	0.509	0.524
	720	0.761	0.869	0.847	0.691	0.901	0.714	0.964	0.746	0.839	0.695	1.195	0.695	1.447	0.941
	Avg.	0.331	0.386	0.360	0.403	0.367	0.404	0.416	0.443	0.354	0.414	0.519	0.429	0.613	0.539

- 앞서 확인한 효율성 비교 결과와 동일하게 EntroPE의 성능이 가장 뛰어난 것을 확인

모델별 효율성 비교



- 기존 모델들과 비교했을 때, 파라미터 수가 적음에도 대부분의 모델보다 성능이 더 뛰어난 모습을 보임

Ablation study

Dataset	T	Full	-EDP			-EDP -APE	-EDP -APE -FD	Baselines	
		Dynamic	Static(1)	Static(8)	Static(16)	Pool + FD	Pool + Flat	PatchTST	iTrans.
ETTh1	336	0.429	0.425	0.441	0.444	0.438	0.519	0.501	0.487
	720	0.439	0.460	0.469	0.477	0.461	0.527	0.503	0.500
ETTh2	336	0.355	0.428	0.435	0.439	0.439	0.462	0.427	0.428
	720	0.397	0.448	0.470	0.463	0.466	0.456	0.436	0.427
ETTm1	336	0.393	0.401	0.409	0.415	0.402	0.427	0.421	0.426
	720	0.445	0.451	0.459	0.452	0.457	0.483	0.462	0.491
Weather	336	0.258	0.261	0.265	0.270	0.262	0.262	0.284	0.278
	720	0.341	0.343	0.353	0.359	0.347	0.346	0.356	0.358

- EntroPE의 핵심 모듈들을 모두 적용했을 때가 성능 향상 폭이 가장 큰 것을 확인