

2025 10 30
발표 자료

광운대학교 로봇학과
FAIR Lab

김한서

이번 주 진행사항

- PatchTST
 - 논문 리뷰
 - 실험 세팅 및 결과
 - 시각화
 - DLinear와 성능 비교
 - 결과 정리

PatchTST (Patch Time-Series Transformer)

A TIME SERIES IS WORTH 64 WORDS: LONG-TERM FORECASTING WITH TRANSFORMERS

Yuqi Nie^{1*}, Nam H. Nguyen²*, Phanwadee Sinthong², Jayant Kalagnanam²

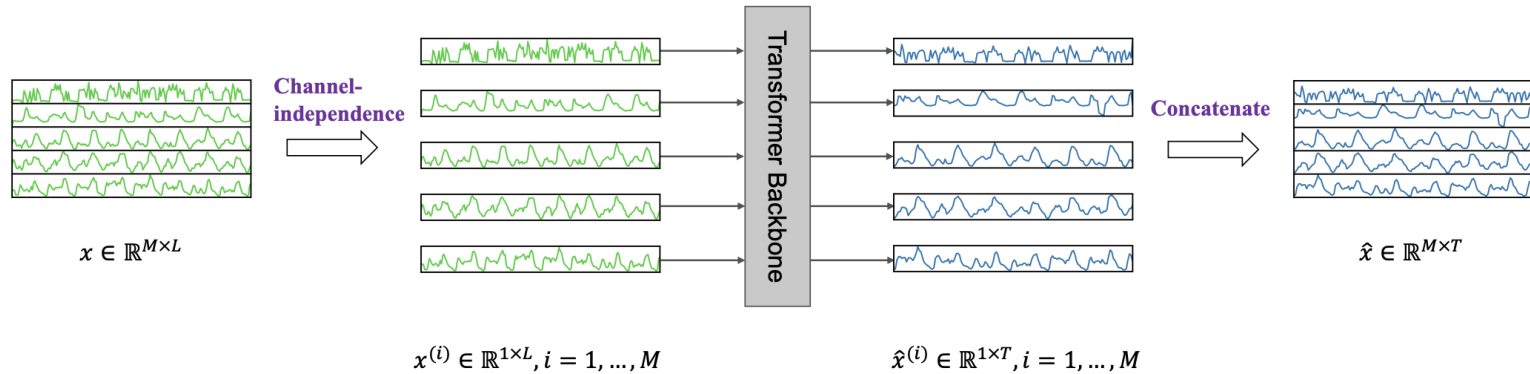
¹Princeton University ²IBM Research

`ynie@princeton.edu, nnguyen@us.ibm.com, Gift.Sinthong@ibm.com, jayant@us.ibm.com`

- arXiv 등록일: 2022-11-27
- 인용 수: 2956회(Google Scholar, 2025-10-27)
- Time Series Library 등록
- Published at ICLR 2023
- 링크: <https://arxiv.org/abs/2211.14730>

PatchTST

Channel-Independence



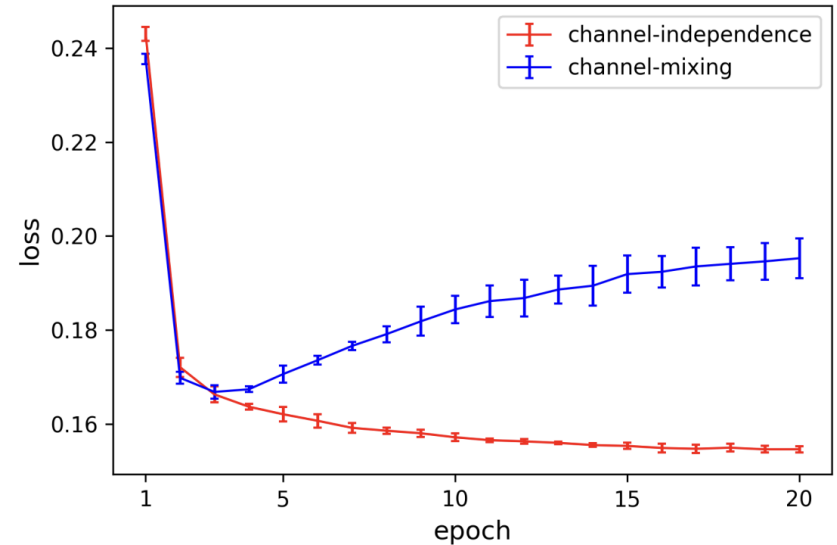
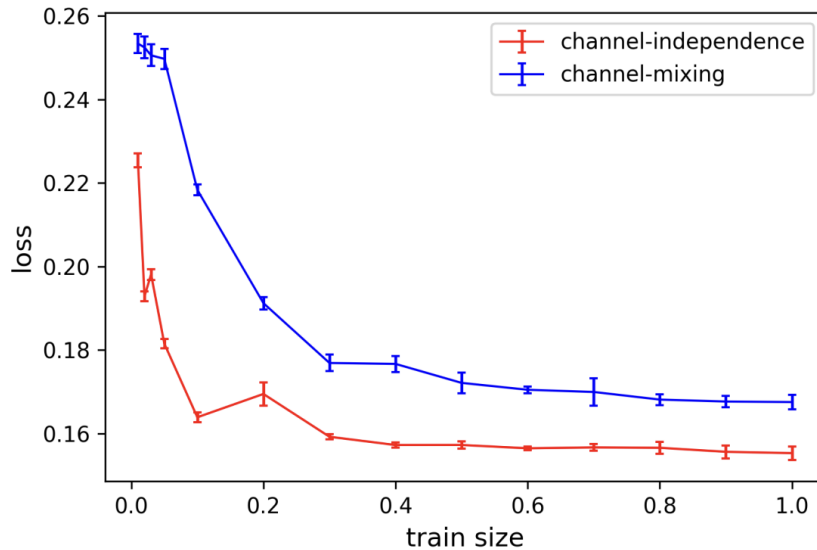
(a) PatchTST Model Overview

Channel-Independence

- 다변량 시계열 데이터에서 각 채널을 분리하여 분석하겠다는 개념, 분석 후 예측 수행
- 분리된 채널들을 동일한 가중치를 공유하는 Transformer Backbone에 통과시키고 나온 채널들을 다시 결합

PatchTST

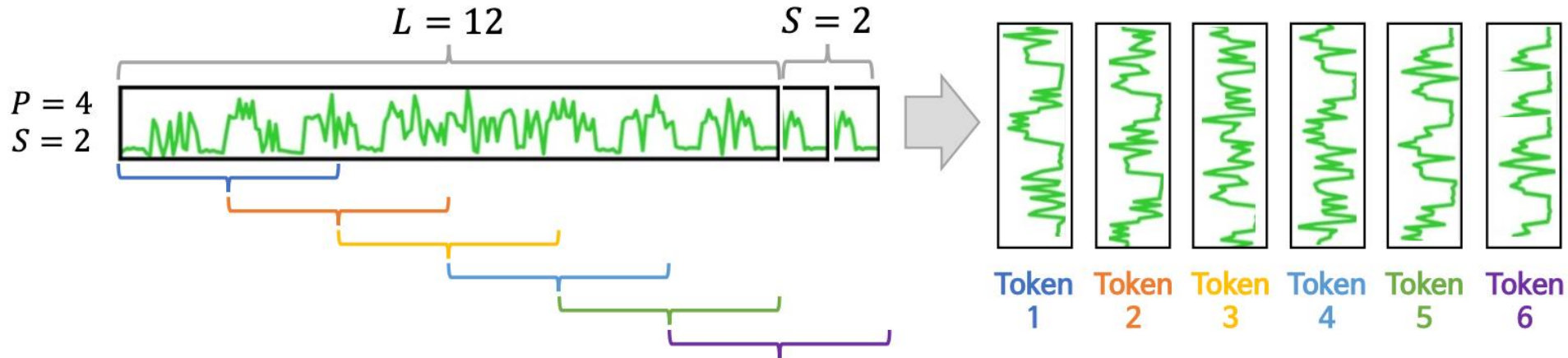
Channel-Independence, Channel-mixing 비교



- Channel-mixing은 여러 채널들을 혼합해 채널들 간의 복잡한 관계를 학습 가능한 반면, Channel-Independence 방식은 여러 종류의 채널들이 동일한 모델을 통과하며 가중치 값들을 공유하므로 채널들 간 상호작용 포착이 어려워 성능이 떨어질 것으로 보았지만, Channel-mixing보다 나은 성능을 보여줌

PatchTST

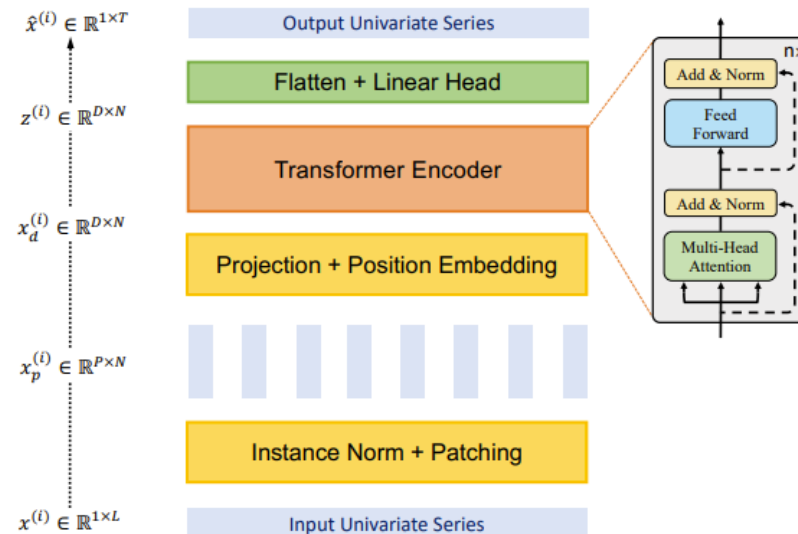
Patching



• Patching

- 인접한 데이터들을 묶어 하나의 단위로 만드는 것
- Input token의 개수를 줄여 시간적, 공간적 복잡도를 낮춤으로써 더 긴 시계열 데이터 학습 가능, 이를 통해 더 나은 예측 결과를 얻음
- 시계열 데이터를 일정한 크기의 패치 단위로 나누어 입력으로 사용하여 학습함으로써 지역 정보 보존

Transformer Backbone (Supervised)



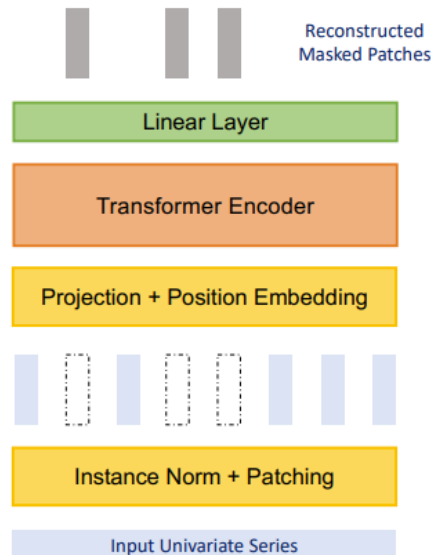
(b) Transformer Backbone (Supervised)

- Supervised(지도학습)

- Channel-Independence로 분리된 채널들을 각 입력으로 사용, 입력들이 Transformer Backbone으로 들어간 뒤 각 분리된 채널에 대한 예측 값을 도출
- Self-supervised와는 달리 마스킹 없이 모든 patch를 입력으로 사용
- Self-supervised에서 학습된 모델을 가져와 실제 예측에 사용하는 단계

PatchTST

Transformer Backbone (Self-supervised)



(c) Transformer Backbone (Self-supervised)

- Self-supervised(자기지도 학습)

- 라벨링이 되어있지 않은 데이터에서 의미 있는 정보를 추출하기 위한 기법
- 일부 patch를 무작위로 선택해 마스킹 하고, MSE 손실을 사용해 모델이 마스킹 된 patch를 복원하도록 학습시킴
- 이 과정을 통해 학습된 모델은 한 채널 내에서 일어나는 패턴과 관계를 더 깊이 이해할 수 있게 됨
- 모델이 시계열의 패턴과 특징을 미리 학습하게 하는 사전 학습 단계

PatchTST

주요 모델 성능 비교

Models	Metric	PatchTST/64		PatchTST/42		DLinear		FEDformer		Autoformer		Informer		Pyraformer		LogTrans	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weather	96	0.149	0.198	<u>0.152</u>	<u>0.199</u>	0.176	0.237	0.238	0.314	0.249	0.329	0.354	0.405	0.896	0.556	0.458	0.490
	192	0.194	0.241	<u>0.197</u>	<u>0.243</u>	0.220	0.282	0.275	0.329	0.325	0.370	0.419	0.434	0.622	0.624	0.658	0.589
	336	0.245	0.282	<u>0.249</u>	<u>0.283</u>	0.265	0.319	0.339	0.377	0.351	0.391	0.583	0.543	0.739	0.753	0.797	0.652
	720	0.314	0.334	<u>0.320</u>	<u>0.335</u>	0.323	0.362	0.389	0.409	0.415	0.426	0.916	0.705	1.004	0.934	0.869	0.675
Traffic	96	0.360	0.249	<u>0.367</u>	<u>0.251</u>	0.410	0.282	0.576	0.359	0.597	0.371	0.733	0.410	2.085	0.468	0.684	0.384
	192	0.379	0.256	<u>0.385</u>	<u>0.259</u>	0.423	0.287	0.610	0.380	0.607	0.382	0.777	0.435	0.867	0.467	0.685	0.390
	336	0.392	0.264	<u>0.398</u>	<u>0.265</u>	0.436	0.296	0.608	0.375	0.623	0.387	0.776	0.434	0.869	0.469	0.734	0.408
	720	0.432	0.286	<u>0.434</u>	<u>0.287</u>	0.466	0.315	0.621	0.375	0.639	0.395	0.827	0.466	0.881	0.473	0.717	0.396
Electricity	96	0.129	0.222	<u>0.130</u>	<u>0.222</u>	0.140	0.237	0.186	0.302	0.196	0.313	0.304	0.393	0.386	0.449	0.258	0.357
	192	0.147	0.240	<u>0.148</u>	<u>0.240</u>	0.153	0.249	0.197	0.311	0.211	0.324	0.327	0.417	0.386	0.443	0.266	0.368
	336	0.163	0.259	<u>0.167</u>	<u>0.261</u>	0.169	0.267	0.213	0.328	0.214	0.327	0.333	0.422	0.378	0.443	0.280	0.380
	720	0.197	0.290	<u>0.202</u>	<u>0.291</u>	0.203	0.301	0.233	0.344	0.236	0.342	0.351	0.427	0.376	0.445	0.283	0.376
ILI	24	1.319	0.754	<u>1.522</u>	<u>0.814</u>	2.215	1.081	2.624	1.095	2.906	1.182	4.657	1.449	1.420	2.012	4.480	1.444
	36	1.579	0.870	<u>1.430</u>	<u>0.834</u>	1.963	0.963	2.516	1.021	2.585	1.038	4.650	1.463	7.394	2.031	4.799	1.467
	48	1.553	0.815	<u>1.673</u>	<u>0.854</u>	2.130	1.024	2.505	1.041	3.024	1.145	5.004	1.542	7.551	2.057	4.800	1.468
	60	1.470	0.788	<u>1.529</u>	<u>0.862</u>	2.368	1.096	2.742	1.122	2.761	1.114	5.071	1.543	7.662	2.100	5.278	1.560
ETTh1	96	0.370	0.400	<u>0.375</u>	<u>0.399</u>	<u>0.375</u>	<u>0.399</u>	0.376	0.415	0.435	0.446	0.941	0.769	0.664	0.612	0.878	0.740
	192	<u>0.413</u>	<u>0.429</u>	<u>0.414</u>	<u>0.421</u>	0.405	0.416	0.423	0.446	0.456	0.457	1.007	0.786	0.790	0.681	1.037	0.824
	336	0.422	0.440	<u>0.431</u>	0.436	0.439	0.443	0.444	0.462	0.486	0.487	1.038	0.784	0.891	0.738	1.238	0.932
	720	0.447	<u>0.468</u>	<u>0.449</u>	0.466	0.472	0.490	0.469	0.492	0.515	0.517	1.144	0.857	0.963	0.782	1.135	0.852
ETTh2	96	0.274	<u>0.337</u>	0.274	0.336	0.289	0.353	0.332	0.374	0.332	0.368	1.549	0.952	0.645	0.597	2.116	1.197
	192	<u>0.341</u>	<u>0.382</u>	0.339	0.379	0.383	0.418	0.407	0.446	0.426	0.434	3.792	1.542	0.788	0.683	4.315	1.635
	336	0.329	<u>0.384</u>	<u>0.331</u>	0.380	0.448	0.465	0.400	0.447	0.477	0.479	4.215	1.642	0.907	0.747	1.124	1.604
	720	0.379	0.422	0.379	0.422	0.605	0.551	0.412	0.469	0.453	0.490	3.656	1.619	0.963	0.783	3.188	1.540
ETTM1	96	<u>0.293</u>	0.346	0.290	0.342	0.299	<u>0.343</u>	0.326	0.390	0.510	0.492	0.626	0.560	0.543	0.510	0.600	0.546
	192	<u>0.333</u>	0.370	0.332	<u>0.369</u>	0.335	0.365	0.365	0.415	0.514	0.495	0.725	0.619	0.557	0.537	0.837	0.700
	336	<u>0.369</u>	<u>0.392</u>	0.366	<u>0.392</u>	<u>0.369</u>	0.386	0.392	0.425	0.510	0.492	1.005	0.741	0.754	0.655	1.124	0.832
	720	0.416	0.420	<u>0.420</u>	<u>0.424</u>	<u>0.425</u>	<u>0.421</u>	0.446	0.458	0.527	0.493	1.133	0.845	0.908	0.724	1.153	0.820
ETTM2	96	<u>0.166</u>	<u>0.256</u>	0.165	0.255	0.167	0.260	0.180	0.271	0.205	0.293	0.355	0.462	0.435	0.507	0.768	0.642
	192	<u>0.223</u>	<u>0.296</u>	0.220	0.292	0.224	0.303	0.252	0.318	0.278	0.336	0.595	0.586	0.730	0.673	0.989	0.757
	336	0.274	0.329	<u>0.278</u>	<u>0.329</u>	0.281	0.342	0.324	0.364	0.343	0.379	1.270	0.871	1.201	0.845	1.334	0.872
	720	0.362	0.385	<u>0.367</u>	<u>0.385</u>	0.397	0.421	0.410	0.420	0.414	0.419	3.001	1.267	3.625	1.451	3.048	1.328

- PatchTST 기반 모델이 DLinear를 포함한 다른 모델들보다 전반적으로 좋은 성능을 보여줌
- 특히 Weather, Traffic, Electricity와 같은 대규모 데이터셋에서의 성능 향상이 두드러짐

실험 세팅

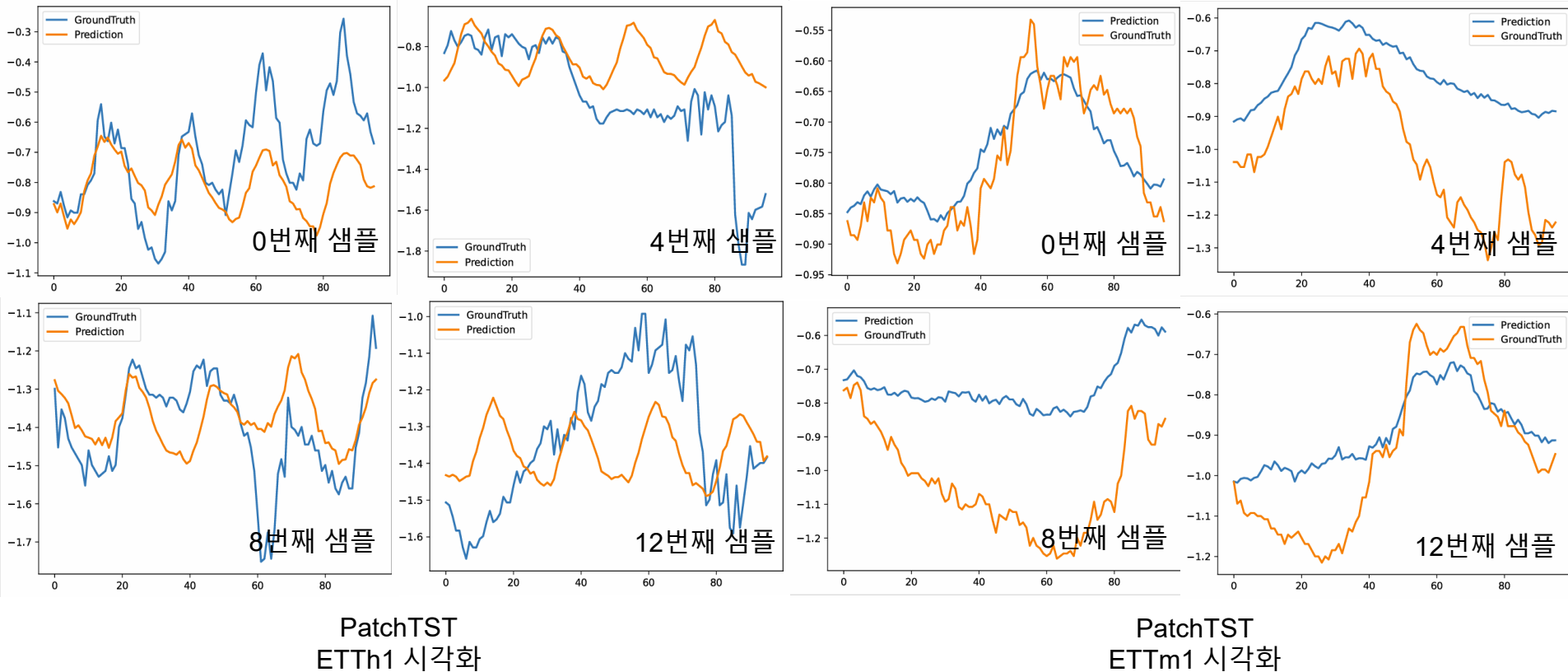
- 사용한 모델: PatchTST
- 재현 실험 데이터셋: ETTh1, ETTm1

Experiment	ETTh1, ETTm1
Learning rate	10^{-4}
Epoch	100
Batch size	128
Loss function	MSE Loss
Sequence Length	512
input_feature	7
Pred_len	96/192/336/720

PatchTST 재현 실험 (ETTh1, ETTm1)

- ETTh1, ETTm1 모두 모든 예측 길이에서 재현 결과가 논문과 차이가 거의 나지 않는다는 것을 확인하였음

	ETTh1		ETTh1		ETTm1		ETTm1	
	Paper		Reproduction		Paper		Reproduction	
Pred len	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
96	0.370	0.400	0.370	0.400	0.293	0.346	0.305	0.358
192	0.413	0.429	0.413	0.429	0.333	0.370	0.339	0.377
336	0.422	0.440	0.421	0.439	0.369	0.392	0.368	0.394
720	0.447	0.468	0.447	0.468	0.416	0.420	0.414	0.425



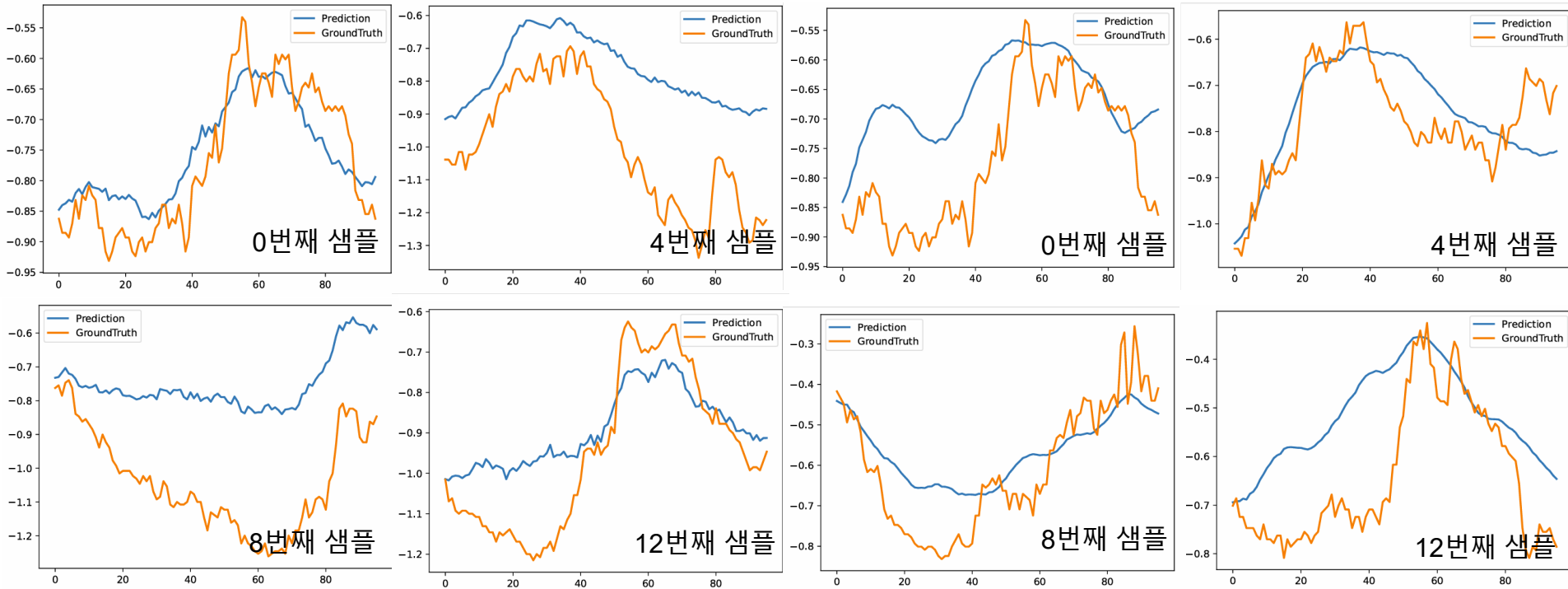
PatchTST

PatchTST, DLinear 성능 비교 (ETTh1, ETTm1)

- ETTh1과 ETTm1 두 데이터셋의 모든 예측 길이에서 PatchTST가 DLinear보다 더 좋은 성능을 보임
- 예측 길이가 길어질수록 성능 차이가 두드러지는 것을 확인

Pred len	PatchTST ETTh1		DLinear ETTh1		PatchTST ETTm1		DLinear ETTm1	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
96	0.370	0.400	0.396	0.410	0.305	0.358	0.345	0.373
192	0.413	0.429	0.445	0.440	0.339	0.377	0.381	0.391
336	0.421	0.439	0.487	0.465	0.368	0.394	0.415	0.415
720	0.447	0.468	0.512	0.510	0.414	0.425	0.472	0.450

PatchTST, DLinear 시각화 (ETTm1)



PatchTST
ETTm1 시각화

DLinear
ETTm1 시각화

실험 결과 정리

- Transformer 계산 방식 Point wise 대신 PatchTST의 Channel-Independence, Patching을 통해 Patch 단위로 연산하여 학습 속도 향상 및 일반화 성능 향상을 확인
- 시각화 결과, ETTm1의 예측값은 정답값을 전체적으로 잘 따라간 반면 ETTh1은 일부 구간에서 발생한 오차를 시각적으로 확인 가능하였음
- DLinear와의 성능 비교 결과, PatchTST가 더 좋은 성능을 보였으며, 이를 통해 PatchTST가 장기 예측에 더 효과적임을 알 수 있음