

2025 11 06
발표 자료

광운대학교 로봇학과
FAIR Lab

김한서

이번 주 진행사항

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

iTransformer

ITRANSFORMER: INVERTED TRANSFORMERS ARE EFFECTIVE FOR TIME SERIES FORECASTING

Yong Liu,* Tengge Hu,* Haoran Zhang,* Haixu Wu, Shiyu Wang[§], Lintao Ma[§], Mingsheng Long[✉]

School of Software, BNRIst, Tsinghua University, Beijing 100084, China

[§]Ant Group, Hangzhou, China

{liuyong21, htg21, z-hr20, whx20}@mails.tsinghua.edu.cn

{weiming.wsy, lintao.mlt}@antgroup.com, mingsheng@tsinghua.edu.cn

- arXiv 등록일: 2023-10-10
- 인용 수: 1880회(Google Scholar, 2025-11-01)
- Time Series Library 등록
- Published at ICLR 2024
- 링크: <https://arxiv.org/abs/2310.06625>

iTransformer

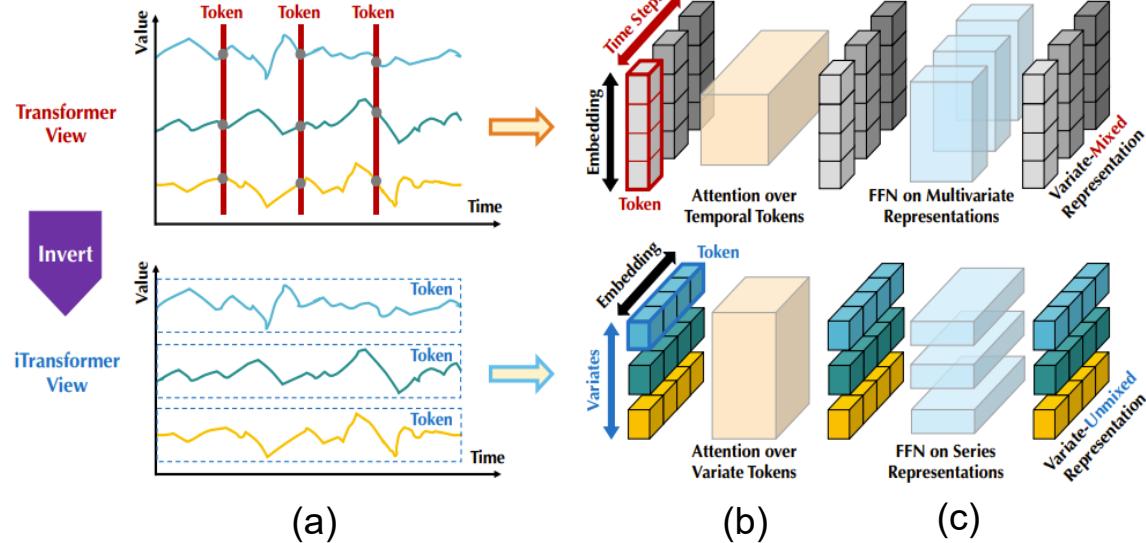
등장 배경

- 기존 Transformer 기반 모델 문제점 2가지
 - 입력 길이에 따라 연산량이 2차적으로 증가
 - 시간을 토큰화하여 변수 간 정보가 혼합되어 다변량 시계열 데이터에서 변수 간 관계를 제대로 학습하지 못함

→ iTransformer는 기존 Transformer 모델과는 달리 시간 축이 아닌 변수 축을 토큰 단위로 처리하여, 변수 간 관계를 명시적으로 학습하고 연산 효율성 개선

iTransformer

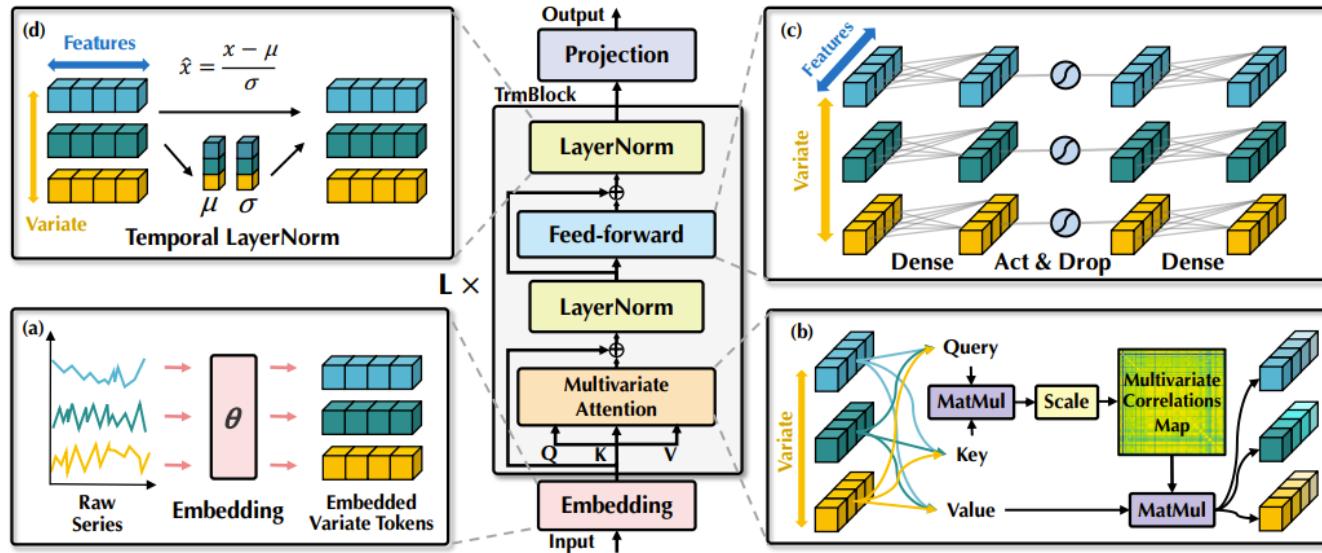
구조 설명



- 각 변수를 하나의 독립적인 토큰으로 변환 (a)
 - 개별 변수(시계열) 전체를 하나의 독립적인 토큰으로 변환
- 변수 간 관계 학습을 위한 Self-Attention (b)
 - Self-Attention을 변수 축에 적용해 변수 간의 관계를 학습
- 변수별 시간적 패턴 학습을 위한 Feed Forward Network (c)
 - Attention을 통과한 각 변수 토큰이 FFN을 개별적으로 통과, 이후에 각 변수별 시간적 패턴 학습

iTransformer

구조 설명



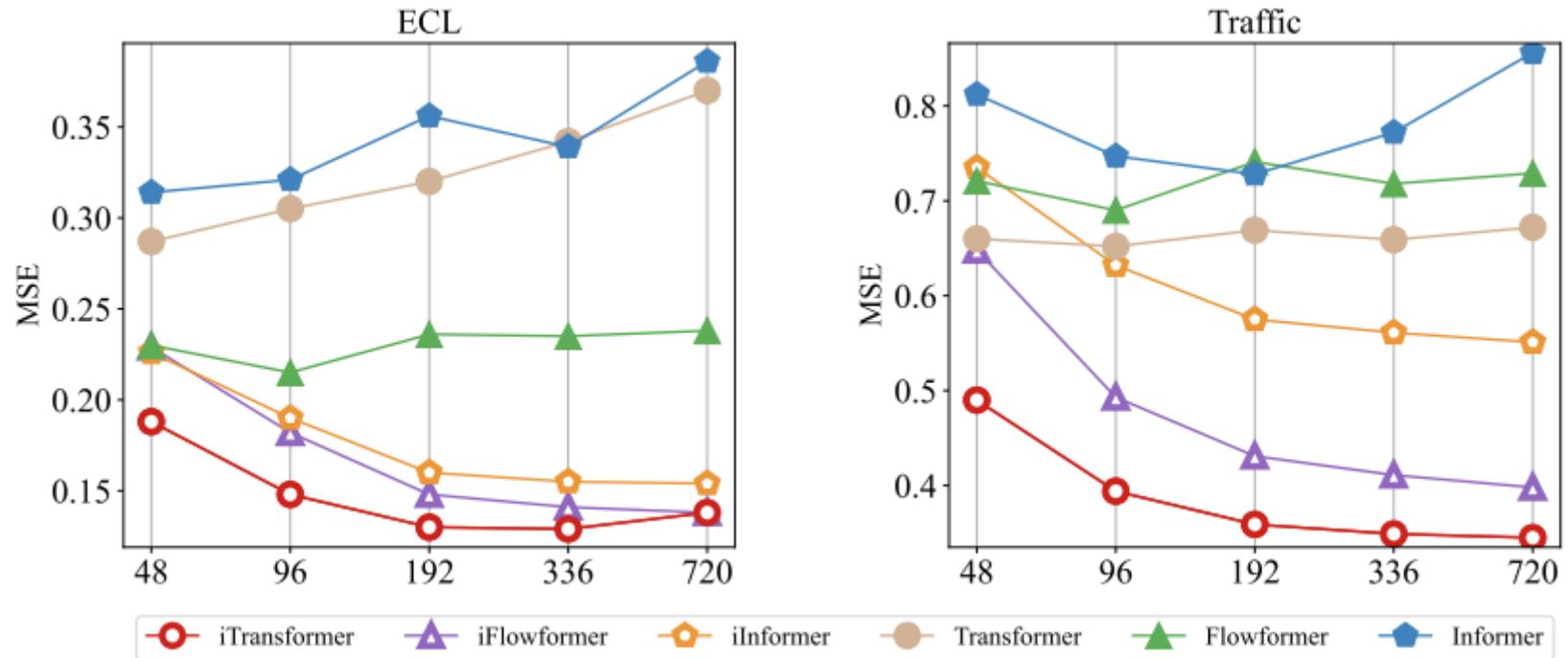
- (a) 개별 변수(시계열) 전체가 임베딩 층을 거쳐 임베딩된 하나의 독립적인 토큰으로 변환
- (b) 변환된 변수 토큰을 기반으로 Q, K, V 생성한 뒤 Self-Attention 연산 수행, 이 과정에서 변수 간의 관계를 학습하는 Multivariate Correlations Map 생성
- (c) Attention을 통과한 각 변수 토큰이 FNN을 개별적으로 통과하는데, 이때 FNN은 각 변수 고유의 압축된 시간적 패턴을 비선형적으로 학습
- 이후 LayerNorm을 거쳐 정규화 진행 후 Projection을 통해 예측 결과 출력

iTransformer

주요 모델 성능 비교

iTransformer

입력 길이에 따른 성능 변화



- 기존 Transformer 기반 모델과는 달리 입력 길이가 길어질수록 성능이 떨어지는 반면, iTransformer는 성능이 꾸준히 향상되는 모습을 보임
- 이를 통해 더 많은 과거 정보를 효과적으로 활용하는 것을 확인

iTransformer

실험 세팅

- 사용한 모델: iTransformer
- 재현 실험 데이터셋: ETTh1, ETTm1, ECL

Experiment	ETTh1, ETTm1
Learning rate	10^{-4}
Epoch	10
Batch size	32
Loss function	MSE Loss
Sequence Length	96
input_feature	7
Pred_len	96/192/336/720
d_model	256/512
d_ff	256/512

iTransformer

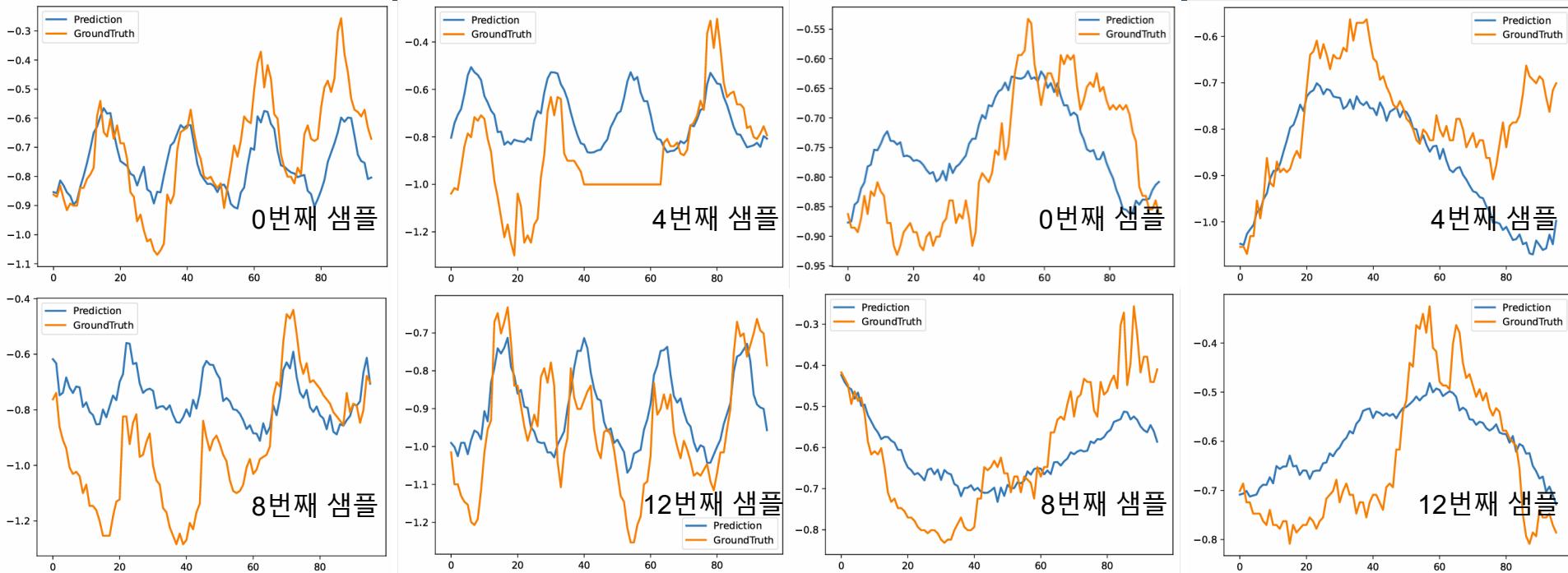
iTransformer 재현 실험 (ETTh1, ETTm1)

- 재현 실험 결과, 오차가 있지만 논문과 거의 유사한 수치가 나온 것을 확인

Pred len	ETTh1		ETTh1		ETTm1		ETTm1	
	Paper		Reproduction		Paper		Reproduction	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
96	0.386	0.405	0.385	0.404	0.334	0.368	0.338	0.373
192	0.441	0.436	0.440	0.436	0.377	0.391	0.381	0.395
336	0.487	0.458	0.491	0.460	0.426	0.420	0.431	0.423
720	0.503	0.491	0.509	0.493	0.491	0.459	0.499	0.464

iTransformer

iTransformer 재현 실험 시각화 (ETTh1, ETTm1)



iTransformer
ETTh1 시각화

iTransformer
ETTm1 시각화

iTransformer

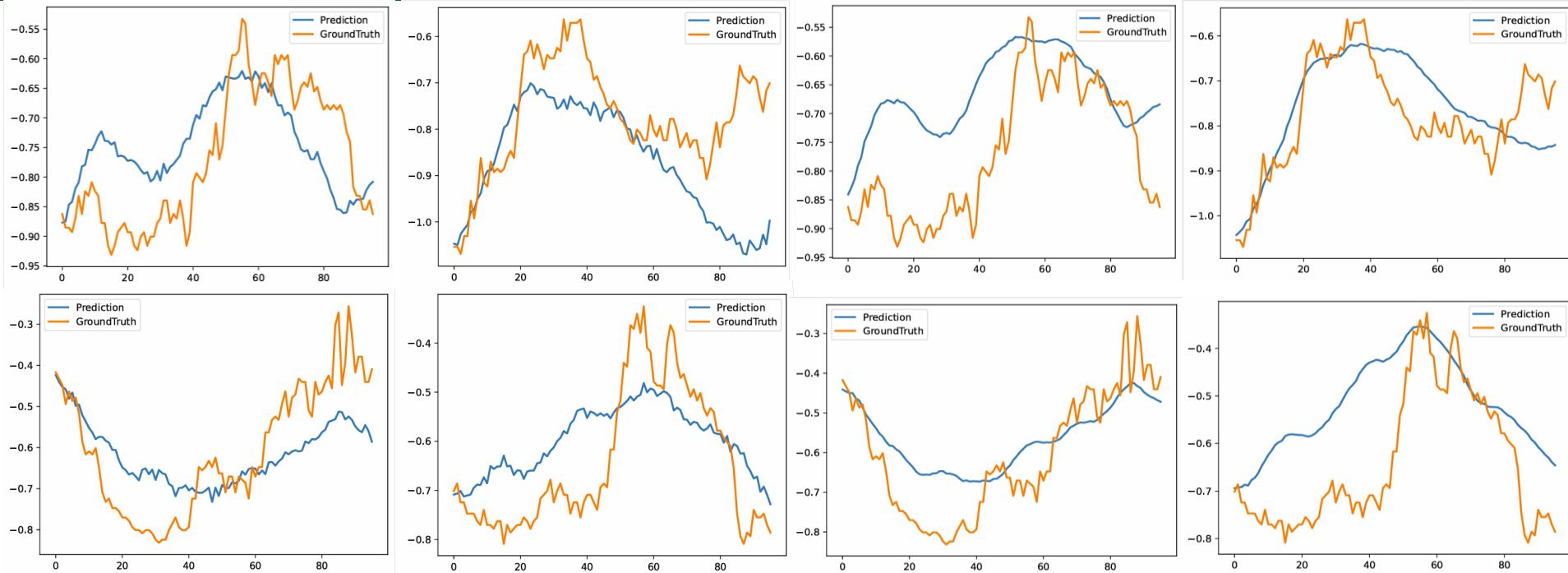
iTransformer, PatchTST, DLinear 성능 비교 (ETTh1, ETTm1)

- 주요 모델 성능 비교에서 확인했듯, ETT와 같은 데이터셋에서 성능이 두드러지지 않는 것을 확인
- ETT 데이터셋에서는 PatchTST 모델 성능이 두드러짐

	iTransformer ETTh1		PatchTST ETTh1		DLinear ETTh1		iTransformer ETTm1		PatchTST ETTm1		DLinear ETTm1	
Pred len	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
96	<u>0.385</u>	<u>0.404</u>	0.370	0.400	0.396	0.410	<u>0.338</u>	<u>0.373</u>	0.305	0.358	0.345	<u>0.373</u>
192	<u>0.440</u>	<u>0.436</u>	0.413	0.429	0.445	0.440	<u>0.381</u>	0.395	0.339	0.377	<u>0.381</u>	<u>0.391</u>
336	0.491	<u>0.460</u>	0.421	0.439	<u>0.487</u>	0.465	0.431	0.423	0.368	0.394	<u>0.415</u>	<u>0.415</u>
720	<u>0.509</u>	<u>0.493</u>	0.447	0.468	0.512	0.510	0.499	0.464	0.414	0.425	<u>0.472</u>	<u>0.450</u>

iTransformer

iTransformer, DLinear 시각화 (ETTm1)



iTransformer
ETTm1 시각화

DLinear
ETTm1 시각화

iTransformer

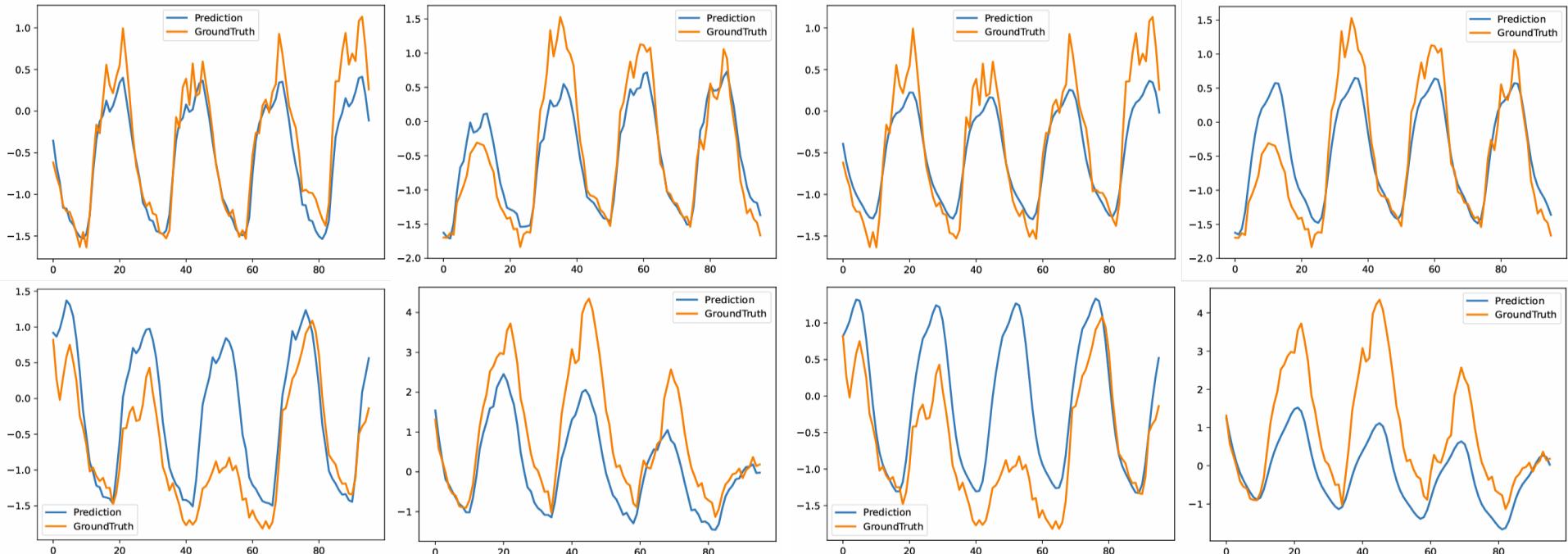
iTransformer, DLinear 성능 비교 (ECL)

- 모든 예측 길이에서 iTransformer가 DLinear보다 수치가 더 낮게 나온 것을 확인

Pred len	iTransformer ECL		DLinear ECL	
	MSE	MAE	MSE	MAE
96	0.163	0.254	0.210	0.301
192	0.175	0.264	0.210	0.304
336	0.192	0.282	0.223	0.319
720	0.231	0.315	0.257	0.349

iTransformer

iTransformer, DLinear 시각화 (ECL)



iTransformer
ECL 시각화

DLinear
ECL 시각화

iTransformer

실험 결과 정리

- ETTh1, ETTm1 재현 실험에서는 논문과 거의 동일한 수치를 확인했으나, 성능 자체는 PatchTST 모델이 더 좋은 것을 확인
- ECL 재현 실험을 통해 Attention을 변수 축에 적용하는 iTransformer에서는 변수가 많은 데이터셋에서의 성능이 두드러진다는 것을 확인
- 입력 길이가 길어질수록 더 많은 과거 정보를 활용해 성능이 꾸준히 향상됨