

Resolution-Dependent Analysis of Time Series Forecasting Performance: A Comparative Study Using Correlative Multi-Datasets

이름 양태규

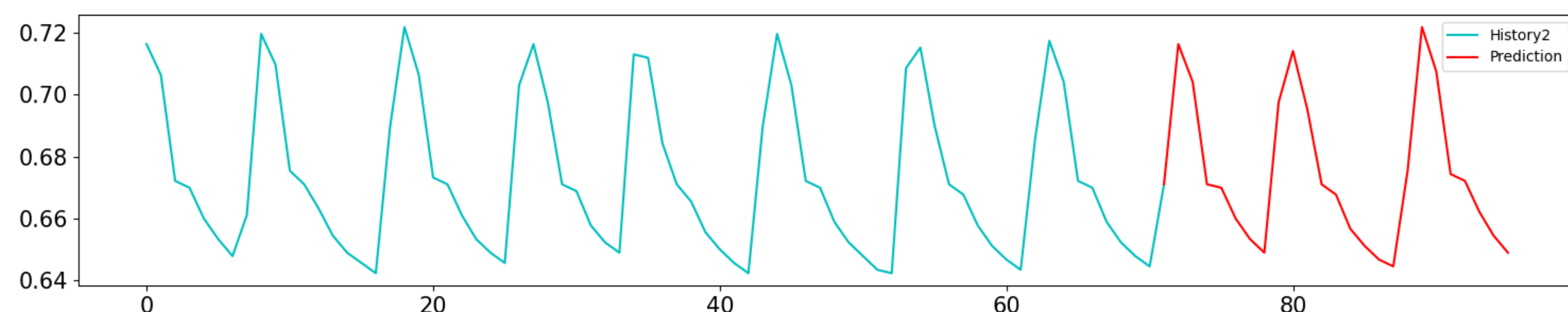
지도교수 손경아 교수님

연구 배경

- 일반적으로 시계열 데이터 예측에는 하나의 긴 센서 데이터를 사용하지만, 의료 데이터에서는 패턴의 유사성을 가지는 여러 환자의 시계열 데이터를 사용함.
- 환자에 따라 signal data의 형태가 다르지만, 일반적으로 좋은 예측이 가능하도록 resolution의 guidance를 조정해 보고자 함.
- 장기간의 시계열 데이터를 토대로 미래 사건을 예측할 수 있다면 분야에 상관없이 위험 관리에 큰 도움이 될 것임.

연구 진행 과정

Dataset:

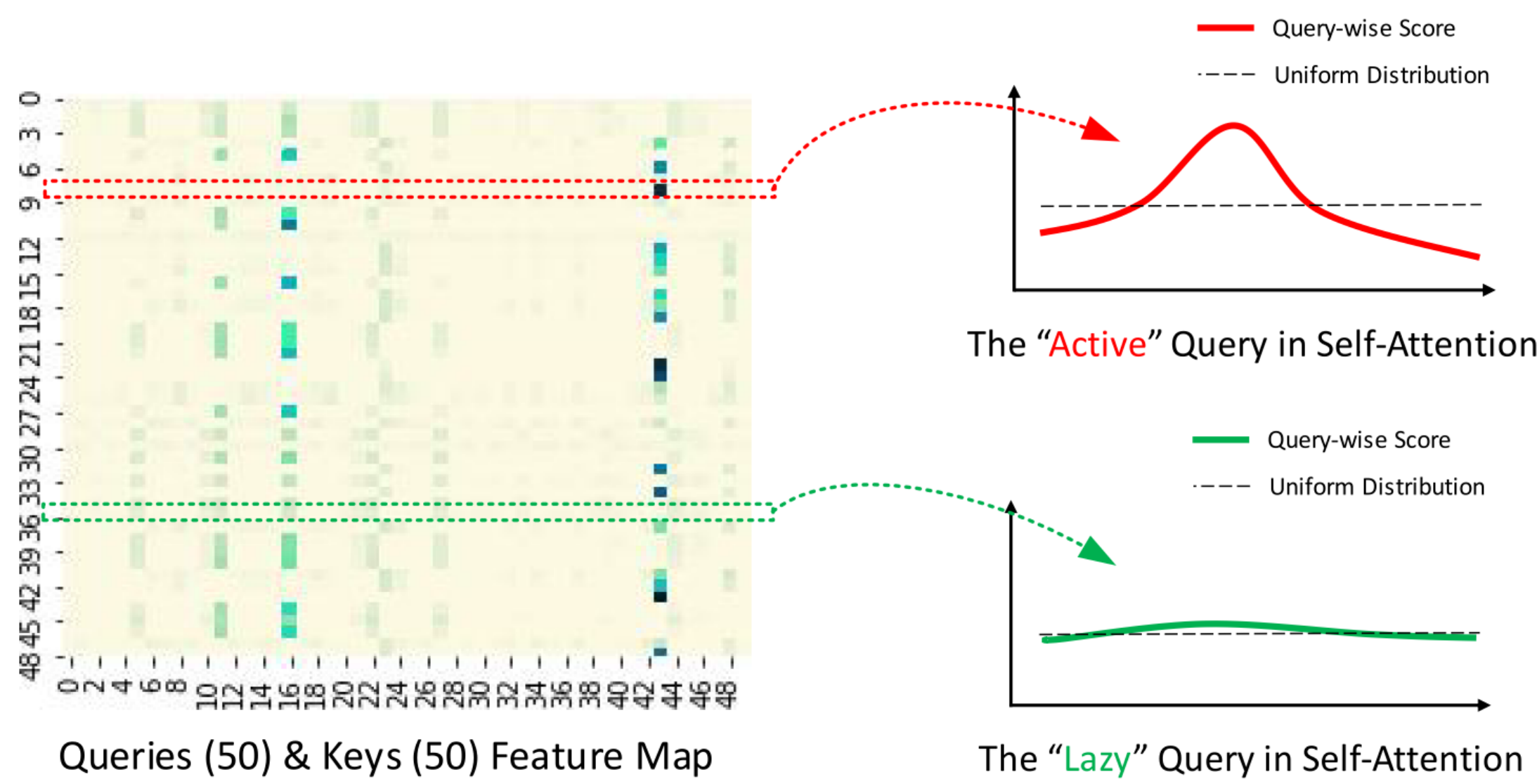


- VitalDB에서 공개된 데이터 중 혈압을 사용함.
- 혈압이 너무 낮거나 너무 높아 학습에 영향을 미치는 부분을 outlier로 제외함.
- 100frame/second의 데이터를 5와 10 frame으로 downsampling하여 학습을 진행함.
- 데이터에 정규화를 수행하고, train, validation, test로 각각 8:1:1으로 나누어 실험을 진행함.

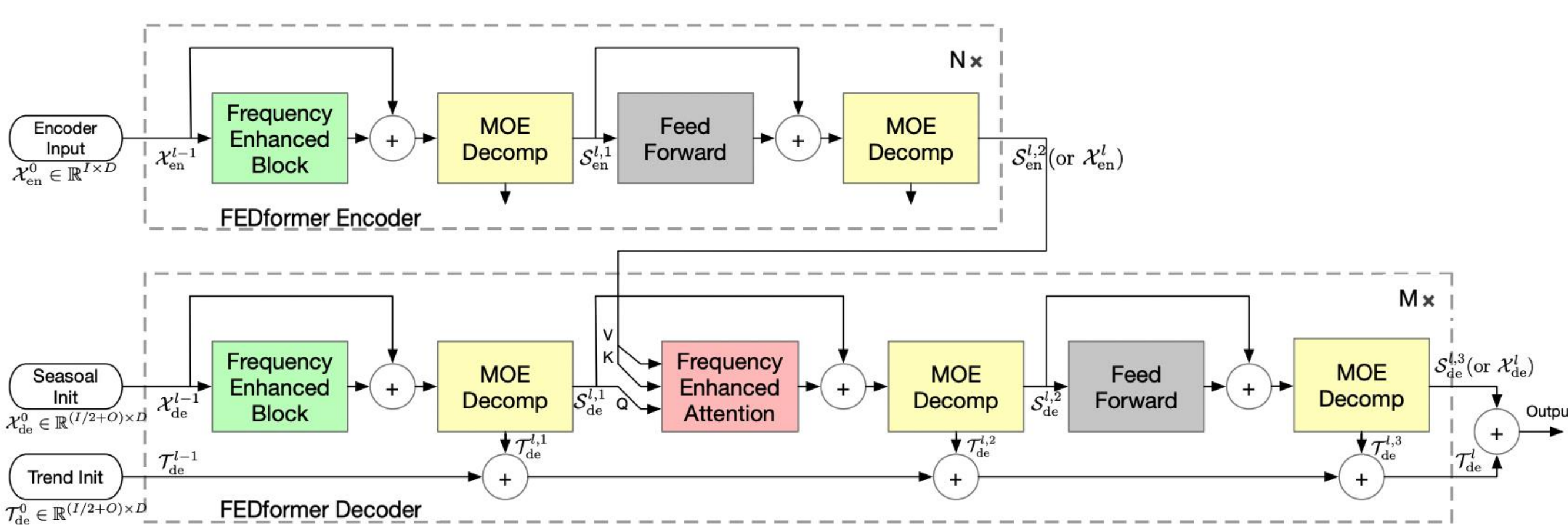
Problem Definition:

- 96 time step의 time series sequence가 주어지면, 그 이후 24 time step만큼의 데이터를 예측하는 것이 목표 (forecasting problem)

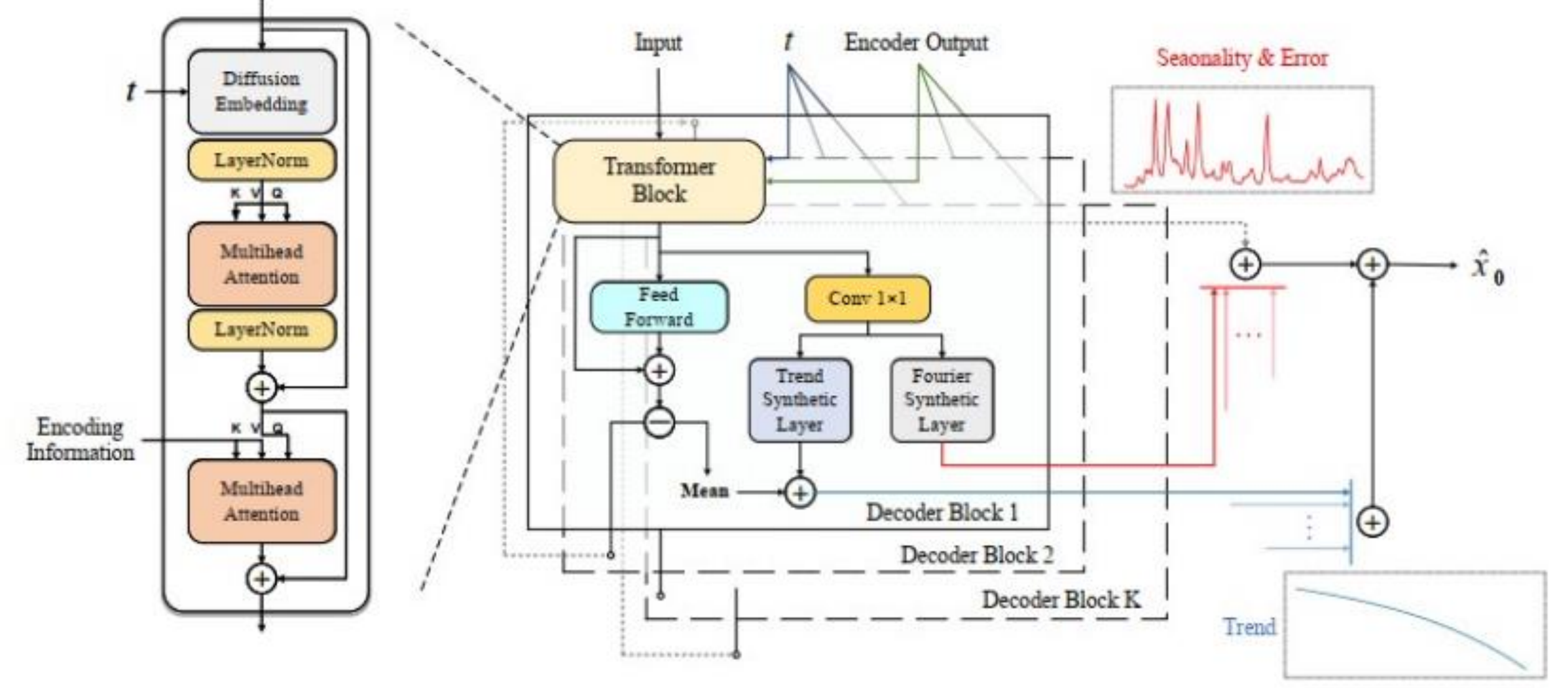
Model:



- Informer[1]: Attention을 수행할 때, Fourier 변환을 통해 dominant한 query-key를 사용함.

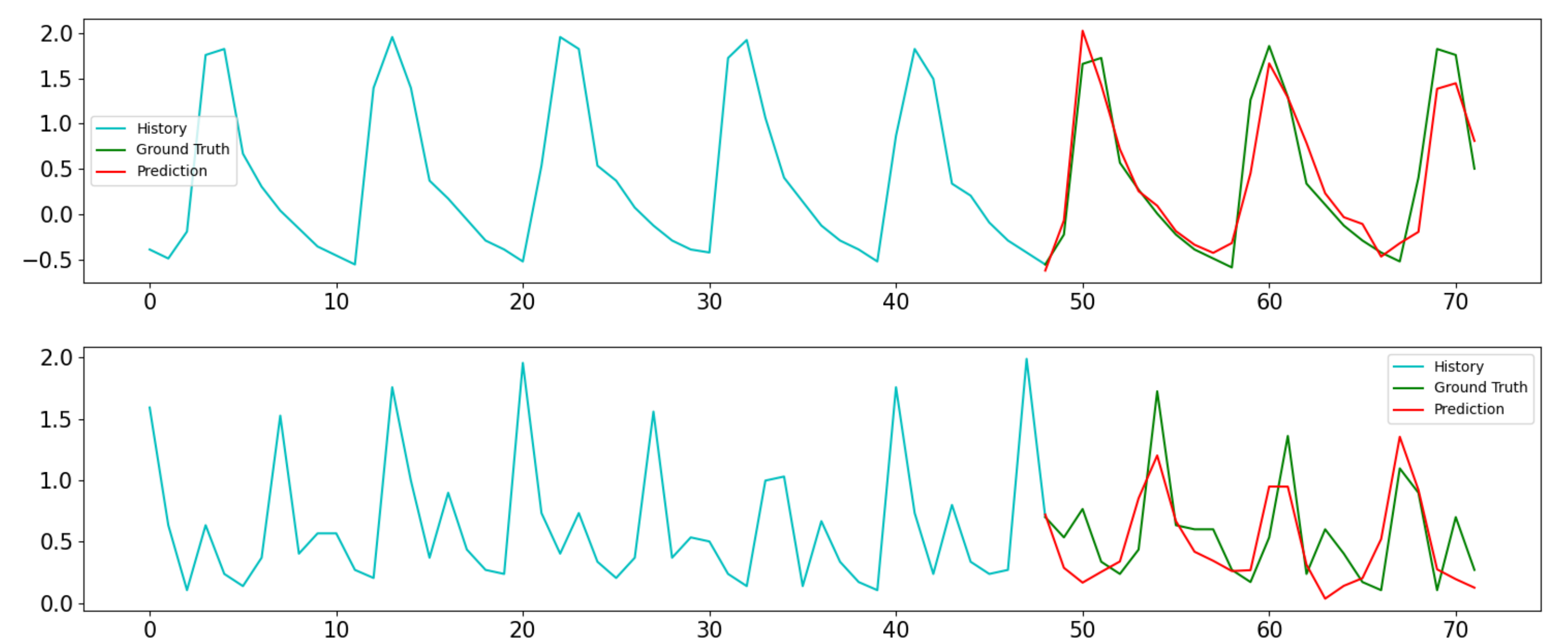


- FEDformer[2]: Signal data를 Fourier/Wavelet 변환을 통해 seasonal-trend로 분해하여 학습함.



- Diffusion-TS[3]: Diffusion-based model에서 FEDformer의 아이디어와 유사한, time series data를 seasonal-trend한 feature으로 분해하여 학습을 진행함.

결과 및 분석



Forecasting 결과 시각화(Informer):

- 주기성이 확실하게 나타나는 sequence의 경우에는 Ground truth와 prediction이 거의 일치함.
- Abnormal한 상황에서는 정확도가 떨어지지만 주기성과 추세성을 어느 정도 따라감.

Model	10frame/s MSE	5frame/s MSE
Vanilla Transformer	0.415	0.143
Informer	0.149	0.095
FEDformer	0.183	0.138
Diffusion-TS	0.583	0.352

실험 결과 분석:

- 동일한 데이터를 어떤 resolution으로 학습하는지에 따라 학습 성능이 다름.
- 이를 일반화하여 데이터에 적절한 guidance를 알아서 부여할 수 있다면, 통상적인 time series data에 모두 좋은 성능을 보이는 모델 개발이 가능할 것.
- Diffusion의 경우 transformer base 모델보다 성능이 좋지 않는데, 학습을 진행하면서 적절한 하이퍼파라미터를 찾지 못한 것으로 보임.

참고 문헌

- [1]Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, Wancai Zhang, "Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting", AACL 2021
- [2]Zhou, Tian and Ma, Ziqing and Wen, Qingsong and Wang, Xue and Sun, Liang and Jin, Rong, "FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting", ICML 2022
- [3]Yuan, Xinyu and Qiao, Yan, "Diffusion-TS: Interpretable Diffusion for General Time Series Generation", ICLR 2024