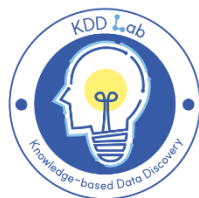


미세먼지 보간 모델을 위한 지역 배치 정규화

Local Batch Normalization for Air Pollution Interpolation Models



Knowledge-based Data Discovery

INHA KDD Lab

Presented by **Jinyong Yun**

INHA Univ. Computer Science and Technology

(12191632Y@inha.edu)

2024.11.21



목차



- 서론
- 미세먼지 보간 모델의 지역 배치 정규화
- 실험 및 평가
- 결론

■ 연구 배경

- 도시 내 미세먼지 농도를 정확히 측정하기 위해 곳곳에 미세먼지 측정소가 설치되고 있지만, 모든 지역에 충분한 미세먼지 측정소를 설치하는 것은 공간적으로 불가능
- 따라서 미세먼지 측정소가 없는 지역의 농도 추정을 위해, 다양한 보간 연구가 활발히 진행중

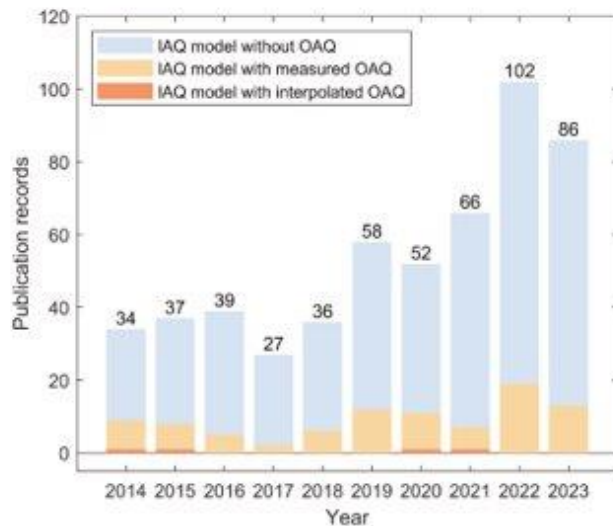
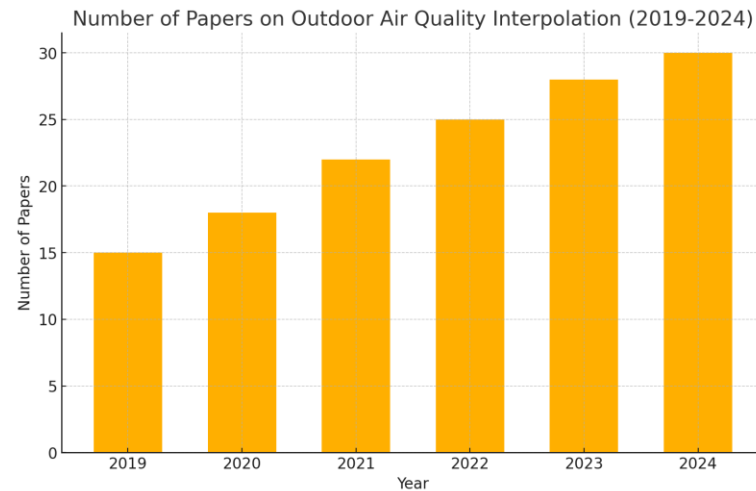


Fig. 1. Publication records related to IAQ modeling considering OAQ information from 2014 to 2023 year.



■ 관련 연구

■ CNN 보간 모델, CNN-LSTM 보간 모델

- ✓ Real-time Air Pollution prediction model based on Spatiotemporal Big data (BIC 2018)
- ✓ CNN 보간 모델에 LSTM 결합하여 시계열적 성능 개선

■ ConvLSTM 보간 모델

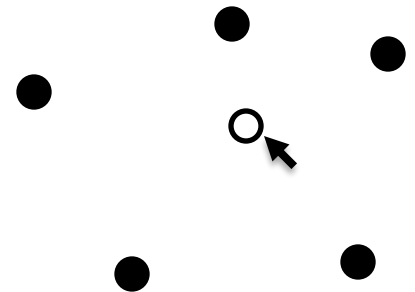
- ✓ Spatiotemporal DL Model for CityWide Air Pollution Interpolation and Prediction (IEEE BIGCOMP 2020)
- ✓ ConvLSTM을 사용하여 도시 내 미세먼지 측정소 데이터 보간

■ 기존 연구의 한계

- 기존 보간 모델 연구에서는 주로 Spatiotemporal 모델이 사용
- 특히 기존 모델은 시계열적 성능 개선(예측)에 초점을 맞춰 보간 성능을 향상
- 공간적 성능 개선으로 보간 성능을 향상시키는 방법 : 배치 정규화
- 시계열적 성능 개선 방식의 배치 정규화
 - ✓ 각 시점에서 개별적으로 평균과 분산을 구해 정규화하여 시간 축에 따른 데이터의 일관성이 깨질 위험
 - ✓ C. S. Laurent, et al. "Batch Normalized Recurrent Neural Networks", in Proc. IEEE ICASSP, October 2015
- 공간적 성능 개선 방안으로 배치 정규화
 - ✓ 각 측정소의 미세먼지 농도의 범위와 변화 패턴을 일정하게 조정하여 미세먼지의 공간적 분포가 왜곡되지 않도록 방지하며, 공간적 특성을 안정화

■ 제안 모델

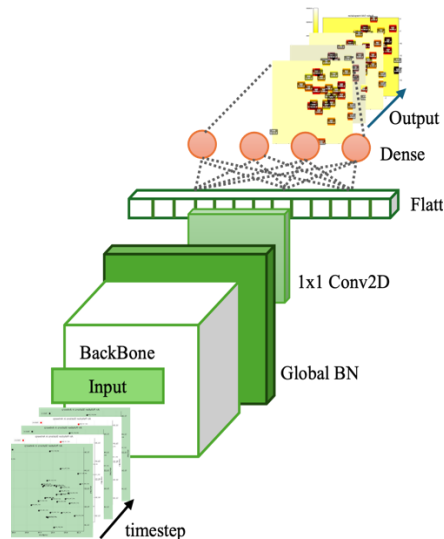
- 기존 보간 모델들은 더 복잡한 모델, 더 많은 채널, 더 많은 데이터로 보간 성능을 향상
- 이는 시간, 자원적으로 비용 소모가 큰 방식
- 좀 더 효율적으로 성능을 높일 수 있는 compact한 방식 필요
- 여러 미세먼지 측정소가 분산되어 있는 상황에서, 미세먼지 보간을 수행할 때
Local BatchNormalization 기법을 도입하여 미세먼지 보간 모델을 개선
- 실험 결과, 다양한 수준에서 미세먼지 보간 모델에 배치 정규화 계층을 적용하면
보간 정확도가 개선됨을 확인했음



미세먼지 보간 모델의 지역 배치 정규화

■ 전역 배치 정규화의 한계

- 전역 배치 정규화(b)는 모델이 전체 데이터를 기준으로 평균과 분산을 계산하여 정규화하는 방식
- 모델의 학습 과정에서 데이터를 균일하게 처리하는 효과
- 하지만 공간적 변동이 큰 데이터에는 적합하지 않을 수 있음
- 미세먼지 농도는 지역별로 환경적 차이, 위치 특성에 따라 변동성이 굉장히 큰 데이터
- 전역 배치 정규화는 이런 지역적 특성을 반영하지 않아, 각 지역의 특수한 분포나 변화를 효과적으로 학습하지 못하는 한계

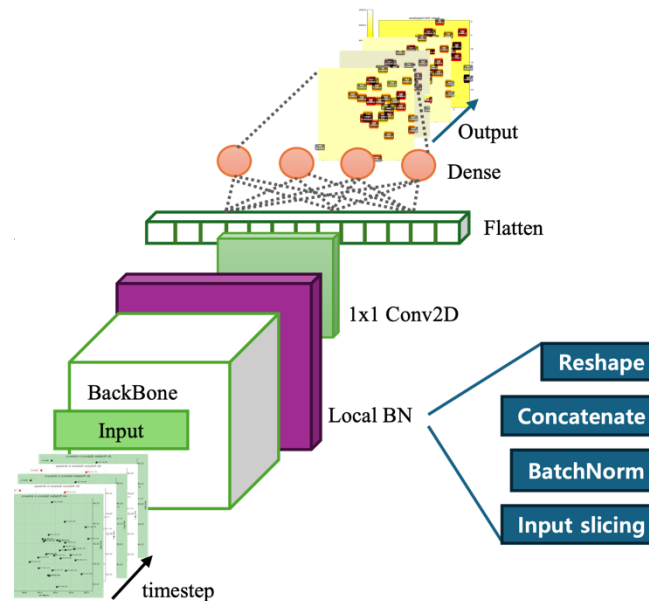


(b)

미세먼지 보간 모델의 지역 배치 정규화

■ 지역 배치 정규화의 강점

- 지역 배치 정규화는 각 측정소마다 독립적으로 배치 정규화를 수행
- 특정 위치의 미세먼지 농도가 갖는 특유의 패턴을 유지하면서 데이터를 학습할 수 있도록 함
- 지역적 특성을 반영한 배치 정규화는 모델이 특정 위치의 평균 농도와 변화 패턴을 더 잘 학습하게 하여, 각 측정소별 예측 정확도를 높이는 데 효과적

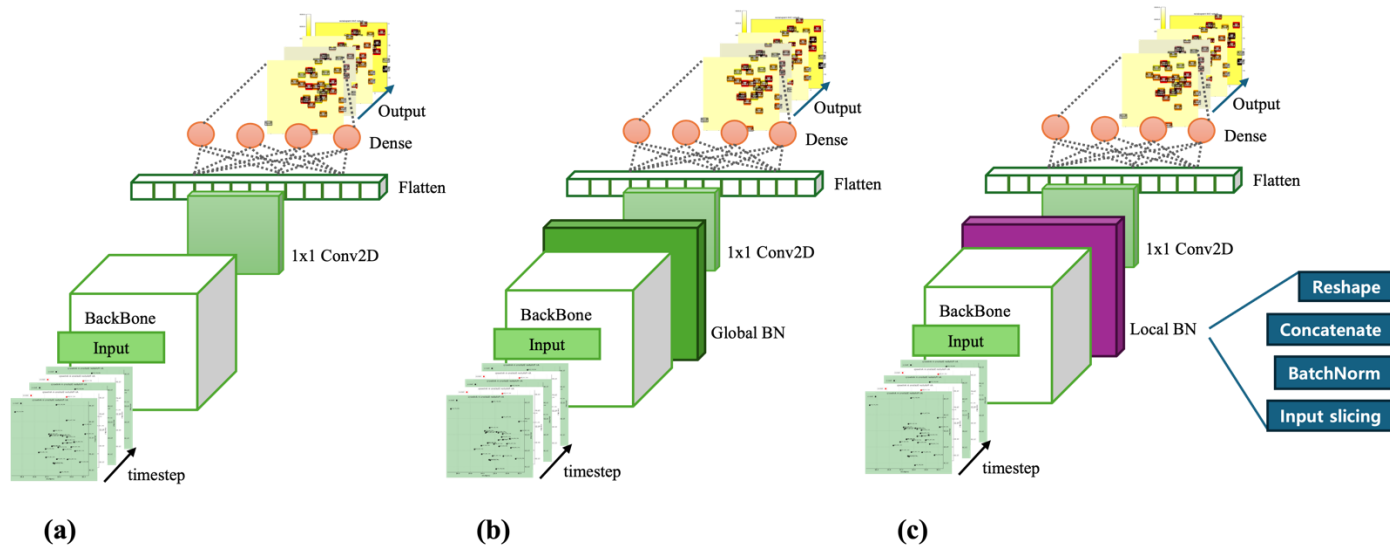


(c)

미세먼지 보간 모델의 지역 배치 정규화

■ 지역 배치 정규화

- 제안 방식은 그림 (c)
- Baseline이 되는 보간 모델에 지역 배치 정규화 층을 추가
- 기존 배치 정규화 적용 방식은 전역 배치 정규화를 적용한 그림 (b)



■ Dataset

■ SensEURCity

- ✓ 총 85개의 센서, 2020년-2021년 안트베르펜, 오슬로, 자그레브의 세 도시에 배치
- ✓ 1분 단위, 센서 위치(경도, 위도), NO, NO2, O3, PM2.5, PM10, PM1, CO2, traffic 등 데이터 포함
- ✓ <https://zenodo.org/records/7669644>

- 본 실험에서 사용한 데이터셋은 Antwerp 지역 32개의 미세먼지 측정소 데이터
- OPCN3PM25 센서를 이용한 PM2.5 데이터
- 모든 측정소가 가동중인 2020-06-29 10:20:00 ~ 2021-02-17 07:18:00
- 각 station의 위도와 경도를 32x32 그리드에 매핑
- Train : Validation : Test = 6:2:2



■ 실험 환경

- Backbone model로 ConvLSTM[1]과 CNN[6] 보간 모델을 사용
- Backbone model에 대해 전역 배치 정규화와 지역 배치 정규화를 수행하고 그 결과를 비교했음
- 보간 평가 지표 spRMSE
 - ✓ 측정소별로 보간 값과 실제 미세먼지 값의 RMSE 값
 - ✓ 모든 측정소에 대해 해당 RMSE값 평균
- 훈련 시 Validation Loss 20 step별로 체크하여 Overfitting 방지
- 기타 실험 환경은 백본 모델에서의 실험 환경에 맞춤

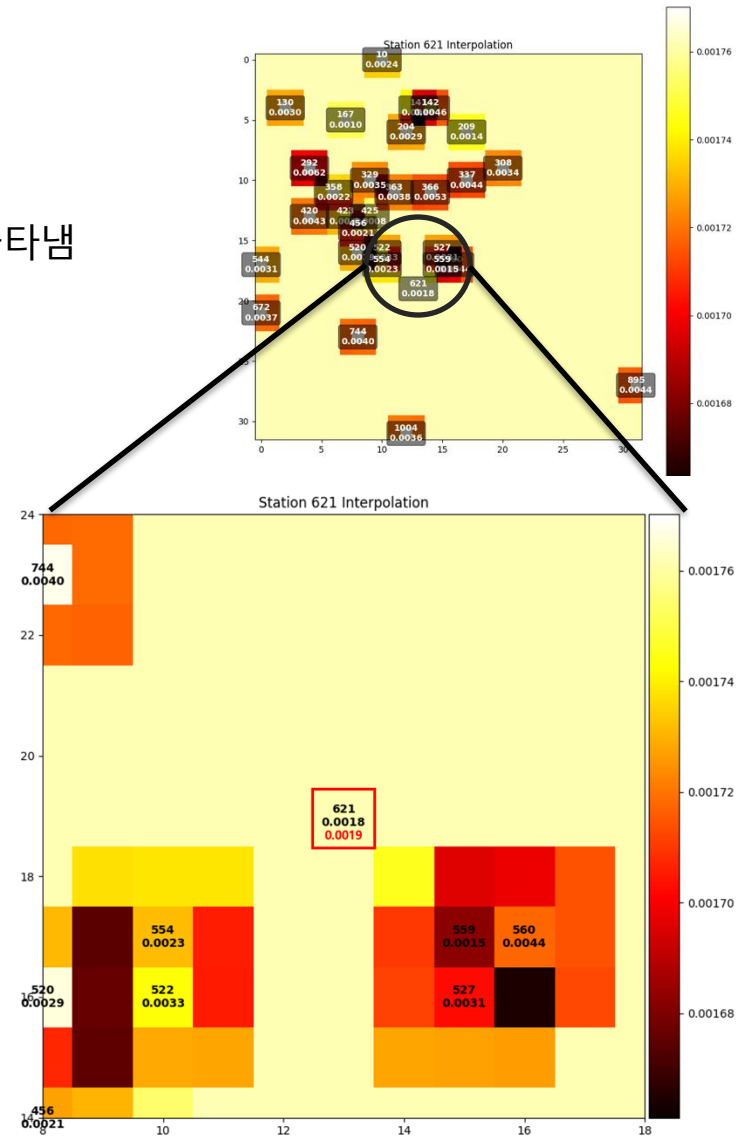
■ 실험 결과

- 두 백본 모델에 대해 배치 정규화를 수행하지 않았을 때 (Baseline), 전역 배치 정규화를 수행했을 때 (Global BatchNorm), 그리고 지역 배치 정규화를 수행했을 때 (Local BatchNorm) spRMSE를 정량적으로 비교
- spRMSE는 Baseline > Global BatchNorm > Local BatchNorm
- 즉 보간 정확도는 Local BatchNorm > Global BatchNorm > Baseline 순으로 높음
- 따라서 미세먼지 보간 모델에 전역 배치 정규화를 적용하는 것이 성능에 유의미한 영향을 끼칠 뿐만 아니라, 지역 배치 정규화를 적용하는 방식이 더 효과적임을 보임

Backbone	ConvLSTM	CNN
Baseline	16.6354	18.6550
Global BatchNorm	13.9992	15.9571
Local BatchNorm	13.7041	14.1488

■ 사례 연구

- ConvLSTM에 지역 배치 정규화를 적용한 모델
- 특정 시점의 621번 측정소의 보간 결과를 히트맵으로 나타냄
- 히트맵의 색이 더 짙을 수록 미세먼지 농도가 높음
- 621번에서 보간값은 0.0018
- 621번에서 실제값은 0.0019



■ 결론

- 배치 정규화를 미세먼지 보간 모델에 활용하면 보간 성능이 올라간다는 가설을 입증
- 특히 각 측정소별로 배치 정규화를 적용하면 측정소별 공간적 특성에 맞게 정규화가 일어나서 더욱 정밀한 보간 결과를 얻을 수 있다

■ 향후 연구

- 측정소 간의 상관관계나 환경적 요인을 고려하는 추가적인 정규화 기법을 연구
- 시계열 흐름 보존을 위한 다층 배치 정규화를 도입하여 시간에 따른 예측 안정성을 보완

Thanks for listening



Knowledge-based Data Discovery
INHA KDD Lab