# 251124
# Project Meeting: Synthetic Data

AAILAB
Department of Industrial and Systems Engineering
KAIST

# 울주-기장 지역 도로망 데이터셋

- 시뮬레이션 도로망 구축에 사용되었던 표준 노드 링크 데이터

- MOCT_NODE
  - Node(지역) 정보를 기록한 데이터
  - # of data: 206,132
  - # of attribute: 8 + 1 (geometry)
  - Core attributes
    - NODE_ID: 노드 식별자
    - Geometry: 위치 정보

- MOCT_LINK
  - Link(도로) 정보를 기록한 데이터
  - # of data: 548,667
  - # of attribute: 21 + 1 (geometry)
  - Core attributes
    - LINK_ID: 링크 식별자
    - F_NODE: 링크 시작 노드
    - T_NODE: 링크 종료 노드
    - LANES: 차로 수

| | NODE_ID | NODE_TYPE | NODE_NAME | TURN_P | UPDATEDATE | REMARK | HIST_TYPE | HISTREMARK | geometry |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1670063400 | 105 | 계양구-김포시 | 0 | 20230519 | None | None | None | POINT (178715.118 553973.737) |
| 1 | 1680000800 | 101 | 가좌IC동측 | 1 | 20230519 | None | None | None | POINT (171319.860 542834.784) |
| 2 | 1680001000 | 104 | 가좌IC고가교 | 1 | 20230519 | None | None | None | POINT (171001.862 542849.426) |
| 3 | 1680002201 | 101 | 가정동581-2 | 0 | 20230519 | None | None | None | POINT (171440.621 546888.447) |
| 4 | 1680002300 | 101 | 서인천IC남측 | 1 | 20230519 | None | None | None | POINT (171571.450 546682.493) |

**Examples of Node dataset**

| | LINK_ID | F_NODE | T_NODE | LANES | ROAD_RANK | ROAD_TYPE | ROAD_NO | ROAD_NAME | ROAD_USE | MULTI_LINK | ... | REST_VEH | REST_W |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2333122000 | 2330041100 | 2330043700 | 1 | 107 | 000 | - | 제암고주로 | 0 | 0 | ... | 0 | 0 |
| 1 | 3080340700 | 3080111800 | 3080111000 | 1 | 107 | 000 | - | 용연길 | 0 | 0 | ... | 0 | 0 |
| 2 | 3070212800 | 3070076600 | 3070077300 | 1 | 107 | 000 | - | 동령길 | 0 | 0 | ... | 0 | 0 |

**Examples of Link dataset**

# 시뮬레이션 로그 데이터

- 2025-kjwj_kaist_common_43200_50400_20250317_{요일}_trajectory.csv
  - 각 vehicle의 simulated path를 요일별로 기록한 데이터
    ➔ 모든 요일을 통합하여 사용
  - # of data: 약 45만 ~ 82만 개
  - # of attribute: 5
  - Core attributes
    - vehID: 경로 식별자
    - linkID: 링크 식별자
    - seq: 각 경로의 링크 순서
      - 해당 vehicle이 어떤 링크를 따라 움직였는지 표현
      - 순서에 따라 enterTime과 leaveTime이 동일함
    - enterTime: 링크 진입 시간 (초)
    - leaveTime: 링크 진출 시간 (초)
      - −1: 링크 진출 전에 시뮬레이션 종료로 시간 기록이 안된 경우

**Example of Path data**

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | vehID | linkID | seq | enterTir | leaveTir |
| 37407 | 15 | 1.96E+09 | 10 | 43731 | 43770 |
| 37408 | 15 | 1.96E+09 | 11 | 43770 | 43779 |
| 37409 | 15 | 1.96E+09 | 12 | 43779 | 43783 |
| 37410 | 15 | 1.96E+09 | 13 | 43783 | 43798 |
| 37411 | 15 | 1.96E+09 | 9 | 43704 | 43731 |
| 37412 | 15 | 1.96E+09 | 8 | 43611 | 43704 |
| 37413 | 15 | 1.96E+09 | 14 | 43798 | 43802 |
| 37414 | 15 | 1.96E+09 | 15 | 43802 | 43867 |
| 37415 | 15 | 1.96E+09 | 7 | 43607 | 43611 |
| 37416 | 15 | 1.96E+09 | 6 | 43527 | 43607 |
| 37417 | 15 | 1.96E+09 | 3 | 43381 | 43389 |
| 37418 | 15 | 1.96E+09 | 4 | 43389 | 43462 |
| 37419 | 15 | 1.96E+09 | 5 | 43462 | 43527 |
| 37420 | 15 | 1.96E+09 | 16 | 43867 | 43935 |
| 37421 | 15 | 1.96E+09 | 1 | 43329 | 43377 |
| 37422 | 15 | 1.96E+09 | 2 | 43377 | 43381 |
| 37423 | 15 | 1.96E+09 | 17 | 43935 | 44045 |
| 37424 | 15 | 1.96E+09 | 0 | 43202 | 43329 |
| 37425 | 15 | 1.96E+09 | 20 | 44060 | 44065 |
| 37426 | 15 | 1.96E+09 | 21 | 44065 | 44084 |
| 37427 | 15 | 1.96E+09 | 30 | 44131 | 44137 |
| 37428 | 15 | 1.96E+09 | 31 | 44137 | 44156 |
| 37429 | 15 | 1.96E+09 | 22 | 44084 | 44096 |
| 37430 | 15 | 1.96E+09 | 18 | 44045 | 44051 |
| 37431 | 15 | 1.96E+09 | 19 | 44051 | 44060 |
| 37432 | 15 | 1.96E+09 | 23 | 44096 | 44098 |
| 37433 | 15 | 1.96E+09 | 29 | 44124 | 44131 |
| 37434 | 15 | 1.96E+09 | 24 | 44098 | 44101 |
| 37435 | 15 | 1.96E+09 | 28 | 44115 | 44124 |
| 37436 | 15 | 1.96E+09 | 27 | 44109 | 44115 |
| 37437 | 15 | 1.96E+09 | 25 | 44101 | 44104 |
| 37438 | 15 | 1.96E+09 | 26 | 44104 | 44109 |

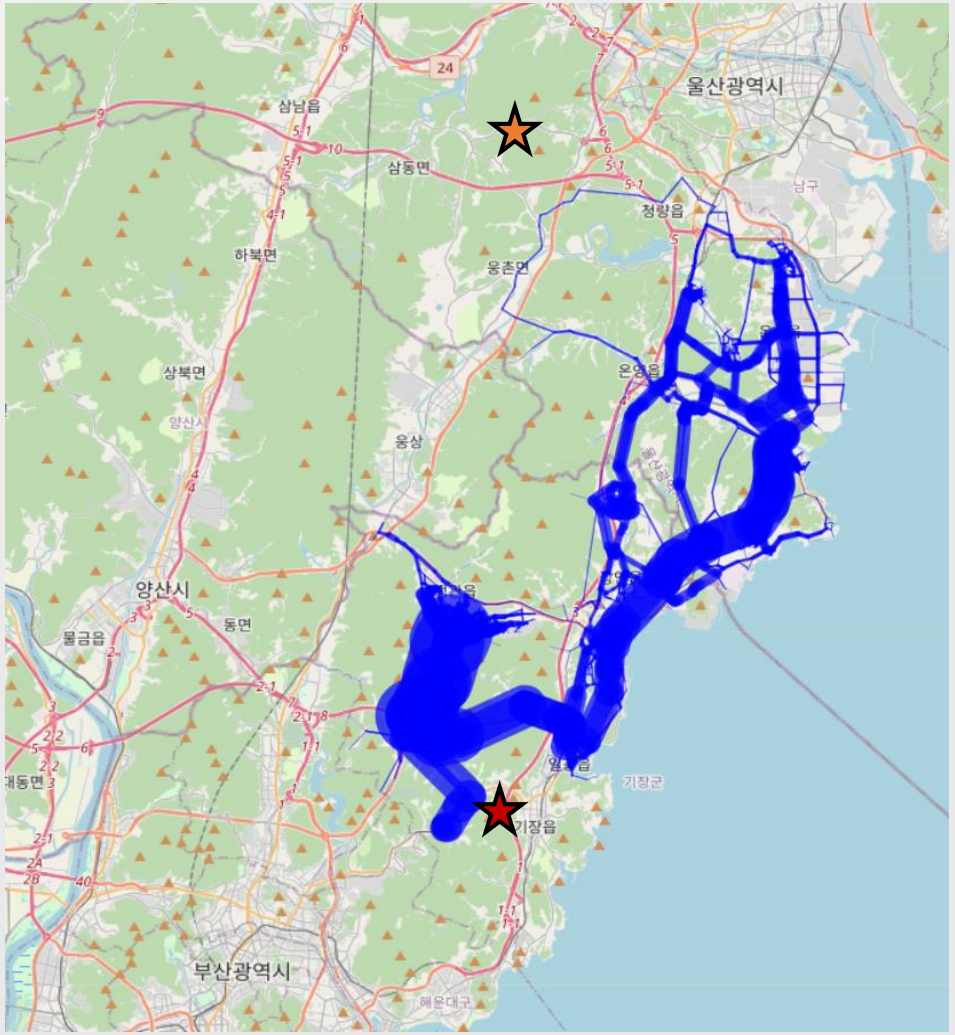
# Dataset Preprocessing

- Node와 Link dataset을 이용하여 Graph 생성
- Considerations
  1. Node and link filtering
     - Node와 Link dataset은 울주-기장 지역 외의 정보도 포함되어 있음
     - ➔ Simulated path에 존재하는 Node와 Link만 추출하여 사용
  2. Path filtering
     - Simulated path의 길이가 0인 경우 제외
       - 첫 link 진입 후, 진출 전에 시뮬레이션이 종료된 경우
     - Disconnected path 제외
       - 이전 link의 종료 node가 다음 link의 시작 node와 다른 경우
  3. Disconnected Subgraph filtering
     - 몇몇 path들로 생성한 Graph는 독립적인 Graph를 구성함
     - ➔ Connected Graph Component 중 Node 수가 많은 Graph를 사용

**Summary stats of Generated datasets**

| | Attributes | # of instances |
|---|---|---|
| 대전 | Nodes | 1,470 |
| | Edges | 2,257 |
| | Path | 87,910 |
| 울주-기장 | Nodes | 1,453 |
| | Edges | 1,965 |
| | Path | 119,589 |

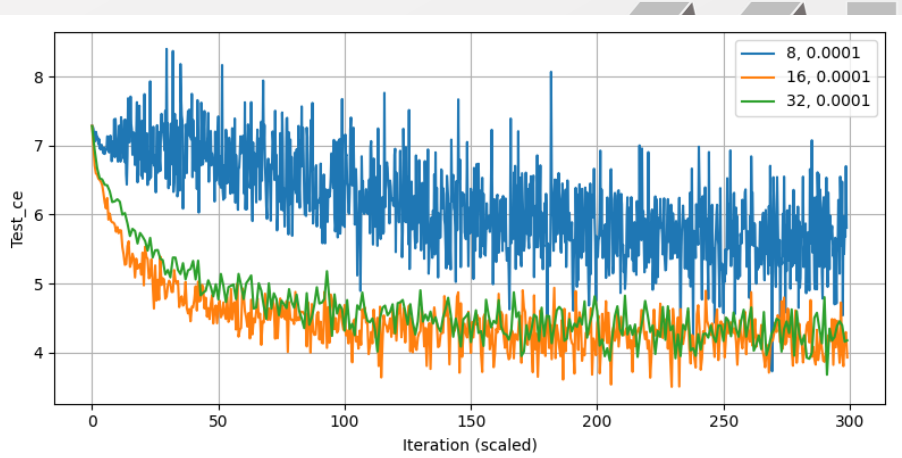
★:울주 ★:기장 —:Paths (통행량 반영)

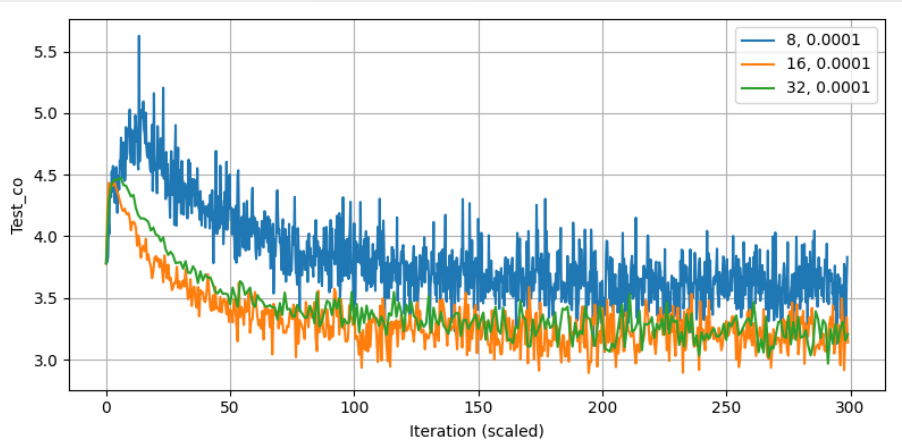


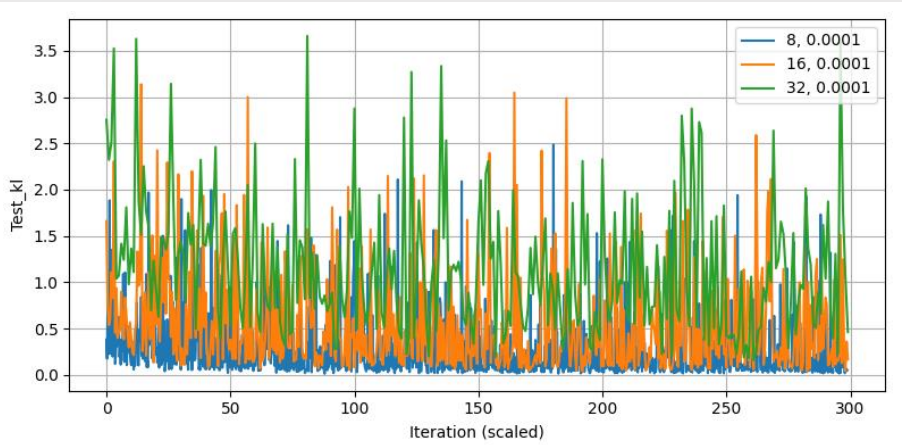
# Hyper-parameter search for unconditior

- Comparing batch size in [8,16, 32]
  - (Training curve) Loss = KL + CE + CO 에서 CE는 학습 시 batch-wise sum, 나머지는 batch-wise average 되어있음, 즉 batch size에 따라 loss weighted sum이 다름. 오른쪽 figure에서 CE의 경우는 scaling 하여 비교 하였음.
    - [16, 32]에서 CE가 잘 줄어든 이유는 loss summation 과정에서 CE의 비중이 큰 것이 이유 일 수 있음. KL은 batch 8에서 가장 안정적으로 감소함.
  - (Evaluation result)
    - KL loss component가 가장 잘 줄어들었던 batch 8이 evaluation metric이 젤 좋았음.

| bs | lr | KLEV | JSEV | nll_avg |
|---|---|---|---|---|
| 8 | 0.0001 | **7.00** | **3.17** | **23.86** |
| 8 | 0.0005 | **7.00** | **3.17** | 27.20 |
| 8 | 0.0010 | **7.00** | **3.17** | 28.85 |
| 16 | 0.0001 | 7.15 | 3.22 | 32.07 |
| 16 | 0.0005 | 7.11 | 3.21 | 23.92 |
| 16 | 0.0010 | 7.06 | 3.19 | 27.68 |
| 32 | 0.0001 | 7.11 | 3.21 | 35.46 |
| 32 | 0.0005 | 7.09 | 3.20 | 28.43 |
| 32 | 0.0010 | 7.10 | 3.21 | 25.71 |

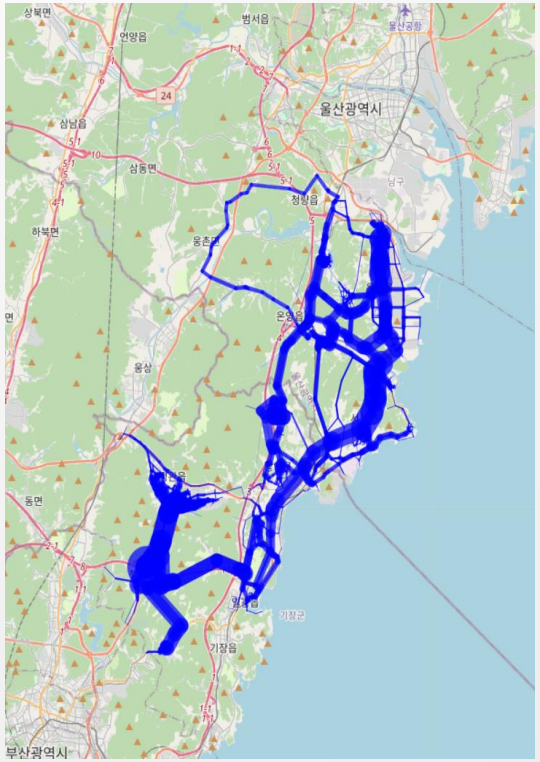


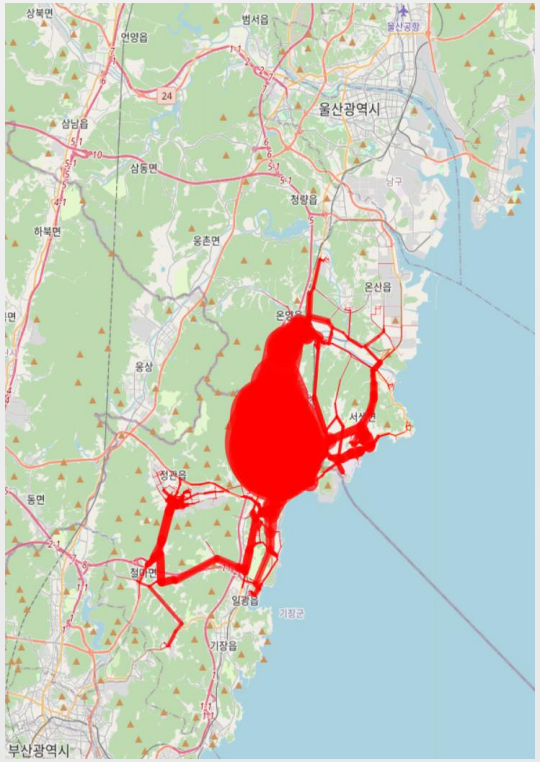
# Generated results

- Generated quality has correlation with NLL metric.

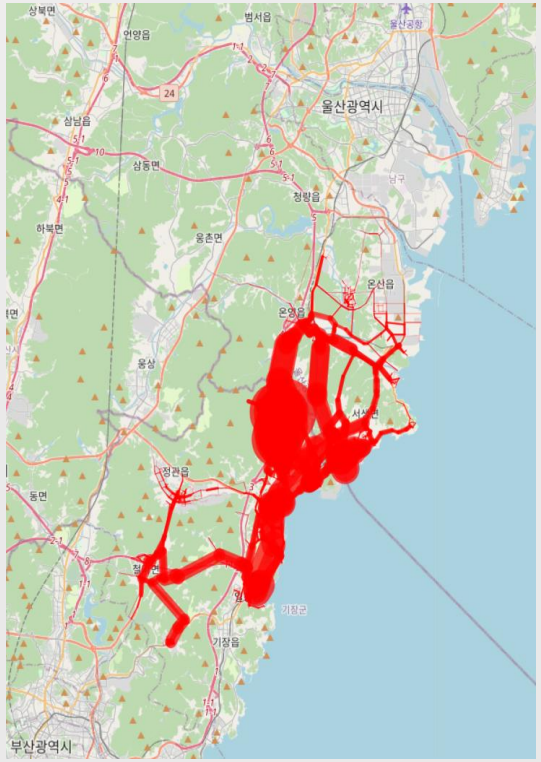


Real sample

(8, 0.0001), NLL= **23.86**



(16, 0.0001), NLL= 32.07



(32, 0.0001), NLL= 35.46

# Planning model 진행사항

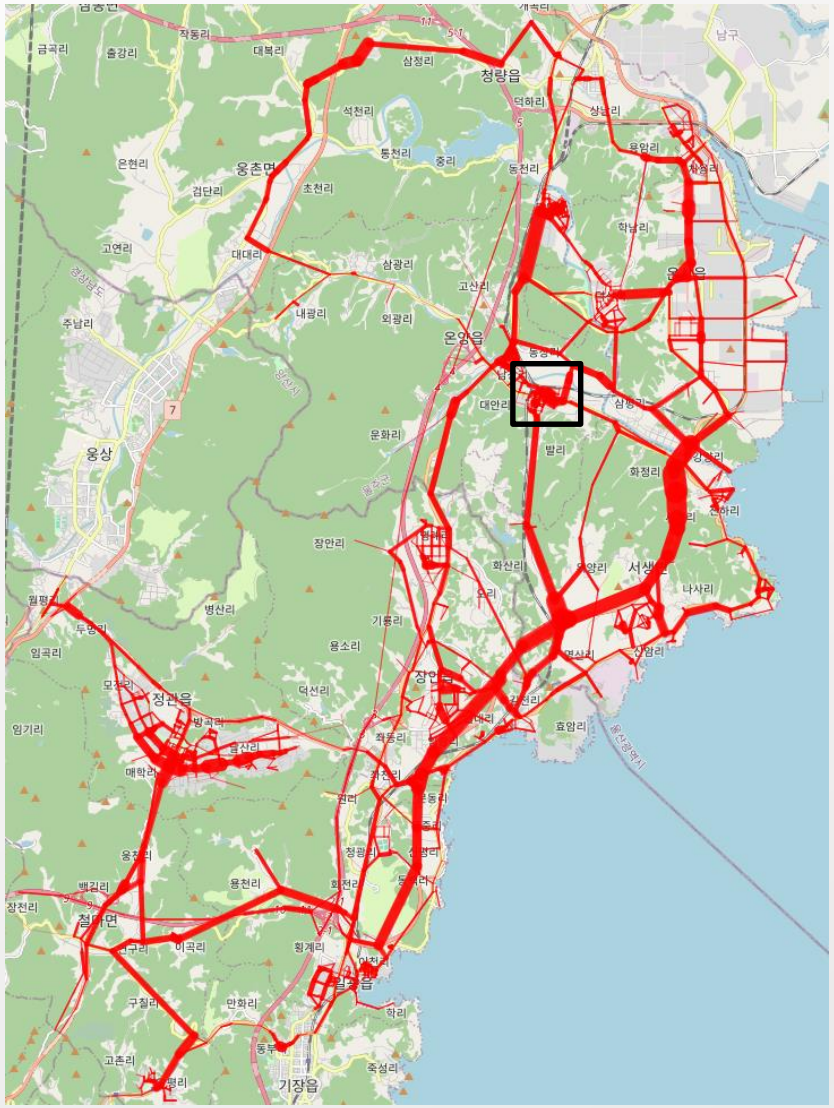
- 학습이 현재 잘 돌아가는 상태이며, 학습 결과 등은 분석이 필요함.

- 주요 디버깅 사항.
  - Planning model 학습 과정에서 A metrics에서 값이 1인 index를 이용한 데이터 처리가 이뤄 지는 것으로 보임.
  - 새롭게 전처리된 데이터는 대전 데이터와는 다르게 A가 연결 되었을 때 1이 아니라 real value로 나와 있어서 오류가 발생했는데, 해당 부분만 잘 처리해주었더니 잘 돌아갔음.

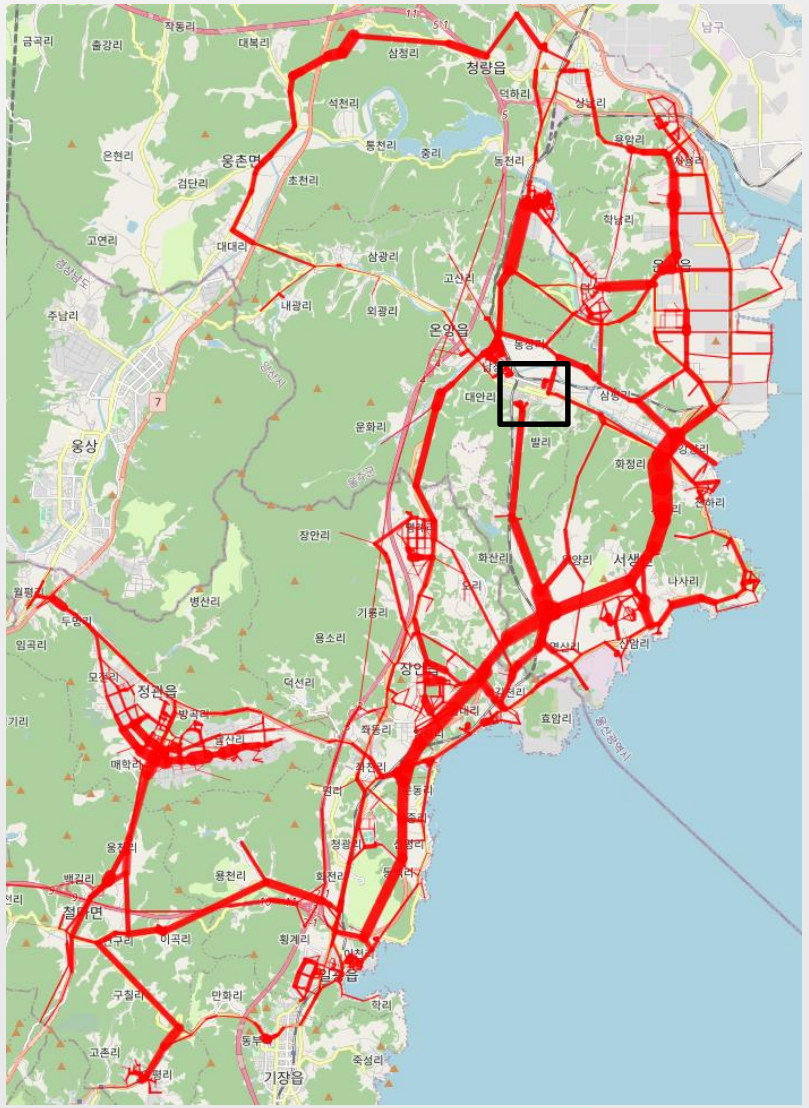
```
A tensor([[0., 0., 3.,  ..., 0., 0., 0.],
        [0., 0., 1.,  ..., 0., 0., 0.],
        [3., 1., 0.,  ..., 0., 0., 0.],
        ...,
        [0., 0., 0.,  ..., 0., 0., 0.],
        [0., 0., 0.,  ..., 0., 0., 0.],
        [0., 0., 0.,  ..., 0., 0., 0.]], dtype=torch.float64)
torch.Size([1453, 1453])
```

```
vertex: 1453
e: 0, i: 1, train loss:  81.6637, test loss:  76.0507
e: 0, i: 100, train loss:  34.5211, test loss:  24.2577
e: 0, i: 200, train loss:  23.3174, test loss:  19.2608
e: 0, i: 300, train loss:  18.4191, test loss:  15.6203
e: 0, i: 400, train loss:  16.5314, test loss:  14.7123
e: 0, i: 500, train loss:  15.7294, test loss:  14.5084
e: 0, i: 600, train loss:  14.9684, test loss:  13.1589
e: 0, i: 700, train loss:  14.0188, test loss:  12.9109
e: 0, i: 800, train loss:  13.9274, test loss:  12.3947
e: 0, i: 900, train loss:  13.0241, test loss:  11.6702
e: 0, i: 1000, train loss:  13.0614, test loss:  11.9751
e: 0, i: 1100, train loss:  13.0016, test loss:  11.7452
e: 0, i: 1200, train loss:  12.5043, test loss:  11.4993
e: 0, i: 1300, train loss:  12.4687, test loss:  11.8190
e: 0, i: 1400, train loss:  12.3357, test loss:  11.2937
e: 0, i: 1500, train loss:  11.5615, test loss:  10.8483
e: 0, i: 1600, train loss:  11.7049, test loss:  11.1007
e: 0, i: 1700, train loss:  11.4263, test loss:  11.3993
e: 0, i: 1800, train loss:  11.3835, test loss:  10.4769
e: 0, i: 1900, train loss:  11.4289, test loss:  10.6850
e: 0, i: 2000, train loss:  11.0827, test loss:  10.5319
e: 0, i: 2100, train loss:  10.8716, test loss:  10.9036
e: 0, i: 2200, train loss:  11.0501, test loss:  10.2465
e: 0, i: 2300, train loss:  11.1572, test loss:  10.1733
e: 0, i: 2400, train loss:  10.7059, test loss:  9.1576
e: 0, i: 2500, train loss:  10.4354, test loss:  10.5867
e: 0, i: 2600, train loss:  10.6141, test loss:  9.1517
e: 0, i: 2700, train loss:  10.5060, test loss:  9.9299
e: 0, i: 2800, train loss:  10.2231, test loss:  9.5562
```

# Sampling of Exceptional Cases



General case



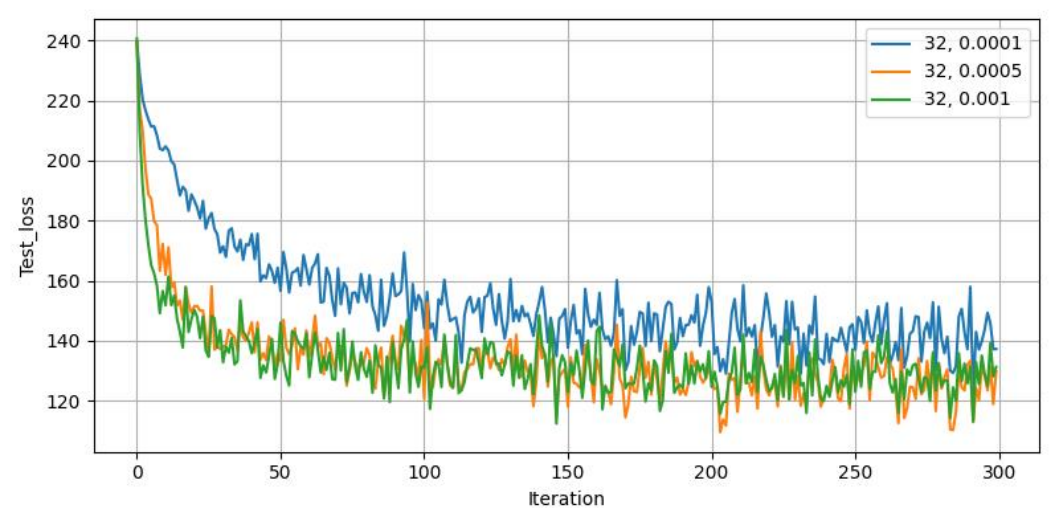
Exceptional case

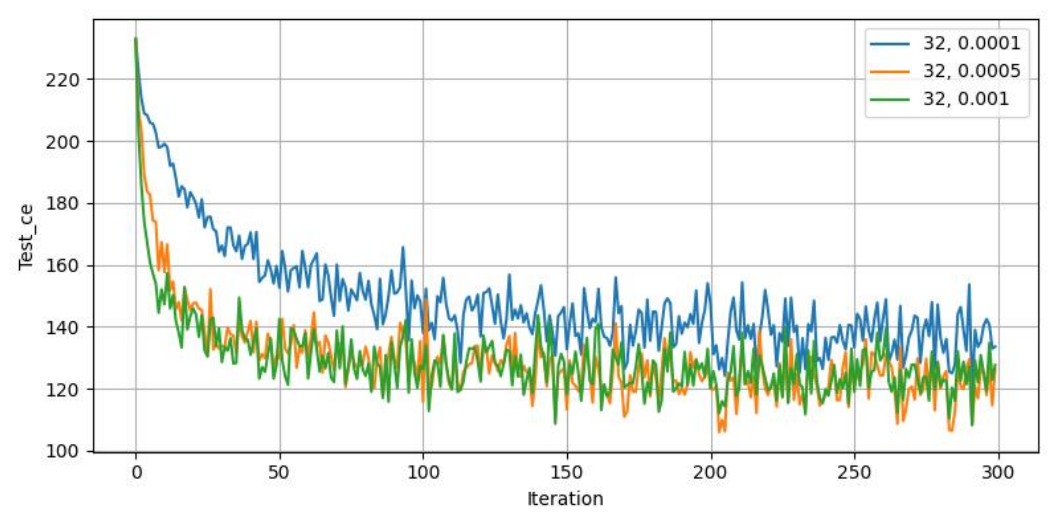
# APPENDIX

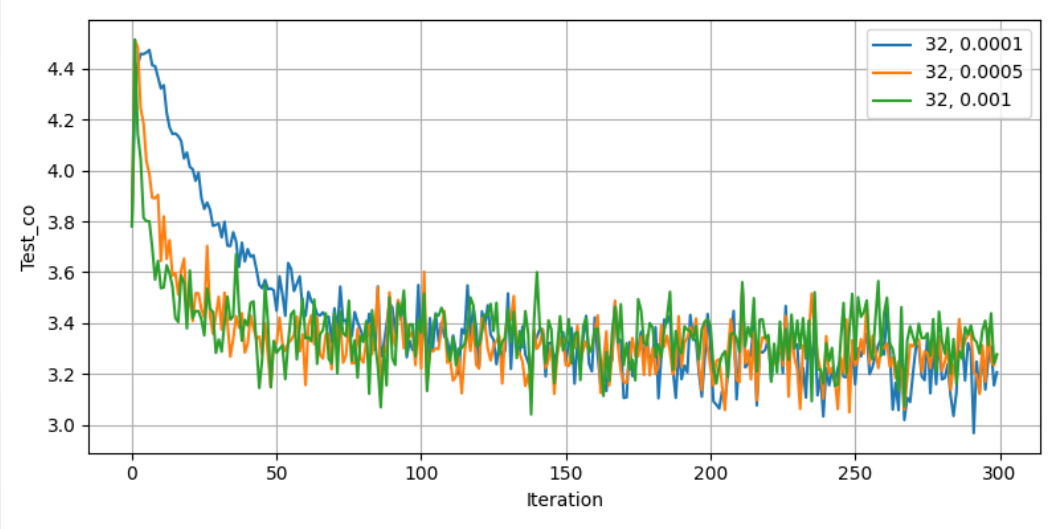
# Hyper-parameter search for unconditional training

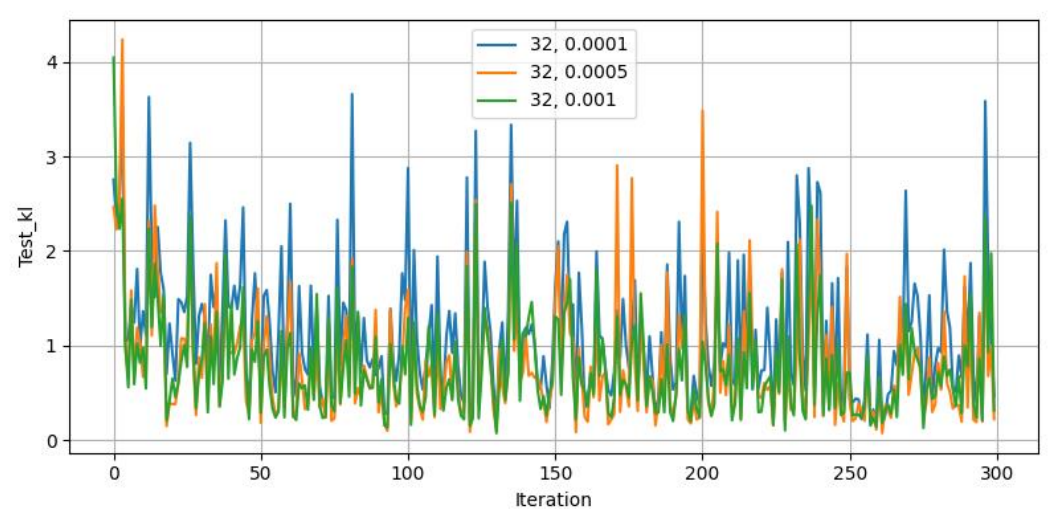
- 32 batch, 10 epoch, learning rate [0.0001, 0.0005, 0.001]

Loss = KL + CE + CO



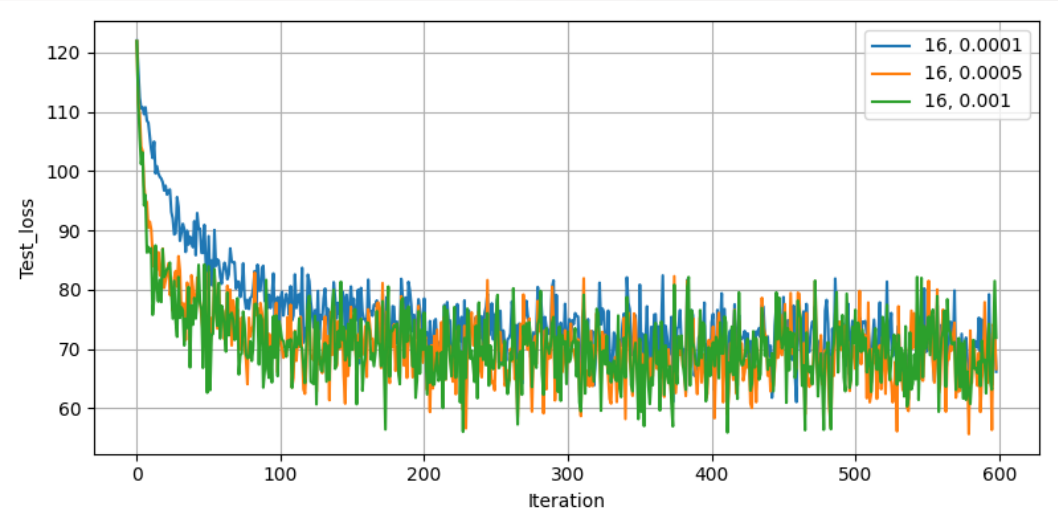



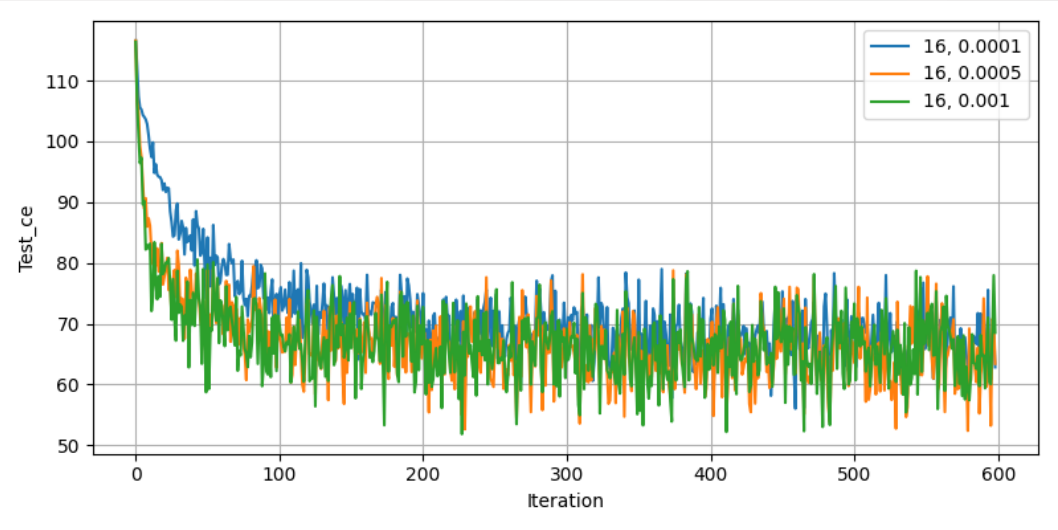



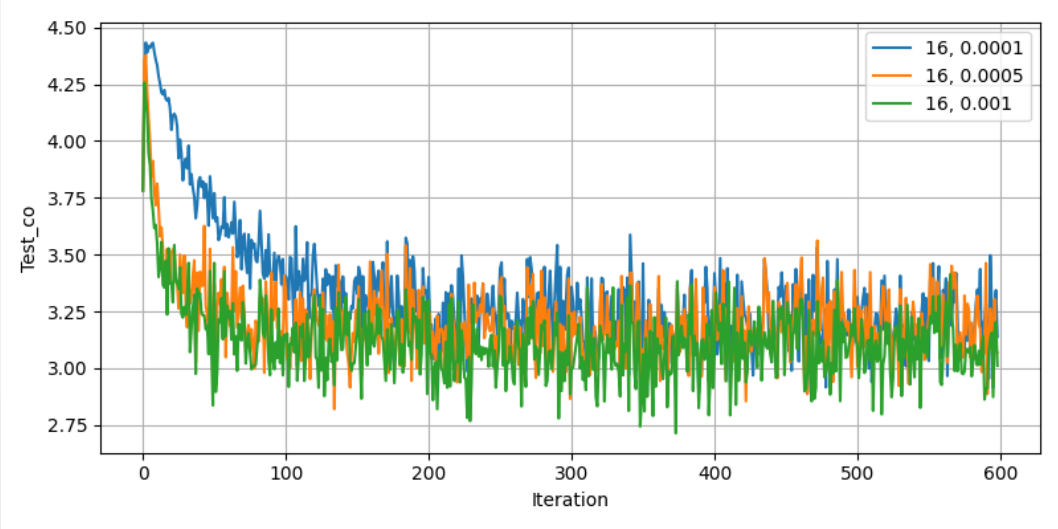
# Hyper-parameter search for unconditional training

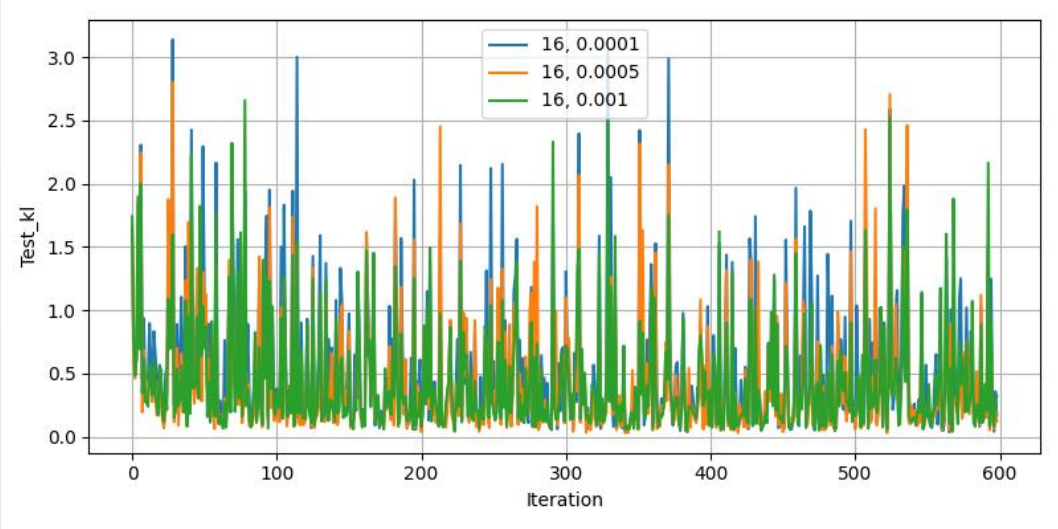
- 16 batch, 10 epoch, learning rate [0.0001, 0.0005, 0.001]

Loss = KL + CE + CO



# Hyper-parameter search for unconditional training

- 8 batch, 10 epoch, learning rate [0.0001, 0.0005, 0.001]

Loss = KL + CE + CO