Learning to Transfer for Traffic Forecasting via Multi-task Learning

Yichao Lu
Layer 6 AI

About Myself

- Senior Machine Learning Research Scientist, Layer 6 AI, Toronto, Canada

- B.Sc. Fudan University 2013 - 2017, M.Sc. University of Toronto 2017 - 2019

- Passionate about competitive machine learning:
  - 1st place winner: ACM RecSys Challenge 2018
  - 1st place winner: Kaggle ICCV 2019 Open Images Challenge (Visual Relationship Track)
  - 1st place winner: The Stanford Question Answering Dataset (SQuAD) 2.0
  - 1st place winner: IEEE BigData Cup 2021 Science4cast Challenge
  - 2nd place winner: ACM RecSys Challenge 2019
  - 2nd place winner: Microsoft MIND News Recommendation Challenge
  - 2nd place winner: CVPR 2021 ActivityNet Challenge (Scene-graph Generation Track)
  - …

layer 6
AI at TD

Task Formulation

- The Traffic4cast competition series challenge participants to predict short-term large-scale traffic states in selected cities.

- Traffic4cast 2021 poses additional challenges for participants, which seek to capture the underlying patterns of traffic flow that are both robust and transferable. The robustness of the solutions is benchmarked through
    - The core challenge, where models need to adapt to a drastic temporal domain shift due to the Covid-19 pandemic.
    - The extended challenge, where models need to predict the traffic flow in entirely new cities.

layer 6
AI at TD

Data Preparation

- The input is a 12x495x436x8 tensor consisting of a stack of 12 consecutive heatmaps of 5 minute interval time bins, spanning a total of 1 hour.
- We concatenate the 12 heatmaps across the channel dimension, which results in a tensor of shape 495x436x96. We additionally concatenate the static information of shape 495x436x9 to the input.
- The model takes as input the 495x436x105 tensor, and outputs a tensor of shape 495x436x48. The output is then reshaped into 6x495x436x8, where the 6 predicted heatmaps correspond to the predicted traffic states 5, 10, 15, 30, 45, 60 minutes in the future, respectively.

Data Preparation

- The input is a 12x495x436x8 tensor consisting of a stack of 12 consecutive heatmaps of 5 minute interval time bins, spanning a total of 1 hour.
- We concatenate the 12 heatmaps across the channel dimension, which results in a tensor of shape 495x436x96. We additionally concatenate the static information of shape 495x436x9 to the input.
- The model takes as input the 495x436x105 tensor, and outputs a tensor of shape 495x436x48. The output is then reshaped into 6x495x436x8, where the 6 predicted heatmaps correspond to the predicted traffic states 5, 10, 15, 30, 45, 60 minutes in the future, respectively.

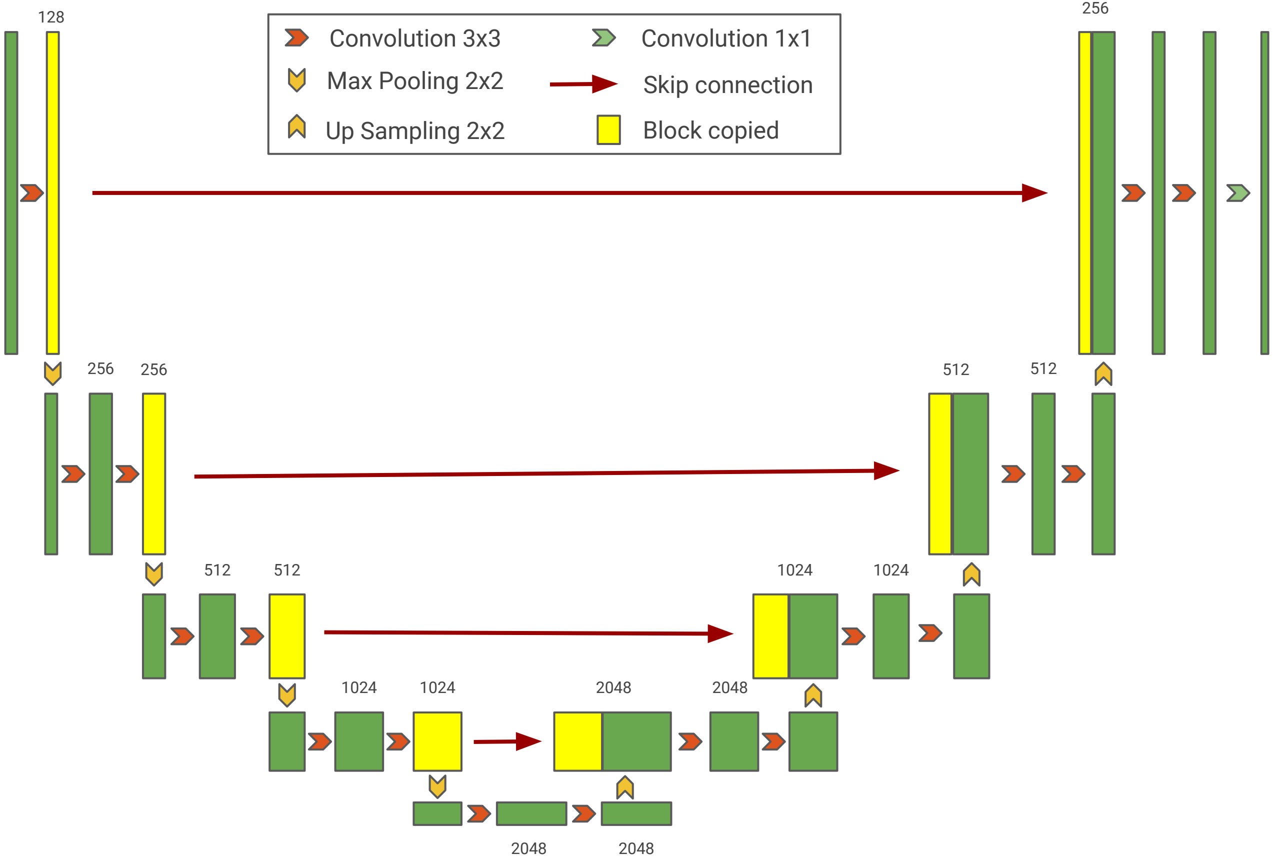
layer 6
AI at TD

Data Preparation

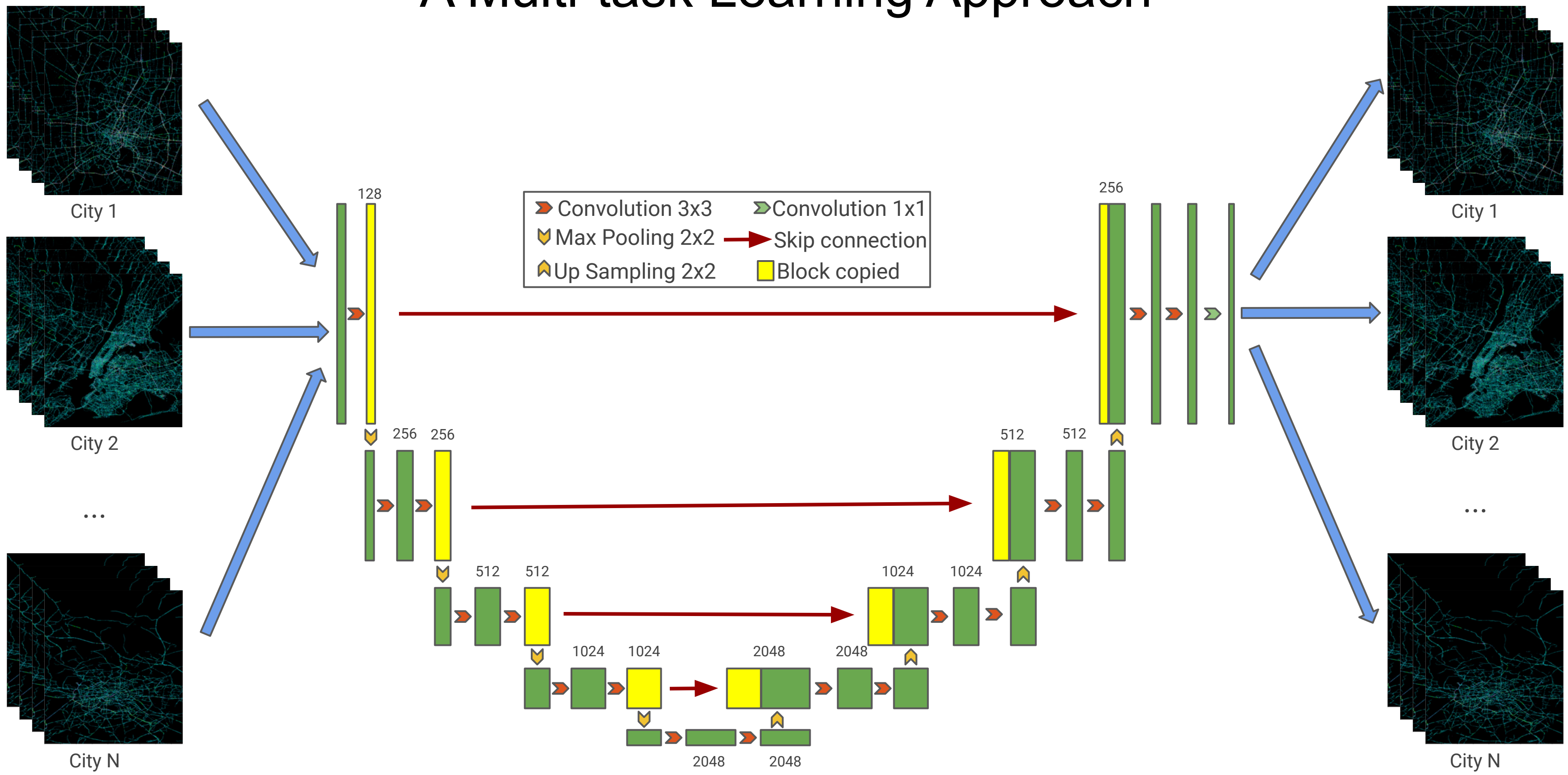
- The input is a 12x495x436x8 tensor consisting of a stack of 12 consecutive heatmaps of 5 minute interval time bins, spanning a total of 1 hour.
- We concatenate the 12 heatmaps across the channel dimension, which results in a tensor of shape 495x436x96. We additionally concatenate the static information of shape 495x436x9 to the input.
- The model takes as input the 495x436x105 tensor, and outputs a tensor of shape 495x436x48. The output is then reshaped into 6x495x436x8, where the 6 predicted heatmaps correspond to the predicted traffic states 5, 10, 15, 30, 45, 60 minutes in the future, respectively.

layer 6
AI at TD

U-Net for Traffic Forecasting



A Multi-task Learning Approach



Implementation Details

Core Competition

- Optimizing with the Adam optimizer, batch size of 8 and a learning rate of 1e-4.
- U-Net architecture with 4 downsampling layers and 4 upsampling layers.
- Training for 5 epochs.

Extended Competition

- Optimizing with the Adam optimizer, batch size of 8 and a learning rate of 1e-4.
- U-Net architecture with 1 downsampling layer and 1 upsampling layer.
- Training for 50,000 steps.

layer 6
AI at TD

Core Challenge Results

Method	MSE	Training time (hours)
Naive Average	53.406	-
U-Net [41]	49.127	24.8
Graph ResNet [27]	49.546	48.4
U-Net + AdaBN [48]	49.257	24.8
U-Net + DaNN [32]	49.096	26.4
U-Net + DDC [33]	49.223	28.8
U-Net + DeepCORAL [34]	49.230	26.8
U-Net + ADDA [35]	49.176	27.6
U-Net + DANN [36]	49.104	26.8
U-Net + DSN [37]	49.072	26.4
U-Net + MAML [50]	49.054	68.4
U-Net + Multi-task Learning (ours)	**48.659**	**20.5**

layer 6 AI at TD

Core Challenge Leaderboard

Rank	Team	MSE
1	**oahciy (ours)**	**48.422**
2	sungbin	48.494
3	sevakon	49.379
4	ai4ex	49.720
5	Bo	50.219

layer 6
AI at TD

Extended Challenge Results

Method	MSE	Training time (hours)
Naive Average	63.140	-
U-Net [41]	60.114	2.6
Graph ResNet [27]	60.537	10.8
U-Net + AdaBN [48]	60.253	2.6
U-Net + DaNN [32]	60.077	2.8
U-Net + DDC [33]	60.274	3.0
U-Net + DeepCORAL [34]	60.361	2.8
U-Net + ADDA [35]	60.157	2.8
U-Net + DANN [36]	60.175	2.8
U-Net + DSN [37]	60.012	3.2
U-Net + MAML [50]	60.249	18.2
U-Net + Multi-task Learning (ours)	**59.732**	**1.2**

layer 6
AI at TD

Extended Challenge Leaderboard

Rank	Team	MSE
1	sungbin	59.559
2	**oahciy (ours)**	**59.586**
3	nina	59.915
4	dninja	60.221
5	HBKU	60.266

layer 6
AI at TD

Temporal Domain Adaptation Results (Bangkok)

Training data	MSE
{Antwerp, Barcelona, Moscow} {2019, 2020} data	35.374
Bangkok 2019 data	35.118
Bangkok 2019 data + {Antwerp, Barcelona, Moscow} 2019 data	34.977
Bangkok 2019 data + {Antwerp, Barcelona, Moscow} 2020 data	34.826
Bangkok 2019 data + Barcelona {2019, 2020} data	34.609
Bangkok 2019 + {Antwerp, Barcelona, Moscow} {2019, 2020} data	**34.315**

Temporal Domain Adaptation Results (Moscow)

Training data	MSE
{Antwerp, Bangkok, Barcelona} {2019, 2020} data	99.946
Moscow 2019 data	99.015
Moscow 2019 data + {Antwerp, Bangkok, Barcelona} 2019 data	97.847
Moscow 2019 data + {Antwerp, Bangkok, Barcelona} 2020 data	96.996
Moscow 2019 data + Barcelona {2019, 2020} data	95.375
Moscow 2019 + {Antwerp, Bangkok, Barcelona} {2019, 2020} data	**95.017**

layer 6
AI at TD

Discussion and Future Work

- Possible avenues of further improvements:
  - Utilizing more sophisticated U-Net architecture.
  - Utilizing other state-of-the-art spatio-temporal learning models.
  - Experimenting with a per-city optimization for the multi-task learning approach.
  - Adding manually designed features such as the location of the pixel in the heatmap, the time of day for the predicted traffic flows, and whether it is a weekend or a holiday.

Thank You

layer 6
AI at TD