Work Report

Deep Learning for Cellular Traffic Prediction

Chuanting Zhang

2018-09-29

Contents

Backgrounds

- Machine learning is a necessity in 5G
  - A fully functional 5G system is not going to happen without machines that can learn and make decisions by themselves

4G → Reactive

5G → Predictive
5G → Proactive
5G → Anticipatory

- ***Huawei***: The road to 5G is paved with AI[1]
- ***Ericsson***: AI is key to fixing network complexity from 5G, IoT[2]
- ***ITU***: Launches new Focus Group to study machine learning in 5G systems[3]
- ……

1. https://www.forbes.com/sites/adrianbridgwater/2018/02/08/huawei-the-road-to-5g-is-paved-with-ai/#4f14a8f17457
2. https://www.techrepublic.com/article/ericsson-ai-is-key-to-fixing-network-complexity-from-5g-iot/
3. https://news.itu.int/itu-launches-new-focus-group-study-machine-learning-5g-systems/

Backgrounds

- Machine learning for wireless communications

Machine Learning in 5G and Beyond

Supervised learning

Regression model, KNN, SVM, Bayesian learning

- Channel identification
- Traffic prediction
- Massive MIMO channel estimation/detection
- User location/behavior learning/classification

Unsupervised learning

Clustering algorithm, PCA, ICA

- MTC devices clustering
- Small cell clustering
- Anomalies detection
- HetNet clustering
- Signal dimension reduction

Reinforcement learning

MDP, POMDP, Q-learning, multi-armed bandit

- Decision making under unknown network conditions
- Energy modeling in energy harvesting
- HetNet selection/association

Jiang C, Zhang H, Ren Y, et al. Machine Learning Paradigms for Next-Generation Wireless Networks. *IEEE Wireless Communications*, vol.24,no.2,pp.98-105,2016.

Backgrounds

☐ Cellular traffic prediction

- Forecasting the future traffic volume based on knowledge of the past, and information from cross domain datasets

☐ Why it matters now?

- ***Improve network management***: dynamic network congestion control
- ***Reduce operating expenditure***: accurate radio resource purchase
- ***Enhance energy efficiency***: intelligent BS on/off

Preliminaries—Categories of Traffic Prediction

☐ Cellular traffic prediction → Methods

- Traffic prediction (time series prediction)
  - Statistical
    - Linear model
      - ARIMA
      - Exponential Smoothing
    - Non-linear model
      - ARCH
      - GARCH
  - Machine learning
    - Shallow learning
      - Linear Regression
      - SVM
    - Deep learning
      - DNN/MLP
      - LSTM
      - CNN

Preliminaries—Categories of Traffic Prediction

☐ Cellular traffic prediction → Factors

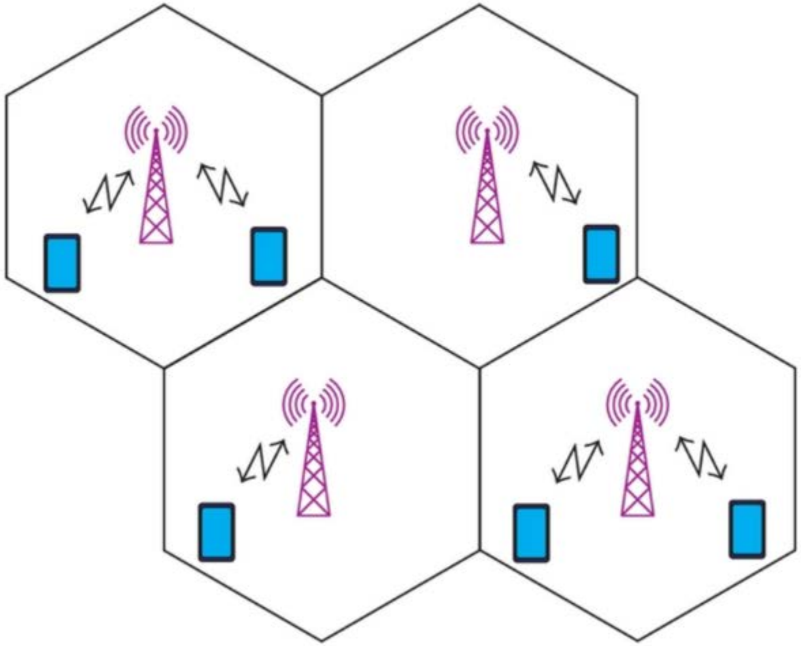
- Traffic prediction (time series prediction)
  - Spatial
    - Gotzner-Eplus-VTC’98
    - Guan-BUPT-ICCP’13
    - Zhang-ZJU-ICT’16
  - Temporal
    - Ge-Globecom’04
    - Shafiq-MSU-SIGMETRICS’11
    - Xu-Tsinghua-IEEE TSC’16
  - Spatiotemporal
    - Xue-ASU-INFOCOM’17
    - Wang-Tsinghua-ICNP’17
    - Cui-UCDavis-CL’18
    - Zhang-SDU-CL’18

Cellular traffic prediction → Tasks

$$X = [x_1; x_2; x_3; x_4]$$

Traffic prediction (time series prediction)

- SISO
- SIMO
- MISO
- <u>MIMO</u>



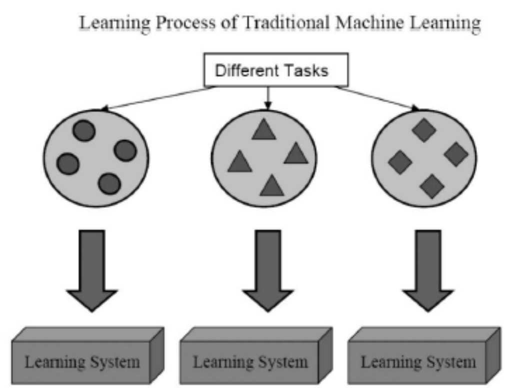
Preliminaries—Transfer Learning

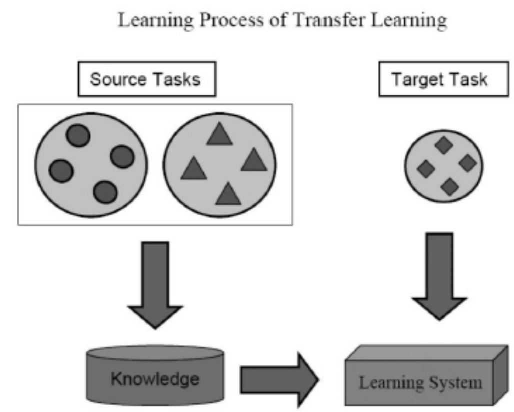
Transfer Learning

Definition: Given a source domain $\mathcal{D}_S$ and learning task $\mathcal{T}_S$, a target domain $\mathcal{D}_T$ and learning task $\mathcal{T}_T$, transfer learning aims to help improve the learning of the target predictive function $f_T(\cdot)$ in $\mathcal{D}_T$ using the knowledge in $\mathcal{D}_S$ and $\mathcal{T}_S$, where $\mathcal{D}_S \neq \mathcal{D}_T$, or $\mathcal{T}_S \neq \mathcal{T}_T$.

迁移学习是指利用**数据**、**任务**、或**模型**之间的**相似性**，将在旧领域学习过的模型，应用于新领域的一种学习过程

⇨





Pan S J, Yang Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 2010, 22(10): 1345-1359.

Preliminaries—Transfer Learning

- Why transfer learning?
  - ***Big data*** vs. ***limited labeled data***
    - ✓ Train a model use the labeled data and apply it to non-labeled datasets
  - ***Big data*** vs. ***limited computing ability***
    - ✓ Use the "big model" trained by the big company to our tasks
  - ***Generalization*** vs. ***personalization***
    - ✓ Fine-tune the generalized model and make it task-dependent
  - ***Specific applications***
    - ✓ Transfer knowledge between different tasks
- Categories of transfer learning
  - Instance based TL
  - Feature based TL
  - Model/parameter based TL
  - Relation based TL

Jingdong Wang. Transfer Learning Tutorial. 2018.

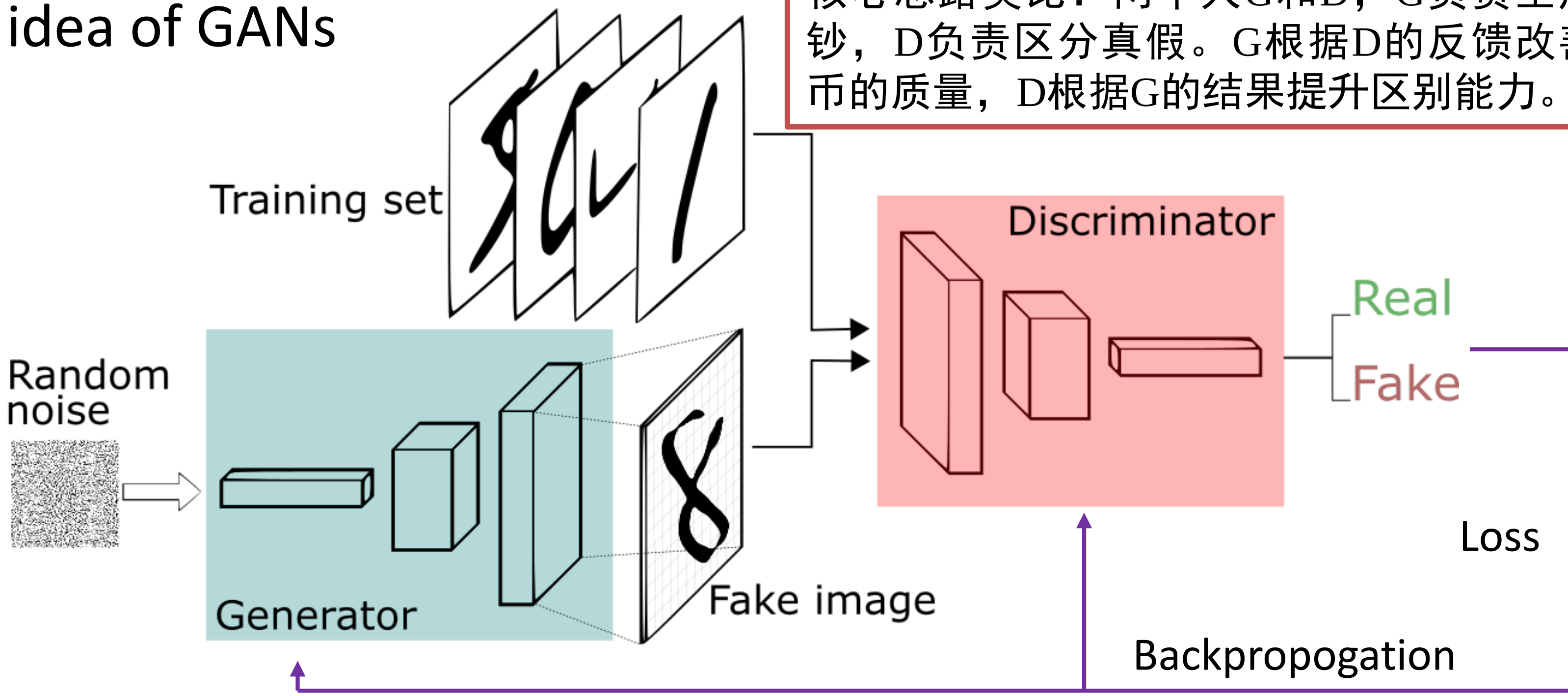
Preliminaries—Transfer Learning

- ***Deep transfer learning***: Pretrain a ConvNet on a very large dataset, and then use the ConvNet either as an initialization or a fixed feature extractor for the task of interest.
  - ConvNet as ***fixed feature extractor***
    - ✓ Take a ConvNet pretrained on ImageNet, remove the last fully-connected, then treat the rest of the ConvNet as a fixed feature extractor for the new dataset.
  - ***Fine-tuning*** the ConvNet
    - ✓ Not only replace and retrain the classifier on top of the ConvNet on the new dataset, but to also fine-tune the weights of the pretrained network by continuing the backpropagation.
  - ***Pretrained models***
    - ✓ Model zoo

Preliminaries—Generative Adversarial Networks

☐ The idea of GANs

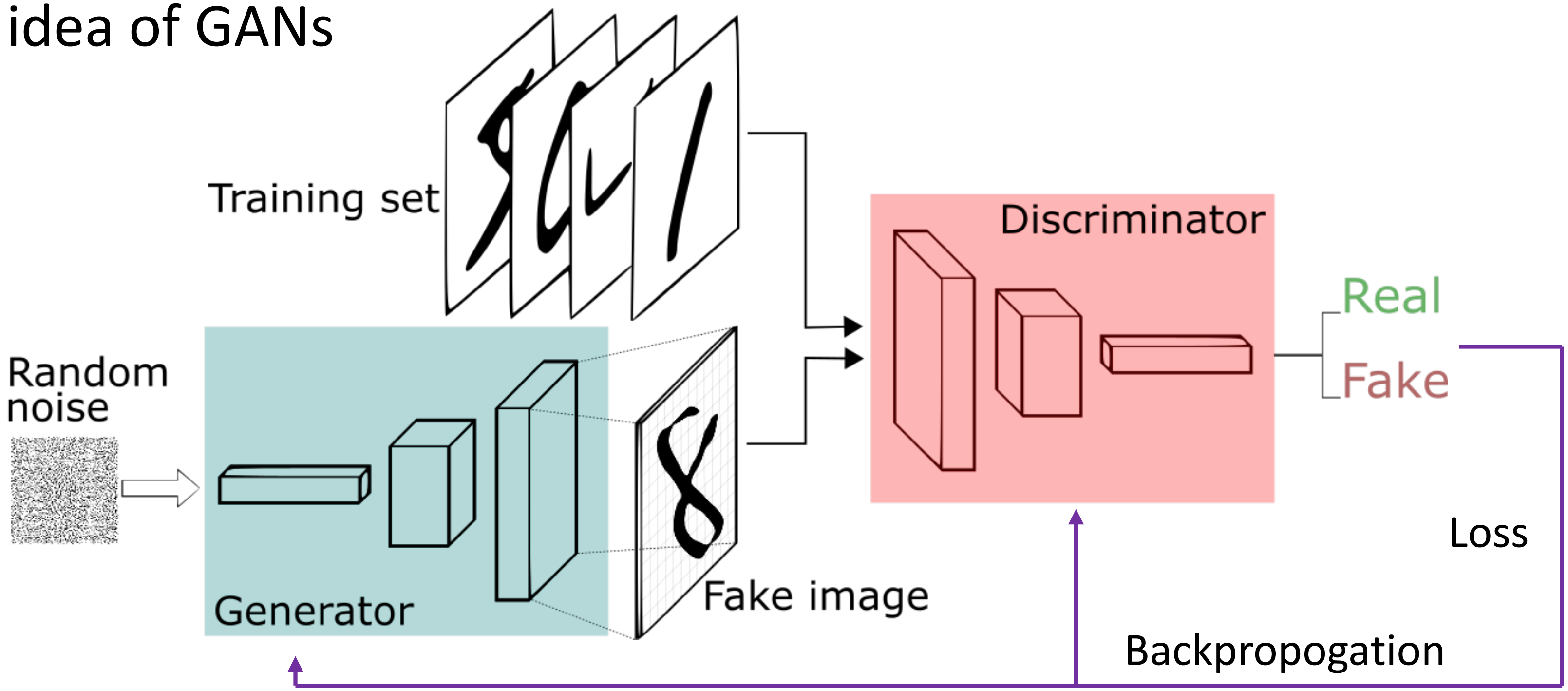
核心思路类比：两个人G和D，G负责生产假钞，D负责区分真假。G根据D的反馈改善假币的质量，D根据G的结果提升区别能力。

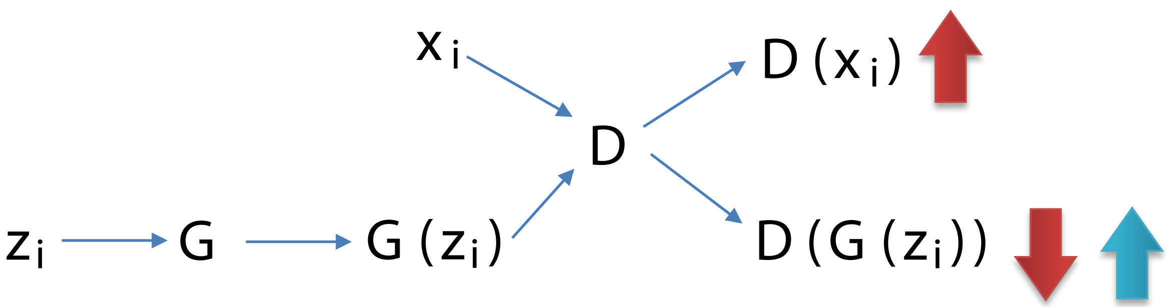


Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C], *Advances in neural information processing systems*. 2014: 2672-2680.

Preliminaries—Generative Adversarial Networks

☐ The idea of GANs

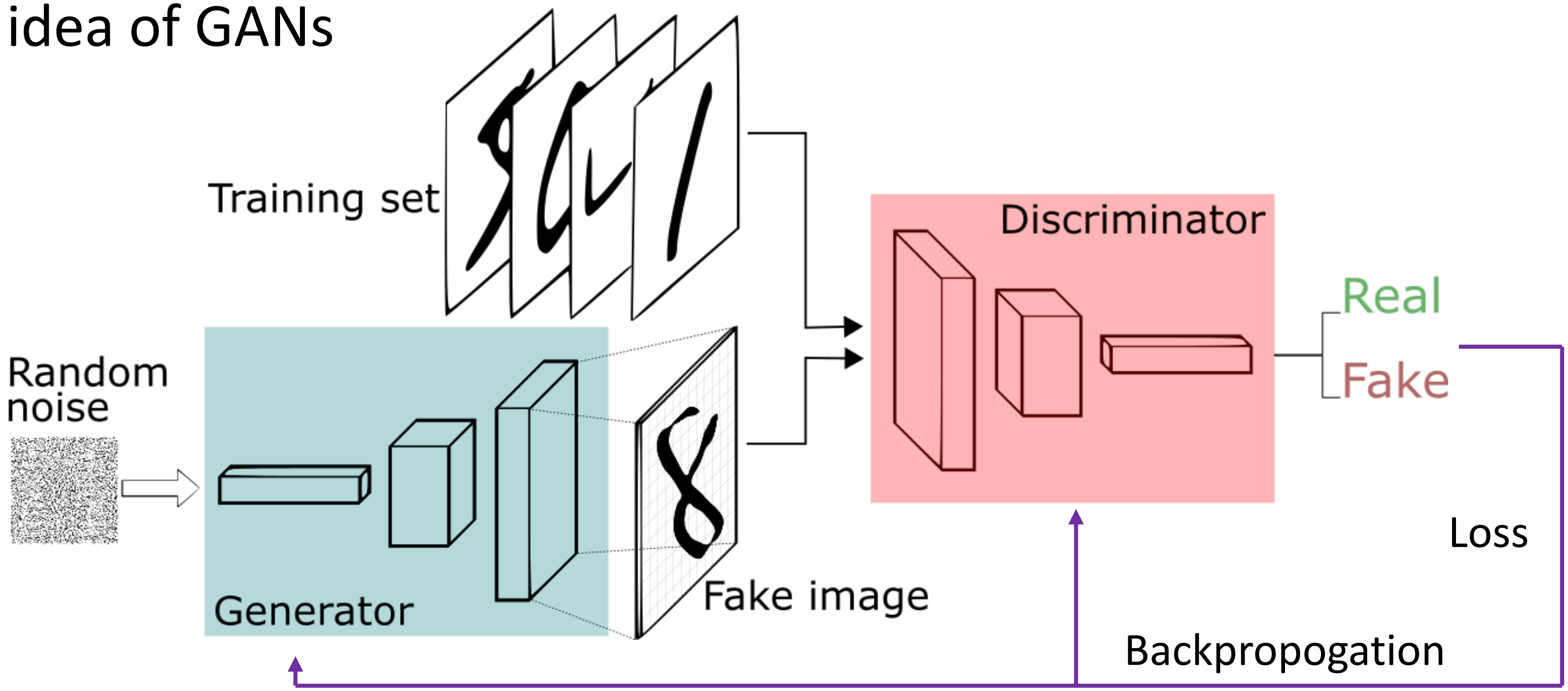




- The ***red*** arrow denotes the purpose of ***D***
- The ***blue*** one denotes the purpose of ***G***

Preliminaries—Generative Adversarial Networks

☐ The idea of GANs



Value function of minimax game **二人零和博弈**

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

Discriminator loss

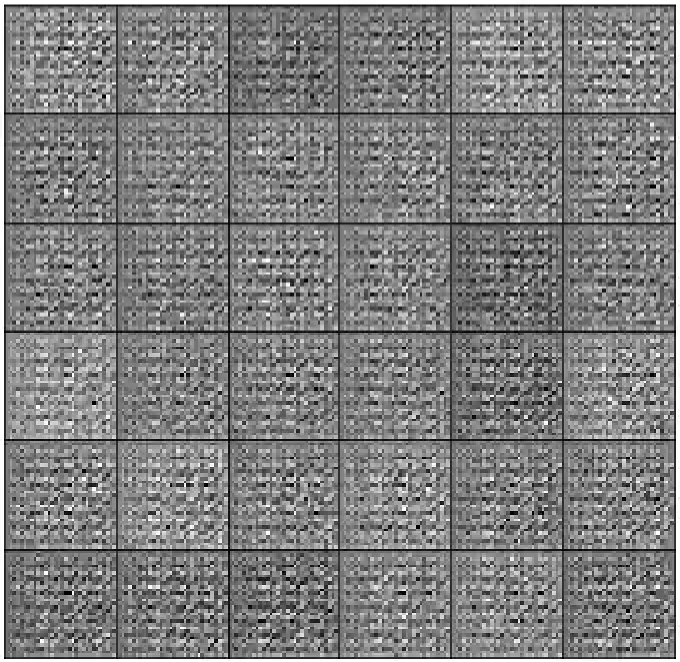
$$\frac{1}{m}\sum_{i=1}^{m}\left[\log D\left(\boldsymbol{x}^{(i)}\right) + \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right)\right]$$

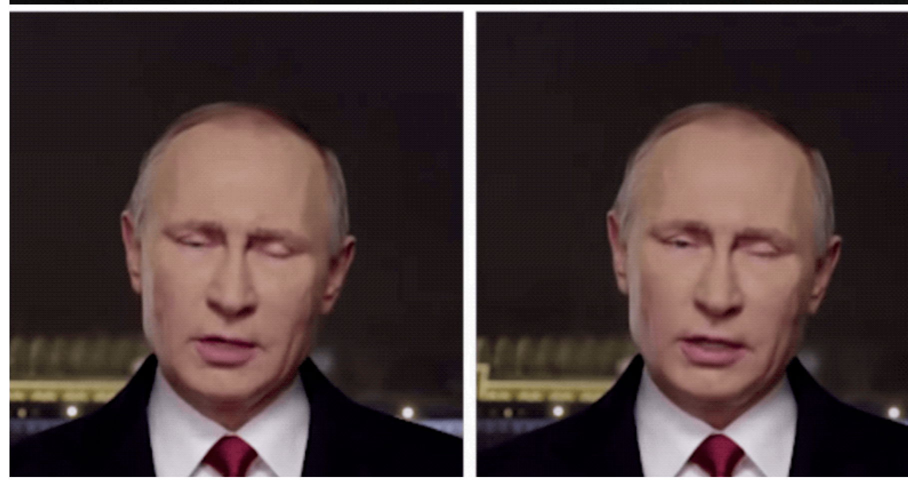
Generator loss

$$\frac{1}{m}\sum_{i=1}^{m}\log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right)$$

Preliminaries—Generative Adversarial Networks

Application of GANs

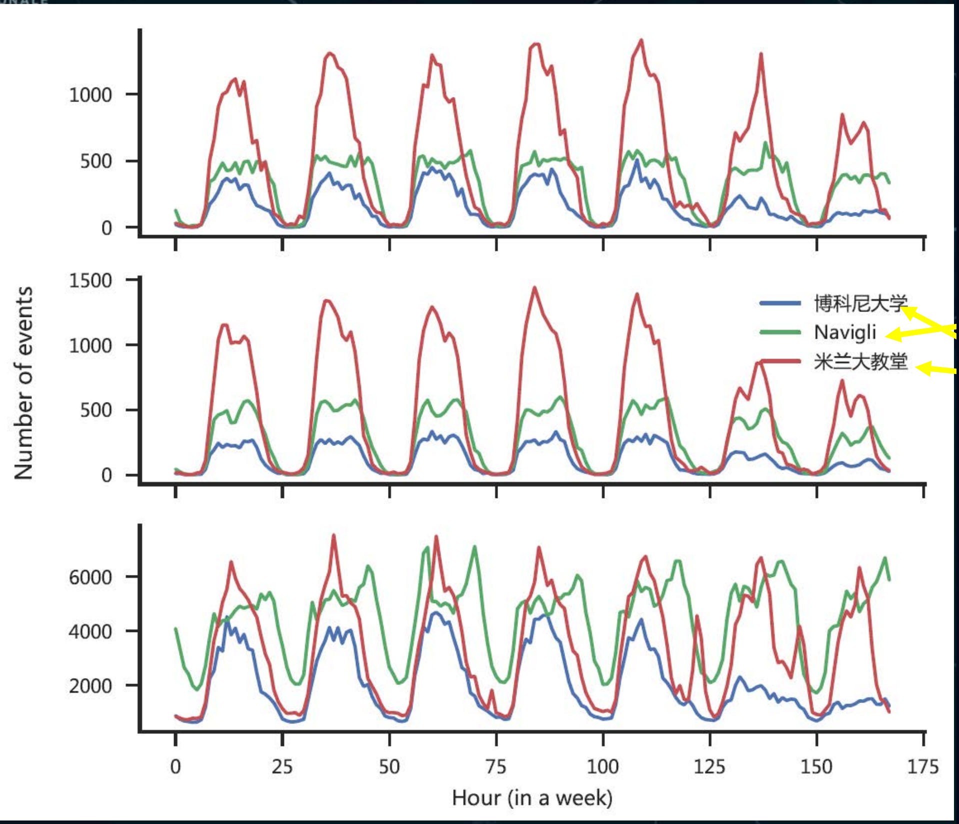


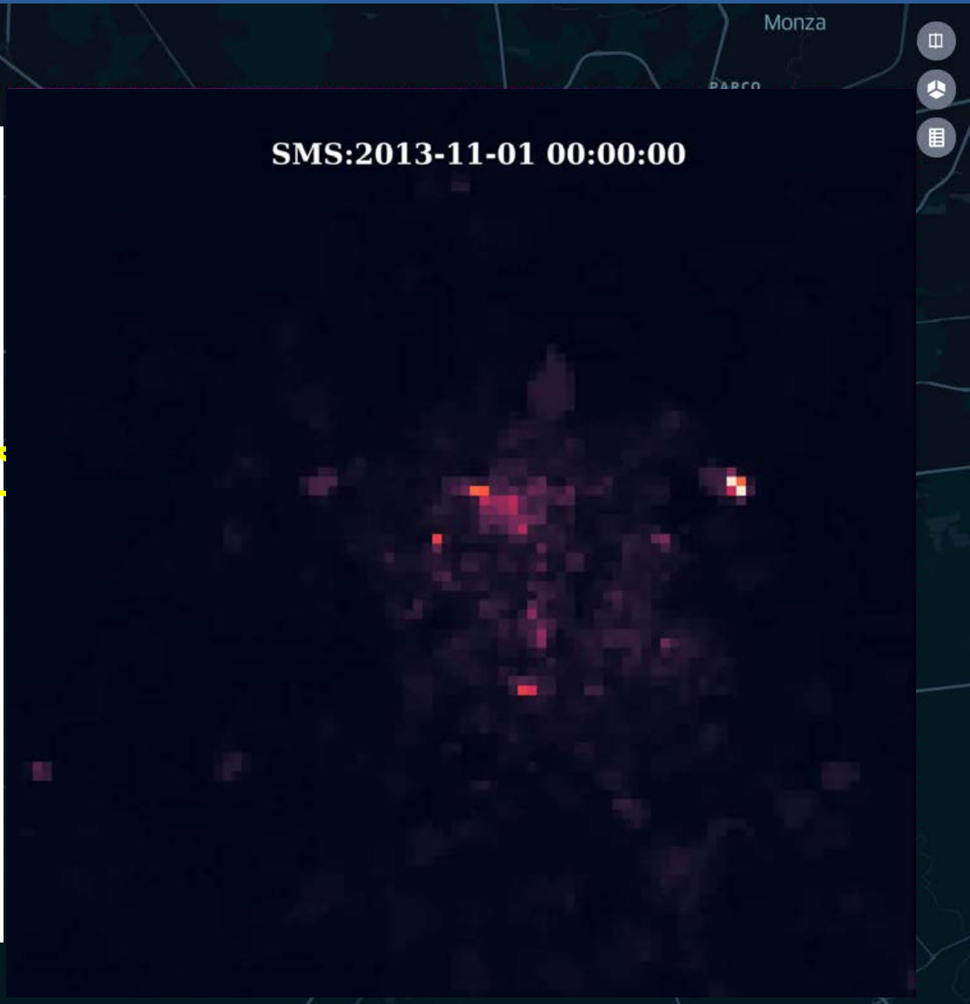


Ground Truth Generated

Contents

Cellular Traffic Prediction



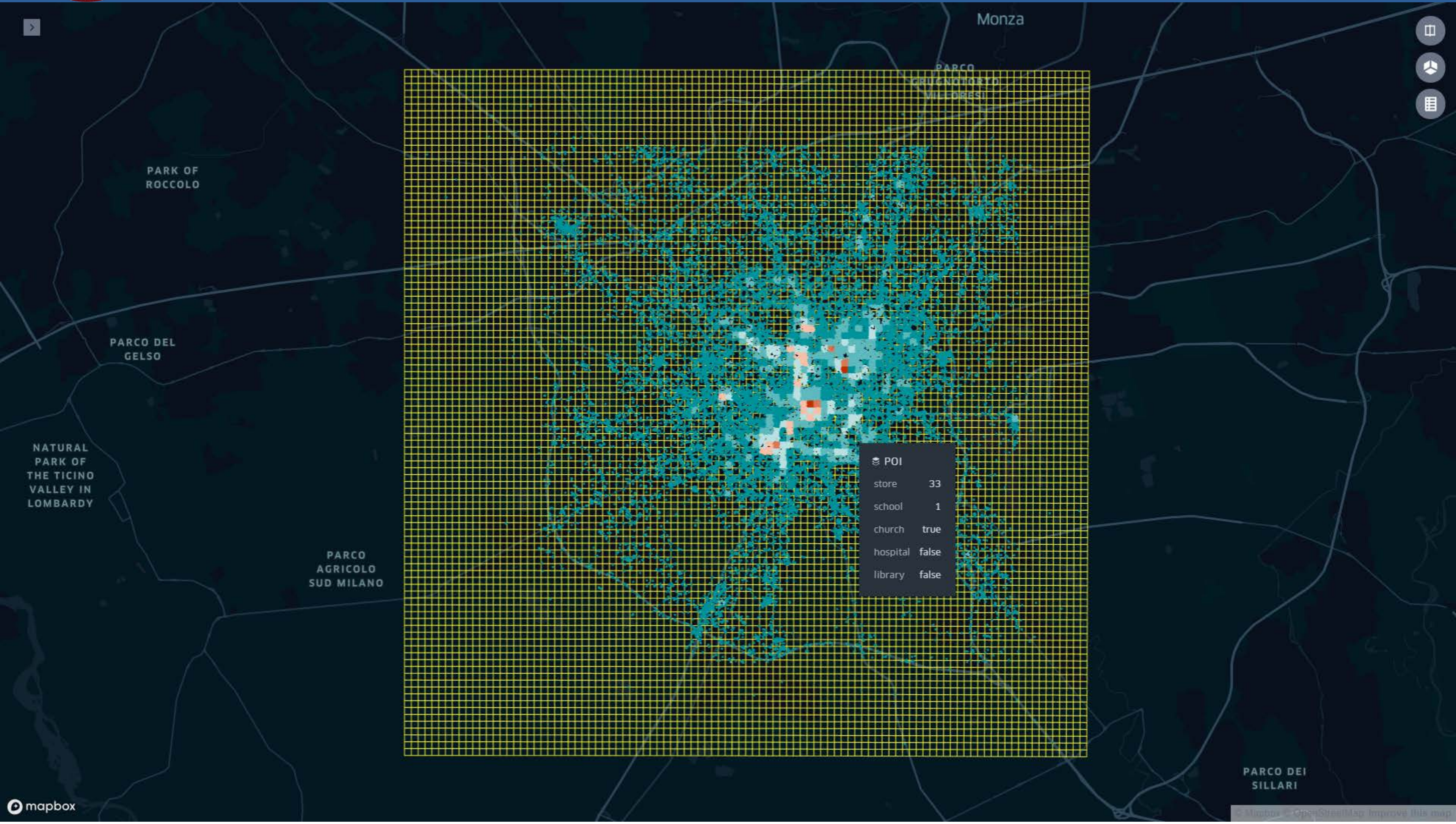


The state-of-the-art on cellular traffic prediction: Almost all of the work consider the cellular traffic itself

Existing Challenges

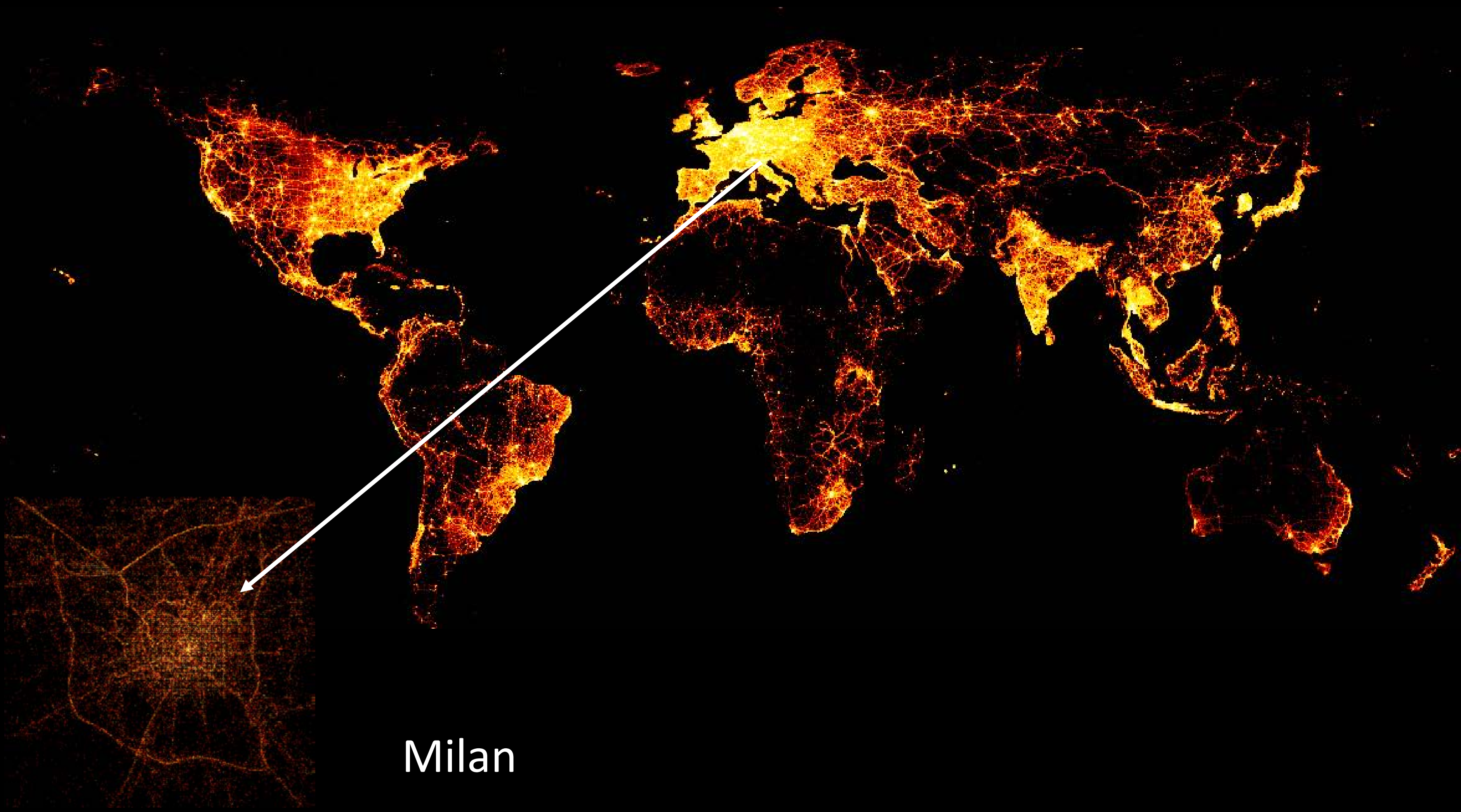
- The ***relationships*** among ***cross-domain datasets*** and different kinds of ***cellular traffic datasets*** are still unexplored in research community
  - Are the BSs distribution and POI information correlated to the traffic volume of specific cells? Is there any data available to model their relationships? How to model?
- The ***different traffic patterns in different places*** are not well captured
  - For places of CBD and a university campus, they have different traffic patterns. How to capture the traffic pattern diversity and reduce the prediction complexity at the same time?
- The ***performance is hard to improve*** using the specific kind of dataset.
  - How to boost the prediction performance using knowledge learned from other traffic/data?

To Solve the First Challenge



Cross-domain datasets

Base Station Distribution of The World

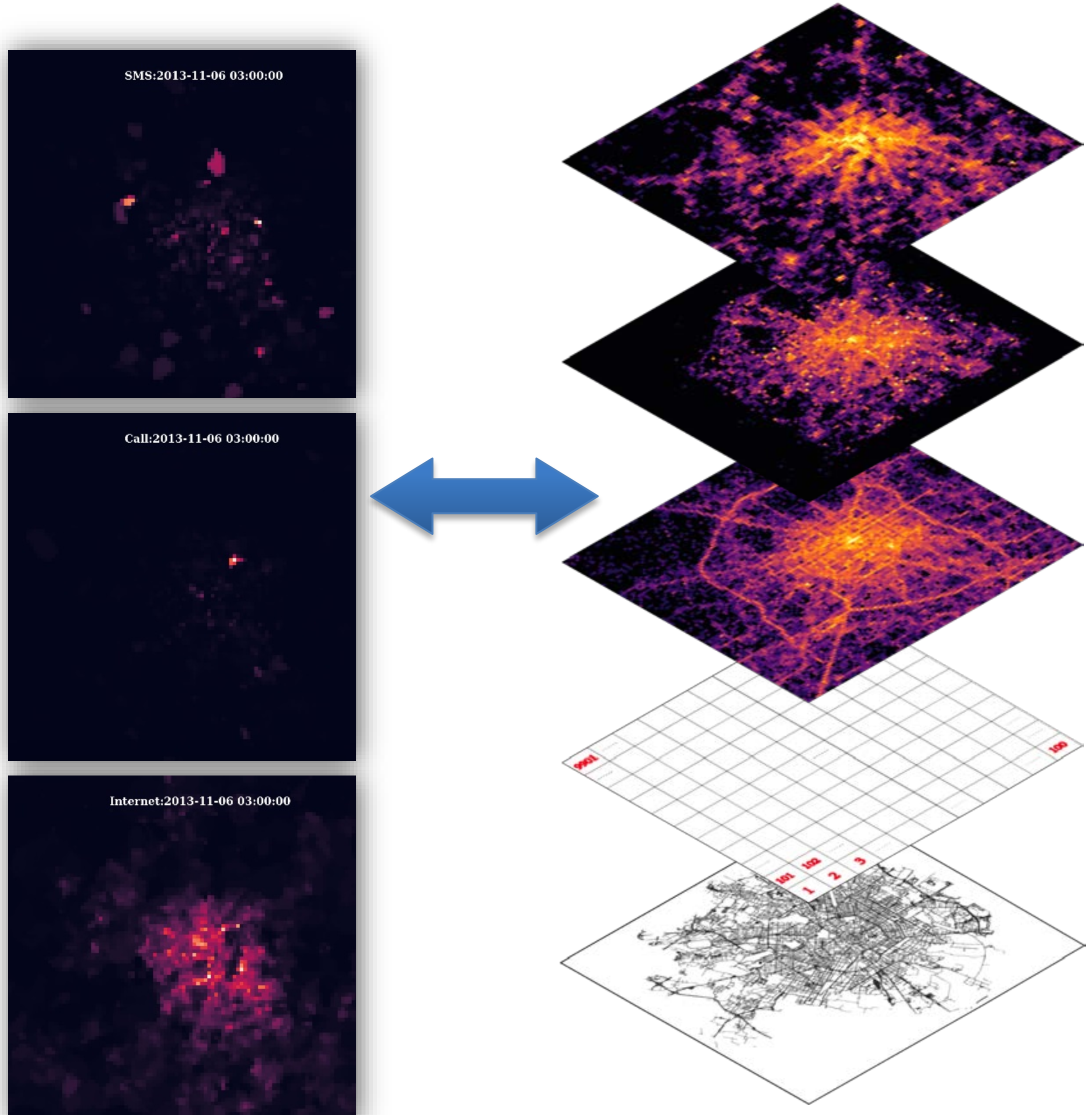


Milan

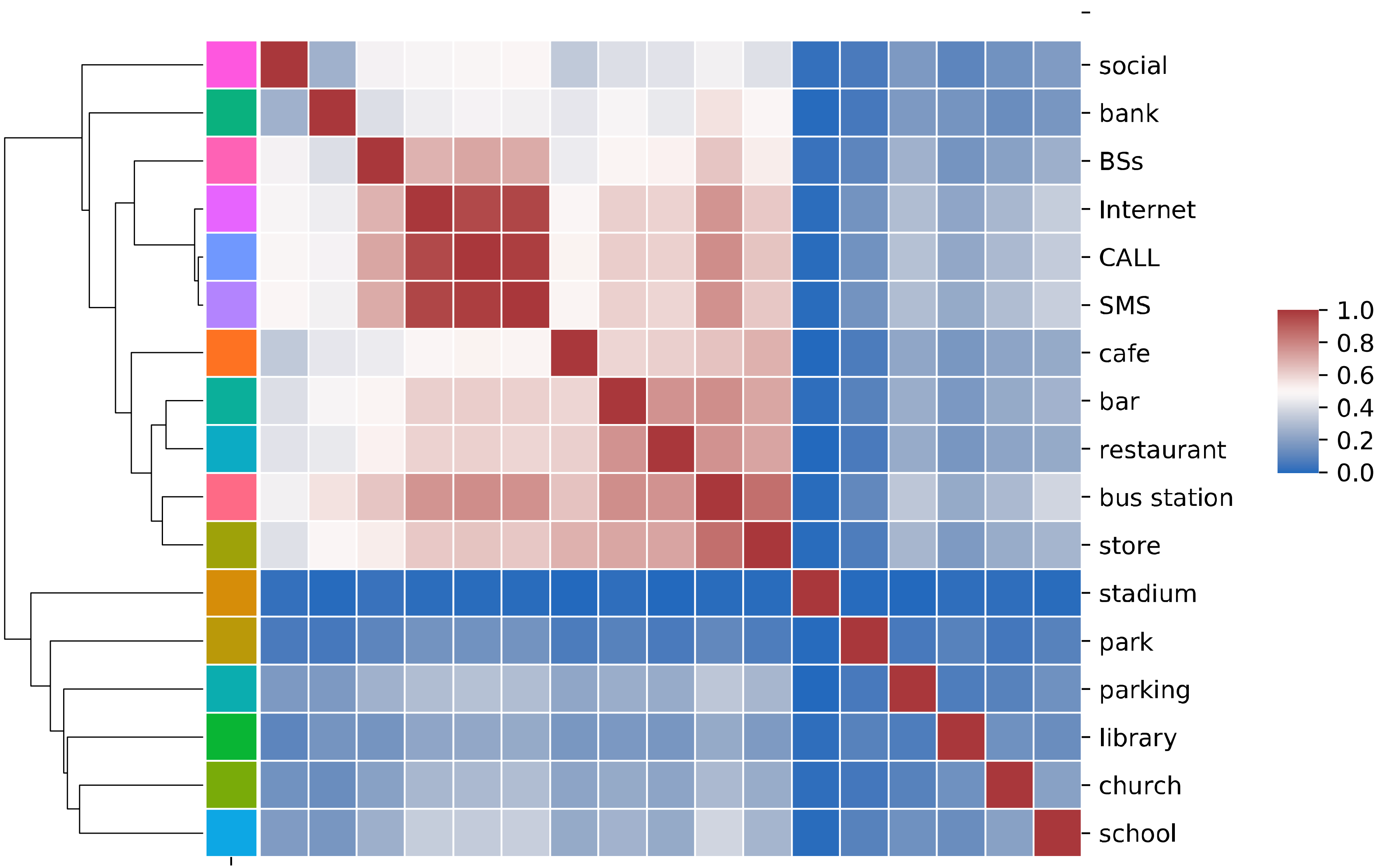
Datasets

- Detailed statistics of crawled data

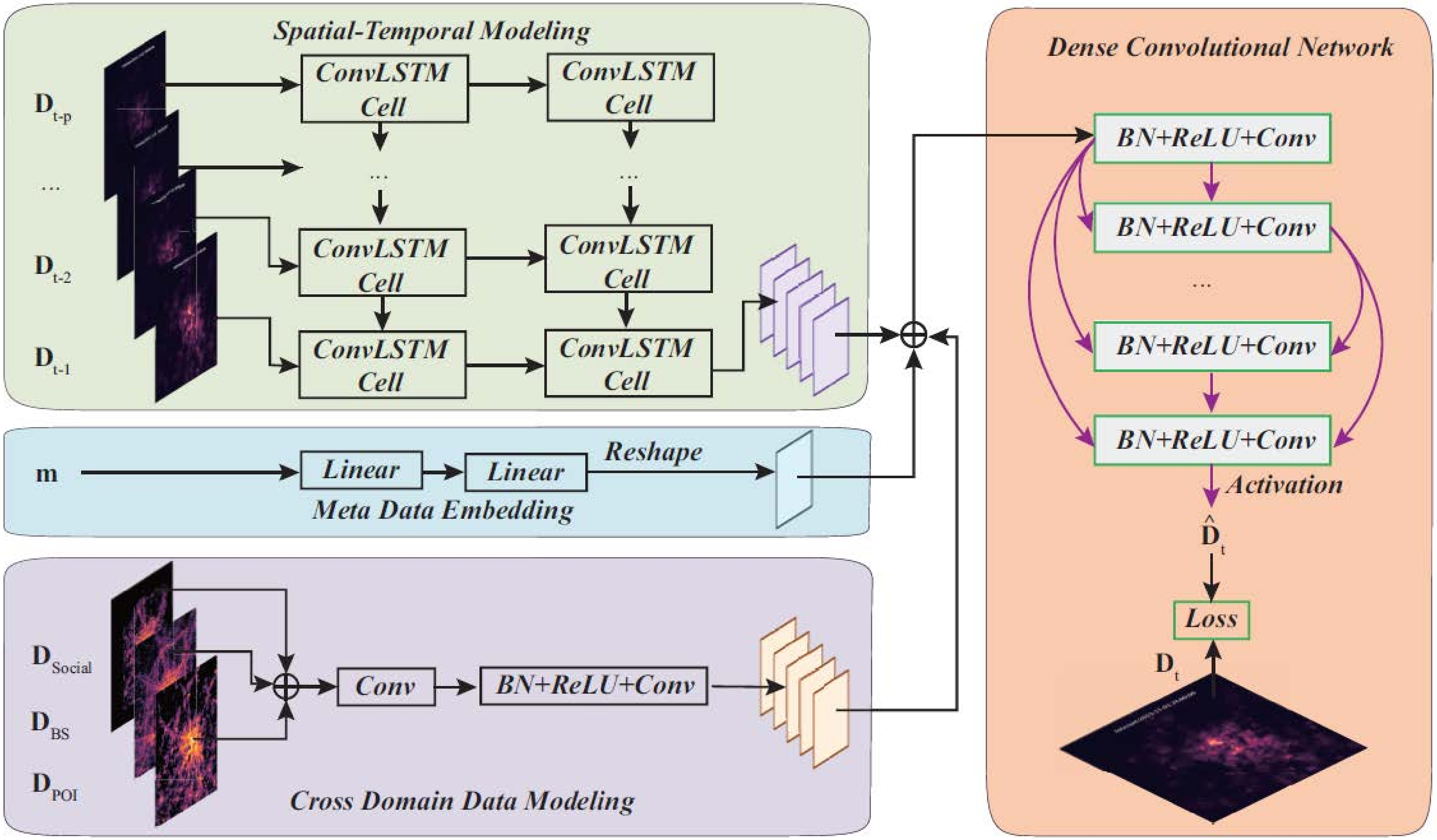
Dataset	Type	# of records
Cellular traffic	SMS / Call / Internet	$\approx$ 300 million
POI	Subway station	104658
	Store	19748
	Church	512
	Cafe	995
	Park	765
	Library	188
	Bank	882
	Bar	3192
	Parking	392
	Restaurant	4666
	School	1284
	Lodging	2922
	Hospital	1585
BSs	GSM / CDMA / LTE	69909
Social activity	Twitter	269290



Correlation Analysis



- STCNet: Spatial-Temporal Cross-domain neural Network



To Solve the Second Challenge

- Identify the city functional zones and train different models for these different areas.

D

$\downarrow$

A

$\downarrow$

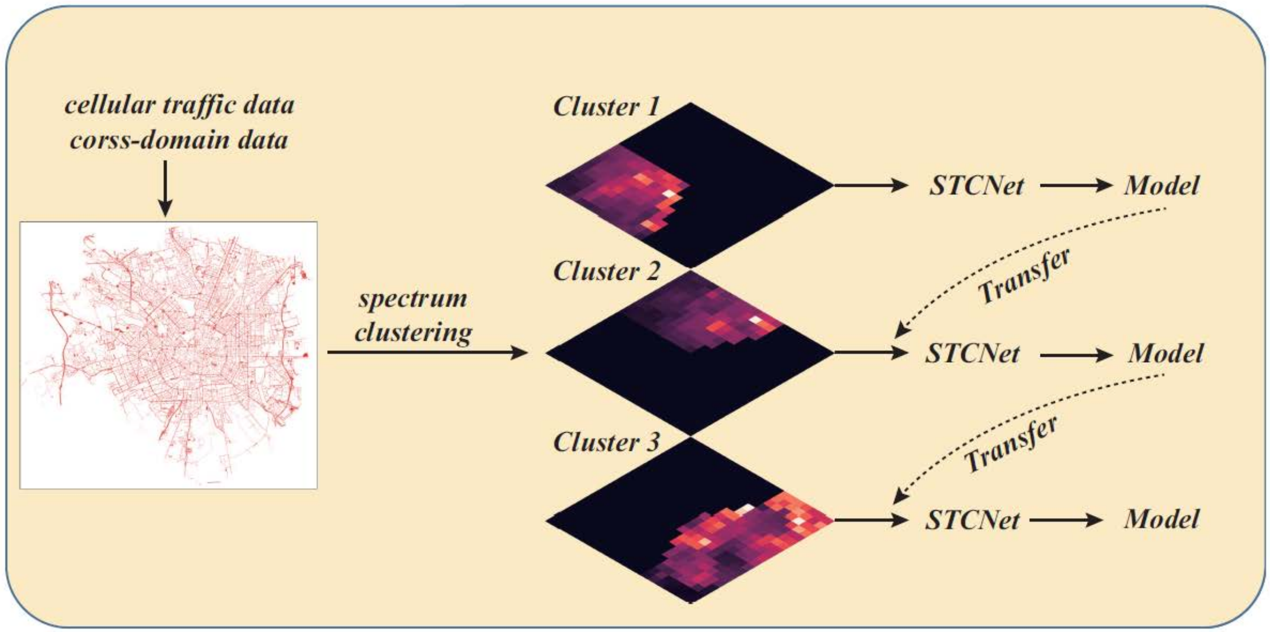
$L = P^{i\ 1=2} A P^{i\ 1=2}$

$\downarrow$

$X = [x_1; ¢¢¢; x_k]$

$\downarrow$

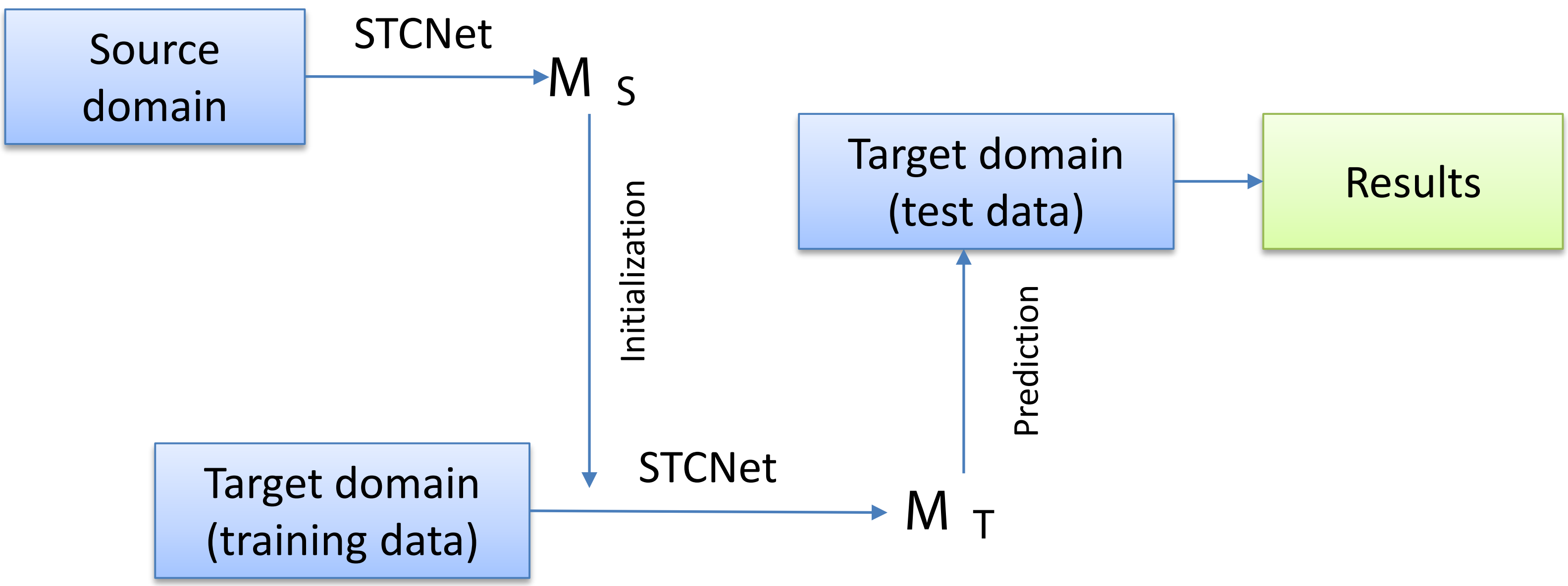
Perform K-Means on X and get the results



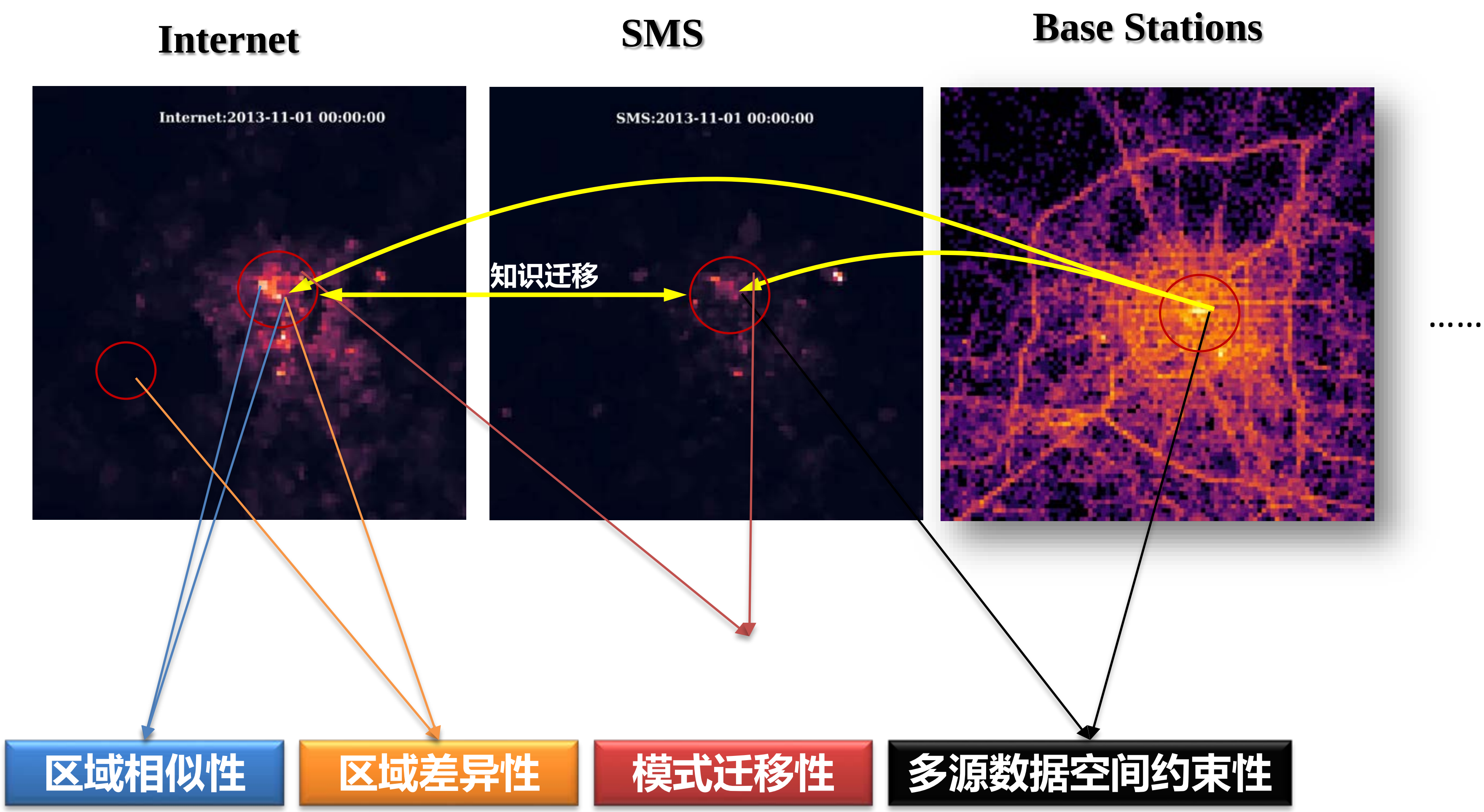
Successive inter-cluster transfer learning strategy

To Solve the Third Challenge

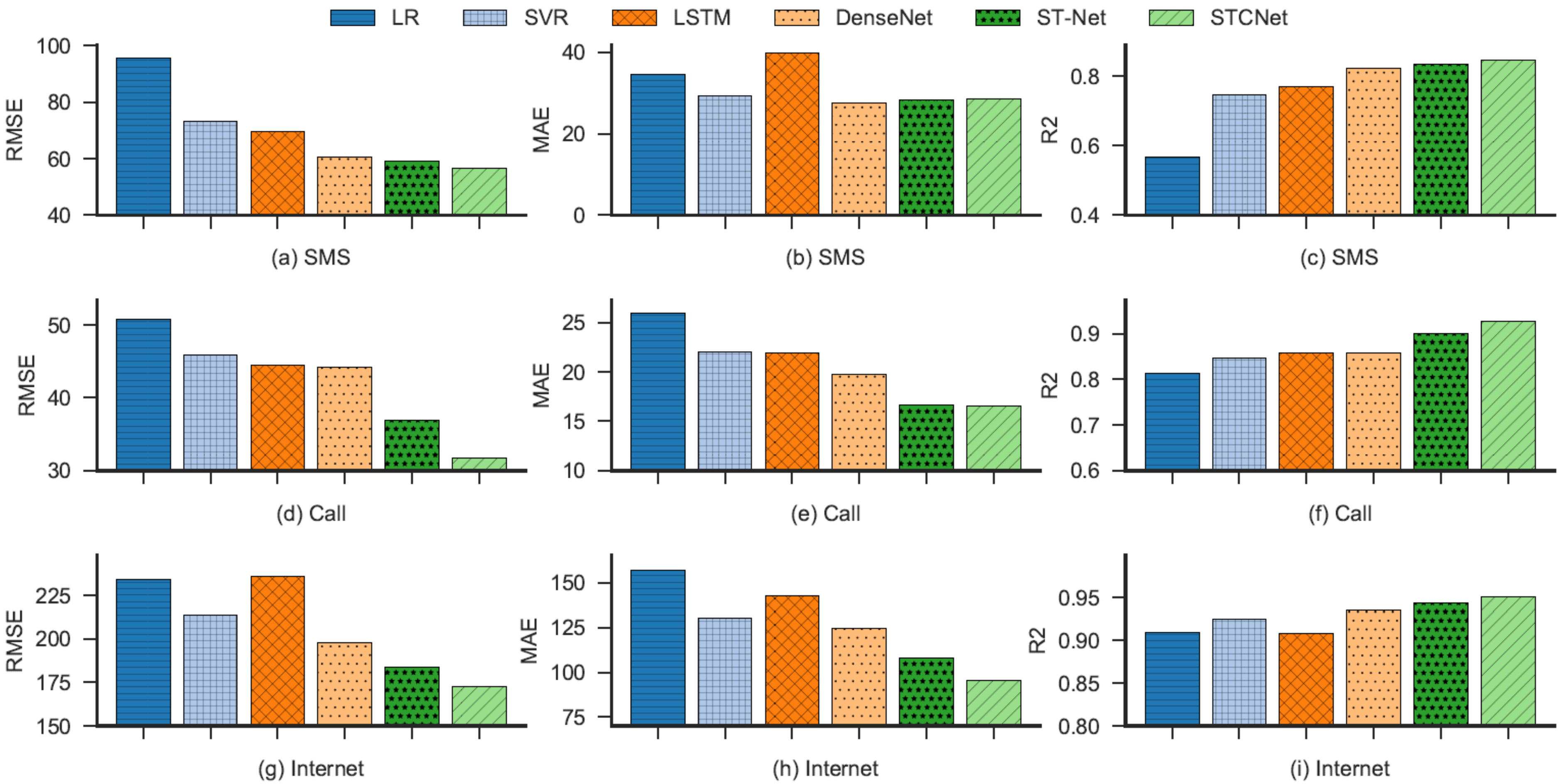
- Train STCNet using the source cellular traffic (SMS) and we get the model M_{SMS}, then use this model (parameters) as initializations and continue training STCNet using the target cellular traffic (CALL) and get the model M_{CALL}, we use the second model to carry out prediction on CALL dataset.



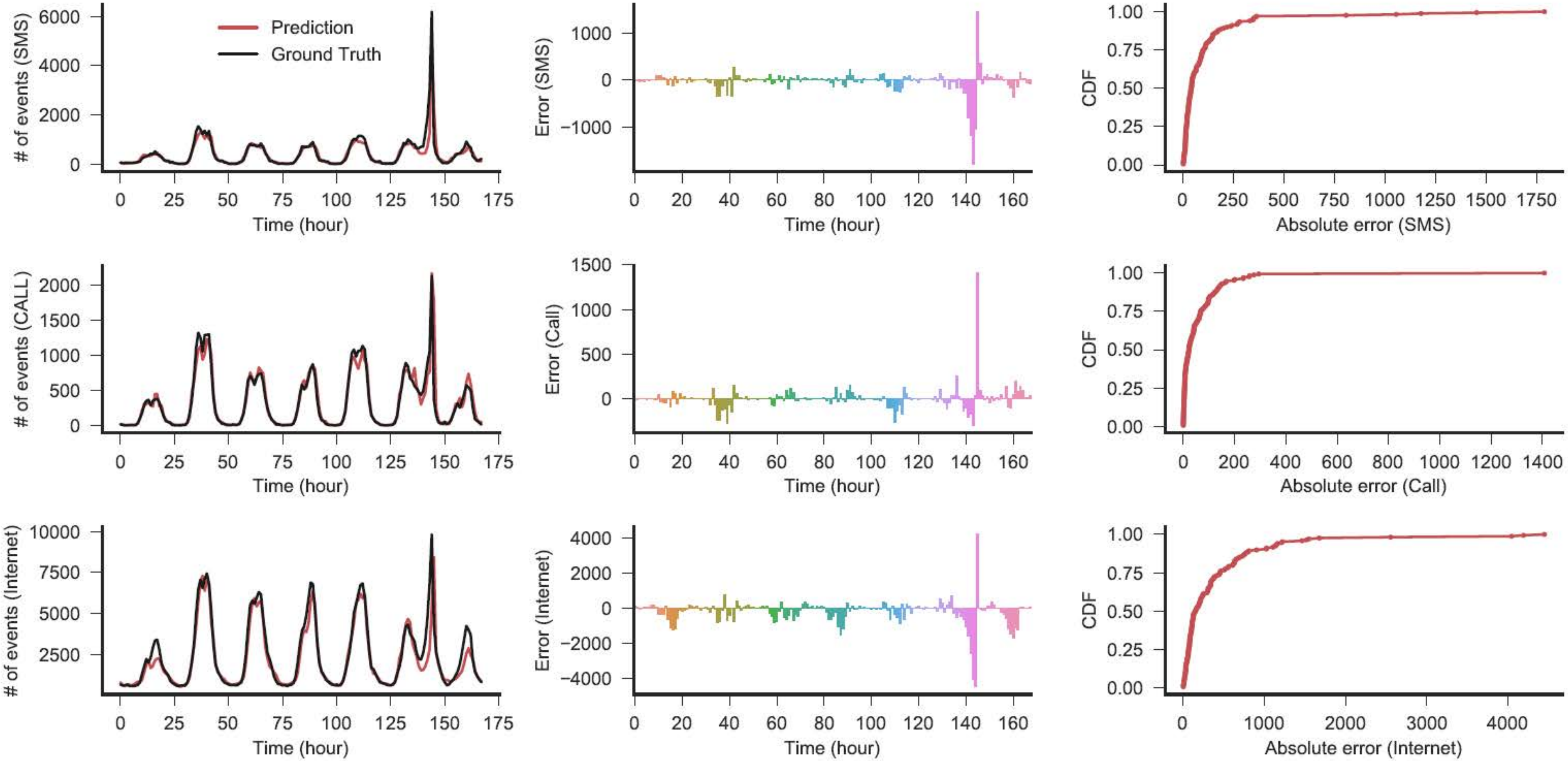
Overall Idea



Prediction Results



Prediction Results



Prediction Results

Dataset	Transfer or Not	RMSE	MAE	R2
SMS	No Transferring	55.0727	28.3204	0.8593
	Transferring with Call	50.9684	25.9039	0.8714
	Transferring with Internet	52.7757	25.4138	0.8593
Call	No Transferring	35.4332	16.8691	0.9163
	Transferring with SMS	33.4663	15.7211	0.9240
	Transferring with Internet	30.8529	14.4174	0.9312
Internet	No Transferring	186.1173	111.7783	0.9411
	Transferring with SMS	168.8695	97.8216	0.9511
	Transferring with Call	169.5268	94.3403	0.9503

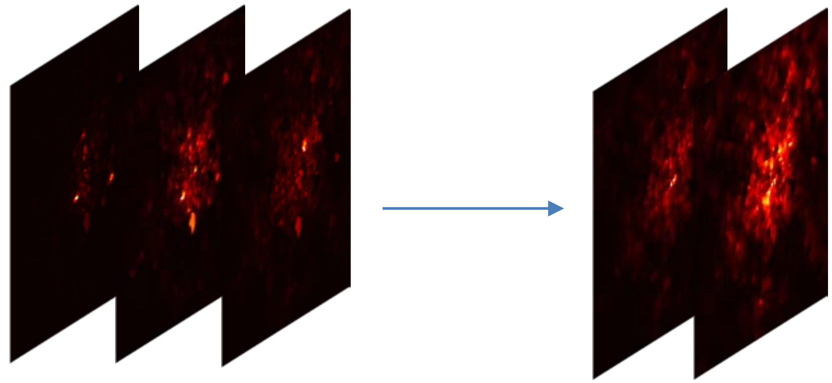
Contents

Cellular Traffic Prediction Beyond MSE

Problem revisit of cellular traffic prediction

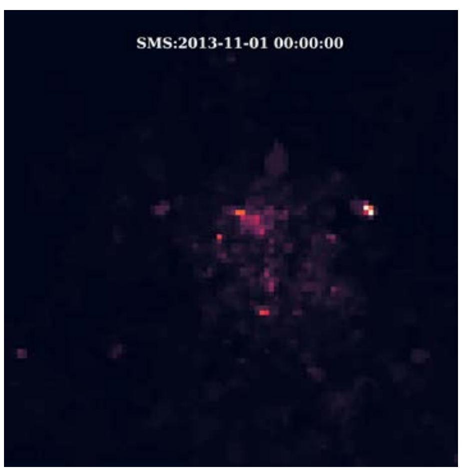
$$\text{Input} = X = \{X_1, \cdots, X_n\}$$

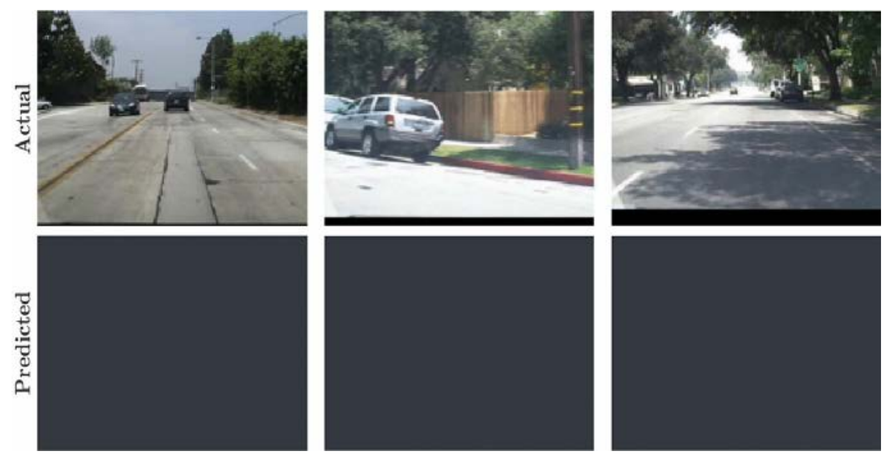
$$\text{Output} = Y = \{Y_1, \cdots, Y_m\}$$



X Y

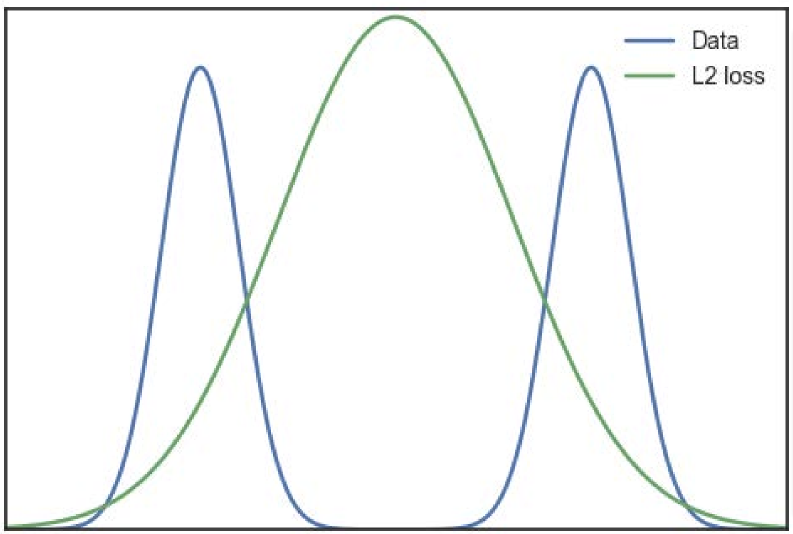
It's video-like data: Multiple frames with single channel image





Cellular Traffic Prediction Beyond MSE

- Most of the work on cellular traffic prediction select ℓ_p norm as their loss function
- But ℓ_p loss produces ***blurry predictions***, increasingly worse when predicting further in the future
  - If the probability distribution for an output pixel/cell has two equally likely modes v_1 and v_2, the value $v_{avg} = (v_1 + v_2)/2$ minimizes the ℓ_2 loss over the data



As the loss have to minimize the distance of reconstruction to both sample types, the model ***trying to satisfy everybody here***. Intuitively, it is because ***the middle ground between both modes is where the distance is minimized*** to both of them.

- Most of the work on cellular traffic prediction select ℓ_p norm as their loss function
- But ℓ_p loss produces ***blurry predictions***, increasingly worse when predicting further in the future
  - If the probability distribution for an output pixel/cell has two equally likely modes v_1 and v_2, the value $v_{avg} = (v_1 + v_2)/2$ minimizes the ℓ_2 loss over the data

How to solve this problem?

Using adversarial loss!

GANs

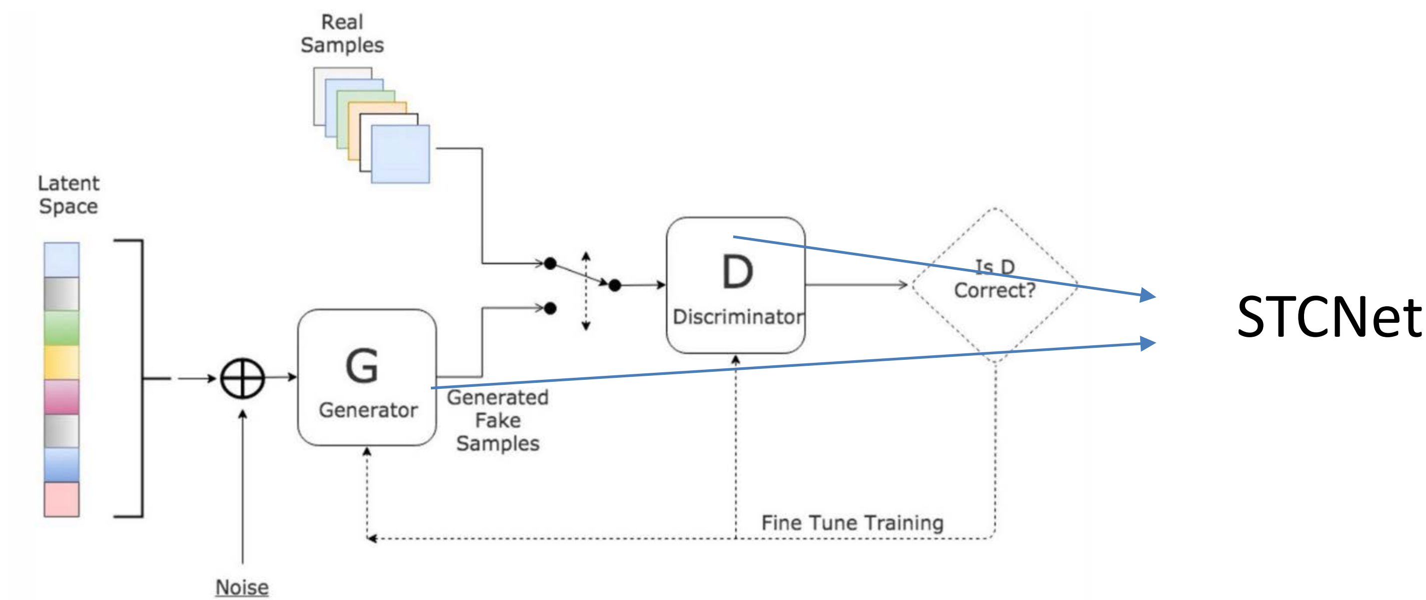
$Y = \{Y_1, \cdots, Y_n\}$

$X = \{X_1, \cdots, X_n\}$ $\xrightarrow{\ell_2}$ (X, Y_{avg}) Not real

$Y' = \{Y_1', \cdots, Y_n'\}$

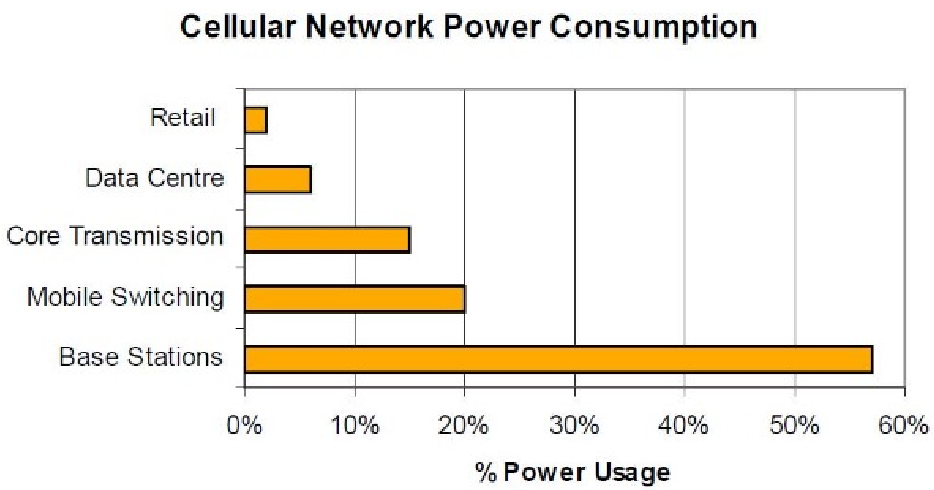
- G can always generate samples that “confuse” D, without being close to Y. In turn, D will learn to discriminate these samples, leading G to generate other “confusing” samples, and so on
- In practice, the combined loss is used to generate predictions closing to Y

$$\mathcal{L}(X;Y) = \lambda_{\ell_p}\mathcal{L}_p(X;Y) + \lambda_{adv}\mathcal{L}^{G}_{adv}(X;Y)$$



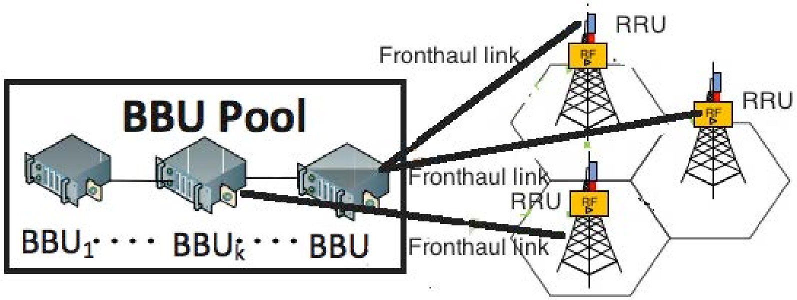
Traffic Prediction for Power Saving

☐ Power saving

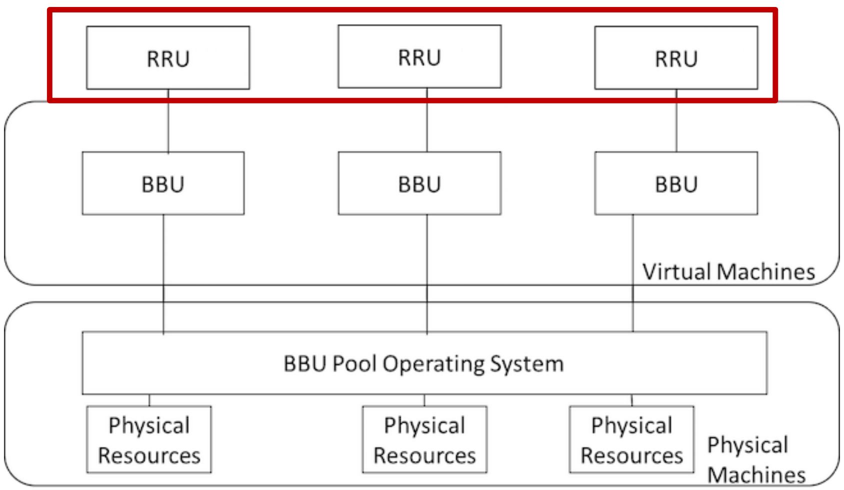


- For the operator, ***57% of electricity use is in radio access***
- Operating electricity is the dominant energy requirement at base stations

Carriers → RRU → Sector → BS



BS (C-RAN) architecture components



Virtualized BBU pool in C-RAN

Han C, Harrold T, Armour S, et al. Green radio: radio techniques to enable energy-efficient wireless networks[J]. ***IEEE communications magazine***, 2011, 49(6).

Traffic Prediction for Power Savings

- Power model of BS

$$P = P_{tx} + P_{misc} = P_{®} ¢L + P_{-} + P_{misc}$$

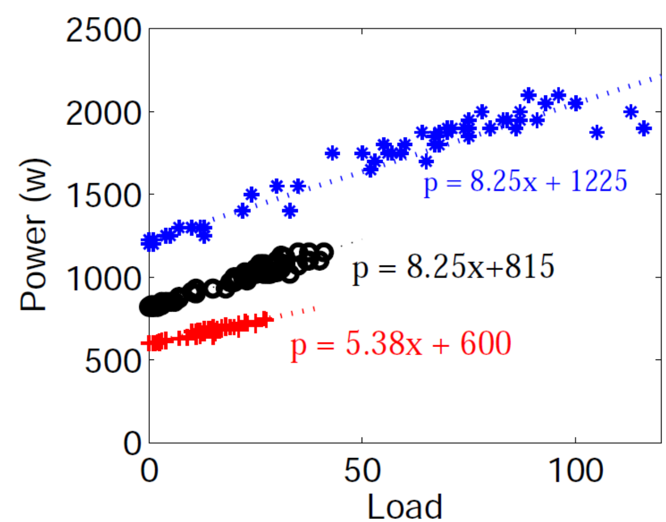
Traffic load level

Power used to provide network access to mobile clients: ***changes with carrier load***

Coefficients of linear model

RRU	**BBU**
Feeder	**RNC transmission**

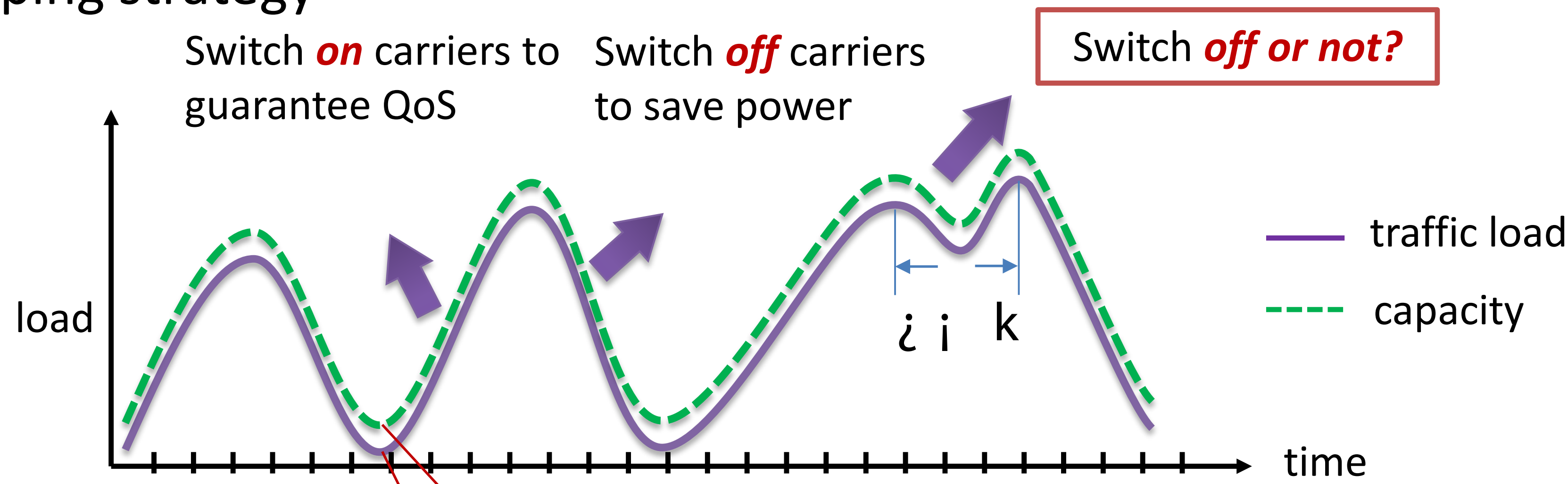
Auxiliary power for cooling, power supply and monitoring: ***remains constant given a fixed operating environment***



Peng C, Lee S B, Lu S, et al. Traffic-driven power saving in operational 3G cellular networks[C]//Proceedings of the 17th annual international conference on Mobile computing and networking. ACM, 2011: 121-132.

Traffic Prediction for Power Savings

Sleeping strategy



Considerations

$$Q(T) = \sum_{k=1}^{N} \frac{\max(L_k - C_k, 0)}{L_k}$$

QoS: reflected by the data missing rate

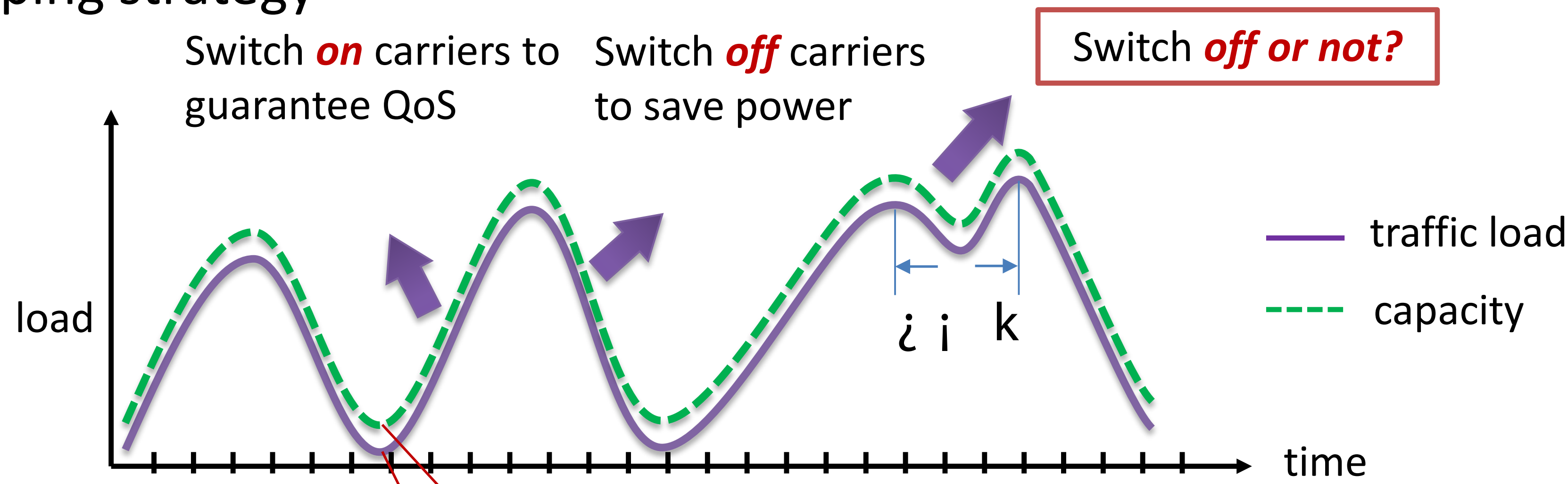
$$W(T) = F\left(\sum_{k=1}^{N} \left\lceil \frac{|C_k - C_{k-1}|}{\Delta C} \right\rceil\right)$$

Operation cost: reflected from the frequency of operations, F(n) symbols for the accumulation function of device damage after n operations

$$E(T) = \sum_{k=1}^{N} P(\min(C_k, L_k))$$

Energy consumption: the total consumed power

- Sleeping strategy



Considerations

$$Q(T) = \sum_{k=1}^{N} \frac{\max(L_k - C_k, 0)}{L_k}$$

$$W(T) = F\left(\sum_{k=1}^{N} \left\lceil \left| \frac{C_k - C_{k-1}}{\Delta C} \right| \right\rceil\right)$$

$$E(T) = \sum_{k=1}^{N} P(\min(C_k, L_k))$$

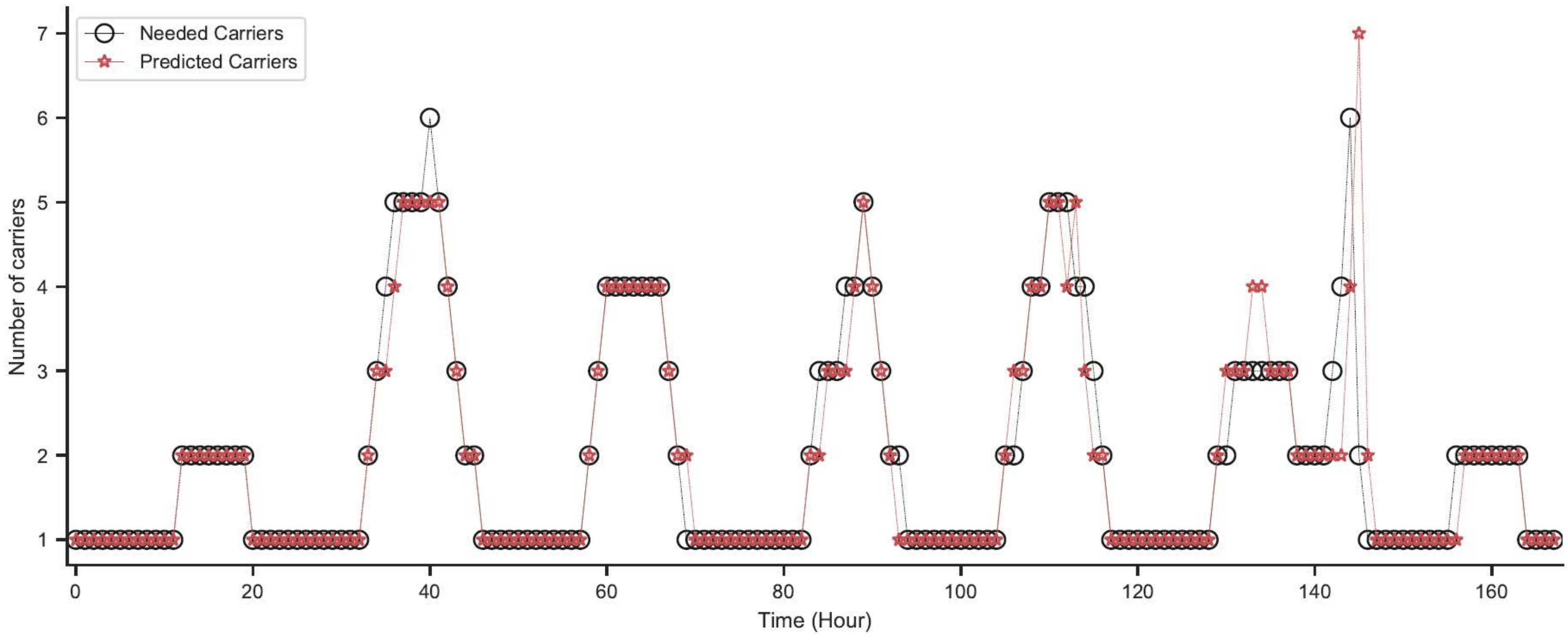
$$\arg\min \sum_{k=1}^{N} P(\min(C_k, M_k)) + F\left(\sum_{k=1}^{N} \left\lceil \left| \frac{C_k - C_{k-1}}{\Delta C} \right| \right\rceil\right)$$

$$\text{s.t.:} \sum_{k=1}^{N} \max(M_k - C_k, 0) = 0;$$

$$C_k = n \cdot \Delta C, n \in \mathbb{N}.$$

Traffic Prediction for Power Savings

- No operation cost situation
  - Each cell has 3 sectors, each sector has 2 carriers
  - Traffic load is divided into 6 levels



谢谢！

Chuanting Zhang

Code: https://github.com/zctzzy

Distributed Base Station DBS3900

Parameters	Specifications
Working Frequency Bands	• 876 to 880 MHz and 921 to 925 MHz • 880 to 915 MHz and 925 to 960 MHz • 1,710 to 1,785 MHz and 1,805 to 1,880 MHz
Capacity	• One BBU supports six RRUs. • Each RRU supports a maximum of six levels of cascading. • Each RRU supports two carriers.
Networking	A maximum of 12 subsites are allowed to serve one cell. Each subsite supports three RRUs.
Transmit Power	• 918 MHz to 925 MHz: 2 x 60W • 925 MHz to 960 MHz: 2 x 80W
Receiver Sensitivity	• 918 MHz to 925 MHz: Single antenna: -112.5 dBm, Double antennas: -115.5 dBm • 925 MHz to 960 MHz: Single antenna: -113.4 dBm, Double antennas: -116.4 dBm
BBU3900 Size (H x W x D)	86 mm x 442 mm x 310 mm
RRU3004 Size (H x W x D)	480 mm x 356 mm x 100 mm
RRU3004 with housing Size (H x W x D)	485 mm x 380 mm x 130 mm

一个BBU支持6个RRUs

一个RRU支持2个载波数

一个cell最多允许12个子站点，每个子站支持3个RRUs