Convolutional neural networks with hierarchical context transfer for high-resolution spatiotemporal predictions

Ksenia D. Mukhina, Alexander A. Visheratin, Denis Nasonov

BIGSPATIAL 2020, 3rd November 2020

acm SIGSPATIAL

Open Lab

ITMO UNIVERSITY

Motivation

Spatiotemporal prediction is a problem where the goal is to use previous and current states of the spatial area to generate a precise next state.

A major challenge is a strong need for precise predictions on large and detailed areas.

Modern solutions require powerful workstations.

Pedro Lastra, Unsplash

Hierarchical area split

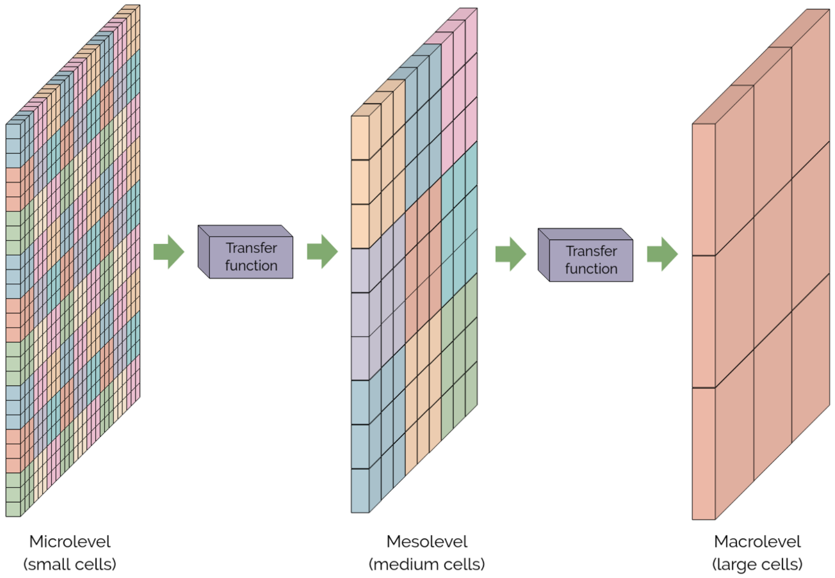
It was designed to significantly reduce the number of trainable parameters in the target model.

$$a^{l}_{\{m,n\}} = F\left(S^{l-1}_{m,n}\right), \{m \in [0 \ldots M^{L-l}], n \in [0 \ldots N^{L-l}]\}$$

where $F\left(S^{l-1}_{m,n}\right)$ is a transfer function that is applied to the subset S of the original matrix.

$$S^{l-1}_{m,n} = \left(a^{l-1}_{i,j}\right)^{(m+1)\cdot M,(n+1)\cdot N}_{i=m\cdot M, j=n\cdot N}$$

In this work, we operate terms macro-, meso-, and micro-level assuming there are three granularity levels.



Hierarchical area split - 2

0	0	1	0	0	0	0	0	0
1	0	5	10	1	0	1	0	0
0	2	0	0	0	0	2	2	0
0	0	1	0	0	1	2	0	1
1	0	0	1	3	1	1	1	1
0	1	1	0	2	4	1	0	0
0	0	1	0	0	23	0	0	1
0	0	0	1	0	0	0	0	0
0	0	0	0	9	0	0	0	3

9	11	5
4	15	7
1	33	4

Context transfer loss function

We transfer context between layers via context transfer loss function:

$$CTL = \frac{\alpha \cdot L(Y', Y) + \ \beta \cdot L(A')}{\alpha + \beta}$$

where α and β are weight coefficients, $L(Y', Y)$ is a standard loss function for predicted values and ground truth. $L(A')$ is a part responsible for actual context transfer:

$$L(A') = \sum_{l=1}^{L-1} \sum_{m=0}^{M^{L-l}} \sum_{n=0}^{N^{L-l}} \left(\left(a_{m,n}^{l}\right)' - \ F\left(\left(S_{m,n}^{l-1}\right)'\right) \right)$$

where $\left(a_{m,n}^{l}\right)'$ is a prediction of the value in position m, n at the level $l,$ and $F\left(\left(S_{m,n}^{l-1}\right)'\right)$ is a transform function over subset of predicted values at the level $l - 1$.

Context transfer loss function - 2

Meso-level prediction

0	0	1	0	0	0	0	0	0
1	0	15	1	1	0	0	5	0
0	2	0	0	0	0	1	1	0
0	0	1	0	0	1	2	0	0
0	1	1	1	0	1	1	0	0
0	0	0	0	2	2	1	0	0
0	0	1	0	0	3	0	0	0
0	1	0	1	0	0	0	7	0
1	1	1	0	9	0	0	0	0

Aggregated data

19	2	7
3	7	4
5	13	7

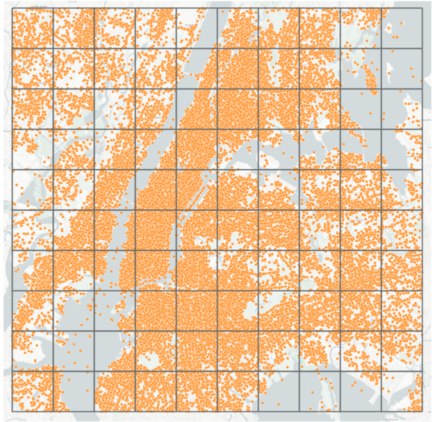
Ground truth

9	11	5
4	15	7
1	33	4

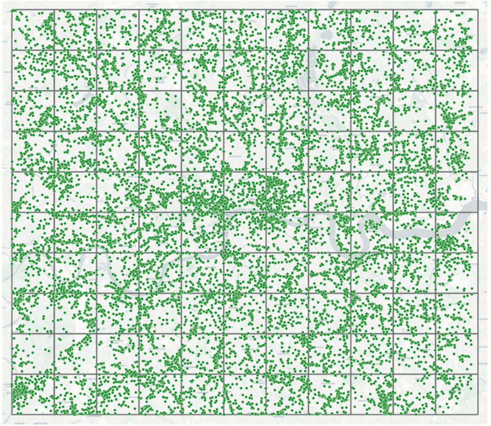
Loss

10	9	2
1	8	3
4	20	3

New York

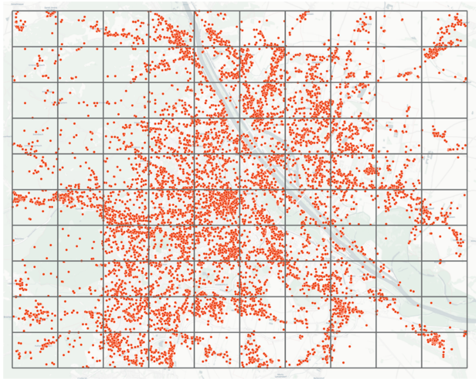


London

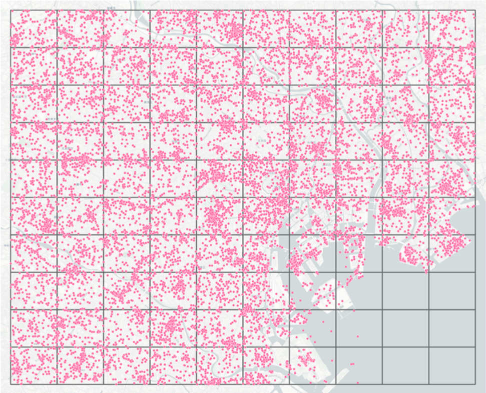


Moscow

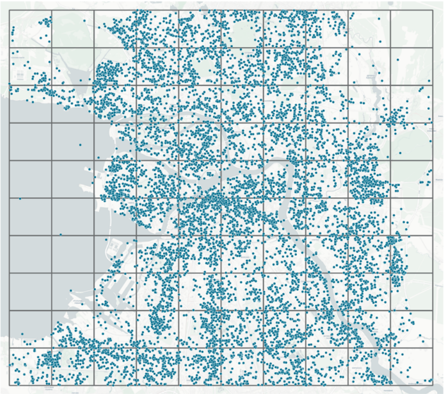
Vienna



Tokyo

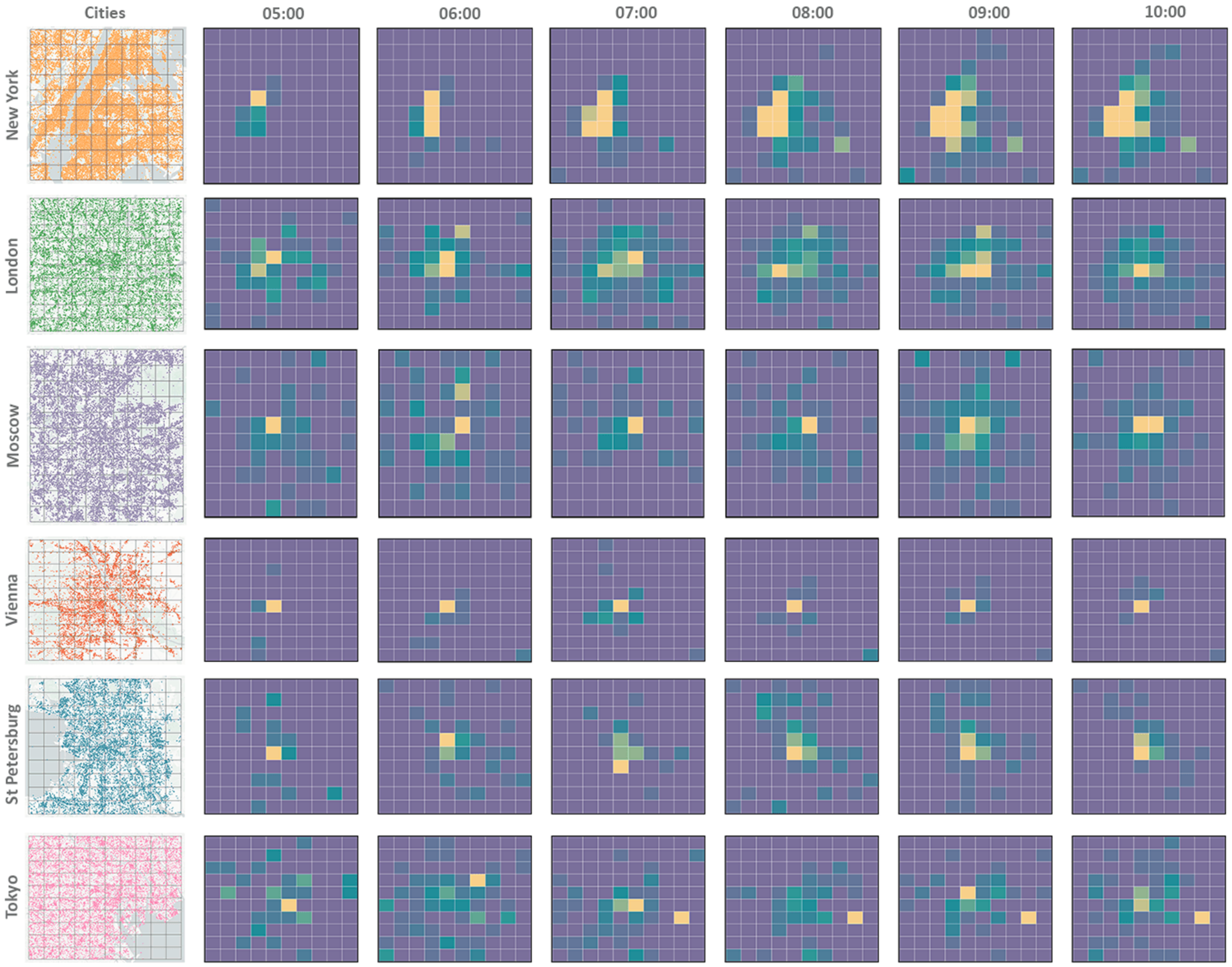


St Petersburg



We used public data from Instagram.

City coverage by LBSN depends on Internet access, popularity of particular service, landscape, etc.



The main challenges of prediction task:

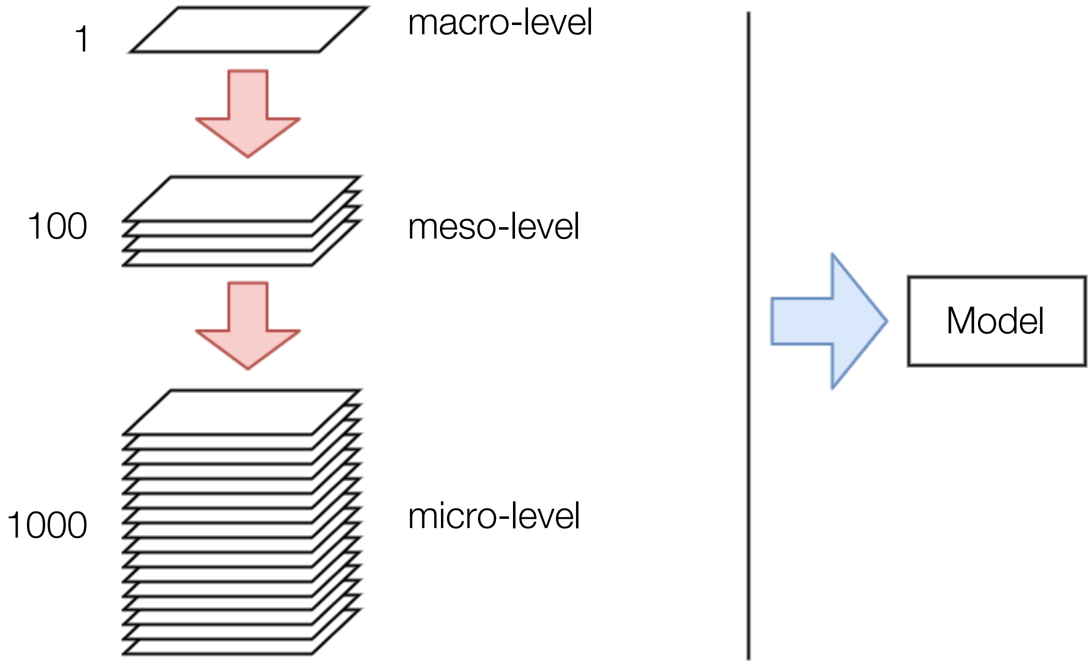
- the absence of a strong influence on sequential hours;
- a lot of zeroes in monitoring area;
- important to predict each cells as precise as possible

Dataset

The data for every hour was aggregated and placed on a geospatial grid.

We split area into three layers using 10x10 grid.

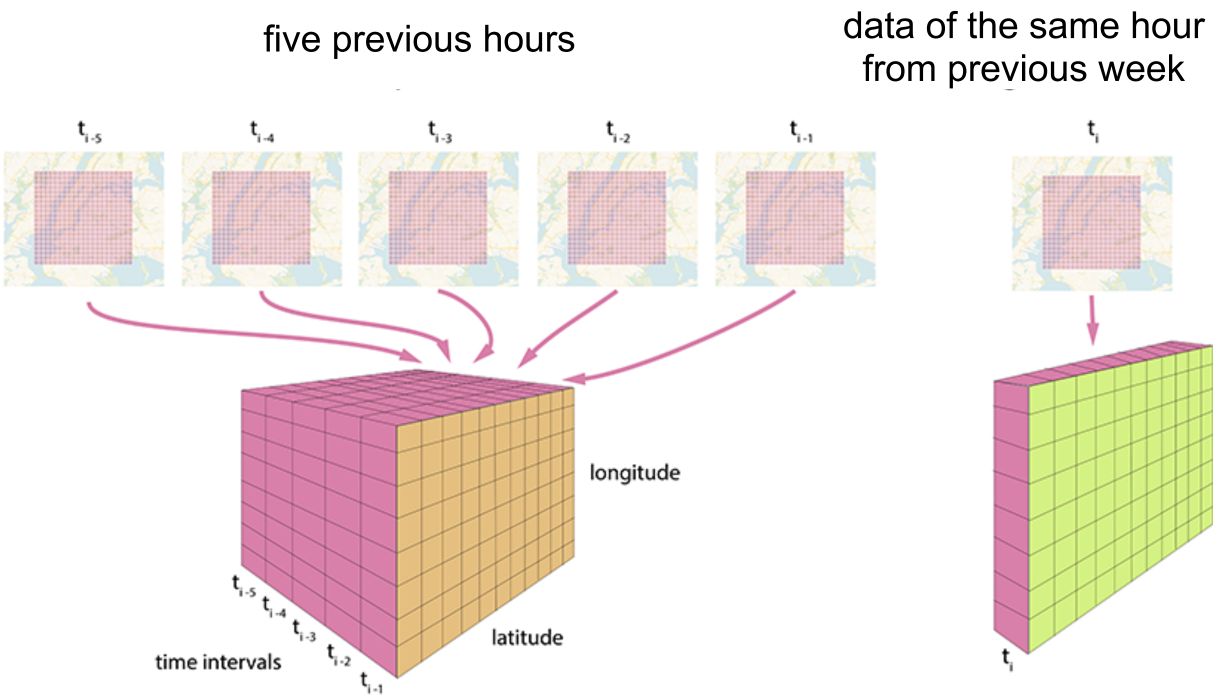
Thus for every hour we obtained 1101 examples.



Dataset - 2

The value of each cell corresponds to the number of Instagram posts.

Input data is formed by using data of five previous hours and data of the same hour from previous week.



Model architecture

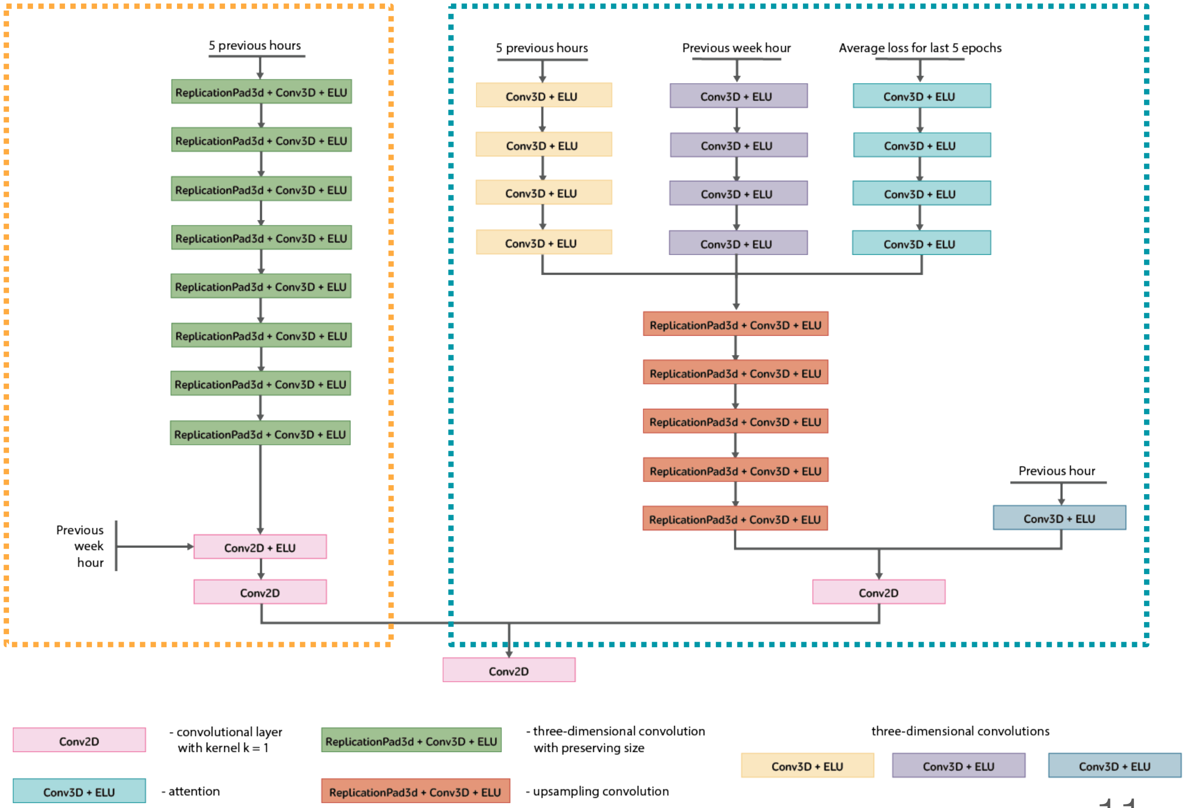
Model consists of convolutional layers:

$$Y_t^k = ELU\left(W^k * Y_t^{\{k-1\}} + b^k\right)$$

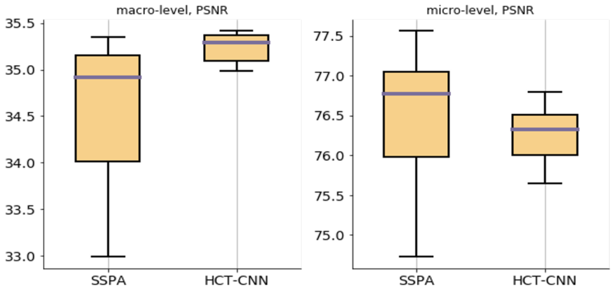
where $*$ denotes convolution operation, t corresponds to timestamp, k is index of convolutional layer, $W^k$is weights and b is bias.

Results of both branches are concatenated:

$$H_t = ELU(W * (Y_t^K \oplus X_t^K) + b)$$



Ablation study



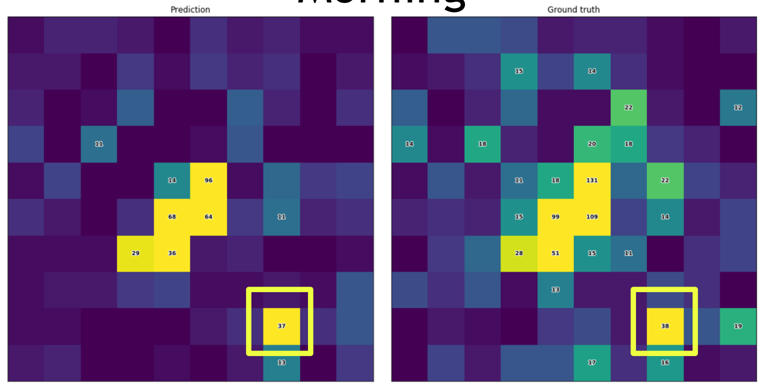
We compare performance of different parts of the model separately to emphasize their influence on the final results.

- S - sequential branch (the first branch of the model);
- SP - the second branch - a sequential-periodic branch;
- SSPA - full model without transfer (sequential and sequential-periodic branches with attention mechanism);
- HCT-CNN - full model with transfer mechanism.

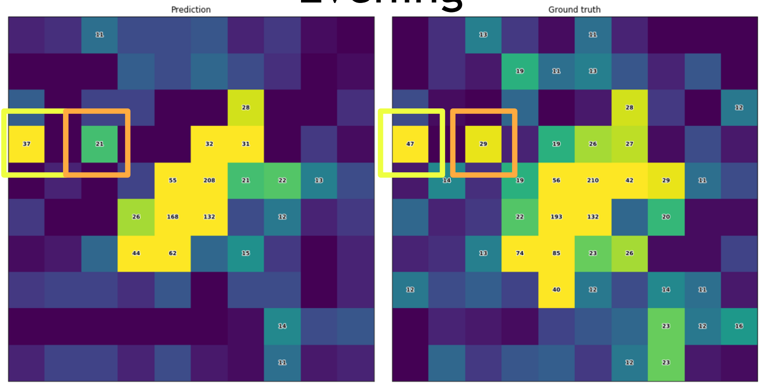
	Macro-level			
	MAE	MSE	PSNR	SSIM
S	5.61	511.6	30.99	0.935
SP	4.16	96.6	34.27	0.981
SSPA	3.97	116.4	34.54	0.982
HCT-CNN	3.61	72.2	35.24	0.984
	Micro-level			
	MAE	MSE	PSNR	SSIM
S	0.00018	0.027	65.84	0.999
SP	0.00015	0.021	65.98	0.999
SSPA	0.00005	0.002	75.98	0.999
HCT-CNN	0.00005	0.002	76.07	0.999

New York: macro-level

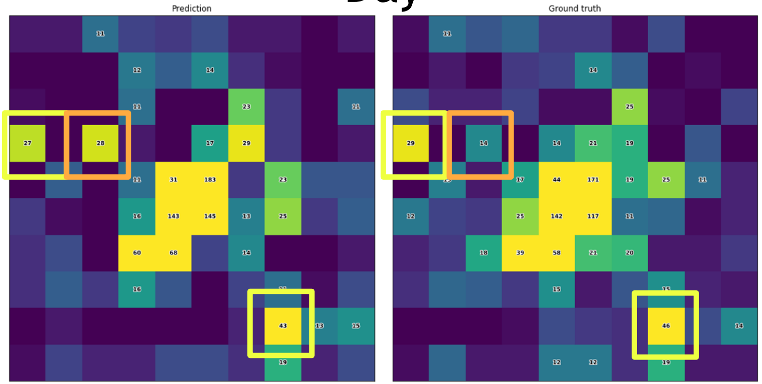
Morning



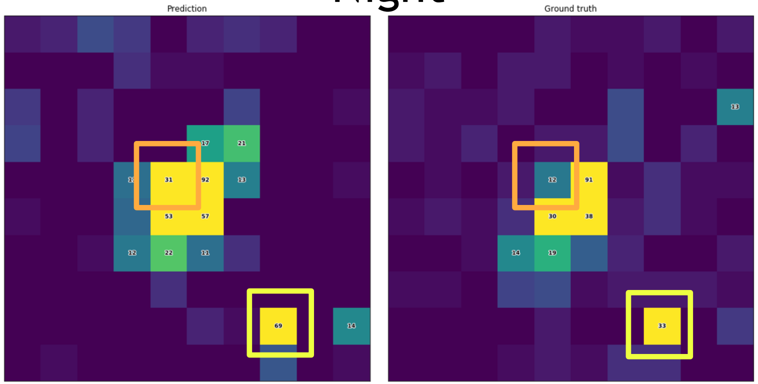
Evening



Day

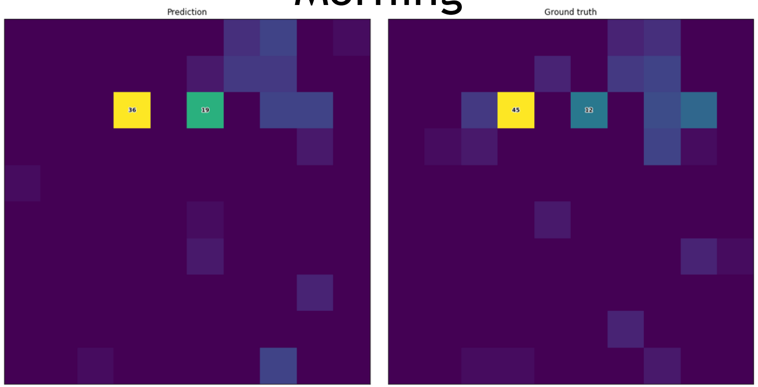


Night



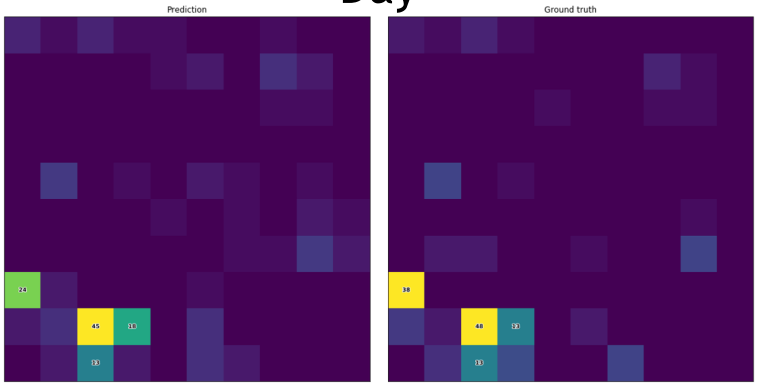
New York: meso-level

Morning

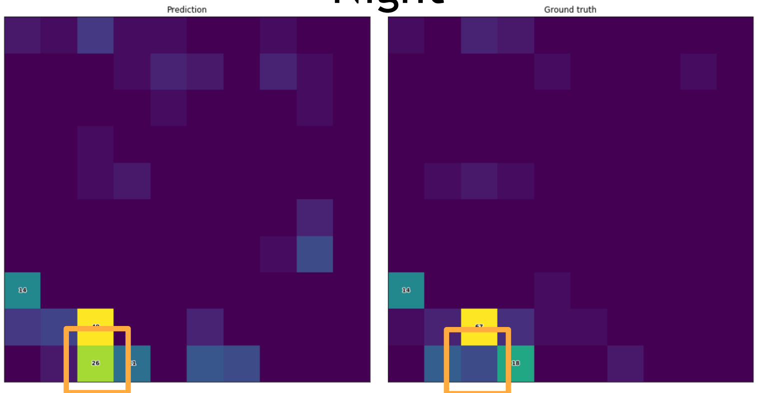


Evening

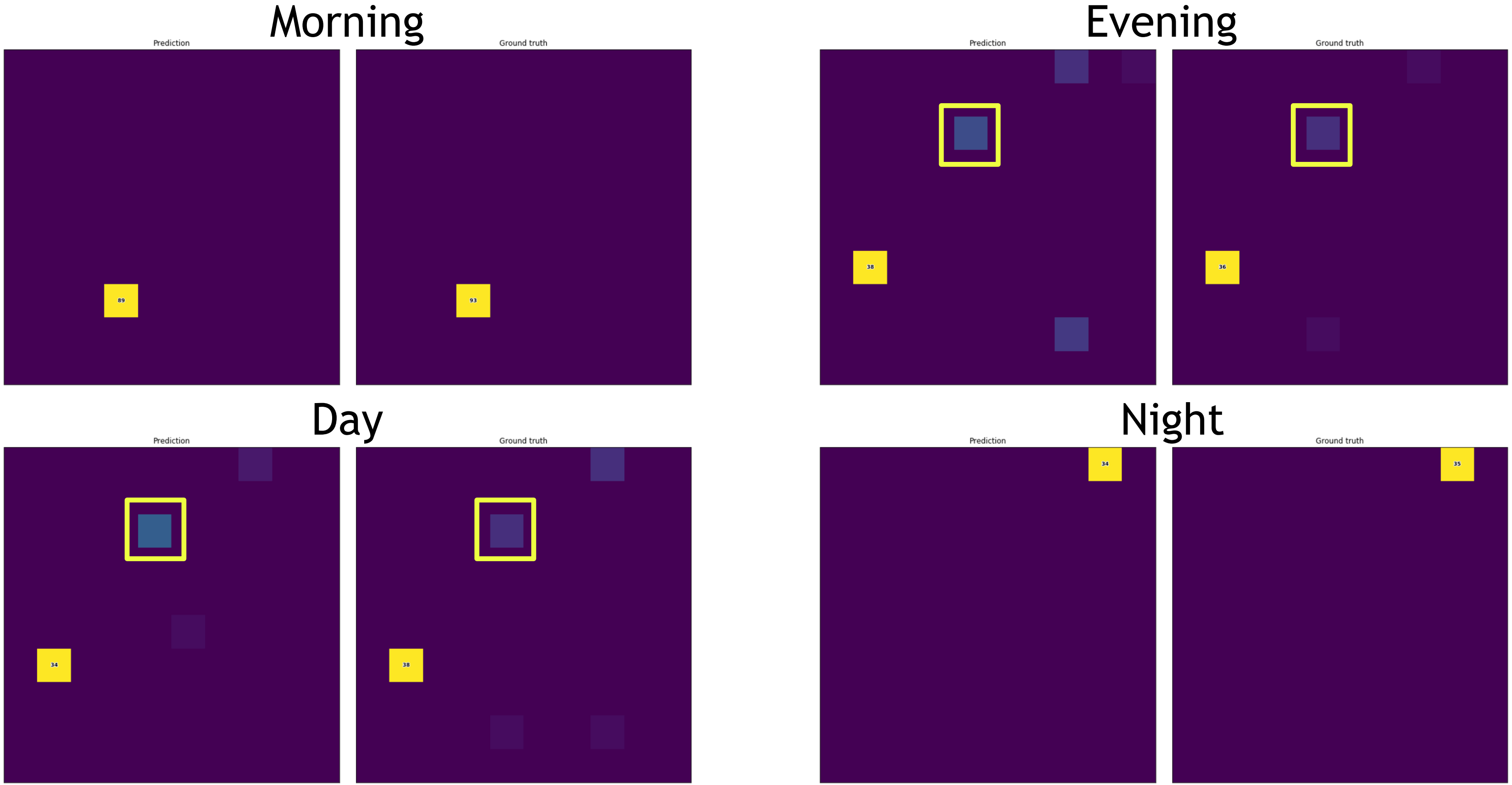
Day



Night



New York: micro-level



Macro-level comparison

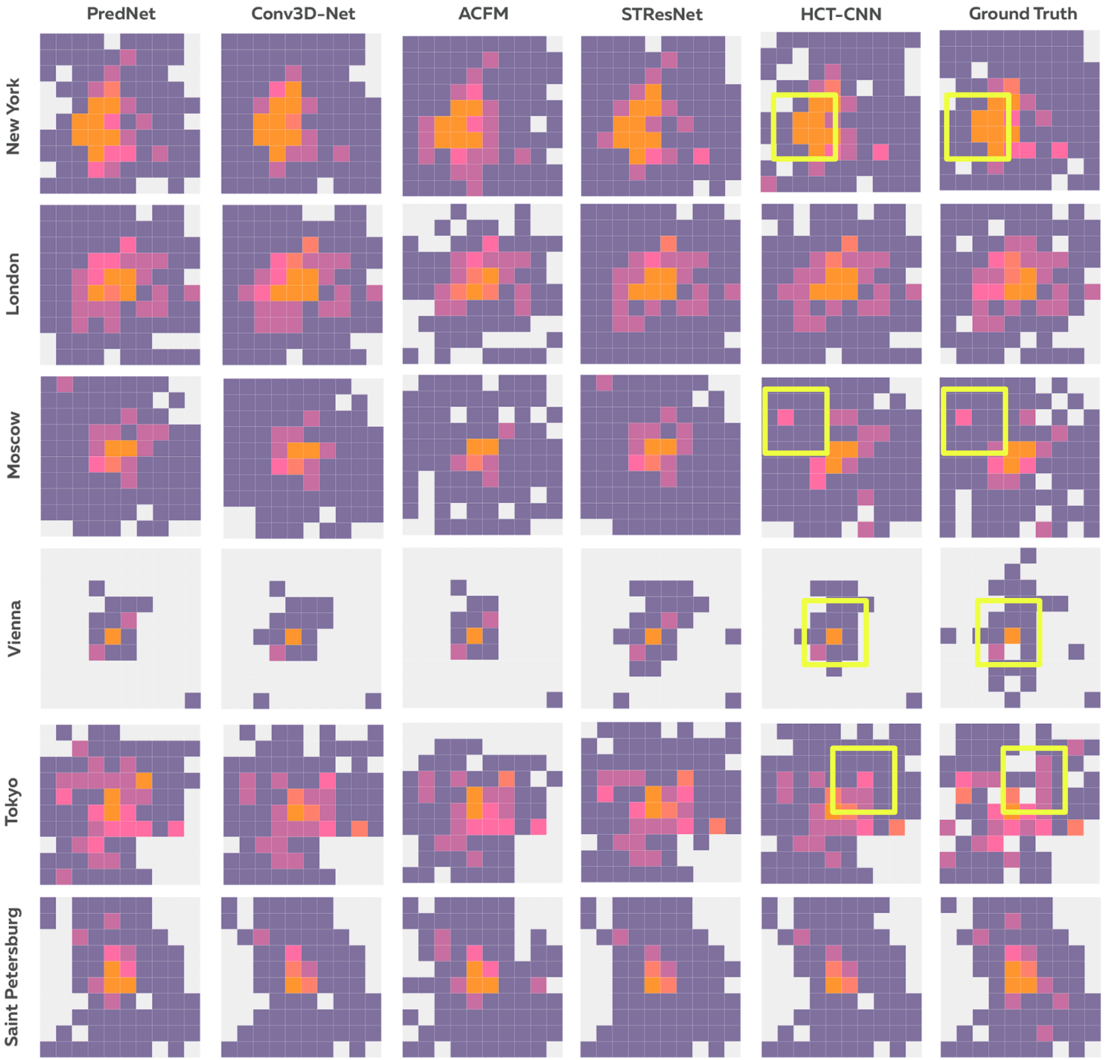
The most difficulties all model have with the most active city - New York, MAE and MSE value are the highest for the all models.

City	New York				London				Moscow			
	MAE	MSE	PSNR	SSIM	MAE	MSE	PSNR	SSIM	MAE	MSE	PSNR	SSIM
Zeros	22.36	5746.9	16.97	0	7.61	273.5	13.82	0	9.65	698.1	15.33	0
PredNet	6.72	574.4	27.44	0.895	3.23	50.6	20.38	0.748	4.36	171.8	22.63	0.771
ACFM	5.37	164.2	31.47	0.929	3.06	38.7	21.16	0.754	3.98	73.8	24.35	0.832
Conv3D	4.68	192.9	32.15	0.965	2.69	39.4	21.62	0.817	2.98	69.9	24.37	0.873
STResNet	3.92	74.6	34.88	0.983	2.39	**20.3**	**23.72**	**0.892**	2.88	39.1	27.29	**0.930**
HCT-CNN	**3.61**	**72.2**	**35.24**	**0.984**	**2.35**	21.2	23.63	0.886	**2.76**	**39.6**	**27.33**	0.928

Macro-level comparison - 2

HCT-CNN for all cities achieves the best or second best results for all metrics among all cities.

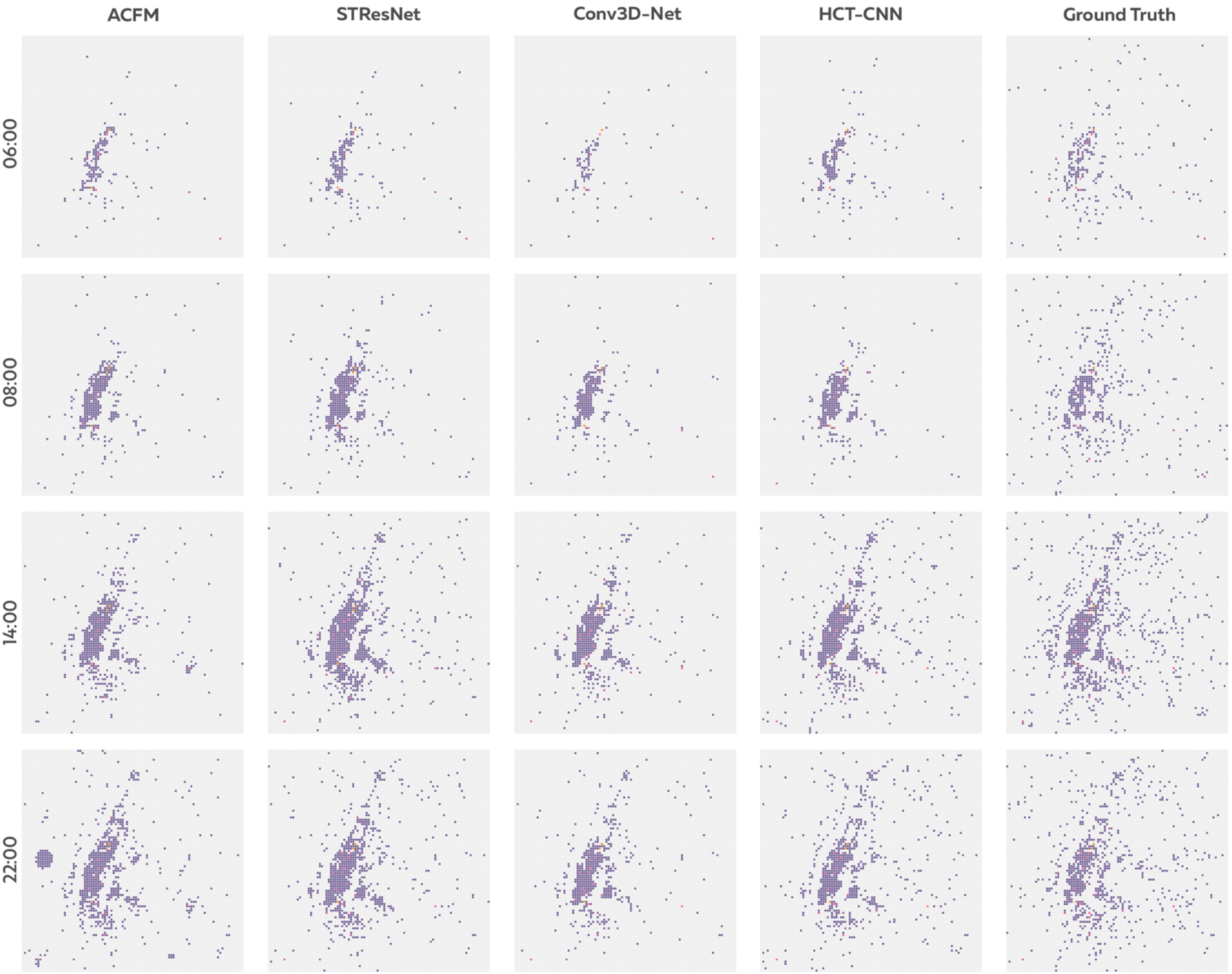
City	Vienna				Tokyo				Saint Petersburg			
	MAE	MSE	PSNR	SSIM	MAE	MSE	PSNR	SSIM	MAE	MSE	PSNR	SSIM
Zeros	1.35	46.7	18.44	0	4.88	119.6	14.04	0	5.99	280.9	15.34	0
PredNet	0.72	10.9	24.90	0.723	2.40	31.3	19.38	0.689	2.51	47.9	22.37	0.765
ACFM	0.67	3.9	27.50	0.826	2.12	17.5	21.56	0.793	2.31	26.5	24.12	0.829
Conv3D	0.51	3.3	25.98	0.811	1.79	15.3	19.98	0.735	1.82	21.2	24.60	0.877
STResNet	0.59	**3.0**	**28.62**	**0.890**	1.83	**12.0**	**22.96**	**0.874**	1.83	**15.4**	**26.38**	**0.916**
HCT-CNN	**0.56**	3.3	28.39	0.874	**1.78**	12.4	22.81	0.857	**1.74**	16.5	26.33	0.914



HCT-CNN is able to correctly capture the high activity areas as well as majority of low activity areas.

Orange color represents the maximum level of activity, grey cells indicate zero posts.

Yellow squares indicate correct predictions of HCT-CNN where other models struggled.



We selected the best models from alternative comparison and trained it again for meso- and micro-levels.

Meso-level is the most challenging to predict.

Memory consumption

	Macro-level			Meso-level			Micro-level		
	MAE	SSIM	Memory, MB	MAE	SSIM	Memory, MB	MAE	SSIM	Memory, MB
ACFM	5.368	0.929	232.24	0.232	0.342	15905.73	-	-	-
Conv3D	4.679	0.965	**0.27**	**0.122**	0.525	**19.12**	0.00198	0.969	1903.62
STResNet	3.916	0.983	13.02	0.147	0.511	2519.89	-	-	-
HCT-CNN	**3.615**	**0.984**	38.04	0.140	**0.539**	38.04	**0.00005**	**0.999**	**38.04**

With increasing of the resolution from macro-level (10x10 grid) to meso-level (100x100 grid) and micro-level (1000x1000 grid) the memory consumption is significantly increasing as well.

The work station was equipped by Intel(R) Core i7-8700 CPU @ 3.20GHz with NVIDIA GeForce RTX 2070.

Conclusion

Main contributions of the papers are the following:

- Convolutional deep learning model integrated with HCT for generating high-resolution predictions.
- Large dataset containing Instagram users activity in six cities for a two year period https://doi.org/10.5281/zenodo.4088833.
- Experimental results that demonstrate that proposed model with HCT outperforms existing solutions on micro-level and shows comparable results on meso- and macro-level with significantly less memory consumption.

Thank you for the attention!

My colleagues

Alexander
Visheratin
Beehive AI

Denis
Nasonov
ITMO
University

Acknowledgements

This research is financially supported by Russian Science Foundation according to the research project #20-11-20264.

K.M. is supported as CUDAN Senior Research Fellow at the ERA Chair for Cultural Analytics, funded through the European Union's Horizon 2020 research and innovation program (Project No. 810961).

Ksenia Mukhina

Tallinn University, Estonia

mukhinaks@gmail.com