

일반논문 (Regular Paper)

방송공학회논문지 제31권 제5호, 2026년 9월 (JBE Vol.31, No.5, September 2026)

<https://doi.org/10.5909/JBE.2026.31.5.1080>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

셀룰러 트래픽 예측을 위한 확장 문맥과 공간적 주의를 활용한 시공간 네트워크

Maryam Qamar^{a)}, 배 성 호^{a)†}

Cellular Traffic Prediction via Spatiotemporal Networks with Dilated Context and Spatial Attention

Maryam Qamar^{a)} and Sung-Ho Bae^{a)†}

요 약

정확한 도시 전역 셀룰러 트래픽 예측은 선제적 자원 관리와 자율 조직형 지능형 셀룰러 네트워크를 실현하는 데 필수적이다. 기존 연구들은 지역적 특징을 효율적으로 추출하지만, 시공간 상호작용에서 발생하는 장거리 전역 문맥을 포착하는 데 어려움을 겪으며, 복잡한 도시 구조 내에서 서로 다른 변화 추세를 보이는 지역들을 효과적으로 우선 순위화하지 못한다. 본 연구에서는 더 넓은 전역 문맥을 포착하기 위한 병렬 다중 스케일 팽창(dilated) 분기와, 중요한 공간적 위치를 적응적으로 강조하는 공간 어텐션 메커니즘을 도입한 개선된 네트워크 구조를 제안한다. 도시 전역 데이터셋(SMS 및 통화 트래픽 벤치마크)을 대상으로 한 실험 평가 결과, 제안된 프레임워크는 기존 모델(baseline)보다 유의미하게 우수한 성능을 보였으며, 평균제곱근오차(RMSE)를 SMS 트래픽에서 12.59, 통화 트래픽에서 6.01까지 감소시켰다.

Abstract

Accurate citywide cellular traffic prediction is essential for proactive resource management and self-organizing intelligent cellular networks. While existing work efficiently extracts local features, it struggles to capture long-range global context of spatiotemporal interactions and fails to prioritize distinct regions of varying trends across a complex city structure. In this work, we present an enhanced architecture that introduces a parallel multi-scale dilated branch to capture broader global context alongside a spatial attention mechanism that adaptively emphasizes critical spatial locations. Experimental evaluations on citywide datasets (SMS and Call traffic benchmarks) show that our proposed framework significantly outperforms baseline, lowering Root Mean Square Error (RMSE) to 12.59 for SMS traffic and 6.01 for Call traffic.

Keyword : Cellular Traffic Prediction, Spatiotemporal Modeling, Dilated Convolutions, Spatial Attention

I . Introduction

With the exponential rise in mobile data services and the rapid evolution of next-generation communication infrastructures, modern cellular traffic exhibits significant fluctuations and highly complex dynamics^[1]. Providing accurate spatiotemporal traffic forecasting across city grids is foundational for intelligent cellular network operations, allowing service providers to execute dynamic resource allocation, minimize infrastructure energy footprints, and prevent cellular congestion beforehand^{[2][3]}.

Cellular traffic modeling naturally possesses dual dimensional dependencies, spatial dependencies, where data traffic correlations depend on functional geographical zones (e.g., residential areas, business centers, or universities), and temporal dependencies, characterized by continuous short-term fluctuations, cyclical daily patterns, and broader weekly trends^[4]. Recent efforts treat citywide traffic grids as grid-based spatial maps (similar to pixel structures in natural images) and employ deep learning methods to model these dynamics^[4]. Among these, densely connected convolutional networks provide powerful inductive biases by recycling feature maps through dense connections, reducing training difficulties while capturing fine-grained structural data.

However, traditional convolutional networks possess specific architectural limitations when applied to citywide scales. Standard convolutional kernels focus purely on lo-

calized receptive fields^[5]. This locality limits the network's capacity to process global interactions across distant areas of the city grid simultaneously. Furthermore, conventional convolutions treat all grid regions uniformly, ignoring the fact that distinct hot-spots dictate urban traffic behavior and require higher emphasis.

To resolve these limitations, this work proposes an enhanced architectural design incorporating a parallel multi-scale dilated convolution branch to expand the receptive field and capture global contextual information. Additionally, it integrates a spatial attention module that generates dynamic attention maps to adaptively highlight complex, high-traffic regions across the urban environment.

We evaluate the proposed method against baseline configurations using a real-world traffic dataset. The empirical results confirm that combining dilated convolutions with spatial attention achieves the lowest prediction errors.

II . Background

The dataset^[6] in this study, sourced from Telecom Italia's "Big Data Challenge," contains aggregated SMS and call traffic time series recorded at 10-minute intervals in Milan from November 1, 2013, to January 1, 2014. The city is spatially structured into a 100 x 100 grid of cells. Following the baseline work^[4], the data is formatted such that at any 10-minute time slot t , the incoming and outgoing traffic across the entire grid is represented as a two-channel, image-like matrix processed via a sliding window scheme to generate training and test samples. To predict the traffic at target time slot, the historical data is split into two temporal fragments: closeness spanning immediately preceding c intervals, and daily periodic dependence capturing intervals from the same time slots in previous days. The matrices within each fragment are then concatenated along their first axis to form the final input

a) 경희대학교 컴퓨터공학부(School of Computing, Kyung Hee University)

‡ Corresponding Author : 배성호(Sung-Ho Bae)

E-mail: shbae@khu.ac.kr

Tel: +82-31-201-2593

ORCID: <https://orcid.org/0000-0003-2677-3186>

※ This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2026-RS-2023-00259004) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

· Manuscript received August 3, 2026; Revised August 7, 2026; Accepted August 7, 2026.

tensors for the model.

III. Related Works

Citywide cellular traffic prediction has evolved from classical statistical models toward advanced deep representation models. Traditional methods^{[7][8][9]} struggle to manage the highly nonlinear, high-dimensional spatiotemporal distributions found in modern big data telecommunication grids^[10].

To overcome this, research moved towards neural networks^[11], succeeded by employment of LSTM networks that significantly outperformed traditional statistical models^[12]. To improve accuracy further, Nie et al.^[13] integrated a discrete wavelet transform to capture low-frequency data components. Addressing spatial-temporal dynamics, a hybrid LSTM-autoencoder scheme was proposed to model geographic relationships among base stations^[14]. However, autoencoder-based approaches suffer from lossy compression, which can fail to fully encapsulate intricate spatial relationships among traffic zones^[15]. Furthermore, because most of these methods forecast traffic on an individual cell-by-cell basis, they become computationally prohibitive when scaled to entire city networks in real-world scenarios.

To overcome these barriers, modern architectures treat spatial urban sectors as multi-channel image frames where deep neural network layers extract feature hierarchies. STDenseNet^[4] optimizes feature extraction by connecting each layer to every other layer in a feed-forward design. This design encourages feature reuse and mitigates vanishing gradient issues. For spatiotemporal fusion, inputs are typically categorized into unique time streams to represent varying temporal rhythms, that is, closeness synonymous to the immediate preceding history and periodicity representing daily patterns.

However, while deep convolutional representations yield significant gains, they struggle to capture wide-area corre-

lations and spatially distinct regions. Attention based spatiotemporal models, for instance att-MCSTCNet^[1], ADNet^[17] etc., have been introduced which model long range dependencies and apply attention to weigh the relative importance. However, attention operates temporally rather than over spatial locations, and the accuracy gains sometimes rely in part on the availability of auxiliary cross-domain data (base station, point-of-interest, and social-activity statistics), which is not always practical. GLSTTN^[19] uses stacked spatial-temporal transformer blocks to capture global correlations while CNN is used to extract local spatiotemporal dependencies.

A separate line of work formulates it as a graph-based problem, Wang et al.^[16] proposed AMF-STGCN, which represents base stations as nodes and jointly models spatial and temporal dependencies through graph convolution, combined with graph attention that adaptively captures the differing receptive fields needed by heterogeneous nodes. Despite good performance, it requires an explicit graph construction step relying on metadata like inter-station distance that may not be available. In addition to design overhead and can be sensitive to how the graph is defined.

Our method doesn't rely on any metadata, utilizes a simple dilation based design for long-range dependencies and attention over spatial locations account for regional heterogeneity.

IV. Methodology

To overcome the limitations of localized feature extraction and uniform spatial processing inherent in traditional convolutional networks, we propose an enhanced spatiotemporal framework. The rationale behind this architecture is to bridge the gap between local variations and urban dynamics.

Instead of relying solely on standard localized convolutions, our proposed method runs a parallel global path

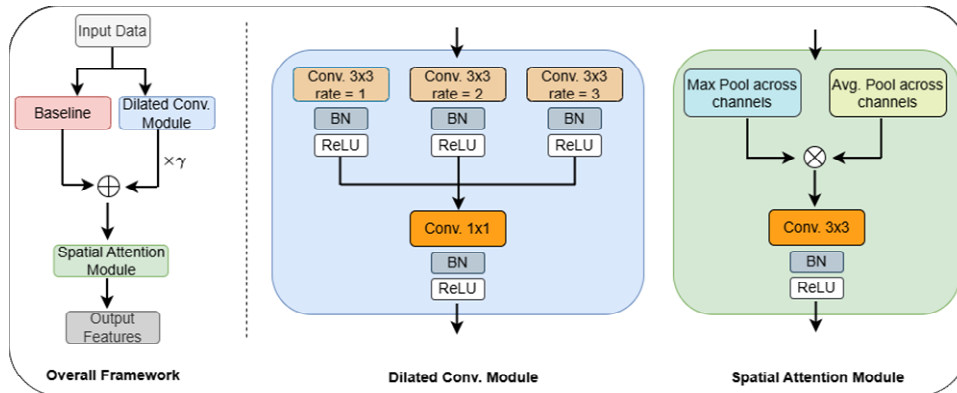


그림 1. 전반적인 프레임워크 및 세부 구성 요소
Fig. 1. Overall Framework and detailed Components

along with the baseline local feature encoder. First, we process the input streams through baseline encoder to maintain fine-grained structural data and concurrently, a global path broadens the model's field of view to capture macro-level interaction trends across distant areas of the city. Finally, because urban traffic is highly non-uniform, an adaptive spatial attention-based module is applied to automatically prioritize high-traffic hotspots over quiet residential zones.

The proposed global dilated convolution pathway and adaptive spatial attention module are detailed next and outline in Figure 1.

1. Global Representation

To resolve the localized field limitations of standard convolutions, we introduce a parallel dilated convolution branch. This module processes input streams across parallel paths, each utilizing a distinct dilation factor. The comparison of standard and dilated convolution operation is shown in figure 2. By expanding the spacing between kernel weights according to dilation rate, the model expands its receptive field to capture long-range contextual dependencies across distant grid coordinates. The multi-scale features are concatenated along the channel axis and down-projected through a channel reduction. This global

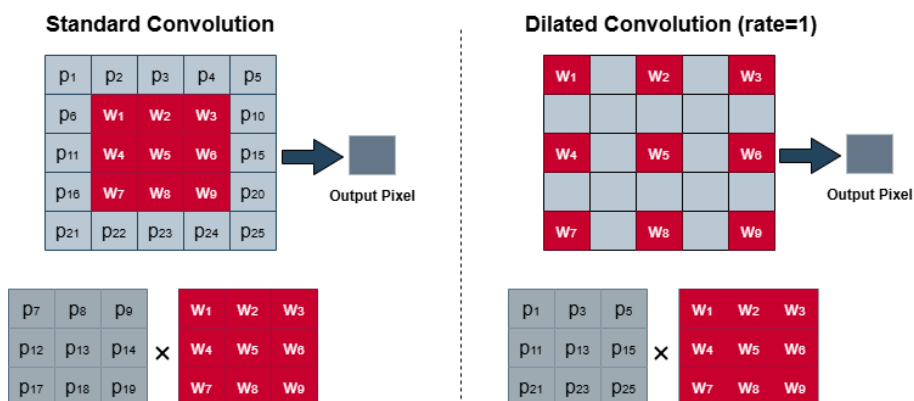


그림 2. 표준 합성곱 연산과 확장 합성곱 연산
Fig. 2. Standard versus dilated convolution operation

representation is scaled by a learnable weight parameter (γ), which determines the influence of global features, and added to the localized branch output.

$$X_{fused} = X_{local} + (\gamma \cdot X_{global}) \quad (1)$$

2. Spatial Importance

Following the fusion of all three branches, the feature maps pass through a spatial attention module to dynamically isolate traffic hot-spots. The module computes spatial statistics by simultaneously running average pooling and max pooling operations across the channel dimension, packing the local energy properties into two distinct spatial maps.

These statistical maps are processed through an intermediate bottleneck convolution and down-sampled to a single channel. Then a Sigmoid function bounds the values between 0 and 1, producing a spatial attention map that rescales the feature maps adaptively.

$$X_{att} = X \otimes Att(X) \quad (2)$$

V. Experiments and Results

The proposed framework was implemented using PyTorch and evaluated on two traffic modalities, SMS traffic and Call traffic. The dataset spans two months recorded at 10-minute intervals, which we aggregate to hourly resolution for more stable training. Following STDenseNet^[4] used as baseline, we hold out the final seven days as the test set and reserve 10% of the remaining data for validation. All reported results are averaged over three runs. We used grid search to find optimal dilation rates using SMS data with r in range of [1,2,3,4] then utilized same rates for all datasets.

In addition to the primary results in Table 1 discussed

in next section, to demonstrate that the performance gains stem from structural and architectural innovations rather than a simple inflation of network capacity, we perform an ablation study in Table 2 comparing our model against several specific configurations. These include the standard localized STDenseNet^[4] architecture (Baseline) and an up-scaled version of the baseline model with increased channels to match or exceed the parameter capacity of our proposed method (Baseline (CP)). Additionally, in Table 3 we evaluate models where the baseline network is augmented solely with the parallel multi-scale dilated convolution branch (Dilation) or solely with the adaptive spatial attention module (Attention). Model performance is quantified across both datasets using three standard statistical validation metrics, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the Pearson Correlation Coefficient (r).

1. Evaluation and Discussion

The empirical results in Table 1 across both data modalities demonstrate that our proposed framework consistently outperforms the baseline along with some graph and attention based methods. For the SMS traffic dataset, the standard Baseline records an RMSE of 14.70, an MAE of 7.25, and a correlation (r) of 0.79 with 85,184 parameters. Scaling up the baseline's capacity to 91,404 parameters in Table 2 for the comparable parameters configuration (CP) yields an improved RMSE of 13.55 and an MAE of 6.78. However, our unified Proposed model achieves the lowest overall error on the SMS dataset, dropping the RMSE to 12.59 and the MAE to 5.42, while pushing the correlation coefficient to 0.84 with only 89,317 parameters. A similar trend is observed in the Call traffic dataset. The Baseline starts at an RMSE of 7.75 and an MAE of 4.90, which the parameter-matched baseline improves to 6.81 and 4.13, respectively. The Proposed framework heavily outperforms both, reducing the Call traffic prediction error to an RMSE

of 6.01 and an MAE of 3.27, while maximizing the correlation at 0.87.

표 1. SMS 및 통화 트래픽 데이터셋에서의 RMSE를 활용한 정량적 평가
Table 1. Quantitative Evaluation using RMSE on SMS and Call Traffic Datasets

Model	SMS	Call
AMF-STGCN ^[16]	57.73	-
GLSTTN ^[19]	41.24	24.44
Att-MCSTCNet ^[1]	38.31	20.96
ADNet ^[17]	24.78	12.49
Baseline ^[4]	14.70	7.75
Proposed	12.59	5.42

Isolating the individual sub-modules via ablation testing in Table 3 confirms that both proposed architectural components provide distinct advantages. Evaluating the dilated branch in isolation (Dilation) reveals a substantial performance jump, particularly in Call traffic, where the RMSE drops sharply from 7.75 to 6.23. This proves that expand-

ing the receptive field allows the network to successfully capture correlations across distant spatial regions that standard localized convolutions miss. Meanwhile, the addition of the spatial attention module alone (Attention) drops the SMS MAE from 7.25 to 6.42. This behavior demonstrates the module's effectiveness in calculating cross-channel spatial statistics, enabling the network to dynamically prioritize highly variable traffic hotspots over quiet regions.

Ultimately, the lowest prediction errors are achieved when both modules are deployed simultaneously. Crucially, the Proposed model improves over the baseline model remaining parameter-efficient. It consistently outperforms the upscaled Baseline (CP) model across all metrics despite utilizing fewer parameters. This confirms that the accuracy gains are driven entirely by our targeted spatiotemporal design rather than raw parameter scaling, offering an optimized, resource-conscious solution for real-world citywide deployments.

For generalization, we validate our method on a recent

표 2. SMS 및 통화 트래픽 데이터셋에서의 비교 가능 파라미터 크기(CP)에 대한 Ablation
Table 2. Ablation study for Comparable Parameter size (CP) on SMS and Call Traffic Datasets

Model	RMSE		MAE		r		Parameters
	SMS	Call	SMS	Call	SMS	Call	
Baseline	14.70	7.75	7.25	4.90	0.79	0.83	85,184
Baseline (CP)	13.55	6.81	6.78	4.13	0.83	0.85	91,404
Proposed	12.59	6.01	5.42	3.27	0.84	0.87	89,317

표 3. SMS 및 통화 데이터셋에서의 개별 구성 요소의 기여도에 대한 Ablation
Table 3. Ablation study for contribution of individual components on SMS and Call Datasets

Model	Dilation	Attention	RMSE	MAE	r
SMS	x	x	14.70	7.25	0.79
	✓	x	13.59	6.64	0.83
	x	✓	13.35	6.42	0.83
	✓	✓	12.59	5.42	0.84
Call	x	x	7.75	4.90	0.83
	✓	x	6.23	3.70	0.86
	x	✓	6.85	3.84	0.85
	✓	✓	6.01	3.27	0.87

표 4. 5G 트래픽 데이터셋에서의 RMSE를 활용한 정량적 평가
Table 4. Quantitative Evaluation using RMSE on 5G Traffic Dataset

Model	Week 6	Week 11
Baseline	0.69	0.92
Proposed	0.66	0.89

5G traffic dataset^[18]. It covers 2,880 beams across 30 base stations, sampled hourly over a 5-week training period (840 hours), with weeks 6 and 11 held out for short- and long-term evaluation, respectively. We use downlink throughput volume (MB) per beam and reshape to a 30×96 grid, compatible with our method. As shown in Table 4, the proposed model consistently outperforms the baseline on the 5G dataset, reducing RMSE from 0.69 to 0.66 on the near-term week 6 split and from 0.92 to 0.89 on the longer-horizon week 11 split. This confirms that the gains from our method can generalize well to other datasets.

VI. Conclusion

In this work, we introduced an enhanced spatiotemporal architecture optimized for citywide cellular traffic prediction. By integrating a parallel multi-scale dilated branch alongside a spatial attention mechanism, our model effectively captures global spatial context and concentrates computational focus onto salient regions. Empirical results on SMS and Call traffic data show that our model achieves notable improvements in prediction accuracy, outperforming parameter-matched baseline. Future extensions will investigate implementing these spatial dependencies across irregular, non-grid city boundaries using Graph Neural Networks.

References

[1] Zeng, Qingtian, et al. "Attention based multi-component spatio-

temporal cross-domain neural network model for wireless cellular network traffic prediction." *EURASIP Journal on Advances in Signal Processing* 2021.1 (2021), 46.

doi: <https://doi.org/10.1186/s13634-021-00756-0>

- [2] Li, Rongpeng, et al. "Intelligent 5G: When cellular networks meet artificial intelligence." *IEEE Wireless communications* 24.5 (2017), 175-183.
doi: <https://doi.org/10.1109/MWC.2017.1600304WC>
- [3] Saxena, Navrati, Bharat JR Sahu, and Young Shin Han. "Traffic-aware energy optimization in green LTE cellular systems." *IEEE Communications Letters* 18.1 (2013), 38-41.
doi: <https://doi.org/10.1109/LCOMM.2013.111213.131809>
- [4] Zhang, Chuanting, et al. "Citywide cellular traffic prediction based on densely connected convolutional neural networks." *IEEE Communications Letters* 22.8 (2018), 1656-1659.
doi: <https://doi.org/10.1109/LCOMM.2018.2841832>
- [5] Bian, Xuan, et al. "TCI-UNet: transformer-CNN interactive module for medical image segmentation." *Biomedical Optics Express* 14.11 (2023), 5904-5920.
doi: <https://doi.org/10.1364/BOE.499640>
- [6] Barlacchi, Gianni, et al. "A multi-source dataset of urban life in the city of Milan and the Province of Trentino." *Scientific data* 2.1 (2015), 150055.
doi: <https://doi.org/10.1038/sdata.2015.55>
- [7] Zhou, Bo, Dan He, and Zhili Sun. "Traffic modeling and prediction using ARIMA/GARCH model." *Modeling and Simulation Tools for Emerging Telecommunication Networks: Needs, Trends, Challenges and Solutions*. Boston, MA: Springer US, 2006. 101-121.
doi: https://doi.org/10.1007/0-387-34167-6_5
- [8] D'Alconzo, Alessandro, et al. "A distribution-based approach to anomaly detection and application to 3G mobile traffic." *GLOBECOM 2009-2009 IEEE Global Telecommunications Conference*. IEEE, 2009.
doi: <https://doi.org/10.1109/GLOCOM.2009.5425651>
- [9] Li, Rongpeng, et al. "The learning and prediction of application-level traffic data in cellular networks." *IEEE Transactions on Wireless Communications* 16.6 (2017), 3899-3912.
doi: <https://doi.org/10.1109/TWC.2017.2689772>
- [10] De Gooijer, Jan G., and Rob J. Hyndman. "25 years of time series forecasting." *International Journal of Forecasting* 22(3), 443 - 473." (2006).
doi: <https://doi.org/10.1016/j.ijforecast.2006.01.001>
- [11] Chabaa, Samira, Abdelouhab Zeroual, and Jilali Antari. "Identification and prediction of internet traffic using artificial neural networks." *Journal of Intelligent Learning Systems and Applications* 2.03 (2010), 147-155.
doi: <https://doi.org/10.4236/jilsa.2010.23018>
- [12] Fu, Rui, Zuo Zhang, and Li Li. "Using LSTM and GRU neural network methods for traffic flow prediction." *2016 31st Youth academic annual conference of Chinese association of automation (YAC)*. IEEE, 2016.
doi: <https://doi.org/10.1109/YAC.2016.7804912>
- [13] Nie, Laisen, et al. "Network traffic prediction based on deep belief

- network in wireless mesh backbone networks.” 2017 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2017.
doi: <https://doi.org/10.1109/WCNC.2017.7925498>
- [14] Wang, Jing, et al. “Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach.” IEEE INFOCOM 2017-IEEE conference on computer communications. IEEE, 2017.
doi: <https://doi.org/10.1109/INFOCOM.2017.8057090>
- [15] Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. “Reducing the dimensionality of data with neural networks.” science 313.5786 (2006), 504-507.
doi: <https://doi.org/10.1126/science.1127647>
- [16] Wang, Xing, et al. “Adaptive multi-receptive field spatial-temporal graph convolutional network for traffic forecasting.” 2021 IEEE Global Communications Conference (GLOBECOM). IEEE, 2021.
doi: <https://doi.org/10.1109/GLOBECOM46510.2021.9685054>
- [17] Oh, Myeongjun, et al. “Comparative Analysis of Attention Mechanisms in Densely Connected Network for Network Traffic Prediction.” Signals 6.2 (2025), 29.
doi: <https://doi.org/10.3390/signals6020029>
- [18] Fechete, Luca-Andrei, et al. “Goal-oriented time-series forecasting: foundation framework design.” Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 40. No. 25. 2026.
doi: <https://doi.org/10.1609/aaai.v40i25.39249>
- [19] Gu, Bo, et al. “A spatial-temporal transformer network for city-level cellular traffic analysis and prediction.” IEEE Transactions on Wireless Communications 22.12 (2023), 9412-9423.
doi: <https://doi.org/10.1109/TWC.2023.3270441>

저 자 소 개



Maryam Qamar

- Sep. 2009 ~ Sep. 2013 : Bachelor's degree in Computer Science at University of Azad Jammu and Kashmir, Pakistan
- Sep. 2014 ~ Mar. 2017 : Master's degree in Computer Science at National University of Sciences and Technology, Pakistan
- Mar. 2022 ~ Present : She is currently pursuing Ph.D. degree with the Department of Artificial Intelligence at Kyung Hee University, South Korea
- ORCID : <https://orcid.org/0000-0003-0774-5411>
- 주관심분야 : Computer Vision and Machine Learning



배 성 호

- Mar. 2004 ~ Feb. 2011 : Bachelor's degree in Department of Computer Engineering and Electronic Engineering (dual majors) at Kyung Hee University, South Korea
- Feb. 2011 ~ Aug. 2016 : Ph.D. in Department of Electrical Engineering at KAIST, South Korea
- Jul. 2016 ~ Aug. 2017 : Postdoc. Associate in MIT Computer Science and Artificial Intelligence Laboratory (CSAIL)
- Sep. 2017 ~ Present : Associate Professor at Kyung Hee University in school of Computing, South Korea
- ORCID : <https://orcid.org/0000-0003-2677-3186>
- 주관심분야 : Efficient Machine Learning and Computer Vision