DBpia

논문 21-46-12-12 The Journal of Korean Institute of Communications and Information Sciences '21-12 Vol.46 No.12
https://doi.org/10.7840/kics.2021.46.12.2184

WiSECam: 무선 카메라를 위한 CSI 기반 딥 러닝 모션 감지

다오딩응웬[•], 무하마드 살만[*], 노 영 태[°]

WiSECam: A CSI-Based Deep Learning Motion Detection for Wireless Cameras

DinhNguyen Dao[•], Muhammad Salman[*], YoungTae Noh[°]

요 약

현재 감시 카메라는 안전과 보안을 개선하는 데 중요한 역할을 한다. Wi-Fi 카메라는 설치가 쉽기 때문에 널리 사용된다. 그러나 이러한 카메라는 장애물 뒤에 있는 물체를 볼 수 없는 등 시각적 제약이 있다. 또한, 먼지, 조명, 습기 등과 같은 환경적 요인도 카메라의 가시도에 부정적인 영향을 끼칠 수 있다. 이 문제를 해결하기 위해, 모니터링 영역의 모든 사각지대를 커버하려면 여러 대의 카메라를 사용해야 한다. 그러나 이 방법은 많은 비용이 든다. 그러므로 우리는 은행 사물함, 박물관 및 상점과 같이 도난에 취약한 곳에서 독립적으로 사용할 수 있는 WiSECam (WiFi-Security Enhanced Camera)이라는 저렴하고 응답력이 높으며 효과적인 방법을 제안한다. 이 방법은 와이파이 신호의 CSI (Channel State Information)를 사용하여 사람의 움직임을 감지한다. 우리는 CSI 데이터를 처리하고 CNN (Convolutional Neural Network)과 LSTM (Long Short-Term Memory)을 활용하여 딥러닝 모델을 구축한다. 우리는 고려된 실제 시나리오에서 이를 구현하고 평가한다. WiSeCam은 다양한 실생활 설정에서 1초 응답 시간으로 약 98%의 평균 정확도를 달성하여 실생활에서 사용할 수 있다.

Key Words : Channel State Information (CSI), Wi-Fi monitoring mode, deep learning, Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM)

ABSTRACT

Nowadays surveillance cameras play an important role in safety and security. In particular, a Wi-Fi camera is gaining popularity because of its easy installation. However, there is a visibility limitation of existing commodity cameras, for instance, they can not see an object behind an occlusion. In addition, environmental factors, such as dust, lighting, and humidity can also adversely affect the camera's visibility. To address these problems, the prevalent solutions use multiple cameras to cover the entire scene. But it comes with a cost. We, therefore, propose a cheaper, responsive, and effective solution called WiSECam (Wi-Fi-Security Enhanced Camera), which can be used as stand-alone in such as bank lockers, museums, and goldsmith stores that are proneto burglary. Our proposed solution leverages the Channel State Information (CSI) from Wi-Fi signals to detect human motion. We devise a deep learning model by using the Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) to handle the sequential CSI data. We implemented and evaluated it in

※ This research was supported in part by ICT Innovative Company Technology Development Support Program through the Institute for Information & Communication Technology Planning & Evaluation (IITP) funded by Ministry of Science and ICT (2020-0-01965), and by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2020R1A4A1018774)

◆ First Author : Inha University Computer Science Engineering, nguyen@nsl.inha.ac.kr, 학생회원
° Corresponding Author : Inha University Computer Science Engineering, ytnoh@nsl.inha.ac.kr, 종신회원
* Inha University Computer Science Engineering, salman@inha.edu,

논문번호 : 202109-247-B-RU, Received September 16, 2021; Revised September 28, 2021; Accepted September 29, 2021

www.dbpia.co.kr

the considered real-life scenarios. WiSECam achieves an average accuracy of around 98% with 1 second response time in different real-life settings.

Ⅰ. Introduction

Surveillance cameras are used for security purposes to monitor people in high secure places or to prevent crime and terrorism. Nowadays, the Wi-Fi camera is extensively used because of its easy installability and portability. However, there are some inherent limitations of commodity cameras. Firstly, a single camera cannot cover the entire scene in the target area and there might be some blind spots (due to the presence of occlusion). Secondly, the camera performance is prone to atmospheric conditions such as dust, lighting, and humidity, etc. To handle these challenges, multiple cameras are installed to cover all the blind spots in the monitoring area. However, such a solution is costly and requires more care to handle all the coordinating cameras.

The commodity cameras use 802.11n or 802.11g Wi-Fi standard. These standards utilize OFDM subcarriers for data transmission and use CSI to ensure reliable communication with high data rates in MIMO systems[1]. The wireless signals when propagating in a multipath-rich environment produce multiple reflected copies of the same signal. Such reflections degrade the channel quality that reflexes in the CSI signal at the receiver. Therefore, the CSI data is stimulated by the movements produced by the people in everyday activities[2]. Fig. 1 presents a sample application where CSI signals is used to detect thief motion even he moves behind the obstacle.

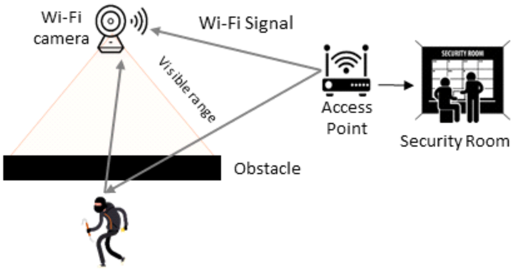

Fig. 1. A scenario where CSI is used to detect thief motion.

As a matter of fact, CSI in the raw form encounters a significant amount of variation and it is hard to choose a correct feature for motion detection in real-time. Many works related to Wi-Fi-based human activity recognition have been developed in the literature. Wang et al[3] use a traditional machine learning approach named CARM to classify human activity. The authors have processed the CSI by filtering its noise using the principal component analysis (PCA) denoising technique and then features have been extracted. These features were then classified for motion detection by different machine learning techniques such as logistic regression, support vector machines (SVMs), hidden Markov model (HMM), and deep learning. In another work, Long Short-Term Memory (LSTM) solves this detection problem using a unique additive gradient structure, which results in a boosted performance for time series with temporal dependency such as Wi-Fi CSI signals[4].

Herein, we propose a cheaper, responsive, and effective solution called WiSECam (WiFi-Security Enhanced Camera), which can operate with a single Wi-Fi camera in places such as bank lockers, museums, and goldsmith stores that are prone to burglary. We exploit the variation of Channel States Information (CSI) from a Wi-Fi camera and use a deep learning model that combines the Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to detect human motion.

The major contributions of our work are summarized as follows:

We propose deep learning models that exploit a combination of CNN and LSTM to improve the effectiveness of CSI-based motion detection. The CNN can learn CSI features automatically whilst LSTM processes the CSI data sequentially.

We thoroughly evaluated the accuracy of WiseCam in diverse and real-life settings. Our

www.dbpia.co.kr

settings include the effect of multipath by assessing it in rooms of different sizes, physical activity with varying intensities, and in the presence (or absence) of a person moving in the Line-of-sight (LoS).

We structure the remainder of this paper as follows. We introduce the design of WiSECam and explain the CSI-based human movement detection in section II. The experimental settings and results are thoroughly discussed in section III. Finally, section IV concludes this paper.

Ⅱ. Overview of Wisecam

In this section, we discuss the building blocks of WiSECam in detail.

2.1 CSI Data Collection

The CSI signal is vigorously affected by the multipath effect in indoor environments. This effect varies with the choice of room (room size), scatters inside the room (e.g., furniture), and human motion. To fully evaluate the performance of WiSECam, we tested our proposed method in 3 rooms. Rooms consist of different sizes and varied furniture arrangements. For each room, we conducted 3 physical activities: (1) light physical activity such as changing a jacket; (2) medium physical activity (i.e., moderate walking); and (3) intensive physical activity (i.e., running, jumping). We collected 40 samples for every activity performed in each room and recorded each sample for 30 seconds. Our collected samples were the combination of both sedentary (initially staying still for 15s) and physical activity. For the sake of sensitivity analysis, we tested our system with outside room movement. The details of the data collection parameters are shown in Table 1.

Table 1. Testing parameters

Room	Physical Activities	Movement
Large (6m×8m)	Light	Inside
Medium (4m×4m)	Medium	Outside
Small (1.5m×2.8m)	Intensive	Inside+Outside

2.2 CSI Cleaning

Raw CSI signals contain significant distortion which should be cleaned before conducting an evaluation. The data sanitizing step involves the removal of pilots and DC subcarriers. In particular. there are a total of 64 CSI subcarriers (52 data subcarriers, 4 pilots, and 8 DC subcarriers) in 20 MHz bandwidth and 2.4 GHz frequency spectrum. After sanitizing, we obtain 52 subcarriers amplitude used for training our deep learning model. Lastly, the magnitude is extracted from the sanitized CSI data.

2.3 Convolutional Neural Networks

CNN is widely used for extracting spatial features from data. It has been successfully applied on CSI data in localizing the indoor IoT devices[5], indoor people counting[6], sign language gestures recognition[7], and human activity recognition[8]. In this paper, we utilize 1D CNN in our model for automatically extracting the CSI features in contrast to statistically defined features that require domain knowledge and exhaustive effort. Fig. 2 illustrates the architecture of 1D CNN, consisting of Convolutional layers, pooling layers, and fully connected layers. The convolution operation uses a single filter to compute one feature map, then the ReLu activation function is used. After that, the pooling blocks are used to reduce the spatial size, keeping only optimal features in the feature map. This makes the system more robust to noise. Based on the specific application, these basic blocks can be increased or reduced.

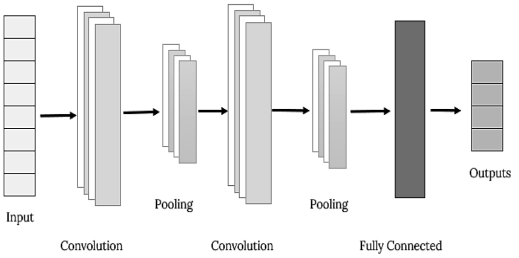

Fig. 2. A sample 1D CNN with 2 convolution layers and 1 fully-connected layer.

www.dbpia.co.kr

2.4 Long Short-Term Memory Networks

To handle the WiFi CSI time series data, we apply the LSTM network. Compared with RNNs, LSTM has memory cells to store and access information over long periods. As a result, they are better at finding and exploiting long-range dependencies in the data. The key component of LSTM is the cell state, which stores and passes information between LSTMs. Each LSTM block has three gates connected to the cell which are Forget gate, Input gate, and Output gate. Fig. 3 presents a single LSTM memory block.

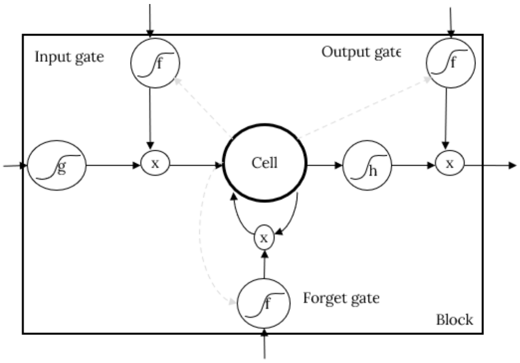

Fig. 3. A sample of a single LSTM block.

2.5 The Model

Fig. 4 shows three proposed deep learning models for CSI-based human motion detection. The input data is the amplitude of the sanitized CSI (i.e. 52 data sub-carriers) recorded in a sliding window of 1 second. To learn features from the CSI signals, three layers stacked 1D CNN are used followed by the Max Pooling layer. The LSTM layer is used to learn the current state from the past. Finally, the fully connected layer takes the feature vector and classifies the user's state (i.e. whether or not the user is moving). In this paper, LSTM and CNN layers are combined to build LSTM model, CNN-LSTM model, and LSTM-CNN model. These models are used for comparison in the next section.

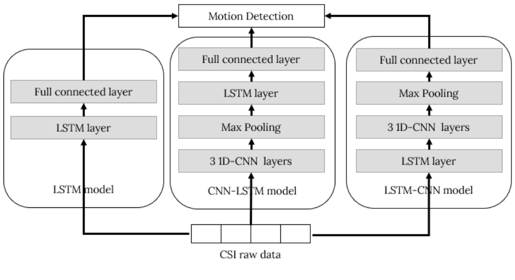

Fig. 4. Block view of LSTM, CNN-LSTM and LSTM-CNN models

For model parameters, we chose 3 1D-CNN layers with 64 filters, which are the optimal parameters obtained from Section IV.1, the kernel size = 3 for extracting the features with information from the neighboring data, and uses a ReLU activation function. The pool size is set to 3 for the Max-pooling layer to highlight the most present features. For the LSTM layer, we use 100 units to store the hidden state of the CSI data. After CNN and LSTM layers, the output is sent to the fully connected output layer with a Softmax activation binary classification output, i.e., Static and Dynamic. For training models, we set the learning rate as 0.0001 for Adam optimizer, minibatch size as 32, and the training epochs as 200, which are got from running an optimization framework called Optuna. We chose these hyper-parameters to be consistent across all models as these hyper-parameters and had little impact on performing the respective models.

Ⅲ. Experiment and Results

3.1 Implementation

We used a Raspberry pi 3 (RP3) to extract CSI raw data from Wi-Fi signals. To handle the CSI extraction, we modified the WLAN firmware of RP3 using the Nexmons' open-sourced GitHub repository[9]. Moreover, we enabled the monitoring mode feature, then listened to the frames at a particular UDP socket. Each captured UDP frame was collected on port 5500 with a source address as 10.10.10.10 and destination address as 255.255.255.255 to extract CSI data.

All the models are built on Keras with TensorFlow backend. Comparisons are made using accuracy. Input data for the model is a 1-second temporal window, which contains 30 frames. Each frame contains data of 52 subcarriers. After applying a sliding window to extract 1-second time-series

www.dbpia.co.kr

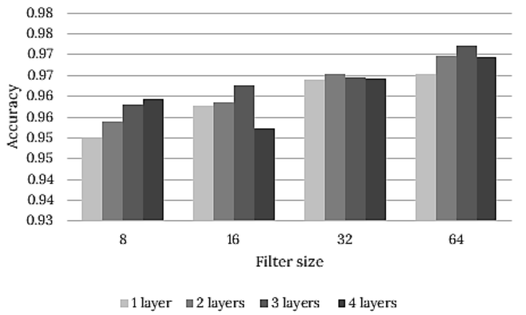

Fig. 5. Impact of number of layers and filter size on accuracy.

data we obtain 40,000 data samples. Those samples are used for training and testing. We perform 10-fold cross-validation for evaluation. Aggregated accuracy was derived from the average of all 10 runs.

3.2 Parameters tuning

To optimize the 1D CNN structure, we investigated the effect of the number of layers and the number of filter sizes on accuracy. It can be seen in Fig. 5 that the overall accuracy improves with the increase in the number of layers (from 1 to 4). It can be also observed that the 2∼3 layers are most suitable for our application. In addition, by applying various output filter sizes, we also found that a filter size of 64 was the best fit for our purpose. As a result, we use 3 layers for our CNN model and a filter size of 64 in all the models' evaluations.

3.3 Overall Accuracy

We present the accuracy attained by various models in Fig. 6. We can observe that our proposed CNN-LSTM (97.6%) model outperformed other DL-based counterparts with LSTM-CNN having 96.8% accuracy and LSTM of 95%. The performance gain of these models is owing to the CNN layer that extracts rich features from the CSI data. We speculate that the CNN-LSTM model achieved better accuracy than the LSTM model because of its efficiency in retaining more information when extracting features from raw CSI data compared to the LSTM model.

3.4 Accuracy with varying room sizes

We exploited the room size due to the effect of multipath reflection in the CSI signal caused by different indoor settings. We conducted our experiment in three rooms with different sizes, namely small, medium, and large sizes as in Table 1. In each room, we recorded 40 samples for each physical activity, and the duration of each sample was 10 seconds static then next 20 seconds, diverse physical activity intensities (such as light, medium, and intensive) were performed. The accuracy results are shown in Fig. 7.

In all the environmental settings, the CNN-LSTM model outperformed the other models. We note that the accuracy in the smaller room is lower than others because the smaller-sized room has a smaller spacing between walls. As a result, the CSI data experiences an unstable behavior due to the severe multipath effect. When we increase the space between walls (i.e. larger rooms), a more stable behavior of CSI leads to higher accuracy.

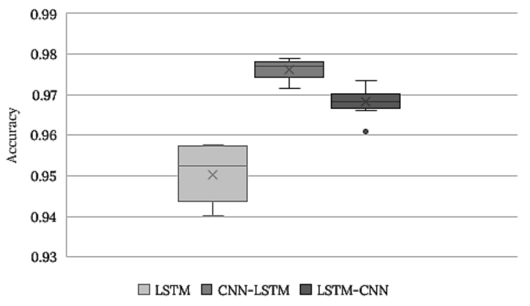

Fig. 6. The motion detection accuracy of LSTM, CNN-LSTM, and LSTM-CNN models.

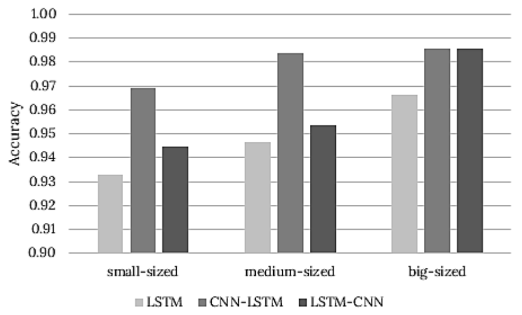

Fig. 7 The detection accuracy in varying room sizes.

www.dbpia.co.kr

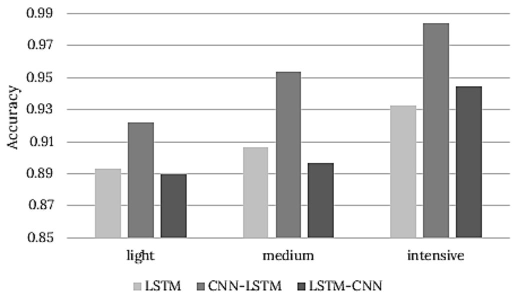

Fig. 8. The detection accuracy with varying physical activities.

3.5 Accuracy with varying physical activities

To measure the sensitivity of WiSECam in response to the intensity of physical activity, we conduct our experiment with three types of activities: (1) light physical activity where the individual is changing the jacket, (2) medium physical activity where the user is walking around the room at a moderate pace, and (3) intensive physical activity where the user is running.

This experiment aims to measure the sensitivity of camera detection in response to the intensity of physical activity. As evident from Fig. 8, the accuracy of the model is proportional to the intensity of physical activity. Clearly, the intensive activity case gets the better accuracy, with the CNN-LSTM model, we achieve around 98% of accuracy, whereas, with the medium and light physical activity, the accuracy is around 95% and 92% respectively. LSTM and LSTM-CNN models are about 2% lower in accuracy than the CNN-LSTM model in each case.

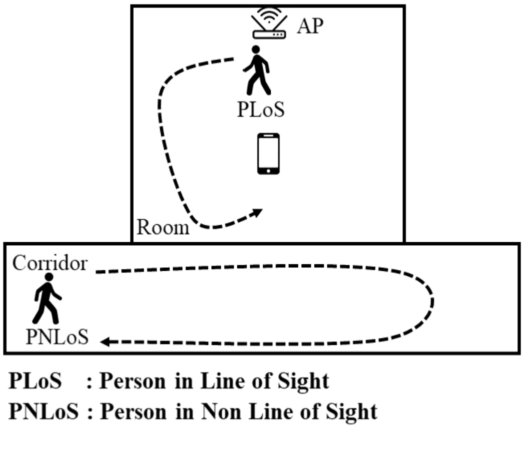
AP
PLoS
Room
Corridor
PNLoS
PLoS : Person in Line of Sight
PNLoS : Person in Non Line of Sight

Fig. 9. Detect motion with a Person in Non Line of Sight.

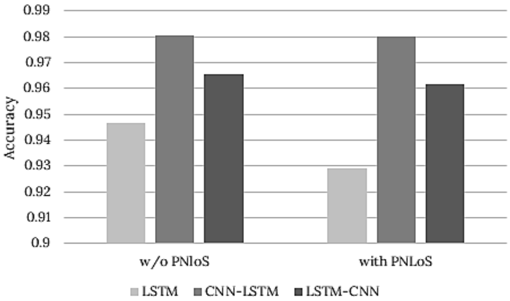

Fig. 10. The detection accuracy with and without and without a Person-in-Non-Line-of-Sight.

3.6 Accuracy with activity outside the room

To check the robustness of WiSECam, we do the experiment with and without the movement of Person in Non-Line-of-Sight (PNLoS). We set up a person continuously moving across the corridor when we detect a movement inside the room that is shown in Fig. 9. It can be seen from the result in Fig. 10 that compared to a case with PNLoS, the accuracy of the case without PNLoS improves a little with a small margin, only 0.2% with the CNN-LSTM model to around 1.5% with the LSTM model. Interestingly, the movement outside the room has little effect on the CSI captured in the room.

Ⅳ. Conclusion

In this paper, we proposed WiSECam, a deep learning-based lightweight approach for detecting human movement by analyzing the fluctuations of CSI data obtained from an AP. Thus, by passively sensing, a single camera can detect the human motion in the target scene, even behind the occlusion. Our extensive results show that WiSECam can attain an average accuracy of 97.6% with 1 second of response time, thus making it feasible to use in real-time.

www.dbpia.co.kr

References

[1] Y. Xiao, "IEEE 802.11 n: Enhancements for higher throughput in wireless LANs," *IEEE Wireless Commun.*, vol. 6, no. 12, pp. 82-91, 2005.

[2] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," *ACM Comput. Surv.*, vol. 52, no. 3, 2019.

[3] W. Wang, A. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE J. Sel. Areas in Commun.*, vol. 35, no. 5, pp. 1118-1131, 2017.

[4] S. Yousefi, H. Narui, and S. Dayal, "A survey on behavior recognition using WiFi channel state information," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 98-104, 2017.

[5] B. Berruet, O. Baala, A. Caminada, and V. Guillet, "DelFin: A deep learning based CSI fingerprinting indoor localization in IoT context," in *Int. Conf. IPIN*, Nantes, France, Sep. 2018.

[6] Y. K. Cheng and R. Chang, "Device-free indoor people counting using wi-fi channel state information for internet of things," in *GLOBECOM 2017*, Singapore, Dec. 2017.

[7] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "Signfi: Sign language recognition using wifi," in *Proc. ACM Interactive Mob. Wearable and Ubiquitous Tech.*, vol. 2, no. 1, pp. 1-21, Mar. 2018.

[8] Y. Chung and Y. Jung, "Design and implementation of CNN-Based human activity recognition system using WiFi signals," in *J. Advanced Navig. Technol.*, vol. 25, no. 4, pp. 299-304, 2021.

[9] F. Gringoli, M. Schulz, J. Link, and M. Hollick, "Free your CSI: A channel state information extraction platform for modern wi-fi chipsets," in *Proc. 13th Int. Wkshp. Wireless Netw. Testbeds, Experimental Evaluation & Characterization*, pp. 21-28, 2019.

다오딩응웬 (DinhNguyen Dao)

2013년 2월 : 하노이백과대학 기계 공학과 졸업

2016년 8월 : 인하대학교 컴퓨터 석사 졸업

2018년 3월~현재 : 인하대학교 컴퓨터공학과 박사과정

<관심분야> 무선 네트워크, 트래픽 모니터링, 딥 러닝. 침입 탐지

[ORCID:0000-0001-6530-5188]

무하마드 살만 (Muhammad Salman)

2010년 8월 : 파키스탄 바루치스탄 정보통신공학 및 경영과학대학교(BUITEMS) 전자공학과 졸업

2014년 8월 : 테크니코 디 토리노(Politecnico di Torino, 토리노공과대학) 전자공학과 석사 졸업

2014년 10월~2019년 6월 : 사우디아라비아 에팟 대학교 전자 컴퓨터 공학과 강사

2019년 3월~현재 : 인하대학교 컴퓨터공학과 박사과정

<관심분야> 버퍼블로트 완화, 무선 네트워크 및 소프트웨어 정의 네트워크

[ORCID: 0000-0003-4754-805X]

노 영 태 (YoungTae Noh)

2004년 8월 : 조선대학교 전계산학과 학사

2007년 2월 : 광주과학기술원정보기술공학부 석사

2007년 9월 : 광주과학기술정보기술원 정보통신공학과 연구원

2012년 6월 : University of California 컴퓨터 과학과박사

2012년 7월~2014년 11월 : Cisco Systems Software Engineer

2015년 2월~2015년 8월 : Purdue University Post-doc

2015년 9월~현재 : 인하대학교 컴퓨터 공학과 교수

<관심분야> 데이터 센터 네트워킹, 무선 네트워킹, HCI(Human-computer interaction), 머신러닝

[ORCID:0000-0002-9173-1575]

www.dbpia.co.kr