디지털산업정보학회 논문지 제19권 제3호-2023년 9월
http://dx.doi.org/10.17662/ksdim.2023.19.3.245

MobileNet과 TensorFlow.js를 활용한 전이 학습 기반 실시간 얼굴 표정 인식 모델 개발

차 주 호*

Development of a Ream-time Facial Expression Recognition Model using Transfer Learning with MobileNet and TensorFlow.js

Cha Jooho

〈Abstract〉

Facial expression recognition plays a significant role in understanding human emotional states. With the advancement of AI and computer vision technologies, extensive research has been conducted in various fields, including improving customer service, medical diagnosis, and assessing learners' understanding in education. In this study, we develop a model that can infer emotions in real-time from a webcam using transfer learning with TensorFlow.js and MobileNet. While existing studies focus on achieving high accuracy using deep learning models, these models often require substantial resources due to their complex structure and computational demands. Consequently, there is a growing interest in developing lightweight deep learning models and transfer learning methods for restricted environments such as web browsers and edge devices. By employing MobileNet as the base model and performing transfer learning, our study develops a deep learning transfer model utilizing JavaScript-based TensorFlow.js, which can predict emotions in real-time using facial input from a webcam. This transfer model provides a foundation for implementing facial expression recognition in resource-constrained environments such as web and mobile applications, enabling its application in various industries.

Key Words : Machine Learning, Transfer Learning, Mobilenet, TensorFlow.js, Facial Expression Recognition

Ⅰ. 서론

얼굴 표정 인식은 인간의 감정 상태를 파악하는데 중요한 역할을 한다. 인간의 감정 인식 기술은 심리학, 인지 과학, 컴퓨터 과학 등 여러 분야에서 관심을 받고 있으며, AI와 컴퓨터 비전 기술이 발전하면서 광범위한 응용 연구들이 수행되고 있다. 이러한 감정

* 청운대학교 공과대학 멀티미디어학과 교수

인식 기술은 효율적인 감정 인식을 통한 고객 서비스 개선, 의료 진단 기술 개발, 교육에서의 학습자의 이해도 판단 등 다양한 산업 분야에 활용될 수 있다.

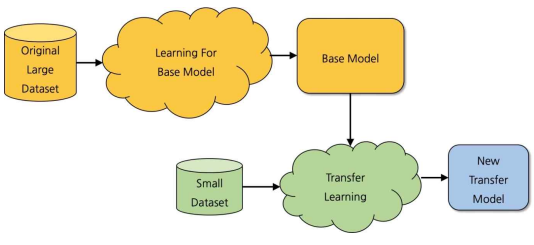
본 연구에서는 TensorFlow.js[1]와 MobileNet[2]을 활용하여 7가지 얼굴의 감정을 전이 학습(Transfer Learning)[3]을 수행함으로써 웹캠으로 부터 실시간으로 입력된 얼굴의 감정을 추론할 수 있는 모델을 개발한다. 기존의 얼굴 감정 인식 연구들은 대부분 딥러닝 모델의 활용을 통해 높은 정확도를 달성하는 것을 목표로 하고 있다. 그러나 이러한 모델들은 딥러닝 모델의 복잡한 구조와 높은 연산량으로 인해 실행에 있어 많은 자원을 필요로 한다. 이로 인해 상대적으로 제한적인 웹브라우저나 엣지 디바이스와 같은 환경에서는 활용에 어려움이 있다. 따라서 웹브라우저나 엣지 디바이스를 위한 보다 경량화된 딥러닝 모델이나 전이학습에 대한 연구가 중요한 주제로 떠오르고 있다. 본 연구에서는 Mobilenet을 베이스 모델로 사용하여 전이 학습을 수행하고, 웹브라우저나 드론과 같은 엣지 디바이스에서도 동작할 수 있도록 자바스크립트 기반의 TensorFlow.js를 활용하여 딥러닝 전이 모델을 개발한다. 본 연구에서 개발된 전이 모델은 웹브라우저 상에서 웹캠을 통해 사용자의 얼굴 정보를 입력으로 받아 실시간으로 감정을 예측할 수 있게 된다. 따라서 본 연구에서 개발된 전이모델은 웹과 모바일 등의 제한된 자원 환경에서도 얼굴 표정 인식 기능을 쉽게 구현할 수 있게 함으로써 다양한 분야에서 적용될 수 있는 기반을 제공할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 MoibleNet, 전이 학습, TensorFlow.js 등 본 연구를 위해 사용된 주요 관련 기술에 대해 기술하고, 3장에서는 MobileNet을 활용한 전이 학습 기반 얼굴 감정 예측 모델의 구축 방법, 학습 데이터 획득 방법, 그리고 TensorFlow.js를 사용한 구현 방법 등에 대하여 설명한다. 4장에서는 본 논문에서 개발한 전이 학습 모델을 테스트하고 평가한다. 마지막 5장에서는 결론 및 향후 과제에 대하여 기술한다.

II. 관련 연구

최근의 얼굴 표정 인식 연구들은 인간의 감정 표현에서 다양한 공간적 및 시간적 패턴을 찾기 위해 주로 머신 러닝과 같은 인공 신경망 기술에 의존한다. 머신 러닝 기반의 모델들은 주로 CNN (Convolutional Neural Networks)[4], RNN (Recurrent Neural Networks)[5], LSTM(Long-Short Term Memory) [6] 등과 같은 다양한 신경망 구조를 활용하여 얼굴 이미지의 특성을 추출하고, 이를 분류하는 방식으로 얼굴 표정을 인식한다.

MobileNet[2]은 ImageNet[7]의 경량화된 딥러닝 모델로, 비교적 낮은 계산 자원을 필요로 하지만, 높은 성능을 달성하는 이미지 분류를 위한 대표적인 딥러닝 모델이다. MobileNet의 모든 층은 배치 정규화와 ReLU 활성화 함수가 적용되며, 마지막 출력 층은 Softmax 활성화 함수가 적용된다. MobileNet은 일반적인 컨볼루션 계층을 깊이별 분리 합성곱(Depthwise Separable Convolution)으로 대체하여 모델의 크기를 최소화하고 필요한 연산량을 줄여준다. 이러한 특징으로 인해 MobileNet은 빠른 속도 제공뿐만 아니라 적은 메모리 사용량을 가짐으로써 엣지 디바이스나 웹 기반 환경에 적합하다.



<그림 1> 전이 학습의 개념

전이 학습은 기존에 미리 학습된 베이스 모델의 가중치(weight)와 구조(layer structure)를 가져와 새로운 문제에 적용하는 머신러닝 기술이다. <그림 1>에서 보인 것과 같이 전이 학습은 대규모 데이터 셋에 대해 학습된 기존 베이스 모델의 지식을 활용하여 작은 데이터 셋과 제한된 자원 상황에서도 높은 성능을 달성할 수 있도록 하는 기술이다. 이러한 전이 학습의 핵심은 이전에 학습한 베이스 모델을 재사용하여 빠르게 새로운 학습을 수행하여 새로운 머신러닝 모델을 생성하는 것이다.

TensorFlow.js는 웹브라우저 환경에서 딥러닝 모델을 실행할 수 있게 해주는 자바스크립트 라이브러리이다. TensorFlow.js를 사용함으로써 사용자는 서버와의 통신 없이도 웹 기반 애플리케이션에서 실시간으로 딥러닝 모델을 사용할 수 있다. 따라서 모델 개발뿐만 아니라 학습 및 추론을 웹브라우저 상에서 수행할 수 있으며, 클라이언트 측 기기에서 빠르게 딥러닝 예측을 처리할 수 있다. TensorFlow.js를 사용하면 자원이 제약된 브라우저와 같은 환경에서 전이 학습을 수행할 수 있다.

본 연구의 목표는 앞에서 설명한 MobileNet을 베이스 모델로 한 전이 학습을 통해 얼굴 표정을 인식할 수 있는 새로운 다중 분류 모델을 TensorFlow.js를 활용하여 자바스크립트로 구현함으로써 웹브라우저 상에서 웹캠을 통해 실시간으로 사용자의 기분을 검출하는 시스템을 개발하는 것이다.

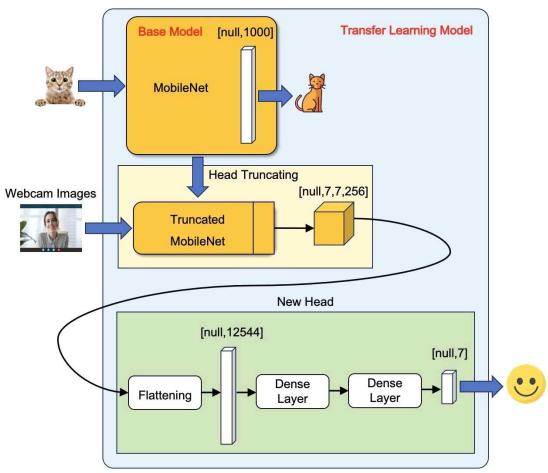
Ⅲ. Mobilenet 기반 전이 학습 모델 개발

3.1 전이 학습 데이터 준비

본 전이 학습 모델에서 사용하는 데이터 셋은 표정 인식에 적합한 특징을 가진 얼굴 이미지로 구성된다. 총 7가지의 감정 레이블로 구성되며, 화남, 행복, 놀람, 슬픔, 무표정, 혐오, 공포 등이 포함된다. 이들 전이 학습에 필요한 데이터 셋은 웹캠을 통해 수집한다. 웹브라우저 상에서 모델이 학습할 데이터를 각 레이블에 맞게 웹캠으로 초당 20에서 30 프레임의 속도로 이미지를 수집하였다. 최상의 전이 학습 품질을 얻기 위해 각 표정(클래스)마다 최소한 100개의 이미지를 수집하였다. 또한 이미지의 다양성을 위해 얼굴을 조금씩 움직이고, 시선의 방향을 조금씩 바꿔가며 데이터를 수집하였다. 웹캠으로부터 수집된 이미지 데이터 중 학습 데이터 셋으로 70%, 검증 데이터 셋으로 15%, 그리고 나머지 15%는 테스트 셋으로 사용한다.

3.2 전이 학습 모델 구조

본 연구에서는 전이 학습을 위한 베이스 모델로 케라스에서 사전 훈련된 MobileNet v1 모델[5]을 사용하였다. 이 모델은 사람의 얼굴을 인식하도록 훈련되지 않았을 뿐만 아니라, 입력 이미지에 대한 출력의 크기(출력 크기 1,000개)가 본 연구에서 목표로 하는 감정 예측의 출력 크기(출력 크기 7)와 일치하지 않는다. 따라서 목표로 하는 얼굴 감정 예측 전이 모델은



<그림 2> MobileNet 기반 전이 학습 모델 구조

베이스 모델인 MobileNet에서 출력 층(헤드, head)을 삭제하고, 헤드가 삭제된 MobileNet(Truncated MobileNet)의 출력을 flatten 층을 통해 새로운 밀집층과 7개의 감정 클래스를 갖는 출력 층을 갖는 새로운 모델과 결합하였다. 7개의 감정 인식 전이 모델 또한 다중 클래스 모델이므로 출력 층은 다중 분류를 위한 소프트맥스(Softmax) 활성화 함수와 범주형 크로스엔트로피(Categorical Crossentropy)를 손실 함수로 사용한다. 또한 원-핫 인코딩(One-hot encoding)을 수행하여 7개의 표정 중 예측된 하나의 표정을 출력한다. 구축된 전이 학습 기반 모델은 웹캠으로 부터 수집된 데이터 셋으로 학습시키고, 미세 조정을 통해 성능을 최적화하였다.

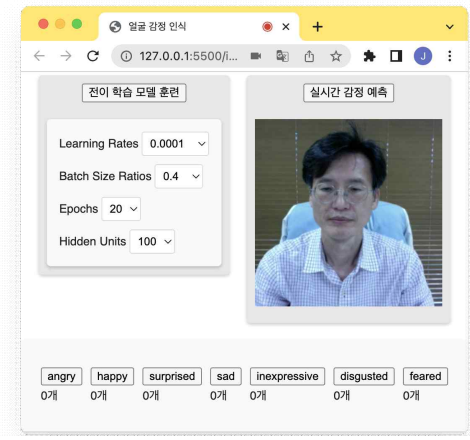
3.2 전이 학습 모델 구현

TensorFlow.js에서 모델 객체의 두 가지 중요한 속성은 입력과 출력이다. 본 연구의 새로운 전이 학습 모델에서는 심볼릭 텐서를 사용해 새로운 모델을 만들었다. 이를 위해 입력과 출력을 심볼릭 텐서 객체을 받는 tf.model() 함수를 사용하였다. 여기에서 입력은 베이스 모델인 MobileNet의 입력이고, 출력은 MobileNet이 갖는 헤드(head) 이전의 특정 층을 지정하였다. 따라서 tf.mode() 함수는 베이스 모델인 MobileNet 모델의 헤드가 제거된 일부분으로 구성된 새로운 모델을 만들어 준다. 우리의 구현에서 사용하는 MobileNet 모델은 헤드 포함 93개 층을 가지고 있으며, 새롭게 생성된 모델은 헤드를 제외한 특성 추출 층들만 갖는 81개 층으로 구성하였다. 웹캠으로부터 수집한 7개의 이미지 셋의 이미지는 각각 [244 × 244 × 3] 크기를 갖는다. 이들 입력 이미지는 헤드가 없는 MobileNet의 predict() 메소드를 사용하여 [7 × 7 × 256] 크기의 출력 임베딩을 얻을 수 있게 된다. 이 출력 이미지는 다시 tf.layers.flatten() 메소드를 사용하여 1차원 텐서로 변경한 후 은닉 밀집층으로 전달된다. 첫 번째 은닉층은 비선형 활성화 함수인 ReLU를 사용하고, 두 번째 은닉층은 새로운 모델의 출력층으로 각 표정에 대한 7개의 클래스를 분류하기 위한 소프트맥스 활성화 함수를 사용하였다.

IV. 전이 학습 모델 성능평가

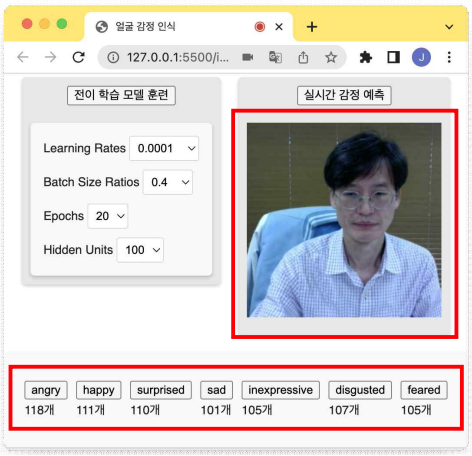
4.1 모델 테스트

먼저 전이 학습을 위한 입력 이미지 데이터 셋을 얻기 위해 웹캠을 통해 7개의 감정에 대한 표정 각각을 약 100개 정도씩 수집한다. 이는 <그림 3>에서 보인 것과 같이 아래쪽에 있는 각 감정에 대한 7개의 버튼에 맞게 웹캠에 나타나는 표정을 지은 후 각 감정 버튼을 누르고 있으면 자동으로 해당 감정에 대한 이미지와 레이블을 갖는 이미지 데이터가 수집된다. 웹캠으로부터 수집된 이미지 데이터의 샘플 개수는 각 감정 버튼 아래에 나타난다.

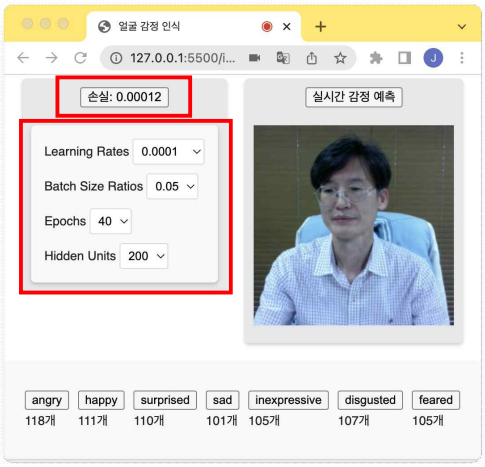


<그림 3> 웹에서 실시간 감정 예측을 위한 브라우저 화면

<그림 4>는 각 감정에 대한 입력 데이터 셋을 웹캠으로부터 수집한 결과를 보인 것이다. 각 감정 버



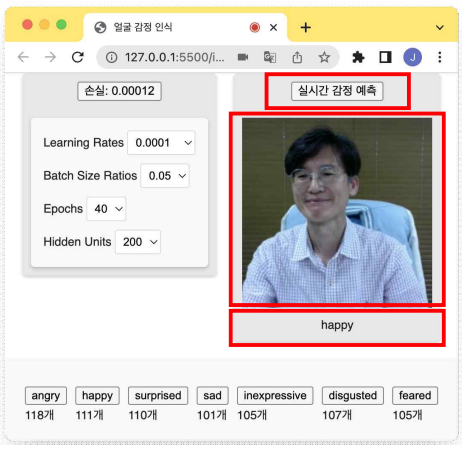
<그림 4> 웹캠으로부터 수집된 각 감정 이미지 데이터 샘플의 수



<그림 5> 전이 학습 실행 결과(손실율 0.00012)

튼의 아래에는 수집된 이미지 샘플의 수가 나타나있다. 이렇게 전이 학습을 위한 입력 데이터 셋이 준비되면 화면의 좌측 상단에 있는 전이학습 훈련 버튼을 클릭하여 학습을 시작할 수 있다. 이러한 전이 학습은 수 초만에 완료되며, 학습이 진행되면서 손실 값이 계속적으로 줄고 변하지 않게 된다(<그림 5> 참조). <그림 5>는 학습의 결과로 손실율이 0.00012인 결과를 보인 것이다. 이 결과를 위해 설정한 학습율은 0.0001, 배치 크기 비율은 0.05, 에포크는 40, 히든 유닛은 200이다. 이제 손실이 0.00012인 전이 학습 모델이 구해졌으므로, 실시간 감정 예측 버튼을 클릭하여 웹캠으로부터 입력되는 영상에 대해 실시간 감정 추론을 수행할 수 있다.

<그림 6>과 같이 전이 학습 모델의 실시간 감정 추론은 웹캠으로부터 받은 이미지 스트림에 대해 실시간으로 추론을 수행할 수 있다. 실시간 감정 예측의 결과는 웹캠의 입력 비디오 프레임마다 실시간으로 전이 학습 모델이 7가지 감정 중 가장 높은 확률 점수를 가진 클래스(감정)를 웹캠 이미지 아래에 출력해준다. 본 전이 학습 모델은 전이 학습으로 인해 많은 데이터와 오랜 학습 시간이 필요하지 않을 뿐만 아니라, 웹브라우저가 지원되는 모든 스마트폰에서도 실시간으로 사용자의 감정 추론을 가능하게 해준다.



<그림 6> 전이 학습 모델의 실시간 감정 예측 수행

4.2 모델 성능평가

본 실시간 얼굴 감정 인식 모델의 경우에는 높은 정확도와 빠른 응답시간이 가장 핵심적인 성능평가 지표가 된다. 다중 클래스 분류를 위한 정확도

($Accuracy$)는 수식(1)과 같이 정확히 예측한 평균 개수로 정의된다.

$$Accuracy = \frac{1}{N}\sum_{k=1}^{G}\sum_{x=k} I(g(x) = \hat{g}(x)) \quad (1)$$

여기에서 $G = \{1, \cdots, K\}$로 7개의 클래스를 나타내므로 K의 최댓값은 7이 된다. 그리고 I는 클래스가 일치하면 1을 반환하고, 그렇지 않으면 0을 반환하는 함수이다.

다중 분류에서는 True Negative(TN)의 개념이 적용되지 않는다. 다중 분류에서는 각 클래스별로 정의된 True Positive(TP), False Positive(FP), 그리고 False Negative(FN)에 초점을 맞추어야 한다. 이러한 이유로 본 논문에서는 다중 클래스 모델을 위한 성능 평가 지표로 각 클래스별 평균 정밀도(P_{macro})와 평균 재현율(R_{macro}), 그리고 평균 F1 Score($F1_{macro}$)를 사용한다. 평균 정밀도(P_{macro}), 평균 재현율(R_{macro}) 및 평균 F1 Score($F1_{macro}$)을 구하기 위한 수식은 아래와 같다.

$$P_{macro} = \frac{1}{G}\sum_{i=1}^{G}\frac{TP_i}{TP_i + FP_i} \quad (2)$$

$$R_{macro} = \frac{1}{G}\sum_{i=1}^{G}\frac{TP_i}{TP_i + FN_i} \quad (3)$$

$$F1_{macro} = 2\frac{P_{macro} \times R_{macro}}{P_{macro} + R_{macro}} \quad (4)$$

<표 1> 성능평가 결과

$Accuracy$	P_{macro}	R_{macro}	$F1_{macro}$
0.97	0.89	0.88	0.885

<표1>에서 보인 것과 같이 정밀도와 재현율 및 F1 Score는 모두 0.85이상의 결과를 보여주고 있다. 일반적으로 본 논문에서 수행하는 실시간 추론의 경우에는 0.85 이상의 값을 가질 때 모델이 높은 성능을 갖는다고 말할 수 있다. 또한 전이 학습이 학습할 데이터 셋이 부족한 상황에서 수행하는 상황적 한계성 측면에서도 매우 좋은 성능을 보여주고 있다.

실제 애플리케이션에서는 모델이 높은 성능뿐만 아니라 빠른 예측 시간을 요구한다. 따라서 감정 예측 모델의 추론 속도가 중요한 평가 요소이다. 본 논문에서 구현된 전이 학습 모델은 웹브라우저 상에서 웹캠을 통해 입력되는 영상에 대해 실시간으로 추론을 수행할 수 있다. 즉 머신러닝 모델을 통한 추론을 수행하기 위해 서버로 영상 데이터를 보내고, 그 결과를 받기 위해 대기할 필요가 없다는 것이다. 본 연구에서의 전이 학습 모델의 추론 속도는 웹브라우저에서 40ms 내외였다. 이는 실시간 애플리케이션에서 충분한 성능이다. 결론적으로, Mobilenet을 활용한 전이 학습 기반 얼굴 감정 추론 모델은 빠른 추론 속도와 범용성을 갖추고 있다. 이와 같은 실시간으로 추론을 수행할 수 있는 전이 학습 모델을 TensorFlow.js를 사용해 강력한 머신러닝 모델을 웹브라우저에서 바로 사용할 수 있어서 다양한 응용 분야에 활용할 수 있다.

Ⅴ. 결론

본 논문에서는 Mobilenet을 활용한 전이 학습 기반 얼굴 감정 예측 모델을 제안하고, TensorFlow.js를 활용하여 프론트엔드 웹에서 구현하고 실험하였다. 실험 결과 웹브라우저 상에서의 실시간 추론을 통해 모델의 빠른 추론 속도와 높은 범용성으로 인해 실시간 애플리케이션에서의 사용 가능성이 충분함을 보였다. 향후 개선 방향으로는 더 다양한 표정 레이블과 얼굴 이미지를 포함하는 대규모 데이터 셋을 사용하여 다양한 환경 조건(조명, 거리, 각도 등)에서의 성

능 향상을 위한 추가 연구가 필요하고, 학습에 사용한 얼굴 이미지가 아닌 처음 보는 얼굴에 대해서도 감정 예측을 잘 수행할 수 있도록 하는 범용성을 확보하는 방안에 대한 연구도 필요하다. 또한 모델의 복잡도와 추론 속도를 최적화하여 다양한 기기에서 사용할 수 있도록 개선할 필요가 있다.

참고문헌

[1] TensorFlow.js, https://www.tensorflow.org/js

[2] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Computing Research Reposit- ory (CoRR), Apr. 2017.

[3] S. Cai, S. Bileschi, E. Nielsen, "Deep Learning with JavaScript," Manning Publications, 2019.

[4] R. Yamashita, M. Nishio, R. K. G. Do, K. Togashi, "Convolutional Neural Networks: an Overview and Application in Radiology," Insights Imaging, Vol. 9, pp. 611-629, 2018.

[5] M. Kaur, A. Mohra, "A Review of Deep Learning with Recurrent Neural Network," International Conference on Smart Systems and Inventive Technology(ICSSIT), IEEE, India, 29-29 Nov. 2019.

[6] Z. Takac, M. Ferrero-Jaurrieta, L. Horanska, N. Krivonakova, G. P. Dimuro, H. Bustince, " Enhancing LSTM for sequential image classification by modifying data aggregation," International Conference on Electrical, Computer and Energy Technologies(ICECET), IEEE, Cape Town, South Africa, 09-10 Dec. 2021.

[7] J. Deng, W. Dong, R. Cocher, L.-J. Li, K. Li, L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20-25 June 2009.

■ 저자소개 ■

차 주 호
(Cha Jooho)

2009년 3월~현재
청운대학교 공과대학
멀티미디어학과 교수
2020년 3월~2021년 2월
Auckland University of
Technology 방문교수
1997년 7월~2000년 2월
대우통신 종합연구소 선임연구원
2004년 2월 광운대학교 컴퓨터과학과
(공학박사)

관심분야 : 네트워크 관리, 차량통신 네트워크, 시맨틱웹, 머신러닝
E-mail : jhcha@chungwoon.ac.kr

논문접수일 : 2023년 6월 30일
게재확정일 : 2023년 8월 16일