

CNN 기반의 이미지 특징과 객체 정보를 결합한 내용 기반 이미지 검색 기법

이기원*, 맹주형**, 송석일***

Content based Image Retrieval Method that Combining CNN based Image Features and Object Recognition Information

Gi-Won Lee*, Ju-Hyeong Maeng**, and Seokil Song***

2020년 한국교통대학교 지원을 받아 수행하였음. 또한, 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로
정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2014-3-00123, 실시간 대규모 영상 데이터 이해·예측을 위한
고성능 비주얼 디스커버리 플랫폼 개발)

요 약

컨텐츠 기반 이미지 검색 방법은 이미지 데이터 집합 중 질의 이미지와 유사한 이미지를 찾는 것이다. 최근 CNN(Convolution Neural Network)과 같은 심층 신경망을 이용하여 이미지의 특징을 추출하고 이를 이용하여 유사 이미지를 검색하는 방법이 제안되었다. 이 방법들은 이미지 전체의 특징을 이용한 검색은 가능하지만, 이미지 내의 특정 객체 정보를 이용한 이미지 검색은 어렵다. 이 논문에서는 CNN 기반 이미지 특징과 객체인식 기법을 이용하여 이미지에서 인식한 객체 정보를 결합하는 새로운 이미지 특징 추출 기법을 제안한다. 또한, 이 논문에서는 DCNN기반 특징과 객체 특징을 결합한 이미지의 특징을 위해서 RBF(Radial Basis Function)과 Jaccard 거리를 결합한 새로운 거리 척도를 제안한다. 제안하는 방법의 우수성을 입증하기 위해서 실험을 통해 제안하는 방법이 CNN 미지 특징만을 이용하는 방법에 비해 정확도가 향상됨을 보인다.

Abstract

Contents-based image retrieval (CBIR) method is to find an similar image to the query image among the image database. Recently, the methods extracting the image features using a deep neural network such as the Convolution Neural Network (CNN) have been proposed. These methods can retrieve similar images by comparing the global deep features of images, but it is difficult to retrieval images using specific object information in the image. In this paper, we proposes a new image feature extraction method that combines the CNN based image features and recognized objects' information. In addition, a new similarity metric combining RBF(Radial Basis Function) and Jaccard Distance for the proposed image feature extraction method. In order to show the superiority of the proposed method, we perform experiments comparing the proposed method and the CNN based CBIR method. Through the experiments, it is shown that the accuracy of the proposed method is more than twice as much as CNN based CBIR method.

Keywords

image retrieval, CNN, object recognition, image feature

* 한국교통대학교 컴퓨터공학과 학사
- ORCID: <https://orcid.org/0000-0002-8277-2690>
** 한국교통대학교 컴퓨터공학과 석사과정
- ORCID: <https://orcid.org/0000-0001-5152-3213>
*** 한국교통대학교 컴퓨터공학과 교수 (교신저자)
- ORCID ID: <https://orcid.org/0000-0002-0110-7155>

• Received: Apr. 30, 2021, Revised: May 25, 2022, Accepted: May 28, 2022
• Corresponding Author: Seokil Song
School of Computer, Korea National University of Transportation,
Daehakro 50, Chungju, Chungbuk 27469, Korea
Tel.: +82-43-841-5349, Email: sisong@ut.ac.kr

1. 서 론

내용 기반 이미지 검색(Content based image retrieval)은 이미지 데이터 베이스에서 주어진 질의 이미지와 가장 유사한 이미지를 검색하는 것이다 [1]. 유사한 이미지를 검색하는 것은 데이터베이스의 이미지들을 질의 이미지와의 유사도를 기반으로 순위를 부여하고 순위가 높은 이미지들을 검색한다. 따라서, 유사도에 기반한 이미지 검색은 이미지들간의 유사도 계산에 의존적이다. 이미지의 유사도는 각 이미지를 나타내는 특징을 기반으로 유사도 척도를 이용하여 계산된다.

기존의 이미지 특징 추출 방법은 심층 컨볼루션 신경망(DCNN, Deep Convolutional Neural Network) 기술이 확산되기 전에는 SIFT(Scale-Invariant Feature Transform)[2]와 같은 수작업 기반의 방법들이 대부분이었다. 2010년경에 DCNN 기술이 발전되고 다양한 분야에 적용되기 시작하면서 이미지 특징 추출에도 DCNN 기술이 적용된 방법들이 다수 제안되었다[3]-[5].

DCNN기반의 이미지 특징 추출 방법들은 CNN의 완전 연결 계층(Fully connected layer) 출력을 특징 벡터로 사용하거나 SPoC[6], MAC[7], RMAC[7] 등과 같이 활성화 맵(Activation map)에 대한 다양한 풀링(Pooling)을 수행해서 출력을 특징 벡터로 사용하는 방법, 지역 특징 벡터를 합성하여 전역 특징 벡터로 사용하는 방법 등 매우 다양하다. 문헌 [8]에서는 DCNN기반의 이미지 특징(Global feature)과 이미지 상의 객체를 인식한 후 각 객체의 영역의 특징(Local Feature)을 추출하여 결합하는 방법을 제

안하고 있다.

이상 언급한 DCNN기반 이미지 특징 추출 방법들은 이미지에서 추출된 객체들의 클래스를 고려하지 않는다. [8]에서 이미지에서 인식한 각 객체의 영역에 대한 특징을 추출하고 있지만, 각 객체의 클래스 정보를 이용하지는 않는다.

그림 1은 질의로 “테니스 치는 사람” 이미지로 주어졌을 때 COCO 테스트 데이터 집합[9]에서 DCNN기반 특징 추출 기법을 이용하여 유사 이미지 검색을 한 결과를 보여준다. 질의 결과 이미지로 부터 볼 수 있듯이 잔디밭에서 손에 운동기구를 들고 운동하는 사람 이미지들이 결과로 검색되었다. 이 예에서, 이미지의 테니스 라켓 객체를 인식하고 테니스 라켓 객체 정보를 이미지 특징에 반영할 수 있다면 이미지 검색 결과의 정확도를 보다 더 높일 수 있을 것이다.

이 논문에서는 DCNN 기반의 이미지 전체에 대한 특징과 객체 인식 기법[10, 11]을 이용하여 이미지에서 인식한 객체들을 결합한 특징 생성 기법을 제안한다. 또한, 이 논문에서는 DCNN기반 특징과 객체 특징을 결합한 이미지의 특징에 대한 새로운 거리 척도를 제안한다. 제안하는 거리 척도는 DCNN 기반의 특징을 위한 RBF(Radial Basis Function) 기반의 유사도와 객체들을 위해 Jaccard 거리 기반 유사도를 가중치와 함께 결합한다.

이 논문의 구성은 다음과 같다. 2장에서는 제안하는 특징 추출 방법과 유사도 척도에 대해서 기술한다. 3장에서는 제안하는 방법과 DCNN기반의 특징 추출 방법과의 성능을 비교하고 4장에서 결론을 맺는다.

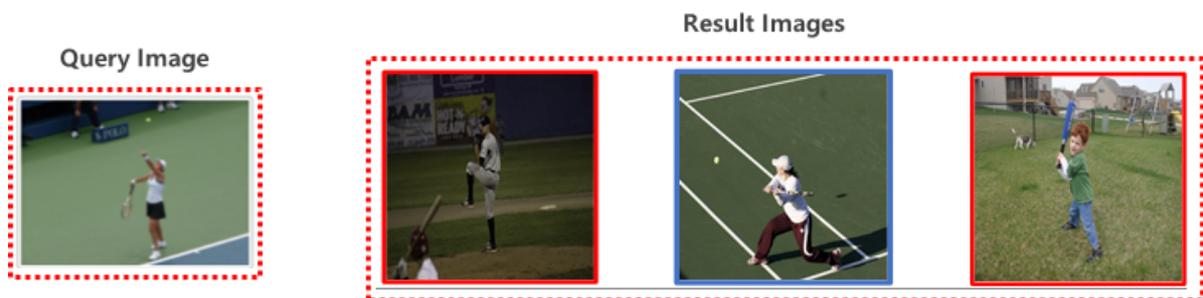


그림 1. COCO 테스트 데이터 집합에서 질의 이미지에 대한 검색 결과
Fig. 1. Retrieval result of query image from COCO test data set

II. 제안하는 특징 추출 방법 및 유사도 척도

앞서 기술한 바와 같이 이 논문에서는 이미지의 객체를 인식한 결과를 DCNN으로 추출한 이미지 특징에 결합하여 검색 정확도를 높이는 이미지 특징 벡터를 생성하는 방법을 제안한다. 이미지에 대한 특징 벡터를 생성하는 전체적인 과정은 그림 2와 같다. 그림에서처럼 VGG16[12]을 이용하여 이미지의 특징을 추출하고 Yolo v4[10]를 이용하여 이미지의 객체를 인식하여 객체 정보를 추출한다.

VGG16을 이용해서 이미지의 특징을 추출할 때는 사전에 학습된(Pretrained) 모델을 이용하였으며, VGG16의 완전 연결 계층 (1×1×4096)의 출력을 그대로 사용하였다. 제안하는 특징 추출 방법에서 이미지에 대한 DCNN기반 특징 추출은 다른 어떤 방법으로도 대체 가능하다.

또한, 이미지의 객체 정보는 대표적인 객체 인식 모델인 Yolo v4를 이용하여 주요 객체 정보를 추출한다. 추출한 객체 정보는 각 객체의 위치 및 클래스이다. 종합적으로 이 논문에서 제안하는 이미지 특징 (*feature*)는 수식 1에서와 같이 DCNN 이미지 특징 (*feature_{DCNN}*)과 객체 정보(*objectlist*)로 표현된다. 다시 객체 정보(*objectlist*)는 객체 위치(*objloc*)과 객체 클래스(*objclass*)로 표현된다.

$$feature = (feature_{DCNN}, objectlist(objloc, objclass)) \quad (1)$$

새롭게 제안하는 이미지 특징 (*feature*)을 이용해 이미지에 대한 유사도 검색을 위해서는 제안하는 이미지 특징을 고려한 유사도 척도가 필요하다. 이 논문에서는 *feature_{DCNN}*들간의 유사도 측정을 위해 RBF기반의 유사도 척도를 이용한다. 또한, 이미지의 *objectlist*들 간의 유사도 측정을 위해서 Jaccard 거리를 이용한다. Jaccard 거리는 수식 2에서와 같이 두 *objectlist*(*objectlist1*, *objectlist2*)가 주어졌을 때 각 *objectlist*에 포함된 동일한 *objclass*를 갖는 객체들의 비율로 계산된다.

$$J(objectlist_1, objectlist_2) = \frac{|objectlist_1 \cap objectlist_2|}{|objectlist_1 \cup objectlist_2|} \quad (2)$$

주어진 두 *feature* (*f1*, *f2*) 간의 유사도(*dist*)는 수식 3에 의해 계산된다. *dist*는 각 *feature*의 *feature_{DCNN}*간의 유사도를 RBF를 이용해서 계산하고 다시 각 *feature*의 *objectlist* 간의 유사도를 Jaccard 거리를 기반으로 계산한다. *dist*는 *feature_{DCNN}*거리와 *objectlist*거리에 각각 가중치를 부여하고 합산하여 계산된다.

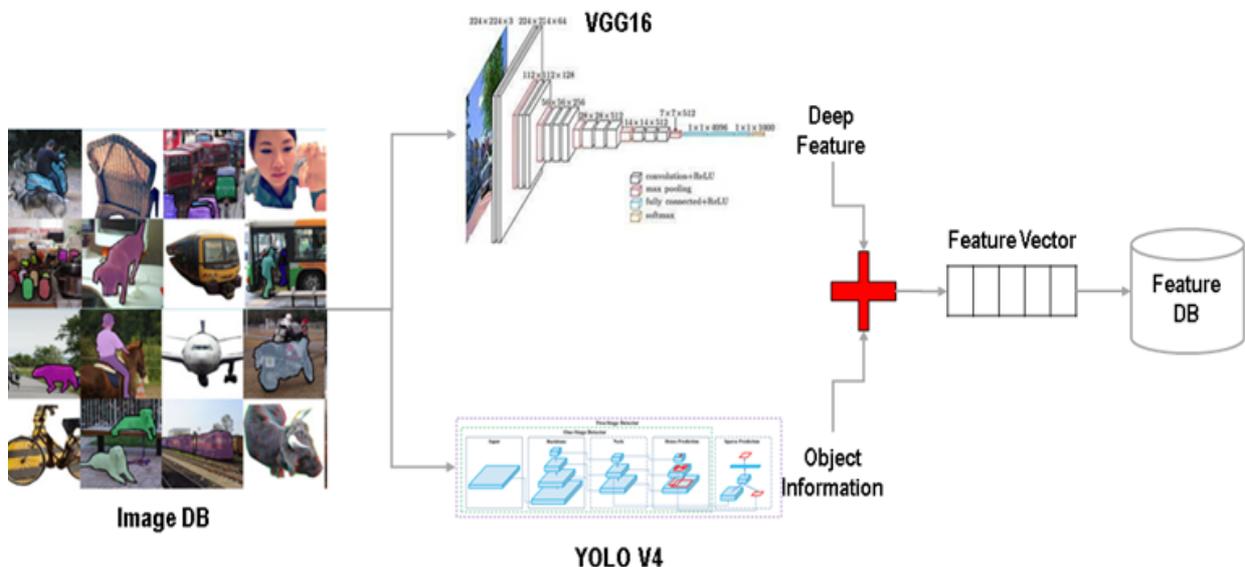


그림 2. 이미지에 대한 특징벡터 추출 과정
Fig. 2. Process of extracting features from images

$$dist(f1, f2) = 1 - \left\{ \begin{array}{l} RBF(f1.feature_{DCNN}, f2.feature_{DCNN}) \\ \times (1-w) + \\ Jaccard(f1.objectlist, f2.objectlist) \times w \end{array} \right\} \quad (3)$$

가중치에 따라 이미지의 유사도 비교 시 객체 정보와 DCNN 기반 이미지 특징 중 어디에 더 중심을 둘지가 결정할 수 있다. 응용 영역에 따라서 인식 가능한 클래스 수가 많고, 인식 정확도가 높은 경우에는 객체 정보에 가중치를 더 주는 것이 이미지 검색의 정확도를 더욱더 높일 수 있다. 이 논문의 실험에서는 가중치를 0.5로 하였다.

III. 성능평가

이 논문에서는 제안하는 방법이 DCNN기반 이미지 특징만을 이용한 유사 이미지 검색에 비해 검색 정확도가 더 높음을 보이기 위해 제안하는 방법을 구현하고 COCO 데이터 집합을 이용하여 이미지 검색 실험을 수행하였다. COCO 데이터 집합은 크게 이미지와 어노테이션(Annotation) 집합으로 구성된다. 어노테이션은 객체들 각각의 영역, 클래스 등과 같은 정보들을 담고 있다. COCO 데이터 집합의 클래스는 80개이며, 2017년 COCO test 데이터 기준으로 학습(Train) 데이터 118k 건, 검증(Validation) 5k건, 테스트(Test) 41k 건으로 구성되어있다. 표 1은 이 논문의 실험 환경을 보여준다.

표 1. 실험 환경
Table 1. Experimental environments

Division	Content
Hardware	CPU Intel i5-11400 2.60GHZ 6 Cores, GPU GeForce RTX 2060 Memory DDR4, 16GB
Software	CUDA 9.0, Python 3.6, Tensorflow 1.12.0
Data set	COCO 2017 test data set, 41K images

실험은 다음과 같이 진행되었다. 총 41k 건의 이미지 데이터 집합에 대해서 80건의 10-NN(Nearest Neighbor) 질의를 수행하였다. 80건의 10-NN 질의 이미지는 41K건의 이미지 데이터 중에서 200건의 이미지를 랜덤(Random)하게 선정하고 선정된 200건

의 이미지 중에서 되도록 서로 다른 클래스의 객체들로 구성되는 80건의 이미지를 질의 이미지로 선정하였다.

80건의 질의 이미지 각각에 대해서 DCNN 기반의 이미지 특징만을 이용한 10-NN 질의를 수행한 결과와 제안하는 특징 추출 방법을 이용한 10-NN 질의를 수행한 결과를 비교하였다. 이 실험에서 이용한 DCNN 기반 특징 추출 방법은 [13]에서 제안한 방법이다. [13]의 특징 추출 방법과 비교한 이유는 DCNN 기반 특징 추출 방법에 객체정보를 결합하였을 때 성능향상 정도를 확인하기 위해서 이다.

10개의 최근접 이미지를 검색하고 검색된 10개의 이미지들 중 질의 이미지와 실제로 유사한 이미지가 몇 개가 포함되었는지를 계산하였다. 실제 유사한 이미지가 포함되었는지는 육안으로 직접 확인하였다. 표 2는 80건의 10-NN 질의 각각에 대해 질의 처리 결과에 대한 평균 정확도를 보여준다. DCNN 기반 특징을 이용한 유사도 검색의 평균 정확도는 39.5%인 반면에 제안하는 방법(Proposed method)의 정확도는 89.13%로 약 2배 이상 정확도 향상이 있었음을 확인하였다.

표 2. 80건의 10-NN 질의 처리 결과에 대한 평균 정확도 비교

Table 2. Average accuracy comparison of 80 10-NN queries' results

Division	Proposed method	DCNN feature
Average Accuracy	89.13%	39.5%

그림 3은 3건의 10-NN 질의에 대한 제안하는 방법과 DCNN 기반 이미지 특징 추출 방법의 질의 처리 결과를 질의 이미지와 실제 검색된 이미지를 가지고 비교한 결과를 보여준다.

1번 질의는 테니스 라켓, 공, 사람 객체가 포함된 이미지이며 배경은 녹색이다. DCNN 이미지 특징만을 이용한 검색 결과에는 5번과 같이 질의 이미지와 전혀 관계 없는 어두운 배경에 컴퓨터를 사용하는 사람 이미지와 6번과 같이 야구하는 사람 이미지 등이 포함되어 있다. 제안하는 방법으로 검색한 10-NN 이미지에는 모두 테니스 라켓, 공, 사람 객체가 포함되어 있다. 다만, 5번, 6번 결과 이미지를 보면 배경이 황토색임을 볼 수 있다.

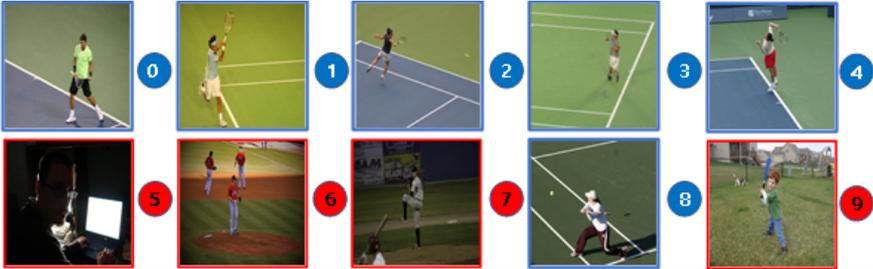
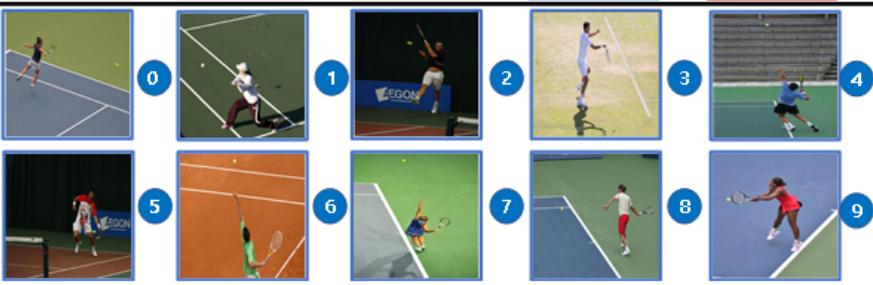
<p>Query 1 Person playing tennis</p> 	<p>Results of DCNN based image feature extracting method</p>	
	<p>Results of proposed method</p>	
<p>Query 2 Stop sign</p> 	<p>Results of DCNN based image feature extracting method</p>	
	<p>Results of proposed method</p>	
<p>Query 3 Person riding a motor cycle</p> 	<p>Results of DCNN based image feature extracting method</p>	
	<p>Results of proposed method</p>	

그림 3. 이미지 질의에 대한 결과 (빨간색 원 : 불일치, 파란색 원 : 일치)
Fig. 3. Results for image queries (Red Circle : non-match, Blue circle : match)

2번 질의는 정지신호 (Stop Sign) 객체를 포함하는 이미지이다. 배경에는 나무와 도로가 포함되어 있다. DCNN 기반 이미지 특징만을 이용한 검색 결과 이미지 중에서 0번과 8번 이미지는 정지신호를 포함하지는 않지만 배경으로 도로와 나무를 포함하고 있음을 볼 수 있다. 제안하는 방법의 검색 결과 이미지에서는 모두 정지신호가 포함되어 있다. 하지만, 5번 결과 이미지의 경우 배경에 도로는 포함되어 있지 않다.

3번 질의는 오토바이와 사람 객체를 포함하며 배경은 도로인 이미지이다. DCNN 기반 이미지 특징만을 이용한 검색 결과에서는 오토바이와 사람 객체를 모두 포함하는 이미지는 3번과 6번뿐이다. 나머지 이미지들은 공통적으로 녹색의 객체를 포함하고 있다. 반면에 제안하는 방법으로 검색된 결과 이미지들에는 모두 오토바이 객체를 포함하고 있으며 0번을 제외한 나머지는 모두 사람 객체도 같이 포함하고 있다. 또한 질의 이미지와 유사하게 도로를 배경으로 하거나 오토바이 색이 녹색인 이미지들이 대부분이다.

IV. 결 론

이 논문에서는 내용 기반 이미지 검색을 위한 새로운 이미지 특징 추출 방법을 제안한다. 제안하는 이미지 특징 추출 방법은 기존에 제안된 DCNN 기반의 이미지 특징 추출 방법에 Yolo 등의 객체 인식 기법을 이용해 추출한 이미지의 객체 정보를 결합하여 특징을 추출하는 방법을 제안하였다. 또한, 제안하는 특징 추출 방법과 함께 제안하는 특징들 간의 유사도를 비교할 수 있도록 Jaccard 거리와 RBF 기반의 유사도를 결합하는 새로운 유사도 비교 방법을 제안하였다.

제안하는 방법이 보다 정확하게 질의 이미지와 유사한 이미지를 검색할 수 있음을 보이기 위해 제안하는 방법과 DCNN 기반 이미지 특징 추출방법을 구현하고 질의 처리 실험을 진행하였다. 41K건의 COCO 테스트 데이터 집합에 대해 80개의 유사도 질의를 수행하여 제안하는 방법과 DCNN기반의 이미지 특징만을 이용한 검색 결과와 비교하였다.

비교 결과 제안하는 방법의 정확도가 89.13%로 DCNN기반 이미지 특징만을 이용한 검색 결과에 비해 약 2배 이상의 정확도 향상을 확인하였다.

향후 연구에서는 객체 정보의 영역을 이용하여 객체간의 상대적인 공간 관계를 추가하는 연구를 진행할 계획이다. 또한, 다양한 DCNN 기반 이미지 특징 추출방법과 결합하는 실험을 진행하여 제안 방법이 다른 DCNN 기반 이미지 특징 추출방법의 성능개선에 기여 할 수 있음을 보인다.

References

- [1] Deselaers, T., Keysers, D. and Ney, H., "Features for image retrieval: an experimental comparison", *Information retrieval*, Vol. 11, No. 2, pp. 77-107, 2008.
- [2] Tang, G., Liu, Z., and Xiong, J., "Distinctive image features from illumination and scale invariant keypoints", *Multimedia Tools and Applications*, Vol. 78, No. 16, pp. 23415-23442, 2019.
- [3] Kalantidis, Y., Mellina, C., and Osindero, S., "Cross-dimensional weighting for aggregated deep convolutional features", *European conference on computer vision*, pp. 685-701, 2016.
- [4] Wan, J., Wang, D., Hoi, S. C. H., Wu, P., Zhu, J., Zhang, Y., and Li, J., "Deep learning for content-based image retrieval: A comprehensive study", *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 157-166, 2014.
- [5] Alzu'bi, A., Amira, A., and Ramzan, N., "Semantic content-based image retrieval: A comprehensive study", *Journal of Visual Communication and Image Representation*, Vol. 32, pp. 20-54, 2015.
- [6] Babenko, A., and Lempitsky, V., "Aggregating local deep features for image retrieval", *Proceedings of the IEEE international conference on computer vision*, pp. 1269-1277, 2015.
- [7] Tolias, G., Sivic, R., and Jégou, H., "Particular

object retrieval with 'integral max-pooling of CNN activations", ICLR, pp. 1-12, 2015.

- [8] Sun, S., Zhou, W., Li, H., & Tian, Q., "Search by detection: Object-level feature for image retrieval", Proceedings of international conference on internet multimedia computing and service, pp. 46-49, 2014.
- [9] MS COCO Dataset: <https://cocodataset.org>
- [10] Jiang, P., Ergu, D., Liu, F., Cai, Y., and Ma, B., "A Review of Yolo Algorithm Developments", Procedia Computer Science, pp. 1066-1073, 2022
- [11] He, K., Gkioxari, G., Dollár, P., and Girshick, R., "Mask r-cnn", Proceedings of the IEEE international conference on computer vision, pp. 2961-2969, 2017.
- [12] Qassim, H., Verma, A., and Feinzimer, D., "Compressed residual-VGG16 CNN model for big data places image recognition", 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), pp. 169-175, 2018.
- [13] Babenko, A., Slesarev, A., Chigorin, A., and Lempitsky, V., "Neural codes for image retrieval", Proceedings of European conference on computer vision, pp. 584-599, 2014.

저자소개

이 기 원 (Ki-Won Lee)



2020년 2월 : 한국교통대학교
컴퓨터공학과(공학사)
관심분야 : 인공지능,
데이터베이스, 컴퓨터 비전,
이동객체 등

맹 주 형 (Ju-Hyeong Maeng)



2020년 2월 : 한국교통대학교
컴퓨터공학과(공학사)
2020년 9월 ~ 2022년 8월 :
한국교통대학교 컴퓨터공학과
(공학석사)
관심분야 : 인공지능,
데이터베이스, 컴퓨터 비전,
이동객체 등

송 석 일 (Seokil Song)



1998년 2월 : 충북대학교
정보통신공학과(공학사)
2000년 2월 : 충북대학교
정보통신공학과(공학석사)
2003년 2월 : 충북대학교
정보통신공학과(공학박사)
2003년 7월 ~ 현재 :
한국교통대학교 컴퓨터공학과 교수
관심분야 : 데이터베이스, 센서 네트워크, 스토리지
시스템 등