

이미지 검색기를 통한 랜드마크 인식

김기덕^o, 이근후^{*}

^o(주)쓰리아이퓨처,

^{*}(주)쓰리아이퓨처

e-mail: lghoo@naver.com^{*}, kimsjpk@naver.com^o

Landmark recognition through image searcher

Gi-Duk Kim^o, Geun-Hoo Lee^{*}

^o3IFuture,

^{*}3IFuture

● 요약 ●

본 논문에서는 이미지 검색기를 통한 랜드마크 인식 방법을 제안한다. 특정 랜드마크 데이터셋에서 라벨링을 하지 않은 비지도 학습을 통해서 이미지에서 랜드마크의 클래스 분류를 위한 특징을 추출한다. 학습된 모델을 랜드마크 데이터셋인 Paris6k 데이터셋과 Oxford5k 데이터셋에 적용하여 랜드마크 인식 정확도를 확인하였다. 성능과 속도를 강화하기 위해 이미지 특징 추출 모델로 ResNet 대신에 YOLO에서 사용된 CSPDarknet-53을 사용하여 모델의 크기를 줄이고 랜드마크 인식 정확도를 높였다. 그리고 모델로부터 추출된 특징의 수를 줄여 이미지 검색 시 소요되는 시간을 감소시켰다. 학습된 모델로 rOxford5k 데이터셋에 적용 시 mAP 80.37, rParis6k에서 mAP 89.07을 얻었다.

키워드: 랜드마크 인식(landmark recognition), 이미지 검색기(image searcher), 비지도 학습(unsupervised learning)

I. Introduction

CNN(convolutional neural network)를 통한 이미지 인식에서 높은 성능을 거둔 이후로 CNN에서 추출한 특징을 통하여 이미지 검색에 CNN을 통한 방법이 적용되었다. Babenko 등[1]은 사전 학습된 모델을 통한 이미지 특징 추출 방법을 적용한 이미지 검색을 제안하였다. 또한 Arandjelovic 등[2]은 사전 학습된 모델에서 추출한 특징과 본 논문에서 적용한 특징 데이터베이스에서 특징 매칭을 통해 특징의 유사도에 따라 매칭된 이미지와 비매칭된 이미지를 구분하고 이를 학습데이터로 사용한 방법을 제안하였다. 본 논문에서는 Radenovic 등[3]이 제안한 방법인 SfM(structure-form-motion) 데이터셋에 클러스터링을 사용하여 유사한 이미지와 유사하지 않은 이미지를 구분하여 라벨링 작업 없이 이미지의 특징을 추출하는 방법을 적용하였다. 그리고 학습된 모델을 rOxford5k 데이터셋[4]과 rParis6k 데이터셋[5]에 적용하였다. Radenovic 등 은 학습 성능을 높이기 위해 Local whitening, Regional whitening과 GeM(Generalized-Mean) Pooling을 적용하였다.

본 논문에서는 Radenovic 등 의 모델에서 이미지 특징 추출 모델을 ResNet [6] 대신에 YOLO [7] 모델에 적용된 CSPDarknet-53을 적용하였고 이를 통해 학습 모델의 크기를 560메가에서 350메가로

줄였다. 그리고 한 이미지에서 특징 추출에 걸리는 시간을 0.016초에서 0.0099초로 감소시켰고 추출된 특징의 수를 2048개에서 1024개로 줄임으로써 특징 데이터베이스의 크기를 줄이고 검색시간 또한 한 쿼리당 0.026초에서 0.022초로 줄일 수 있었다. 그리고 rOxford5k에서 mAP(mean Average Precision)값이 79.34에서 80.37로 성능이 향상됨을 확인하였다.

II. Preliminaries

1. Related works

CNN 기반 이미지 특징 추출 후 이미지 검색에 적용된 방법으로 지역 특징 기반 방법인 Fisher vectors [8], VLAD [9]이 연구되었고 추출된 특징을 가공하는 방법으로 visual codebooks [10], spatial verification [11], query expansion [12] 방법이 연구되었다. 그리고 이미지 검색 네트워크로 two-branch siamese [13], triplet networks [14] 방법 등이 연구되었다. 이 외에도 Radenovic 등 은 풀링 방법과 descriptor whitening 방법을 제안하였다.

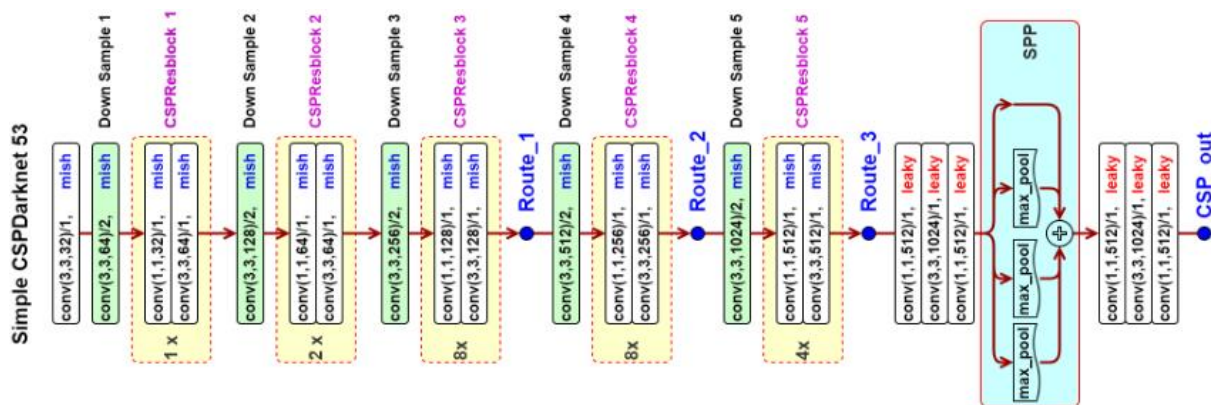


Fig. 1. CSPDarknet-53 구조

III. The Proposed Scheme

Radenovic 등 이 제안한 방법을 사용하여 클러스터링을 통해 SfM 데이터셋에서 클러스터링을 통해 별도의 라벨링 작업 없이 데이터를 학습하였다. 학습의 성능을 향상하기 위해 이미지 특징 추출 모델을 ResNet에서 CSPDarknet-53 모델을 사용하였다. CSPDarknet-53 구조는 그림 1과 같다.

Radenovic 등 은 ResNet을 통해 추출된 특징에 Local, Regional Whitening을 적용하였으나 CSPDarknet-53에 추출한 특징에 Local, Regional Whitening을 적용 시 오히려 성능이 감소하는 결과를 얻었다. 아마도 CSPDarknet-53에서 얻은 특징과 조합이 맞지 않은 것으로 추측한다. 그리고 ResNet에 Local, Regional Whitening을 적용하지 않았을 시 학습할 때 Loss값이 발산하여 성능이 감소하는 결과를 얻었다. CSPDarknet-53과 GeM Pooling을 적용한 모델은 서론에서 언급한 대로 학습 모델의 크기를 감소시키고 랜드마크 데이터셋인 rOxford5k 데이터셋에 대한 mAP 값을 증가시킬 수 있었다. 언급한 랜드마크 데이터셋뿐만 아니라 부산대에서 제공하는 랜드마크 데이터셋(https://github.com/PNUCSE/2022_DLC_LandmarkClassification)에 랜드마크 인식 정확도를 구했을 때 쿼리 이미지 15장 중에 14장의 클래스 일치 결과를 얻을 수 있었다. 그림 2는 부산대 랜드마크 데이터셋 중 일부이다.



Fig. 2. 부산대 랜드마크 이미지

학습된 모델을 가지고 전체 데이터셋의 특징을 추출한 후 특징 데이터셋을 만들고 쿼리 이미지에서 추출한 특징과 비교하여 가장 가까운 특징의 이미지 목록 5장을 뽑았다. 5장의 클래스 중 가장 많은 클래스 번호의 수와 쿼리 이미지의 클래스 번호를 비교해서 일치한 지를 검사하였다. 이를 통해서 학습한 모델이 다른 데이터셋

에 적용 가능함을 확인하였다.

IV. Conclusions

본 논문에서 이미지 검색기를 통한 랜드마크 인식 방법을 제안하였다. 이미지 검색기의 성능을 향상하기 위해 이미지 특징 추출 모델을 CSPDarknet-53을 적용하였고 이를 통해 서론에서 언급한 대로 모델의 크기를 줄이고 분류 성능을 높일 수 있었다. 그리고 부산대 랜드마크 데이터셋에 적용하여 결과를 얻음으로써 다른 랜드마크 인식 데이터셋에도 적용 가능함을 보였다. Radenovic 등 이 제안한 모델뿐만 아니라 다른 이미지 검색 모델 내부의 이미지 추출 모델에 본 논문의 방법을 적용하면 더욱 성능을 높일 수 있을 것이다. 논문에 사용된 코드는 블로그에 올려두었다. (<https://blog.naver.com/kimsjpk/223077122709>)

REFERENCES

- [1] BABENKO, Artem, et al. Neural codes for image retrieval. In: Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13. Springer International Publishing, 2014. p. 584-599.
- [2] ARANDJELOVIC, Relja, et al. NetVLAD: CNN architecture for weakly supervised place recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 5297-5307.
- [3] RADENOVIC, Filip; TOLIAS, Giorgos; CHUM, Ondrej. Fine-tuning CNN image retrieval with no human annotation. IEEE transactions on pattern analysis and machine intelligence, 2018, 41.7: 1655-1668.
- [4] PHILBIN, James, et al. Object retrieval with large vocabularies and fast spatial matching. In: 2007 IEEE

- conference on computer vision and pattern recognition. IEEE, 2007. p. 1-8.
- [5] PHILBIN, James, et al. Lost in quantization: Improving particular object retrieval in large scale image databases. In: 2008 IEEE conference on computer vision and pattern recognition. IEEE, 2008. p. 1-8.
- [6] HE, Kaiming, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770-778.
- [7] REDMON, Joseph; FARHADI, Ali. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.
- [8] PERRONNIN, Florent, et al. Large-scale image retrieval with compressed fisher vectors. In: 2010 IEEE computer society conference on computer vision and pattern recognition. IEEE, 2010. p. 3384-3391.
- [9] JÉGOU, Hervé, et al. Aggregating local descriptors into a compact image representation. In: 2010 IEEE computer society conference on computer vision and pattern recognition. IEEE, 2010. p. 3304-3311.
- [10] PHILBIN, James, et al. Object retrieval with large vocabularies and fast spatial matching. In: 2007 IEEE conference on computer vision and pattern recognition. IEEE, 2007. p. 1-8.
- [11] SHEN, Xiaohui, et al. Spatially-constrained similarity measure for large-scale object retrieval. IEEE transactions on pattern analysis and machine intelligence, 2013, 36.6: 1229-1241.
- [12] CHUM, Ondřej, et al. Total recall II: Query expansion revisited. In: CVPR 2011. IEEE, 2011. p. 889-896.
- [13] CHOPRA, Sumit; HADSELL, Raia; LECUN, Yann. Learning a similarity metric discriminatively, with application to face verification. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). IEEE, 2005. p. 539-546.
- [14] WANG, Jiang, et al. Learning fine-grained image similarity with deep ranking. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014. p. 1386-1393.