

# 딥러닝 기반의 의류 속성을 고려한 패션 이미지 검색

임보영, 남종호

서강대학교 컴퓨터공학과

bylim@sogang.ac.kr, jhnang@sogang.ac.kr

## A Fashion Image Retrieval using Clothing Attributes based on Deep Learning

Boyoeng Lim, Jongho Nang

Department of Computer Science and Engineering, Sogang University

### 요 약

기존의 의류 이미지 검색 방법은 텍스트 기반의 키워드 검색을 주로 사용하였으며, 최근 출시된 의류 이미지 검색 서비스는 유사도 측정에 있어 정확성이 낮은 한계점을 가지고 있다. 이를 위해 본 논문에서는 이미지 검색의 정확도를 높이기 위하여 딥러닝 기반의 의류 속성을 고려한 패션 이미지 검색 알고리즘을 제안한다. 제안된 알고리즘은 전역적인 시각적 특성에 기반한 이미지 검색 모델의 검색 결과와 의류 속성을 고려한 이미지 분류의 결과를 결합하여 검색 결과를 재순위화 한다. 실험을 통해 이미지 검색 모델의 결과와 이미지 분류의 결과를 결합한 재순위화 된 결과를 각각 비교 분석한다. 질의 데이터 14,213장과 갤러리 데이터 12,608장의 패션 이미지를 이용한 실험 결과에 의하면 기존 성능에 비하여 약 4% 향상한 82.8%의 정확도로 유사 이미지를 검색할 수 있음을 알 수 있었다. 본 연구는 기존 이미지 검색의 한계점을 해결할 수 있으며, 상품 검색 서비스뿐 만 아니라 상품 추천 서비스와 같은 다양한 서비스에 활용될 것으로 기대된다.

## 1 서 론

의류, 패션 온라인 시장 규모는 매년 성장하고 있으며, 특히 온라인 쇼핑 증가에 따라 패션 시장에서는 온라인 상에서 원하는 상품을 찾을 수 있는 의류 검색 서비스, 의류 추천 서비스 등 다양한 서비스 개발과 연구들이 진행되고 있다. 하지만 현재 대부분의 의류 검색 서비스는 이미지의 특성이 아닌, 텍스트 기반의 키워드로 검색이 이루어지며 의류 추천 서비스 또한 사용자의 이전 검색 기록이나 다수의 사용자가 선호하는 데이터만을 이용하여 추천해주는 방식을 주로 사용하고 있다. 또한 최근 의류 검색 서비스 [1,2,3]들이 출시되었지만, 검색 시 사용되는 유사도 측정에 있어서 의류 속성을 고려하지 않아 정확한 검색 결과가 나오지 않는다는 한계점을 가지고 있다.

본 논문에서는 이러한 한계점을 해결할 수 있는 딥러닝 기반의 의류 속성을 고려한 패션 이미지 검색 알고리즘을 제안한다. 이 과정에서 이미지를 구성하는 특징들을 기반으로 한 이미지 검색, 의류 속성과 이미지 특징을 고려한 이미지 분류, 이미지 검색 후 재순위화 기법을 수행한다. 질의 데이터 14,213장과 갤러리 데이터 12,608장의 패션 이미지 실험 결과에 의하면 기존 이미지 검색 정확도보다 높은 82.8%의 정확도로 유사 이미지를 검색할 수 있음을 알 수 있었다.

## 2. 관련 연구

### 2.1 이미지 검색에 관한 연구

내용 기반 이미지 검색이란 검색하고자 하는 이미지를 넣었을 때, 검색 결과인 유사한 이미지를 반환해주는 기술이다. 과거에는 전통적인 방법인 Bag-of-Features[4], large vocabularies[5] 등을 사용하였다. 최근에는 CNN을 이용한 이미지 검색 방법으로 R-MAC[6], NetVLAD[7] 방법을 사용하여 이미지 검색에 관한 연구가 진행되고 있다.

### 2.2 DeepFashion

DeepFashion[3]은 패션 이미지에 대한 데이터로 4개의 벤치마크로 분류되어 있다. 각각의 벤치마크는 매장에서 촬영된 이미지에서부터 소비자들 촬영한 이미지까지 80만장의 이미지들로 구성되어 있으며, 이미지들은 모델이 옷을 입고 촬영한 이미지에서부터 옷만 촬영된 이미지까지 다양하다.

## 3. 제안한 알고리즘

본 논문에서는 딥러닝 기반의 패션 이미지 검색 알고리즘을 제안한다. 알고리즘은 <그림 1>의 의사코드와 같이 이미지의 시각적 특징을 기반으로 한 이미지 검색, 의류 속성 기반의 이미지 분류, 검색 결과 기반 재순위화 3단계로 구성된다.

<b>Top-k Image Retrieval Algorithm</b> input: Image, Databaselmages, topk output: retrieval result
<b>Start of Algorithm</b>
<pre> // 패션 이미지 검색 1. feats = extractFeature(Databaselmages) 2. input_feat = extractFeature(image) 3. M = getNum(Databaselmages) 4. for each i &lt; M 5.   cosine_dist[i] = calculateDistance(feats[i], input_image_feat) 6. cosine_dist_result = sort(cosine_dist) 7. result_topk = cosine_dist_result [:topk]  // 의류 속성 기반의 이미지 분류 8. Query_attribute = getClassification(image) 9. for each j &lt; topk 10.  topk_classification[j] = getClassification(result_topk[j])  // 검색결과 기반의 재순위화 11. for each k &lt; topk 12.  Retrieval_score[k] = getScore(result_topk[k]) 13.  Classifier_score[k] = getScore(topk_classification[k]) 14.  Total_score[k] = Retrieval_score[k] + Classifier_score[k] 15. results = sort(Total_score) 16. return results </pre>
<b>End of Algorithm</b>

<그림 1> 제안하는 알고리즘의 의사코드

### 3.1 패션 이미지 검색 모델 학습

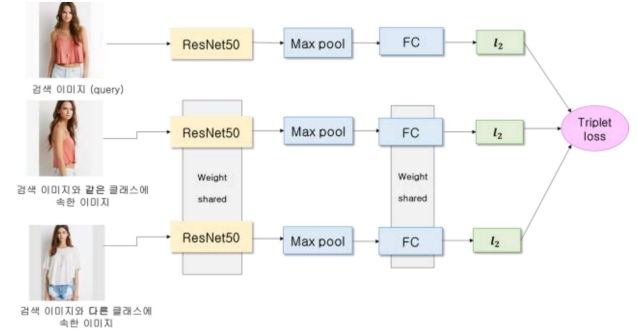
기존 FashionNet[3]의 경우 모델 구조가 복잡하여 학습을 시키기 어려운 단점이 있다. 본 연구에서는 보다 단순한 학습 모델인 <그림 2>의 삼(Siamese) 네트워크[8]를 사용하였으며, 입력으로 3개의 이미지인 ‘질의 이미지’, ‘같은 클래스 이미지’, ‘다른 클래스의 이미지’ 들을 입력 받게 된다. 같은 클래스의 이미지란 질의 이미지와 동일한 옷이 촬영된 이미지들이다. 다른 클래스의 이미지들은 질의 이미지와 다른 옷이 촬영된 이미지이며, 학습 시 질의 이미지와 다른 클래스 중 랜덤으로 추출하였다. 삼 네트워크[8]는 ResNet50[9] 모델과 max pool, fully-connected 레이어를 거쳐, 최종적으로 l2 정규화를 한 후 결과를 얻게 된다. 손실 함수는 triplet loss[8]를 사용하여, 질의 이미지와 같은 클래스인 이미지와의 거리는 작아지고, 다른 클래스의 이미지 사이의 거리가 멀어지도록 학습을 시켰다. 코드는 파이썬 기반의 Pytorch 프레임워크를 사용하여 구현하였다.

### 3.2 의류 속성 기반의 이미지 분류 모델 학습

의류의 속성에는 카테고리, 스타일, 패턴, 소재 등 463개의 속성들로 구성되어 있다. 각각의 이미지들은 의류 속성에 해당되면 1, 해당되지 않으면 -1의 값을 가지도록 지정되어 있다. 두 번째 단계인 이미지 분류는 <그림 3>의 학습 모델로 vgg16[10] 네트워크에 3개의 fully-connected 레이어가 더해져, 최종적으로 463개의 의류 속성 클래스에 대하여 분류된다.

### 3.3 검색결과 기반의 재순위화

이미지 검색 모델에서 나온 결과는 이미지의 전역적인 특징만을 고려하고, 세부적인 스타일이나 패턴 같은 지역적인 특징을 고려하지 않는다. 보다 높은 정확성을 얻기 위해 이미지 검색 모델의 결과 순위와 지역적인



<그림 2> 패션 이미지 검색 학습 모델[8]

특성인 의류 속성 기반의 이미지 분류 모델에서 나온 결과 순위를 융합하여 재순위화 하는 단계를 수행한다. 재순위화 하는 알고리즘은 두 결과 순위에 가중치를 각각 부여하여, 점수화 한 후 높은 점수 순으로 순위를 재순위화 한다.

## 4. 실험 및 결과

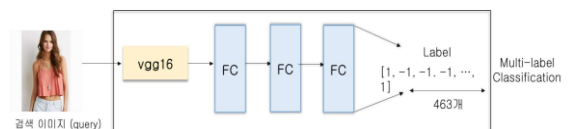
본 연구에서는 이미지 검색 모델을 위한 벤치마크인 ‘In-shop Clothes Retrieval’를 사용하였다. 학습 데이터로 25,866장, 463개의 클래스, 실험 데이터로 26,821장을 사용하였으며, 실험 데이터는 질의 데이터 14,213장과 갤러리 데이터 12,608장으로 분류하였다. 정확도 측정은 성능 비교 한 모델 WTB[1], DARN[2], FashionNet[3]에서 사용한 식을 사용하였으며, 식(1)을 따른다.

$$\frac{1}{|Q|} \sum_{q \in Q} \text{hit}(query, K) \quad (1)$$

식(1)의 Q는 전체 질의 이미지의 개수, Hit(query, K)=1은 검색 결과(K)에서 질의 이미지(q)의 클래스와 동일한 이미지가 있는 경우, 없는 경우는 Hit(query, K)=0이다.

### 4.1 패션 이미지 검색 모델 별 성능 측정

성능은 <그림 4>와 같이 검색 결과 이미지 수를 20개 뽑았을 때는 약 74.4%의 정확도를 나타냈으며, FashionNet[3]보다 2% 정도 낮았다. 하지만 top-k가 20보다 많을 경우 3개의 모델 중 가장 높은 성능을 보였으며, top-k가 50 일 때 82.8% 성능을 보였다.

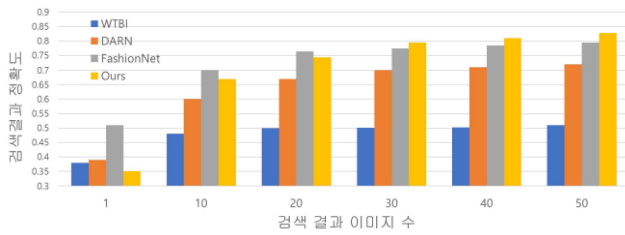


<그림 3> 의류 속성 기반의 이미지 분류 학습 모델

### 4.2 재순위화 한 모델 성능 측정

이미지 분류를 결합한 모델은 점수로 순위를 결정하였고, 점수 계산식은 다음과 같다.

## 참고문헌



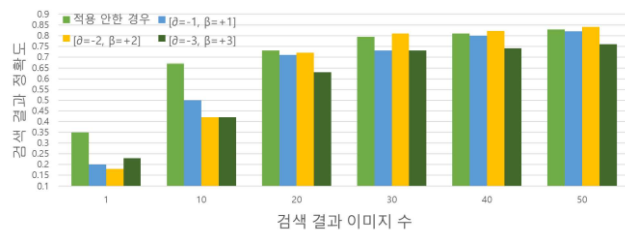
&lt;그림 4&gt; 패션 이미지 검색 모델 성능 비교

$$\sum_{i=1}^{topk} Retrieval\_score[i] = -Retrieval\_ranking[i] \quad (2)$$

$$\sum_{j=1}^{topk} Classifier\_score[j] = |Query\_attribute \cap gallery\_attribute[j]| \quad (3)$$

$$\sum_{k=1}^{topk} Total\_score[k] = \alpha \cdot Retrieval\_score[k] + \beta \cdot Classifier\_score[k] \quad (4)$$

실험 시 검색 점수는 식(2), 분류 점수는 식(3), 최종 점수는 식(4)로 정의하였으며, 가중치  $\alpha$ 와  $\beta$ 는 각각 -1부터 -3, +1부터 +3을 주었다. 그 결과 <그림 5>와 같이 [ $\alpha=-2$ ,  $\beta=+2$ ]를 주었을 때는 성능이 1.1% 정도 오른 것을 확인하였다. 이는 카테고리 분류에 비해 패턴, 스타일 분류의 정확성은 상대적으로 낮아, 성능이 검색 모델에 비해 낮게 오른 것으로 분석되었다.



&lt;그림 5&gt; 이미지 분류 결합한 모델 성능

## 5. 결론 및 추후 연구

이번 연구에서는 딥러닝 기반의 의류 속성을 고려한 패션 이미지 검색을 수행하는 연구를 수행하였다. 전역적인 시각적 특성을 고려한 이미지 검색과 지역적인 특성을 고려한 이미지 분류, 검색과 분류를 결합한 재순위화 3가지 단계의 검색 알고리즘을 제안하였다. 실험을 통해 전역적인 특성만을 고려한 이미지 검색의 경우, 기존 가장 높은 성능을 가진 모델의 정확도보다 약 4% 향상됨을 확인하였다. 또한 두 검색 결과를 융합하여 재순위화 한 결과 성능이 1.1%정도 오른 것을 확인하였다. 추후에는 더 많은 학습 이미지와 분류 정확도를 향상 시켜 검색 결과의 정확성을 더 높이는 방법에 대해 연구할 예정이다.

## 6. 감사의 글

이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No. 2017-0-00271, 아카이브 솔루션 및 콘텐츠 개발)

- [1] H. Chen, A. Gallagher, and B. Girod, "Describing Clothing by Semantic Attributes," in *Proceedings of European Conference on Computer Vision*, pp. 609–623, 2012.
- [2] J. Huang, R.S. Feris, Q. Chen, and S. Yan, "Cross-Domain Image Retrieval with a Dual Attribute-Aware Ranking Network," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 1062–1070, 2015.
- [3] Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang, "Deepfashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 1096–1104, 2016.
- [4] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual Categorization with Bags of Keypoints," in *Proceedings of Workshop on Statistical Learning in Computer Vision*, pp. 1–2, 2004.
- [5] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object Retrieval with Large Vocabularies and Fast Spatial Matching," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007.
- [6] G. Tolias, R. Sivic, and H. Jégou, "Particular Object Retrieval with Integral Max-Pooling of CNN Activations," *arXiv preprint arXiv:1511.05879*, 2015.
- [7] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "NetVLAD: CNN Architecture for Weakly Supervised Place Recognition," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 5297–5307, 2016.
- [8] A. Gordo, J. Almazán, J. Revaud, and D. Larlus, "Deep image retrieval: Learning global representations for image search," in *Proceedings of European Conference on Computer Vision*, pp. 241–257, 2016.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.
- [10] K. Simonyan, and A. Simonyan, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint arXiv:1409.1556*, 2014.