논문 20-45-05-06 The Journal of Korean Institute of Communications and Information Sciences '20-05 Vol.45 No.05
https://doi.org/10.7840/kics.2020.45.5.801

의류 인식을 위한 의류 랜드마크 특징 탐색 네트워크

이 수 민◆, 오 성 찬*, 정 찬 호°, 김 창 익**

Fashion Landmark-Driven Feature Exploitation Network for Clothing Recognition

Sumin Lee◆, Sungchan Oh*, Chanho Jung°, Changick Kim**

요 약

의류 영상은 non-rigid 성질을 가져서 알고리즘 적용이 힘들다. 이를 극복하여 Fabric과 Part의 분류 성능 향상을 위해서, 본 논문에서는 의류 랜드마크 특징 정보를 활용한 의류 인식 방법을 제안한다. 제안하는 방법은 의류의 문맥적 특징 정보를 활용하기 위해서 랜드마크 국소화 과정에서 생성되는 특징 정보를 사용한다. DeepFashion 데이터셋에 대한 실험을 통해 제안하는 방법을 분석하고 해당 방법의 우수성을 보여준다.

Key Words : clothing recognition, fashion landmark localization, global-local embedding module

ABSTRACT

Due to non-rigid deformations, it is hard to apply algorithms on fashion images. To overcome this problem and imporve performances of Fabric and Part attributes, in this paper, we propose a clothing recognition method with fashion landmark features. The proposed method exploits contextual features of a fashion item, which is generated in fashion landmark localization. Through experimental results on the DeepFashion dataset, we demonstrate that the proposed method has an excellent ability to learn deep feature representation for fashion recognition.

Ⅰ. 서 론

최근 온라인 쇼핑 시장이 커지면서, 대량의 의류 영상 데이터를 처리를 위한 의류 영상 분석에 대한 연구가 주목받고 있다. 주어진 의류 영상에 대해서 옷의 종류 (예: Tee, Shirt, Pants)와 특성 정보(예: V-neck, Denim, Abstract, Lace)를 예측하고 수행하는 의류 인식은 의류 영상 분석 분야 중 하나이다. (그림 1). 최근 대량의 의류 데이터셋이 등장하고 영상처리에서 딥러닝이 큰 성능 향상을 이끌면서, 딥러닝을 사용한 의류 인식 방법들이 제안되고 있다. 의류 영상은 일반 물체 영상과 달리 non-rigid 특성으로 영상에서 변형이 심하게 일어나 알고리즘 적용이 힘들다. 이 때문에 의류 랜드마크를 활용하여 의류의 시멘틱 정보를 이해하기 위한 방법들이[1,2,4] 제시되어 왔다. 하지만 기존 방법들은 국소화한 랜드마크 좌표만 사용할 뿐 국소화 과정에서 생성된 특징 정보를 활용하지 않는다. 랜드마크 국소화 과정에서 생성되는 특징 정보는 의류 영상의 문맥적인 정보를 포함하고 있다. 문맥적인 정보란 영상 속 상황에 따라 다양한 형태로 존재할 수 있는 의류를 적응적으로 이해하기 위한 정보이다. 예를 들어, 옷이 상의, 하의 혹은 전신 옷인지, 어떠한

그림 1. 의류 인식의 예시
Fig. 1. Example of fashion recognition

◆ First Author : (ORCID:0000-0003-1490-1355)School of Electrical Engineering, Korea Advanced Institute of Science and Technology, suminlee94@kaist.ac.kr, 학생(박사), 학생회원
° Corresponding Author : (ORCID:0000-0003-3145-6732) Department of Electrical Engineering, Hanbat National University, peterjung@hanbat.ac.kr, 부교수, 정회원
* (ORCID:0000-0003-3700-6492) Electronics and Telecommunications Research Institute. sungchan.oh@etri.re.kr, 선임연구원
** (ORCID:0000-0001-9323-8488) School of Electrical Engineering, Korea Advanced Institute of Science and Technology, changick@kaist.ac.kr, 정교수
논문번호 : 202004-077-A-LU, Received April 6, 2020; Revised April 21, 2020; Accepted April 24, 2020

형태로 존재하는지, 어디서 겹침이 일어났는지 등의 정보가 해당된다. 이러한 정보는 랜드마크 검출 뿐만 아니라 의류 영상을 이해하는데 중요한 역할을 할 수 있다.

따라서 본 논문에서는 Fabric과 Part에 해당하는 특성 정보의 분류 능력을 향상시키기 위해서 의류 랜드마크 국소화 과정에서 생성되는 특징 정보를 활용하는 의류 인식 방법을 제안한다. 그림 2는 Fabric과 Part의 특성 정보 레이블의 예시를 나타낸다. 제안하는 방법은 의류 인식과 의류 랜드마크 국소화를 동시에 진행하며, 랜드마크 국소화 과정에서 생성된 특징 정보와 의류 인식 특징 정보를 조합하여 풍부한 정보를 가지는 의류 인식을 위한 특징 정보를 생성한다. 랜드마크 국소화 과정에서는 Global-local Embedding Module (GLEM)[3]을 활용하여 의류의 문맥적인 정보를 탐색한다. GLEM은 입력 영상의 지역 정보를 비교 및 분석하고 이 정보를 융합하여 특징맵 (Feature map)이 전역적인 문맥 정보 뿐만 아니라 국소적인 문맥 정보를 포함할 수 있도록 한다. 대량 의류 영상 데이터셋인 Deepfashion 데이터셋에 대한 실험을 통해 제안하는 방법을 분석하고 해당 방법의 우수성을 보여준다.

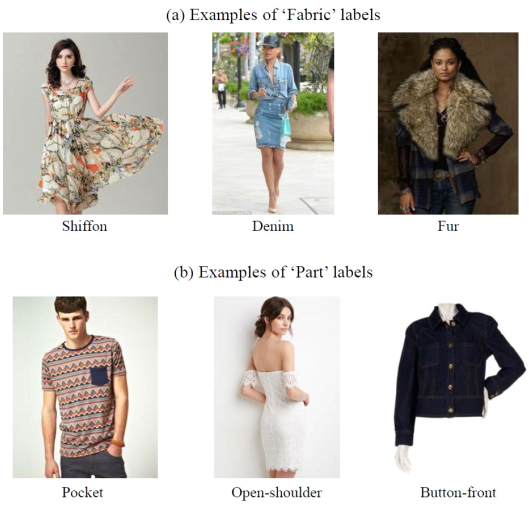

그림 2. 특성 정보 Fabric (a)과 Part의 예시
Fig. 2. Examples of 'Fabric'(a) and 'Part'(b) attribute labels

Ⅱ. 관련 연구

그림 3는 의류 랜드마크 예시를 보여준다. 의류 랜드마크는 칼라, 소매 등 옷의 구조적으로 중요한 부분

그림 3. Deepfashion 데이터셋의 의류 랜드마크 예시
Fig. 3. Example of fashion landmark on Deepfashion dataset

들에 해당된다. Deepfashion 데이터셋의 경우 좌우 칼라, 소매, 허리라인, 옷 끝 8개의 의류 랜드마크를 제공한다. 의류 랜드마크의 중요성이 커지면서 다양한 의류 랜드마크 국소화에 대한 방법[1,2,5]들이 제시되고 있다.

그림 4은 Global-local embedding module (GLEM)의 구조이다. GLEM [3]은 비국소 연산 (Non-local operation)과 컨볼루션 연산(Convolutional operation)의 구성된다. 비국소 연산으로 모든 영역 간의 의존성을 고려하고, 두 개의 콘볼루션 연산을 통해서 장거리 의존성 (long-range dependecy)이 고려된 정보를 국소적으로 종합한다. 이 모듈의 랜드마크 검출 성능은 [3]에서 증명되었다.

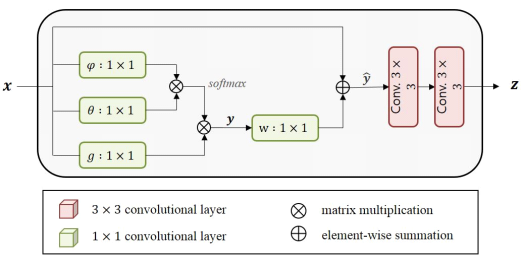

그림 4. Global-local Embedding Module (GLEM)[5]의 구조
Fig. 4. The architecture of the global-local embedding module (GLEM)[5]

Ⅲ. 제안하는 방법

그림 5은 의류 인식을 위한 의류 랜드마크 특징 탐색 네트워크 구조를 나타낸다. 네트워크는 의류 인식 분기와 의류 랜드마크 국소화 분기로 이루어진다. 먼저, 특징 추출기 (Feature Extractor)를 통해서 의류 인식과 랜드마크 국소화 분기에 대한 입력 영상의 공통 특징맵을 추출한다. 특징 추출기는 VGG-19[6]의 가장 아랫단 레이어부터 Conv.4 레이어까지 사용하였다.

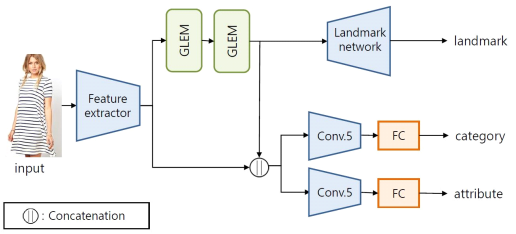

그림 5. 제안하는 방법
Fig. 5. Illustration of our proposed method

Conv.4 특징맵은 랜드마크 국소화 분기에서 Global-local Embedding Module (GLEM)[5]을 통과하여 의류 영상의 문맥적인 정보를 담게 된다. GLEM은 장거리 종속성을 고려하여 추출하는 비국소적(Non-local) 연산과 국소적인 방법으로 정보를 취합하는 컨볼루션 연산으로 구성된다. GLEM을 통해서 특징맵은 옷의 전역적인 정보 뿐만 아니라 지역적인 정보를 포함하게 되어 의류의 문맥 정보를 담게 된다. 랜드마크 국소화 네트워크를 통해 GLEM으로 생성된 특징맵으로부터 랜드마크 히트맵을 생성한다. GLEM으로 생성된 특징맵은 Conv.4 특징맵과 합쳐져 의류 인식을 위한 특징맵이 생성된다. 의류 인식을 위한 특징 맵은 카테고리 분류와 의류 특성 인식을 위해 각각 독립적인 컨볼루션 레이어와 분류기를 거쳐 테스크를 수행하게 된다.

제안하는 네트워크는 의류 인식 분기와 랜드마크 국소화 분기를 동시에 학습한다. 카테고리 분류를 위해서는 일반적인 교차 엔트로피 손실함수 $L_{category}$를 사용하였으며, 다중 레이블인 특성정보 분류를 위해서는 가중 교차 엔트로피 (Weighted Cross-Entropy) 손실함수 L_{attr}를 사용하였다. $\boldsymbol{x}_j$, $\boldsymbol{c}_j$, $\boldsymbol{a}_j$는 각각 j번째 의류 영상, 그것의 카테고리 레이블과 특성 정보 레이블이다.

$$L_{category} = \sum_j \left(\begin{array}{l} \boldsymbol{c}_j \log p(\boldsymbol{c}_j|\boldsymbol{x}_j) + \\ (1-\boldsymbol{c}_j)\log(1-p(\boldsymbol{c}_j|\boldsymbol{x}_j)) \end{array} \right) \quad (1)$$

$$L_{attr} = \sum_j \left(\begin{array}{l} w_{pos}\boldsymbol{a}_j \log p(\boldsymbol{a}_j|\boldsymbol{x}_j) + \\ w_{neg}(1-\boldsymbol{a}_j)\log(1-p(\boldsymbol{a}_j|\boldsymbol{x}_j)) \end{array} \right) \quad (2)$$

랜드마크 국소화 학습을 위해서는 랜드마크 점수 지도(Score Map)에 대한 l_2 손실함수 L_{lm}를 사용하였다. 랜드마크 점수 지도는 해당 위치가 랜드마크가 될 수 있는 정도를 점수화한 것이다. 이때, M_{ij}는 i번째 의류 영상의 j번째 정답 랜드마크 히트맵을 의미하며 $M_{ij}^{'}$는 예측한 랜드마크 히트맵이다.

$$L_{lm} = \sum_i \sum_j \| M_{ij} - M_{ij}^{'} \|_2^2 \quad (3)$$

Ⅳ. 실 험

4.1 Dataset.

Deepfashion[1] 데이터셋에 대해 제안하는 방법의 성능 평가를 수행한다. Deepfashion 데이터셋은 209,222장의 학습 데이터셋과 40,000장의 확인(Validation) 데이터셋, 40,000장의 테스트 데이터셋으로 이루어져 있다. 각 의류 영상에 대해서 랜드마크, 50개의 카테고리, 1000개의 특성정보 등의 레이블을 제공한다. 특성정보는 Texture, Fabric, Shape, Part, Style 5개의 소분류로 나뉜다.

4.2 Evaluation metrics.

카테고리 분류 성능 측정을 위해서는 Top-k 정확도를 측정하고, 특성정보 분석을 위해서는 Top-k 리콜(Recall) 정확도를 사용한다.

4.3 Performance comparison

표 1과 표 2는 랜드마크 좌표를 사용하는 의류 인식 방법[1,2,4] 및 랜드마크 정보를 사용하지 않는 방법[3]과 제안하는 방법의 성능 비교를 나타낸다. 표 1은 카테고리 분류 성능과 특성정보 전체 성능을 나타내며, 표 2는 특성정보 각 소분류에 대해서 측정한 성능을 나타낸다. 제안하는 방법은 Fabric과 Part뿐만 아니라 Texture와 Shape에서도 가장 높은 성능을 달성하였다. 카테고리 분류에서 기존의 가장 높은 성능 대비 top-3와 top-5에서 각각 98.5%, 98.7%를 달성하였다. 다만, Style에서 상당히 낮은 성능을 보였고 이는 특성정보 전체 성능이 낮아지는 결과를 보였다.

표 1. Deepfashion 데이터셋 [1]에 대한 카테고리 및 특성정보 분석 성능 비교
Table 1. Performance comparison for category classification and attribute prediction on the Deepfashion dataset[1]

Method	Category		Attribute	
	top3	top5	top3	top5
[1]	82.58	90.17	45.52	54.61
[2]	90.99	95.78	51.53	60.95
[3]	91.37	95.26	47.70	57.28
[4]	91.16	96.12	54.69	63.74
Ours	90.46	95.27	27.30	35.61

표 2. Deepfashion 데이터셋 [1]에 대한 소분류 특성정보 분석 성능 비교. 가장 높은 성능은 파란색, 두 번째 높은 성능은 빨간색으로 표시.
Table 2. Performance comparison for attribute prediction of subcategory on the Deepfashion dataset[1]. The best scores are marked in blue, and the second best scores are marked in red.

Method	Texture		Fabric		Shape		Part		Style	
	top3	top5	top3	top5	top3	top5	top3	top5	top3	top5
[1]	37.46	49.52	39.30	49.84	39.41	48.59	44.13	54.02	66.43	73.16
[2]	50.31	65.48	40.31	48.23	53.32	61.05	40.65	56.32	68.70	74.25
[3]	56.95	66.24	44.03	54.21	56.87	66.25	44.89	55.15	33.98	42.21
[4]	56.17	65.83	43.20	53.52	58.28	67.80	46.97	57.42	68.82	74.13
Ours	57.07	66.47	44.52	55.17	59.52	68.62	47.93	58.28	33.67	42.57

4.4 Analysis & Future Work

카테고리 분류에서 약간의 성능 손실과 특성 정보의 Style에서 성능 저하가 있지만, 특성 정보 Texture, Fabric, Shape, Part에서 성능이 향상되어 가장 높은 성능을 달성하였다. 성능이 오른 네 개의 특성 정보 소분류들은 비교적 국소적인 영역과 관계된 특성 정보(예: Pocket, chiffon, Cotton, Lace)를 포함한다. 반면, 카테고리나 Style은 Retro, Relaxed, Ornate, Wild 등 주관적이고 전체적인 특징 정보를 많이 요한다. 이러한 특징을 가지는 Style이 충분히 학습되기에는 랜드마크 국소화 과정에서 추출한 특징맵과 특징 추출기에서 얻은 공통 특징맵 사이의 균형이 적합하지 않았던 것으로 분석된다. 향후 공통 특징맵과 문맥 특징 정보의 균형을 조절하여 Style 특성 정보의 특징을 적절히 학습할 수 있는 방법을 연구할 계획이다.

V. 결 론

본 논문에서는 Fabric과 Part 성능 향상을 위해 의류 랜드마크 국소화 과정에서 발생하는 특징맵을 활용한 의류 인식 방법을 제안하여 의류 인식을 수행하였다. Global-local embedding module을 활용하여 추출한 문맥 특징 정보를 활용하여 의류의 non-rigid 특성을 극복하고자 하였다. 실험을 통해 제안하는 방법이 다른 방법들과 비교할만한 성능을 나타냄을 보여주었다. 공통 특징 정보와 문맥 특징 정보의 가중치를 조절하고 Style의 특성 정보 특성을 반영할 수 있는 방법을 고안한다면 더 높은 성능을 얻을 수 있을 것으로 판단된다.

References

[1] Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang, "Deepfashion: Powering robust clothes recognition and retrieval with rich annotations," *IEEE Conf. Comput. Vision and Pattern Recognition*, pp. 1096-1104, Jun. 2016.

[2] W. Wang, Y. Xu, J. Shen, and S. C. Zhu, "Attentive fashion grammar network for fashion landmark detection and clothing category classification," *IEEE Conf. Comput. Vision and Pattern Recognition*, pp. 4271-4280, Jun. 2018.

[3] S. Lee, H. Eun, S. Oh, W. Kim, C. Jung, and C. Kim, "Landmark-free clothes recognition with a two-branch feature selective network," *Electron. Lett.*, vol. 55, no. 13, pp. 745-747, Jun. 2019.

[4] J. Liu and H. Lu, "Deep fashion analysis with feature map upsampling and landmark-driven attention," in *Proc. ECCV 2018*, vol. 11131, pp. 30-36, 2018.

[5] S. Lee, S. Oh, C. Jung, and C. Kim, "A global-local embedding module for fashion landmark detection," in *Proc. IEEE ICCV Workshop*, Nov. 2019.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *ICLR 2015*, Apr. 2015.