배경이 이미지 유사도 학습에 미치는 영향 분석*

선주형 ○, 안국원 ○, 윤성의
한국과학기술원 전산학부
munuwazzi@kaist.ac.kr, anguoyuan111@gmail.com, sungeui@kaist.edu

Empirical Analysis of the Effects of Image Backgrounds in Deep Metric Learning

Ju-Hyeong Seon○, Guoyuan An○, Sung-Eui Yoon
School of Computing, KAIST

요 약

이미지 검색, 신원 확인 등에 사용되는 메트릭 학습(Deep Metric Learning)은 신경망이 이미지의 시각적 유사성을 반영하는 공간에 데이터를 인코딩하도록 지도한다. 그러나 이미지에는 중요한 정보인 물체와 그렇지 않은 정보인 배경이 일종의 편향성(bias)을 띠며 혼합되어 있기 때문에, 신경망이 이 편향을 사용하여 물체 간 유사도를 계산할 수 있다. 본 논문은 실험을 통해 기존의 메트릭 학습 모델이 배경을 사용하여 이미지 유사도를 계산하는 문제점을 보이고, 실험 결과를 바탕으로 새로운 해결책을 제시한다.

1. 서 론

이미지의 시각적인 유사도 정보는 물체 검색 (object retrieval), 인간 자세 추정 (pose estimation), 신원 인식 (face verification) 등의 문제에서 중요한 역할을 한다. 이때, 이미지 유사성을 판단하는 일반적인 방법으로 딥러닝 기반의 메트릭 학습(Deep Metric Learning) 모델이 사용된다.

메트릭 학습법은 데이터 간 거리가 클래스 유사성을 의미하는 인코딩 공간에 모델이 이미지를 배치하도록 지도한다. 이미지는 유사도 판단의 대상인 물체와 그 외의 배경으로 이루어지는데, 이미지의 클래스 유사성은 물체의 시각적 정보를 바탕으로 판단되어야 한다. 하지만, 이미지 데이터셋에는 물체와 배경이 예측할 수 없는 편향 (bias)을 띄며 혼합되어 있다. 예를 들면, 새 데이터셋에서 펭귄 이미지의 배경은 대부분 빙하라는 점은 편향의 한 예시이다. 이는 모델이 빙하 배경의 이미지를 모두 펭귄 클래스로 판단하는 오류를 유도할 수 있다. 즉, 특별한 지도가 없다면 신경망이 물체와 배경을 모두 사용하여 유사도를 판단하므로, 의도치 않은 편향에 의해 모델의 성능이 변하는 문제가 발생한다. 또한 이 편향성은 데이터셋마다 다르기 때문에, 데이터셋에 따라 모델의 성능이 크게 변하는 문제도 발생할 수 있다.

본 논문은 실험을 통해, 기존의 메트릭 학습 모델이 물체와 배경의 편향성을 이용하여 시각적 유사도를 계산한다는 문제점을 제기한다. 또한, 실험 결과를 바탕으로 모델의 배경 편향 의존성을 줄이고 성능을 개선하는 새로운 학습법을 제시한다.

2. 관련 연구

2.1 메트릭 학습 (Deep Metric Learning)

메트릭 학습 모델은 의미상 유사한 이미지를 인코딩 공간상에서 가깝게 배치하므로, 물체를 검색하거나 신원을 확인하는 데 사용될 수 있다. 딥러닝이 이미지 분류 성능을 극적으로 개선한 후, 물체 분류에 사전 훈련된 신경망이 메트릭 학습의 기본 모델로 사용된다. 딥러닝 모델이 공간상의 같은 클래스 데이터를 끌어당기고 다른 클래스의 데이터를 밀어내도록 손실 함수를 설계하는 것이 메트릭 학습 연구의 주류를 이루었다[1].

초기에는 특정 쿼리 이미지와 클래스가 같은 이미지와 다른 이미지로 쌍을 이루어 모델을 학습시키는 트리플렛 (Triplet) 방식이 사용되었다. 많은 데이터 쌍을 만들 수 있기 때문에, 효율적인 학습을 위한 트리플렛 샘플링 기법이 활발히 연구되었다[2]. 한 편, 트리플렛의 높은 복잡도를 구조적으로 해결하기 위해 데이터의 부분 집합을 대표하는 프록시(Proxy)와 쿼리 간 유사도를 학습하는 방식도 제안되었다[3]. 현재의 메트릭 학습법은 위의 2가지 방식으로 분류되는데, 이후의 실험에서는 두 방식 모두 이미지 배경 편향 정보를 사용함을 보인다.

2.2 픽셀 분류 (Semantic Segmentation)

Semantic Segmentation은 자율 주행이나 보안 카메라에 적용되는 기술로, 이미지의 모든 픽셀을

*이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (IITP-2015-0-00199, 대용량 이미지 검색과 시제품 렌더링을 위한 근접질의 SW 개발)

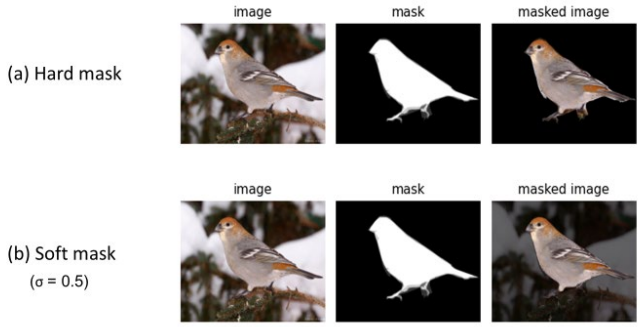

그림 1. Hard mask & Soft mask 방법. Mask(Segmentation map)를 사용하여 배경의 RGB 값을 (a) 0으로 바꾸거나 (b) 일정 비율 σ로 낮춘다.

클래스별로 분류하는 작업이다. 픽셀마다 레이블을 얻는 비용이 크기 때문에, 최근에는 적은 양의 데이터로 분류를 학습하는 연구(few-shot segmentation)가 활발히 진행되고 있다[4]. 이후의 실험에서는 Semantic segmentation map이 이미지의 배경을 가리기 위해 사용된다.

3. 메트릭 학습의 배경 편향성 확인 실험

본 실험의 목적은 기존의 메트릭 학습 모델이 물체의 클래스 유사성을 계산할 때 이미지의 배경 정보를 사용함을 확인하는 것이다. 배경을 모두 가린 훈련 & 테스트 데이터셋에서 모델의 성능이 배경이 있는 경우보다 낮다면, 기존의 모델이 배경 정보를 활용하여 물체의 유사도를 판단함을 확인할 수 있다.

배경을 가리기 위해 앞서 소개한 Segmentation map을 사용한다. Segmentation map으로 이미지상에서 배경을 나타내는 픽셀을 찾고, 2가지의 방법으로 배경을 가린다 (그림 1). (a) Hard mask는 배경의 RGB 값을 0으로 바꾸고, (b) Soft mask는 배경 RGB 값을 일정한 비율 σ로 줄인다 ($0<\sigma<1$). 두 방식으로 가려진 데이터셋으로 학습하고 테스트한 모델의 성능을 비교한다. 성능 측정에는 메트릭 학습의 벤치마크 데이터셋 CUB200-2011을 사용하여 모델의 이미지 검색 정확도를 측정한다. 기준 모델로는 트리플렛 기반의 Multi-similarity Loss 모델[2]과 프록시 기반의 Proxy-Anchor Loss 모델[3]이 사용되었다. Segmentation map은 CUB200-2011 홈페이지에 제공된 것을 사용하였다.

4. 결과 및 분석

그림 2는 배경이 가려진 데이터셋을 학습한 모델의 이미지 검색 결과 예시이다. No-mask는 배경이 가려지지 않은 원본 데이터셋을 의미한다. No-mask

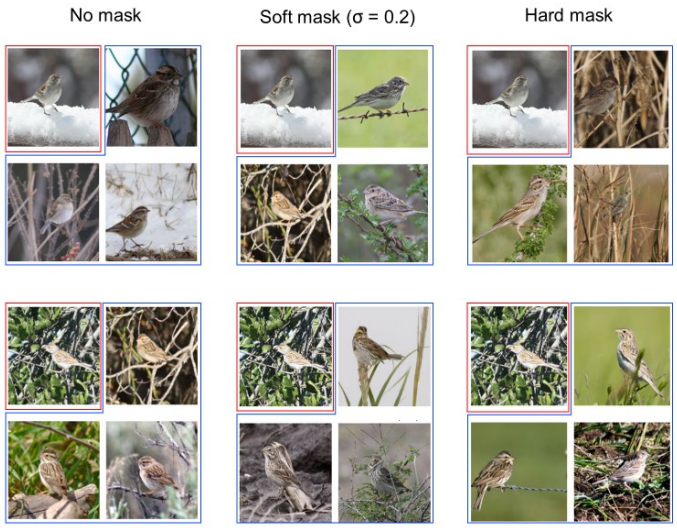

그림 2. Mask 데이터로 학습한 모델 별 이미지 검색 결과. 빨강: 쿼리 이미지, 파랑: 모델이 가장 유사하다고 판단한 이미지. No mask 데이터셋 학습 모델이 응답한 이미지는 쿼리와 배경이 유사하지만 (1열), Soft mask 또는 hard mask 모델은 유사하지 않다 (2, 3열).

Proxy Anchor Loss	Recall@1 ↑	Recall@2 ↑	Recall@4 ↑	Recall@8 ↑
No mask	0.658	0.761	0.841	0.904
Soft mask (σ = 0.2)	**0.663**	**0.765**	**0.848**	**0.905**
Soft mask (σ = 0.5)	0.653	0.764	0.840	0.904
Hard mask	0.638	0.753	0.834	0.892
Multi Similarity Loss	Recall@1 ↑	Recall@2 ↑	Recall@4 ↑	Recall@8 ↑
No mask	0.657	**0.770**	**0.863**	**0.912**
Soft mask (σ = 0.2)	0.656	0.765	0.849	0.909
Soft mask (σ = 0.5)	**0.660**	0.764	0.846	0.910
Hard mask	0.648	0.752	0.833	0.900

표 1. Mask 데이터로 학습한 모델 별 검색 테스트 정확도 (Recall@k ↑). 프록시 기반 모델(상단)은 Soft mask를 사용할 때, 트리플렛 기반 방법(하단)은 mask가 없을 때 가장 높은 성능을 보였다. 두 방식 모두 hard mask를 사용할 때 정확도가 가장 낮다.

모델이 응답한 이미지는 쿼리와 배경이 유사하지만 (1행: 흰색 눈), soft mask 또는 hard mask 모델은 유사하지 않다 (1행: 녹색 풀, 갈색 나뭇가지). 즉, 배경을 가리고 학습한 모델은 이미지 유사도 판단에서 배경의 영향을 적게 받았음을 확인할 수 있다.

표 1은 배경이 가려진 데이터셋을 학습한 모델의 검색 테스트 성능을 보여준다. 프록시 기반과 트리플렛 기반 모델 모두 hard mask 데이터셋에서 정확도가 가장 낮다. 즉, 기존의 학습법(no-mask)은 배경 정보를 사용해 성능을 높였음을 알 수 있다. 이는 기존의 훈련 데이터와 테스트 데이터가 비슷한 이미지 배경 편향을 가지고 있기 때문으로, 편향이 없는 테스트 데이터셋에서는 기존 모델의 성능이 낮아질 것을 예상할 수 있다. 이후에는 메트릭 학습 모델의 편향성 의존을 줄이면서 유사도 판단 성능을 개선하는 새로운 방안을 제시한다.

5. 개선 방안

실험을 통해, 기존의 모델은 물체와 배경의 상관관계를 이용하여 성능을 개선함을 확인했다. 이는 모델이 훈련 데이터에서 학습한 물체와 배경 간 관계가 테스트 데이터에서도 확인되었다는 뜻이다. 즉, 훈련 중 학습되는 편향은 테스트 데이터에 따라 도움이 되거나 그렇지 않을 수 있다. 다만, 문제는 모델이 훈련 중 어떤 편향을 학습하는지 모른다는 점이다. 이때, 본 논문에서 제시한 실험을 응용하면 모델이 학습한 편향을 파악할 수 있다. No mask와 hard mask 데이터셋마다 모델을 학습하고, 클래스별로 검색 테스트에 실패한 데이터 수를 파악한다. 특정 클래스 A에서 테스트에 실패한 데이터의 수를 비교할 때, 배경을 사용하여 학습했을 때(no mask)가 그렇지 않을 때보다(hard mask) 많다고 하자. 클래스 A는 no mask 모델이 배경으로부터 학습한 편향의 방해를 받아 테스트에 실패하였음을 유추할 수 있다. 이 클래스 A의 데이터에서 배경과 물체 관계의 패턴을 발견한다면, 모델이 학습한 편향을 파악하는 데에 도움이 된다. 하지만 이 방법은 테스트 데이터셋의 결과를 이용해야 한다는 단점이 있다. 따라서, 훈련 중에도 모델의 편향성을 파악할 수 있는 지도법의 연구가 이루어질 필요가 있다.

만약 모델이 사용될 테스트 데이터셋을 알 수 없다면, 특정 편향의 유용성을 예측하기 어렵다. 이 경우에는 편향 의존성을 줄이고 물체에 집중하는 것이 다양한 데이터셋에서의 일관된 성능에 도움이 된다. 모델이 물체에 집중하도록 돕기 위해서는 실험에서 제안된 segmentation map을 활용할 수 있다. 그림 2에서 나타나듯이 soft mask 모델은 hard mask와 마찬가지로 유사도 계산에서 배경의 영향을 적게 받지만, 일부 상황에서 no mask 모델보다 높은 정확도를 보인다. 이는 soft mask가 배경의 RGB 값을 낮추어 모델이 물체에 더 집중하도록 돕기 때문이다. Soft mask 방식이 아니더라도, segmentation map을 이미지와 함께 전달하면 모델은 이미지상에서 물체와 배경을 구분할 수 있다. 배경보다 물체에 집중하는 방법은 최근 주목받는 어텐션 (attention) 모듈을 사용한 비전 트랜스포머 (Vision Transformer) 모델의 강점이다[5]. 하지만, 트랜스포머는 학습에 필요한 데이터가 많고 계산 비용이 많이 들며, 이는 크기가 작은 데이터셋이 사용되는 메트릭 학습 분야에서 큰 단점이다[6]. 이와 달리, few-shot segmentation 모델은 적은 데이터로도 다양한 데이터셋에서 정확한 segmentation map을 얻을 수 있다[4]. 이 segmentation map을 사용하는 유사도 계산 모델은 물체 정보에 집중하여 다양한 데이터셋에서 우수한 성능을 달성할 것이다. 이러한 맥락에서, 계산된 Segmentation map을 메트릭 학습 모델에 효과적으로 전달하는 기법은 향후 연구의 주제가 될 것이다.

6. 결론

본 논문은 딥러닝 기반 메트릭 학습 연구를 소개하고, 실험을 통해 기존의 모델이 이미지의 배경을 사용해 물체의 시각적 유사성을 계산함을 보였다. 이는 의도하지 않은 편향이 모델의 이미지 검색 성능을 높인다는 문제를 확인한 의의가 있다. 또한, 모델의 배경 편향 의존성을 줄이기 위해 few-shot segmentation 모델을 사용하는 새로운 학습법을 제안하였다.

참 고 문 헌

[1] Mahmut KAYA et al. Deep Metric Learning: A Survey. Symmetry. 2019.

[2] Xun Wang et al. Multi-Similarity Loss with General Pair Weighting for Deep Metric Learning. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2019.

[3] Sungyeon Kim et al. Proxy Anchor Loss for Deep Metric Learning. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2020.

[4] Xiangwen Shi et al. Multi-similarity based Hyperrelation Network for few-shot segmentation. arXiv preprint. 2022. (arXiv:2203.09550)

[5] Samuel Black et al. Visualizing Paired Image Similarity in Transformer Networks. IEEE/CVF Winter Conference on Applications of Computer Vision. 2022.

[6] Salman Khan et al. Transformers in Vision: A Survey. ACM. 2021.