벡터 데이터베이스 최적화를 위한 이미지 특징값 추출 모델 설계

송진혁, 조용성
한국전자통신연구원

song020@etri.re.kr

Design of an Image Feature Extraction Model for Vector Database Optimization

JinHyuk Song and YongSeong Cho
ETRI

요 약

최근 대용량 이미지 기반 검색 및 중복 검출 기술이 다양한 산업에서 활용되면서, 벡터 데이터베이스의 효율적인 운용이 중요해지고 있다. 하지만 기존 MobileNetV2 기반 모델은 1,280 차원의 임베딩 벡터를 출력하여, GPU 메모리 소모와 검색 속도 측면에서 한계를 드러낸다. 본 논문에서는 임베딩 차원을 320 으로 축소하면서도 파라미터 수 증가를 최소화한 경량 이미지 특징 추출 모델을 제안한다. 제안된 모델은 기존 대비 유사도 분포 분리도가 향상되었으며, 잘못 분류되는 비율도 크게 감소하는 등 벡터 검색의 정확도와 효율성을 동시에 확보할 수 있음을 실험적으로 입증하였다.

Ⅰ. 서론

최근 다양한 분야에서 이미지 기반 검색 시스템의 활용이 급증하면서, 고속 이미지 유사도 검색을 위한 벡터 데이터베이스(Vector Database)[1, 2]의 중요성이 부각되고 있다. 특히 대규모 이미지 데이터셋을 활용하는 환경에서는 개별 이미지에 대한 특징값(임베딩)을 벡터 형태로 추출하여 GPU 기반 데이터베이스에 저장하고, 벡터 간 유사도 계산을 통해 유사 이미지를 검색하는 방식이 일반화되고 있다. 그러나 기존의 대표적인 경량화 모델인 MobileNetV2[3] 기반 특징값 추출 방식은 출력 임베딩의 차원이 1,280 으로 고정되어 있어, 수백만 개 이상의 이미지에 대한 벡터를 저장할 경우 벡터 데이터베이스의 크기가 기하급수적으로 증가한다. 이러한 대용량 벡터를 GPU 메모리에 탑재하여 실시간 검색을 수행할 경우, 처리 속도 저하와 함께 인프라 비용 증가 문제가 발생하게 된다. 이와 같은 문제를 해결하기 위해서는, 벡터 데이터베이스의 검색 정확도는 유지하면서도 임베딩 차원을 줄여 저장 공간을 최소화하고, GPU 메모리 효율을 개선할 수 있는 특징값 추출 모델의 설계가 필요하다. 이에 본 논문에서는 기존 CNN 기반 특징 추출 모델 구조를 개선하여, 출력 임베딩 차원을 축소하면서도 유사도 기반 검색에서의 성능을 유지하거나 향상시킬 수 있는 새로운 이미지 특징값 추출 모델을 제안한다.

Ⅱ. 이미지 특징값 추출 모델 제안

기존의 이미지 특징 추출 모델 중 하나인 MobileNet V2 는 경량화 구조에도 불구하고, 마지막 단계에서 1,280 차원의 고차원 임베딩 벡터를 출력한다. 이로 인해 대규모 이미지 데이터셋에 대해 벡터 데이터베이스를 구축할 경우, 전체 저장 용량이 증가하고 GPU 메모리 자원이 과도하게 소모되는 문제가 발생한다. 또한 벡터 간 유사도 계산 과정에서 고차원 연산이 필요하므로 검색 속도에도 부정적인 영향을 미친다. 임베딩 차원을 축소하기 위한 일반적인 접근 방식으로는 기존 출력 벡터에 Fully Connected (FC) Layer 를 추가하여 차원을 낮추는 방법이 있다. 그러나 이 방식은 모델의 전체 파라미터 수를 크게 증가시키는 단점이 존재한다. 실제로 FC 방식 적용 시 모델 파라미터는 약 1 백만 개 이상 증가하여, 경량화를 중시하는 모바일 환경이나 실시간 시스템에 적합하지 않다. 이에 본 논문에서는 MobileNetV2 의 구조를 수정하여 파라미터 수 증가를 최소화하면서도 임베딩 차원을 효과적으로 줄일 수 있는 특징값 추출 모델을 제안한다. 제안된 방식은 기존의 320 → 1280 경로의 Convolution 대신, 320 → 640 의 중간 차원으로 확장한 뒤, 이후에 640 → 480 → 320 으로 점진적으로 차원을 축소하는 DNN 구조를 적용하는 것이다. 이와 같은 단계적 축소 구조는 정보 손실을 최소화하면서도, 파라미터 수를 기존보다 효율적으로 관리할 수 있는 장점이 있다. 아래 <표 1>은 기존 MobileNetV2, FC 기반 차원 축소 모델, 그리고 본 논문에서 제안하는 모델의 파라미터 수와 출력 임베딩 크기를 비교한 결과이다. 제안된 모델은 FC 방식에 비해 약 760,000 개의 파라미터를 줄이면서도 동일한 320 차원의 임베딩 벡터를 출력한다. 이를 통해 저장 공간을 절약하고 벡터 데이터베이스의 검색 성능을 유지하면서도, 경량성과 효율성을 동시에 달성할 수 있다.

표 1. 모델별 파라미터 수 및 임베딩 벡터 크기 비교

	파라미터	임베딩 크기
기존 MobileNetV2	2,223,872	1,280
FC 기반 차원 축소 모델	3,248,832	320
제안된 모델	2,482,392	320

Ⅲ. 학습 환경 및 데이터셋 구성

이미지 특징값 추출 모델은 벡터 간의 유사도 기반 검색 성능을 극대화하기 위해 대조 학습(Contrastive Learning)[4] 기법을 적용하여 학습되었다. 대조

학습은 임베딩 공간 상에서 anchor 이미지와 유사한 이미지(positive)는 가깝게, 비유사한 이미지(negative)는 멀어지도록 모델을 학습시키는 방식으로, 이미지 간 관계성을 효과적으로 반영할 수 있다. 대조 학습에는 margin 기반 contrastive loss 함수를 사용하였으며, margin 값은 0.45 로 설정하였다. 이는 positive 쌍과 negative 쌍 간의 거리 차이를 명확히 구분하기 위한 기준값으로 사용된다. 최적화에는 Adam 옵티마이저를 적용하였으며, 학습률(learning rate)은 0.000004 로 설정하여 안정적인 수렴을 유도하였다. 학습은 batch size 64 로 수행되었으며, negative 샘플 벡터의 길이는 10,240 으로 설정하였다. 또한 positive 쌍 간 임베딩 거리의 목표값(target distance)은 0.8 로 설정하여, 유사 이미지 간의 임베딩 거리가 해당 값 이하가 되도록 학습되었다

표 2. 학습용 이미지 데이터셋 구성

구분	이미지수	설명
Anchor	50,000	기준 이미지
Positive	500,000	Anchor 에 대한 변형 이미지 (15 종)
Negative	500,000	유사하지 않은 비교용 이미지

Ⅳ. 제안된 모델 학습 및 결과 분석

제안된 이미지 특징값 추출 모델의 성능을 검증하기 위해 학습을 진행하였으며, 학습 손실값(Train Loss)과 검증 손실값(Validation Loss)의 변화를 <그림 1>과 같이 시각화하였다. 학습 초기에는 두 손실값이 모두 급격히 감소하며, 이후 에폭이 증가함에 따라 점진적으로 완만하게 감소하는 양상을 보였다. 특히 에폭 30 이후부터는 학습 손실과 검증 손실이 안정적인 수렴을 보이며, 오버피팅 없이 학습이 성공적으로 이루어졌음을 확인할 수 있다. 또한, 학습된 모델의 임베딩 벡터 성능을 평가하기 위해 anchor 이미지와 positive 이미지 간의 평균 유사도(A_P), anchor 와 negative 이미지 간의 평균 유사도(A_N), 각 유사도의 표준편차, 유사도 차이(A_P - A_N), 그리고 A_N > A_P 인 사례의 개수를 측정하여 표 3 에 정리하였다. 세 가지 에폭 구간(10, 30, 60)에 대해 측정된 결과는 표 3 과 같다. A_P 유사도는 에폭이 증가함에 따라 소폭 감소하지만, A_N 유사도는 지속적으로 낮아지는 경향을 보인다. 이는

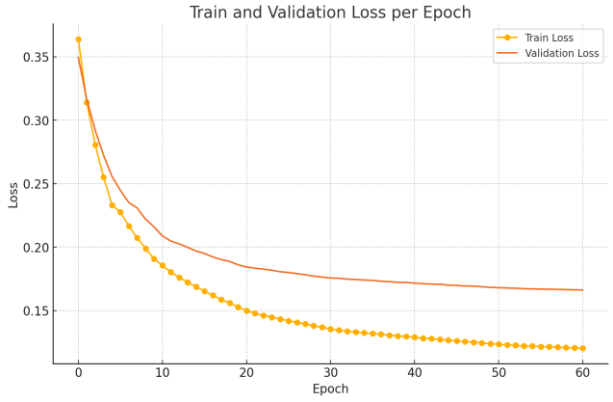

그림 1. Train 및 Validation Loss 변화 추이

모델이 negative 이미지와의 임베딩 간 거리를 효과적으로 확장해나가고 있음을 의미한다. 또한, 유사도 차이는 에폭 60 에서 가장 크게 나타났으며 잘못 분류된 사례 수 역시 가장 적었다(27 건). 이를 통해 에폭 60 시점에서 학습된 모델이 positive-negative 구분 성능 측면에서 가장 우수함을 확인할 수 있다.

표 3. epoch 당 유사도 성능 비교

	epoch		
	10	30	60
A_P 유사도	0.921	0.92	0.919
A_N 유사도	0.437	0.404	0.392
A_P 표준편차	0.00381	0.004	0.004
A_N 표준편차	0.0201	0.01968	0.0191
유사도 차이	0.484	0.516	0.527
A_N > A_P 개수	40	32	27

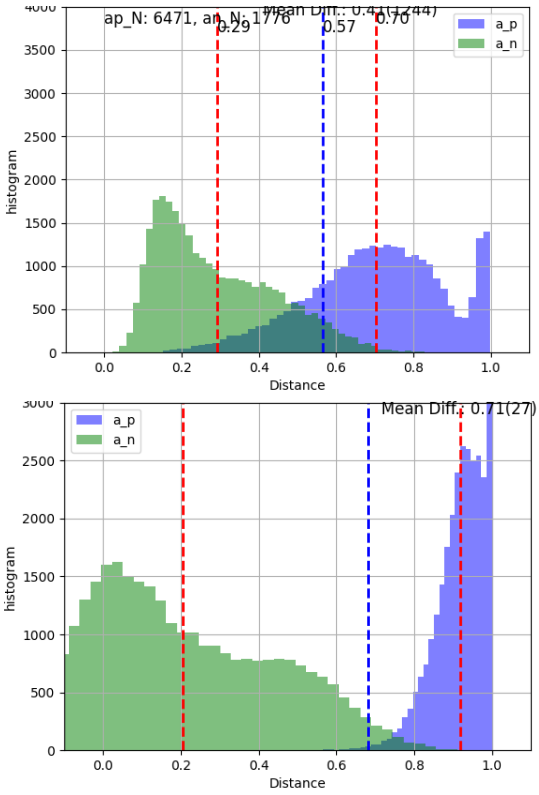

그림 2. A_P, A_N 거리 분포 비교
(학습 전(상), 학습 후(하))

학습의 효과를 보다 직관적으로 확인하기 위하여 A_P 쌍과 A_N 쌍 간의 거리 분포를 히스토그램 형태로 시각화하였다. <그림 2>는 학습 초기(상단)와 학습 완료 이후(하단)의 임베딩 거리 분포를 비교한 것이다. 각 분포의 중심값을 기준으로 양자 간의 평균 차이와 임계값 위치도 함께 표시하였다. 상단 히스토그램은 모델이 학습되기 이전의 거리 분포를 나타낸다. 이 시점에서는 A_P 와 A_N 의 거리 분포가 비교적 넓게 겹쳐져 있어, positive 와 negative 간의 임베딩 구분이 명확하지 않다. 특히 중첩 영역이 두드러지며, 임계값(threshold) 기준으로 잘못

분류되는 샘플 수가 상당함을 알 수 있다. 반면, 하단의 학습 완료 후 히스토그램에서는 A_P 와 A_N 의 거리 분포가 뚜렷하게 분리되었으며, 평균 거리 차이가 0.71 로 크게 증가하였다. A_P 는 대부분 0.7 이상의 고유사도로 집중된 반면, A_N 은 0.5 이하의 분포로 이동하여 두 군간의 중복 영역이 거의 사라졌다. 이로 인해 임베딩 구분이 명확해졌으며, 잘못 분류되는 샘플 수도 대폭 감소한 것을 확인할 수 있다. 이와 같은 분포 변화는 contrastive learning 기반 학습의 효과를 입증하는 결과로, 제안된 모델이 유사도 기반 검색에서 더욱 정밀하고 신뢰도 높은 특징 추출이 가능하다는 점을 뒷받침한다. 결과적으로, 본 모델은 벡터 데이터베이스 검색의 정확도 향상과 오탐률 감소에 효과적인 특징값을 제공할 수 있음을 실증적으로 보여준다.

V. 결론

본 논문에서는 벡터 데이터베이스의 효율성을 높이기 위해, MobileNetV2 의 구조를 경량화하고 임베딩 차원을 1,280 에서 320 으로 줄인 이미지 특징값 추출 모델을 제안하였다. 제안된 모델은 파라미터 수 증가를 최소화하면서도 유사도 기반 검색 성능을 유지하도록 설계되었으며, contrastive learning 기반 학습을 통해 효과적으로 임베딩 표현을 학습하였다. 학습 결과, 손실 함수는 안정적으로 수렴하였고, anchor-positive 와 anchor-negative 간 유사도 차이도 에폭 60 기준으로 가장 크게 나타났다. 또한 히스토그램 분석을 통해 학습 전후의 거리 분포가 뚜렷하게 분리됨을 확인하였으며, 잘못 분류되는 샘플 수 역시 감소하였다. 본 모델은 저장 공간 절약과 정확도 향상을 동시에 달성할 수 있으며, 향후 다양한 분야에서의 적용 가능성이 기대된다.

ACKNOWLEDGMENT

이 논문은 2025 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. RS-2023-00224740, 디지털성범죄 피해 예방을 위한 불법촬영물 이미지 유포 차단 및 추적 기술 개발)

참 고 문 헌

[1] Y. Han, C. Liu, and P. Wang, "A Comprehensive Survey on Vector Database: Storage and Retrieval Technique, Challenge," arXiv preprint arXiv:2310.11703, Oct. 2023. [Online]. Available: https://arxiv.org/abs/2310.11703

[2] J. J. Pan, J. Wang, and G. Li, "Survey of Vector Database Management Systems," The VLDB Journal, vol. 33, no. 2, pp. 123- 145, Feb. 2024. [Online]. Available: https://doi.org/10.1007/s00778-024-00864-x

[3] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, Jun. 2018, pp. 4510-4520.

[4] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," in Proceedings of the 37th International Conference on Machine Learning (ICML), Jul. 2020, pp. 1597- 1607