KIPS Trans. Softw. and Data Eng.
Vol.12, No.12 pp.519~524
ISSN: 2287-5905 (Print), ISSN: 2734-0503 (Online)

https://doi.org/10.3745/KTSDE.2023.12.12.519

Comparative Analysis of Self-supervised Deephashing Models for Efficient Image Retrieval System

Kim Soo In† · Jeon Young Jin†† · Lee Sang Bum††† · Kim Won Gyum††††

ABSTRACT

In hashing-based image retrieval, the hash code of a manipulated image is different from the original image, making it difficult to search for the same image. This paper proposes and evaluates a self-supervised deephashing model that generates perceptual hash codes from feature information such as texture, shape, and color of images. The comparison models are autoencoder-based variational inference models, but the encoder is designed with a fully connected layer, convolutional neural network, and transformer modules. The proposed model is a variational inference model that includes a SimAM module of extracting geometric patterns and positional relationships within images. The SimAM module can learn latent vectors highlighting objects or local regions through an energy function using the activation values of neurons and surrounding neurons. The proposed method is a representation learning model that can generate low-dimensional latent vectors from high-dimensional input images, and the latent vectors are binarized into distinguishable hash code. From the experimental results on public datasets such as CIFAR-10, ImageNet, and NUS-WIDE, the proposed model is superior to the comparative model and analyzed to have equivalent performance to the supervised learning-based deephashing model. The proposed model can be used in application systems that require low-dimensional representation of images, such as image search or copyright image determination.

Keywords : Deephashing, Image Retrieval, Variational Inference, Self-supervised Learning, Attention Mechanism

효율적인 이미지 검색 시스템을 위한 자기 감독 딥해싱 모델의 비교 분석

김 수 인† · 전 영 진†† · 이 상 범††† · 김 원 겸††††

요 약

해싱 기반 이미지 검색에서는 조작된 이미지의 해시코드가 원본 이미지와 달라 동일한 이미지 검색이 어렵다. 본 논문은 이미지의 질감, 모양, 색상 등 특징 정보로부터 지각적 해시코드를 생성하는 자기 감독 기반 딥해싱 모델을 제안하고 평가한다. 비교 모델은 오토인코더 기반 변분 추론 모델들이며, 인코더는 완전 연결 계층, 합성곱 신경망과 트랜스포머 모듈 등으로 설계된다. 제안된 모델은 기하학적 패턴을 추출하고 이미지 내 위치 관계를 활용하는 SimAM 모듈을 포함하는 변형 추론 모델이다. SimAM은 뉴런과 주변 뉴런의 활성화 값을 이용한 에너지 함수를 통해 객체 또는 로컬 영역이 강조된 잠재 벡터를 학습할 수 있다. 제안 방법은 표현 학습 모델로 고차원 입력 이미지의 저차원 잠재 벡터를 생성할 수 있으며, 잠재 벡터는 구분 가능한 해시코드로 이진화 된다. CIFAR-10, ImageNet, NUS-WIDE 등 공개 데이터셋의 실험 결과로부터 제안 모델은 비교 모델보다 우수하며, 지도학습 기반 딥해싱 모델과 동등한 성능이 분석되었다.

키워드 : 딥해싱, 이미지 검색, 변분 추론, 자기 지도 학습, 어텐션 메커니즘

1. 서 론

딥해싱(deephashing)은 이미지의 지각적 표현을 이용해 이진 해시코드를 매핑하는 AI 기법으로 이미지 콘텐츠 인식과 검색, 워터 마킹 등 다양한 응용에 활용되고 있다[1]. 그러나 이미지 조작은 멀티미디어 인증 및 보안을 보장하기에 어려움이 있다[2,3]. 이미지 편집 소프트웨어는 색상 보정, 개체 수정 및 복제와 같은 조작을 용이하게 하며 변형된 이미지는 원본 이미지와 다른 해시코드를 갖기 때문에 이미지 검색 시스템의 성능을 저하한다. 따라서 이미지 조작에도 강건한 디지털 이미지 검색 시스템의 필요성이 강조된다.

기존의 해싱 방법은 고차원 입력 이미지를 고정 길이의 저차원 해시코드로 변환한다. 조작된 이미지는 원본과 서로 다른 해시코드를 생성하며 원본 이미지의 해시코드와 비교시 높은 비유사도(dissimilarity)를 갖는다[4]. 최근 이미지의 질감, 형태와 색상 등 지각적 특징 정보를 이용해 유사한 이미

※ 이 논문은 2023년도 정부(문화체육관광부)의 재원으로 한국콘텐츠진흥원 의 지원을 받아 수행된 연구임(No.2021-ec-9500S2, 교육 콘텐츠에 대한 인공지능 기반 저작권 침해 의심요소 검출 및 대체 재료 콘텐츠 추천 기술 개발).
† 준 회 원 : 단국대학교 인공지능융합학과 석사과정
†† 비 회 원 : ㈜에이아이딥 책임연구원
††† 종신회원 : 단국대학교 소프트웨어학과 정교수
†††† 비 회 원 : ㈜에이아이딥 연구소장
Manuscript Received : August 24, 2023
First Revision : October 31, 2023
Accepted : November 28, 2023
* Corresponding Author : Kim Soo In(kimsooinj7740@gmail.com)

지를 검색하는 딥해싱(deephashing) 연구가 진행되었다[5]. 딥해싱은 고차원 데이터를 저차원 이진코드로 매핑하는신경망 모델이다[6].

지도학습 기반 딥해싱은 학습 데이터의 레이블 정보를 이용하며 학습된 네트워크의 중간층에서 이진 해시코드가 생성된다. 반면, 비지도학습 기반 딥해싱은 클러스터 내 또는 클러스터 간의 비유사성을 최대화하는 손실 함수를 정의하여 이진 해시코드를 생성한다. 따라서 매핑된 해시코드는 입력 이미지의 저차원 표현과 유사한 이미지를 재구성한다[7].

본 논문은 변분 추론 오토인코더(Variational Autoencoder, VAE[8]), 합성곱 필터(Convolutional VAE, CVAE[9]), 트랜스포머(transformer[10]) 기반 변분 오토인코더(VTE)와 CSVAE (Simple Parameter-free Attention Module[11] based VAE, CSVAE) 등 딥해싱 모델을 제안하고 평가한다. 2절에서는 학습 패러다임에 따른 딥해싱 모델들을 살펴보고 비교한다. 3절에서 제안 딥해싱 모델의 구성과 학습 방법 등을 기술하며, 4절에서 딥해싱 모델의 일반화 성능을 비교 분석한다. 마지막으로 딥해싱 모델의 이미지 검색과 저작권 이미지 판별에서 효과적인 대안 가능성을 기술한다.

2. 관련 연구

지도학습 기반 딥해싱 모델은 데이터셋의 레이블에 가깝게 표현된 해시 벡터를 갖도록 학습되며, 비지도 학습 기반 딥해싱 모델은 이미지 자체의 구조에서 추출된 특징을 바탕으로 해시 벡터를 갖도록 학습된다. 따라서 딥해싱 알고리즘 구성 시 백본 모델의 선택과 지도, 비지도 학습 기반 딥해싱 모델의 비교를 통해 개선된 모델 설계가 가능하다.

VGG-F[12]를 백본(backbone)으로 사용하는 딥해싱 모델 DPN(deep polarized network)이 제안되었다[13]. 제안 모델은 해시층이 추가되었으며 손실함수는 연속 해시코드의 양자화 과정을 학습하도록 설계되었다. 생성된 해시코드는 데이터와 레이블의 매핑 시 나타나는 특징 정보가 반영된다. 유사하게, AlexNet[14]에서 추출된 이미지 특징을 두 개의 완전 연결층으로 전달해 태그 정보와 해시코드를 생성하는 WDHT가 제안되었다[15]. 각 입력 데이터는 여러 개의 태그 정보를 갖는 출력 데이터로 구성되어 있다. 태그층은 힌지 손실(hinge loss)을 사용하며 해시층은 양자화 손실과 쌍별 유사성 손실을 결합했다. DPN과 WDHT는 독립된 출력층과 해시층으로 구성되며 백본 네트워크를 통한 전이 학습으로 이용된다.

전이학습의 특징 벡터와 딥해싱 모델을 융합하는 접근 방법인 DSH[16], CSQ[17], HashNet[18] 등이 제안되었다. 딥해싱 모델은 해시코드 생성을 위해 사용된다. DSH는 3개의 CNN층으로 구성되며 두 개의 입력 이미지에 대해 해시코드 길이의 출력을 갖는다. 출력된 두 개의 해시 벡터 간 유사도 검증 손실을 통해 해시코드의 분별력을 높인다. CSQ는 CNN층에서 출력된 벡터의 중앙값을 이용한다. 유사한 입력 이미지에 대한 출력 벡터는 공통된 중앙값으로 매핑되며 다른 이미지들에 대해 상이한 중앙값으로 매핑된다. 그러므로 입력 이미지에 대해 지각적으로 유사한 이미지를 반환할 수 있는 분별력을 갖는 해시코드를 생성한다.

비지도 학습 기반 딥해싱 모델은 고차원 입력 데이터를 더 낮은 차원으로 매핑하며 입력과 출력이 동일한 데이터로 훈련된다. 이미지 딥해싱에서 인코더 및 디코더를 합성곱 계층으로 대체한 모델 SCA-H가 제안되었다[19]. 합성곱 계층은 입력 데이터의 공간 정보, 계층적 표현 등을 잠재 벡터로 매핑한다. 제안 모델은 NUS-WIDE[20]과 Flickr30k[21]으로 사전 학습 되었으며, 중간층의 잠재 벡터로부터 생성된 해시코드의 판별과 이진화 오류를 최소화 시키는 학습을 한다. 또한 인코더와 디코더를 구성하는 합성곱 계층을 특징 피라미드 형태(feature pyramid network, FPN[22])로 구성하고 이진 코드를 희소 벡터로 표현하는 연구가 진행되었다. FPN는 입력 데이터의 계층적 특징을 저차원 잠재 벡터로 매핑하고 희소 표현된 해시코드를 출력한다. 이진화를 위한 임계값은 잠재 벡터에서 이진 이완(binary relaxation)을 통해 결정된다. 학습된 임계값은 입력 데이터에 의해 조정되어 이진 해시코드의 품질을 향상시킨다.

지도학습 기반 딥해싱 모델은 해시코드를 학습하기 위해 레이블이 지정된 데이터가 필요하며 이를 통해 해시코드를 생성할 수 있다. 학습 모델은 백본 네트워크의 출력을 딥해싱 모델의 입력으로 사용하거나 해시층을 결합한 다중 출력 구조 등으로 구성된다. 반면, 비지도학습 기반 딥해싱 모델은 레이블이 지정된 데이터가 제한적이거나 사용할 수 없을 때 선택되며 입력 이미지 자체에서 데이터 구조를 추출하도록 학습된다. 또한 지도학습 기반 딥해싱 모델은 비지도 학습 기반 모델보다 대부분 높은 일반화 성능을 보이나 높은 데이터 라벨링 비용이 발생한다.

제안 모델은 데이터 라벨링 비용이 발생하지 않는 비지도 학습 기반 변분 추론 모델로, 인코더 및 디코더 구조를 갖는다. 인코더는 합성곱 신경망과 트랜스포머, SimAM 등으로 블록 단위 모델 구성이 용이하며, 이미지 검색 또는 이미지 판별에 표현 학습의 도입이 가능하다.

3. 제안 방법

데이터셋 $D=\left\{ X_i \in \mathbb{R}^{128\times128} \mid i=1,...,N \right\}$의 각 이미지는 128×128의 흑백 이미지이며 [0, 1]의 픽셀 값을 가지도록 스케일링(Scaling)되었다. 제안된 딥해싱 모델은 인코더와 디코더로 구성된다: $M=(M_e, M_d)$. 인코더 M_e는 Equation (1)과 같이 입력 이미지 $X_i \in D$의 저차원 잠재 벡터 Z_i로의 매핑을 학습하며, 디코더 M_d는 잠재 벡터 $Z_i : Z_i = M_e(X_i)$로부터 출력 이미지 $\hat{X}$: $\hat{X_i} = M_d(Z_i)$의 재구성 과정을 학습한다. 딥해싱 모델은 X_i와 $\hat{X_i}$의 차이를 최소화하도록 학습한다. 해시 함수 $Q(Z_i)$는 Equation (2)를 통해 이미지 X_i로부터 d크기의 해시코드 $h_i = [h_i^j]_d$를 생성한다. 임계값 θ_i는 잠재 벡터의 평균이며 Equation (3)으로 이진화시킨다.

$$\theta_i = \frac{1}{N}\sum_{i=1}^{N} Z_i = \frac{1}{N}\sum_{i=1}^{N} M_e(X_i) \tag{1}$$

$$h_i = Q(M_e(X_i)) = [h_1^i, h_2^i, ..., h_d^i] \tag{2}$$

$$h_j^i = \begin{cases} 1 \text{ if } Z_j^i > \theta_j \\ 0 \; otherwise \end{cases} \tag{3}$$

Variational autoencoder(VAE)[8]는 표준 오토인코더 모델에 확률적 구성 요소를 도입한 모델이다. 유사한 데이터를 잠재 벡터의 가까운 포인트에 매핑하도록 학습되며 입력 데이터의 분포와 잠재 벡터의 분포간 KL-Divergence를 최소화 하도록 훈련된다. VAE의 인코더는 입력 데이터를 확률 분포의 매개변수에 해당하는 잠재 벡터로 매핑하는 과정을 학습한다. 학습 모델은 완전 연결층 구조를 갖는 대칭 인코더 및 디코더로 구성된다. 해시 비트는 학습된 잠재 벡터의 원소에 해당하며 이진코드의 유사성은 이미지 간의 관계를 반영한다. 따라서 VAE 기반 딥해싱 모델은 입력 데이터에 강건한 해시코드를 추출할 수 있다.

Fig. 1은 이미지 딥해싱을 위한 VAE 모델 구조이다. VAE의 인코더는 4개의 완전 연결 계층으로 구성되며 입력 이미지의 평탄화 크기 이후 뉴런의 수가 반으로 감소한다. 디코더는 잠재 벡터의 크기부터 출력층까지 완전 연결 계층으로 구성된 층의 뉴런 수를 두 배 증가한다. 해시코드 h_i는 입력 이미지 X_i에 대한 분포 평균 층의 활성화 값 Z_i를 이용해 이진화시킨다.

완전 연결 계층으로 구성된 VAE의 입력은 이차원 이미지 데이터를 일차원 벡터로 평탄화한다. 완전 연결은 학습 모델의 계산량과 학습 매개변수를 증가시킨다. 이는 학습 모델의 과적합 양상으로 나타나 이미지 검색 성능을 저하하는 요인으로 작용한다. 또한 평탄화 과정은 공간 구조를 반영하지 않는 단일 픽셀을 독립 변수로 취급해 이미지 내 기하학적인 패턴과 추상화 특징을 학습하는데 제한될 수 있다.

CVAE는 VAE의 인코더 및 디코더 구성을 CNN으로 대체한 딥해싱 모델이다. CNN의 커널을 이용한 저수준과 고수준의 특징 추출과 추상화가 가능하며, 커널 파라미터 공유는 학습 파라미터의 수를 감소시킨다. 디코더의 업풀링층은 추상화된 특징이 입력 이미지와 동일한 크기의 출력을 낼 수 있도록 확장한다. CVAE의 CNN층은 2차원 이미지의 로컬 영역의 정보를 추출하고 저차원 벡터 공간으로의 매핑을 학습한다. 디코더는 벡터에서 입력 데이터를 재구성하는 역할을 하며 인코더에 의해 매핑된 저차원 특징으로부터 이미지를 재구성하기 위해 다수의 업샘플링 계층으로 구성된다.

Fig. 2는 CVAE 모델 구조이다. CVAE의 인코더는 여섯 계층의 CNN으로 구성되며 각 합성곱 필터의 크기는 2×2와 스트라이드(stride) 2로 구성된다. 패딩을 고려하지 않은 이유는 동일 이미지 크기를 유지하기 위해서이다. 디코더는 인코더의 대칭 구조를 가지며 여러 개의 업샘플링 층으로 구성된다(Fig. 2).

비전 트랜스포머(Vision Transformer, ViT)는 셀프 어텐션 메커니즘을 사용해 독립된 패치 간의 전역 종속성을 반영하는 모델로 다중 헤드 어텐션(multi-head attention)을 통한 높은

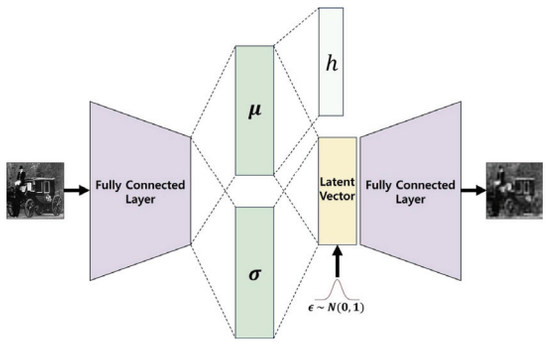

Fig. 1. VAE Based Deephashing Model

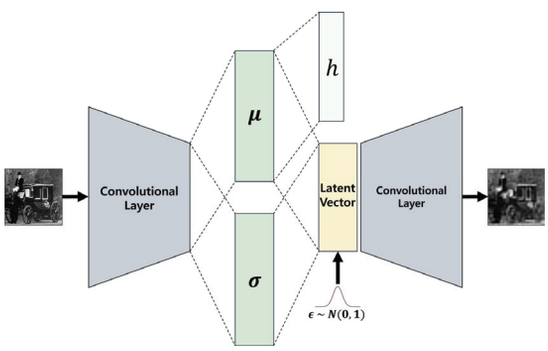

Fig. 2. CVAE Based Deephashing Model

병렬성을 지원한다. ViT는 이미지 내 포함된 객체들의 위치적 관계를 학습할 수 있다. 입력 이미지의 패치에서 위치 정보를 포함한 전역 특징을 학습할 수 있으며 전역적인 맥락을 고려해 입력 이미지를 잠재 변수 공간으로 매핑한다. 이를 통해 입력 이미지 내 대표 객체 영역은 잠재 변수 공간에서 배경과 구분되도록 매핑되며, 이미지 변형 등에 강직한 해시코드를 생성할 수 있다.

Fig. 3은 VTE 모델의 구조이다. VTE는 표준 ViT를 인코더로 사용한 VAE 모델이다. 학습된 ViT의 출력은 입력 이미지의 저차원 표현이며 VAE의 잠재 벡터이다. 디코더는 CVAE의 디코더로 구성된다.

제안 모델은 SimAM 모듈을 사용한 VAE 모델이다. SimAM [10]은 뇌과학 이론을 바탕으로 뉴런과 주변 뉴런의 활성화 값들을 이용해 정의한 에너지 함수 기반 어텐션 메카니즘을 적용한다. 활성 뉴런은 하나의 대상 뉴런과 이웃 뉴런 사이의 선형 분리가 가능한 에너지 함수로 표현된다. 그러므로 학습 시 두드러지는 객체 또는 로컬 영역이 강조되는 효과를 제공한다.

SimAM의 도입으로 잠재 벡터는 추상화된 특징 간 선형 분리된 저차원 벡터를 표현하며 입력 이미지 내 대표 객체 영역을 두드러지게 하는 효과를 제공한다. 따라서 대표 객체에 대한 특징 벡터가 높게 반영되며 노이즈로 간주되는 배경에 대한 중요도가 낮게 매핑되어 대표 객체에 강조된 표현을 갖는 해시코드를 생성할 수 있다. Fig. 4는 제안 모델의 구조이다. CNN 계층과 동일한 크기의 SimAM 모듈은 CVAE 인코더의 두 번째 및 세 번째 계층에 삽입된다. 디코더 구조는 CVAE와 동일하게 여섯 계층의 업샘플링 계층으로 구성된다.

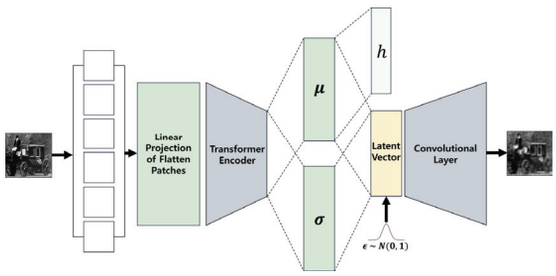

Fig. 3. Transformer Based Deephashing Model

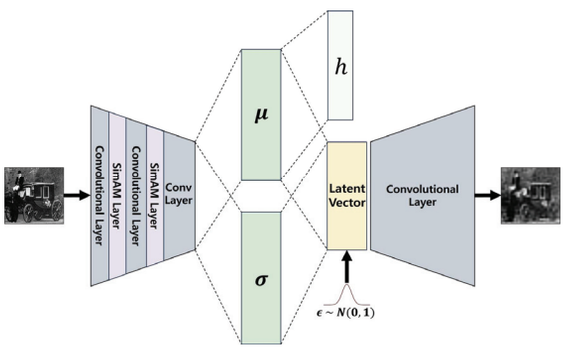

Fig. 4. SimAM Module Based Deephashing Model

4. 실험 결과

제안 모델의 평가를 위해 공개 데이터셋인 CIFAR-10[23], ImageNet[20]과 NUS-WIDE[21]를 사용했다. 학습 데이터는 균등 무작위 샘플링을 통해 학습 데이터셋 20,000장과 평가 데이터셋 10,000장으로 구성된다. 평가 데이터셋은 검색 데이터베이스 내 이미지 5,000장과 쿼리 이미지 5,000장이며 검색 데이터베이스는 원본 이미지, 해싱 모델에서 추출된 이진 해시코드와 레이블이 저장된다. 검색은 쿼리 이미지의 해시코드와 검색 데이터베이스 내 가장 낮은 해밍 거리를 갖는 이미지를 선택한다. 각 알고리즘의 학습 에폭 수는 2,000이며 배치 크기는 64이다. 학습은 코사인 어닐링 스케쥴러(cosine annealing scheduler)가 적용했으며 Adam 옵티마이저(optimizer)를 이용한다.

모델 평가 지표는 쿼리 이미지의 레이블과 검색 이미지의 레이블의 정밀도(precision), 재현률(recall), F1-score, mAP 등이다. mAP는 임계값 변화에 따른 평가 데이터셋의 정밀도-재현률 곡선 아래 영역의 평균이며 해당 클래스의 평균 정밀도이다. mAP 값이 1에 근접할 수록 딥해싱 성능이 더 높다고 분석한다.

해시 충돌(hash collision)은 모델에서 두 개의 다른 이미지가 동일 해시코드를 생성시켜 이미지 검색 시스템의 성능 저하의 원인이 된다. 해시 충돌률의 계산식은 Equation (4)와 같다.

$$Collision\,rate = \frac{No.\,of\,collided\,hashes}{No.\,of\,images} \tag{4}$$

Table 1은 평가 데이터셋에 대한 검색 성능 테스트 결과이다. 평가 지표는 16, 32, 64, 128 비트 크기의 해시코드에서 계산되었다. 평가에서 VAE 모델은 CVAE, 제안 모델과 VTE 모델보다 약 0.5 가량 낮은 mAP 지표를 보였다. 이미지 공간 정보 유실, 제한된 수용 필드 등 이미지 평탄화 시 발생하는 요인으로 인해 낮은 성능이 나타났다.

CNN으로 구성된 CVAE와 제안모델은 입력의 공간 정보를 유지하며 학습되어 향상된 품질의 해시코드를 생성했으며 공간 구조 및 정보 반경과 더불어 적은 수의 CNN 커널 파라미터가 도움 되었다. VTE의 어텐션 구조 또한 공간 구조를 반영해 CNN 계열의 딥해싱 모델과 유사한 성능을 보인다. VAE, CVAE, 제안 모델과 VTE 모델에 대해 64 비트 해시 공간에서 가장 높은 성능이 나타났다. Table 2는 제안된 딥해싱 모델의 해시 충돌률 비교이다. 10,000개의 ImageNet 테스트에서 VTE 모델은 해시 크기 128

Table 1. Performance Evaluation of the Proposed Deephashing Model

Method		CIFAR10@10000				ImageNet@10000				NUS-WIDE@10000			
		16b	32b	64b	128b	16b	32b	64b	128b	16b	32b	64b	128b
VAE	F1	0.624	0.647	0.662	0.647	0.543	0.532	0.662	0.647	0.537	0.555	0.564	0.583
	Precision	0.607	0.647	0.663	0.647	0.512	0.534	0.663	0.647	0.538	0.555	0.565	0.584
	Recall	0.608	0.648	0.664	0.648	0.527	0.515	0.664	0.648	0.536	0.554	0.563	0.587
	mAP	0.410	0.455	0.475	0.456	0.319	0.394	0.475	0.456	0.348	0.356	0.407	0.425
CVAE	F1	0.895	0.980	0.979	0.978	0.850	0.835	0.836	0.837	0.869	0.873	0.873	0.873
	Precision	0.896	0.976	0.979	0.978	0.832	0.819	0.820	0.822	0.844	0.867	0.864	0.879
	Recall	0.894	0.981	0.978	0.978	0.894	0.870	0.870	0.872	0.884	0.890	0.884	0.869
	mAP	0.812	0.962	0.960	0.959	0.807	0.819	0.820	0.822	0.810	0.821	0.822	0.815
VTE	F1	0.888	0.970	0.979	0.977	0.854	0.832	0.824	0.833	0.869	0.873	0.880	0.882
	Precision	0.890	0.954	0.979	0.976	0.846	0.814	0.814	0.819	0.844	0.867	0.881	0.881
	Recall	0.874	0.979	0.980	0.979	0.897	0.872	0.830	0.861	0.884	0.890	0.878	0.883
	mAP	0.826	0.953	0.959	0.959	0.810	0.816	0.820	0.818	0.809	0.813	0.823	0.823
Proposed	F1	0.925	0.979	0.981	0.982	0.851	0.840	0.842	0.843	0.870	0.872	0.871	0.873
	Precision	0.924	0.974	0.980	0.981	0.833	0.821	0.820	0.821	0.845	0.866	0.865	0.880
	Recall	0.874	0.926	0.983	0.982	0.870	0.859	0.868	0.868	0.885	0.889	0.886	0.862
	mAP	0.842	0.966	0.969	0.970	0.784	0.820	0.821	0.822	0.809	0.821	0.824	0.824

Tabel 2. Hash Collision Ratio Comparison

Method	Collision Ratio(%)											
	CIFAR@10000				ImageNet@10000				NUS-WIDE@100000			
	16	32	64	128	16	32	64	128	16	32	64	128
VAE	0.34	0.32	0.29	0.27	0.39	0.37	0.33	0.33	0.36	0.29	0.26	0.30
CVAE	0.28	0.23	0.21	0.21	0.24	0.16	0.14	0.13	0.26	0.18	0.18	0.15
VTE	0.24	0.23	0.21	0.21	0.22	0.15	0.13	0.11	0.23	0.20	0.18	0.15
Proposed	0.23	0.23	0.21	0.21	0.22	0.23	0.15	0.13	0.13	0.22	0.19	0.16

Table 3. Comparative Results of Deephashing Models in Related Studies

Method	CIFAR@10000			ImageNet@10000			NUS-WIDE@100000		
	16b	32b	64b	16b	32b	64b	16b	32b	64b
DHN	0.693	0.645	0.588	0.311	0.482	0.573	-	0.748	-
HashNet	0.748	0.778	0.626	0.506	0.631	0.684	0.622	0.699	0.716
DPN	0.744	0.803	0.812	0.608	0.691	0.727	0.810	0.822	0.839
TransH	0.908	0.911	0.917	0.920	0.932	0.933	0.726	0.739	0.749
Proposed	0.842	0.966	0.969	0.784	0.820	0.821	0.809	0.821	0.824

비트에서 0.11%(11장)의 해시 충돌이 발생했으며 제안 모델, CVAE, VAE과 비교 시 각각 0.01%(1장), 0.02%(2장), 0.2%(20장) 낮은 충돌률이다. NUS-WIDE 데이터셋을 통해 4바이트 환경에서 해시 충돌률을 비교했다. VAE 모델의 해시 충돌률은 CVAE, 제안 모델, VTE에 대해 각각 0.12%(12장), 0.28%(28장), 0.28%(28장)으로 높은 충돌률을 보인다. 동일한 해시 크기에서 VTE 모델을 제외한 모델의 해시 충돌률은 비슷하다.

연구된 딥해싱 모델과 비교를 위해 실험이 진행되었다. 기존에 연구된 모델은 DHN[24], HashNet[18], DPN[13]과 TransH[25] 등이다. Table 3은 연구된 모델과 제안 모델의 성능 비교이다. 제안된 모델 중 가장 높은 성능을 보인 제안 모델과 기존에 연구된 지도 학습 기반 해싱 알고리즘의 mAP 비교이다. 제안된 딥해싱 모델은 높은 성능을 보여 간단한 모델 재구성만으로 분별력이 높은 해시코드를 생성할 수 있는 것으로 기대된다. ImageNet 데이터셋에서 TransH는 제안 모델보다 약 0.112 높은 mAP 지표를 보인다. CIFAR-10과 Nus-Wide에서 각각 0.052와 0.075의 낮은 mAP 지표가 나타났다.

지도학습 모델 DPN과 비교 시 CIFAR-10과 ImageNet 데이터셋에 대해 64비트의 해시에서 각각 0.157, 0.094 높은 mAP를, NUS-WIDE 데이터셋에 대해 0.069 낮은 mAP를 보였다. 제안 모델은 전반적으로 DHN, HashNet, DPN 등 지도학습 기반 알고리즘보다 우수한 성능을 보인다.

5. 결 론

본 논문에서 이미지 검색을 위한 비지도 변분 추론 기반 딥해싱 모델이 제안되었다. 이미지 인식에서 널리 사용되는 신경망, CNN, 트렌스포머 구조를 적용해 오토 인코더 형태로 구성했으며 알고리즘들은 종단간(end-to-end) 학습이 가능하다. 제안 모델의 테스트는 CIFAR-10, ImageNet, NUS-WIDE 등에서 수행되었으며, 지도 학습 기반 딥해싱 모델과 비슷한 일반화 성능을 보였다. 딥해싱 모델로부터 생성된 해시코드는 입력 이미지의 특징 벡터를 포함할 뿐만 아니라 데이터셋이 가지는 높은 차원의 분포를 저차원으로 매핑이 가능하다. 따라 지각적으로 유사한 이미지가 특정 영역에서 군집화되며, 추출된 해시코드는 이미지가 가지는 변형에 강건하게 구분될 수 있다. 이미지 해싱에 대한 접근 방식은 요구사항, 사용 가능한 리소스, 데이터셋 등에 따라 다르게 선택된다. 실험 결과로부터 제안 모델은 저차원의 표현 학습이 가능하여 이미지 검색 또는 저작권 이미지 판별 등 응용 시스템에 활용될 수 있다.

References

[1] L. Du, A. T. Ho, and R. Cong, "Perceptual hashing for image authentication: A survey," *Signal Processing: Image Communication*, Vol.81, pp.115713, 2020.

[2] J. Mao, D. Zhong, Y. Hu, W. Sheng, G. Xiao, and Z. Qu, "An image authentication technology based on depth residual network," *Systems Science & Control Engineering*, Vol.6, No.1, pp.57-70, 2018.

[3] D. Kim, S. Heo, J. Kang, H. Kang, and S. Lee, "A photo identification framework to prevent copyright infringement with manipulations," *Applied Sciences*, Vol.11, No.19, pp. 9194, 2021.

[4] S. Zhu, C. Zhu, and W. Wang, "A new image encryption algorithm based on chaos and secure hash sha-256," *Entropy*, Vol.20, No.9, pp.716, 2018.

[5] L. W. Kang, C. S. Lu, and C. Y. Hsu, "Compressive sensing-based image hashing," in *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, pp.1285-1288, 2009.

[6] K. P. Murphy, "Probabilistic Machine Learning: An introduction," MIT Press, 2022.

[7] G. E Hinton, A. Krizhevsky, and S. D. Wang, "Transforming auto-encoders," in *Artificial Neural Networks and Machine Learning-ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo*, Finland, June 14-17, 2011, Proceedings, Part I 21. Springer, pp.44-51, 2011.

[8] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *2nd International Conference on Learning Representations*, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings, 2014.

[9] X. Chen, Y. SUN, M. Zhang, and D. Peng, "Evolving deep convolutional variational autoencoders for image classification," *IEEE Transactions on Evolutionary Computation*, Vol.25, No.5, pp.815-829, 2020.

[10] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[11] L. Yang, R. Y. Zhang, L. Li, and X. Xie, "Simam: A simple, parameter-free attention module for convolutional neural networks," in *Proceedings of the 38th International Conference on Machine Learning,* Marina Meila and Tong Zhang, Eds. 18-24 Jul 2021, Vol. 139 of Proceedings of Machine Learning Research, pp.11863-11874, PMLR.

[12] H. Venkateswara, J. Eusebio, S. Chakraborty, and S. Panchanathan, "Deep hashing network for unsupervised domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.5018-5027, 2017.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, Vol.25, 2012.

[14] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[15] V. Gattupalli, Y. Zhuo, and B. Li, "Weakly supervised deep image hashing through tag embeddings," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.10375-10384, 2019.

[16] H. Liu, R. Wang, S. Shan, and X. Chen, "Deep supervised hashing for fast image retrieval," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[17] Z. Cao, M. Long, J. Wang, and P. S. Yu, "Hashnet: Deep learning to hash by continuation," *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

[18] S. R. Dubey, S. K. Singh, and W. T. Chu. "Vision transformer hashing for image retrieval," *2022 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2022.

김 수 인

https://orcid.org/0009-0004-1695-6223
e-mail : kimsooinj7740@gmail.com
2023년 단국대학교 소프트웨어학과(학사)
2023년 ~ 현 재 단국대학교
인공지능융합학과 석사과정
관심분야 : 인공지능, 딥러닝, 영상처리

전 영 진

https://orcid.org/0009-0007-7423-9240
e-mail : yjjeon@aideep.ai
2017년 단국대학교 컴퓨터과학과(학사)
2019년 ~ 현 재 (주)에이아이딥 책임연구원
2023년 ~ 현 재 단국대학교
인공지능융합학과 석사과정
관심분야 : 블록체인, 모바일 보안

이 상 범

https://orcid.org/0009-0003-3801-2285
e-mail : sblee@dankook.ac.kr
1983년 한양대학교(학사)
1989년 Louisiana State Univ(석사)
1992년 Louisiana State Univ
소프트웨어공학(박사)
현 재 단국대학교 소프트웨어학과 정교수
관심분야 : 데이터 베이스, 소프트웨어 공학, 유무선 인터넷

김 원 겸

https://orcid.org/0000-0003-3022-6230
e-mail : wgkim@aideep.ai
1992년 충남대학교 전산학과(학사)
1994년 충남대학교 전산학과(석사)
2001년 충남대학교 컴퓨터과학과(Ph.D.)
1995년 ~ 1997년 (주)SK하이닉스 대리
2002년 ~ 2006년 한국전자통신연구원(ETRI) 선임연구원
2007년 ~ 2008년 마크애니(주) 부장
2009년 ~ 2016년 (사)한국저작권단체연합회 저작권보호센터 팀장
2016년 ~ 현 재 (주)에이아이딥 연구소장
관심분야 : 인공지능, 딥러닝, 영상처리, 스마트홈 IoT