

스케치 기반 이미지 검색의 최신 연구 동향

Recent advances in sketch based image retrieval: a survey

오 세 홍*, 석 호 식**

Sehong Oh*, Ho-Sik Seok**

Abstract

A sketch is an intuitive means to express information, but compared to actual images, it has the problem of being highly abstract, diverse, and sparse. Recent advances in deep learning models have made it possible to discover features that are common to images and sketches. In this paper, we summarize recent trends in sketch-based image retrieval (SBIR) but it is not limited to SBIR. Besides SBIR, we also introduce sketch-based image recognition and generation studies. Zero-shot learning enables models to recognize categories not encountered during training. Zero-shot SBIR methods are also discussed. Commonly used free-hand sketch datasets are summarized and retrieval performance based on these datasets is reported.

요 약

스케치는 직관적으로 정보를 표현할 수 있는 수단이나 실제 이미지와 비교하였을 때 추상화가 심하고 동일한 대상에 대한 스케치 임에도 변화가 매우 크다는 문제점이 있다. 따라서 스케치에 기반한 이미지 인식에 도움이 되는 특성을 표현하기 어려웠는데, 딥뉴럴 모델의 발전은 이미지와 스케치라는 두 도메인에 공통으로 존재하는 특성을 발견할 수 있게 해 주었다. 본 논문에서는 스케치에 기반한 이미지 검색 기술의 최신 연구 동향을 소개한다. 이미지 검색 외에 스케치에 기반한 이미지 인식과 이미지 생성의 연구 동향도 함께 요약하였으며, 스케치 기반 연구에서 주로 활용되는 데이터셋 및 해당 데이터셋에 기반하여 측정된 검색 성능을 포함하였다.

Key words : sketch recognition, sketch based image retrieval, zero-shot learning, image generation, image recognition

1. 서론

인간은 스케치의 형태로 주어진 정보에서 원본의 이미지를 쉽게 재생성할 수 있다. 그러나 스케치에 기반하여 원본의 이미지를 검색하거나[1], 사진 수준의 정밀한 이미지를 생성하는 작업[2]은 현재 만족할만한 성능을 달성하지 못한 상태이다. 이것은 스케치가 나타내는 정보와 원본 사진이 표현하는 정보를 단순하게 비교하여 일

치 여부를 판단할 수 없기 때문이다. 그러나 스마트폰과 태블릿의 보편화로 다양한 형태의 스케치가 생성되고 활용되면서, 정보 처리 과정에서 스케치의 중요성은 증대하고 있으며 그 결과 기계 학습 기술을 적용하여 스케치를 분석하려는 연구가 활발히 진행되고 있다.

기계학습/패턴 인식 관점에서 살펴보면 스케치 기반의 정보는 동일한 객체의 표현이 작성자마다 틀리고, 배경이 없으며, 매우 추상적으로 표현된다는 특징으로 인하

* (Lecturer, Assistant Professor) Dept. of Artificial Intelligence and Data Science, Korea Military Academy

★ Corresponding author

E-mail : hosik.seok@gmail.com, Tel : +82-2-2197-2873

※ Acknowledgment

This study was supported by research fund of Korea Military Academy. (Future Strategy and Technology Research Institute). (RN: 24-AI-Center-03).

Manuscript received Jun. 10, 2024; revised Jun. 20, 2024; accepted Jun. 25, 2024.

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

여 기존의 이미지 분석 방법과 다른 독자적인 처리 방법을 요구한다[3].

본 논문에서는 스케치 기반 이미지 검색(Sketch-based image retrieval, SBIR)의 최신 기술 수준을 소개하고, 주요 데이터셋의 특징을 정리하여 SBIR 문제를 해결하려는 연구자들에게 도움을 제공하고자 한다. SBIR의 주요 접근법 외에도 스케치 기반 이미지 생성이나 객체 인식의 최신 접근법을 함께 정리하였다.

II. 스케치 관련 이미지 분석 연구

스케치는 상대적으로 간단한 표현을 통해 풍부한 정보를 전달할 수 있는 의사소통 방법이다. 따라서 스케치를 활용한 이미지 검색에 대하여 활발한 연구가 이루어지고 있다[4-6]. 딥러닝 모델의 놀라운 발전은 SBIR에도 큰 영향을 주고 있으므로, 먼저 딥러닝에 기반하여 스케치 이미지를 분석하려는 다양한 접근법을 살펴본 후 SBIR의 주요 접근법을 살펴보도록 한다.

1. 스케치 기반 인식

텍스처와 색상 정보가 존재하지 않는다는 스케치의 특성은 일반적인 이미지 분석 기법의 적용을 어렵게 한다. 실제 이미지와 스케치 이미지 사이에 공통적으로 존재하는 은닉 구조(latent structure)를 분류에 이용하기 위하여 실제 이미지에서 도출한 특성과 스케치에서 도출한 특성 중 공통되는 특성을 컨볼루션 뉴럴넷으로 찾으려는 시도가 보고되었다[7].

일반적인 이미지 인식에서는 색상이나 질감과 같은 외형적인 특징을 고려하는 것으로 충분한 성능을 달성할 수 있으나, 스케치 인식에서는 외형(appearance)과 모양(shape)을 함께 분석하는 것이 도움이 된다는 결과가 보고되었으며 이 접근법은 2개의 CNN을 결합한 하이브리드 넷의 형태로 구현되었다[8].

스케치 이미지는 상대적으로 부족하기 때문에 실제 이미지에서 학습한 특성을 스케치 이미지 인식에 전이하려는 접근법도 존재한다[9]. 또한 제안 방법에서는 학습 과정에서 특성에 대하여 적대적 학습을 적용하여 특성 학습을 시도하였다.

트랜스포머 구조를 이용하여 스케치의 구조적 특성을 분석할 수 있도록 오토인코더로 은닉 벡터를 추출한 후 트랜스포머를 적용한 네트워크도 발표되었다[10].

스케치 이미지 인식에서 중요한 역할을 하는 획(stroke)을 분석하는 과정에서 기존의 CNN 기반 모델이 취약점

을 가지고 있다는 관찰은, 어텐션(attention) 기제를 활용하여 중요한 획에 집중하는 네트워크 구조의 제안으로 이어졌다[11].

스케치에 기반하여 객체를 인식하는 접근법도 소개되었다. [12]에서는 RPN(region proposal network) 모델에 기반하여 이미지에서 질의로 주어진 스케치에 해당하는 객체를 찾는 방법을 소개하였다. 제안된 방법에서는 어텐션 모듈을 훈련시켜 질의로 주어진 스케치와 유사한 이미지 영역에 높은 스코어를 부여하는 가중치 맵을 생성하였다.

기반 모델(CLIP)과 SBIR 접근법을 결합하여 바운딩 박스나 클래스 레이블 없이 객체를 인식하려는 접근법도 보고되었다[13]. 모델의 훈련 과정에서 SBIR은 객체 단 위에서 학습이 진행되었고 객체 인식은 이미지 수준에서 발생한다는 난점이 발생하는데, 이의 해결을 위해 n ($1 \leq n \leq 7$)개의 객체 수준 사진을 무작위로 배치한 캔버스를 생성한 후 객체와 이에 대응하는 스케치를 쌍으로 묶어 서포트 셋(support set)을 구성하였다. 서포트 셋을 이용하여 카테고리 및 인스턴스 수준에서 객체가 존재하는 영역을 탐색할 수 있는 능력을 학습하였다.

비디오 영상에서 스케치가 표현하는 객체를 찾는 연구도 소개되었다[14]. 해당 연구에서는 ResNet-50(비디오 프레임 처리용)과 ResNet-18(스케치 처리용)을 백본으로 한 후 Sketch-Video Cross-Attention, Content Self-Attention, Token Self-Attention, Content-Token Cross Attention의 연산을 수행하는 크로스-모달 트랜스포머를 결합하여 스케치 정보와 비디오 프레임의 정보를 연관시키려고 시도하였다.

2. 스케치와 생성

스케치와 관련된 생성은 실제 객체에 기반하여 스케치를 생성하는 접근과 스케치에 기반하여 실제 객체의 이미지를 생성하는 접근으로 나누어 볼 수 있다.

객체에 기반한 스케치 생성은 이미지에서 이미지로 변환하는 문제로 접근할 수 있다. [15]에서는 감독학습과 무감독학습을 결합한 접근법을 통해 스케치를 분석할 때 발생하는 실제값의 불명확성을 처리하였다. 제안 모델은 사진을 스케치로 변환하는 감독 서브 모델, 스케치를 다시 사진으로 매핑하는 감독 서브 모델, 사진을 재생성하는 무감독 서브 모델, 스케치를 재생성하는 무감독 서브 모델로 구성되었다. 실제값의 불명확성 문제는 손실 함수에 사이클 일관성(cycle-consistency)[16]의 개념을 도입함으로써 해결하였다.

Sketch-RNN은 인간이 생성한 스케치에 기반하여 새로운 스케치를 생성하는 모델이다[17]. 해당 모델은 Sequence-to-sequence VAE(variational autoencoder)로 스케치를 입력으로 받아 은닉 벡터를 출력한다. Sketch-RNN은 스케치 시퀀스와 해당 시퀀스의 역시퀀스를 모두 활용하였다.

주어진 텍스트에 해당하는 이미지를 생성하는 딥러닝 모델은 활발한 연구 대상이다. Sketchforme는 텍스트 캡션이 부여된 데이터셋으로부터 스케치 조합의 원칙을 학습한 후 구체적인 객체를 생성하는 접근법을 취하여 텍스트에 해당하는 스케치를 생성하였다[18].

DiffSketcher는 주어진 텍스트에 대응하는 스케치를 생성하는 과정에서 Bezier 커브를 활용함으로써 벡터 이미지 방식의 스케치를 생성할 수 있었다. 또한 획의 초기화 과정에서 디퓨전 모델의 어텐션 맵을 활용하여 스케치 생성의 효율성을 향상하였다[19].

Visuomotor 메모리 측면에서 스케치 생성 프로세스를 학습한 접근법도 존재한다[20]. 제안 모델에서는 이미지와 스케치 생성 동작을 모두 특성으로 간주하여 학습을 진행하였다.

DoodlerGAN은 인간의 스케치를 합성하는 것이 아니라 기존에 없었던 새로운 스케치를 생성하는 것을 목표로 하였다. 해당 방법에서는 객체를 구성하는 부분을 생성한 후 생성된 부분을 조합하는데, StyleGAN2 구조에 기반하여 부분을 생성한 후(part generator), 다음에 생성될 파트 카테고리를 선택하는 과정(part selector)을 반복하여 스케치를 생성한다[21].

짧은 직선의 연속으로 스케치를 생성하면 스케치의 확장성에 문제가 발생한다. Cloud2Curve 모델은 Bezier 커브의 연속으로 스케치를 해석하였다[22]. 제안 모델에서는 디코더가 Bezier 커브의 파라미터를 생성함으로써 벡터 이미지 방식의 스케치를 생성할 수 있었다.

스케치에 존재하는 획의 끝점(stroke point)의 분포를 학습하여 스케치를 생성하는 접근법도 소개되었다[23]. 제안 모델은 디퓨전 모델(Diffusion model)의 포워드/백워드 단계를 개선하여 생성 속도를 향상하였다. 디퓨전 모델의 분포 학습을 스케치의 끝점 분포 학습에 적용함으로써 생성된 스케치의 식별 가능성(recognizability)을 수치화할 수 있었다.

스케치 생성과 더불어 주어진 스케치로부터 고품질의 이미지를 생성할 수 있는지 여부도 중요한 주제이다.

[2]에서는 기본적으로 StyleGAN[24]과 ResNet50[25]을 활용하여 고품질의 이미지를 생성한다. 그러나 은닉

코드의 예측 과정에서 은닉 코드간의 순차적 의존성을 유지하도록 모델링함으로써, 입력으로 주어진 스케치가 생성될 이미지의 대략적 구조에 영향을 주고, 세밀한 사항은 StyleGAN에 학습된 분포로부터 샘플링 하도록 구현할 수 있었다.

스케치에 기반하여 3D로 얼굴 표정을 생성하는 방법도 소개되었다[26]. 해당 방법에서는 2D 스케치로부터 3D 얼굴 모습을 묘사할 수 있도록, StyleGAN2[27]를 이용하여 입력으로부터 세 개의 직교 평면 특성을 생성하였다. 2D 이미지와 3D 형상의 매핑 관계를 바로 학습하기 보다는 StyleGAN2가 생성한 직교 평면 특성을 EG3D[28]의 은닉 변수 공간에 프로젝션하여 결과물의 품질을 개선하였다.

텍스트에 기반하여 이미지를 생성하는 과정에서 공간적 특성의 조합이 자연스럽게 않은 문제가 지적된다. [29]에서는 은닉 변수의 매핑 과정을 제어하는 다층신경망 모델을 추가하여 캡션과 함께 주어진 스케치로부터 이미지를 생성하는 모델을 소개하였다.

스케치를 이용하여 3D 이미지를 편집하는 모델은 텍스트만으로 명시하기 어려운 이미지의 특성을 스케치라는 직관적인 방법으로 표현할 수 있도록 해주었다. 제안 방법에서는 3D 이미지 표현을 위하여 NeRF(Neural Radiance Field)를 사용하였으며, 손실함수의 설계를 통해 뉴럴 필드의 편집 과정에서 기본 객체의 특성이 유지되면서 스케치로 추가된 특성이 반영되도록 하였다 [30].

III. 스케치 기반 이미지 검색

1. 전반적인 연구 동향

스케치에 기반한 이미지 검색은 꾸준한 관심을 받아온 주제이다[31, 32]. 스케치는 상대적으로 간단하고 풍부한 정보를 전달할 수 있는 수단인 반면, 스케치를 기반한 이미지 검색은 기존 이미지 기반 검색과 다른 새로운 접근법을 요구한다. 검색 매체 관점에서 분석하였을 때 스케치 기반 이미지 검색의 어려움은 다음의 문제들에 의해 발생한다: (1) 일반적인 이미지와 비교했을 때 특징 추출이 상대적으로 어려우며 (2) 도메인 교차 검색의 상황이 발생하고 (3) 스케치 이미지를 구하기가 쉽지 않으며 (4) 편리한 사용자 인터페이스 구축이 어렵다[32]. 스케치에 기반하여 스케치를 검색하는 응용[33]도 흥미롭지만 본 논문에서는 스케치에 기반하여 실제 이미지를 검색하는 접근법에 집중한다.

어텐션(attention)을 적용하여 차별성을 부여할 수 있는 특정 지역에 집중하면서, 스케치의 특성 반영을 위하여 숏컷 연결(shortcut connection) 구조를 반영한 접근법이 소개되었다[34]. 스케치를 이용하여 이미지를 검색하는 모델에서 특성 벡터가 원소간 정렬되었다고 가정하는 경우가 많으나, 이런 가정이 성능 저하를 초래하기도 한다. 제안 방법에서는 학습할 수 있는 고차 에너지 함수를 제안하여 정렬의 문제를 해결하고자 하였다.

스케치에 기반하여 이미지를 검색하고자 할 때 스케치와 도메인이 공유하는 특성을 탐색하고자 하지만, 스케치에 존재하는 변화와 특성 생략으로 인하여 공유 특성에 기반한 검색 방법의 성능은 그다지 우수하지 못하다. [35]에서는 DSH(deep sketch hashing)라는 방법을 제안하였다. DSH에서는 먼저 이미지에서 “스케치 토큰”을 획득한다. 이미지와 스케치 토큰을 위한 네트워크가 별도로 존재하며, 스케치의 경우 스케치 토큰과 가중치를 공유하는 구조가 사용되었다. 손실함수 계산 과정에서는 스케치와 이미지를 이진 분류 문제로 취급하면서 발생하는 문제를 해결하기 위해 워드 임베딩 기법을 응용하였다.

주어진 스케치 질의로부터 추가적인 콘텐츠를 추정하여 검색에 활용하려는 접근법이 제안되었다[36]. cGAN(conditional GAN)에 기반하여 콘텐츠를 생성하였으며 2개의 인코더를 이용하여 일관성을 유지하였다.

사전 획득한 스케치 표현 요소와 재순위 조정 기법을 결합하여 스케치 기반 이미지 검색을 시도하는 접근법은 근접한 이미지를 검색하는 기존 접근법을 개선한 방법이다[37]. 제안 방법에서는 SBIR 시스템의 결과를 받아 k개의 클러스터로 나눈 후 클러스터 구성 요소의 유사도 점수를 평균하여 클러스터에 순위를 부여하였다. 클러스터 구성요소는 유사도 점수에 기반하여 정렬되었다.

스케치와 이미지에 포함된 정보를 계층별로 분석하기 위하여, 먼저 스케치와 이미지를 여러 계층의 시각적 표현으로 분할한 후 계층별로 신경망을 학습하여 특징을 추출하고, 추출된 특징에 기반하여 유사도를 측정하여 이미지를 검색하는 방법이 제안되었다[38].

계층별로 학습을 진행하는 동시에 가중치 공유가 검색 성능에 미치는 영향을 확인한 연구도 존재한다[39]. 해당 연구에서는 CNN의 상위층을 공유하여 이미지와 스케치의 시맨틱을 조율하고자 시도하였다. 이미지와 스케치를 담당하는 별도의 네트워크 브랜치를 먼저 독립적으로 훈련한 후, 공유되지 않는 레이어의 가중치를 고정된 상태에서 공유 레이어를 학습하였으며 다시 모든 레이어의 학습을 진행하는 접근법을 취하였다.

스케치와 이미지의 상이한 모달리티를 다루기 위하여 이미지를 스케치와 유사한 형태로 변환한 후 검색을 시도하는 방법이 개발되었다[40]. 비주얼 특성 탐색과 학습을 위해서는 BoVW(bag of visual word) 프레임워크가 선택되었으며 지역 컨볼루션 특성 자체의 탐색에는 CKN(convolutional kernel network)를 사용하였다.

동일한 대상에 대한 스케치에는 상당한 변화가 존재한다. 이러한 변화에 대응할 수 있는 랜드마크를 학습하는 방법이 제안되었다[41]. 랜드마크는 객체의 에지로부터 샘플링된 키포인트의 위치에 오프셋을 더하여 키포인트를 조정하는 방법으로 발견되었는데, 오프셋의 결정 과정에서는 스케치 이미지의 랜드마크를 특성 맵으로 프로젝션하는 방법을 사용하였다.

스마트 기기 앱의 동작을 검증하는 과정에서 인터페이스 구성 요소의 검색은 생각 외로 쉽게 진행되지 않는다. Swire는 인터페이스 구성 요소 검색을 위한 모델로 유사한 스케치-이미지 페어에 대해서는 유사한 출력을, 상이한 스케치-이미지 페어에 대해서는 상이한 출력을 생성하도록 훈련되었으며 손실 함수는 트리플렛(Triplet) 손실 함수가 채택되었다[42].

스케치에서의 특성 파악은 일반적인 이미지의 특성 파악과 다른 접근법을 요구한다. [8]에서는 외형과 모양 모두 고려하는 모델을 제안하여, 외형 특성에 집중된 기존 모델의 한계를 극복하려고 하였다. 모양 학습의 경우 기하학적 변환에 대한 강인성이 요구되는데, 제안 모델에서는 아핀 변환을 예측하는 네트워크를 포함하여 강인성을 갖추고자 하였다.

전이학습은 스케치와 이미지라는 유사하면서도 상이한 모달리티에 적용할 수 있는 접근법이다. 문제는 데이터의 부족인데 먼저 답뉴럴넷을 이용하여 이미지로부터 스케치를 확보하고 전이학습을 이용하여 이미지를 검색하는 방법이 제안되었다[43]. 새로 생성된 스케치 데이터 집합의 분포는 VAE(variational auto-encoder)를 이용하여 학습되었으며, 학습된 인코더 블록을 전이하여 검색에 활용하였다.

스케치와 이미지에 존재하는 차이를 줄이기 위하여 스케치를 이미지로 전환하려는 접근법이 소개되었다[44]. 먼저 DMGAN(domain migration GAN)을 이용하여 이미지와 스케치를 생성하는데, DMGAN은 적대적 손실(adversarial loss)과 사이클 일관성(cycle consistency) 손실을 사용하여 GAN을 훈련하였다. 그리고 유사성 학습 네트워크가 DMGAN에 추가된다. 유사성 학습 네트워크는 DMGAN이 생성한 이미지와 실제 이미지의 유사

도 정보를 학습하는 것을 목표로 하였고, Siamese 네트워크의 구조에 기반하고 있으며 분류 성능과 유사도 손실을 활용하여 학습하였다.

VAE모형을 이용하여 스케치 기반의 이미지 검색을 시도한 방법으로 SketchGCN도 존재한다[45]. 제안 모델에서는 먼저 인코딩 네트워크를 이용하여 스케치와 이미지를 공통의 시멘틱 스페이스로 임베딩한다. 그리고 스케치와 이미지 사이의 도메인 격차를 줄이기 위하여 그래프 근접행렬 계산 과정에서 매트릭 학습과 그래프 컨볼루션 네트워크를 모두 활용하였다.

스케치에 색상과 같이 새로운 정보가 부가되는 경우 현재까지 축적된 스케치 분석 방법으로 대처하기 어렵다. Sketch-QNet은 색상이 부가된 경우에 대처하기 위하여 쿼드러플렛(quadruplet) 구조를 제안하였다[46].

SBIR 문제는 일반적으로 스케치와 이미지가 공유하는 의미적 콘텐츠가 공유되는 임베딩 공간을 학습하는 방식으로 접근한다[1]. [1]에서는 스케치에 존재하는 스타일 변화에 강인한 검색 모델을 개발하기 위하여 특성 변환 층과 레귤러라이저를 추가하였다.

손실 함수는 최적화 과정에서 중요한 역할을 한다. 스케치와 이미지의 도메인 특성을 반영하기 위하여 [47]에서는 MEMS(multiplicative Euclidean margin softmax) 함수를 제안하였다. 해당 손실함수는 클래스 내부 거리의 최대값이 클래스간 거리의 최소값보다 작아지도록 설계되었다. MEMS함수를 이용한 학습 과정에서는 먼저 클래스별 프로토타입을 획득한 후 손실함수를 계산하게 된다.

트리플렛 손실함수를 이용하여 스케치와 이미지의 도메인 갭을 줄여보려는 기존 접근법이 세밀한 특성을 무시하고 있다는 문제점을 극복하려는 접근법이 제안되었다[48]. 제안 모델의 스케치 브랜치와 이미지 브랜치는 이전 방법과 달리 가중치를 공유하지 않으며, 개별 지역 특성에 독립적인 노달라이제이션 연산을 적용하였다. 이에 더하여 컨볼루션 연산과 유사하게 스케치의 지역 특성을 이미지의 특성 맵에 매칭시킴으로써 오정렬 문제를 해결하였다.

스케치를 구성하는 획 중 이미지 검색에 도움이 되는 획만을 선정하여 이미지 검색 성능을 향상하려는 접근법이 제안되었다[49]. 획 선택 모듈은 2단계 계층 구조로 획 수준의 임베딩 네트워크와 관계 네트워크로 구성되었다. 검색에 도움이 되는 획 레이블이 존재하지 않기 때문에 주어진 획의 레이블 분포를 사전 학습된 SBIR 시스템을 이용하여 추정하였다.

스케치는 직관적인 형태로 정보를 전달할 수 있으나 스케치의 작성 자체가 간단한 작업은 아니다. 이와 같은 단점을 극복하고자 획이 추가됨에 따라 이미지 검색을 진행하는 방법이 제안되었다[50]. 제안 방법에서는 연속된 스케치 사이의 거리를 최소화하면서 동시에 타겟 이미지와의 거리를 줄이는 것을 목표로 학습을 진행하였다. 연속된 스케치 사이의 임베딩 벡터 매핑은 간단한 선형 함수로 표현되었다.

[51]에서 제안한 방법은 딥뉴럴넷의 레벨 정보와 위치 정보를 모두 감안하여 중요성을 부여할 수 있는 트랜스포머 기반 방법이다. 제안 방법은 기본적으로 트리플렛 손실 함수를 사용하지만, 겹치는 영역에 페널티를 부여하는 손실 함수를 추가하였다.

이미지와 스케치외에 이미지의 배경을 제거한 후 객체의 가장자리를 감지하여 보조 모달리티를 생성하는 접근법이 제안되었다[52]. 모달리티에 따라 변하지 않는 특성의 학습을 위해 이미지와 스케치의 특성을 보조 모달리티에 맞춰 정렬하였다. 획득된 특성은 스케치와 이미지 분포 학습에 활용되었다.

스케치 추상화 문제를 다루기 위하여 StyleGAN으로부터 모양 및 구조와 관련된 레이턴트 벡터를 추출한 후 활용하는 방법이 제안되었다[53]. 제안 방법에서는 특성 행렬을 구축하는데 해당 행렬의 행의 변화는 특성의 세밀함/자세함의 변화를 의미하여 거리 계산 과정에서 추상화 정도를 제어할 수 있도록 고안하였다.

2. 데이터 부족의 해결

제로샷 학습 등을 이용하여 SBIR의 데이터 부족 문제를 해결하려는 연구도 활발하게 진행되고 있다.

SAKE(Semantic-Aware Knowledge prEservation)는 사전 훈련된 넷에서 획득할 수 있는 차별화 특성의 보존을 위해 클래스 간 관계를 활용하는 방법이다[54]. SAKE는 오토인코더 구조를 활용하여 이미지 도메인과 스케치 도메인의 특성을 하나의 네트워크로 학습한다. 학습한 네트워크는 두 개의 분류기와 연결되는데, 첫 번째 분류기는 벤치마크 레이블을 예측하며, 두 번째 분류기는 ImageNet의 레이블을 예측하도록 하여 SBIR을 위한 재훈련 과정에서 기억 상실을 예방하고자 설계하였다.

유사한 이미지와 스케치를 임베딩 공간에서 근접시키는 과정에서 의미 유사도 행렬(semantic similarity matrix)을 요구하지 않는 방법이 제안되었다[55]. 제안 모델은 훈련 과정에서 정답(ground-truth)을 요구하지 않는데, 아이템 별로 추가 정보를 제공하는 것이 아니라

클래스에 대한 평균 활성화 분포를 제공함으로써 제로샷 학습을 실현하였다.

BDA-SketRet은 여러 수준에서 모달리티 적응을 진행하여 스케치와 이미지에 존재하는 모달리티 갭을 극복하려고 하였다[56]. 스케치 및 포지티브 이미지에 대한 백본 네트워크의 마지막 텐스 레이어에서 획득한 특성 분포가 공유하는 정보는 최대화하고, 네가티브 이미지와 공유하는 정보는 최소화하도록 손실함수가 설계되었다. 이미지와 스케치가 공유하는 추상적인 특성은 적대적 도메인 적응 프레임워크를 활용하였는데, 결정경계의 설계를 통해 도메인 불변성을 획득하였다.

SBIR에 제로샷 학습을 적용한 대부분의 접근법이 도메인 정렬과 일반화로 문제를 분할하고 있는데, Sketch-an-Anchor 모델에서는 사전 학습된 모델의 이미지 스페이스와 매칭되는 스케치 표현을 학습하고자 하였다[57]. 제로샷 SBIR 모델에서는 Word2Vec 등을 통해 확보한 시멘틱 임베딩을 활용하는데 제안 모델은 워드 임베딩과 훈련 데이터의 임베딩 평균값을 모두 시멘틱 임베딩으로 활용하였다.

[6]에서 소개한 모델은 트랜스포머를 제로샷 SBIR에 적용하려는 모델로 가장 도움이 되는 로컬 영역을 찾는 어텐션 모듈, 이미지와 스케치의 비주얼 토큰의 대응 관계를 계산하는 어텐션 모듈과 전반적인 유사도를 계산하는 관계 네트워크로 구성되어 있다.

스케치 기반 이미지 검색에서도 훈련 데이터를 요구하지 않는 접근법이 소개되고 있다[58]. 스케치의 경우 획득 가능한 데이터의 양이 제한적이라는 것이 현실적으로 문제가 되고 있는데, 다양한 사전 훈련 모델들의 존재가 데이터 부족을 극복할 가능성을 제시해주고 있다. 단, 스케치와 이미지의 모달리티가 상이하다는 문제점을 해결하기 위하여 분류기와 인코더의 표현 공간의 통합, 모달리티에 특화된 정보를 삭제할 수 있는 방법이 제안되었다.

사용할 수 없는 스케치의 부족으로 스케치 기반 이미지 검색의 연구와 실용화가 제약을 받고 있다. [59]에서는 액티브 러닝을 이용하여 스케치의 부족을 해결하려고 하였다. 제안 방법에서는 훈련 과정에서 접하지 못한 이미지와 임베딩 공간에서 다른 스케치 및 이미지와의 거리를 인덱스로 표시한 후, 접하지 못할 가능성이 큰 이미지를 샘플링하여 학습에 활용하였다.

IV. 주요데이터 및 성능

스케치 데이터는 일반적인 이미지 데이터에 비하여 생

성하기도 어려우며, 데이터셋의 규모도 작다. 4장에서는 스케치 기반 이미지 검색에서 주로 사용하는 데이터셋의 특징을 요약하고, 주요 방법의 검색 성능을 제시한다.

1. 주요데이터의 특징

[60]에서는 스케치 기반 이미지 검색의 초기 연구 과정에서 성능 평가를 위해 벤치마크를 설계하였다. 해당 벤치마크는 31개의 벤치마크 리스트로 101,240개의 이미지에 대하여 검색을 수행하는 과정으로 구성되어 있다. TU-Berlin sketch는 250개 카테고리에 대하여 사람들이 그린 20,000개의 스케치로 구성된 데이터셋이다[61]. Tu-Berlin extended 데이터셋은 TU-Berlin sketch에 ImageNet에서 획득한 실제 이미지를 추가한 것으로 191,067개의 실제 이미지를 포함하고 있다[62]. Sketchy는 125개의 카테고리에 속하는 12,500개의 객체에 대하여 그려진 75,471개의 스케치를 모은 데이터셋이다. Sketchy extended(Sketchy ext)은 ImageNet 데이터셋에서 획득한 60,502개의 이미지를 포함하고 있는 데이터셋이다[63]. QuickDraw는 온라인 게임인 Quick, Draw!에서 획득한 스케치의 집합으로, 345개의 카테고리에 대한 스케치 약 5천만개로 구성되어 있다[64]. MECD(museum exhibition collection dataset)는 중국 박물관에 수집된 이미지로부터 생성된 데이터셋이다. 30개의 스케치 질의, 900개의 타겟 이미지, 100,000개의 변형된 이미지로 구성되어 있다[40]. [65]에서는 Flickr30k-Sketchy 인터섹션 데이터셋과 MSCOCO-Sketchy 인터섹션 데이터셋을 소개하였다. Flickr30k-Sketchy 인터섹션 데이터셋은 7,643개의 이미지와 4,127개의 스케치로 구성되어 있으며 인터섹션 카테고리는 총 8개이다. MSCOCO-Sketchy 인터섹션 데이터셋은 56,999개의 이미지와 16,908개의 스케치로 구성되어 있으며 인터섹션 카테고리는 총 32개이다. QMUL 데이터셋도 스케치 관련 이미지 연구에서 많이 사용된다. 해당 셋은 QMUL Shoe, QMUL-Shoe V2, QMUL Chair, QMUL-Chair V2, QMUL Handbag, QMUL FG-SBIR으로 나눌 수 있는데 각각 419개, 2000개, 297개, 400개, 568개, 8721개의 스케치로 구성되어 있다[66]. Quick-DrawCOCO-92c는 QuickDraw를 확장한 데이터셋이다[67]. 총 92개의 카테고리에 대하여 115,800개의 스케치가 존재한다. [68]에서는 이미지 검색 연구를 위해 의류를 대상으로 2000개의 스케치-이미지 쌍의 데이터셋을 소개하였다.

2. 주요 방법별 검색 성능

표 1에서 주요 데이터셋에 대한 최신 SBIR 방법 및 제로샷 접근법의 성능을 정리하였다. 성능 측도로는 mAP@all(mean average precision)과 Prec@200(precision at 200), Prec@100(precision at 100) 등이 이용되었는데, 논문들이 동일한 벤치마크 데이터셋에 대하여 동일한 측도로 성능을 측정하지 않았기 때문에 각 측도별로 높은 성능을 달성한 모델들을 소개하였다. 주로 SBIR 모델의 성능을 조사하였으나, ★로 표시된 모델은 제로샷 학습법에 기반한 SBIR 모델이다. 딥뉴럴 모델에 기반한 이미지 검색 모델들의 검색 성능과 비교하였을 때 SBIR 모델의 검색 성능은 여전히 만족스러운 수준은 아니라는 것을 알 수 있다.

Table 1. Performance comparison.

표 1. 성능 비교

Dataset	Model	Measure	Performance
Sketchy	[58]	mAP@all	0.827
		Prec@200	0.831
Sketch ext	[6]	mAP@all	0.736
		Prec@100	0.808
TU-Berlin	[58]	mAP@all	0.680
		Prec@200	0.693
	[57]★	mAP@all	0.495
		Prec@100	0.608
Tu-Berlin ext	[6]	mAP@all	0.569
		Prec@100	0.657
QuickDraw	[58]	mAP@all	0.400
		Prec@200	0.410

V. 결론

본 리뷰 논문에서는 최근 활발히 연구되고 있는 SBIR 뿐만 아니라 스케치 기반 이미지 생성 및 인식 관련 연구들을 소개하였다. 사전 학습된 딥뉴럴 모델이 많이 공개되면서 대부분의 접근법들이 사전 학습된 모델에서 유사한 이미지와 스케치의 임베딩 벡터를 근접시키려는 목적의 손실 함수를 설계하여 사용하고 있으며 주로 트리플렛 손실 함수가 사용됨을 알 수 있었다. 그러나 벤치마크 데이터셋에 대한 실험 결과에서 알 수 있듯 아직 만족스러운 검색 성능을 달성하지는 못하였다. 지금까지 보고된 만족스럽지 못한 성능은 이미지와 스케치의 모달

리티 격차 해결에 있어 임베딩 벡터의 선택과 Word2Vec 임베딩 벡터의 가이드라인으로써의 적절성에 대하여 의문을 던지게 한다. 이미지 인식 및 검색과 달리 최신 SBIR 방법이 달성한 성능은 스케치라는 추상화된 정보를 처리하기 위하여 더 많은 추가 연구가 필요함을 보여주고 있다.

References

- [1] A. Sain, A. K. Bhunia, Y. Yang, T. Xiang, and Y.-Z. Song, "StyleMeUp: towards style-agnostic sketch-based image retrieval," in *Proc. of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, 2021. DOI: 10.48550/arXiv.2103.15706
- [2] S. Koley, A. K. Bhunia, A. Sain, P. N. Chowdhury, T. Xiang, Yi-Zhe Song, "Picture that sketch: photorealistic image generation from abstract sketches," in *Proc. of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023)*, 2023. DOI: 10.48550/arXiv.2303.11162
- [3] P. Xu, T. M. Hospedales, Q. Yin, Y.-Z. Song, T. Xiang, and L. Wang, "Deep learning for free-hand sketch: a survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol.45, no.1, pp.285-312, 2022. DOI: 10.48550/arXiv.2001.02600
- [4] R. Hu and J. Collomosse, "A performance evaluation of gradient field HOG descriptor for sketch based image retrieval," *CVIU*, vol.117, no.7, pp.790-806, 2013. DOI: 10.1016/j.cviu.2013.02.005
- [5] S. K. Yelamarthi, S. K. Reddy, A. Mishra, and A. Mittal, "A zero-shot framework for sketch based image retrieval," in *Proc. the 15th European Conference on Computer Vision (ECCV2018)*, 2018. DOI: 10.48550/arXiv.1807.11724
- [6] F. Lin, M. Li, D. Li, T. Hospedales, Y.-Z. Song, Y. Qi, "Zero-shot everything sketch-based image retrieval, and in explainable style," in *Proc. of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023)*, 2023. DOI: 10.48550/arXiv.2303.14348

- [7] H. Zhang, S. Liu, C. Zhang, W. Ren, R. Wang, and X. Cao, "SketchNet: sketch classification with web images," in Proc. of *2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, 2016. DOI: 10.1109/CVPR.2016.125
- [8] X. Zhang, Y. Huanga, Q. Zoua, Y. Pei, R. Zhanga, and S. Wang, "A Hybrid convolutional neural network for sketch recognition," *Pattern Recognit. Lett.*, vol.130, pp.73-82, 2020. DOI: 10.1016/j.patrec.2019.01.006
- [9] K. Zhang, W. Luo, L. Ma, and H. Li, "Cousin network guided sketch recognition via latent attribute warehouse," in Proc. of the *AAAI Conference on Artificial Intelligence (AAAI2019)*, pp. 9203-9210, 2019. DOI: <https://doi.org/10.1609/aaai.v33i01.33019203>
- [10] G. Jain, S. Chopra, S. Chopra, and A. S. Parihar, "TransSketchNet: attention-based sketch recognition using transformers," in Proc. of the *24th European Conference on Artificial Intelligence (ECAI 2020)*, 2020.
- [11] L. Li, C. Zou, Y. Zheng, and Q. Su, "Sketch-R2CNN: An Attentive Network for Vector Sketch Recognition," *IEEE Trans. Vis. Comput. Graph.*, vol.27, no.9, pp.3745-3754, 2021.
- [12] A. Tripathi, R. R. Dani, A. Mishra, and A. Chakraborty, "Sketch-Guided Object Localization in Natural Images," in Proc. of the 16th European Conference on Computer Vision (ECCV2020), 2020. DOI: 10.48550/arXiv.2008.06551
- [13] P. N. Chowdhury, A. K. Bhunia, A. Sain, S. Koley, T. Xiang, and Y.-Z. Song, "What can human sketches do for object detection?," in Proc. of *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023)*, 2023. DOI: 10.48550/arXiv.2303.15149
- [14] S. Woo, S.-Y. Jeon, J. Park, M. Son, S. Lee, and C. Kim, "Sketch-based video object localization," in Proc. of *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV2024)*, 2024. DOI: 10.48550/arXiv.2304.00450
- [15] J. Song, K. Pang, Y.-Z. Song, T. X., and T. M. Hospedales, "Learning to Sketch with Shortcut Cycle Consistency," in Proc. of *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018)*, 2018. DOI: 10.48550/arXiv.1805.00247
- [16] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in Proc. of *2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, 2017. DOI: 10.48550/arXiv.1703.10593
- [17] D. Ha and D. Eck, "A neural representation of sketch drawings," in Proc. of the *6th International Conference on Learning Representations (ICLR2018)*, 2018.
- [18] F. Huang and J. F. Canny, "Sketchforme: Composing Sketched Scenes from Text Descriptions for Interactive Applications," in Proc. of the *32nd Annual ACM Symposium on User Interface Software and Technology (UIST19)*, 2019. DOI: 10.48550/arXiv.1904.04399
- [19] X. Xing, C. Wang, H. Zhou, J. Zhang, Q. Yu, and D. Xu, "DiffSketcher: text guided vector sketch synthesis through latent diffusion models," in Proc. of the *28th Annual Conference on Neural Information Processing Systems (NeurIPS 2024)*, 2024. DOI: 10.48550/arXiv.2306.14685
- [20] K. Sasaki and T. Ogata, "Adaptive drawing behavior by visuomotor learning using recurrent neural networks," *IEEE Trans. Cogn. Dev. Syst.*, vol.11, no.1, pp.119-128, 2019. DOI: 10.1109/TCDS.2018.2868160
- [21] S. Ge, V. Goswami, C. L. Zitnick, and D. Parikh, "Creative sketch generation," in Proc. of *International Conference on Learning Representations (ICLR2021)*, 2021. DOI: 10.48550/arXiv.2011.10039
- [22] A. Das, Y. Yang, T. Hospedales, T. Xiang, and Y.-Z. Song, "Cloud2Curve: generation and vectorization of parametric sketches," in Proc. of *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, 2021.
- [23] Q. Wang, H. Deng, Y. Qi, D. Li, and Y.-Z. Song, "SketchKnitter: vectorized sketch generation

- with diffusion models,” in Proc. of *International Conference on Learning Representations (ICLR2023)*, 2023.
- [24] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” in Proc. of *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019)*, 2019.
DOI: 10.48550/arXiv.1812.04948
- [25] K. He, X. Zhang, S. Ren, and Ji. Sun, “Deep residual learning for image recognition,” in Proc. of *2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, 2016.
- [26] L. Gao, S.-Y. Chen, K. Jiang, C.-P. Li, Y.-K. Lai, and H. Fu, “SketchFaceNeRF: sketch-based facial generation and editing in neural radiance fields,” *ACM Trans. Graph.*, vol.42, no.4, pp.1-17, 2023. DOI: 10.1145/3592100
- [27] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, “Analyzing and improving the image quality of stylegan,” in Proc. of *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020)*, 2020.
DOI Bookmark: 10.1109/CVPR42600.2020.00813
- [28] E. R. Chan, C. Z. Lin, M. A. Chan, K. Nagano, B. Pan, S. De Mello, O. Gallo, L. J Guibas, J. Tremblay, S. Khamis, et al., “Efficient geometry-aware 3D generative adversarial networks,” in Proc. of *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, 2022.
DOI: 10.48550/arXiv.2112.07945
- [29] A. Voynov, K. Aberman, and D. Cohen-Or, “Sketch-guided text-to-image diffusion models,” in Proc. of *Special Interest Group on Computer Graphics and Interactive Techniques (SIGGRAPH '23)*, 2023.
- [30] A. Mikaeili, O. Perel, M. Safaee, D. Cohen-Or, and A. Mahdavi-Amiri, “SKED: sketch-guided text-based 3D editing,” in Proc. *2023 IEEE/CVF International Conference on Computer Vision (ICCV 2023)*, 2023. DOI: 10.48550/arXiv.2303.10735
- [31] Y. Li and W. Li, “A survey of sketch-based image retrieval,” *Mach. Vis. Appl.*, vol.29, pp. 1083-1100, 2018.
- [32] F. Yang, N. A. Ismail, Y. Y. Pang, V. R. Kebande, A. AL-Dhaqm, AND T. W. Koh, “A systematic literature review of deep learning approaches for sketch-based image retrieval: datasets, metrics, and future directions,” *IEEE Access*, vol.12, pp.14847-14869, 2024.
DOI: 10.1109/ACCESS.2024.3357939
- [33] A. Creswell and A. A. Bharath, “Adversarial training for sketch retrieval,” in Proc. of *the 14th European Conference on Computer Vision (ECCV 2016)*, 2016.
DOI: 10.1007/978-3-319-46604-0_55
- [34] J. Song, Q. Yu, Y.-Z. Song, T. Xiang, and T. M. Hospedales, “Deep spatial-semantic attention for fine-grained sketch-based image retrieval,” in Proc. of *2017 IEEE International Conference on Computer Vision (ICCV2017)*, 2017.
DOI: 10.1109/ICCV.2017.592
- [35] L. Liu, F. Shen, Y. Shen, X. Liu, and L. Shao, “Deep Sketch Hashing: Fast Free-Hand Sketch-Based Image Retrieval,” in Proc. of *2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, 2017.
DOI: 10.48550/arXiv.1703.05605
- [36] L. Guo, J. Liu, Y. Wang, Z. Luo, W. Wen, and H. Lu, “Sketch-based image retrieval using generative adversarial networks,” in Proc. of *the 25th ACM International Conference on Multimedia (MM17)*, 2017. DOI: 10.1145/3123266.3127939
- [37] T. Portenier, Q. Hu, P. Favaro, and M. Zwicker, “SmartSketcher: sketch-based image retrieval with dynamic semantic re-ranking,” in Proc. of *the Symposium on Sketch-Based Interfaces and Modeling (SBIM'17)*, 2017.
DOI: 10.1145/3092907.3092910
- [38] D. Yu, Y. Liu, Y. Pang, Z. Li, and H. Li, “A multi-layer deep fusion convolutional neural network for sketch based image retrieval,” *Neuro-computing*, vol.296, pp.23-32, 2018.
DOI: 10.1016/j.neucom.2018.03.031
- [39] T. Bui, L. Ribeiro, M. Ponti, and J. Collomosse, “Sketching out the details: Sketch-based image

retrieval using convolutional neural networks with multi-stage regression,” *Comput. Graph.*, vol.71, pp.77-87, 2018.

DOI: 10.1016/j.cag.2017.12.006

[40] H. Fei, J. Cheng, Z. Yuejie, W. Kangnian, Z. Tao, and F. Weiguo, “Sketch-based image retrieval with deep visual semantic descriptor,” *Pattern Recognit.*, vol.76, pp.537-548, 2018.

DOI: 10.1016/j.patcog.2017.11.032

[41] H. Zhang, P. She, Y. Liu, J. Gan, X. Cao, and H. Foroosh, “Learning structural representations via dynamic object landmarks discovery for sketch recognition and retrieval,” *IEEE Trans. Image Process.* vol.28, no.9, pp.4486-4499, 2019.

DOI: 10.1109/TIP.2019.2910398

[42] F. Huang, J. F. Canny, and J. Nichols, “Swire: sketch-based user interface retrieval,” in *Proc. of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19)*, 2019.

DOI: 10.1145/3290605.3300334

[43] M. Bhattarai, D. Oyen, J. Castorena, L. Yang, and B. Wohlberg, “Diagram image retrieval using sketch-based deep learning and transfer learning,” in *Proc. of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020. DOI: 10.1109/CVPRW50498.2020.00095

[44] C. Bai, J. Chen, Q. Ma, P. Hao, and S. Chen, “Cross-domain representation learning by domain-migration generative adversarial network for sketch based image retrieval,” *J. Vis. Commun. Image R.* vol.71, 102835, 2020.

DOI: 10.1016/j.jvcir.2020.102835

[45] Z. Zhang, Y. Zhang, R. Feng, T. Zhang, and W. Fan, “Zero-shot sketch-based image retrieval via graph convolution network,” in *Proc. of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*, 2020.

DOI: 10.1609/aaai.v34i07.6993

[46] A. Fuentes and J. M. Saavedra, “Sketch-QNet: a quadruplet convnet for color sketch-based image retrieval,” in *Proc. of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2021.

DOI: 10.48550/arXiv.2104.11130

[47] P. Lu, G. Huang, H. Lin, W. Yang, G. Guo, and Y. Fu, “Domain-aware SE network for sketch-based image retrieval with multiplicative euclidean margin softmax,” in *Proc. of the 29th ACM International Conference on Multimedia (MM '21)*, 2021. DOI: 10.48550/arXiv.1812.04275

[48] J. Xu, H. Sun, Q. Qi, J. Wang, C. Ge, L. Zhang, and J. Liao, “DLA-Net for FG-SBIR: dynamic local aligned network for fine-grained sketch-based image retrieval,” in *Proc. of the 29th ACM International Conference on Multimedia (MM '21)*, 2021. DOI: 10.1145/3474085.3475705

[49] A. K. Bhunia, S. Koley, A. F. Khilji, A. Sain, P. N. Chowdhury, T. Xiang, and Y.-Z. Song, “Sketching without worrying: noise-tolerant sketch-based image retrieval,” in *Proc. of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, 2022.

[50] Dai, X. Tang, Y. Liu, S. Xia, and G. Wang, “Multi-granularity association learning for on-the-fly fine-grained sketch-based image retrieval,” *Knowl.-Based Syst.*, vol.253, no.109447, 2022.

DOI: 10.1016/j.knosys.2022.109447

[51] Z. Ling, Z. Xing, J. Li, and L. Niu, “Multi-level region matching for fine-grained sketch-based image retrieval,” in *Proc. of the 30th ACM International Conference on Multimedia (MM '22)*, 2022. DOI:10.1145/3503161.3548147

[52] X. Liu, X. Cheng, H. Chen, H. Yu, and G. Zhao, “Differentiable auxiliary learning for sketch re-identification,” in *Proc. of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)*, 2024. DOI:10.1609/aaai.v38i4.28165

[53] S. Koley, A. K. Bhunia, A. Sain, P. N. Chowdhury, T. Xiang, and Y.-Z. Song, “How to handle sketch-abstraction in sketch-based image retrieval?,” accepted in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2024)*, 2024.

DOI: 10.48550/arXiv.2403.07203

[54] Q. Liu, L. Xie, H. Wang, and A. L. Yuille, “Semantic-aware knowledge preservation for zero-

- shot sketch-based image retrieval,” in Proc. of *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019)*, 2019. DOI: 10.48550/arXiv.1904.03208
- [55] O. Tursun, S. Denman, S. Sridharan, E. Goan, and C. Fookes, “An efficient framework for zero-shot sketch-based image retrieval,” *Pattern Recognit.*, vol.126, 108528, 2022. DOI: 10.1016/j.patcog.2022.108528
- [56] U. Chaudhuri, R. Chavan, B. Banerjee, A. Dutta, and Z. Akata, “BDA-SketRet: Bi-level domain adaptation for zero-shot SBIR,” *Neurocomputing*, vol.514, pp.245-255, 2022. DOI: 10.1016/j.neucom.2022.09.104
- [57] L. S. F. Ribeiro and M. A. Ponti, “Sketch-anchor: sub-epoch fast model adaptation for zero-shot sketch-based image retrieval,” arXiv: 2303.16769, 2023.
- [58] A. Chaudhuri, A. K. Bhunia, Y.-Z. Song, and A. Dutta, “Data-free sketch-based image retrieval,” in Proc. of *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023)*, 2023. DOI: 10.48550/arXiv.2303.07775
- [59] H. Thakur and S. Chattopadhyay, “Active learning for fine-grained sketch-based image retrieval,” in Proc. of the 34th British Machine Vision Conference (BMVC2023), 2023. DOI: 10.48550/arXiv.2309.08743
- [60] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, “Sketch-based image retrieval: benchmark and bag-of-features descriptors,” *IEEE Trans. Vis. Comput. Graphics.*, vol.17, no.11, pp.1624-1636, 2011. DOI: 10.1109/TVCG.2010.266
- [61] M. Eitz, J. Hays, and M. Alexa, “How do humans sketch object?,” *ACM Trans. Graph.*, vol. 31, no.4, pp.1-10, 2012.
- [62] H. Zhang, S. Liu, C. Zhang, W. Ren, R. Wang, and X. Cao, “SketchNet: sketch classification with web images,” in Proc. of *2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, 2016. DOI: 10.1109/CVPR.2016.125
- [63] P. Sangkloy, N. Burnell, C. Ham, and J. Hays, “The sketchy database: learning to retrieve badly drawn bunnies,” *ACM Trans. Graph.*, vol.35, no.4, pp.1-12, 2016. DOI: 10.1145/2897824.29259
- [64] D. Ha and E. Eck, “A neural representation of sketch drawings,” in Proc. of *2018 International Conference on Learning Representations (ICLR 2018)*, 2018.
- [65] Y. Wang, F. Huang, Y. Zhang, R. Feng, T. Zhang, and W. Fan, “Deep cascaded cross-modal correlation learning for fine-grained sketch-based image retrieval,” *Pattern Recognit.*, vol.100, no. 107148, 2020. DOI: 10.1016/j.patcog.2019.107148
- [66] Q. Yu, J. Song, Y.-Z. Song, T. Xiang, and T. H. Hospedales, “Fine-grained instance-level sketch-based image retrieval,” *Int. J. Comput. Vis.*, vol. 129, no.2, pp.484-500, 2021. DOI: 10.1007/s11263-020-01382-3
- [67] L. S. Ferraz Ribeiro, T. Bui, J. Collomosse, and M. Ponti, “Scene designer: a unified model for scene search and synthesis from sketch,” in Proc. of *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021.
- [68] X. Zhang, M. Shen, X. Li, and F. Feng, “A deformable CNN-based triplet model for fine-grained sketch-based image retrieval,” *Pattern Recognit.* vol.125, 108508, 2022. DOI: 10.1016/j.patcog.2021.108508

BIOGRAPHY

Sehong Oh (Member)



2015 : BS degree in Weapon System Engineering, Korea Military Academy, Seoul, South Korea.

2023 : MS degree in Electrical and Computer Engineering, University of Michigan - Ann Arbor, U.S.

2023~present: He is lecturing on Introduction to AI and Reinforcement learning and working for AI R&D center of Korea Military Academy.

Ho-Sik Seok (Member)

1999 : BS degree in Computer Engineering, Seoul National University.

2001 : MS degree in Electrical Engineering and Computer Science, Seoul National University.

2012 : PhD degree in Electrical Engineering and Computer Science, Seoul National University.

2016~2020.2 : Assistant professor, Dept. of Computer and Communications Engineering, Kangwon National University.

2020.3~2022.1 : Assistant professor, Dept. of Computer Science and Engineering, Kangwon National University.

2022.2~present : Assistant professor, Dept. of Artificial Intelligence and Data Science, Korea Military Academy.