일반논문 (Regular Paper)
방송공학회논문지 제27권 제6호, 2022년 11월 (JBE Vol.27, No.6, November 2022)
https://doi.org/10.5909/JBE.2022.27.6.897
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

대조적 학습을 활용한 주요 프레임 검출 방법

박 경 태[a], 김 원 준[a]‡, 이 용[b], 장 래 영[b], 최 명 석[b]

Key Frame Detection Using Contrastive Learning

Kyoungtae Park[a], Wonjun Kim[a]‡, Ryong Lee[b], Rae-young Lee[b], and Myung-Seok Choi[b]

요 약

비디오 영상 내 주요 프레임(Key Frame) 검출은 컴퓨터 비전 분야에서 꾸준히 연구되고 있는 분야 중 하나이다. 최근 심층학습(Deep Learning) 기술의 발전으로 비디오 영상에서의 주요 프레임 검출 성능이 향상 되었으나, 다양한 종류의 영상 콘텐츠 및 복잡한 배경으로 인해 여전히 효과적인 학습이 어려운 문제점이 있다. 본 논문에서는 대조적 학습(Contrastive Learning)과 메모리 뱅크(Memory Bank)를 통해 영상의 주요 프레임을 검출하는 새로운 방법을 제안한다. 제안하는 방법은 입력 프레임과 같은 영상 내 이웃하는 프레임 간 차이와 다른 영상 내 프레임과의 차이를 기반으로 특징 추출 신경망을 학습한다. 이와 같은 대조적 학습을 통해 메모리 뱅크에 주요 프레임을 저장 및 갱신하여 영상의 중복성을 효과적으로 제거한다. 비디오 영상 데이터셋에서의 실험 결과를 통해 제안하는 방법의 성능을 검증하였다.

Abstract

Research for video key frame detection has been actively conducted in the fields of computer vision. Recently with the advances on deep learning techniques, performance of key frame detection has been improved, but the various type of video content and complicated background are still a problem for efficient learning. In this paper, we propose a novel method for key frame detection, witch utilizes contrastive learning and memory bank module. The proposed method trains the feature extracting network based on the difference between neighboring frames and frames from separate videos. Founded on the contrastive learning, the method saves and updates key frames in the memory bank, witch efficiently reduce redundancy from the video. Experimental results on video dataset show the effectiveness of the proposed method for key frame detection.

Keyword : contrastive learning, self-supervised learning, key frame detection

a) 건국대학교 전기전자공학부(Department of Electrical and Electronics Engineering, Konkuk University)
b) 한국과학기술정보연구원(Korea Institute of Science and Technology Information)
‡ Corresponding Author : 김원준(Wonjun Kim)
E-mail: wonjkim@konkuk.ac.kr
Tel: +82-2-450-3396
ORCID: https://orcid.org/0000-0001-5121-5931
※ 본 연구는 한국과학기술정보연구원(KISTI) ‘Data/AI 기반 문제해결 체계 구축(K-22-L04-C05-S01)’사업 지원으로 수행되었습니다.
※ This work was supported by a Research and Development project, Building a Data/AI-based Problem-solving System of Korea Institute of Science and Technology Information (KISTI), South Korea, under Grant K-22-L04-C05-S01
· Manuscript August 22 2022; Revised October 24, 2022; Accepted October 24, 2022.

Ⅰ. 서 론

최근 비디오 영상을 이용한 다양한 솔루션 개발이 활발히 진행되고 있으며, 이에 따라 비디오 영상 내 주요 프레임을 검출하는 방법에 대한 수요 또한 증가하고 있다. 주요 프레임 검출 방법은 주어진 비디오 영상 요약을 통해 행동 인식이나 동시적 위치추정 및 지도 작성(Simultaneous Localization and Mapping, SLAM)과 같은 응용 솔루션 개발 시 메모리 사용량과 알고리즘의 수행 속도를 효과적으로 개선할 수 있다. 그러나 프레임의 중요도를 판단하는 기준은 사용자마다 상이하고, 영상 요약에 많은 시간이 요구된다. 따라서, 각 프레임의 특징을 추출하고 이를 활용해 비디오 영상의 주요 프레임을 검출하는 영상처리 기술이 꾸준히 연구되어왔다.

초기 연구는 비디오 영상을 분할하기 위해 군집(Clustering) 기법을 적용하였다. Ngo[1] 등은 움직임 주의(Motion Attention) 모델링 및 장면 군집화를 통해 그래프 모델을 생성하고 그래프 컷 알고리즘을 기반으로 주요 프레임을 검출하였다. Mundur[2] 등과 Kuanar[3] 등은 들로네 삼각분할 알고리즘을 활용해 프레임 군집화를 수행하여 주요 프레임 검출을 하였다. Furini[4] 등은 HSV 히스토그램을 군집 분석에 이용하여 실시간 주요 프레임 검출을 제안하였다. 이후 프레임의 특징을 활용하여 주요 프레임을 검출하는 알고리즘의 연구가 진행되었다. Almeida[5] 등은 원본 이미지를 64배 압축한 DC image[6]와 Zero-mean Normalized Correlation 알고리즘을 통해 아핀 광량 변환(Affine Photometric Transform)에 효과적인 알고리즘을 선보였다. Guan[7] 등은 SIFT[8]를 기반으로 주요 포인트(Key Point)를 검출하고 주변 영역의 지역적 정보를 이용하여 주요 프레임을 검출하는 방법을 제안하였다.

최근 심층학습 기술의 발전으로 심층 신경망을 활용하여 주요 프레임을 검출하는 방법이 활발하게 연구되고 있다. Zhang[9] 등은 장단기 메모리(Long Short-Term Memory, LSTM)를 기반으로 복수의 프레임 간 시간 관계를 적용하여 주요 프레임을 검출하는 방법을 제안하였다. Mahasseni[10] 등은 예측한 주요 프레임을 기반으로 비디오 영상을 복원하고 적대적 생성신경망(Generative Adversarial Network, GAN) 학습을 통해 원본과의 차이를 최소화하여 주요 프레임 검출 정확도를 향상시켰다. Zhao[11] 등은 장면의 경계를 예측하는 장단기 메모리와 각 장면의 중요도를 판단하는 장단기 메모리를 분리하여 비디오 영상을 계층적으로 해석하고 이를 바탕으로 주요 프레임을 검출하는 방법을 제안하였다. 그러나 많은 수의 프레임을 학습에 이용하기 때문에 메모리 운용에 어려움이 있으며, 콘텐츠 종류가 다양해짐에 따라 검출 정확도가 저하되는 문제점이 있다.

본 논문에서는 대조적 학습(Contrastive Learning)과 메모리 뱅크를 이용한 심층 신경망 기반 주요 프레임 검출 방법을 제안한다. 입력 프레임을 특징 벡터로 압축하여 군집화를 수행하였으며, 메모리를 효과적으로 사용하였다. 또한 제안하는 방법은 메모리 뱅크 구조를 채용하여 비디오 영상의 이전 프레임과의 비교를 진행하였다. 프레임을 K-평균 군집화 알고리즘을 활용해 군집 분석을 수행하여 목표하는 프레임 수에 맞게 메모리 뱅크 내 프레임 수를 조정한다. 주요 프레임 검출에 자주 사용되는 VSUMM 데이터셋[12]를 기반으로 성능평가를 진행하여 기존 알고리즘 대비 더 좋은 성능을 가짐을 확인하였다. 본 논문의 구성은 다음과 같다. 2장에서는 제안하는 심층 신경망 구조에 대해 자세히 설명하며, 3장에서는 다양한 실험을 통해 제안하는 방법이 기존 주요 프레임 검출 방법보다 성능이 뛰어남을 검증한다. 마지막으로 4장에서는 본 논문의 결론을 서술한다.

Ⅱ. 제안하는 방법

제안하는 방법은 대조적 학습을 통해 학습한 신경망을 기반으로 프레임의 특징을 압축하고 이를 메모리 뱅크에 저장한다. 또한, 제안하는 방법은 메모리 뱅크에 저장한 각 프레임 간의 유사도를 추론하여 프레임의 군집화와 주요 프레임 검출을 수행한다. 본 장에서는 먼저 대조적 학습 방법과 제안하는 주요 프레임 검출 방법에 대해 자세히 설명한 후 대조적 학습에 사용된 손실 함수에 관해 설명한다.

1. 대조적 학습 및 주요 프레임 검출 방법

제안하는 방법은 크게 두 모듈로 구성되어있다. 첫 번째 모듈은 유사도 검출 신경망이다. 유사도 검출 신경망은 두 개의 프레임 쌍에 대하여 유사한 정도를 0과 1사이의 값으

로 정량화하는 신경망으로 대조적 학습을 이용해 학습하였다. 제안하는 방법은 유사도 검출 신경망을 통해 입력 프레임과 이전 주요 프레임과 유사한 정도를 판단한다. 두 번째 모듈은 메모리 뱅크이다. 메모리 뱅크는 주요 프레임의 저장 및 갱신에 이용되며 K-평균 군집화 알고리즘을 통해 메모리 뱅크 안의 저장된 특징 벡터의 수를 조절한다. 비디오 영상의 모든 프레임을 입력한 후 메모리 뱅크 내 저장된 특징 벡터의 프레임을 주요 프레임으로 판단한다.

본 논문에서는 비디오 영상의 연속적인 프레임의 유사성을 활용하여 대조적 학습을 수행한다. 제안하는 방법은 현재 프레임과 동일한 영상의 근접한 프레임을 양의 쌍(Positive Pair)으로 정의하고, 현재 프레임과 다른 영상의 프레임을 음의 쌍(Negative Pair)으로 정의한다. 신경망은 현재 프레임과 비교 프레임의 유사도를 수치화하고, 해당 신경망은 비슷한 형태를 가진 양의 쌍의 유사도가 음의 쌍의 유사도에 대비하여 높게 추론하도록 학습된다. 구체적으로, 해당 방법은 그림 1과 같이 ImageNet[13] 데이터셋을 기반으로 사전 훈련된 신경망 ResNet-50[14] 구조를 압축기로 사용하여 프레임의 특징을 128차원의 벡터로 압축한다. 압축된 특징 벡터는 다른 프레임의 특징 벡터와 256차원의 벡터로 결합 되어 3개의 완전 연결 층(Fully Connected Layer)으로 이루어진 판별기(Discriminator)의 입력값으로 사용된다. 판별기는 입력된 두 프레임 쌍의 유사도를 0과 1사이의 값으로 추론하기 위해 시그모이드 함수를 마지막에 사용한다. 신경망은 양의 쌍의 특징 벡터 결합과 음의 쌍의 특징 벡터 결합의 유사도를 각각 1과 0으로 추론하도록 학습된다.

메모리 뱅크의 크기는 검출하고자 하는 주요 프레임 수의 2배로 설정하며, 대조적 학습에 사용된 뼈대 신경망을 이용하여 첫 프레임에서 추출한 특징 벡터를 먼저 메모리 뱅크에 저장한다. 이후, 입력 영상의 현재 프레임에서 추출한 특징 벡터와 메모리 뱅크 내 가장 최근에 저장된 특징 벡터 간 유사도를 계산하여 정해진 임계값보다 작은 경우, 메모리 뱅크에 해당 특징 벡터를 저장한다. 임계값은 두 프레임을 유사한 프레임으로 판단하는 기준을 나타내는 매개변수이며 같은 비디오 영상에 일정한 값을 적용한다. 임계값은 0과 1사이의 값으로 설정하며 높은 값으로 설정할수록 더 많은 프레임이 메모리 뱅크에 저장된다. 실험은 임계값을 0.9로 설정하여 주요 프레임 검출을 수행한다. 낮은 임계값을 적용하면 적은 주요 프레임 검사를 통해 빠른 수행속도를 얻을 수 있다. 그러나 장면마다 적은 프레임만 선별하여 각 장면 내 충분한 프레임을 얻지 못하는 어려움이 있

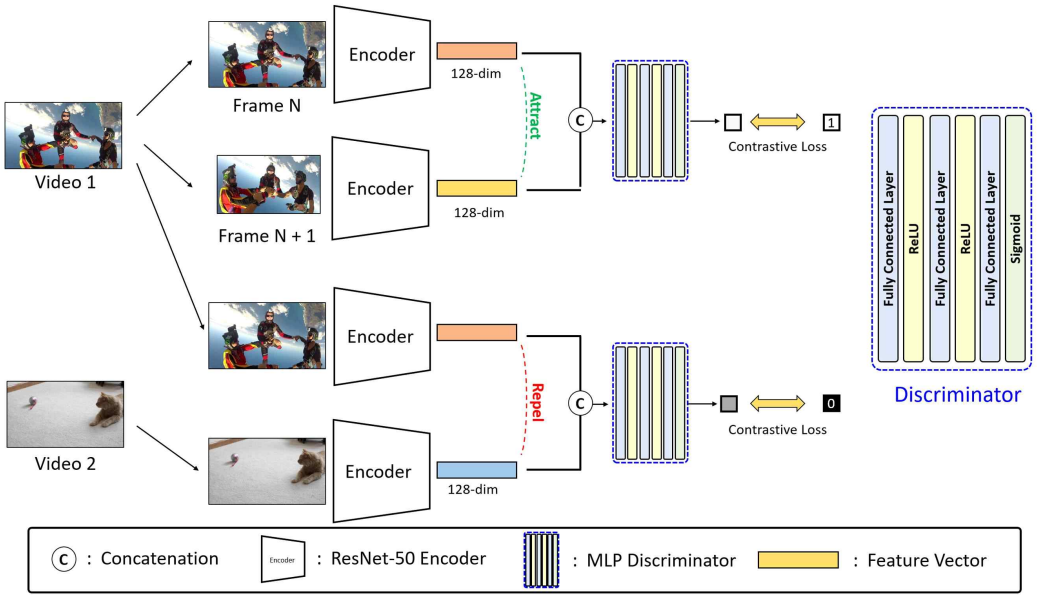

그림 1. 제안하는 대조적 학습 모듈(왼쪽)과 판별기 상세 구조(오른쪽)
Fig. 1. Entire structure of the proposed contrastive learning module(left) and details of the discriminator(right)

다. 제안하는 방법은 비디오 영상을 입력으로 받으며 주요 프레임 후보의 특징 벡터를 메모리 뱅크에 저장한다. 메모리 뱅크 내 특징 벡터가 일정량 이상 차면 K-평균 군집화 알고리즘을 통해 원하는 프레임의 수만큼 조절한다. 특징 벡터 간 거리 계산은 대조적 학습과 동일한 방식을 이용하여 계산하였으며, 기존 K-평균 군집화 알고리즘과 달리 빠른 수행 속도를 위해 군집 중심을 반복적 갱신하는 동작은 적용하지 않았다. 입력 영상의 모든 프레임에 대한 검사가 완료되면, 이때 생성된 각 군집의 중심 벡터와 가장 유사한 특징 벡터를 가진 프레임을 주요 프레임으로 검출한다. 수행 속도 향상을 위해 각 군집 내 첫 번째 프레임을 주요 프레임으로 검출하는 방법도 큰 성능 저하 없이 적용 가능하다.

2. 대조적 학습을 위한 손실함수 설계

본 논문에서는 대조적 학습에 널리 사용되는 손실 함수를 변형하여 사용한다. 제안하는 방법은 메모리 뱅크에 저장된 다수의 특징 벡터를 활용해 음의 쌍을 생성한다. 다량의 음의 쌍을 활용하기 위해 대조적 학습에 사용되는 Triplet loss[15] 손실 함수 대신 이진 크로스 엔트로피 손실 함수를 변형하여 계산한다. 간단히 살펴보면, 신경망의 최종 예측 값에 적용되는 손실함수 L은 양의 쌍의 유사도를 계산할 때 사용하는 손실함수 L_{pos}와 음의 쌍의 유사도를 계산할 때 사용하는 손실함수 L_{neg}의 합으로 정의되며 다음과 같이 표현할 수 있다.

$$L = L_{pos} + \lambda L_{neg}. \tag{1}$$

L_{pos}와 L_{neg}를 계산하기 위해 이진 크로스 엔트로피(Binary Cross Entropy) 손실 함수를 사용하고 있으며 다음과 같이 계산한다.

$$L_{pos} = -\log(D(E(x), E(y))), \tag{2}$$

$$L_{neg} = -\log(1 - D(E(x), E(z))). \tag{3}$$

여기서 x는 현재 프레임, y는 현재 프레임과 이웃하는 프레임, z는 다른 영상의 프레임을 의미한다. $E(\cdot)$는 뼈대 신경망을 통해 압축된 특징 벡터이며, $D(\cdot)$는 판별기를 통해 계산된 유사도이다.

Ⅲ. 실험 결과 및 분석

본 논문에서는 제안하는 방법의 학습과 성능평가를 위

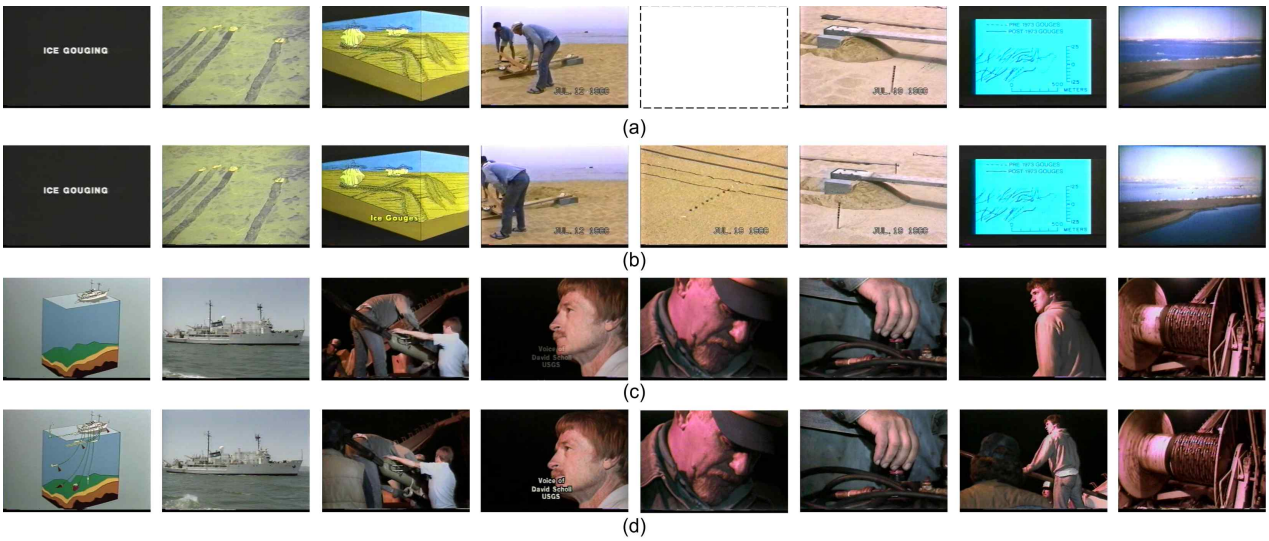

그림 2. VSUMM 데이터셋[12]에 대한 비디오 영상 주요 프레임 검출 결과. (a), (c): 제안하는 방법의 주요 프레임 검출 결과. (b), (d): 정답 주요 프레임
Fig. 2. Examples of video key frame detection on VSUMM dataset[12] (a), (c): Results of the proposed key frame detection method (b), (d): ground truth

해 두 개의 비디오 영상 데이터셋을 사용하였다. 먼저, 대조적 학습을 수행하기 위하여 비디오 영상 데이터셋으로 자주 활용되는 YouTube-VOS 데이터셋[16]을 사용하였다. YouTube-VOS 데이터셋[16]은 다양한 콘텐츠의 영상으로 구성되어있는 대규모 비디오 영상 데이터셋이며, 학습을 위해 3471개의 영상을 제공한다. 성능평가를 위해 주요 프레임 검출에 널리 사용되는 VSUMM 데이터셋[12]와 Summe 데이터셋[17]을 사용하였다. 제안하는 방법은 PyTorch[18] 프레임워크에 기반하여 구현되었다. 본 논문은 신경망 가중치를 최적화하기 위한 알고리즘으로 확률적 경사 하강법(Stochastic Gradient Descent, SGD)을 사용하였고, 가속도(Momentum) 값은 9×10^{-3}을 사용하였다. 학습 속도(Learning Rate)는 1×10^{-3}부터 시작하여 학습의 진행도의 50%와 80%에서 각각 1×10^{-4}와 1×10^{-5}으로 감소하며, 전체 에포크(Epoch)는 1000으로 설정하였다. 제안하는 심층 신경망에 대한 학습 소요 시간은 7일이며 학습과 성능평가에는 Intel i9-10980XE@3.00GHz CPU와 NVIDIA RTX 3080TI GPU 2대가 사용되었다.

제안하는 방법의 효율성을 검증하기 위해 본 논문에서는 VSUMM 데이터셋[12]를 기반으로 해당 논문의 방법을 이용해 다른 주요 프레임 검출 방법과 성능을 비교한다. 다양한 비디오 영상에 대한 주요 프레임 검출 결과를 그림 2에 나타내었다. 그림 2의 결과를 통해 제안하는 방법이 다양한 영상 콘텐츠에 관계없이 성공적으로 주요 프레임을 검출할 수 있음을 확인할 수 있다. 그러나, 제안하는 방법은 그림 3과 같이 장면 전환이 발생하는 경우 두 장면의 속성

그림 3. 제안하는 주요 프레임 검출 실패 예시. (a)-(b): 정답 주요 프레임. (c): 제안하는 방법의 주요 프레임 검출 결과

Fig. 3. Examples of fail key frame detection of the proposed method. (a)-(b): ground truth. (c): detected key frames of the proposed method

을 모두 가지고 있는 프레임을 주요 프레임으로 검출하는데 종종 어려움이 있다. 다음으로는 주요 프레임 검출 성능의 정량적 평가를 위해 널리 사용되는 F-점수(F-measure)를 기반으로 기존 방법과 성능 비교를 수행하였으며 그 결과를 표 1에 나타내었다. 표 1에서 알 수 있듯이 제안하는 방법은 VSUMM 데이터셋[12]에서 비교 방법 중 가장 높은 성능을 달성하였다. 비교 대상의 방법은 HSV 히스토그램 혹은 ZNCC 등 하나의 특징을 통해 주요 프레임 검출을 수행하는 반면 제안하는 방법은 전역적인 특징을 압축하여 프레임을 선별하기 때문에 높은 Recall과 Precision을 기록한다.

표 1. VSUMM 데이터셋[12]에서의 정량적 평가 비교
Table 1. Quantitative evaluations on the VSUMM dataset[12]

	VSUMM dataset		
Method	Recall	Precision	F-measure
Mundur et al. [2]	0.53	0.64	0.57
Furini et al. [4]	0.72	0.55	0.62
Almeida et al. [5]	0.77	0.56	0.65
Kuanar et al. [3]	0.57	0.60	0.43
Proposed Method	0.61	0.75	0.67

추가로 딥러닝을 사용한 주요 프레임 검출 방법과의 비교를 위해 SUMME 데이터셋[17]을 사용하여 표 2에 나타냈다. 해당 데이터셋은 비디오 영상을 장면 단위로 요약하여 다양한 비디오 요약 성능평가에 사용된다. Zhang[9] 등이 소개한 변환 방법을 통해 프레임 단위 주요 프레임 검출을 주요 장면 단위 검출로 변환하였다. 그러나, 제안하는 방법은 타 딥러닝 방법에 비해 낮은 성능을 보여준다. 이는 시간적 정보를 전혀 반영하지 않아 동일한 내용을 가진 장면 내 프레임을 구별에 어려움이 있다. (예를 들어, 그림 4의 예시)

표 2. SUMME 데이터셋[17]에서의 정량적 평가 비교
Table 2. Quantitative evaluations on the SUMME dataset[17]

SUMME dataset	
Method	F-measure
Zhang et al.[9]	0.38
Mahasseni et al.[10]	0.39
Zhao et al.[11]	0.43
Proposed Method	0.33

판별기를 다른 방법과 비교하여 진행한 실험 결과를 표 3에 나타냈다. 코사인 유사도를 이용하여 유사도를 계산한 방법은 신경망을 이용한 제안하는 방법보다 낮은 성능을 보인다. 다른 비교 방법 대비 유사도 값의 일관성이 낮아 정확한 주요 프레임을 검출에 어려움이 있다. 이진 크로스 엔트로피 손실 함수 대신 Triplet Loss[15] 손실함수를 사용한 신경망은 제안하는 신경망보다 좋은 결과를 얻을 수 있으나 학습에 요구되는 시간이 더 길어 많은 양의 데이터를 학습에 어려움이 있다.

표 3. 판별 방법에 따른 VSUMM 데이터셋[12]에서의 정량적 평가 비교
Table 3. Quantitative evaluations of different redundant evaluating methods on the VSUMM dataset[12]

VSUMM dataset	
Method	F-measure
Cosine Distance	0.53
Triplet Loss	0.68
Proposed method	0.67

표 4. 메모리 뱅크의 크기에 따른 VSUMM 데이터셋[12]에서의 정량적 평가 비교
Table 4. Quantitative evaluations of Different Memory Bank Size on the VSUMM dataset[12]

VSUMM dataset	
Method	F-measure
Proposed Method with Memory Size of 2*K	0.67
Proposed Method with Memory Size of 3*K	0.69
Proposed Method with Memory Size of 4*K	0.69
Proposed Method with Memory Size of 5*K	0.69

마지막으로 메모리 뱅크의 크기 변화에 따른 실험의 결과를 표 4에 나타냈다. 본 알고리즘은 메모리 뱅크의 크기에 따라 K-평균 군집화 알고리즘에 사용한 샘플 수가 변하고 이에 따라 주요 프레임 선별 성능이 변한다. 표 4를 통해 메모리 뱅크의 크기가 커지면 성능이 소폭 증가함을 확인할 수 있다.

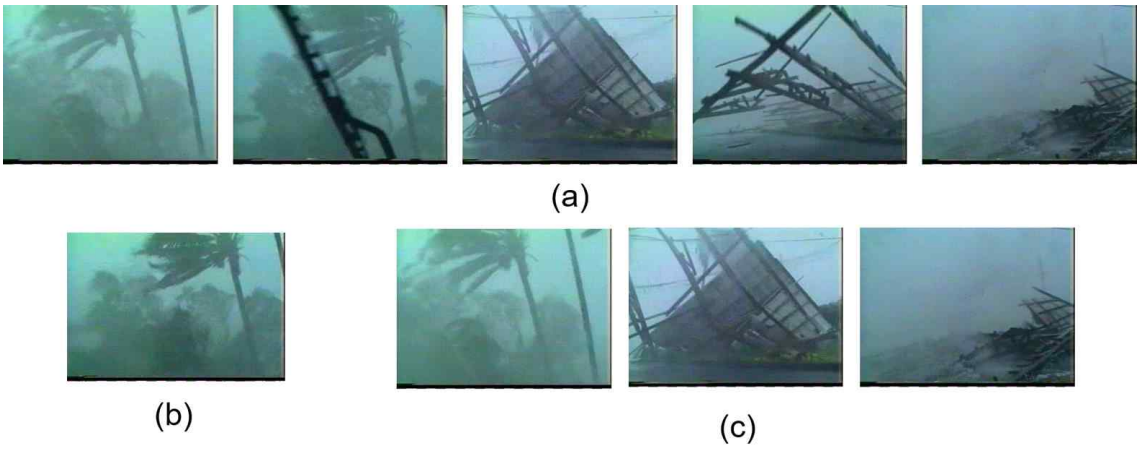

그림 4. 제안하는 주요 프레임 검출 실패 예시. (a): 동일한 내용의 장면 속 프레임 (b): 정답 주요 프레임. (c): 제안하는 방법의 주요 프레임 검출 결과
Fig. 4. Examples of fail key frame detection of the proposed method. (a): Frames with identical context (b): ground truth. (c): detected key frames of the proposed method

Ⅳ. 결 론

본 논문에서는 대조적 학습을 이용한 심층 신경망 기반 주요 프레임 검출 방법을 제안하였다. 제안하는 방법은 대조적 학습을 통해 이웃하는 프레임간의 유사한 특징을 학습하여 두 이미지 쌍의 유사도를 추론하는 신경망을 학습하고, 이를 기반으로 메모리 뱅크에 저장된 특징 벡터의 군집화를 통해 주요 프레임을 검출하였다. 다양한 실험 결과를 통해 제안하는 방법이 주요 프레임 검출에 효과적임을 확인하였다.

참 고 문 헌 (References)

[1] C.-W. Ngo, Y.-F. Ma, and H.-J. Zhang, “Video summarization and scene detection by graph modeling,” IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 2, pp. 296 – 305, Feb. 2005.
doi: https://doi.org/10.1109/TCSVT.2004.841694

[2] P. Mundur, Y. Rao, and Y. Yesha, “Keyframe-based video summarization using delaunay clustering,” International Journal on Digital Libraries, vol. 6, no. 2, pp. 219-232. Apr. 2006.
doi: https://doi.org/10.1007/s00799-005-0129-9

[3] S. K. Kuanar, R. Panda, and A. S. Chowdhury, “Video key frame extraction through dynamic delaunay clustering with a structural constraint,” J. Vis. Commun. Image Represent., vol. 24, no. 7, pp. 1212 – 1227, Apr. 2013.
doi: https://doi.org/10.1016/j.jvcir.2013.08.003

[4] M. Furini, F. Geraci, M. Montangero, and M. Pellegrini, “STIMO: StIll and moving video storyboard for the web scenario,” Multimed. Tools Appl., vol. 46, no. 1, pp. 47 – 69, Dec. 2010.
doi: https://doi.org/10.1007/s11042-009-0307-7

[5] J. Almeida, N. J. Leite, and R. D. S. Torres, “VISON: VIdeo Summarization for Online applications,” Pattern Recognit. Lett., vol. 33, no. 4, pp. 397 – 409, Sep. 2012.
doi: https://doi.org/10.1016/j.patrec.2011.08.007

[6] B. L. Yeo, and B. Liu, “Rapid scene analysis on compressed video.” IEEE Transactions on circuits and systems for video technology, vol 5, no. 6, pp. 533-544. Dec. 1995.
doi: https://doi.org/10.1109/76.475896

[7] G. Guan, Z. Wang, S. Lu, J.D. Deng, and D.D. Feng, “Keypoint-based keyframe selection”, IEEE Trans. Circuits Syst. Video Technol., vol. 23, no. 4, pp. 729-734, Apr. 2013.
doi: https://doi.org/10.1109/TCSVT.2012.2214871

[8] G. LoweDavid, “Distinctive image features from scale-invariant keypoints.” International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, Nov. 2004.
doi: https://doi.org/10.1023/B:VISI.0000029664.99615.94

[9] K. Zhang, W. L. Chao, F. Sha, and K. Grauman, “Video summarization with long short-term memory,” in Proc. Eur. Conf. Comput. Vis., pp. 766 – 782, Oct. 2016.
doi: https://doi.org/10.1007/978-3-319-46478-7_47

[10] B. Mahasseni, M. Lam, and S. Todorovic, “Unsupervised video summarization with adversarial lstm networks,” in Proc. IEEE Comput. Vis. Pattern Recognit., pp. 202 – 211, Jul. 2017.
doi: https://doi.org/10.1109/CVPR.2017.318

[11] B. Zhao, X. Li, and X. Lu, “HSA-RNN: Hierarchical structure-adaptive RNN for video summarization,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 7405 – 7414, Jun. 2018.
doi: https://doi.org/10.1109/CVPR.2018.00773

[12] S. E. F. De Avila, A. P. B. Lopes, A. Luz Jr, and A. de Albuquerque Araújo, “VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method.” Pattern Recognit. Lett., vol. 32, no. 1, pp. 56-68, Sep. 2011
doi: https://doi.org/10.1016/j.patrec.2010.08.004

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. Advances in Neural Information Processing Systems, pp. 1097 – 110, Dec. 2017.

[14] K. He, Z. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition.” in Proc. IEEE Comput. Vis. Pattern Recognit., pp. 770-778, Jun. 2016.
doi: https://doi.org/10.1109/CVPR.2016.90

[15] F. Schroff, D. Kalenichenko, D. and J. Philbin, “Facenet: A unified embedding for face recognition and clustering.” In Proc. IEEE Comput. Vis. Pattern Recognit., pp. 815-823, Jun. 2015.
doi: https://doi.org/10.1109/CVPR.2015.7298682

[16] N. Xu, L. Yang, Y. Fan, J. Yang, D. Yue, Y. Liang, and T. Huang, “Youtube-vos: Sequence-to-sequence video object segmentation.” in Proc. Eur. Conf. Comput. Vis., pp. 585 – 601, Sep. 2018.
doi: https://doi.org/10.1007/978-3-030-01228-1_36

[17] M. Gygli, H. Grabner, H. Riemenschneider, and L. V. Gool, (2014, September). “Creating summaries from user videos.” in Proc. Eur. Conf. Comput. Vis., pp. 505-520, Sep. 2014
doi: https://doi.org/10.1007/978-3-319-10584-0_33

[18] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. Devito, Z. Lin, A. Desmaison, L. Antiga, A. Lerer, “Automatic differentiation in pytorch”. in Proc. Conference and Workshop on Neural Information Processing Systems, pp. 1-4, Dec. 2017.
https://openreview.net/pdf?id=BJJsrmfCZ

저 자 소 개

박 경 태

- 2022년 2월 : 건국대학교 학사
- 2022년 3월 ~ 현재 : 건국대학교 전자정보통신공학과 석사과정
- ORCID : https://orcid.org/0000-0002-2565-2238
- 주관심분야 : 컴퓨터 비전, 객체 검출, 기계학습

김 원 준

- 2012년 8월 : 한국과학기술원(KAIST) 박사
- 2012년 9월 ~ 2016년 2월 : 삼성종합기술원 전문연구원
- 2016년 3월 ~ 2020년 2월 : 건국대학교 전기전자공학부 조교수
- 2020년 3월 ~ 현재 : 건국대학교 전기전자공학부 부교수
- ORCID : https://orcid.org/0000-0001-5121-5931
- 주관심분야 : 영상이해, 컴퓨터 비전, 기계학습, 패턴 인식

이 용

- 2003년 : 일본교토대학교 정보학과(공학박사)
- 2013년 9월 ~ 현재 : 한국과학기술정보연구원 기계학습데이터연구단 책임연구원
- 2021년 3얼 ~ 현재 : 과학기술연합대학원대학교 교수
- ORCID : http://orcid.org/0000-0001-5142-6106
- 주관심분야 : 인공지능, 시각지능, 사물인터넷, 공간데티어, 등

저 자 소 개

장 래 영

- 2018년 8월 : 한남대학교 컴퓨터공학과 (공학박사)
- 2019년 9월 ~ 현재 : 한국과학기술정보연구원 기계학습데이터연구단 선임연구원
- ORCID : http://orcid.org/0000-0001-9391-4028
- 주관심분야 : MLOps. Kubernets, 인공지능 등

최 명 석

- 2005년 : 한국과학기술원 전산학과 (공학박사)
- 2005년 ~ 현재 : 한국과학기술정보연구원 기계학습데이터연구단 단장
- ORCID : http://orcid.org/0000-0003-4821-3390
- 주관심분야 : 오픈사이언스, 연구데이터관리, 인공지능, 등