Journal of The Institute of Electronics and Information Engineers Vol.58, NO.12, December 2021 https://doi.org/10.5573/ieie.2021.58.12.35
ISSN 2287-5026(Print) / ISSN 2288-159X(Online)

논문 2021-58-12-5

주기적 행동 검출을 위한 멀티스케일 U-Net

(Multi-scale U-Net for Periodic Motion Detection)

유 철 환*, 김 호 원*, 한 병 옥*, 장 재 윤*, 유 장 희**

(Cheol-Hwan Yoo, Ho-Won Kim, Byung-Ok Han, Jae-Yoon Jang, and Jang-Hee Yoo©)

요 약

동영상에 포함된 반복적, 주기적 구간을 검출하기 위한 기술은 컴퓨터 비전 분야에서 활발히 연구되고 있다. 기존의 기법들은 일반적으로 반복적 구간 검출을 위한 중간 표현으로서 자기 유사성 행렬(SSM)을 생성하여 활용한다. 그러나 기존의 기법들은 단일 스케일에서의 자기 유사성 행렬의 활용으로 인해 다양한 길이 및 스케일의 반복적 행동을 포함한 동영상에 대해 검출 정확도가 떨어지는 한계점을 갖는다. 이러한 한계점을 극복하기 위해 제안하는 네트워크의 인코더에서는 먼저 3차원 합성곱 신경망의 여러 계층에서 추출된 특징 벡터를 활용하여 다양한 시간적 스케일에 대한 정보를 갖는 자기 유사성 행렬을 생성한다. 이렇게 생성된 자기 유사성 행렬들을 멀티 스케일 특징 앙상블 모듈을 통해 멀티 스케일 U-Net의 입력으로 제공함으로써 동영상 내 다양한 길이의 반복적 구간을 효율적으로 검출한다. 제안하는 기법은 Countix, PERTUBE 데이터셋에서의 실험을 통해 기존의 핸드 크래프트 특징 기반의 기법들뿐만 아니라 딥러닝을 활용한 최신 기법들보다 우수한 검출 성능을 보였다.

Abstract

Recently, techniques for detecting repetitive and periodic segments in a video have been extensively studied in the field of computer vision. Conventional methods typically generate and utilize a self-similarity matrix as an intermediate representation for identifying repetitive segments in a video. However, these methods rely on a single-scale self-similarity matrix(SSM) and thus have a limitation that classification accuracy drops for videos including repetitive segments with various lengths and scales. To solve these problems, the encoder of the proposed network firstly generates self-similarity matrices, which incorporate information on various temporal scales by utilizing feature vectors extracted from multiple layers of the 3D CNN. By providing generated self-similarity matrices as input of a multi-scale U-Net through a multi-scale feature ensemble module, repetitive segments of various lengths in the video can be efficiently detected. Extensive experiments on the Countix and PERTUBE datasets demonstrate that the proposed network not only outperforms most hand-craft feature-based methods but also the latest deep learning-based methods.

Keywords : Multi-scale U-Net, 3D CNN, Periodicity, Repetition

Ⅰ. 서 론

걷기, 손 흔들기, 다리 떨기, 숨쉬기와 같이 반복적이고 주기적인 행동(repetitive and periodic motion)은 일상생활에서 흔하게 발생하는 현상이다[1, 2]. 비디오에 포함된 이러한 반복성에 대한 검출 기술은 변화 검출(change detection), 주기 예측(period estimation) 등의 컴퓨터 비전 분야에서의 응용뿐만 아니라 자폐 아동의 제한적이고 반복적인 행동(restricted and repetitive behaviors) 검출 등의 의료분야에 이르기까지 다양한 분야에서의 수요로 인해 활발히 연구되고 있다.

이러한 주기적 행동 검출에 관한 연구들은 크게 2가지 접근 방법으로 분류될 수 있다. 첫 번째 방법은 핸드 크래프트 특징(hand-craft features)을 활용한 비 딥러

*비회원, **정회원, 한국전자통신연구원 인공지능연구소(AI Research Laboratory, ETRI)
© Corresponding Author(E-mail : jhy@etri.re.kr)
※ 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2019-0-00330, 영유아/아동의 발달장애 조기선별을 위한 행동·반응 심리인지 AI 기술 개발).
Received ; September 2, 2021 Revised ; September 24, 2021 Accepted ; September 30, 2021

Copyright © The Institute of Electronics and Information Engineers.

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.

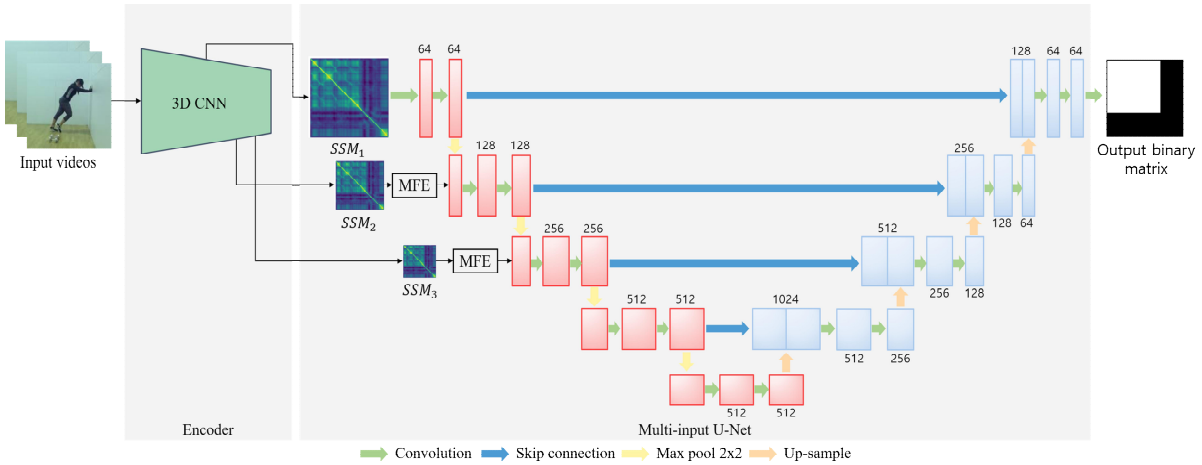


그림 1. 제안하는 주기적 행동 검출 방법의 전체 구조도
Fig. 1. The overall architecture of the proposed method for periodic motion detection.

닝 기반의 기법들[3, 5]이다. P-MUCOS[3]는 주기적 행동 검출을 시퀀스 내의 공통적인 서브 시퀀스(common sub-sequence)를 찾는 문제로 재정의하였다. 이를 위해 비디오의 각 프레임에 대해 향상된 조밀 궤적(improved dense trajectories)에 기반한 특징을 추출하여 거리 행렬을 생성하였고, 생성된 거리 행렬에 MUCOS[4]를 적용하여 주기적 행동 구간을 검출하는 방법을 제안하였다. Panagiotakis[5]는 P-MUCOS 알고리즘이 주 대각선(main diagonal)을 주요한 공통성(major commonality)으로 검출하는 자명해(trivial solution) 문제를 해결하기 위해 거리 행렬에 대한 필터링 및 최적화 기반의 검출 결과 개선의 2단계로 구성된 P-MUCOS-S2를 제안하였다. 이러한 핸드 크래프트 특징 기반의 기법들은 별도의 훈련데이터 없이 비지도적(unsupervised) 방법을 통해 효율적으로 주기적 행동 구간을 검출할 수 있다는 장점이 있다.

두 번째 방법으로서, 딥러닝의 우수한 성능에 힘입어 최근에는 딥러닝 기반의 기법들이 활발히 연구되고 있다. ReActNet[6]는 주기적 행동 검출 분야에 처음으로 딥러닝을 도입하였다. ReActNet은 기존의 비 딥러닝 기법들과 마찬가지로 비디오 각 프레임의 특징 표현(feature representation)에 기반한 거리 행렬 생성과 이러한 거리 행렬을 이용한 반복적 구간 검출의 프레임워크 구성을 그대로 채택하였다. 즉, 기존의 핸드 크래프트 특징과 달리 VGG-19[7]를 이용한 특징 표현을 제안하였으며, 생성된 거리 행렬에 대해 오토인코더(Auto-encoder) 형태의 딥러닝 구조를 활용하여 반복적 구간 검출의 성능을 향상시켰다. 또한 딥러닝 네트워크의 효율적인 학습을 위해 기존의 반복적 구간의 시작과 끝의 형태로 구성된 비디오에 대한 어노테이션(annotation)을 정사각형 형태의 이진 행렬로 재구성하여 활용하는 방법을 제안하였다. 이와 유사하게 최근의 RepNet[8]는 ResNet-50[9] 기반의 특징 표현을 활용하여 자기 유사성 행렬(Self-Similarity Matrix, SSM)을 생성하였으며, 합성곱 필터(convolutional filter)와 Transformer[10]로 구성된 주기 예측기(period predictor) 모듈을 통해 프레임당 주기 길이(per frame period length) 및 프레임당 주기성 분류(per frame periodicity classification)를 동시에 수행하는 구조를 제안하였다. 더불어 주기적 행동 검출 분야에서의 대규모 데이터셋의 부재를 해결하기 위해 행동 인식 분야에서 널리 쓰이는 Kinetics[11] DB 중 반복적인 행동이 나타나는 비디오를 선별하여 Countix[8] DB를 새롭게 구성하였다.

그러나 이러한 딥러닝을 이용한 주기적 행동 검출 방법들이 주목할만한 성능 향상을 가져왔음에도 불구하고 기존의 기법들은 단일 스케일(single-scale)에서 생성된 거리 행렬 혹은 자기 유사성 행렬의 활용으로 인해 다양한 길이 및 스케일의 주기적 행동을 포함한 동영상에 대해 검출 정확도가 떨어지는 한계점을 갖는다. 또한 Transformer 기반의 딥러닝 네트워크가 자연어 처리, 이미지 분류 등의 분야에서 우수한 성능을 보임에도 불구하고 inductive bias를 효율적으로 학습하기 위해서는 대규모의 데이터셋이 필요하며, 안정적인 학습을 위해 사전 학습된 모델에 대한 의존성, optimizer와 학습 스케줄에 대한 세심한 고려가 필요하다는 한계점이 여전히 존재한다[12].

본 논문에서는 이러한 기존 기법들의 한계점을 극복하기 위하여, 그림 1에 제시된 것과 같이 3차원 합성곱

신경망(3D CNN) 기반 백본 네트워크(back bone network)와 멀티 스케일 U-Net[13]의 결합을 통해 다양한 길이의 반복적 구간을 포함한 동영상에 대해서도 우수한 검출 성능을 내는 방법을 제시한다. 본 논문의 구성은 2장에서 제안하는 방법에 관하여 기술하고, 3장에서는 제안하는 방법의 성능을 검증하기 위한 실험 방법 및 실험 결과를 기술한다. 마지막으로 4장에서는 결론 및 향후 연구 방향에 대해 논의한다.

Ⅱ. 본 론

1. 인코더 네트워크

제안하는 주기적 행동 구간 검출 구조의 인코더 네트워크(encoder network)는 그림 1에서 볼 수 있듯이 3D CNN으로 구성되어 있다. 기존의 핸드 크래프트 특징 기반 기법들[3, 5]은 비디오를 구성하는 프레임마다 향상된 조밀 궤적과 같은 특징 표현을 사용하였으며, 최근의 딥러닝 기반 기법들[6, 8] 또한 VGG-19나 ResNet-50과 같은 2D CNN을 사용하여 각각의 프레임마다 특징을 추출하였다. 그러나 이러한 2D CNN 기반의 인코더는 비디오를 구성하는 프레임 간의 시간적 관계(temporal relationship)를 제대로 활용하지 못하는 한계점이 있다. 따라서 본 연구에서는 시간적 패턴(temporal pattern)을 학습하기에 용이한 3D CNN 기반의 인코더를 제안하였으며, 특히 네트워크의 효율과 성능 간의 균형을 고려하여 3D ResNet-18[14]을 활용하였다. 3D ResNet-18은 1개의 7×7×7 합성곱 계층(convolutional layer), 2개의 풀링 계층(pooling layers), 그리고 8개의 레지듀얼 블록(residual block)으로 구성되며, 각각의 레지듀얼 블록은 2개의 3×3×3 합성곱 계층과 단축 경로(shortcut-path)로 구성된다.

2. 멀티 스케일 U-Net

동영상 내의 반복적인 동작(repeating motion)은 짧게는 몇 초, 길게는 수분에 걸쳐 일어날 수 있다. 기존의 방법들[6, 8]에서 제안한 딥러닝 네트워크는 이러한 시간적 스케일(temporal scale)에 대한 명백한 고려의 부족으로 인해 다양한 길이의 주기적 행동을 포함한 동영상에 대해 정확도가 떨어지는 한계점을 지닌다. 이를 해결하기 위해 시간적 멀티 스케일 특징(temporal multi-scale features)을 충분히 활용할 수 있는 딥러닝 네트워크를 그림 1과 같이 제안하였다. 먼저 멀티 스케일 U-Net(multi-scale U-net)의 활용을 위해 앞서 언급

그림 2. (a) 자기 유사성 행렬, (b) 네트워크 출력, (c) 이진 행렬 참값
Fig. 2. (a) Self-similarity matrix, (b) Network output, (c) Binary matrix ground truth.

된 인코더의 여러 계층에서 특징 벡터를 추출하여 자기 유사성 행렬을 생성함으로써 동영상 내의 다양한 스케일의 주기적, 반복적 행동에 대한 특징을 학습할 수 있도록 하였다. 구체적으로 인코더는 공간 크기(spatial size) 224×224 및 N개의 프레임으로 구성된 비디오를 입력으로 받아 멀티 스케일 특징 표현을 위해 conv1, conv3_2, 그리고 마지막 average pooling 계층의 출력으로부터 각각의 잠재적 특징 벡터(latent feature vector)를 추출한다. k번째 스케일의 n번째 프레임에서 추출된 특징 벡터를 x_k^n라고 할 때 k번째 스케일의 자기 유사성 행렬은 S_k는 다음과 같이 정의된다.

$$S_k^{i,j} = F(x_k^i, x_k^j),\ i,j = 1,\cdots,N, \tag{1}$$

여기서 $F(\cdot)$는 특징 벡터 간의 유사도 함수(similarity function)이며, 특징 벡터 간의 크기와 각도를 동시에 고려하여 유사도를 측정하기 위해 벡터 간의 내적, $F(\vec{a},\vec{b}) = \vec{a} \cdot \vec{b}$을 유사도 함수로 사용하였다. 최종적으로 그림 1과 같이 N×N, $N/2$×$N/2$, 그리고 $N/4$×$N/4$ 공간 크기를 가지는 3가지 레벨의 자기 유사성 행렬이 생성되어 멀티 스케일 U-Net의 입력으로 사용된다. 이러한 멀티 스케일 특징 혹은 멀티 스케일 이미지를 단일 U-Net의 입력으로 활용하는 것이 효율적이라는 것은 깊이 맵 초고해상도(depth map super-resolution) 및 객체 검출과 같은 다양한 컴퓨터 비전 분야에서 입증되었다[15, 16].

첫 번째 스케일의 자기 유사성 행렬 S_1을 제외한 다른 스케일의 유사성 행렬을 U-Net의 입력으로 활용하기 위해 멀티 스케일 특징 앙상블(Multi-scale Feature Ensemble, MFE) 모듈을 활용하였다. MFE 모듈은 효율성을 고려하여 하나의 3×3 합성곱 계층, 배치 정규화(batch normalization), 그리고 활성화 함수(activation function)로 구성하였으며, 합성곱 계층의 출력 차원은

U-Net의 특징 맵과의 원소별 덧셈(element-wise summation)을 위해 각각 64, 128로 설정하였다. 멀티 스케일 유사성 행렬에 대한 MFE 모듈의 출력은 동영상 내 다양한 길이의 반복적인 동작에 대해 상호보완적인 정보를 갖고 있을 것으로 예상되며, 이러한 멀티 스케일에 대한 고려가 주기적 행동 구간 검출의 성능 향상에 효과적이라는 것을 다음 장에서 실험을 통하여 보다 자세히 제시될 것이다. 최종적으로 그림 1과 같은 멀티 스케일 U-Net 구조를 거쳐 그림 2(b)와 같은 $N \times N$ 공간 크기의 0과 1 사이의 값을 갖는 이진 행렬을 출력한다.

3. 손실 함수

제안하는 네트워크를 학습하기 위해 ReActNet[6]에서 제안하는 훈련 방법을 채택하였다. ReActNet에서는 주기적 행동 구간 검출 문제를 프레임별 이진 분류 문제로 정의하였으며, 효율적인 네트워크 학습을 위해 기존의 비디오 어노테이션을 입력 거리 행렬과 같은 크기의 정사각형 형태의 이진 행렬 A로 표현하였다. 이진 행렬 A의 값 $a_{i,j}$은 i번째 프레임과 j번째 프레임이 모두 반복적 구간에 포함될 때는 1, 그렇지 않을 때는 0의 값으로 할당되었다. 입력 비디오로부터 생성된 자기 유사성 행렬, 네트워크 출력 및 이진 행렬 참값은 그림 2의 예시와 같다. 이렇게 구성된 이진 행렬 어노테이션을 활용하여 네트워크를 학습하기 위해 다음과 같은 이진 크로스 엔트로피 손실 함수(binary cross entropy loss function)를 사용하였다.

$$L_{total} = E(y, \hat{y}) = -y log \hat{y} - (1-y)\log(1-\hat{y}), \quad (2)$$

여기서 $y \in \{0, 1\}$은 픽셀별 클래스 레이블 참값(ground truth class label)이고, $\hat{y}$ 은 네트워크 예측 확률값이다.

Ⅲ. 실 험

1. 딥러닝 학습

제안하는 네트워크를 학습하기 위해 PyTorch 프레임워크를 사용하였으며, 학습의 반복(iteration)마다 각각의 비디오 영상을 256×256 크기로 조정 후 중심 크랍(center crop)을 통해 획득한 224×224 해상도의 비디오를 사용하였다. 비디오를 구성하는 프레임 개수 N은 태스크의 특성을 고려하여 별도의 샘플링 등의 처리 없이 그대로 사용하였으며, 배치 크기는 1을 사용하였다. 제안하는 구조가 충분히 수렴할 수 있도록 총 20 에폭(epochs) 동안 학습을 진행하였다. 본 논문에서는 아담 옵티마이저(Adam optimizer)[17]를 사용하였으며, 학습율(learning rate)은 0.00001로 설정하였다. 더불어 안정적인 학습을 위해 L2 regularization을 추가하였으며 regularization 파라미터(parameter) λ는 실험적으로 0.0001로 설정하였다.

2. 데이터셋 및 평가지표

제안하는 기법의 우수성을 보이기 위해 Countix[8], PERTUBE[3] 데이터셋을 사용하였다. PERTUBE 데이터셋은 다양한 반복적, 주기적 행동(사람 활동, 물체 움직임 등)을 포함하고 있으며, YouTube에서 수집된 50여 개의 비디오로 구성되어 있다. 각각의 비디오 프레임은 프레임이 반복적 구간에 속하는지 여부에 대해 태깅되어있다. Countix 데이터셋은 반복 횟수 계산(repetition counting) 및 주기적 행동 구간 검출에 사용되는 가장 큰 규모의 데이터셋이다. Countix 데이터셋은 Kinetics[11] 데이터셋의 부분 데이터셋으로서 반복적 행동 구간과 총 반복 횟수에 대해 태깅되어있다. Countix 데이터셋은 학습, 검증, 테스트에 각각 4,588, 1,450, 그리고 2,719장의 비디오로 구성되어 있으며, 본 논문에서는 주기적 행동 검출 학습을 위해 Countix의 학습 데이터를 사용하였고 최신 알고리즘들과의 비교 및 대조 실험을 위해 PERTUBE 데이터셋을 테스트셋으로 활용하였다. 평가지표에 대해서는 기존 방법들[3, 5, 6, 8]의 성능 평가 프로토콜을 따라 주기적 행동 검출 태스크를 프레임별 이진 분류 문제로 간주하였으며, 따라서 recall R, precision P, F_1-score, 그리고 overlap O에 대한 성능을 측정하였다.

3. 실험결과

Power spectrum baseline[18], P-MUCOS[3], P-MUCOS-S2[5]를 포함한 기존의 핸드 크래프트 특징 기반의 기법들과 최근의 딥러닝 기반의 ReActNet[6] 기법에 대해 제안하는 알고리즘과의 비교실험을 진행하였다. 표 1에서 제시된 것과 같이, 제안하는 기법은 PERTUBE 데이터셋에 대해 모든 평가지표 면에서 기존의 방법들에 비해 우수한 성능을 보였다. ReActNet의 경우 논문에 제시된 수치에 비해 recall이 높고 precision이 낮게 측정되었는데 이는 ReActNet을 구성하는 합성곱 계층의 차원이 낮아 네트워크의 캐패시티(capacity)가 Countix 데이터셋에서 학습되기에 충분하지 못한 것으로 간주되었다.

또한 제안하는 구조에서 멀티 스케일 자기 유사성 행

표 1. PERTUBE 데이터셋에서의 주기적 행동 검출 결과
Table 1. Results of Periodic motion detection on the PERTUBE Dataset.

Method	Recall	Precision	F1	Overlap
Power spectrrum baseline	0.793	0.611	0.668	0.573
P-MUCOS	0.841	0.757	0.77	0.677
P-MUCOS-S2	0.925	0.802	0.839	0.759
ReActNet	0.962	0.667	0.767	0.653
Proposed	**0.983**	**0.830**	**0.893**	**0.817**

표 2. PERTUBE 데이터셋에서의 멀티 스케일 SSM의 효과에 대한 대조 실험. 'Mk', 'M123'은 각각 k-th 레벨의 SSM 및 멀티 스케일 SSM이 U-Net의 입력으로 사용된다는 것을 나타냄
Table 2. Ablation studies for effectiveness of multi-scale SSM on the PERTUBE Dataset. 'Mk', 'M123' denotes SSM at k-th level and multi-scale SSM is fed into a U-net, respectively.

Method	Recall	Precision	F1	Overlap
3D ResNet-18 + M_1	0.976	0.827	0.882	0.808
3D ResNet-18 + M_2	0.978	**0.831**	0.891	0.814
3D ResNet-18 + M_3	0.899	0.776	0.819	0.712
3D ResNet-18 + M_{123}	**0.983**	0.830	**0.893**	**0.817**

럴과 멀티 스케일 U-Net의 활용을 통한 효과를 확인하기 위한 대조 실험을 진행하였다. 실험을 위해 제안한 방법, 그리고 3D ResNet-18의 k번째 단계에서 추출된 특징으로부터 생성된 각각의 자기 유사성 행렬을 입력으로 사용한 결과를 비교하였다. 2번째, 3번째 단계의 스케일에서 생성된 유사성 행렬은 3D ResNet-18의 시간적 풀링(temporal pooling)으로 인해 각각 $N/2\times N/2$, 그리고 $N/4\times N/4$ 공간 크기를 갖는다. 따라서 공정한 비교실험을 위한 이중 선형보간법을 통해 $N\times N$의 공간 크기로 해상도를 높여주었다. 표 2에서 볼 수 있듯 3D ResNet-18의 k번째 단계에서 추출된 특징으로 생성된 자기 유사성 행렬을 단독으로 사용할 때에 비해 제안하는 멀티 스케일 U-Net 기반의 방법이 대부분의 평가지표에서 더 우수한 성능을 보임을 확인할 수 있다. 이는 다양한 스케일에서의 특징 표현과 멀티 스케일 U-Net의 활용을 통해 동영상 내에 존재하는 다양한 길이의 주기적 행동을 효과적으로 검출할 수 있음을 확인하였다. 제안하는 방법을 이용하여 주기적 행동 구간 검출에 대한 결과 예시는 그림 3 및 그림 4와 같다. 비주기적 구간은 빨간색, 주기적 구간은 초록색으로 표시되었

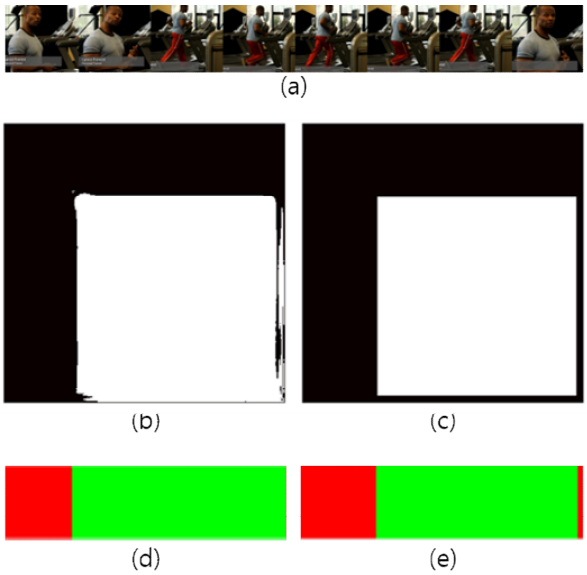


그림 3. (a) PERTUBE 데이터셋의 "TreadmillCut" 입력 비디오, (b) 네트워크 출력, (c) 이진 행렬 참값, (d) 프레임별 주기성 예측값, (e) 프레임별 주기성 참값
Fig. 3. (a) Input videos of "treadmillCut" from the PERTUBE datasets, (b) Network output, (c) Binary matrix ground truth, (d) Per frame periodicity prediction, (e) Per frame periodicity ground truth.

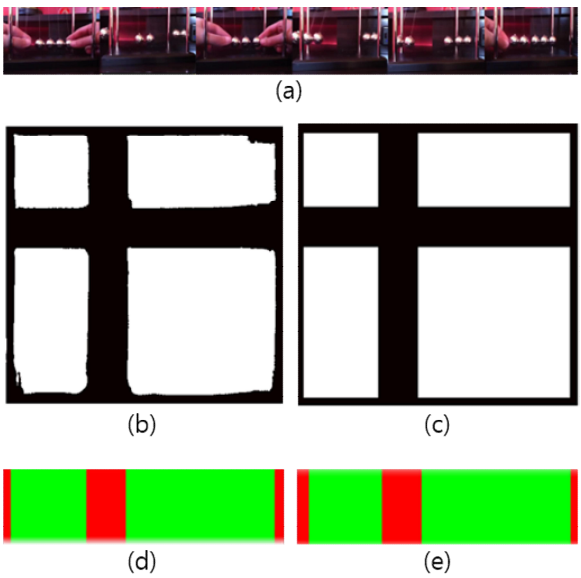


그림 4. (a) PERTUBE 데이터셋의 "NewtonBallCut" 입력 비디오, (b) 네트워크 출력, (c) 이진 행렬 참값, (d) 프레임별 주기성 예측값, (e) 프레임별 주기성 참값
Fig. 4. (a) Input videos of "NewtonBallCut" from the PERTUBE datasets, (b) Network output. (c) Binary matrix ground truth, (d) Per frame periodicity prediction, (e) Per frame periodicity ground truth.

다. 제안하는 네트워크를 사용했을 때 그림 3(d) 및 그림 4(d)의 결과와 같이 주기적 행동 구간을 정확하게 검출함을 확인할 수 있다.

Ⅳ. 결 론

본 논문에서는 동영상에 포함된 다양한 길이와 스케일을 가진 주기적 행동 구간을 효율적으로 검출하기 위해 3차원 합성곱 신경망과 멀티 스케일 U-Net의 결합을 통한 새로운 딥러닝 네트워크를 제안하였다. 3차원 합성곱 신경망의 여러 계층에서의 특징 추출을 통해 다양한 시간적 정보를 포함한 자기 유사성 행렬을 생성하였으며, 멀티 스케일 U-Net과의 결합을 통해 기존의 방법들에 비해 우수한 성능을 보임을 확인하였다. 이러한 주기적 행동 구간 검출 기술의 발전을 통해 컴퓨터 비전 및 의료 등의 다양한 분야에 활용될 수 있을 것으로 기대된다.

REFERENCES

[1] Z. Li, B. Ding, J. Han, R. Kays, and P. Nye, “Mining periodic behaviors for moving objects,” in Proc. of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1099-1108, Phuket, Thailand. Jan. 2010.

[2] A. K. Chanda, C. F. Ahmed, M. Samiullah, and C. K. Leung, “A new framework for mining weighted periodic patterns in time series databases,” Expert Systems with Applications, Vol. 79, pp. 207-224, Aug. 2017.

[3] C. Panagiotakis, G. Karvounas, and A. Argyros, “Unsupervised detection of periodic segments in videos,” in Proc. of IEEE International Conference on Image Processing, pp. 923-927, Athens, Greece, Oct. 2018.

[4] C. Panagiotakis, K. Papoutsakis, and A. Argyros, “A graph-based approach for detecting common actions in motion capture data and videos,” Pattern Recognition, Vol. 79, pp. 1-11, Feb. 2018.

[5] C. Panagiotakis, and A. Argyros, “A two-stage approach for commonality-based temporal localization of periodic motions,” in Proc. of International Conference on Computer Vision Systems, pp. 366-375, Thessaloniki, Greece, Sep. 2019.

[6] G. Karvounas, I. Oikonomidis, and A. Argyros, A, “ReActNet: Temporal Localization of Repetitive Activities in Real-World Videos,” *arXiv preprint arXiv:1910.06096.*

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

[8] D. Dwibedi, Y. Aytar, J. Tompson, P. Sermanet, and A. Zisserman, “Counting out time: Class agnostic video repetition counting in the wild,” in Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10387-10396, Virtual, Jun. 2020.

[9] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, Las Vegas, USA, Jun. 2016.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in Proc. of the 31th Conference on Neural Information Processing Systems (NIPS), pp. 5998-6008, Long Beach, USA, Dec. 2017.

[11] J. Carreira, and A. Zisserman, “Quo vadis, action recognition? a new model and the kinetics dataset,” in Proc. of the IEEE/CVFConference on Computer Vision and Pattern Recognition (CVPR), pp. 6299-6308, Hawaii, USA, Jul. 2017.

[12] L. Liu, X. Liu, J. Gao, W. Chen, and J. Han, “Understanding the difficulty of training transformers,” *arXiv preprint arXiv:2004.08249.*

[13] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Proc. of International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 234-241, Munich, Germany, Oct. 2015.

[14] H. Kataoka, T. Wakamiya, K. Hara, and Y. Satoh, “Would mega-scale datasets further enhance spatiotemporal 3D CNNs?,” *arXiv preprint arXiv:2004.04968.*

[15] C. Guo, C. Li, J. Guo, R. Cong, H. Fu, and P. Han, “Hierarchical features driven residual learning for depth map super-resolution,” IEEE Transactions on Image Processing, Vol. 28, No. 5, pp. 2545-2557, May 2019.

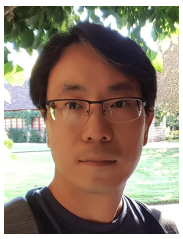
[16] Y. Pang, T. Wang, R. M. Anwer, F. S. Khan, and L. Shao, L, “Efficient featurized image pyramid network for single shot detector,” in Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7336-7344, Long Beach, Jun. 2019.

[17] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:*

1412.6980, 2014.
[18] R. Cutler, and L. S. Davis, "Robust real-time periodic motion detection, analysis, and applications," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, pp. 781-796, Aug. 2000.

저 자 소 개

유 철 환(비회원)
2014년 고려대학교 전기전자전파 공학부 학사 졸업.
2020년 고려대학교 전기전자공학과 석·박사통합과정 박사 졸업.
2020년～현재 한국전자통신연구원 인공지능연구소 연구원
<주관심분야 : 컴퓨터비전, 영상처리, 딥러닝>



김 호 원(비회원)
1997년 경북대학교 전자공학과 학사 졸업.
1999년 한국과학기술원 전기및전자공학과 석사 졸업.
2004년 한국과학기술원 전기및전자공학과 박사 졸업.
2004년～2005년 LG전자 모바일멀티미디어연구소 선임연구원
2006년～현재 한국전자통신연구원 인공지능연구소 책임연구원
<주관심분야 : 컴퓨터비전, 딥러닝, 그래픽스>

한 병 옥(비회원)
2008년 중앙대학교 컴퓨터공학과 학사 졸업.
2010년 한국과학기술원 로봇공학학제전공 석사 졸업.
2016년 한국과학기술원 전산학과 박사 졸업.
2016년～현재 한국전자통신연구원 인공지능연구소 선임연구원
<주관심분야 : 컴퓨터비전, 딥러닝, 영상처리>

장 재 윤(비회원)
2013년 국립한밭대학교 제어계측공학과 학사 졸업.
2019년 과학기술연합대학원대학교 정보통신기술 컴퓨터 소프트웨어 박사 졸업.
2019년～현재 한국전자통신연구원 인공지능연구소 연구원
<주관심분야 : 머신러닝, 영상처리, 얼굴인식>

유 장 희(정회원)-교신저자
1988년 한국외국어대학교 물리학과 학사 졸업.
1990년 한국외국어대학교 전산학과 석사 졸업.
2004년 영국 University of Southampton 전자 및 컴퓨터과학 박사 졸업.
1989년～현재 한국전자통신연구원 인공지능연구소 책임연구원
2005년～현재 한국저작권위원회 감정인, 현)감정전문위원
2007년～현재 과학기술연합대학원대학교 전임교수, 캠퍼스 대표교수
2014년～현재 경찰청 과학수사자문위원
2014년～2015년 University of Washington 방문학자
2017년 한국SW감정평가학회 회장
2018년～2020년 국가지식재산위원회 전문위원
<주관심분야 : 컴퓨터 비전, 인공지능, 생체인식, 휴먼 모션분석, HCI 및 지능형 로봇>