

일반논문 (Regular Paper)

방송공학회논문지 제27권 제1호, 2022년 1월 (JBE Vol.27, No.1, January 2022)

<https://doi.org/10.5909/JBE.2022.27.1.124>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

## 실시간 온라인 수업 및 시험 태도 데이터 세트 설계 및 구현

김 준 식<sup>a)</sup>, 이 찬 휘<sup>a)</sup>, 송 혁<sup>b)</sup>, 권 순 철<sup>c)†</sup>

### Real-time Online Study and Exam Attitude Dataset Design and Implementation

Junsik Kim<sup>a)</sup>, Chanhwi Lee<sup>a)</sup>, Hyok Song<sup>a)</sup>, and Soonchul Kwon<sup>c)†</sup>

#### 요 약

최근 코로나바이러스감염증-19(COVID-19)로 인해 온라인 원격 수업과 비대면 시험으로 인해 수업 태도 및 시험 부정행위에 대한 관리가 어려움을 겪고 있다. 따라서 온라인으로 학생들의 행동을 자동으로 인식하고 검출하는 시스템이 필요하다. 사람의 행동을 인식하는 행동 인식의 경우 컴퓨터 비전에서 많이 연구되는 기술 중 하나이다. 이러한 시스템을 개발하기 위해서는 온라인 수업 및 시험에서 주요 정보가 될 수 있는 사람의 팔 움직임 정보와 주변 물체에 대한 정보를 포함하는 데이터가 필요하다. 기존 데이터 세트는 여러 분야에 대해 분류를 하거나 일상생활 행동으로 구성되어 있어 본 시스템에 적용시키기에 어려움이 있다. 본 논문에서는 실시간으로 진행되는 온라인 시험 및 수업에서 태도를 분류할 수 있는 데이터 세트를 제시한다. 또한, 기존의 행동 인식 데이터 세트와의 비교를 통해 제안된 데이터 세트가 올바르게 구성되었는지를 보여준다.

#### Abstract

Recently, due to COVID-19, online remote classes and non-face-to-face exams have made it difficult to manage class attitudes and exam cheating. Therefore, there is a need for a system that automatically recognizes and detects the behavior of students online. Action recognition, which recognizes human action, is one of the most studied technologies in computer vision. In order to develop such a technology, data including human arm movement information and information about surrounding objects, which can be key information in online classes and exams, are needed. It is difficult to apply the existing dataset to this system because it is classified into various fields or consists of daily life action. In this paper, we propose a dataset that can classify attitudes in real-time online tests and classes. In addition, it shows whether the proposed dataset is correctly constructed through comparison with the existing action recognition dataset.

Keyword : dataset construction, online study and exam attitude, deep neural network

## I. 서론

COVID-19로 인한 국내 원격 수업 통계(2020년 9월, 교육부)에 따르면, 유, 초, 중, 고등학교 중 8,000개의 학교가 원격 수업을 진행하였다. 또한, 대학교(2020년 9월, 교육부)의 경우 59%가 전면 비대면, 24.1%가 거리두기에 따른 비대면을 진행하였다<sup>[1]</sup>. 이러한 원격 수업에 따라 학생들의 행동을 자동으로 인식하여 관리하는 기술이 필요하다.

컴퓨터 비전 중 행동 인식 분야는 사람의 행동을 인식하고 분류한다. 대표적인 행동 인식 방식은 두 가지 접근 방식이 존재한다. 첫 번째로 최근 많이 시도되는 Spatial Temporal Graph Convolutional Network(ST-GCN), Channel-wise Topology Refinement Graph Convolution Network(CTR-GCN) 등 관절 정보를 추출하여 관절 정보를 가지고 행동을 인식하는 방식이 있다<sup>[2,3]</sup>. 하지만 온라인 시험 부정행위 행동의 경우 정상적인 행동과의 관절 정보의 차이가 거의 없으므로 특징점을 찾기 어렵다. 이 때문에 우리는 두 번째 방식인 이미지 자체를 학습하는 방식을 채택하였다. 하지만 행동의 특성상 단일 프레임을 보고 어떤 행동인지 예측하는 것은 어렵기 때문에 연속된 프레임을 보고 행동을 인식하여야 한다. 연속된 이미지를 가진 온라인 시험 및 수업 행동 인식을 위한 데이터 세트는 현재 존재하지 않고, 대표적으로 알려진 행동 인식 데이터 세트에 존재하는 클래스들은 온라인 시험 및 수업 행동 인식을 위한 클래스와는 거리가 멀다. 이 때문에 우리는 본 논문에서 온라인 시험 및 수업 행동 인식을 위한 행동 동영상 데이터 세트를 제안한다.

현재 연속된 이미지 학습 방식을 통해 높은 성능을 내는

Two-stream Inflated 3D ConvNet(I3D) 모델과 Two-stream fusion 모델의 근간이 되는 Convolutional 3D(C3D) 모델을 통해 우리가 제안한 데이터 세트의 활용 가치를 확인하였다<sup>[4,5,6]</sup>. 데이터 세트는 총 2가지로 제작되었다. 첫 번째로 온라인 수업 태도 데이터 세트이다. 온라인 수업 태도 데이터 세트는 오픈드려 자기, 노트북 조작, 하품, 공부하기, 스마트폰 조작, 태블릿 조작, 줄기 총 7가지 클래스로 정의 하였고, 총 6,510개의 동영상 데이터 세트이다. 두 번째로 온라인 시험 태도 데이터 세트는 책 넘기기, 양팔 내리기, 노트북 조작, 문제 풀기, 스마트폰 조작, 태블릿 조작 총 6가지 클래스로 정의 하였고, 총 4,042개의 동영상 데이터 세트이다.

본론에서는 우리가 제안한 데이터 세트에 대한 더 자세한 내용을 확인할 수 있고, 데이터 세트의 촬영 환경에 대한 소개가 있다. 각 행동마다의 예시 이미지를 확인할 수 있다. 실험 및 실험 결과에서는 우리가 제안한 데이터 세트를 C3D 모델에 적용하여 실험한 방법과 그 결과를 확인할 수 있다. C3D 모델에 학습하기 위해 데이터 세트를 가공하였고, 대표적인 행동 인식 데이터 세트인 UCF101과 HMDB51과 그 결과를 비교하였을 때, 더 높은 정확도를 보이거나 비슷한 수준의 정확도를 보이는 것을 확인하였다.<sup>[7,8]</sup>

## II. 본론

### 1. 데이터 세트의 구성

온라인 수업 및 시험 태도를 분석하기 위해 데이터 세트는 다음과 같은 조건들을 고려하였다. 첫째는 상황에서 제시될 수 있는 행동으로 구성되어야 한다. 수업 및 시험에 관련되지 않는 행동(예: 자전거 타기)은 모델의 학습을 방해할 뿐만 아니라 실제 행동 분류할 때 난이도를 어려워지게 한다. 따라서 필기, 책 넘기기와 같은 상황과 관련된 행동들로 각 데이터 세트를 구성하였다. 둘째는 상황에서 사용될 수 있는 물품을 고려해야 한다. 구성한 데이터 세트는 수업과 시험에서 사용될 수 있는 책, 스마트폰, 태블릿, 노트북을 추가하여 구성하였다. 셋째는 정면에서 카메라가 감독한다고 가정하며 손과 머리가 모두 보일 수 있도록 구성한다. 마지막으로 행동만 하는 영상과 행동 이외의 물건

a) 광운대학교 전자공학과(Department of Electronic Engineering, Kwangwoon University)

b) 한국전자기술연구원(Korea Electronics Technology Institute)

c) 광운대학교 스마트융합대학원(Graduate School of Smart Convergence, Kwangwoon University)

‡ Corresponding Author : 권순철(Soonchul Kwon)

E-mail: ksc0226@kw.ac.kr

Tel: +82-2-940-5112

ORCID:https://orcid.org/0000-0001-6595-6415

※ This work was supported by institute of information & communications Technology Planning & Evaluation(IITP) Grant funded by the Korea Government(MSIT) (No. 2021-0-00804, Development of online exam fraud prevention and class concentration improvement technology).

· Manuscript received November 8, 2021; Revised December 21, 2021; Accepted December 21, 2021.



그림 1. (a) 데이터 세트 획득 환경 예시 (b) 촬영 환경에서 획득되는 데이터의 예시  
 Fig. 1. (a) Example of dataset acquisition environment (b) Example of data acquired in shooting environment

을 추가하는 영상을 추가 촬영하였다. 추가 물품으로는 책, 스마트폰, 태블릿이 있다. 그림 1(a)는 데이터를 획득하는 촬영 환경이며 그림 1(b)는 환경 예시에서 획득되는 데이터의 예시이다. 1080x1080 해상도로 촬영하였으며 폐색을 피하도록 구성하였다. 그림 2는 데이터 세트의 획득 환경 세부 사항을 도식화한 그림이다. 사용자와 카메라의 거리는 가로 1m, 책상으로부터 높이 75cm로 구성하였으며, 카메라의 각도는 30° 아래를 향하도록 촬영하였다. 추가적으로 모든 촬영에는 펜과 시험지를 가장한 종이도 포함되어 시험 및 수업의 환경을 조성하였다. 데이터 세트 구성을 위해 촬영한 인원은 27명이며 관련 지식이 있는 인원으로 2달 동안 진행하였다. 각 행동에 대해 옷과 배경을 바꾸어 동일한 행동에 대해 4번 촬영을 진행하였다. 장기 시간 컨볼루션

(Long-term Temporal Convolution)을 사용한 네트워크에서는 100프레임과 같이 긴 인풋 클립(input clip)을 가질수록 높은 정확도를 가짐을 보여주었다.<sup>[9]</sup> 따라서 본 논문의 데이터 세트 구성의 경우, 모든 동영상은 5~6초의 30fps로 구성되었다. 데이터 수집 이후 수작업으로 명확한 행동으로 구성된 데이터를 선별하였다. 선별 기준은 사람의 머리와 상반신이 영상 내에 모두 보여야 하며, 행동이 정확하고, 추가 물품이 존재 시 잘려 보이지 않는다는 기준이다.

## 2. 온라인 공부 태도 데이터 세트 OSAD(Online Study Attitude Dataset)

표 1은 공부 태도 데이터 세트(OSAD)의 클래스와 클래스당 동영상 개수를 나타낸다. 수업 중에 가능한 행동 중 손동작과 얼굴에 중점을 두어 행동을 구성하였다. 공부하기(필기하기 932개, 화면 응시하기 910개) 1,842개, 스마트폰 조작 1,162개, 태블릿 조작 818개, 노트북 조작 890개, 하품 482개, 졸기 408개, 엎드려 자기 908개로 구성되어 있으며 전체 비디오 수는 6,510개이다. 각 클래스에 대한 예시 이미지는 그림 3에 있다. 공부 태도에 대한 데이터 세트의 경우 사람의 행동 외 화면 내의 물건이 포함될 수 있으므로 모든 클래스에 대해서 추가 물품(책, 스마트폰, 태블릿)을 무작위로 1개 추가하였다. 추가 물품에 대한 예시는 그림 4에 있다. 해당 클래스와 동일한 추가 물품은 사용하지 않았다. 예를 들어, 스마트폰 조작 클래스의 경우 추가 물품은 스마트폰을 제외하였다.

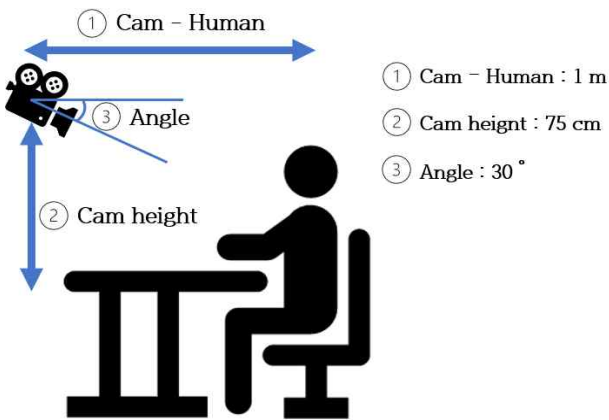


그림 2. 데이터 세트 획득 환경 세부 사항  
 Fig. 2. Dataset acquisition environment details

표 1. 온라인 공부 태도 데이터 세트의 각 클래스의 비디오 수  
 Table 1. Number of videos for each class in the Online Study Attitude Dataset(OSAD)

class	Sleep Down	Using Laptop	Yawn	Study	Using Smartphone	Using Tablet	Doze off
no.	908	890	408	1,842	1,162	818	482

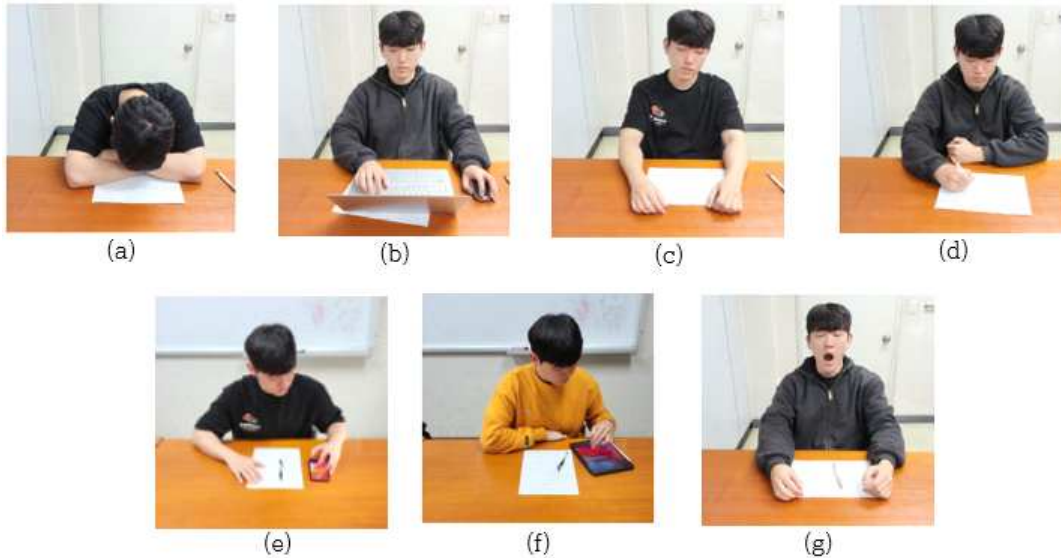


그림 3. OSAD의 각 클래스 예시 데이터 (a) 엎드려 자기 (b) 노트북 조작 (c) 졸기 (d) 공부하기(필기하기) (e) 스마트폰 조작 (f) 태블릿 조작 (g) 하품하기  
 Fig. 3. Example data of each class for OSAD (a) Sleep Down (b) Using Laptop (c) Doze off (d) Study (e) Using Smartphone (f) Using Tablet (g) Yawn



그림 4. OSAD의 추가 물품 예시 (a) 추가 물품이 태블릿인 졸기 클래스 (b) 추가 물품이 스마트폰인 엎드려 자기 클래스 (c) 추가 물품이 책인 하품하기 클래스  
 Fig. 4. Examples of additional items from OSAD (a) Doze off with tablet (b) Sleep down with smartphone (c) Yawn with book

### 3. 온라인 시험 태도 데이터 세트 OEAD(Online Exam Attitude Dataset)

표 2는 시험 태도 데이터 세트(OEAD)의 클래스와 클래스당 동영상 개수를 나타낸다. 일반적인 시험 동작과 부정행위로 간주할 수 있는 동작으로 행동을 구성하였다. 문제 풀

기 510개, 양팔 내리기 168개, 스마트폰 조작 1,162개, 태블릿 조작 818개, 노트북 조작 890개, 책 넘기기 494개로 전체 비디오 수는 4,042개이다. 각 클래스에 대한 예시 이미지는 그림 5에 있다. 시험 태도 행동 인식의 경우 추가 물품이 부정행위가 될 수 있으므로 문제 풀기 클래스에 대해 추가하지 않았다. 추가 물품에 대한 예시는 그림 6에 있다.

표 2. OEAD의 각 클래스의 비디오 수

Table 2. Number of videos for each class in the OEAD

class	Flipping Pages	Hand Down	Using Laptop	Solving Exam Question	Using Smartphone	Using Tablet
no.	494	168	890	510	1,162	818

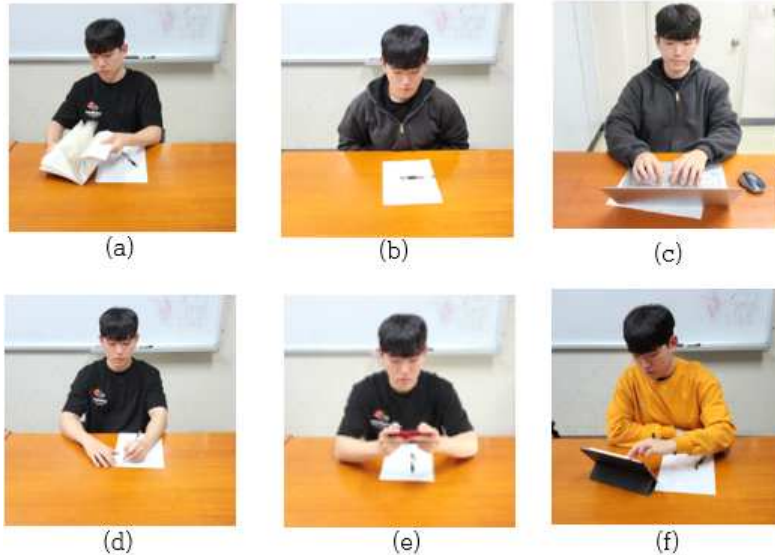


그림 5. OEAD의 각 클래스 예시 데이터 (a) 책 넘기기 (b) 양팔 내리기 (c) 노트북 조작 (d) 문제 풀기 (e) 스마트폰 조작 (f) 태블릿 조작  
 Fig. 5. Example data of each class for OEAD (a) Flipping Pages (b) Hand Down (c) Using Laptop (d) Solving Exam Question (e) Using Smartphone (f) Using Tablet

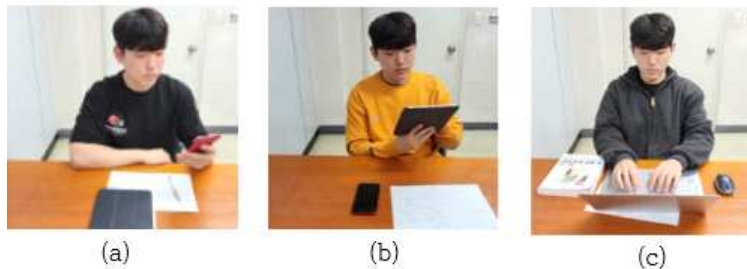


그림 6. OEAD의 추가 물품 예시 (a) 추가 물품이 태블릿인 스마트폰 조작 (b) 추가 물품이 스마트폰인 태블릿 조작 (c) 추가 물품이 책인 노트북 조작  
 Fig. 6. Examples of additional items from OEAD (a) Using Smartphone with tablet (b) Using Tablet with smartphone (c) Using Laptop with book

### III. 실험 및 평가

#### 1. UCF101과 HMDB51 데이터 세트

성능 평가를 위해 UCF101 데이터 세트와 HMDB51 데이터 세트를 사용하였다. UCF101 데이터 세트는 UCF50 데이터 세트의 확장으로 Youtube에서 수집된 101개의 클

래스를 가진 데이터 세트다.<sup>[10]</sup> 사람의 행동과 관련된 클래스로 구성되어 있으며, 동작 범주는 다음과 같다: 1) 사람과 물체의 상호작용 2) 사람의 행동 3) 사람과 사람의 상호작용 4) 악기 연주 5) 스포츠. HMDB51 데이터 세트는 영화 및 공개 데이터베이스(유튜브 및 구글)에서 수집된 51개의 클래스를 가진 데이터 세트다. 데이터 세트는 인터넷 소스와 영화의 비디오를 보고 단일 인간의 행동을 나타내는

모든 부분에 주석을 달았다. 하지만 카메라 모션이 존재할 수 있어 인접한 두 프레임에서 두드러진 특징을 감지하여 일치 후 배경 평면을 추정하는 표준 이미지 스티칭(stitching) 기술을 사용하여 클립 프레임을 정렬하였다. 동작 범주는 다음과 같다: 1) 일반적인 얼굴 행동 2) 물체 조작과 관련된 얼굴 동작 3) 일반적인 신체 동작 4) 물체와 상호작용을 위한 신체 동작 5) 사람과 상호작용을 위한 신체 동작. 표 3은 UCF101과 HMDB51 데이터 세트의 비디오 수이다.

표 3. UCF101과 HMDB51 데이터 세트의 비디오 수  
 Table 3. The number of videos for UCF101 and HMDB51

Dataset	UCF101	HMDB51
Videos	13,320	6,849

## 2. C3D Model

C3D 모델은 비디오 영상의 시공간적 정보를 모두 학습하기 위해 3D 컨볼루션을 사용하여 구성된 모델이다. 표 4는 C3D 모델의 구조이다. 8개의 3D 컨볼루션 레이어와 3개의 완전 연결 레이어(fully connected layer)를 사용하였고, 풀링 레이어(pooling layer)는 최대 풀링 레이어(max

표 4. C3D 모델 구조  
 Table 4. C3D model structure

conv1	conv2	conv3	conv4	conv5	fc6	fc7
[3x3x3, 64] max pool	[3x3x3, 128] max pool	[3x3x3, 256] x 2 max pool	[3x3x3, 512] x 2 max pool	[3x3x3, 512] x 2 max pool	4096	4096

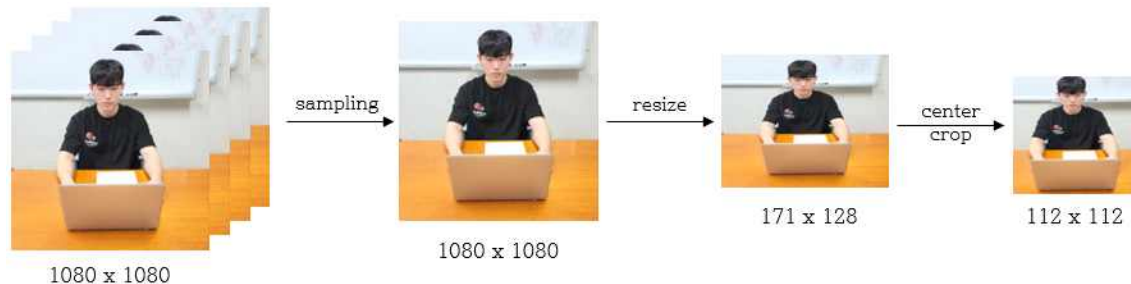


그림 7. C3D 모델의 데이터 가공 과정  
 Fig. 7. Data processing process of C3D model

pooling layer)를 사용하였다. 활성화 함수는 ReLU 함수를 사용하였다.

학습 파라미터로는 동등한 실험을 위해 모든 데이터 세트에 대해 동일하게 적용하였다. 사전 학습 가중치로는 Sport-1M 데이터 세트로 학습한 가중치를 사용하였다.<sup>[11]</sup> 학습률은 0.0001에서 시작하여 10 에폭(epoch)마다 0.1을 곱해주었다. 손실 함수는 CrossEntropy 손실 함수를 사용하였고, 최적화 함수는 SGD 함수를 사용하였다.<sup>[12]</sup> 학습 중 과적합을 방지하기 위해 드롭아웃(dropout)을 사용하였다.<sup>[13]</sup> 전체 에폭은 20회 진행하였다. 그림 7은 데이터 가공 과정을 도식화한 그림이다. 데이터 가공은 C3D 논문과 동일하게 적용하였다. 학습에서 사전 학습(pre-trained) 파라미터 사용을 위해 모든 입력 동영상을 4프레임 당 1프레임을 샘플링하여 171x128 이미지로 이미지 크기를 조정하였다. 그 후 112x112 이미지로 센터 크롭(center crop)을 하여 모델의 입력으로 사용하였다.

## 3. 실험 결과

실험은 우분투 18.04, NVIDIA GeForce RTX 2080 Ti 2개, 파이썬 3.8, 파이토치 1.7.1에서 진행되었다. 성능 평가



표 5. 각 데이터 세트에 대한 정확도  
Table 5. Accuracy for each dataset

Dataset	UCF101	HMDB51	OSAD	OEAD
train acc.	0.9923	0.9151	0.9901	0.9903
val. acc.	0.9698	0.6757	0.976	0.9706
test acc.	0.9744	0.6511	0.9862	0.9691

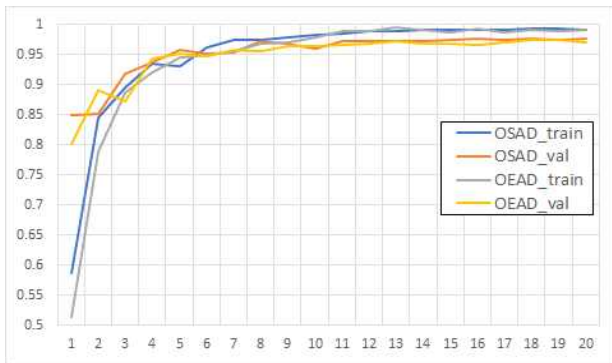


그림 8. OSAD와 OEAD의 에폭 당 정확도  
Fig. 8. Accuracy per epoch for OSAD and OEAD

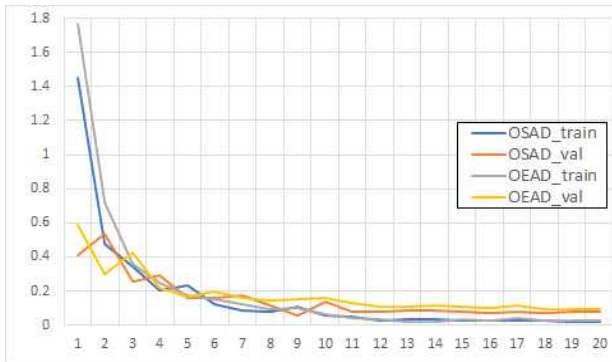


그림 9. OSAD와 OEAD의 에폭 당 로스  
Fig. 9. Loss per epoch for OSAD and OEAD

는 정확도를 사용하였다. 트레인(train), 밸리데이션(validation), test(테스트) 세트는 0.64:0.16:0.2의 비율로 무작위 샘플링을 하였다. 표 5는 UCF101, HMDB51, 공부 태도 및 시험 태도 데이터 세트에 대한 정확도 결과표이다. UCF101 데이터 세트의 경우 트레인 세트에서 0.9923, 밸리데이션 세트에서 0.9698, 테스트 세트에서 0.9744의 정확도를 보였다. HMDB51 데이터 세트의 경우 트레인 세트는 0.9151, 밸리데이션 세트는 0.6757, 테스트 세트는 0.6511의 정확도

를 보였다. 본 논문에서 제안된 OSAD의 경우 트레인 세트에서 0.9901의 정확도, 밸리데이션 세트에서 0.976, 테스트 세트에서 0.9862의 정확도를 달성하였고, OEAD의 경우 트레인 세트는 0.9903, 밸리데이션 세트는 0.9706, 테스트 세트는 0.9691의 정확도를 달성하였다. 그림 8과 9는 OSAD와 OEAD에 대해 에폭 당 로스(loss)와 정확도다.

#### IV. 결론

본 논문에서는 실시간 온라인으로 진행될 수 있는 공부 및 시험 태도 데이터 세트를 구성하고 C3D 모델을 통해 성능을 실험하였다. 각 목적에 맞게 클래스를 구성하였으며 추가 물품을 통해 영상 내에 물품을 보고 클래스를 구별하는 것이 아닌 사용 여부에 따라 구별하도록 학습할 수 있는 데이터 세트를 구성하였다. 기존 데이터 세트의 경우 배경의 특징이나 특정 클래스를 유추할 수 있는 행동을 가져 클래스의 구분이 명확한 데이터 세트와 달리 본 논문에서 제안된 배경의 특징 없이 손과 상반신의 미세한 동작의 차이로 구성된 클래스는 학습을 어려워지게 만들 수 있는 장점이 존재한다. 또한, 각 데이터 세트의 트레인, 밸리데이션, 테스트 세트의 정확도를 통해 기존의 우수한 데이터 세트와 비슷하거나 우수한 성능을 가짐을 확인하였으므로 제안하는 데이터 세트가 올바르게 구축되었다는 것을 확인할 수 있다. 하지만 제한적인 데이터 획득과 수작업의 데이터 선별로 인해 부족한 데이터 세트의 클래스와 동영상 수의 한계가 있다. 또한, 화각의 제한으로 카메라와 사용자의 거리가 1m로 제한하였다. 더 넓은 화각을 가진 카메라를 이용하여 사람의 머리와 팔의 행동 범위를 포착할 수 있다면 촬영 거리를 더 좁힐 수 있을 것이다. 향후 연구계획으로는 데이터의 질적 수준을 유지하면서 데이터 세트의 클래스 및 비디오 수의 양적 향상과 실시간이 가능한 성능이 더 우수한 모델 개발이 있다.

#### 참고 문헌 (References)

[1] School Remote Class Operation Status due to COVID-19, [https://kess.kedi.re.kr/post/6711307?itemCode=03&menuId=m\\_02\\_03\\_03](https://kess.kedi.re.kr/post/6711307?itemCode=03&menuId=m_02_03_03)

- (accessed Oct. 20, 2020)
- [2] Yan, S., Xiong, Y., Lin, D., "Spatial temporal graph convolutional networks for skeleton-based action recognition," American Association for Artificial Intelligence Conference, Artificial intelligence, New Orleans, USA, pp. 7444-7452, 2018
  - [3] Chen, Y., Zhang, Z., Yuan, C., Li, B., Deng, Y., Hu, W., "Channel-wise Topology Refinement Graph Convolution for Skeleton-Based Action Recognition," Proceedings of the IEEE/CVF International Conference, Computer Vision, Paris, France, pp. 13359-13368, 2021.
  - [4] Carreira, J., & Zisserman, A.. "Quo vadis, action recognition? a new model and the kinetics dataset," Proceeding of the IEEE Conference, Computer Vision and Pattern Recognition, Honolulu, USA, pp. 6299-6308, 2017. <https://doi.org/10.1109/cvpr.2017.502>
  - [5] Feichtenhofer, C., Pinz, A., & Zisserman, A., "Convolutional two-stream network fusion for video action recognition," Proceedings of the IEEE conference, Computer Vision and Pattern Recognition, Las Vegas, USA, pp. 1933-1941, 2016. <https://doi.org/10.1109/cvpr.2016.213>
  - [6] Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M., "Learning spatiotemporal features with 3d convolutional networks," Proceedings of the IEEE international conference, Computer Vision, Santiago, Chile, pp. 4489-4497, 2015. <https://doi.org/10.1109/iccv.2015.510>
  - [7] Soomro, K., Zamir, A. R., & Shah, M., "UCF101: A dataset of 101 human actions classes from videos in the wild," arXiv preprint arXiv:1212.0402. 2012.
  - [8] Kuehne, H., Jhuang, H., Garrote, E., Poggio, T., & Serre, T. (2011, November). "HMDB: a large video database for human motion recognition," Proceeding of the IEEE International conference, Computer Vision, Barcelona, Spain, pp. 2556-2563, 2011. <https://doi.org/10.1109/iccv.2011.6126543>
  - [9] Varol, G., Laptev, Cordelia Schmid "Long-term temporal convolutions for action recognition," Proceeding of the IEEE transactions on pattern analysis and machine intelligence, pp. 1510-1517, 2017. <https://doi.org/10.1109/tpami.2017.2712608>
  - [10] Reddy, K. K., & Shah, M., "Recognizing 50 human action categories of web videos," Machine Vision and Applications, Vol.24, No.5, pp.971-981, July 2013. <https://doi.org/10.1007/s00138-012-0450-4>
  - [11] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L., "Large-scale video classification with convolutional neural networks," Proceedings of the IEEE conference, Computer Vision and Pattern Recognition, Columbus, USA, pp. 1725-1732, 2014. <https://doi.org/10.1109/cvpr.2014.223>
  - [12] Bottou, L., Neural networks: Tricks of the trade, Springer, pp. 421-436, Berlin, Heidelberg. 2012.
  - [13] N. Srivastava, G. Hinton, A. Krizhevsky and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," J. of Machine Learning Research, Vol. 15, No. 1, pp. 1929-1958, June 2014.

---

## 저 자 소 개



김 준 식

- 2021년 2월 : 광운대학교 전자공학과 학사
- 2021년 3월 ~ 현재 : 광운대학교 전자공학과 석사과정
- ORCID : <https://orcid.org/0000-0003-3989-2336>
- 주관심분야 : 딥러닝, 영상처리



이 찬 휘

- 2020년 2월 : 광운대학교 전자공학과 학사
- 2020년 3월 ~ 현재 : 광운대학교 전자공학과 석사과정
- ORCID : <https://orcid.org/0000-0001-5422-8383>
- 주관심분야 : 딥러닝, 영상처리



---

저 자 소 개



**승 혁**

- 2012년 : 광운대학교 전자공학과 공학박사
- 2000년 ~ 현재 : 한국전자기술연구원 수석연구원
- ORCID : <https://orcid.org/0000-0003-0376-9467>
- 주관심분야 : 영상분석, 영상인식, 보안시스템



**권 순 철**

- 2012년 8월 : 광운대학교 정보디스플레이 공학박사
- 2012년 9월 ~ 2013년 2월 : 광운대학교 초빙교수
- 2013년 ~ 현재 : 광운대학교 스마트융합대학원 부교수
- ORCID : <https://orcid.org/0000-0001-6595-6415>
- 주관심분야 : 3D그래픽스, 3D영상처리