

## 한국어 BERT 모델을 활용한 청각 정보 기반 광고 영상 분류 방법론

박연지<sup>1</sup> · 양근제<sup>2</sup> · 손채봉<sup>3\*</sup>

<sup>1</sup>광운대학교 전자통신공학과 박사과정

<sup>2</sup>광운대학교 전자통신공학과 석사과정

<sup>3\*</sup>광운대학교 전자통신공학과 부교수

# Advertisement Video Classification Methodology Based on Auditory Information Using the Korean BERT Model

Yeon-Ji Park<sup>1</sup> · Geun-Je Yang<sup>2</sup> · Chae-Bong Sohn<sup>3\*</sup>

<sup>1</sup>Ph.D. Candidate, Department of Electronics and Communications Engineering, KwangWoon University, Seoul 01897, Korea

<sup>2</sup>Master's Course, Department of Electronics and Communications Engineering, KwangWoon University, Seoul 01897, Korea

<sup>3\*</sup>Professor, Department of Electronics and Communications Engineering, KwangWoon University, Seoul 01897, Korea

### [요약]

본 연구는 동영상 광고의 효과적인 분류 방법론을 제안하며, 한국어 기반의 자연어 처리 모델인 BERT(Bidirectional Encoder Representations from Transformers)를 중심으로 두고 있다. 연구 과정에서는 인터넷 크롤링을 통해 광고 영상 데이터를 수집하고, 이를 정답 라벨링 및 전처리 과정을 거쳐 분석에 활용할 수 있는 형태로 만들었다. 특히 전처리 과정에서 영상에서 청각적 정보, 즉 음성 데이터를 추출하여 텍스트로 변환하는 작업을 수행하였다. 연구 방법으로는 KoBERT(Korean BERT) 모델을 활용하여 실험을 진행하였으며, 광고 분류의 효율성과 정확성을 향상하게 시키는 방법론을 검증하였다. 본 연구를 통해 광고 콘텐츠의 더욱 정밀한 분류를 할 수 있게 되며, 이를 통해 광고 산업의 효율성이 크게 향상될 것으로 기대된다.

### [Abstract]

In this paper, we propose an effective classification methodology for video advertisements, focusing on the Korean language-based natural language processing model Bidirectional Encoder Representations from Transformers (BERT). During the research process, advertisement video data are collected through internet crawling and processed through answer labeling and preprocessing to make them analyzable. Specifically, in the preprocessing stage, auditory information, that is, voice data, is extracted from the videos and converted into text. The Korean BERT (KoBERT) model is used for the experiment as the research method, and the methodology is verified to enhance the efficiency and accuracy of advertisement classification. This study is expected to enable more precise classification of advertising content, significantly enhancing the efficiency of the advertising industry.

**색인어** : 광고 분류, 웹 크롤링, 자연어 처리, 청각적 정보 추출, BERT

**Keyword** : Advertisement Classification, Auditory Information Extraction, BERT, Natural Language Process, Web Crawling

<http://dx.doi.org/10.9728/dcs.2024.25.1.121>



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.

**Received** 08 November 2023; **Revised** 12 December 2023

**Accepted** 15 December 2023

**\*Corresponding Author; Chae-Bong Sohn**

**Tel:** 

**E-mail:** [cbsohn@kw.ac.kr](mailto:cbsohn@kw.ac.kr)

## I. 서론

### 1-1 연구배경

온라인 광고 시장, 특히 동영상 광고 분야는 근래 몇 년간 빠르게 발전하였다. 이는 Statista의 연구에서 확인할 수 있듯이, 디지털 비디오 광고 시장이 2023년에는 1,766억 달러의 지출을 기록했으며, 2027년까지 2,298억 달러에 이를 것으로 예상된다는 사실에서 드러난다[1]. 또한, Wyzowl의 2023년 비디오 마케팅 트렌드 보고서에 따르면, 기업들 중 91%가 비디오를 마케팅 도구로 사용하고 있으며, 마케터의 96%가 비디오를 마케팅 전략의 중요한 부분으로 여기고 있다[2]. 이러한 시장 환경에서, 광고 동영상의 효과적인 분류 및 관리는 매우 중요하다.

본 연구는 동영상 광고 분야의 이러한 성장과 더불어, 광고 동영상 분류의 새로운 방법론을 탐구한다. 특히, 기존 연구가 주로 시각적 요소에 초점을 맞춘 반면[3]-[5], 본 연구는 청각적 정보의 활용에 중점을 두고 있다. 이는 광고 분류에 있어 청각적 정보의 중요성을 강조하며, 기존의 시각 중심 접근법과의 차별화를 시도한다. 이러한 접근법은 데이터 세트의 크기가 커짐에 따라 증가하는 학습 리소스의 요구를 최소화하는 데에도 기여할 수 있다.

또한, 본 연구는 한국어 광고 동영상 데이터를 효과적으로 분류하기 위해 BERT 모델을 활용하는 방법론을 제시한다. 복잡한 문맥과 의미를 지닌 한국어의 구조적 특성을 고려하여, 세심한 데이터 전처리와 함께 BERT 모델을 적용함으로써, 광고 비디오의 의미를 보다 정확하게 분석하고자 한다. BERT 모델은 양방향 문맥 정보를 동시에 고려하는 구조로 인해 텍스트의 복잡한 의미를 더욱 효과적으로 이해할 수 있으며, 이러한 점은 텍스트 분류, 감정 분석, 관계 추출 등의 다양한 자연어 처리 작업에서 그 효과성이 검증되었다[6]-[9]. 더 나아가, BERT의 선행 학습된 모델들은 특정 작업에 맞게 미세 조정될 수 있으며, 이는 새로운 도메인이나 작업에 신속하게 적용될 수 있음을 의미한다[10]. 따라서 본 연구는 한국어 기반의 자연어 처리 모델인 KoBERT를 활용한 새로운 비디오 분류 기법을 제안한다.

본 연구는 이러한 BERT 모델의 강점을 동영상 광고 분류에 활용하여, 광고 비디오에서 추출된 청각적 정보를 통해 잠재적 의미를 포착하는 분류 알고리즘을 개발하고자 한다. 이 알고리즘의 개발은 광고 산업의 효율성을 크게 향상시킬 뿐만 아니라, 사용자에게 더욱 맞춤형 광고 콘텐츠를 제공하는 기반을 마련할 것이다.

본 연구의 주요 기여점은 다음과 같다.

- (1) KoBERT 모델을 활용한 새로운 동영상 광고 분류 방법론 제시.
- (2) 딥러닝 모델과 자연어 처리를 결합한 광고 동영상 분류

알고리즘 개발.

- (3) 광고 산업에서의 광고 비디오 분류의 효율성 및 정확도 향상에 기여.

마지막으로, 본 연구는 현존하는 방법론의 한계를 극복하고, 동영상 광고 분류의 미해결 문제를 해결하는 데 기여할 수 있는 혁신적인 접근 방법을 탐구한다. 논문의 후반부에서는 연구 방법론을 상세히 설명하며, 실험 과정과 결과를 통해 제안된 방법의 유효성을 검증한다.

### 1-2 논문의 구성

본 논문의 구성은 1장 서론으로 시작하여, 2장에서는 기존 연구 동향 및 본 연구와의 차별 점에 대해 설명한다. 3장에서는 데이터 수집 및 전처리 방법, 모델 설계 및 학습 접근법 설명하고, 4장에서는 실험 설정, 데이터 세트 구조 및 모델 성능 평가 결과를 소개한다. 마지막 장에서는 연구 결과의 요약하고 한계점 및 향후 연구 방향을 제안한다.

## II. 관련 연구

본 연구는 기존 연구를 기반으로 광고 동영상 분류의 새로운 방법론을 탐구한다. 광고 동영상에서 청각적 정보를 추출하여 활용하며, 이는 텍스트 형태로 변환하여 자연어 처리 모델 중 하나인 BERT 기반 모델을 통해 동영상 분류를 수행한다. 따라서 본 장에서는 동영상 분석과 BERT 기반 모델을 활용한 연구 동향을 살펴보고 본 연구와의 차이점을 검토한다.

### 2-1 동영상 분석 연구 동향

동영상 분류는 다양한 연구 분야에서 중요성을 가지며, 국내외에서 이에 대한 다양한 연구가 활발히 진행되고 있다. 특히 광고 분야에서는 비디오 분석의 중요성이 부각되며 주목할 만한 발전을 이루고 있다. 이러한 연구들은 주로 시각적 데이터에 초점을 맞추고 있다.

Deldjoo 등[4]은 비디오의 시각적 특징을 이용하여 추천 시스템을 개선하는 연구를 수행하였다. Luo 등[5]은 딥러닝을 기반으로 한 실시간 객체 탐지 모델인 Yes-Net을 개발하고, 이를 광고 데이터 분류에 적용하여 사용자에게 효과적으로 전달하는 비디오 광고 프레임워크를 구축하였다. 또한, Schwenzow 등[3]은 동영상 콘텐츠에서 시각적 및 부수적 특징을 추출하는 분석하는 도구를 제안하였다. 이 도구에서 시각적 차이를 추출하는 경우, 딥러닝 기반 Siamese 네트워크 접근 방식[11]을 통해 프레임 간 색상, 모양, 개체 등 다양한 정보를 추출한다. 또한, PySceneDetect[12],[13]을 통해 장면 컷 특징을 탐지하는 연구를 수행하여 0.98의 정밀도

와 0.82의 재현율을 보여주었다.

이와 대조적으로, 본 연구는 청각적 정보에 초점을 맞추어 동영상 분류를 수행하며, 이는 기존 연구와의 주요한 차별점이다. Dhakal[14]은 정치 캠페인 광고를 효과적으로 분류하기 위해 딥러닝 기술을 활용한 연구를 수행하였다. 이 연구에서는 시각적 정보로 이미지 프레임 내의 광학 문자 인식(OCR; Optical Character Recognition) 결과를, 청각적 정보로는 오디오에서 추출한 음성을 텍스트로 변환한 결과를 사용했다. 광고 분류는 이 두 종류의 데이터의 정확도를 계산하고, 그 평균을 이용해 최종적으로 판단한다. 이는 청각 정보가 광고 분류에 있어 중요한 역할을 한다는 것을 보여주었다. Althnian 등[15]에서 언급된 바와 같이, 데이터 세트의 크기가 커질수록 학습에 필요한 리소스가 증가하는 문제가 있다.

이와 같은 연구들을 고려할 때, 본 연구에서는 청각적 정보만을 사용한 동영상 광고 분류 방법을 탐구하여 데이터 세트 양의 증가에 따른 리소스 요구 증가 문제에 대응하고, 그 방안에 대한 효율성을 검증하고자 한다.

## 2-2 BERT 기반 모델을 활용한 연구 동향

본 연구에서 사용되는 데이터 세트는 한국어 광고 동영상으로, 한국어의 구조적 특성을 고려한 데이터 전처리가 중요하다. 한국어는 주어-목적어-동사의 기본 어순을 가지지만, 문맥에 따라 유연하게 변할 수 있으며, 어근과 다양한 접사의 결합으로 새로운 의미나 문법적 기능이 형성된다. 이러한 특성으로 인해 문장의 의미를 파악하고, 문맥적 분석을 수행하기 위해 세심한 데이터 전처리가 요구된다.

최근 몇 년 동안, 데이터 전처리는 기계학습과 딥러닝 모델의 성능을 향상하는 핵심 요소로 간주되었다[16]. 딥러닝 모델이 대규모 언어 모델을 더 효과적으로 학습하기 위해 깨끗하고 품질 높은 데이터가 필요하기 때문이다[17]. 이러한 맥락에서 본 연구는 한국어 데이터 전처리의 중요성을 인식하고, 한국어에 특화된 전처리 방법을 적용하여 광고 비디오의 의미를 더 정확하게 분석하고자 한다. 최근의 연구에서도 한국어 데이터의 효율적인 전처리 방법에 대해 강조하며, 한국어 자연어 처리의 성능 향상을 위한 중요한 전제로 제시하고 있다. 예를 들어 Liu 등[18]은 다양한 기술적 접근 방법을 탐구하여 한국어 데이터의 효율적인 전처리 방법을 강조한다. 본 연구에서도 한국어에 특화된 전처리 방법을 적용하여 데이터의 정확한 분석을 목표로 한다.

본 연구에서는 BERT 모델을 활용하여 한국어 데이터를 처리하고자 한다. BERT는 트랜스포머 아키텍처의 인코더 구조를 기반으로 하는 딥러닝 기반의 자연어 처리 모델로, 언어 이해 및 처리에서 뛰어난 성능을 보여주었으며, 자연어 처리 분야의 패러다임 전환을 주도하였다[17]. 특히 한국어와 같이 복잡한 언어 구조에서 BERT의 문맥에 기반한 언어 이해 능력은 매우 유용하다. 한국어 의료 자연어 처리 분야에서 BERT의 적용 사례를 통해 이 모델은 한국어 데이터의 정확

한 이해와 처리에 효과적임을 보여주었다[19]. 또한, 다양한 연구들은 BERT 기반 모델을 통해 한국어 텍스트의 자연어 처리 성능을 향상하게 시키는 방안을 제시하였다[20],[21].

이어서, 여러 연구에서 BERT를 기반으로 한 다양한 변형 및 확장이 소개되었다. Sun 등[22]은 BERT의 성능을 향상시키기 위한 훈련 전략을 제안했다. 이러한 연구의 결과로, RoBERTa(Robustly optimized BERT approach) 모델이 개발되었으며 이 모델은 BERT의 원래 구조를 유지하면서도 학습 데이터와 학습 방법을 최적화하여 성능을 향상했다. Lan 등[23]은 파라미터의 수를 줄이면서도 BERT의 성능을 유지하는 ALBERT(A Lite BERT) 모델을 제안했다. 이 모델은 기존의 BERT 모델보다 훨씬 더 효율적인 학습이 가능하며, 메모리 사용량도 줄일 수 있었다. 그 외에도, Lewis 등[24]은 BART(Bidirectional and Auto-Regressive Transformers) 모델을 제안하여 BERT를 향상하게 시키기 위한 노력을 계속해 왔으며, 이러한 연구들이 BERT의 원리와 강점을 극대화하는 방향으로 진행돼왔다.

이러한 배경을 바탕으로, 본 연구는 BERT 모델을 기반한 한국어 데이터 처리를 통해 광고 동영상 데이터의 분석 효율성과 정확도를 향상시키는 것을 목표로 한다.

## III. 본 론

본 장에서는 광고 비디오 데이터를 수집 및 전 처리하여 자연어 처리 모델에 적용하는 과정을 설명한다. 그림 1에는 이를 수행하기 위한 제안 모델의 기본 파이프라인이 도식화되어 있다. 논문의 후속 절에서는 그림 1에 제시된 각 단계에 대한 자세한 방법론을 서술한다.

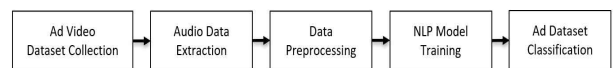


그림 1. 제안하는 모델의 파이프라인

Fig. 1. A pipeline of our proposed model

### 3-1 데이터 수집

본 연구에서는 광고 데이터 세트를 구축하기 위해 한국광고총연합회 광고정보센터 크리에이티브 사이트[25]에서 광고 영상 데이터를 수집한다. 데이터 수집을 위한 흐름도는 그림 2와 같고, 이 데이터 수집에 사용된 도구 및 기술에 대해 아래에서 상세히 설명한다.

데이터 수집은 파이썬 언어를 이용하여 수행되며, 웹 크롤링을 위한 셀레니움(Selenium) 라이브러리와 웹 드라이버인 크롬 드라이버를 활용한다. 크롬 드라이버를 통해 광고 영상 데이터가 그리드 형태로 배열된 사이트에 접속하였으며, 이를 그림 3에서 볼 수 있다.

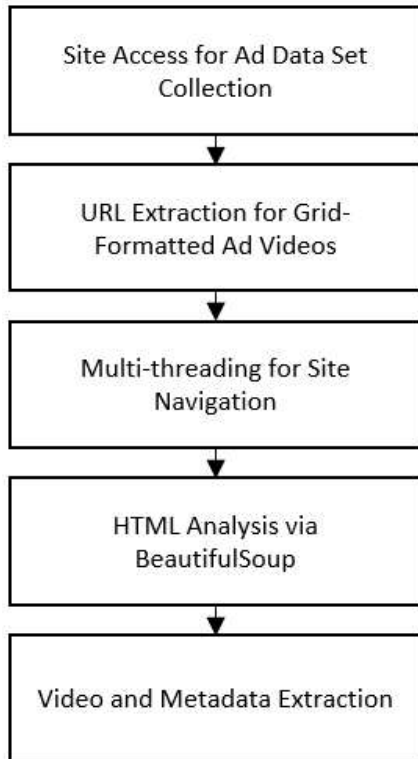


그림 2. 데이터 수집의 흐름도  
 Fig. 2. A flowchart of data collection



\*This image is a screenshot from a Korean website, therefore, it inevitably contains Korean text  
 그림 3. 광고 영상 데이터 세트 수집을 위한 웹 사이트  
 Fig. 3. A website for collecting advertisement video dataset

각 데이터는 XPath 정보를 활용하여 웹 페이지에서 추출된다. XPath는 XML 문서 내에서 특정 요소나 속성을 탐색하고 선택하는 데 사용되는 언어로, HTML 문서에서도 유사하게 작동한다. 이를 통해 웹 페이지의 특정 부분을 정확히 지정하고, 필요한 데이터가 담겨 있는 URL로 접근할 수 있다. 해당 URL에 접근하면 각종 메타 데이터(영상 분류, 등록 날짜 등)와 함께 광고 영상을 확인할 수 있는 페이지인 그림 4로 이동하게 된다. 페이지의 HTML 문서 분석은 BeautifulSoup 라이브러리를 활용하여 이루어진다.

데이터 수집의 효율을 높이기 위해, multi-threading 기법을 적용하여 병렬 처리를 가능하게 한다. 이 방법으로 인해 데이터 수집 시간을 대폭 줄인다. 수집된 데이터 세트는 이후 병합 과정을 거쳐 하나의 데이터프레임으로 구성되었으며, 결측값 제거 등의 전처리 과정을 거치게 된다.



\*This image is a screenshot from a Korean website, therefore, it inevitably contains Korean text  
 그림 4. 하나의 영상 데이터 클릭 화면의 일부분  
 Fig. 4. A partial view of a single video data click screen

표 1. 수집된 카테고리 데이터 세트의 샘플  
 Table 1. A sample of the collected category dataset

Video-name	Tag	Time
1	Government offices and organizations/central and local government offices/local government	0.386265
2	Computer and information communication/service/internet service	0.423995
3	Computer and information communication/computer games/mobile games	0.370578
4	Computer and information communication/communication information service/internet service	0.823157
5	Distribution/small retail distribution/small retail distribution others	0.669001

데이터 세트는 영상 데이터와 카테고리 정보로 구성되며, 영상 데이터는 청각적 정보를 추출하기 위해 사용된다. 카테고리 정보는 모델 학습 시 정답 레이블로 활용된다. 표 1은 최종적으로 수집된 카테고리 정보 데이터 세트의 샘플을 보여준다. 여기서 'video\_name'은 영상을 다운로드할 때 사용된 파일명, 'tag'는 영상의 카테고리 정보를, 'time'은 데이터를 추출하는데 소요된 시간을 나타낸다.

이러한 방법을 통해 광고 영상 데이터 세트를 구축하였으며, 다음 섹션에서는 이를 활용하여 청각적 정보를 분석하는 방법론을 설명한다.

한국표준산업분류 제11차 개정 분류초안 분류체계 및 변경 내역

제10차 분류		제11차 분류		구분	변경 내역
코드	코드 명칭	코드	코드 명칭		
A	농업, 임업 및 어업(01-03)	A	농업, 임업 및 어업(01-03)		
01	농업	01	농업		
011	작물 재배업	011	작물 재배업		
0111	곡물 및 기타 식량작물 재배업	0111	곡물 및 기타 식량작물 재배업		
01110	곡물 및 기타 식량작물 재배업	01110	곡물 및 기타 식량작물 재배업		
0112	채소, 화훼작물 및 증묘 재배업	0112	채소, 화훼작물 및 증묘 재배업		
01121	채소작물 재배업	01121	채소작물 재배업		
01122	화훼작물 재배업	01122	화훼작물 재배업		
01123	종자 및 묘목 생산업	01123	종자 및 묘목 생산업		
0113	과실, 음료용 및 향신료 작물 재배업	0113	과실, 음료용 및 향신료 작물 재배업		
01131	과실작물 재배업	01131	과실작물 재배업		
01132	음료용 및 향신료 작물 재배업	01132	음료용 및 향신료 작물 재배업		
0114	기타 작물 재배업	0114	기타 작물 재배업		
01140	기타 작물 재배업	01140	기타 작물 재배업		
0115	사실작물 재배업	0115	사실작물 재배업		
01151	콩나물 재배업	01159	기타 사실작물 재배업	통합	<ul style="list-style-type: none"> <li>01151(콩나물 재배업)을 01159(기타 사실작물)에 통합</li> <li>통계 시계열자료 분석결과 반영</li> <li>(0115C) 작물재배업을 노지 및 시설로 구분하지 않으며, KSIC 독자 세세분류로 운영</li> </ul>

한국표준산업분류 개정 분류체계(제11차 기준)	
대분류	
항목명	
A 농업, 임업 및 어업	
B 광업	
C 제조업	
D 전기, 가스, 증기 및 공기조절 공급업	
E 수도, 하수 및 폐기물 처리, 원료재생업	
F 건설업	
G 도매 및 소매업	
H 운수 및 창고업	
I 숙박 및 음식점업	
J 정보통신업	
K 금융 및 보험업	
L 부동산업	
M 전문, 과학 및 기술 서비스업	
N 사업시설 관리, 사업지원 및 임대 서비스업	
O 공공행정, 국방 및 사회보장 행정	
P 교육서비스업	
Q 보건업 및 사회복지 서비스업	
R 예술, 스포츠 및 여가관련 서비스업	
S 협회 및 단체, 수리 및 기타 개인 서비스업	
T 가구 내 고용활동 및 달리 분류되지 않은 자가소비 생산활동	
U 국제 및 외국기관	

\*The image is part of a task involving the matching of revised Korean Standard Industrial Classification codes, and therefore, it may inevitably contain Korean labels.

그림 5. 한국표준산업분류(KSIC) 11차 개정코드 기반 라벨링 기법

Fig. 5. A labeling strategy based on the 11th revision of the Korea Standard Industrial Classification (KSIC)

3-2 데이터 전처리

1) 데이터 라벨링

본 절에서는 광고 영상 분류를 위한 정답 라벨링 방법론에 대하여 상세히 설명한다. 본 연구에서는 한국표준산업분류(KSIC) 11차 개정 코드를 기반으로 광고 영상 데이터를 범주화하는 작업을 수행하였다. 한국표준산업분류는 총 21개의 대분류 항목으로 이루어져 있으며, 각 항목을 'A'부터 'U'까지 알파벳으로 라벨링한 것을 그림 5에서 확인할 수 있다.

수집한 영상 정보 데이터 세트의 태그는 대분류, 중분류, 소분류로 구분되며, 이러한 체계를 바탕으로 한국표준산업분류 대분류 항목과 매칭 작업을 수행한다. 데이터 세트는 22개의 대분류와 이에 속하는 165개의 소분류 카테고리 구성되어 있다. 본 연구에서는 수작업을 통해 대분류의 소분류와 KSIC의 분류 항목을 일일이 대조, 검토하여 최적의 매칭을 도출한다. 이 과정에서, 데이터 세트 중에서 표준 분류 체계에 적합한 데이터가 없거나 부족한 레이블은 연구에서 제외한다. 따라서, 최종적으로 A(농업, 임업 및 어업), B(광업), D(전기, 가스, 증기 및 공기조절 공급업), E(수도, 하수 및 폐기물 처리, 원료재생업), O(공공행정, 국방 및 사회보장 행정), T(가구 내 고용 활동, 자가소비 생산활동), U(국제 및 외국기관) 클래스는 분류 작업에서 제외된다. 표 2에는 이러한 방법론을 통해 정리된 데이터 세트의 수집 통계를 나타낸다.

이러한 방법으로 얻은 라벨링 체계는 광고 영상 데이터의 분류 및 분석에 필수적인 기초를 제공하며, 광고 산업의 다양한 항목을 정밀하게 구분하고 분석할 수 있는 기반을 마련한다. 이를 통해 광고 영상 데이터의 체계적인 분류 및 분석을 할 수 있게 한다.

표 2. 데이터셋 수집 통계

Table 2. Dataset collection statistics

Class	Number of videos
C	30,386
F	894
G	10,467
H	939
I	1,982
J	8,732
K	6,335
L	1,011
M	197
N	562
P	2,637
Q	4,638
R	8,689
S	11,812
Sum	89,281

2) 데이터 추출 및 변환

본 연구는 광고 비디오에서 청각적 정보만을 추출하여 분류하는 새로운 접근 방식을 채택한다. 이러한 접근 방식을 채택한 이유로는 다음과 같다. 첫째, 음성 데이터만을 활용하는 접근법은 처리 속도 및 효율성 면에서 영상 데이터를 처리하는 방식보다 더 뛰어나다. 대규모 데이터 세트에 특히 유용하며, 빠른 분석과 효과적인 처리가 가능하다. 둘째, 영상의 해상도가 낮거나 영상이 깨진 경우와 같은 특수한 상황에서도 음성 데이터를 통해 유의미한 분석을 진행할 수 있다. 셋째, 청각적 정보를 활용한 자연어 처리 기술의 발전은 복잡한 영

상 분석 기술과는 다른, 새로운 분류 방법론을 가능하게 한다. 이러한 접근은 특히 언어 기반의 내용이 중요한 광고에서 효과적이며, 광고의 음성 메시지만으로도 충분한 정보를 제공할 수 있다. 이러한 방법론은 광고 동영상 분류의 새로운 지평을 열어주며, 기존 영상 기반의 분류 방법론과는 다른 차별화된 접근을 제공한다. 영상 데이터에서 음성 데이터를 추출하고 텍스트로 변환하는 과정의 흐름도는 그림 6과 같고, 본 절에서 상세히 기술한다.

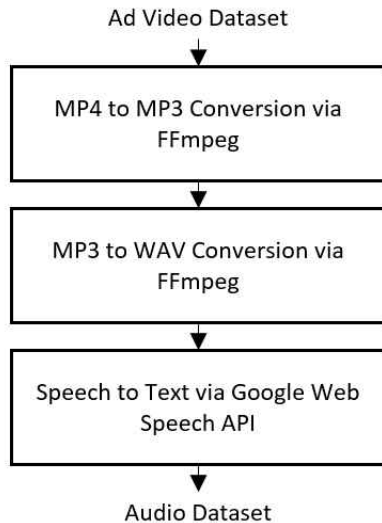


그림 6. 데이터 추출 및 변환의 흐름도  
 Fig. 6. A flowchart of data extraction and conversion

수집한 영상에서 청각적 정보를 추출하고, 청각적 정보를 학습용 데이터로 활용하기 위해 텍스트 데이터로 변환하는 과정은 오픈소스 기반의 강력한 영상 인코딩 소프트웨어인 FFmpeg을 활용하여 이루어진다. FFmpeg은 디지털 오디오 및 비디오 레코딩, 변환, 스트리밍을 가능하게 하는 라이브러리 및 툴 집합으로 구성된 오픈 소스 프로젝트이다. 이 프로젝트는 개발자들에게 비디오 및 오디오 관련 작업을 수행할 수 있는 도구를 제공하며, 거의 모든 미디어 파일 형식과 호환되는 변환 옵션을 제공한다. 본 연구에서는 FFmpeg의 툴을 활용하여 광고 영상 데이터의 형식인 MP4를 MP3로 변환한다. 이어서 MP3 파일을 WAV 형식으로 변환함으로써 음성 데이터를 추출한다. 이렇게 추출된 음성 데이터는 후속 작업에서 텍스트 데이터로 변환되어, 광고 영상 분석의 핵심 입력 데이터로 활용된다. 이러한 데이터 전처리 과정은 학습용 데이터를 생성하는 데 근본적인 단계로 작용하며, 더 나아가 본 연구에서 제안하는 광고 분류 방법론의 효율성과 정확도를 높이는 데 기여한다.

추출한 음성 데이터를 자연어 처리 모델의 학습 데이터로 활용하기 위한 중요한 다음 단계는 텍스트 변환 과정이다. 우리는 이 음성 데이터를 텍스트로 변환하기 위해 파이썬의 SpeechRecognition 패키지를 사용한다. 이 패키지는 다양

한 API를 통한 음성 인식을 용이하게 지원한다. 그러나 이 API들 중 일부는 한국어를 포함한 비영어권 언어 처리에 제한이 있다. 더욱이, 일부 API는 일정 사용량을 초과할 경우 유료화되어, 본 연구에서 구축한 데이터 세트를 변환하는 데 드는 비용에 한계가 있다. 이에 따라, 높은 정확도와 다양한 언어를 지원하는 Google Web Speech API를 선택하였다. 한국어를 포함한 광범위한 언어 데이터 처리가 가능한 이 API는 선택에 있어 결정적인 역할을 하며, 데이터 전처리 과정을 효율적으로 수행할 수 있는 핵심 메커니즘이다.

표 3. 데이터 추출 및 변환의 의사 코드  
 Table 3. Pseudocode for data extraction and conversion

```

Set default_path, class_name, data_dir_path
Create directories for mp3 and wav files
Open data list file, write header 'video_name,text,time'

For each file in data directory:
  # Extract file name and set paths
  Extract file_name
  Set ori_file_path, mp3_file_path, wav_file_path

  # Convert file: mp4 -> mp3 -> wav
  Convert mp4 to mp3 using ffmpeg
  Convert mp3 to wav using ffmpeg

  # Extract text through speech recognition
  Start timer
  Initialize speech recognizer
  Try:
    Recognize speech from wav_file_path in Korean
    Write extracted text to data list file
  Catch and handle any errors
  Record execution time

# Close data list file
Close data list file
    
```

표 4. 변환된 텍스트 데이터 세트의 샘플  
 Table 4. A sample of the converted text dataset

Video-name	Text	Time	Class
40522	뚜레쥬르는 다릅니다 자연의 순수한 재료 건강한 빵을 위한 정성에 시간이 모든 것이 지켜져야 뚜레쥬르 빵이 됩니다 뚜레쥬르	2.33499932 2891236	I
65982	2010년 밴쿠버 10억엔 그들의 경기를 지켜봤다 그 중 실제 경기를 본 사람은 몇이나 될 것인가 언제나 현대카드로부터 현대카드의 놀라운 플레이는 계속됩니다 현대카드	4.22204208 37402335	K
34198	틀니를 사용하는 많은 분들이 잘못된 틀니 관리로 고통 받고 있습니다 이제 폴리덴트와 함께하세요 틀니 관리 날을 맞아 무료샘플을 보내드립니다 지금 전화하세요	2.23200011 25335693	Q

\*This table represents a sample of a Korean text dataset and, as such, inevitably contains Korean text.

표 3은 영상 데이터에서 음성 데이터를 추출하고 텍스트로 변환하는 과정을 간단하게 나타낸 의사 코드이다. 이러한 데이터 전처리 단계는 광고 분류 방법론의 효율성과 정확도를 향상하게 시키는 데 기본적으로 작용하며, 본 연구에서 제시하는 광고 분류 알고리즘의 성능 향상에 중요한 역할을 담당한다.

표 4는 위의 과정을 거쳐 텍스트로 변환되어 학습에 사용된 데이터 세트의 샘플이다. 여기서 'video\_name'은 영상을 다운로드할 때 사용된 파일명, 'text'는 영상에서 변환된 텍스트를, 'time'은 텍스트 데이터를 추출하는데 소요된 시간을, 'class'는 정답 레이블인 분류 클래스를 나타낸다. 예를 들어, 'video\_name'이 40522에 해당하는 광고 영상 데이터에서 추출된 텍스트는 '뚜레쥬르는 다릅니다 자연의 순수한 재료 건강한 빵을 위한 정성에 시간이 모든 것이 지켜져야 뚜레쥬르 빵이 됩니다 뚜레쥬르'와 같다. 해당 영상은 '서비스>음식 및숙박'의 카테고리 정보를 가지고 있어 I(숙박 및 음식점) 클래스로 분류된 것을 확인할 수 있다.

### 3-3 학습 모델

본 연구에서 제안하는 모델은 한국어 기반의 BERT 모델인 KoBERT를 사용하여 광고 동영상의 세부적인 클래스를 분류하는 것이다. BERT는 Google에서 개발된 사전 훈련된 언어 이해 모델로, 텍스트 데이터의 문맥을 파악하는 데 뛰어난 성능을 보여주었다. KoBERT는 이러한 BERT의 구조를 기반으로 하면서 한국어 텍스트 데이터로 사전 훈련된 모델로, 한국어 텍스트 데이터의 문맥을 파악하는 데 특화되어 있다.

제안된 모델의 구조는 크게 특성 추출기(feature extractor)와 분류기(classifier)로 구성된다. 특성 추출기는 각 동영상에서 청각적 정보를 추출하는 역할을 한다. 이 정보는 동영상의 오디오 트랙에서 추출된다. 추출된 특성은 다음 단계에서 입력으로 사용된다. 분류기는 추출된 특성을 입력으로 받아, 해당 동영상의 광고 클래스를 예측하는 역할을 한다. 본 연구에서는 분류기로 KoBERT를 사용한다. KoBERT는 입력된 특성을 분석하여, 각 동영상을 해당하는 광고 클래스로 분류한다. 이 과정에서 최적화 알고리즘으로 AdamW를 선택한 것은 그 장점 때문이다. AdamW는 표준 Adam 최적화 방법에 비해 가중치 감소를 통합하는 방식이 개선되어 있어, 과적합을 방지하고 모델의 일반화 능력을 향상시킨다. 이는 학습 과정의 안정성을 증진시키고, 장기적인 성능 유지에 기여하여, 복잡한 한국어 광고 분류 작업에 있어 KoBERT의 성능을 극대화한다. 본 연구에서 사용된 모델의 주요 파라미터는 아래 표 5와 같다.

표 5. 하이퍼 파라미터

Table 5. Hyper parameters

Parameter	Value
Epochs	100
Classes	14
Optimizer	AdamW
Max Length	64
Batch Size	64
Learning Rate	5e-5

## IV. 실험 및 결과

### 4-1 실험 환경

본 연구의 실험 환경은 표 6과 같다. 파이썬을 통하여 모델을 실험하였으며, 라이브러리는 Pytorch를 사용하였다.

표 6. 실험 환경

Table 6. Experiment environment

Items	Values
Compute Engine	AMD 5600x RTX 3060 16GB Memory
Language	Python 3.9.18
Deep learning library	Pytorch 1.10.1 Sentencepiece 0.1.96 Transformers 4.8.1

### 4-2 데이터 세트

본 연구는 한국어 기반 BERT 모델인 KoBERT를 활용하여 동영상 광고 분류의 효과를 실험적으로 검증한다. 이를 위해, 주로 국내 광고에 초점을 맞춘 광고 비디오 데이터 세트를 구축하고, 이 동영상들로부터 청각적 정보를 추출한다. 추출된 정보를 바탕으로 태그를 생성하고, 이들을 KoBERT 모델에 입력하여 광고 동영상을 분류하는 방식으로 진행된다. 본 연구에서 제안하는 분류 방법은 한국어 텍스트를 중심으로 수행되며, 사용 가능한 한국어 광고 영상 데이터 세트의 규모가 한정적이다. 따라서, 연구의 목적에 맞게 특화된 데이터 세트를 구축하여 사용하는 것이 필수적이다.

표 7. 데이터셋 태그 및 라벨링 개요

Table 7. Overview of dataset tags and labeling

Class	Name	Label
C	Manufacturing	0
F	Construction industry	1
G	Wholesale and retail	2
H	Transportation industry	3
I	Accommodation and restaurant industry	4
J	Information service	5
K	Finance and insurance	6
L	Real estate industry	7
M	Professional, scientific and technological services	8
N	Business facility management, business support and rental services	9
P	Education service industry	10
Q	Health and social services	11
R	Arts, sports and leisure-related services	12
S	Associations and organizations, repair and other personal services	13

표 7은 광고 영상의 14개 클래스와 각각의 라벨을 표시한 것이다. 광고 영상의 분류 체계는 한국표준산업분류 11차 개정안에 따라 21개 클래스 중 분류 체계와 맞는 광고가 없는 클래스 6종과 데이터 세트가 적은 1종을 제외한 14개 클래스를 사용하였다. 광고 분류의 학습 및 테스트를 위해, 표 2에 나타낸 바와 같이 수집된 데이터 세트에서 각 클래스 별로 70%는 학습용 데이터 세트로, 나머지 30%는 테스트용 데이터 세트로 활용하였다.

### 4-3 실험 결과

모델의 성능 검증을 위해 학습 과정 중, 테스트 데이터 세트에 대한 정확도를 측정하였으며, 우리의 모델은 이 검증에서 91.1%라는 높은 정확도를 달성하였다. 이 결과는 표 8에서 다른 모델과 비교하였을 때 경쟁력 있는 성능을 보여준다. 특히, 기존에 이미지 처리 기법을 사용한 영상 분류 방식의 정확도와 비교해 볼 때, 우리의 자연어 처리 기법을 통한 광고 영상 분류 방법이 유사한 수준의 성능을 보여줄 수 있는 것을 확인하였다.

표 8. 영상 분류 방법론 비교

Table 8. Comparison of video classification method

Methods	Test accuracy
Our Proposed	91.1%
LRCN [26]	82.9%
DT+MVSV [27]	83.5%
LSTM-Composite [28]	84.3%
FSTCN [29]	88.1%
C3D [30]	85.2%
iDT+HSV [31]	87.9%
Two-Stream [32]	88.0%
RNN-FV [33]	88.0%
LSTM [34]	88.6%
MultiSource CNN [35]	89.1%
Image-Based [36]	89.6%
TDD [37]	90.3%
Multilayer and Multimodal Fusion[38]	91.6%
Transformation CNN [39]	92.4%

그러나, 음성만을 사용한 광고 분류 방법은 건설 산업, 도매 및 소매, 숙박 및 음식점 산업, 부동산 산업, 그리고 예술, 스포츠 및 여가 관련 서비스와 같은 특정 산업 분야에서의 광고를 정확하게 분류하는 데 어려움을 겪었다. 이러한 카테고리의 광고들은 시각적 요소가 중요한 역할을 하므로, 음성 정보만으로는 제품의 특징이나 전체적인 메시지를 충분히 전달

하기 어렵다. 이는 분류 정확도에 영향을 미칠 가능성이 있음을 시사하며, 특정 산업 분야의 광고에서 시각적 요소의 중요성을 고려할 때 음성 정보만을 사용한 분류 방법의 한계를 나타낸다.

표 9는 본 연구에서 사용된 데이터 세트의 용량을 비교 분석한 것을 나타낸다. 원본 데이터는 초기 광고 동영상 데이터 세트를 나타내며, 그 크기를 총 366GB에 달한다. 반면 추출된 오디오 텍스트 데이터는 원본 데이터에서 추출된 음성을 텍스트로 변환한 데이터로, 그 크기는 기존 데이터 세트보다 훨씬 줄어든 7.09MB이다. 이와 같이 오디오만을 추출하여 텍스트로 변환하는 방식의 접근은 처리해야 할 데이터의 양을 상당히 줄였으며, 이는 처리 시간의 단축과 컴퓨팅 자원 요구의 감소로 이어질 수 있다.

표 9. 데이터 세트 용량 비교

Table 9. Comparison of dataset capacity

Dataset	Capacity
Original data	368GB
Extracted audio text data	7.09MB

본 연구는 기존의 영상 기반 분류 방법론과는 다르게 자연어 처리 기법을 활용하여 광고 영상을 분류하는 새로운 접근법을 제시한다. 이러한 접근법은 고성능 영상 처리 모델과 비슷한 수준의 성능을 달성하였음에도 불구하고, 컴퓨팅 자원의 효율적 사용이라는 중요한 이점을 제공한다. 또한, 품질이 낮거나 손상된 영상에서도 효과적으로 분류 작업을 수행할 수 있는 능력을 보여준다. 이러한 측면에서 본 연구는 광고 영상 분류 분야에 새로운 대안적 접근법을 제공하며, 향후 이 분야의 다양한 연구와 응용에 기여할 수 있을 것으로 기대된다.

## V. 결론

본 연구는 이전에 연구되지 않았던 온라인 광고 동영상의 클래스 분류 문제에 대한 해결책을 제시하였다는 점에서 큰 의의가 있다. 한국어 기반 BERT 모델인 KoBERT를 활용하여 약 91%의 정확도를 달성하였음을 검증하였다. 이러한 결과는 제안된 방법이 광고 동영상 분류에 실제로 적용할 수 있음을 보여주며, 광고 산업의 효율성을 향상하고, 추후에는 사용자에게 맞춤형 광고 콘텐츠를 제공할 수 있을 것으로 기대된다.

본 연구의 중요한 특징 중 하나는 BERT 기반의 모델을 통한 문맥의 중의성을 해결 능력이다. BERT 모델은 양방향으로 문맥을 분석하여 각 단어의 의미를 더욱 정확히 파악할 수 있다. 예를 들어, '시원하다'라는 단어가 음료, 목욕탕 등 다양한 광고 문맥에서 사용될 때, KoBERT는 주변 단어와의 관계를 분석하여 정확한 의미를 파악할 수 있다. 이는 광고 동영상



상의 내용을 더욱 정밀하게 분류하고 이해하는 데 크게 기여한다.

그러나, 본 연구에는 여전히 몇 가지 한계점이 존재한다. 첫째, 본 연구에서 사용된 데이터 세트는 상대적으로 작아, 추가적인 데이터를 활용하여 모델의 성능을 더욱 향상할 필요가 있다. 둘째, 본 연구에서는 광고 동영상의 음성 데이터만을 이용하였기 때문에, 시각적 정보까지 활용한다면 더 높은 성능을 달성할 수 있을 것으로 예상된다.

향후 연구 방향으로는, 더 큰 데이터 세트를 활용하여 모델의 성능을 향상하고, 다양한 언어의 광고 동영상에 대한 분류 성능을 검증하는 것을 제안한다. 또한, 동영상의 시각적 정보를 포함하는 연구도 필요하다. 이러한 방법들은 광고 동영상 분류의 성능을 더욱 향상할 수 있을 것으로 기대된다.

## 감사의 글

본 연구는 2020년도 중소벤처기업부의 중소기업기술혁신개발사업(수출지향형)사업 지원에 의한 연구임 [S2879452]

## 참고문헌

- [1] Statista. Video Advertising [Internet]. Available: <https://www.statista.com/outlook/dmo/digital-advertising/video-advertising/worldwide>.
- [2] Wyzowl. Video Marketing Statistics 2023 [Internet]. Available: <https://www.wyzowl.com/video-marketing-statistics/>.
- [3] J. Schwenzow, J. Hartmann, A. Schikowsky, and M. Heitmann, "Understanding Videos at Scale: How to Extract Insights for Business Research," *Journal of Business Research*, Vol. 123, pp. 367-379, February 2021. <https://doi.org/10.1016/j.jbusres.2020.09.059>
- [4] Y. Deldjoo, M. Elahi, P. Cremonesi, F. Garzotto, P. Piazzolla, and M. Quadrana, "Content-Based Video Recommendation System Based on Stylistic Visual Features," *Journal on Data Semantics*, Vol. 5, No. 2, pp. 99-113, June 2016. <https://doi.org/10.1007/s13740-016-0060-9>
- [5] C. Luo, Y. Peng, T. Zhu, and L. Li, "An Optimization Framework of Video Advertising: Using Deep Learning Algorithm Based on Global Image Information," *Cluster Computing*, Vol. 22, No. 4 supplement, pp. 8939-8951, July 2019. <https://doi.org/10.1007/s10586-018-2024-3>
- [6] S. Aftan and H. Shah, "A Survey on BERT and Its Applications," in *Proceedings of the 2023 20th Learning and Technology Conference (L&T)*, Jeddah, Saudi Arabia, pp. 161-166, January 2023. <https://doi.org/10.1109/LT58159.2023.10092289>
- [7] Q. Zhang, M. Li, W. Dong, M. Zuo, S. Wei, S. Song, and D. Ai, "An Entity Relationship Extraction Model Based on BERT-BLSTM-CRF for Food Safety Domain," *Computational Intelligence and Neuroscience*, Vol. 2022, 7773259, April 2022. <https://doi.org/10.1155/2022/7773259>
- [8] Y. Wen, Y. Liang, and X. Zhu, "Sentiment Analysis of Hotel Online Reviews Using the BERT Model and ERNIE Model-Data from China," *PloS ONE*, Vol. 18, No. 3, e0275382, March 2023. <https://doi.org/10.1371/journal.pone.0275382>
- [9] B. Yang, X. Luo, K. Sun, and M. Y. Luo, "Recent Progress on Text Summarisation Based on BERT and GPT," in *Proceedings of the 16th International Conference on Knowledge Science, Engineering and Management (KSEM 2023)*, Guangzhou, China, pp. 225-241, August 2023. [https://doi.org/10.1007/978-3-031-40292-0\\_19](https://doi.org/10.1007/978-3-031-40292-0_19)
- [10] K. Kaur and P. Kaur, "Improving BERT Model for Requirements Classification by Bidirectional LSTM-CNN Deep Model," *Computers and Electrical Engineering*, Vol. 108, 108699, May 2023. <https://doi.org/10.1016/j.compeleceng.2023.108699>
- [11] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese Neural Networks for One-Shot Image Recognition," in *Proceedings of the 32nd International Conference on Machine Learning*, Lille, France, July 2015.
- [12] M. AlMousa, R. Benlamri, and R. Khoury, "NLP-Enriched Automatic Video Segmentation," in *Proceedings of the 6th International Conference on Multimedia Computing and Systems (ICMCS)*, Rabat, Morocco, pp. 1-6, May 2018. <https://doi.org/10.1109/ICMCS.2018.8525880>
- [13] Z. Shou, J. Pan, J. Chan, K. Miyazawa, H. Mansour, A. Vetro, ... and S.-F. Chang, "Online Detection of Action Start in Untrimmed, Streaming Videos," in *Proceedings of the 15th European Conference on Computer Vision (ECCV 2018)*, Munich, Germany, pp. 551-568, September 2018. [https://doi.org/10.1007/978-3-030-01219-9\\_33](https://doi.org/10.1007/978-3-030-01219-9_33)
- [14] A. M. Dhakal, Political-Advertisement Video Classification Using Deep Learning Methods, Master's Thesis, Iowa State University, Ames, IA, 2019.
- [15] A. Althnian, D. AlSaeed, H. Al-Baity, A. Samha, A. Bin Dris, N. Alzakari, ... and H. Kurdi, "Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain," *Applied Sciences*, Vol. 11, No. 2, 796, January 2021. <https://doi.org/10.3390/app11020796>
- [16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, ... and I. Polosukhin, "Attention is All You

- Need,” in *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach: CA, pp. 6000-6010, December 2017. <https://doi.org/10.48550/arXiv.1706.03762>
- [17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding,” arXiv:1810.04805, October 2018. <https://doi.org/10.48550/arXiv.1810.04805>
- [18] C. Liu, Y. Zhao, X. Cui, and Y. Zhao, “A Comparative Research of Different Granularities in Korean Text Classification,” in *Proceedings of 2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, Dalian, China, pp. 486-489, August 2022. <https://doi.org/10.1109/AEECA5500.2022.9919047>
- [19] Y. Kim, J.-H. Kim, J. M. Lee, M. J. Jang, Y. J. Yum, S. Kim, ... and S. Song, “A Pre-Trained BERT for Korean Medical Natural Language Processing,” *Scientific Reports*, Vol. 12, 13847, August 2022. <https://doi.org/10.1038/s41598-022-17806-8>
- [20] D. Kim, D. Lee, J. Park, S. Oh, S. Kwon, I. Lee, and D. Choi, “KB-BERT: Training and Application of Korean Pre-Trained Language Model in Financial Domain,” *Journal of Intelligence and Information Systems*, Vol. 28, No. 2, pp. 191-206, June 2022. <https://doi.org/10.13088/JIIS.2022.28.2.191>
- [21] S. Y. Yoo and O. R. Jeong, “Korean Contextual Information Extraction System Using BERT and Knowledge Graph,” *Journal of Internet Computing and Services*, Vol. 21, No. 3, pp. 123-131, June 2020. <https://doi.org/10.7472/JKSII.2020.21.3.123>
- [22] C. Sun, X. Qiu, Y. Xu, and X. Huang, “How to Fine-Tune BERT for Text Classification?,” in *Proceedings of the 18th China National Conference on Chinese Computational Linguistics (CCL 2019)*, Kunming, China, pp. 194-206, October 2019. [https://doi.org/10.1007/978-3-030-32381-3\\_16](https://doi.org/10.1007/978-3-030-32381-3_16)
- [23] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, “ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations,” arXiv:1909.11942, September 2019. <https://doi.org/10.48550/arXiv.1909.11942>
- [24] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, ... and L. Zettlemoyer, “BART: Denoising Sequence-to-Sequence Pre-Training for Natural Language Generation, Translation, and Comprehension,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, pp. 7871-7880, July 2020. <https://doi.org/10.18653/v1/2020.acl-main.703>
- [25] Korea Federation of Advertising Associations Advertising Information Center. TVCF Creative [Internet]. Available: <https://www.adic.or.kr/ad/category/tv/list.cjsp>.
- [26] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, T. Darrell, and K. Saenko, “Long-Term Recurrent Convolutional Networks for Visual Recognition and Description,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Boston: MA, pp. 2625-2634, June 2015. <https://doi.org/10.1109/CVPR.2015.7298878>
- [27] Z. Cai, L. Wang, X. Peng, and Y. Qiao, “Multi-View Super Vector for Action Recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus: OH, pp. 596-603, June 2014. <https://doi.org/10.1109/CVPR.2014.83>
- [28] N. Srivastava, E. Mansimov, and R. Salakhutdinov, “Unsupervised Learning of Video Representations Using LSTMs,” in *Proceedings of the 32nd International Conference on Machine Learning*, Lille, France, pp. 843-852, July 2015. <https://doi.org/10.48550/arXiv.1502.04681>
- [29] L. Sun, K. Jia, D.-Y. Yeung, and B. E. Shi, “Human Action Recognition Using Factorized Spatio-Temporal Convolutional Networks,” in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, pp. 4597-4605, December 2015. <https://doi.org/10.1109/ICCV.2015.522>
- [30] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, “Learning Spatiotemporal Features with 3D Convolutional Networks,” in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, pp.4489-4497, December 2015. <https://doi.org/10.1109/ICCV.2015.510>
- [31] X. Peng, L. Wang, X. Wang, and Y. Qiao, “Bag of Visual Words and Fusion Methods for Action Recognition: Comprehensive Study and Good Practice,” *Computer Vision and Image Understanding*, Vol. 150, pp. 109-125, September 2016. <https://doi.org/10.1016/j.cviu.2016.03.013>
- [32] K. Simonyan and A. Zisserman, “Two-Stream Convolutional Networks for Action Recognition in Videos,” in *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS '14)*, Montreal, Canada, December 2014. <https://doi.org/10.48550/arXiv.1406.2199>
- [33] G. Lev, G. Sadeh, B. Klein, and L. Wolf, “RNN Fisher Vectors for Action Recognition and Image Annotation,” in *Proceedings of the 14th European Conference on*

*Computer Vision (ECCV 2016)*, Amsterdam, The Netherlands, pp. 833-850, October 2016. [https://doi.org/10.1007/978-3-319-46466-4\\_50](https://doi.org/10.1007/978-3-319-46466-4_50)

- [34] J. Y.-H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond Short Snippets: Deep Networks for Video Classification," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston: MA, pp. 4694-4702, June 2015. <https://doi.org/10.1109/CVPR.2015.7299101>
- [35] E. Park, X. Han, T. L. Berg, and A. C. Berg, "Combining Multiple Sources of Knowledge in Deep CNNs for Action Recognition," in *Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Placid: NY, pp. 1-8, March 2016. <https://doi.org/10.1109/WACV.2016.7477589>
- [36] S. Zha, F. Luisier, W. Andrews, N. Srivastava, and R. Salakhutdinov, "Exploiting Image-Trained CNN Architectures for Unconstrained Video Classification," arXiv:1503.04144, March 2015. <https://doi.org/10.48550/arXiv.1503.04144>
- [37] L. Wang, Y. Qiao, and X. Tang, "Action Recognition with Trajectory-Pooled Deep-Convolutional Descriptors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston: MA, pp. 4305-4314, June 2015. <https://doi.org/10.1109/CVPR.2015.7299059>
- [38] X. Yang, P. Molchanov, and J. Kautz, "Multilayer and Multimodal Fusion of Deep Neural Networks for Video Classification," in *Proceedings of the 24th ACM international conference on Multimedia*, Amsterdam, The Netherlands, pp. 978-987, October 2016. <https://doi.org/10.1145/2964284.2964297>
- [39] X. Wang, A. Farhadi, and A. Gupta, "Actions ~ Transformations," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas: NV, pp. 2658-2667, June 2016. <https://doi.org/10.1109/CVPR.2016.291>

### 박연지(Yeon-Ji Park)



2020년 : 광운대학교 컴퓨터소프트웨어학과 (공학사)

2020년~현 재: 광운대학교 전자통신공학과 대학원 박사수료  
※ 관심분야 : 인공지능, 자연어처리, 영상처리 등

### 양근제(Geun-Je Yang)



2022년 : 광운대학교 전자통신공학과 (공학사)

2022년~현 재: 광운대학교 전자통신공학과 대학원 석사과정  
※ 관심분야 : 인공지능, 자연어처리, 영상처리 등

### 손채봉(Chae-Bong Sohn)



1993년 : 광운대학교 전자공학과 (공학사)

1995년 : 광운대학교 전자공학과 대학원 (공학석사)

2006년 : 광운대학교 전자공학과 대학원 (공학박사)

1991년~1993년: 삼성전자 소프트웨어 멤버십 1기

2014년~2017년: 광운대학교 정보통신처 처장

2006년~현 재: 광운대학교 전자통신공학과 부교수

※ 관심분야 : 인공지능, 스마트신호처리시스템, 병렬프로그래밍, 유헬스시스템, 디지털방송시스템 등