한국언론학보
2024년 68권 2호 100~139쪽
https://doi.org/10.20879/kjjcs.2024.68.2.003
http://www.comm.or.kr

시위 뉴스 영상에서 폭력 프레이밍의 작동 기제 분석

비전 트랜스포머(Vision Transformer)를 활용한 폭력 이미지 분류를 통해

이문혁 경희대학교 미디어학과 박사수료
이종혁 경희대학교 미디어학과 교수

Analyzing Violence Framing Mechanisms in Protest News Videos*,**

Classifying Violent Images using Vision Transformer

Moon Hyuk Lee***
(Ph.D. Candidate, Department of Media, Kyung Hee University)
Jong Hyuk Lee****
(Professor, Department of Media, Kyung Hee University)

Protests are acts in which citizens exercise their basic rights. However, citizen rallies opposing government policies or labor strikes demanding wage increases are often suppressed as illegal and portrayed negatively by the media. Journalism research challenges the media's reporting techniques, known as the 'protest paradigm', by pointing out that protests are frequently described as disturbances and confrontations. Most studies about pretest news have focused on textual analysis, with little in-depth analysis on news videos. In this regard, this study examined the video editing strategies used to frame violence in broadcast news coverage of protests. Specifically, editing strategies that emphasize violence in videos can be discussed from two perspectives: the location and duration of violence-related shots. From the first standpoint, it is expected that violence-related scenes will be put early in the news story to attract viewers' attention. From the second perspective, it is predicted that violence scenes are expected to be edited in such a way that the number of brief shots is maximized in order to enhance tension and capture viewers' attention. As a result, it is reasonable to expect that shots involving violence will be brief. To verify these hypotheses, this study developed a classifier to determine the presence of violence in images based on the Vision Transformer (ViT). The

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(이 논문은 2022년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임)[NRF-2022S1A5C2A03093660].

** 논문의 발전에 큰 도움을 주신 익명의 심사위원들께 감사를 드립니다.

*** dalvitt@gmail.com

**** jonghhhh@khu.ac.kr, corresponding author

researchers fine-tuned the publicly available vit-large-patch16-224 model on Hugging Face by replacing the output class into violent/non-violent categories. The classifier achieved high levels of accuracy (97.12%) and F1 score. Subsequently, the researchers collected 335 news videos (from 9 broadcasters) on "Labor Day protests" from Naver News between 2003 and 2023. From these, 13,156 keyframes were identified as violent or non-violent using the developed violence classifier. The results showed that more violent scenes were observed in keyframes located in the early parts of the news story, and more violent scenes were observed in keyframes with shorter durations. Moreover, there was a significant interaction effect between the location and duration of keyframes. This indicates that media emphasizing violent scenes tend to place such scenes at the beginning of the video and employ various similar shots for rapid editing. This editing strategy may be designed to capture the audience's attention by highlighting the deviance news value. The protest paradigm in media coverage of protests includes not only riot and confrontation frames but also discussion frame. Korean media, mindful of viewer ratings, tends to use violent frames including riots and confrontations. Moving forward, it is essential for the media to focus on the themes of protests as socially significant issues and to facilitate the exchange of opinions among societal members through discussion frames.

Keywords: Protest Paradigm, Violence, Framing, Image Classification, Vision Transformer

국문초록

집회나 시위는 국민의 기본권 행사 행위이다. 그럼에도 불구하고 정부 정책에 반대하는 시민 집회나 임금 인상을 요구하는 노동자 시위가 불법으로 단속되며 언론에 의해 부정적으로 다뤄지는 경우가 많다. 본 연구는 이런 문제의식을 바탕으로 방송뉴스의 시위 보도에서 폭력 프레이밍에 사용되는 영상 편집 전략을 살펴보았다. 구체적으로, 영상에서 폭력성을 강조하는 편집 전략은 화면의 위치와 숏의 지속시간이라는 두 가지 관점에서 논의될 수 있다. 첫 번째 관점에서 폭력 관련 화면은 시청자의 관심을 끌 수 있으므로, 뉴스 스토리 내에서 초반부에 배치될 것으로 예측됐다. 두 번째 관점에서 폭력 화면은 생동감과 긴장감을 높여 시청자 관심을 끌기 위해 최대한 많은 화면이 짧게 구성되는 방식으로 편집될 것으로 예측된다. 따라서 폭력 화면의 숏 지속시간이 상대적으로 짧을 것으로 예측할 수 있다. 본 연구에서는 위 가설을 검증하기 위해 Vision Transformer(ViT)를 바탕으로 이미지의 폭력 여부를 판단하는 분류기를 개발했다. 구체적으로, 연구진은 허깅페이스(Hugging Face)에 공개된 vit-large-patch16-224 모델에 최종 출력을 폭력/비폭력로 전환하는 미세조정(fine-tuning)을 실시해 분류기를 개발했다. 사용된 학습데이터셋은 로보플로우(Roboflow)에 공개된 이미지 데이터(Dinesh Narianir의 Violence&not_violence Computer Vision Project)였다. 분류기의 정확도(accuracy)와 F1 값은 모두 97.12%로 대체로 높은 수준을 기록했다. 이어서 본 연구진은 네이버 뉴스에서 '노동절 시위'로 2003년~2023년 검색된 뉴스 영상 335건(9개 방송사)을 수집했다. 여기에서 추출된 키프레임 13,156개는 앞서 개발된 폭력 여부 분류기를 통해 폭력과 비폭력으로 분류됐다. 분석 결과, 뉴스 스토리의 초반부에 위치한 키프레임에서 (후반부 키프레임에 비해) 더 많은 폭력 장면이 관찰됐으며, 지속시간이 짧은 키프레임에서 (긴 키프레임에 비해) 더 많은 폭

력 장면이 나타났다. 또한, 키프레임의 위치와 지속시간 사이에 상호작용 효과도 유의미하게 나타났다. 이는 폭력적 장면을 중시하는 언론이 이런 장면을 영상의 초반에 위치시키고 다양한 촬영 장면을 동원해 빠르게 편집한다는 것이다. 시위 관련 화면은 대체로 집회, 연설, 구호, 행진, 퍼포먼스의 장면으로 구성되며, 때때로 몸싸움, 화염병, 기물 파손, 점거, 소동 등 폭력적 장면을 동반한다. 이 가운데 폭력적 장면이 영상의 초반부에 배치돼 시청자의 즉각적 관심을 끄는 역할을 하고 있는 것이다. 또한 다양한 폭력적 장면이 짧게 여러 컷 배치되면서 시청자의 관심을 증폭시키는 것이다. 이와 같은 영상 편집 전략에는 일탈성 뉴스가치를 앞세워 시청자의 관심을 유도하고 시청률을 올리려는 목적이 엿보인다. 이런 편집은 시청자에게 시위의 내용과 목표를 충분히 전달하지 못한다. 시위 관련 취재보도 관행인 '시위 패러다임'에는 폭동 프레임과 대치 프레임뿐 아니라 토론 프레임도 있다. 우리 언론이 시청률을 의식해 폭동과 대치 등 폭력 관련 프레임을 사용하는 관행을 개선하고, 시위 내용에 주목하고 사회적 토론을 유도하는 역할을 맡아야 하겠다.

핵심어 : 시위 패러다임, 폭력, 프레임, 이미지 분류, Vision Transformer

1. 서론

집회나 시위는 국민의 기본권(헌법 제21조 제1항) 행사 행위임에도, 언론 보도에 그 내용과 의미가 충분히 전달되지 않는 경우가 많다. 정부 정책에 반대하는 시민 집회나 임금 인상을 요구하는 노동자 시위가 불법으로 단속되며, 언론은 이를 폭력 프레임을 사용해 부정적으로 다루기 일쑤다(이화연·윤순진, 2013; 임양준, 2009; 홍주현·나은경, 2015). 언론의 시위에 대한 보도는 '시위 패러다임(Protest paradigm)'으로 저널리즘 연구에서 익히 논의돼 왔다(Arpan et al., 2006; Brasted, 2005; McLeod & Hertog, 1992; Mourão, Brown, & Sylvie, 2021). 시위 패러다임은 언론이 시위 현장에서 어떤 장면을 선택하고 어떻게 묘사하느냐와 관련된 취재 보도 관행을 의미한다. 여기에는 시위를 바라보는 관점으로 폭동, 대치, 스펙터클, 토론 프레임 등이 제시된다(Harlow & Bachmann, 2023). 문제는 이 가운데 폭력과 관련되는 폭동과 대치 프레임이 가장 빈번하게 사용된다는 점이다. 이런 시위 패러다임에 갇힌 언론 보도에서는 시위 주제(내용)에 대한 사회적 논란보다 시위대의 폭력적 행위나 경찰과의 충돌이 강조된다(Brasted, 2005). 그루버(Gruber, 2023)는 1980년대 이후 시위 보도 연구들을 분석했는데, 대부분의 연구에서 이와 같은 유형의 시위 패러다임이 나타났음을 확인했다.

언론이 일탈적이며 부정적 장면에 관심을 가지며 더 높은 뉴스가치를 부여하는 관행이 시위 보도에도 적용되고 있는 것이다(Shoemaker & Cohen, 2006; Shoemaker, Danielian, & Brendlinger, 1991). 사람이 폭력과 같은 일탈성에 본능적으로 끌리는 것은 원시 인류로부터 전승된 심리적 기제라고 한다(Shoemaker & Cohen, 2006). 언론이 시위의 폭력성에 특별한 관심을 갖는 이유도 폭력 장면에서 시청률이 잘 나오기 때문이다. 이에 대한 문제의식 없이 언론 현장에서는 폭력 중심의 시위보도가 관행이 되고, 대부분의 기자가 이런 관행을 바탕으로 제약된 시간과 자원 속에서 사건을 쉽게 구성한다(Berkowitz, 1992; McLeod & Hertog, 1992). 구체적으로, 시위 현장에 나선 기자는 마감 시간을 앞두고 시위 내용을 충분히 취재할 여유가 부족한 데에다, 시위 내용의 이해를 도와줄 취재원도 충분히 확보하기 어렵다. 이런 상황에서 기자는 폭력적 행위와 같이 시청자의 관심을 끌 자극적 요소를 중심으로 쉽게 기사를 구성하고, 이런 관행이 지속되는 것이다.

그동안 선행연구들은 언론이 시위보도에서 폭력 프레임을 사용하고 폭력 관련 표현을 동원해, 시위대를 반사회 불법 집단으로 묘사하고 있음을 보여줬다(변영수, 2016; 이화연·윤순진, 2013; 임양준, 2009; 홍주현·나은경, 2015). 대부분의 연구에서는 기사 텍스트에 대한 단어, 표현, 프레임에 대한 분석이 이뤄졌다. 반면, 시위 관련 뉴스영상에 대한 심층분석은 거의 실시

되지 않았다. 영상의 시각적 요소는 시청자에게 즉각적이며 감정적 반응을 유발한다는 점에서 때로는 텍스트보다 강한 영향을 미친다(Rodriguez & Dimitrova, 2011; Wischmann, 1987). 이 때문에 언론이 뉴스 영상에서 폭력 장면을 어떻게 다루느냐는 시위보도의 문제를 규명하는 데에 필수 질문이 될 것이다. 영상 이미지를 텍스트로부터 분리해 독립적 영향력을 분석한 연구가 요구되는 이유이다(Arpan et al., 2006).

이와 같은 연구의 부족을 메우기 위해 본 연구진은 '노동절 시위' 관련 뉴스 영상을 수집해 폭력 장면 분석을 실시했다. 특별히 이미지 분류용 딥러닝 모델인 Vision Transformer를 바탕으로 폭력/비폭력 자동 분류기를 개발해 분석에 적용했다. 연구 결과는 우리나라 언론의 시위보도 영상에서 폭력 장면에 대한 편집 전략과 실제를 보여줬다. 이 연구가 우리 사회에서 시위 관련 방송보도의 문제점을 확인하고, 개선책을 논의하는 계기가 되길 기대한다.

2. 이론적 배경

1) 언론 보도에서 '시위 패러다임'과 폭력성

언론은 매일 발생하는 다양한 사건을 취재하고 보도한다. 이를 위해 기자들이 새로운 사건을 구성하는 중요한 구성 요소를 선택적으로 취재하고, 요소들을 구성하고 설명할 관점을 따라 기사를 작성한다. 기자들은 어떤 기준에 따라 이렇게 쉽지 않은 작업을 매일 수행할까? 찬과 리(Chan & Lee, 1984)는 기자들이 관심 대상인 사건의 성격을 규정하는 일종의 형이상학적 세계관(metaphysical world-view)을 가지고 있다고 설명하며, 이를 '저널리스틱 패러다임(journalistic paradigms)'이란 개념으로 소개한다. 저널리스틱 패러다임은 기자들로 하여금 사건의 뉴스가치를 평가하고 중요한 부분을 관행에 따라 구성하도록 돕는다. 이와 같은 암묵적 관행적 취재보도 가이드라인 덕분에 개별 기자들은 저널리즘 활동에 쉽게 적응하고 언론사와 이용자의 기대에 부합하는 기사를 어렵지 않게 생산해내는 것이다. 저널리스틱 패러다임이 없었다면, 기자들은 짧은 마감 시간과 제한된 자원으로 많은 정보를 수집하고 중요한 요소를 선별하는 작업을 제대로 수행하지 못할 것이다. 이를 감안하면, 저널리스틱 패러다임은 기자들이 새롭게 등장한 사건에 대해 실용적인 취재 절차와 허용 가능한 해석의 범위를 설정해 "예상치 못한 것의 일상화(routinizing the unexpected)"(Tuchman, 1973, p. 110)에 성공하게끔 돕는 셈이다.

이러한 설명은 뉴스보도가 기자 개인의 의도와 능력을 넘어 다양한 관행과 환경의 영향을 받는다는 미디어사회학적 관점과 부합한다. 구체적으로, 슈메이커와 리즈(Shoemaker &

Reese, 1996)는 뉴스 콘텐츠의 생산에 영향을 미치는 요인으로 개인, 관행, 미디어 조직, 조직 외부 기관, 이데올로기의 5가지를 제안했다. 이 요인들은 위계적으로 구성되며, 기자 개인의 영향력은 가장 미약하며 다른 요인들에 의해 제약받는 상황으로 그려진다. 예를 들면, 노동자 시위에 대한 보도에서 기자 개인이 시위대가 주장하는 노동 관련 법 개정에 초점을 맞추었다고 하자. 이 때 개인보다 상위의 여러 요인이 작동해 최종 보도를 변질시킨다. 취재보도 관행 차원에서 새롭거나 일탈적인 현상을 찾아 사건의 뉴스가치를 확보하는 것이 요구된다. 미디어 조직과 광고주의 입장에서는 반기업적인 노동자 시위에 대해 부정적이거나 적어도 비판적 관점을 유지하도록 압박한다. 사회의 이데올로기 차원에서 자본주의 유지와 사회 안전이라는 가치가 취재보도의 방향에 영향을 미칠 수 있다. 이와 같이 다양한 차원의 요인이 기자 개인의 취재보도에 가이드라인처럼 영향을 미치며, 기자는 교육과 경험을 통해 이러한 영향을 저널리스틱 패러다임으로 받아들이고 스스로 활용하게 된다.

시위 관련 취재보도에는 시위 패러다임(protest paradigm)이 있다. 맥클라우드와 헤어토그(McLeod & Hertog, 1992)는 "언론이 사회적 시위(protest) 사건을 취재보도하는 일상화된 저널리스틱 패러다임"이라고 정의한다(260쪽). 이는 기자가 시위를 취재보도할 때 따르는 오래된 관행이 있으며, 지금도 이에 따른 선택과 강조의 판단이 이뤄지고 있음을 시사한다. 이 때문에 처음 접하는 시위에 대해서도 누구를 주인공으로 어떻게 이야기를 구성해야 할지에 대한 인지적 전략이 쉽게 구성된다는 것이다(Berkowitz, 1992). 결국, 시위 패러다임은 언론이 시위 사건을 다루는 관행이면서, 기자가 제약된 시간과 자원 속에서 시위를 뉴스로 쉽게 구성하도록 도와주는 인지적 사용 목록(mental catalogue)인 셈이다(Berkowitz, 1992; McLeod & Hertog, 1992).

위의 시위 패러다임 사례들을 유형화한 연구도 있다. 할로우와 바크만(Harlow & Bachmann, 2023)은 언론의 시위 패러다임의 작동이 4가지 프레임의 사용으로 나타난다고 보았다. 첫째, 폭동 프레임(riot frame)은 시위에서 발생한 소동, 파괴, 폭력, 충돌 등을 다룬다. 이를 통해 시위대의 반사회적 성격을 부각하고, 시위의 불법성을 강조한다. 둘째, 대치 프레임(confrontation frame)에서는 시위 참가자들과 경찰이나 정부 당국 사이의 충돌이나 맞대응이 다뤄진다. 직접적 폭력이 나타나지 않더라도 이에 준하는 갈등의 순간들이 묘사된다. 셋째, 스펙터클 프레임(spectacle frame)에서는 시위대의 의상, 움직임, 퍼포먼스, 반응 등 대중의 관심을 끄는 장면들이 나타난다. 시위 사건을 드라마화하고 선정적 측면을 강조해 시청자의 감정을 자극한다. 넷째, 토론 프레임(debate frame)에서는 시위의 배경, 목표, 요구사항 등을 전달하며 참여자들의 입장을 설명한다. 제기된 문제에 대한 토론을 유도하고 해법을 모색하는 계기를

제공할 수도 있다.

이 가운데 폭동 프레임과 대치 프레임은 언론의 시위 보도에서 가장 흔하게 등장한다. 이를 통해 시위 참가자들을 불법화(delegitimization), 주변화(marginalization), 악마화(demonization)하는 효과를 낳을 수 있다(McLeod & Hertog, 1992). 불법화는 시위 개최가 적법한 절차를 거치지 않았으며 참여자들의 과격한 활동이 법에 저촉된다는 점을 강조하는 방식으로 이뤄진다. 주변화는 시위 참여자들이 일반 시민과 다른 소수에 불과하며 이들의 의견도 일반 시민의 주류 여론과 동떨어져 있다고 보도되는 경우에 나타난다. 악마화는 시위 참여자들이 반사회적이며 위협적 존재라며 부정적인 묘사를 집중하는 언론 보도에서 관찰된다. 용산참사 뉴스를 분석한 임양준(2009)은 언론이 철거민 시위를 불법폭력, 과격시위, 강경투쟁 등에 초점을 맞춰 보도했으며, 사회적 안정과 도덕성을 강조하는 편향적 태도를 보였다고 비판했다. 이화연과 윤순진(2013)도 밀양 송전탑 반대 시위 보도를 분석해, 보수 언론이 폭력과 대립, 주민의 위법성, 분신 등을 다루는 폭력 프레임을 주로 사용하며, 사건의 배경에 대한 환경정의 프레임을 사용하지 않았다고 지적했다. 변영수(2016)는 뉴스 텍스트 분석을 통해 보수 언론이 2008년 미국산 쇠고기 수입 반대 촛불집회를 과격하고 폭력적인 집회로 구성했음을 밝혔다. 언론이 사용한 '차로를 완전 점거하고', '경찰에게 폭력을 휘두르는', '청와대 진격 투쟁' 등의 표현이 근거로 제시됐다. 또한, 시위 참가자를 피해자 관점에서 분석한 홍주현과 나은경(2015)은 언론이 참가자들을 '일탈적 행동을 하는 피해자'와 '합법적 지각하는 피해자'로 나눠 차별적으로 다룬다고 비판했다. 특히 전자로 분류된 시위 참가자에 대해서는 반정부 세력, 집회 불법성, 갈등 불법성, 유가족 폭행 등의 프레임을 동원해 폭력성을 강조하는 태도를 취했다고 밝혔다. 언론의 시위에 대한 폭력적 부정적 묘사는 사실 오래된 국제적 관행이라고도 할 수 있다. 할린(Hallin, 1986)에 따르면, 1960~1970년대 학생 시위 보도를 분석한 연구에서 시위대는 질서 파괴자로, 경찰은 법 수호자로 묘사됐다고 한다. 언론에 나타난 시위는 대체로 일탈 행위로 규정됐고, 시위의 이슈보다 이로 인한 피해가 주로 다뤄졌다는 것이다. 언론의 시위에 대한 피상적 부정적 보도는 1960년대 미국 대학생들의 베트남 전쟁에 대한 반전시위(Gitlin, 1980)와 원자력 발전에 여론 갈등(Gamson & Modigliani, 1989)의 경우에서도 관찰됐다.

그렇다면, 언론은 시위에서 폭력적 장면에 왜 주목할까? 대부분의 언론이 사회적 갈등 사건에 대해 기존 질서와 체계를 옹호하는 보수적 태도를 가지고 있어, 폭력이나 충돌 등의 일탈적 현상에 대해 법과 질서 프레임을 주로 사용한다는 설명이 있다(장용호, 1987). 주류 언론이 기존의 권력 집단을 옹호하는(예를 들면, 노동자보다 고용주를, 시위대보다 경찰을 긍정적으로 보는) 보수적 성향을 가지고 있다는 점도 고려할 만하다(양정혜, 2001). 그러나 언론의 보수성만

으로는 시위 보도에서 폭력 장면이 선호되는 근원적 원인을 찾기 어렵다. 본 연구에서는 폭력에 대한 인간의 본능적 관심과 이에 따라 언론이 폭력 장면에 부여하는 높은 뉴스가치를 부여하는 관행에 주목한다. 인지심리학 관점에서 폭력과 같은 일탈성은 인간의 관심을 자연적으로 유도하는 속성으로 이해된다. 구체적으로, 오만(Ohman, 2000)의 공포 가동 모델(fear activation model)은 인간이 공포 탐색 지각 장치(fear detection perceptional system)를 가지고 있으며, 이에 의해 위험한 사물(threatening feature)에 특별한 주의를 보인다고 한다.

이런 아이디어에 착상해 언론학자 슈메이커(Shoemaker)는 진화심리학적 설명을 바탕으로 뉴스가치 모형(newsworthiness model)을 제시했다(Shoemaker, 1996; Shoemaker & Cohen, 2006; Shoemaker et al., 1991). 여기에서 사건의 뉴스가치 판단 기준은 일탈성(deviance)과 사회적 중요성(social significance)으로 크게 구분된다. 일탈성은 자주 발생하지 않으며 사회 질서를 위협하거나 규범에 어긋나는 사건에서 높게 나타난다. 사회적 중요성은 한 사회에서 구성원들이 공통적으로 중요하다고 인식하는 사건에 높게 부여된다. 흥미로운 점은 일탈성 높은 사건에 대해 언론은 물론 뉴스 이용자들도 본능적으로 관심을 가진다는 것이며, 이런 성향은 인류가 원시시대부터 물려받은 보편적 심리기제라는 것이다. 원시 시대에 맹수의 습격이나 다른 종족의 공격과 같은 일탈적 사건을 빨리 인지한 인간은 도망쳐 생명을 보존했지만, 일탈성에 부주의했던 인간은 생명을 잃었다. 이는 환경에 의한 자연선택 과정이며, 이로 인해 일탈성에 관심을 가지는 유전적 형질이 후세대로 이어졌다는 것이다. 이후 인류는 이와 같은 일탈적 사건을 미리 인지하고자 보초(sentinel)를 세웠으며, 누구보다 빨리 일탈적 사건을 알아채고 전체에게 알리는 보초의 일이 현재의 언론의 역할로 진화했다는 것이다.

시위에서 폭력이 발생하면 길 가던 행인들도 걸음을 멈추고 관심을 갖는다. 취재하던 기자들은 카메라를 들고 현장으로 뛰어든다. 위 논의에 따르면, 행인들은 인간의 본능에 의해, 기자들은 뉴스가치 판단에 의해 폭력 장면에 집중하는 것이다. 결국, 시위에서 폭력 장면은 뉴스가치가 높게 부여되는 언론 관행에 의해 집중취재 대상이 되었으며, 신입 기자도 이러한 관행에 따라 폭력적이며 자극적인 장면에 특별한 관심을 갖도록 학습된다는 것이다(McLeod & Hertog, 1992). 이제 시위 패러다임 활용에 익숙한 기자들은 시위 현장에서 어떤 측면에 주목해야 하는지를 이해한다. 예를 들면, 도로 행진 중인 시위대에서 과격한 언사, 경찰과의 몸싸움, 도로 점거 등 폭력적 장면을 우선적으로 취재하게 만든다. 이렇게 취재된 장면은 편집에서 폭력 장면이 더욱 돋보이게 재구성될 가능성이 있다. 폭력 장면은 시청자로 하여금 본능적으로 관심을 가지는 할 것이며, 이는 뉴스의 시청률 증가로 이어질 수 있기 때문이다.

지금까지 언론이 왜 시위 속에서 폭력적 장면을 선택하고 폭력 프레임을 사용하는지를 설

명했다. 폭력과 같은 일탈성은 인간의 본능적 관심을 유발하며, 이는 취재에서 뉴스가치 판단 기준으로 작용한다. 이러한 시위 패러다임(취재보도 관행)이 오랫동안 자리 잡았고, 이에 익숙한 기자들이 시위의 내용보다 비정상적이거나 일탈적인 장면을 집중적으로 쫓는다. 이에 따라 시위에 대한 언론 보도는 폭동과 대치 등에 집중하는 폭력 프레임으로 주로 구성되며, 시위 참가자들은 폭력의 행위자나 불법 집단으로 규정되곤 한다. 문제는 이러한 관행이 시위의 내용을 사회에 충분히 알리지 못하고, 선의를 가진 시위 참가자들을 악마화하는 결과를 낳는다는 것이다. 폭력 장면으로 시청률을 높이려는 뉴스 제작자의 의도는 문제를 더욱 악화할 수 있다. 본 연구에서는 이런 문제의식을 바탕으로 시위 관련 뉴스 영상에서 폭력적 장면의 구성과 배치가 실제로 어떻게 나타나는지 살펴보고자 한다.

2) 영상의 선택과 편집: 위치와 지속시간을 중심으로

영상 제작은 촬영과 편집으로 나눠진다. 촬영은 카메라의 운용에 있어 움직임 및 화면의 크기와 위치 등을 선택하는 것이고, 편집은 촬영된 영상을 효과적으로 전달하기 위한 장면의 선택과 배열을 의미한다(설진아, 2007). 무엇을 어떻게 영상으로 담을 것인가가 촬영의 영역이라면, 무엇을 어떻게 보여줄 것인가가 편집의 역할이라고 할 수 있다. 영상이 담아내는 일차적 소재가 현실이라고 하더라도, 어떤 시각을 통해 재구성됐는가에 따라 그 의미는 충분히 달라진다. 이 때문에 제작자의 관점이 영상의 촬영 및 편집에 있어 매우 중요하다(최이정, 2013). 즉, 제작자의 기획 의도, 시청자와 매체 특성에 대한 고려, 편집자의 판단이 편집에서 '무엇을 사용하고 버릴 것인가?'와 '무엇을 강조하거나 강조하지 않을 것인가?'에 영향을 미치는 것이다(Zettl, 2016).

영상 편집에도 수용자의 관심과 기억을 극대화하기 위해 관습적으로 사용하는 영상 선택과 강조의 관행이 있다. 예를 들면, 영상 제작자들은 허리케인이나 테러와 같은 위기 상황의 이미지가 시청자의 관심을 끈다는 것을 알고 있으며, 이런 이미지를 '최고의 장면' 혹은 '돈 되는 장면'이라고 부르기도 한다. 실제, 사람은 생존 본능에 기인해 긍정적 메시지보다 부정적 메시지에 잘 반응하며(Zillmann, 2002), 분노나 두려움 유발 영상을 잘 기억하는(Newhagen & Reeves, 1992) 경향을 가지고 있다. 영상 제작자들도 뉴스 영상에 대한 수용자의 반응과 기억을 높이기 위해 이런 효과를 낳는 콘텐츠를 선호한다. 시위 보도에서 폭력 장면은 시청자로 하여금 사건에 대한 부정적 이해와 두려운 감정을 유발할 것으로 예측된다. 이에 따라 시청자의 관심과 반응을 필요로 하는 영상 제작자들에게 폭력 장면은 매우 중요하게 부각돼야 할 요소인 셈이다. 이창훈(2012)은 자극적이고 볼거리 있는 뉴스 아이템의 편성 경쟁이 폭력과 관련된 CCTV 영상의 사용 증가로 이어졌다고 지적했다.

영상 편집에 있어 프로그램 내 장면의 배치 순서는 시청자에게 중요한 영향을 미치는 프레이밍 방식이다. 전 워싱턴포스트의 저널리스트 트레비스 폭스(Travis Fox)는 시각적으로 강렬하고 청중을 사로잡을 강력한 화면으로 초반을 구성하라고 강조하면서(Lancaster, 2013), 가장 좋은 것부터 시작한다는 영상 편집 규칙을 제시했다. 이는 뉴스 리포트 초반에 영상과 기사의 일치도가 높게 구성해 시청자 집중도를 높이는 방식(김성환, 2022)과 함께 영상 편집의 관행으로 자리 잡았다. 장석호(1994)는 TV의 작은 화면을 고려할 때 영상의 도입부에서 과감하게 주제에 뛰어들어 시청자의 관심에 있어 기선을 제압하는 것이 무엇보다 중요하다고 주장한다.

화면 배치 순서의 효과는 수용자의 인지적 정보처리 과정에서도 유의미하게 나타나는 것으로 밝혀졌다. 최윤정(2008)은 초두효과(primacy effect, Asch, 1946)를 소개하며, 영상의 초반부에 배치된 화면이 수용자의 평가와 투표 의향에 강한 영향을 미치고 있음을 확인했다. 실제로 시청자의 주목도를 높이기 위해 자극적 이미지를 뉴스 초반에 배치하는 경향이 관찰되고 있다(최민재, 2005). 이처럼 자극적 화면의 초반부 배치 전략은 분노와 공포 등 불쾌감을 유발하는 이미지가 잘 기억된다는 점과(Newhagen, 1998) 부정적 이미지의 삽입이 인지 자원의 배분을 촉진해 후속 정보의 효율적 처리를 돕는다는 점(Newhagen & Reeves, 1992)에서 시청자 유입과 유지에 효과적이라고 할 수 있다.

영상 편집에서는 숏(shot)의 지속시간(길이)도 효과적인 프레이밍 전략으로 사용된다. 숏이란 영상 구성의 최소 단위이며, 이러한 숏이 컷[1](cut)이나 여타의 방법으로 다른 영상과 연결되는 사이의 시간적 간격을 숏의 지속시간이라고 부른다(Zettl, 2016). 숏의 지속시간을 정하는 것은 숏과 숏을 나누는 컷의 속도라고 할 수 있는데, 빠르거나 느린 컷은 사건의 리듬을 결정할 뿐 아니라 사건의 밀도를 조정하며, 빠른 컷일수록 사건의 밀도가 증가하고 강렬함이 더해진다(Zettl, 2016). 할린(Hallin, 1992)은 1968년부터 1988년까지 미국 대통령 선거 관련 뉴스에서 평균 사운드바이트[2]가 43초에서 9초로 줄어들었다고 밝혔다. 이에 대한 원인으로 뉴스 산업

1) 컷(cut)의 정의는 하나의 이미지에서 다른 이미지로의 전환이다(Zettl, 2016). 예를 들어 2개의 숏(shot)으로 구성된 영상은 하나의 컷(cut)으로 연결된 것이다. 하지만, 이러한 컷과 숏을 엄밀히 구분하지 않고 컷을 숏의 의미로 사용하기도 한다. 백선기·최경진·윤호진(2011)은 방송 뉴스를 국제 비교하면서 '영상 및 그래픽 컷 수'를 분석했는데, 이는 엄밀히 말하면 뉴스 영상의 숏 수를 의미하는 것이며, 김수정(2003)의 연구에서도 "한국 뉴스가 '짧은 컷'의 사용 빈도가 높다"고 표현하고 있는데 이 또한 '짧은 숏'을 의미한다. 본 연구에서는 본래의 의미로 영상과 영상을 나누는 것을 그 기법과 상관없이 '컷'으로, 이러한 컷으로 나눠지지 않은 최소 단위의 영상을 '숏'으로 정의해서 사용한다.

2) 사운드바이트는(sound bite)는 라디오에서 유래된 용어로 뉴스에서 누군가가 말하는 것을 보여주는 것을 의미한다(Hallin, 1992). 이종수(1999)는 사운드바이트를 뉴스의 구조적 요소의 하나로 설정하면서 1)공식적인 인터뷰 2)비공식적인 인터뷰 3)공식석상 발표 4)현장소리화면의 4가지로 구분한다.

의 경쟁이 심화하면서 우위를 점하기 위한 뉴스 제작자들의 전략으로 사운드바이트 시간의 축소가 나타났다는 것이다. 이와 관련해 커밍스(Cummings, 2014)는 하나의 뉴스 영상을 편집하는 데 필요한 평균 숏의 수가 대체로 30개이며, 최근에는 컷 수가 증가하고 사운드바이트가 짧아지는 경향이 관찰된다고 했다. 한정된 시간만 허락되는 뉴스 영상에서 컷 수의 증가는 사운드바이트뿐 아니라 전체적인 숏 지속시간의 감소를 의미한다.

뉴스 영상에서 숏의 지속시간이 짧아지는 경향은 국내 뉴스의 경우 더욱 두드러진다. 이종수(1999)는 지상파 3사의 저녁 뉴스의 평균 숏 지속시간이 4.3초로 1995년의 9.7초에 비해 현저히 짧아졌다고 밝혔다. 이는 뉴스 제작자의 개입을 통한 작위성(artificiality, maniplulation)의 정도가 높아지고 있음을 시사한다. 국제 비교 연구에서도 KBS와 SBS 뉴스가 해외 방송사의 뉴스에 비해 영상에서 월등히 많은 숏을 사용하고 있는 것으로 나타났다(백선기 등, 2011). 특히, KBS와 SBS에서 인터뷰 영상의 평균 숏 지속시간이 각각 7.9초와 8.4초로 나타났는데, 이는 인터뷰보다 긴박하게 편집되는 현장 영상의 숏 지속시간은 더욱 짧을 수 있음을 시사한다. 뉴스 영상의 평균 숏 지속시간보다 긴 인터뷰 숏 지속시간이 유지되기 위해서는 여타 영상의 숏이 상대적으로 짧아야 하기 때문이다. 또한, 김수정(2003)은 국내 뉴스의 영상 구성의 특징으로 많은 화면이 편집을 통해 신속히 보이고, 짧은 숏의 사용 및 부단한 카메라의 조작과 움직임이 동원된다는 점을 들었다. 이런 편집을 통해 뉴스 제작자가 얻고자 하는 것은 극적 효과와 함께 생동감과 현장감이라는 것이다. 이화섭(2016)도 짧은 컷 중심의 반복적 영상 구성이 국내 지상파 방송의 주된 편집 방식이라며, 이는 짧게 반복되는 영상이 역동성을 유지하고 주목 효과를 높여주기 때문이라고 설명한다. 특히, 재난 사건이나 폭력적 영상을 사용하는 사건·사고 담당 기자의 경우에는 이런 자극적 영상에 대한 선호가 더 심하다는 것이다. 심지어 사건 담당 데스크가 10초로 편집된 영상을 4초 이내의 숏으로 나눌 것을 지시하기도 한다고 설명한다. 이런 자극적 편집의 이유는 기대 시청률의 상승 때문이며, 뉴스 제작자들은 느슨하고 편안한 편집을 하면 시청자가 떠날 것이라는 두려움을 갖고 있다고 지적한다. 이와 같은 연구 결과들은 국내 뉴스 영상에 있어서 자극적이고 폭력적인 영상일수록 숏 시간이 짧게 편집될 가능성이 있음을 보여준다.

3. 연구가설

본 연구에서는 시위 관련 영상 뉴스에서 키프레임(화면)이 어떻게 다뤄지는지 살펴봄으로써 폭력 프레이밍의 기제를 규명해 보고자 한다[3]. 이를 위해 우리나라 방송이 2003년부터 2023년까

지 국내외 노동절(labor day) 시위에 대해 보도한 뉴스 영상을 분석 대상으로 설정했다. 노동절 시위는 오랜 역사, 전 세계적인 참여 범위, 참여자의 다양성 등에서 다른 시위들에 비해 높은 대표성을 가질 것으로 판단됐다. 1890년 시작한 노동절 시위는 인종(세계 인종차별 철폐의 날, 1960년 시작), 여성(세계 여성의 날, 1910년 시작), 장애인(세계 장애인의 날, 1981년 시작) 등 다른 유형의 시위보다 오래 전에 시작돼 현재까지 이어지고 있다(신재길, 2023). 또한 노동절은 전 세계 160개국 이상에서 기념되고 있으며, 노동 시위 건수는 미국의 경우에 최근 더욱 증가하고 있다(Heilman, 2023; Mutikani, 2024). 참여 인원들도 제조업뿐 아니라 금융, 의료, 영화계 등 다양한 분야의 노동자들로 구성된다. 어떤 시위는 노동절 시위보다 큰 규모로 나타나지만(예를 들면, 한국의 박근혜 대통령 퇴진 촛불집회나 프랑스의 노란조끼 운동 등), 일시적 현상으로 그치는 경우가 많다. 이런 이유로 본 연구진은 노동절 시위를 분석 대상으로 선정했다. 또한, 노동절 시위 관련 보도는 국내와 국외의 시위를 모두 포함해 강한 수준의 폭력부터 폭력이 동반되지 않은 경우까지 다양한 사례를 보여줄 수 있을 것이다.

위 시위 관련 뉴스 영상의 취재와 보도는 앞서 설명한 저널리스트의 시위 패러다임에 따라 예상해 볼 수 있다. 갈등적 자극적 소재에 뉴스가치를 부여하고 중요하게 다뤄 시청자의 관심을 끄는 관행은 우리 언론에서도 나타날 것으로 예측된다(McLeod & Hertog, 1992). 폭력 관련 사안을 영상에 담아 시위를 폭동이나 대치 프레임으로 전달하는 관행도 예상해 볼 수 있다(Harlow & Bachmann, 2023).

영상에서 폭력성을 강조하는 편집 전략은 키프레임 단위에서 화면의 위치와 숏의 지속시간이라는 두 가지 관점에서 논의될 수 있다. 일단, 화면 선택은 편집자가 주어진 영상 속에서 특정 화면(영상 속 프레임)을 선택하거나 배제하는 방식으로 이뤄진다(Zettl, 2016). 앞에서의 예측처럼 폭력 관련 화면은 시청자의 관심을 끌 수 있으므로, 다른 화면에 비해 선택될 가능성이 높을 것이다. 구체적으로, 영상 편집자는 시위 패러다임의 관행에 따라 폭력 화면을 중요하게 다루고자 할 것이며, 이를 위한 전략으로 폭력적 화면(키프레임)을 뉴스 스토리 내에서 초반부에 배치할 것으로 예측된다.[4] 폭력적 화면은 시청자의 관심을 끌려는 목적에 적합하며, 스토리의 초

---

3) 본 연구에서 사용하는 용어를 다음과 같이 설명한다. 어떤 TV 뉴스 프로그램(예를 들면 KBS 9시 뉴스)이 하루에 30개 꼭지의 리포트(대략 2분 이하)를 제공한다고 할 때, '뉴스 스토리'는 이 꼭지 하나에 해당하는 영상을 의미한다. 뉴스 스토리는 프레임(이미지 스틸 컷)으로 구성되는데, 이 프레임을 '화면'이라고 하겠다.

4) 영상을 구성하는 정지된 이미지를 프레임이라고 하며(뉴스 영상은 초당 30개의 프레임으로 구성), 이 가운데 영상의 내용을 대표할 수 있는 프레임들을 키프레임이라고 한다. 키프레임은 영상에서 중요한 장면을 찾아내거나 전체 내용을 요약하는 데에 필수적으로 사용된다(하종우·윤성웅·김민수·안창원, 2022). 방송뉴스 연구에서 백지연·윤호영(2021)은 영상의 키프레임을 추출해 성범죄 피해 여성의 묘사 방식에 대해 분석한 바 있다. 본 연구에서는 하나의 뉴스 스토리의

반부 이미지가 시청자의 관심 유발과 기억 저장에 효과적이기 때문이다(최민재, 2005; 최윤정, 2008; Newhagen, 1998). 중요한 화면을 과감하게 초반부에 배치해 시청자를 사로잡아야 한다는 영상 편집 실무상의 조언도 많다(장석호, 1994; Lancaster, 2013). 이에 따라 폭력 수준이 높은 화면(키프레임)이 스토리의 초반부에 배치될 것으로 예측해 볼 수 있다. 폭력성을 부각하려는 영상 편집의 입장에서는 폭력장면이 많으면서 폭력장면을 스토리의 초반부에 배치하려고 할 것이다. 이에 따라 화면(키프레임) 배치와 관련된 다음 가설을 제시한다.

연구 가설 1(키프레임 위치). 뉴스 스토리를 구성하는 키프레임 가운데 (후반부에 비해) 초반부 키프레임에서 폭력적 이미지가 나타날 가능성이 높을 것이다.

다음으로 숏의 지속시간과 관련된 가설을 제시한다. 시위 상황에서 시위대와 경찰의 충돌은 뉴스 영상 판단에 중요한 요소이며(김학희·이재경, 2009), 영상 편집자는 뉴스 시청자의 주목을 끌고 영상의 극적 효과와 생동감을 높이기 위해 최대한 많은 영상을 짧게 사용할 것이다(김수정, 2003). 특히 인터뷰를 제외한 현장 영상의 숏 지속시간은 상대적으로 더욱 짧을 것이며(백선기 등, 2011; Cummings, 2014), 자극적인 영상에 대한 뉴스 제작자들의 선호와 제작 관행 및 시청률에 대한 기대(이화섭, 2016), 그리고 영상의 숏 지속시간이 수용자에게 주는 효과(Zettl, 2016)를 고려할 때, 뉴스 영상 중 폭력적인 영상의 숏 지속시간은 더욱 짧을 것으로 예측할 수 있다.

숏의 지속시간은 본 연구에서 키프레임의 지속시간으로 조작적으로 정의한다. 영상 편집에서 숏과 숏을 컷으로 나눈다는 의미는 서로 다른 내용의 영상이 컷의 형태로 연결된다는 것이다. 영상을 구성하는 정지 이미지(화면)인 프레임을 기준으로 하면, 숏이 컷으로 나눠져 있을 때 앞의 숏의 마지막 프레임과 뒤의 숏의 첫 번째 프레임이 다르다는 것을 의미한다. 이러한 숏과 프레임의 특성을 활용하면, 프레임의 텐서 값을 기준으로 영상의 숏과 숏을 나누는 컷의 위치를 추정할 수 있으며, 기술적으로 키프레임을 추출해 숏을 대표하는 이미지로 사용할 수 있다. 이를 통해 키프레임의 지속시간으로 숏의 지속시간을 측정할 수 있다. 이에 따라 다음 가설을 제시한다.

연구 가설 2(키프레임 지속시간). 뉴스 스토리를 구성하는 키프레임 가운데 지속시간이 상대적으

내용을 대표할 수 있는 키프레임을 추출하고, 폭력으로 분류되는 키프레임의 위치와 지속시간의 관찰을 통해 폭력적 장면의 편집 방식을 살펴보고자 한다.

로 짧은 키프레임에서 폭력적 이미지가 나타날 가능성이 높을 것이다.

위 연구가설 1과 2를 종합하면, 키프레임의 위치가 초반부일수록, 키프레임의 지속시간이 짧을수록 폭력적 이미지가 나타날 가능성이 높을 것으로 예측할 수 있다. 이는 키프레임의 폭력 여부 예측에서 키프레임의 위치와 지속시간 사이에 상호작용 효과가 나타날 수 있음을 시사한다. 구체적으로, 키프레임이 초반부에 위치하고 지속시간이 짧은 경우에 폭력 이미지가 나타날 가능성이 가장 높을 것으로 예측해 볼 수 있다. 이를 반영해 연구가설 3을 제시한다.

연구 가설 3(프레임 위치 X 지속시간). 키프레임의 위치와 지속시간 사이에 상호작용 효과가 나타날 것이다. 구체적으로, 초반부에 위치하고 지속시간이 짧은 키프레임에서 폭력적 이미지가 나타날 가능성이 높을 것이다.

4. 연구 방법

본 연구에서는 위 가설을 검증하기 위해 딥러닝모델 기반 전이학습을 통해 폭력 이미지 분류기를 개발하고, '노동절 시위'에 대한 방송뉴스를 수집해 분석했다. 노동절 시위는 매년 5월 1일 전후로 나타나는 국내외 노동자 시위에 대한 보도로, 장기간 우리 언론의 시위 보도 관행에 대해 살펴볼 기회가 될 수 있다. 여기에서는 이미지 분류, 폭력에 대한 규정, 데이터 수집과 키프레임 추출, 주요 변인의 측정 등에 대해 설명한다.

1) 이미지 분류에서 Vision Transformer(ViT)의 활용

본 연구에서 필요한 딥러닝 모형은 영상에서 추출된 키프레임 이미지들을 폭력과 비폭력으로 분류하는 것이다. 이는 이미지를 다루는 컴퓨터비전 분야 가운데 '이미지 분류(image classification)'에 해당하는 것으로, 전통적으로 CNN(Convolutional Neural Network)을 바탕으로 여러 가지 모형들이 선을 보였다. 선행연구를 참조해 이에 대한 간단한 설명을 제공한다(참조: 박대민, 2022; 오일석, 2022; 윤호영, 2021; 최윤정 등, 2020; 위키독스, 2023a, 2023b; Dosovitskiy et al., 2020; Joo & Steinert-Threlkeld, 2022; Park & Kim,

2022; Steiner et al., 2021).

우선, 2012년 AlexNet이 이미지 분류 대회(ILSVRC, ImageNet Large-Scale Visual Recognition Challenge)에 나타나 CNN 기반 분류기 개발의 대중화에 앞장섰다. 이어서 연산 층 수를 늘린 VGGNet과 다양한 필터를 적용한 GoogLeNet(Inception)이 등장했다. 2015년 이미지 분류 대회(ILSVRC)에서 우승한 ResNet은 입력 데이터를 연산 층을 건너뛰어 출력에 직접 더하는 스킵 연결(skip connection)을 통해 좋은 성능을 보여주었다. 다음으로 DenseNet은 모든 연산 층과 연결하는 밀집 연결(dense connection)을 도입했고, EfficientNet는 네트워크 깊이(depth), 필터 수(width), 이미지 해상도(resolution)에서 최적의 조합으로 성능을 극대화하기도 했다

2021년 발표된 Vision Transformer(ViT)는 주로 자연어처리에 사용되던 딥러닝 모델인 Transformer를 이미지 분류에 사용해 관심을 끌었다(Dosovitskiy et al., 2020). 이미지 분류에서 거의 빠짐없이 동원된 CNN 핵심구조(backbone) 없이 더 뛰어난 성능을 보인 것이다. 현재 여러 벤치마크 데이터셋에서 ViT 계열의 모델이 우수한 성능을 기록하고 있다(https://paperswithcode.com/task/image-classification, Chen, Hsieh, & Gong, 2021; Smirnov, 2022).

ViT에서는 자연어처리에서 사용하던 BERT(Bidirectional Encoder Representations from Transformers) 또는 Transformer의 Encoder 구조와 attention 메커니즘을 유지하면서, 이미지 입력 단계에 변화를 주었다. 주어진 이미지를 패치(patch)로 나누고 순서대로 배치해 자연어의 시퀀스 데이터처럼 만들어 입력하는 것이다. 예를 들어, 영상의 한 프레임(이미지)을 폭행(assault), 강도(robbery), 비폭력(non-violence) 등의 클래스(class)로 분류한다고 하자(Rendón-Segador, Álvarez-García, Salazar-González, & Tommas, 2023, 〈Figure 1〉 참조). ViT에서는 주어진 프레임을 16x16(픽셀) 크기의 작은 패치로 나눈다. 그림에서는 이 과정을 통해 9개의 패치가 생성됐고, 각 패치는 transformer가 사용하는 768차원의 벡터로 재구성된다. 하나의 이미지가 9(패치)X768(벡터) 차원의 시퀀스 형태로 입력되는 것이다. BERT 활용 자연어 처리에서 한 문장이 여러 개의 토큰으로 나뉘고, 각 토큰이 768 차원의 벡터로 재구성되는 방식과 유사하다. 시퀀스 데이터로 변환된 이미지는 클래스를 나타내는 특별한 토큰과 위치 정보를 담은 임베딩과 함께 ViT 모델에 입력된다. ViT는 BERT와 유사한 처리 단계로 구성되는데, 멀티헤드 셀프어텐션, 층 정규화, 잔차 연결 등이 포함된다(그림의 오른쪽 부분). 이 단계에서 셀프어텐션을 바탕으로 각 패치의 정보가 다른 패치와 비교 및 조합되면서 이미지의 중요한 특징을 포착한다. 이러한 정보는 피드포워드 신경망을 거쳐 처리되며, 네트워크

의 가중치는 역전파를 통해 업데이트된다. 이러한 학습을 위해서는 모델의 마지막 단계(그림의 윗 부분)에 두 개의 완전 연결된 층과 GeLU 활성화 함수를 포함하는 다층 퍼셉트론(MLP)이 분류기(classifier)로 추가된다. 여기에서 학습된 특징을 기반으로 이미지를 서전에 설정된 클래스에 매핑하게 된다. 연구자들은 이렇게 만들어진(사전학습된) ViT 모델을 허깅페이스(https://huggingface.co/)와 같은 아카이브에서 다운로드한 뒤 모델의 최종 단계에 연구 목적에 적합한 다층퍼셉트론(multilayer perceptron, MLP)과 소프트맥스층(softmax layer)을 추가하는 방식으로 자신만의 분류기를 개발할 수 있다.

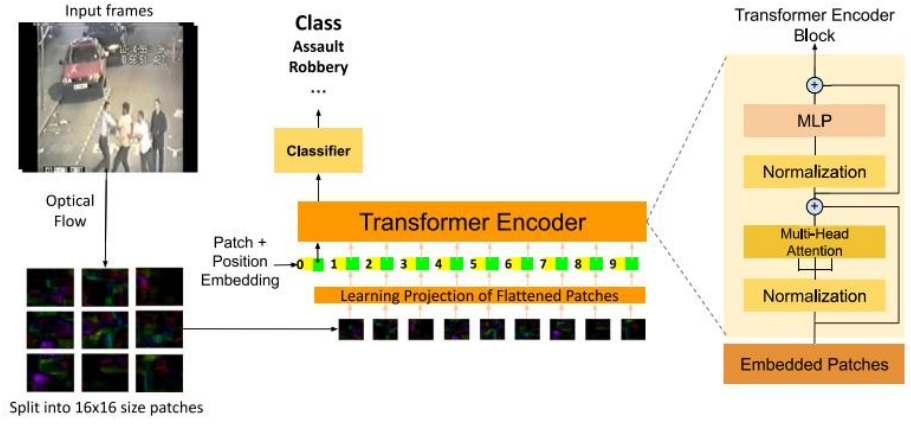


Figure 1. The structure of Vision Transformer (Rendón-Segador et al., 2023, p. 323)

2) 이미지 분류에서 폭력에 대한 정의

이미지를 폭력과 비폭력으로 분류하려면 무엇이 폭력인지에 대한 정의가 필요하다. 세계보건기구(World Health Organization, WHO)는 폭력을 개인이나 집단에 대해 의도적으로 물리적 힘을 사용해 상해, 사망, 심리적 피해 등을 유발하거나 그럴 가능성이 높은 행위로 정의하고 있지만, 이런 정의는 범위가 넓고 추상적이어서 분류기 개발을 위한 지침으로 사용하기 어렵다(Constantin et al., 2020). 이 보다 구체적인 폭력에 대한 정의로는 피를 흘리게 하는 인간 행위(Chen, Hsu, Wang, & Su, 2011), 맥락이나 참여 인원수와 관계없이 싸움이 포함된 장면(Bermejo Nievas, Deniz Suarez, Bueno García, & Sukthankar, 2011), 의도를 가지고 위협하거나 실제로 신체적 피해를 유발하는 행위(Giannakopoulos, Makris, Kosmopoulos, Perantonis, & Theodoridis, 2010), 폭발-총성-싸움 등이 포함되어 빠르게 진행되는 장면(Gong, Wang, Jiang, Huang, & Gao, 2008) 등이 있다. 미디어 콘텐츠의 자동 분류를 위한

국제 연구모임인 MediaEval Benchmarking Initiative for Multimedia Evaluation에서는 폭력 장면 탐지 과제(Violent Scenes Detection task)를 설정하고 딥러닝 모델을 개발했다(http://www.multimediaeval.org/mediaeval2013/violence2013/). 여기에서 학습데이터는 피(blood), 자동차 추격(car chase), 찌르고 때리는 종류의 무기(cold arms), 싸움(fights), 화재(fire), 총기(firearms), 피비린내(gore), 폭발(explosions), 총성(gunshots), 비명(screams) 등 10가지 클래스로 분류됐다(Ionescu, Schlüter, Mironica, & Schedl, 2013).

이와 같은 폭력에 대한 정의와 분류는 대체로 물리적 충돌과 신체적 피해가 명확하게 나타나는 이미지에서 잘 작동할 것으로 보인다. 문제는 이보다 약한 수준의 폭력이 다양한 장르의 미디어 콘텐츠에서 많이 나타나며, 위 분류로는 포착되지 않는다는 점이다. 위에서 제시된 10가지 폭력 클래스 분류도 미국 할리우드 영화에 적용된 것으로, 본 연구 대상인 방송뉴스 분석에는 적합하지 않을 것으로 판단된다. 뉴스보도에서 엄격한 취재보도 윤리 기준에 따라 피 흘리는 싸움이나 총과 칼 등의 무기를 보여주지 않기 때문이다. 그렇다면 방송 영상의 폭력성 판단에 대해 더 넓은 범위를 포괄하면서 실효성 있는 기준이 필요하다. 최근 최윤정, 정유진, 그리고 정금희(2023)는 영상 분석에서 공격성(폭력성) 판단 기준 8가지를 제시했다. 여기에는 ▲시각적으로 관찰되는 공격적 행동(때리기, 차기, 휘두르기, 던지기, 쏘기 등) ▲인물 간 가까운 접촉 ▲인물의 부정적 표정 ▲행동의 빠른 속도 ▲피 흘리는 결과 ▲무기 사용 ▲무생물(건물이나 기물) 파손 ▲역동적 카메라 움직임과 빠른 편집 등이 포함됐다. 이를 종합하면, 폭력적 장면에는 관련 인물들이 가깝게 위치해서 공격적 행동을 보여주고 부정적 표정을 지어야 한다. 행동은 대체로 빠른 속도로 나타나며, 피와 무기가 동반될 수 있다. 여기에 인간이 아닌 건물이나 기물 파손 행위도 폭력적으로 분류된다. 또한, 공격적 행동이 대체로 빠르고 역동적이므로 카메라 촬영 각도가 다양해지고, 시청자 관심을 끌기 위한 빠른 편집이 수반된다.

본 연구진은 위 영상 분석의 폭력성 판단 기준을 언론의 시위 보도의 맥락에서 재구성하고자 했다. 앞서 언론은 시위보도에서 폭동, 대치, 스펙터클, 토론 등 4가지 프레임을 주로 사용한다고 제시했다(Harlow & Bachmann, 2023). 이 가운데 폭력 장면이 나타날 가능성은 폭동과 대치 프레임이라고 할 수 있다. 신체적 충돌, 방화와 약탈, 기물 파손, 점거와 소동 등으로 구성되는 폭동 프레임에서는 위 최윤정 등(2023)의 기준 가운데 여럿을 충족할 것으로 판단된다. 한편, 대치프레임에서는 시위대가 경찰과 매우 가까이 근접한 장면을 바탕으로 할 것이다. 여기에서 관찰되는 행동은 대체로 고성이 오가거나 몸싸움이 벌어지는 정도가 되며, 시위대 표정에는 분노와 슬픔 등의 부정적 패턴이 나타날 것이다. 대치가 격화해 경찰의 강제 진압과 체포가 진행

되면, 명확한 공격적 행동(때리기, 차기, 휘두르기 등)이 나타날 수도 있다. 이를 감안해 본 연구진은 시위 관련 이미지에서 폭력 분류 기준으로 ▲공격적 행동(때리기, 차기, 휘두르기 등) ▲약탈과 방화 ▲기물 파손 ▲점거와 소동 ▲인물 간 근접 거리에 고성, 몸싸움, 부정적 표정이 추가되는 경우로 모두 5가지를 제시한다. 폭동 프레임은 폭력 수위가 비교적 강한 앞의 3가지 기준에서 탐지될 것이며, 대치 프레임은 폭력 수위가 비교적 약한 뒤의 2가지 기준에서 탐지될 것으로 예상한다.

3) 학습데이터와 분류기 학습

다음으로 연구진은 폭력 여부를 클래스로 가진 기존의 공개된 학습데이터를 살펴보았다. 학습데이터를 제공하는 Roboflow(https://universe.roboflow.com/), Kaggle(https://www.kaggle.com/), Papers With Code(https://paperswithcode.com/) 등에서 관련 학습데이터를 검색한 결과, Dinesh Narianir의 Violence&not_violence Computer Vision Project로 제공되는 데이터를 사용하기로 했다(https://universe.roboflow.com/dinesh-nariani-rmnpr/violence-not_violence-ziv7b). 학습-검증-평가에 사용할 이미지 수가 데이터셋 2개 합쳐 모두 39,992개로, 다른 데이터에 비해 매우 많았기 때문이다. 더욱 중요한 선정 이유는 이 데이터가 본 연구가 앞서 제시한 시위 관련 폭력 기준을 충족할 다양한 이미지를 포함하고 있었으며, 비폭력 이미지에서도 다양한 행위, 상황, 장소를 담고 있었기 때문이다. 〈Appendix 1〉은 이 데이터에 포함된 폭력 이미지와 비폭력 이미지 사례를 보여주고 있다.

연구진은 위 학습데이터를 활용해 Vision Transformer(ViT) 사전학습 모델을 바탕으로 분류기를 개발했다. 사용된 모델은 vit-large-patch16-224로 허깅페이스(https://huggingface.co/google/vit-large-patch16-224)에서 쉽게 가져올 수 있다. 허깅페이스의 설명에 따르면, 이 모델은 ImageNet-21k(1,400만 이미지, 21,843개 클래스) 데이터셋에서 224 × 224 해상도로 사전학습됐으며, 이후 ImageNet 2012(100만 이미지, 1,000개 클래스) 데이터셋에서 같은 해상도로 미세조정(fine-tuning)된 것이다. 따라서 이 모델에서는 하나의 이미지가 16 × 16 크기의 패치 196(224 × 224를 16 × 16으로 나눈 값)개로 분할된다. 또한 이 모델은 Large 유형에 속하므로 패치당 768차원이 아니라 1024차원의 벡터를 사용한다. 결국 이미지는 196(패치) × 1024(벡터 차원)의 시퀀스 데이터로 재구성돼 입력된다. 이 ViT 모델을 거쳐 산출되는 클래스 토큰의 1024 차원의 벡터가 이미지의 특징을 대표하는 값이 된다. 본 연구진은 분석 목표에 따라 이 모델의 최종 단계에 있는 MLP(MultiLayer Perceptron)의 출력층을 클래스 2개(폭력/비폭력)로 구성된 새로운 층으로 교체했다. 이를 통해 클래스 토큰의 값이 2차원 벡터로 변환되고, 소

프트맥스 함수를 거쳐 폭력/비폭력 클래스에 속할 확률을 산출하게 조정됐다.

이어서 연구진은 조정된 모델의 학습을 위해 앞서 준비한 Dinesh Narianir의 데이터를 훈련용(train), 검증용(validation), 평가용(test)에 따라 70%:15%:15% 비율로 무작위로 분리했다. 각 집단에 속한 최종 이미지 수는 각각 27,994개, 5,998개, 6000개였다. 이미지 크기(image size)는 사전학습 모델에서 사용된 224 × 224(pixels)로 설정됐다. 학습 과정에서는 다양한 초모수(hyper-parameter) 조건에서 분류기의 정확도를 비교 관찰했다. 가장 높은 성능을 보여준 조건은 학습률(learning rate) 0.00005, 배치 크기(batch size) 32, 에포크(epoch) 20(당초 50으로 설정됐으나, early stopping 적용됨)이었다. 최종 분류기의 정확도(accuracy)와 F1 값은 모두 97.12%로 대체로 높은 수준을 기록했다. 이 과정에서 본 연구진은 파이썬 패키지 transformers, torch, torchvision, timm, sklearn 등을 사용했다.

4) 데이터 수집과 키프레임 추출

위 연구가설 검증을 위해 본 연구진은 네이버 뉴스에서 '노동절 시위'를 키워드로 동영상으로 한정해 검색을 진행했다. 수집 대상 미디어는 지상파 3곳(KBS, MBC, SBS), 종합편성채널(JTBC, TV조선, 채널A, MBN) 4곳, 뉴스전문채널(YTN, 연합뉴스TV) 2곳 등 방송사 9곳으로 설정됐다. 수집 기간은 검색 전체 기간으로 설정됐고, 최종 2003년~2023년으로 나타났다. 연구진은 파이썬의 동적 크롤링 패키지인 selenium을 사용해 검색 결과로 나타나는 기사의 네이버 url과 기본 정보(미디어, 날짜, 제목 등)를 수집했다. 이어서 개별 기사의 네이버 url로 접속해 해당 영상을 짧게 작동시키고, 이 때 크롬 개발자도구 창의 네트워크 탭에 축적되는 영상 파일 정보를 활용해 해당 파일을 다운로드했다. 이 과정은 연구진이 자체 개발한 파이썬 코드로 수행됐다. 이후 수집된 기사와 영상 가운데 중복 케이스를 제거하고, 뉴스 리포트 형식이 아닌 전문가 대담이나 패널 해설 등의 영상도 제거했다. 리포트 뉴스만 선택한 이유는 이 유형이 국내 뉴스의 약 67%를 차지하는 일반적인 방송뉴스 형식이라는 점(반현·홍원식, 2009)과 본 연구 결과를 방송뉴스 일반에 적용해 해석하기 위해서였다. 이 과정을 거쳐 노동절 시위 관련 방송뉴스 335건이 분석 대상으로 결정됐다(〈Table 1〉 참조).

다음으로 연구진은 기사의 영상으로부터 키프레임(keyframe)을 추출했다. 키프레임 추출에서 고려할 사항은 추출 기준이다. 파이썬 코드를 통해 모든 키프레임은 3차원의 텐서(가로 픽셀 수 × 세로 픽셀 수 × 채널RGB)로 변환되며, 이 가운데 어느 정도의 변화가 나타나야 키프레임으로 추출할지 결정해야 한다. 연구진은 같은 영상으로부터 30%, 35%, 40%의 변화율을 적용해 추출된 키프레임과 영상을 비교해 보았다. 방송PD 출신 연구자의 관찰 결과, 40% 변화율

을 적용할 때 키프레임 간 내용 중복이 없으며 전체 영상을 충분히 반영할 수 있다고 판단됐다. 앞 프레임의 텐서에서 뒤 프레임 텐서를 뺀 값의 합이 앞 프레임 텐서 합의 40%를 넘을 때 뒤 프레임을 추출하는 방식이다. 이 과정을 통해 노동절 시위 관련 영상으로부터 키프레임 13,156개가 추출됐다(〈Table 1〉 참조).

Table 1. Number of Video Vews Items and Keyframes Extracted by Media

미디어 유형	미디어	기사	키프레임	키프레임/기사
지상파	KBS	46	2,406	52.30
	MBC	38	1,723	45.34
	SBS	37	1,500	40.54
종합편성	JTBC	19	1,000	52.63
	MBN	10	315	31.50
	TV조선	15	761	50.73
	채널A	5	130	26.00
보도전문	YTN	114	4,204	36.88
	연합뉴스TV	51	1,117	21.90
합계		335	13,156	39.27

5) 키프레임의 폭력 여부, 위치, 지속시간 측정

다음으로 연구진은 개발한 폭력 분류기를 사용해 수집된 모든 키프레임을 폭력와 비폭력으로 분류했다. 전체 13,156개의 키프레임 가운데 2,826개(21.48%)가 폭력으로 분류됐으며, 10,330개(78.52%)가 비폭력으로 분류됐다. 폭력으로 분류된 키프레임을 위에 제시한 폭력 유형별로 살펴보면 〈Figure 2〉와 같은 사례가 관찰된다. 각각의 사례들을 유형별로 살펴보면 먼저 '공격적 행동'의 유형에서는 시위대가 돌을 던지거나 경찰이 물대포를 사용하는 장면 등이 나타났으며, '약탈과 방화'의 유형에는 주로 외국의 시위대가 화염병을 던지거나 차량을 불태우는 장면이 포함됐다. '기물 파손' 유형에서는 외국 시위대가 상점 물건을 부수거나 국내 시위대가 경찰차 차체를 훼손하는 장면에서 관찰됐으며, '점거와 소동' 유형은 시위대가 도로를 점거하고 구호를 외치는 장면에서 나타났다. 마지막으로, '근접 몸싸움' 유형에는 시위대 대치하던 경찰에 근접해 몸싸움을 벌이거나 경찰이 시위대 속에서 시위 참가자를 연행하는 장면에서 관찰됐다. 반면, 비폭력으로 분류된 키프레임에서는 〈Figure 3〉과 같은 사례 유형이 나타났다. 우선, 많은 참가자들이 모인 집회라고 하더라도 움직임이 역동적이지 않은(과격하지 않은) 장면은 비폭력으로 분류

된 것으로 관찰된다. 구호와 피케팅의 경우에도 움직임이 약하거나 경찰과의 대치가 나타나지 않으면 비폭력으로 분류되는 경향도 나타났다. 또한, 시위대 참가자, 경찰, 시민의 발언이나 인터뷰도 비폭력 장면으로 분류됐다.

▲공격적 행동		
	SBS (2023.05.02.) 화염병 던지고 최루탄 터지고…"연금개혁 반대" 격화	JTBC (2015.05.02.) '세월호 집회' 밤새 격렬 충돌…집회 참가자 42명 연행
▲약탈과 방화		
	YTN (2018.05.06.) "대통령 규탄한다!" 러시아 프랑스 대규모 시위	SBS (2016.05.03.) 美, 노동절 맞아 곳곳 격렬한 시위…부상자 속출
▲기물 파손		
	SBS (2016.05.03.) 美, 노동절 맞아 곳곳 격렬한 시위…부상자 속출	TV조선 (2015.05.05.) 〔뉴스쇼 판〕 노동절 '폭력 시위' 연행된 2명 구속 1명 기각
▲점거와 소동		
	JTBC (2019.05.01.) 베네수엘라 무력충돌…최소 1명 사망, 70여 명 부상	SBS (2016.05.03.) 美, 노동절 맞아 곳곳 격렬한 시위…부상자 속출

▲ 근접 + 고성, 몸싸움, 부정적 표정		
	SBS (2009.05.02.) 도심 곳곳서 '노동절 시위' 충돌…70여명 연행	연합뉴스TV (2023.05.31.) 민주노총 도심 대규모 집회…기습 분향소 설치에 충돌도

Figure 2. Examples of keyframes classified as violence

▲ 집회와 구호	▲ 구호와 피케팅	▲ 피케팅
		
KBS (2015.07.15.) 경찰, 세월호 집회 피해 손해배상 청구	JTBC (2015.11.16.) '광우병 파동' 이후 최대 규모 집회…'물대포' 후폭풍	SBS (2009.05.02.) 도심 곳곳서 '노동절 시위' 충돌…70여명 연행
▲ 연설	**▲ 참기자 인터뷰**	**▲ 경찰 인터뷰**
		
MBC (2023.05.17.) 건설노조 상경 투쟁‥"건폭 몰이' 중단하라"	YTN (2018.05.06.) "대통령 규탄한다!" 러시아·프랑스 대규모 시위	MBC (2008.05.01.) 양대노총 등 노동절 기념 행사‥'평화적 진행'

Figure 3. Examples of keyframes classified as non-violence

이와 같은 분류 결과를 검증하기 위해 연구진은 335개의 수집된 기사를 제목과 본문을 검토하며 폭력성 상-중-하 집단으로 수동 분류했다. 폭력성 상 집단에는 격렬 충돌, 연행, 방화, 약탈, 화염병, 살수차, 물대포, 차벽 파손 등이 주요 키워드로 관찰되는 기사들이 포함됐다. 반면, 폭력성 하 집단에는 충돌 없는 집회를 보여주거나 수사나 재판 관련 내용이거나 정부의 시위 대응 방침만 전달하는 등의 기사들이 포함됐다. 폭력성 중 집단에는 양쪽에 속하지 않는 일반적인 집회와 시위에 대한 기사로 판단된 경우가 속했다. 분류 결과, 폭력성 상-중-하 집단에는 각각 기사 117건, 95건, 123건이 포함됐다.[5] 이를 바탕으로 각 집단별 폭력과 비폭력 키프레임의 비율을 비교해 보았다(〈Table 2〉 참조). 폭력성 상-중-하 집단 기사에서 나타나는 폭력 키프레임

비율은 각각 31.93%, 21.46%, 11.39%로 유의미하게 차이나는 것으로 나타났다 (x2=577.500, df=2, p<.001). 이는 분류기에 의한 키프레임의 폭력 결과가 신뢰할 수 있음을 시사한다고 할 수 있다.

Table 2. Ratio of Violent to Non-Violent Keyframes by News Videos with High-Medium-Low Violence

기사의 폭력성	상		중		하	
	개수	비율(%)	개수	비율(%)	개수	비율(%)
비폭력 키프레임	3094	68.07	3074	78.54	4162	88.61
폭력 키프레임	1451	31.93	840	21.46	535	11.39
합계	4545	100	3914	100	4697	100

다음으로 연구진은 수집된 키프레임이 해당 뉴스 스토리 안에서 차지하는 위치와 지속시간을 측정했다. 모든 키프레임은 영상 파일로부터 추출될 때에 해당 파일을 구성하는 모든 프레임 가운데 몇 번째 프레임인지를 기록한 정보와 함께 생성됐다. 예를 들면, 2분짜리 뉴스 영상은 모두 3600개의 프레임(120초 × 30프레임)으로 구성돼 있다. 이 가운데 프레임의 텐서 변화율 40% 기준에서 키프레임 30개가 추출됐다고 하자. 이 때 모든 키프레임은 100, 250, 445, … , 3125, 3300, 3510 등으로 전체 프레임에서 자신의 순서를 가지게 된다. 여기에서 첫 번째 키프레임의 위치는 전체 3600개 가운데 100번째이므로 0.028(100/3600)으로 측정되며, 마지막 키프레임의 위치는 전체 3600개 가운데 3510번째이므로 0.975(3510/3600)이 된다. 이런 방식으로 측정된 키프레임의 위치는 0에서 1 사이의 수치로 측정되며, 초반부(후반부) 프레임일수록 위치 수치가 낮아진다(높아진다).

---

5) 폭력성에 따라 분류된 기사 사례는 다음과 같다.
폭력성 상 집단:
- YTN (2005.05.01.) 건설노조, SK 본사 앞 격렬 시위.
- JTBC (2015.11.14.) 광우병 촛불집회 이후 최대 규모…물대포·차벽 등장.
- SBS (2023.05.02.) 화염병 던지고 최루탄 터지고…"연금개혁 반대" 격화.
폭력성 중 집단:
- TV조선 (2023.05.17.) 건설노조, 밤샘 노숙 농성…오늘도 대규모 집회 이어져.
- 연합뉴스TV (2018.05.01.) 노동절 맞아 세계 곳곳 함성…"노동자 목소리 들어라".
- SBS (2017.05.01.) 근로자의 날 집회…비정규직 철폐·최저임금 인상 요구.
폭력성 하 집단:
- 연합뉴스 (2011.05.02.) 세계 곳곳서 노동절 시위..충돌은 없어.
- SBS (2009.10.14.) 서울경찰청장, 직접 과잉진압 지휘 국감 공방.
- MBC (2021.08.11.) 양경수 민주노총 위원장 구속 여부 오늘 결정.

이어서 키프레임의 지속시간에 대한 조작적 정의를 살펴보자. 하나의 키프레임은 다음 키프레임이 나타날 때까지 해당 시간의 영상 내용을 대표한다고 할 수 있다. 이 시간 동안 많은 프레임들이 영상을 구성하겠지만, 추출된 키프레임의 내용에서 큰 차이가 없이 지속됐다고 볼 수 있다. 따라서 키프레임의 지속시간은 앞선 숏의 시간을 측정한다고 할 수 있다.[6] 이를 측정하기 위해서 본 연구진은 해당 키프레임의 순서 값과 다음 프레임의 순서 값의 차이를 계산했다. 예를 들어, 위 30개 키프레임 가운데 첫 번째 키프레임의 지속시간은 해당 프레임의 순서 값 100과 다음 프레임의 순서 값 250의 차이인 150이라고 할 수 있다. 150을 실제 시간으로 환산하면 (뉴스 영상에서 초당 프레임이 30개인 점을 감안해) 5초에 해당한다(〈Figure 4〉 참조). 마지막 키프레임의 지속시간은 해당 프레임의 순서 값과 전체 프레임 수의 차이로 측정됐다. 위에서 마지막 키프레임은 자신의 순서 값 3510과 전체 프레임 수 3600의 차이인 90을 지속시간으로 갖게 된다. 키프레임의 지속시간은 이처럼 프레임 수나 초 단위의 시간 값으로 측정될 수 있으나, 뉴스 스토리의 시간(길이)에 따라 이 수치가 크게 달라질 수 있으므로 비율로 환산돼야 한다. 앞 사례의 첫 번째 키프레임의 지속시간은 150 프레임 수 또는 5초로 측정됐지만, 전체 뉴스 스토리의 시간 대비 해당 키프레임의 지속시간을 구하면 0.042(150/3600)로 환산된다. 이는 해당 키프레임의 지속시간이 전체 뉴스 스토리의 시간에서 약 4.2%를 차지한다는 뜻이다. 본 연구진은 뉴스 스토리의 시간에 따른 변화를 통제하기 위해 키프레임의 지속시간을 전체 뉴스 스토리 시간에서 해당 키프레임이 차지하는 비율로 최종적으로 측정했다.

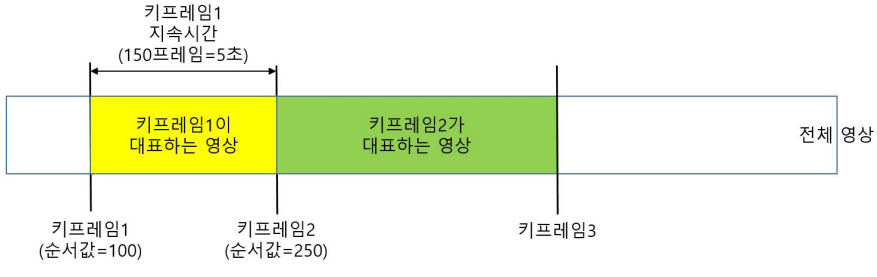


Figure 4. Visualization of keyframe duration

6) 키프레임의 지속시간이란 여러 프레임의 집합인 연속장면의 지속시간을 의미하는 것이 아니라, 전체 영상 중 특정한 부분을 대표하는 하나의 키프레임과 다른 특정한 부분을 대표하는 또 다른 키프레임 사이의 일종의 거리를 의미한다. 이 거리는 키프레임과 키프레임 사이의 프레임 수로 측정할 수 있으며, TV 영상의 경우 1초가 30프레임으로 구성되기 때문에 키프레임의 지속시간은 프레임 수로 표현될 수 있다.

영상의 키프레임의 위치와 지속시간 및 폭력 여부에 대한 분석이 이뤄지면, 이를 바탕으로 가설 검증이 실시됐다. 가설 검증을 위한 분석에서는 교차분석(카이제곱 검증)과 로지스틱회귀분석이 사용됐다. 본 연구진은 연구가설 1을 위해 키프레임의 위치를 도입-전개-심화-결론의 4개 집단으로 나눈 뒤 각 부분의 폭력 프레임 비율을 비교했다. 연구가설 2의 검증에서는 연구진이 키프레임 지속시간을 상-중-하의 3개 집단으로 나눈 뒤 폭력 프레임 비율의 차이를 확인했다. 두 가지 검증에는 교차분석(카이제곱 검증)이 사용됐다. 이어서 연구가설 3을 위해 본 연구진은 키프레임의 위치와 지속시간 및 이들의 상호작용 변인을 독립변인으로 설정하고 키프레임의 폭력 여부를 종속변인으로 설정한 뒤 로지스틱회귀분석을 실시했다. 이를 통해 키프레임의 위치와 지속시간의 주효과는 물론 상호작용 효과까지 검증이 가능했다.

5. 연구 결과

본 연구는 노동절 시위에 대한 뉴스 영상 분석을 통해 우리나라 방송이 시위에 대해 폭력성을 중심으로 보도하는가를 검증하고자 한다. 이를 위해 335건의 뉴스 스토리로부터 13,156개의 키프레임을 추출해 분석했다. 앞서 연구가설에 사용된 주요 변인에 대한 탐색적 통계치를 살펴보면, 우선 키프레임 가운데 폭력으로 분류된 경우는 전체의 21.48%로 나타났다. 이어서 키프레임의 위치는 평균 0.539과 표준편차 0.258로 측정됐다. 이 변인에서는 모든 키프레임 값이 0에서 1 사이로 위치하는 성격을 가지므로, 평균이 중간 값에 가까운 것은 당연하다고 할 수 있다. 다음으로 키프레임의 지속시간은 평균 0.021과 표준편차 0.0366로 나타났다. 이는 키프레임의 시간이 해당 뉴스 스토리 전체 시간에서 차지하는 비율이 평균 2.1%라는 뜻으로, 하나의 뉴스 스토리가 평균 50개의 키프레임으로 구성되고 있다는 것을 의미한다.

연구가설 1에서는 뉴스 스토리를 구성하는 키프레임 가운데 (후반부에 비해) 초반부 키프레임에서 폭력적 이미지가 나타날 가능성이 높을 것으로 예측됐다. 이를 검증하기 위해 우선 키프레임 위치 변인을 낮은 값에서 높은 값까지 4등분해 범주형 변인으로 변환했다. 방송 뉴스가 대체로 도입-전개-심화-결론의 4개 시퀀스로 구성된다는 점을 고려한 것이다(김문환, 2018). 이를 바탕으로 4개 시퀀스별 폭력과 비폭력 키프레임의 비율을 비교하면 유의미한 차이를 확인할 수 있다($x^2 = 35.202$, $df = 3$, $p < .001$)(〈Table 3〉 참조). 폭력 키프레임 비율이 가장 높은 부분은 도입(24.81%)이며, 나머지 전개-심화-결론 부분의 비율은 도입에 비해서 낮게 나타났다. 연구가설 1이 대체로 지지된다고 볼 수 있지만, 이후 키프레임의 위치와 지속시간 및 두

변인 간 상호작용까지 포함한 로지스틱 회귀분석에서 가설 검증 여부를 최종 확인하겠다.

연구가설 2에서는 지속시간이 상대적으로 짧은 키프레임에서 폭력적 이미지가 나타날 가능성이 높을 것으로 예측됐다. 이를 검증하기 위해 키프레임 지속시간 변인을 상-중-하로 3등분한 뒤 집단에 따른 폭력과 비폭력 키프레임의 비율을 비교했다. 〈Table 4〉에 따르면 집단별로 폭력 키프레임 비율에 유의미한 차이가 나타났다(x^2 = 75.438, df = 3, p 〈 .001). 지속시간 집단 가운데 키프레임의 폭력 비율이 가장 높은 곳은 하 집단(23.18%)으로 나타났다. 이는 지속시간이 짧은 키프레임일수록 폭력 장면일 가능성이 높음을 보여주고 있다. 연구가설 2에 대한 최종 검증은 변인들을 종합해 키프레임의 폭력을 예측할 추후 분석에서 살펴보겠다.

앞서 교차분석을 통해 연구가설 1과 2의 검증 가능성을 관찰했는데, 이런 방식은 연속형 변인을 범주화하면서 정보 손실을 초래하고 변인들의 영향력을 (상호작용까지 포함해) 종합적으로 분석하지 못한다는 한계가 있다. 따라서 본 연구에서는 키프레임의 위치와 지속시간 및 이들의 상호작용 변인으로 키프레임의 폭력 여부를 예측하는 로지스틱회귀분석을 실시했다. 〈Table 5〉에 따르면, 키프레임의 위치는 폭력 여부를 유의미하게 예측하는 것으로 나타났다(b = -0.442, p 〈 .001). 키프레임의 위치 변인이 1단위 늘어날 때 해당 키프레임이 비폭력일 확률에 비해 폭력일 확률이 0.364배로 줄어드는 것으로 해석된다. 다시 말하면, 뉴스 스토리에서 (후반부보다) 초반부의 키프레임에서 폭력적 장면이 나타날 가능성이 높다는 것이다. 이는 연구가설 1을 지지하는 결과이다. 키프레임의 지속시간도 폭력 여부를 유의미하게 예측하는 변인이었다(b = -6.240, p 〈 .001). 키프레임의 지속시간이 1단위 늘어날 때 해당 키프레임이 비폭력일 확률에 비해 폭력일 확률이 0.643배 줄어드는 것으로 나타났다. 다시 말하면, 뉴스 스토리에서 지속시간이 상대적으로 짧은 키프레임에서 폭력적 장면이 나타날 가능성이 높다고 할 수 있다. 이로써 연구가설 2도 지지됐다.

다음으로, 키프레임의 위치와 지속시간 사이에 유의미한 상호작용 효과도 관찰됐다(b = 6.883, p 〈 .05). 효과의 방향을 살펴보기 위해 키프레임의 지속시간을 표준편차의 -1배 값, 평균 값, 표준편차의 +1배 값으로 설정한 뒤, 각 경우에 대해 키프레임의 위치가 해당 키프레임이 폭력일 확률을 예측하는 그래프를 시각화했다. 〈Figure 5〉에 따르면, 키프레임의 위치가 해당 키프레임이 폭력일 확률에 부정적 효과를 미치는데(키프레임 위치가 초반부에 있을수록 폭력일 확률이 높아짐), 그 영향력의 크기가 키프레임의 지속시간이 늘어날수록 작아지고 있음을 볼 수 있다. 달리 말하면, 폭력 키프레임이 가장 많이 관찰되는 경우는 키프레임의 위치가 초반부에 있고 지속시간이 상대적으로 짧은 조합에서이다. 이는 키프레임의 위치와 지속시간의 상호작용을 가정한 연구가설 3을 지지하는 결과이다.

Table 3. Ratio of Violent to Non-violent Keyframes by Location of Keyframe

위치	도입		전개		심화		결론	
	개수	비율(%)	개수	비율(%)	개수	비율(%)	개수	비율(%)
폭력	816	24.81	628	19.09	713	21.65	669	20.37
비폭력	2473	75.19	2661	80.91	2580	78.35	2616	79.63
합계	3289	100	3289	100	3293	100	3285	100

Table 4. Ratio of Violent to Non-violent Keyframes by Duration of Keyframe

지속시간	상		중		하	
	개수	비율(%)	개수	비율(%)	개수	비율(%)
폭력	925	21.09	884	20.17	1017	23.18
비폭력	3460	78.91	3499	79.83	3371	76.82
합계	4385	100	4383	100	4388	100

Table 5. Results of Logistic Regression Analysis Predicting the Presence of Violence Based on the Location and Duration of Keyframes[7)]

	b	*se*	*p*	*exp(b)*
상수항	-1.011	0.056	〈.001	0.364
위치	-0.442	0.098	〈.001	0.643
지속시간	-6.240	1.681	〈.001	0.002
위치 X 지속시간	6.883	2.979	0.021	975.333

*모델 x^2 = 40.327, *df* = 3, *p* 〈. 001, Nagelkerke R^2 = 0.005

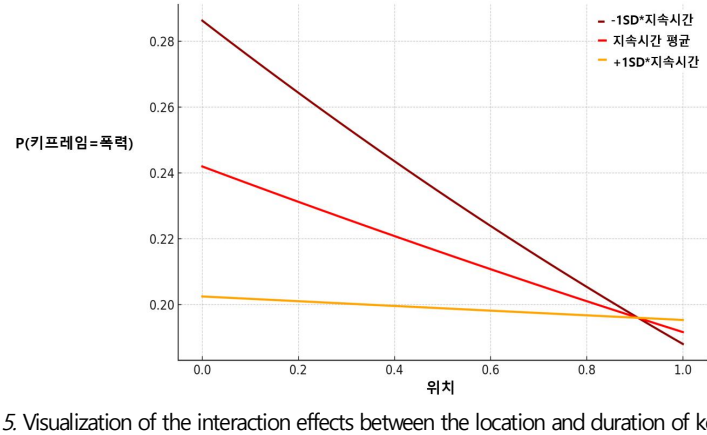


Figure 5. Visualization of the interaction effects between the location and duration of keyframes

7) 모델의 다중공선성에는 문제가 없는 것으로 나타났다. 위치의 VIF=1.41, 지속시간의 VIF=5.61, 상호작용 변인의 VIF=6.16.

6. 논의

본 연구는 방송뉴스의 시위 관련 영상 보도에서 폭력 프레이밍의 작동을 관찰하고, 이러한 프레이밍에 사용되는 편집 전략을 살펴보았다. 분석 결과, 노동절 시위 관련 뉴스 스토리에서 초반부에 위치한 키프레임에서 (후반부 키프레임에 비해) 더 많은 폭력 장면이 관찰됐으며, 지속시간이 짧은 키프레임에서 (긴 키프레임에 비해) 더 많은 폭력 장면이 나타났다. 또한, 키프레임의 위치와 지속시간 사이에 상호작용 효과도 유의미하게 분석됐다. 초반부에 위치하면서 지속시간이 짧은 키프레임에서 폭력 장면이 가장 많이 관찰된 것이다. 이 결과를 통해 본 연구의 가설이 모두 지지됐으며, 이는 우리나라 언론이 시위 보도 영상에서 폭력 장면을 매우 중요시하고 있음을 시사한다. 언론이 시위 보도에서 폭력성에 주목하는 경향은 일부 선행연구에서 밝혀졌지만(변영수, 2016; 임양준, 2009; 이화연·윤순진, 2013; 홍주현·나은경, 2015), 폭력성 부각을 위해 영상 편집 전략이 사용된다는 점은 충분히 알려지지 않았다. 이 연구 결과를 바탕으로 영상 뉴스에서 교묘하게 이뤄지는 폭력 프레이밍에 대한 더 많은 후속연구가 이어지길 기대한다.

우선, 본 연구에서는 시위 현장의 폭력 장면이 뉴스 스토리의 초반부에 배치되고 있음을 확인했다. 뉴스 영상은 대체로 도입-전개-심화-결론의 4개 부분으로 순차적으로 구성된다(김문환, 2018). 이 가운데 도입부는 리포트의 주제를 가장 잘 드러낼 수 있는 현장이나 사례로 보통 구성된다. 시청자의 관심을 끌고 채널이 돌아가지 않도록 하는 것도 도입부의 역할이다. 결국, 도입부는 주제를 선명하게 드러내면서 시청자의 관심을 끄는 두 가지 역할을 동시에 수행해야 한다. 본 연구에서 관찰된 대로, 시위는 대체로 집회, 연설, 구호, 행진, 퍼포먼스의 장면으로 구성되며, 때때로 몸싸움, 화염병, 기물 파손, 점거, 소동 등 폭력적 장면을 동반한다. 기자는 이 가운데 무엇을 선택해(selection) 어떻게 묘사하느냐(description)를 고민해야 하며, 영상 편집에서도 시위를 어떻게 묘사한 장면을 얼마나 많이 선택해 어떤 위치에 배치하느냐를 고민해야 한다(Smith, McCarthy, McPhail, & Augustyn, 2001). 특히 도입부에서 시위의 주제를 명확히 하려면 시위대의 주장을 담은 플래카드나 인터뷰 화면이 포함돼야 한다. 반면, 시청자의 관심을 우선시한다면 이보다 시위대의 폭력적 장면이 더 선호될 것이다. 본 연구 결과는 후자에 가까운 편집이 많았음을 보여준다. 언론이 시위의 내용을 전달하기보다 자극적 이미지로 시청률을 유지하는 편집 전략을 사용했다는 것이다(최민재, 2005).

시위 보도에서 폭력적 장면의 지속시간이 짧았다는 것도 흥미로운 발견이다. 짧은 숏의 사용과 부단한 카메라 움직임을 특징으로 하는 우리나라 뉴스 영상에서도 폭력 장면의 지속시간은 다른 장면의 지속시간에 비해 더욱 짧았다(김수정, 2003). 장면을 짧게 구성한다는 것은 피사체

의 크기, 배경, 촬영 각도를 달리하는 다양한 장면을 많이 동원해 해당 구간을 역동적이며 긴박하게 보이도록 편집한다는 뜻이다. 이화섭(2016)은 이러한 편집으로 시청자의 주목 효과를 높일 수 있다고 한다. 본 연구 결과에서 폭력적 장면은 비폭력 장면에 비해 짧게 구성됐다. 대부분의 뉴스 스토리에서 폭력적 장면 여러 개가 이어지고 있음을 감안하면, 영상 편집자가 시위의 폭력 현장을 다양한 배경과 각도로 촬영된 장면들로 꼼꼼하게 구성했음을 알 수 있다. 폭력이 시위의 핵심 내용이 아님에도 이에 대한 특별한 편집 노력이 동원됐고, 그 이유는 시청자의 관심을 끌려는 전략 때문이라고 추론해 볼 수 있다. 한편으로는 이 결과에 대해 추가 분석을 통해 다양한 해석의 가능성을 열어둬야 한다. 본 연구진의 분석 영상 대부분은 시위 현장을 중심으로 하고 있는데, 관련자 인터뷰를 포함해 시위 이슈를 진단하는 영상에 대한 추가 분석이 필요하다. 다양한 내용과 포맷의 시위 뉴스 영상에 대해 폭력 키프레임의 지속시간을 비교 분석하는 후속연구가 이어져야 하겠다. 본 연구의 또 다른 아쉬운 점은 키프레임의 폭력 강도를 데이터화하지 않고 폭력 여부를 분류하는 데에 그쳤다는 점이다. 폭력 키프레임 가운데에도 강도가 센 경우와 약한 경우에 프레임의 위치나 지속시간의 차이가 나타날 수 있다. 후속연구에서는 키프레임의 폭력 여부뿐 아니라 폭력 강도를 계산해 보다 면밀한 분석이 이루어져야 하겠다.

방송뉴스 취재보도에서 어떤 장면을 중요하게 다루느냐는 기자나 편집자가 어떤 뉴스가치를 중요하게 여기느냐와 관련된다. 앞서 슈메이커(Shoemaker)의 뉴스가치 모형(newsworthiness model)에서는 일탈성과 사회적 중요성이 대표적 뉴스가치 판단 기준으로 제시됐다(Shoemaker, 1996; Shoemaker & Cohen, 2006; Shoemaker et al., 1991). 폭력 장면은 일탈성이 높아 인간의 본능적 관심을 끌며, 이 때문에 시위 보도에서 폭력 장면이 중요하게 편집되는 것이다. 이번에는 사회적 중요성 관점에서 시위의 뉴스가치를 살펴보자. 본 연구에서 분석된 국내외 시위는 최저임금 인상, 일자리 대책, 연금개혁, 반이민정책, 인종차별 등 다양한 주제로 벌어졌다. 이와 같은 시위의 주제로부터 언론은 사회적 중요성 차원에서 사건의 뉴스가치를 판단할 수 있다. 연금개혁이 당시 국민적 관심사였다면, 이 시위는 사회적 중요성 때문에 중요하게 다뤄져야 한다. 영상 편집에서 시위대의 폭력 장면이 아니라 연금개혁에 대한 인터뷰나 피케팅 장면들이 초반부에 위치하고 빠르게 구성되어야 한다. 일탈성에 대한 관심이 유전적 진화의 결과라면, 사회적 중요성에 대한 관심은 문화적 진화로 나타난다(Lee & Choi, 2017). 인간은 한 사회의 구성원으로 학습과 교류 등 사회화 과정을 거치며 자신이 속한 사회에서 중요하게 다뤄야 할 사건이 무엇인지 판단하는 능력을 갖춘다. 이런 맥락에서 언론의 역할은 사회 구성원들이 중요하게 여기는 사건을 앞서 인지하고, 관련 정보를 제공하고 의견 교환을 유도하는 것이 된다. 이를 종합하면, 언론은 시위 보도에서 일탈성 관점에서 폭력 현장을 중시할

것인지, 사회적 중요성 관점에서 시위 주제(내용)를 중시할 것인지 선택할 수 있다. 두 가지 모두 뉴스가치 판단의 과정이지만, 전자는 시청자의 즉각적 관심을 불러일으키며, 후자는 시청자의 사회에 대한 이해를 넓혀준다. 본 연구 결과는 현재 우리 언론의 시위 보도는 전자를 중요시하는 영상 구성을 보여주고 있다.

언론의 시위 패러다임에는 폭동 프레임과 대치 프레임뿐 아니라 토론 프레임도 나타날 수 있다(Harlow & Bachmann, 2023). 그동안 폭동과 대치 프레임이 가장 자주 등장한 점은 국내외 언론 모두 폭력을 중시하는 문제를 안고 있다는 것을 시사한다. 시위를 일탈성이 아닌 사회적 중요성의 렌즈로 바라보면 토론 프레임을 사용할 수 있다. 이 프레임은 시위의 내용과 참가자 입장을 전달하며 사회적 토론을 유도하는 장점을 가지고 있다. 시위 보도의 문제 개선을 위해 사회적 중요성의 관점과 토론 프레임의 활성화가 필요해 보인다. 본래 시위나 집회는 대의민주주의의 단점을 보완해 시민의 정치참여를 보장하는 민주주의 제도의 중요한 요소이다. 시위를 통해 시민과 국가권력 사이에 의견이 자유롭게 교환되어야 하며, 특히 사회적 약자의 처지와 요구가 사회에 알려져야 한다. 이런 점에서 우리 언론은 시위에 대한 취재보도의 초점을 폭력적-자극적-피상적 에피소드에서 논리적-이성적-맥락적 주제로 옮기는 실천에 나서야 하겠다.

References

Arpan, L. M., Baker, K., Lee, Y., Jung, T., Lorusso. L., & Smith, J. (2006). News coverage of social protests and the effects of photographs and prior attitudes. *Mass Communication & Society*, *9*(1), 1-20.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*(3), 258-290.

Baek, G., & Yoon, H. Y. (2021). Key frame analysis of network TV news covering female sexual crime victims: Victimization & sensationalism of visual image. *Korean Journal of Journalism & Communication Studies*, *65*(2), 75-113. [백지연·윤호영 (2021). 방송 뉴스가 재현하는 성범죄 피해 여성 이미지에 대한 키프레임 분석: 가상물, 자료화면을 통한 피해자다움의 재생산과 익명-실명 보도의 차이를 중심으로. <한국언론학보>, 65권 2호, 75-113.]

Baek, S.-G., Choi. K.-J., & Yoon, H.-J. (2011). *International comparative study of broadcast news*. Seoul: Koreal Press Foundation. [백선기·최경진·윤호진 (2011). <방송뉴스의 국제 비교 연구>. 서울: 한국언론진흥재단.]

Ban, H., & Hong, W. (2009). A content analysis of korean broadcasting news format: Focusing on the evening main news of KBS, MBC and SBS. *Studies of Broadcasting Culture*, *21*(1), 9-38. [반현·홍원식 (2009). 국내 지상파 방송 뉴스 포맷 연구: KBS, MBC, SBS 저녁 메인 뉴스를 중심으로. <방송문화연구>, 21권 1호, 9-38.]

Berkowitz, D. (1992). Non-routine news and newswork: Exploring a what-a-story. *Journal of Communication*, *42*(1), 82-94.

Bermejo Nievas, E., Deniz Suarez, O., Bueno García, G., & Sukthankar, R. (2011, August). *Violence detection in video using computer vision techniques*. Paper presented at the 14th International Conference on Computer Analysis of Images and Patterns (CAIP 2011), Seville, Spain.

Brasted, M. (2005). Protest in the media. *Peace Review: A Journal of Social Justice*, *17*(4), 383-388.

Byun, Y. (2016). Describe aspects of conflict issues in the newspaper. *Gyoeraeomoonhak*, *57*, 149-183. [변영수 (2016). 신문의 갈등 이슈 기술 양상 -'촛불집회'의 뉴스 프레임 강조 장치를 중심으로. <겨레어문학>, 57호, 149-183.]

Chan, J. M., & Lee, C. C. (1984). Journalistic paradigms of civil protests: A case study of Hong Kong. *The News Media in National and International Conflict*, 183-202.

Chen, L. H., Hsu, H. W., Wang, L. Y., & Su, C. W. (2011, August). *Violence detection in movies*. Paper

presented at the 2011 Eighth International Conference Computer Graphics, Imaging and Visualization (CGIV 2011), Singapore.

Chen, X., Hsieh, C. J., & Gong, B. (2021). When vision transformers outperform resnets without pre-training or strong data augmentations. *arXiv preprint arXiv:2106.01548*.

Cho, Y., Chung, Y., Yoon, H. Y., Kim, M., Kim, N. Y., Chen, L., ... & Kang J. (2020). Analysis of the 19th presidential TV debate using deep learning based video processing algorithms: Analysis of the frequency, facial expression and gaze. *Korean Journal of Journalism & Communication Studies*, *64*(5), 319-372. [최윤정·정유진·윤호영·김민정·김나영·첸루·신주연·이주희·김나영·여은·강제원 (2020). 딥 러닝(Deep learning)기반 동영상 처리 알고리즘을 통한 19대 대선 TV토론 영상분석: 후보자들의 등장빈도, 표정, 응시방향에 대한 분석. <한국언론학보>, 64권 5호, 319-372.]

Choi, E.-J. (2013). *Video production theory.* Seoul: Communicationbooks. [최이정 (2013). <영상 제작론>. 서울: 커뮤니케이션북스.]

Choi, M.-J. (2005). A study of visual representation paradigm in TV news : Focusing on recognition of TV news cameramen. *Journal of Broadcasting and Telecommunications Research*, *60*, 323-349. [최민재 (2005). TV뉴스의 영상구성에 대한 패러다임 연구: TV카메라기자의 인식을 중심으로. <방송통신연구>, 60호, 323-349.]

Choi, Y. J. (2008). Order and proportion effects of scenes in a broadcasting news story, and a moderating role of image-issue. *Korean Journal of Broadcasting and Telecommunication Studies*, *22*(3), 365-396. [최윤정 (2008). 방송 뉴스에서 신(scene)의 순서효과 및 비중효과 검증과 이미지-이슈의 조절기능에 대한 연구. <한국방송학보>, 22권 3호, 365-396.]

Choi, Y. J., Chung, Y., & Jung, K. H. (2023). Transition in measuring media violence: Automated detection of violent scenes through computer vision. *Studies of the Broadcasting Culture*, *35*(2), 5-59. [최윤정·정유진·정금희 (2023). 미디어 폭력성 측정방식의 전환: 컴퓨터 비전을 통한 자동화된 폭력장면 검출. <방송문화연구>, 35권 2호, 5-59.]

Constantin, M. G., Ştefan, L. D., Ionescu, B., Demarty, C. H., Sjöberg, M., Schedl, M., & Gravier, G. (2020). Affect in multimedia: Benchmarking violent scenes detection. *IEEE Transactions on Affective Computing*, *13*(1), 347-366.

Cummings, D. (2014). The DNA of a television news story: Technological influences on TV news production. *Electronic News, 8*(3), 198-215.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N.

(2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Gamson, R. M., & Modigliani, A. (1989). Media discourse and public opinion on nuclear power: A constructionist approach. *American Journal of Sociology*, *95*(1), 1-37.

Giannakopoulos, T., Makris, A., Kosmopoulos, D., Perantonis, S., & Theodoridis, S. (2010, May). *Audio-visual fusion for detecting violent scenes in videos*. Paper presented at the 6th Hellenic Conference on Artificial Intelligence (SETN 2010), Athens, Greece.

Gitlin, T. (1980). *The whole world is watching.* Berkely, CA: University of California Press.

Gong, Y., Wang, W., Jiang, S., Huang, Q., & Gao, W. (2008, December). *Detecting violent scenes in movies by auditory and visual cues*. Paper presented at the 9th Pacific Rim Conference on Multimedia (PCM 2008), Tainan, Taiwan.

Gruber, J. B. (2023). Troublemakers in the streets? A framing analysis of newspaper coverage of protests in the UK 1992–2017. *The International Journal of Press/Politics*, *28*(2), 414-433.

Ha, J.-W., Noh, J.-D., Yoon, S., Kim, M.-S., & Ahn, C. (2022, January). Finding focused key frames of a given meaning on video data. Paper presented at the Korean Society of Computer Information Conference, Daejeon. [하종우·노정담·윤성웅·김민수·안창원 (2022, 1월). <영상의 특정 의미를 반영하는 Key Frame의 추출 방법>. 한국컴퓨터정보학회 동계학술대회. 대전: 대전창조경제혁신센터.]

Hallin, D. (1986). *The uncensored war: The media and Vietnam.* New York, NY: Oxford University press.

Hallin, D. (1992). Sound bite news: Television coverage of elections, 1968–1988. *Journal of Communication*, *42*(2), 5-24.

Harlow, S., & Bachmann, I. (2023). Police, violence, and the “logic of damage”: Comparing us and chilean media portrayals of protests. *Mass Communication and Society*, *27*(2), 254-277.

Heilman, G. (2023, April 30). *Which countries celebrate International Labor Day on 1 May?* Diario AS. Retrieved 11/20/23 from https://en.as.com/latest_news/which-countries-celebrate-international-labor-day-on-1-may-n/#tooltip_autores

Hong, J.-H., & Na, E.-K. (2015). Victim blaming of Sewal-ferry disaster on news in conservative total TV programming: categorization of victims and word network analysis. *Korean Journal of Journalism & Communication Studies*, *59*(6), 69-106. [홍주현·나은경 (2015). 세월호 사건 보도의 피해자 비난 경향 연구: 보수 종편 채널 뉴스의 피해자 범주화 및 단어 네트워크 프레임 분석. <한국언론학보>, 59권 6호, 69-106.]

Im, Y.-J. (2009). A comparative analysis of news frame of social disputes on the selected TV news: The 2009 Youngsan accident through MBC, KBS, SBS News. *Korean Journal of Journalism & Communication Studies*, *53*(5), 55-79. [임양준 (2009). 집단적 갈등 이슈에 대한 방송뉴스 프레임 비교연구: 용산참사에 대한 MBC, KBS, SBS 저녁뉴스를 중심으로. <한국언론학보>, 53권 5호, 55-79.]

Ionescu, B., Schlüter, J., Mironica, I., & Schedl, M. (2013, April). A naive mid-level concept-based fusion approach to violence detection in Hollywood movies. Paper presented at the International Conference on Multimedia Retrieval (ICMR 2013), Dallas, TX.

Jang, S. H. (1994). *The theory and practice of TV news footage*. Seoul: Kidari. [장석호 (1994). <TV 보도 영상의 이론과 실제>. 서울: 기다리.]

Jang, Y. H. (1988). Social movements and the media: a study on the social construction of social movements by the mass media. *The Korean Journal of Humanities and the Social Sciences*, *11*(4), 37-72. [장용호 (1987). 사회운동과 언론: 대중매체에 의한 사회운동의 사회적 구성에 관한 연구. <현상과 인식>, 41호, 37-72.]

Joo, J., & Steinert-Threlkeld, Z. C. (2022). Image as data: Automated content analysis for visual presentations of political actors and events. *Computational Communication Research*, *4*(1), 11-67.

Kim, H. H., & Lee, J. K. (2009). *Broadcasting news reporting & writing.* Seoul: Namuwasup. [김학희·이재경 (2009). <방송보도>. 서울: 나무와 숲.]

Kim, M. H. (2018). *How to write a TV news story.* Seoul: Communicationbooks. [김문환 (2018). <TV 뉴스 기사 작성법>. 서울: 커뮤니케이션북스.]

Kim, S. H. (2022). *Video and TV journalism*. Seoul: Publius. [김성환 (2022). <영상과 TV 저널리즘>. 서울: 푸블리우스.]

Kim, S.-J. (2003). Visualizing news objectivity: A cmparative case ctudy of environmental television news in the US and Korea. *Korean Journal of Journalism & Communication Studies*, *47*(5), 363-384. [김수정 (2003). 뉴스 객관성의 영상화. <한국언론학보>, 47권 5호, 363-384.]

Lancaster, K. (2013). *Video journalism for the web: A practical introduction to documentary storytelling.* New York, NY: Routledge.

Lee, C. (2012). A study on characteristics, sensationalism and reality representation of CCTV video on TV news. *Broadcasting & Communication*, *13*(4), 5-43. [이창훈 (2012). CCTV영상의 보도 특성과 선정성, 현실 재현에 관한 연구. <방송과 커뮤니케이션>, 13권 4호, 5-43.]

Lee, H. S. (2016). *Korea broadcasting newsroom.* Paju: Nanam. [이화섭 (2016). <한국방송 뉴스룸>. 파주: 나남.]

Lee, H.-Y., & Yun, S.-J. (2013). An analysis of news coverage on conflicts concerning transmission line construction in Miryang - From a perspective of environmental justice. *Economy and Society*, *98*, 40-76. [이화연·윤순진 (2013). 밀양 고압 송전선로 건설 갈등에 대한 일간지 보도 분석: 환경정의 관점에서. <경제와사회>, 98호, 40-76.]

Lee, J. (1999). Visual representation in television news: An analysis of the relationship between visual images and verbal texts. *Korean Journal of Broadcasting and Telecommunication Studies*, *12*, 219-252. [이종수 (1999). 텔레비전 뉴스영상 구성. <한국방송학보>, 12호, 219-252.]

Lee, J., & Choi, Y. (2017). Network analyses of attention to deviance and social significance based on gene and culture co-evolution theory. In C. M. Liebler & T. P. Vos (Eds.), *Media scholarship in a transitional age* (pp. 175-191). New York, NY: Peter Lang.

McLeod, D. M., & Hertog, J. K. (1992). The manufacture of public opinion by reporters: informal cues for public perceptions of protest groups. *Discourse & Society*, *3*(3), 259-275.

Mourão, R. R., Brown, D. K., & Sylvie, G. (2021). Framing Ferguson: The interplay of advocacy and journalistic frames in local and national newspaper coverage of Michael Brown. *Journalism*, *22*(2), 320-340.

Mutikani, L. (2024, February 22). *US labor strikes jump to 23-year high in 2023*. Reuters. Retrieved 11/20/23 from https://www.reuters.com/world/us/us-labor-strikes-jump-23-year-high-2023-2024-02-21/

Newhagen, J. E. (1998). TV news images that induce anger, fear, and disgust : Effects on apprach-avoidance and memory. *Journal of Broadcasting & Electronic Media*, *42*(2), 265-276.

Newhagen, J., & Reeves, B. (1992). The evening's bad news: Effects of compelling negative television news images on memory. *Journal of Communication*, *42*(2), 25-41.

Oh, I. (2022). *Computer vision & deep learning.* Seoul: Hanbit Academy. [오일석 (2022). <컴퓨터 비전과 딥러닝>. 서울: 한빛아카데미.]

Ohman, A. (2000). Fear and anxiety: Evolutionary, cognitive and clinical perspectives. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 573-593). New York, NY: Guilford Press.

Park, D. (2022). A study on the applicability of media videos of deep learning models related to computer vision. *Communication Theories*, *18*(1), 111-154. doi: 10.20879/ct.2022.18.1.111 [박대민 (2022). 미

디어 인공지능: 컴퓨터 비전 관련 딥러닝 모델의 미디어 동영상 분야 적용 가능성에 관한 연구. <커뮤니케이션이론>, 18권 1호, 111-154.]

Park, N., & Kim, S. (2022). *How do vision transformers work? arXiv preprint arXiv:2202.06709.*

Rendón-Segador, F. J., Álvarez-García, J. A., Salazar-González, J. L., & Tommasi, T. (2023). Crimenet: Neural structured learning using vision transformer for violence detection. *Neural Networks*, *161*, 318-329.

Rodriguez, L., & Dimitrova, D. V. (2011). The levels of visual framing. *Journal of Visual Literacy*, *30*(1), 48-65.

Seol, J. (2007). *Fundamentals of broadcast production.* Seoul: Communicationbooks. [설진아 (2007). <방송기획제작의 기초>. 서울: 커뮤니케이션북스.]

Shin, J. G. (2023). May Day 2023: World Federation of Trade Unions (WFTU) statement. *Situations & Labor*, *192*, 115-117. [신재길 (2023). 2023년 노동절: 세계노동조합연맹(WFTU) 성명. <정세와노동>, 192호, 115-117.]

Shoemaker, P. J. (1996). Hardwired for news: Using biological and cultural evolution to explain the surveillance function. *Journal of Communication*, *46*(3), 32-47.

Shoemaker, P. J., & Cohen, A. A. (2006). *News around the world: Content, practitioners, and the public.* New York, NY: Routledge.

Shoemaker, P. J., & Reese, S. D. (1996). *Mediating the message: Theories of influences on mass media content.* White Plains, NY: Longman.

Shoemaker, P. J., Danielian, L. H., & Brendlinger, N. (1991). Deviant acts, risky business and U.S. interests: *The newsworthiness of world events. Journalism Quarterly*, *68*(4), 781-795.

Smirnov, R. (2022, October 3). *Comparing ViT and EfficientNet in terms of image classification problems*. Medium. Retrieved 11/20/23 from https://medium.com/exness-blog/comparing-vit-and-efficientnet-in-terms-of-image-classification-problems-605dfdd843c7

Smith, J., McCarthy, J. D., McPhail, C., & Augustyn, B. (2001). From protest to agenda building: Description bias in media coverage of protest events in Washington, D.C. *Social Forces*, *79*(4), 1397-1423.

Steiner, A., Kolesnikov, A., Zhai, X., Wightman, R., Uszkoreit, J., & Beyer, L. (2021). How to train your vit? Data, augmentation, and regularization in vision transformers. *arXiv preprint arXiv:2106.10270.*

Tuchman, G. (1973). Making news by doing work: Routinizing the unexpected. *American Journal of*

Sociology, 79(1), 110-131.

Wikidocs (2023a). *Deep Learning Bible – 2. Classification U_01. Understanding of vision transformer.* Retrieved from 11/20/23 https://wikidocs.net/164842 [위키독스 (2023a). Deep learning bible – 2. Classification U_01. Understanding of vision transformer.]

Wikidocs (2023b). *Encyclopedia of deep learning computer vision 2.3.4. Vision transformer.* Retrieved 11/20/23 from https://wikidocs.net/137253 [위키독스 (2023b). 한땀한땀 딥러닝 컴퓨터 비전 백과사전 2.3.4. Vision transformer.]

Wischmann, L. (1987). Dying on the front page: Kent State and the Pulitzer Prize. *Journal of Mass Media Ethics*, *2*(2), 67-74.

Yang, J.-H. (2001). Media framing of a social conflict - A case study of medical doctors' strike in Korea. *Korean Journal of Journalism & Communication Studies*, *45*(2), 284-315. [양정혜 (2001). 사회갈등과 의미구성하기. <한국언론학보>, 45권 2호, 248-315.]

Yoon, H. Y. (2021). From human coding to automated detection: Detecting visual images of female body objectification and sexualized poses from TV music programs using YOLO4 and MediaPipe. *Korean Journal of Journalism & Communication Studies*, *65*(6), 452-481. doi: 10.20879/kjjcs.2021.65.6.011 [윤호영 (2021). 사람에서 컴퓨터 자동화로의 연결을 위한 탐색: 객체 인식(Object Detection) 딥러닝 알고리즘 YOLO4, 자세 인식(Pose Detection) 프레임워크 MediaPipe를 활용한 음악 프로그램의 여성 신체 대상화, 선정적 화면 검출 연구. <한국언론학보>, 65권 6호, 452-481.]

Zettl, H. (2016). *Sight, sound, motion: Applied media aesthetics* (8th ed.). Boston, MA: Cengage Learning. 박덕춘 (역) (2016). <영상 제작의 미학적 원리와 방법>. 서울: 커뮤니케이션북스.

Zillmann, D. (2002). Exemplification theory of media influence. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (2nd ed., pp. 19-42). Mahwah, NJ: Lawrence Elbaum Associates.

최초 투고일 2023년 10월 12일
게재 확정일 2024년 03월 29일
논문 수정일 2024년 04월 01일

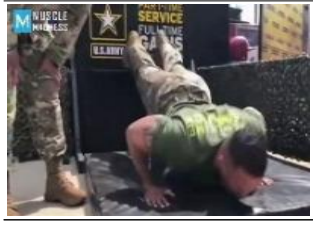
부록

Appendix 1. Examples of Training Data Used for Violence Image Classification in This Study

# 폭력 이미지		
공격적 행동(때리기, 차기, 휘두르기 등)		
		
약탈과 방화		
		
기물 파손		
		
점거와 소동		
		

인물 간 근접 거리 + 고성, 몸싸움, 부정적 표정		
		

 # 비폭력 이미지

질문	악수	포옹
		
대화	인터뷰	표정(긍정적)
		
스포츠		
		
운동	관광	게임
		

공연	공공행사	거래
		
식사	요리	연주
		
경찰서	응급상황	단순 모임