

(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)(11) 공개번호 10-2024-0146499  
(43) 공개일자 2024년10월08일

(51) 국제특허분류(Int. Cl.)

G06V 10/80 (2022.01) G06V 10/774 (2022.01)  
G06V 10/776 (2022.01) G06V 20/40 (2022.01)  
G06V 20/68 (2022.01) G06V 40/16 (2022.01)

(52) CPC특허분류

G06V 10/809 (2023.08)  
G06N 3/08 (2023.01)

(21) 출원번호 10-2023-0041597

(22) 출원일자 2023년03월29일

심사청구일자 2023년03월29일

(71) 출원인

이화여자대학교 산학협력단

서울특별시 서대문구 이화여대길 52 (대현동, 이화여자대학교)

(72) 발명자

정익중

서울특별시 강남구 선릉로126길 22, 111동 301호

강희주

서울특별시 마포구 백범로37길 12, 111동 304호

(74) 대리인

한상수

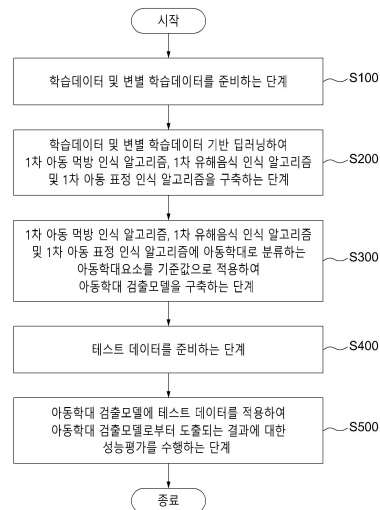
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 디지털 환경 내 아동학대 검출 방법

## (57) 요약

본 발명은 (a) 학습데이터 및 변별 학습데이터를 준비하는 단계, (b) 학습데이터 및 변별 학습데이터 기반 딥러닝하여 1차 아동 먹방 인식 알고리즘, 1차 유해음식 인식 알고리즘 및 1차 아동 표정 인식 알고리즘을 구축하는 단계, (c) 1차 아동 먹방 인식 알고리즘, 1차 유해음식 인식 알고리즘 및 1차 아동 표정 인식 알고리즘에 아동학대로 분류하는 아동학대요소를 기준값으로 적용하여 구축된 2차 아동 먹방 인식 알고리즘, 2차 유해음식 인식 알고리즘 및 2차 아동 표정 인식 알고리즘을 기반으로 아동학대 검출모델을 구축하는 단계, (d) 테스트 데이터를 준비하는 단계 및 (e) 아동학대 검출모델에 테스트 데이터를 적용하여 아동학대 검출모델로부터 도출되는 결과에 대한 성능평가를 수행하는 단계를 포함하는 것을 특징으로 하는 디지털 환경 내 아동학대 검출 방법을 제공한다.

## 대표도 - 도1



(52) CPC특허분류

*G06Q 50/22* (2024.01)  
*G06V 10/774* (2023.08)  
*G06V 10/776* (2023.08)  
*G06V 20/41* (2022.01)  
*G06V 20/49* (2022.01)  
*G06V 20/68* (2023.08)  
*G06V 40/174* (2022.01)  
*G06V 40/178* (2022.01)

---

## 명세서

### 청구범위

#### 청구항 1

- (a) 학습데이터 및 변별 학습데이터를 준비하는 단계;
- (b) 상기 학습데이터 및 상기 변별 학습데이터 기반 딥러닝하여 1차 아동 먹방 인식 알고리즘, 1차 유해음식 인식 알고리즘 및 1차 아동 표정 인식 알고리즘을 구축하는 단계;
- (c) 상기 1차 아동 먹방 인식 알고리즘, 상기 1차 유해음식 인식 알고리즘 및 상기 1차 아동 표정 인식 알고리즘에 아동학대로 분류하는 아동학대요소를 기준값으로 적용하여 아동학대 검출모델을 구축하는 단계;
- (d) 테스트 데이터를 준비하는 단계; 및
- (e) 상기 아동학대 검출모델에 상기 테스트 데이터를 적용하여 상기 아동학대 검출모델로부터 도출되는 결과에 대한 성능평가를 수행하는 단계;를 포함하는 것을 특징으로 하는 디지털 환경 내 아동학대 검출 방법.

#### 청구항 2

제1 항에 있어서,

상기 (a) 단계에서,

상기 학습데이터는 디지털 환경에 노출되는 다수의 동영상 중 아동이 노출되는 아동 얼굴 동영상 및 성인이 노출되는 성인 얼굴 동영상을 포함하는 아동먹방 학습데이터, 상기 다수의 동영상 중 유해음식이 노출되는 유해음식 동영상 및 무해음식이 노출되는 무해음식 동영상을 포함하는 유해음식 학습데이터 및 상기 다수의 동영상 중 아동의 부정적 정서표현이 노출되는 부정적 정서표현 동영상을 포함하는 아동 표정 학습데이터를 포함하고,

상기 변별 학습데이터는 상기 다수의 동영상 중 상기 아동먹방 학습데이터를 제외한 아동먹방 변별 학습데이터, 상기 다수의 동영상 중 상기 유해음식 학습데이터를 제외한 유해음식 변별 학습데이터 및 상기 다수의 동영상 중 상기 아동 표정 학습데이터를 제외한 아동 표정 변별 학습데이터를 포함하는 것을 특징으로 하는 디지털 환경 내 아동학대 검출 방법.

#### 청구항 3

제2 항에 있어서,

상기 (b) 단계는,

- (b1) 상기 아동먹방 학습데이터 및 상기 아동먹방 변별 학습데이터를 딥러닝하여 1차 아동먹방 인식 알고리즘을 구축하는 단계;
- (b2) 상기 유해음식 학습데이터 및 상기 유해음식 변별 학습데이터를 딥러닝하여 1차 유해음식 인식 알고리즘을 구축하는 단계; 및
- (b3) 상기 아동 표정 학습데이터 및 상기 아동 표정 변별 학습데이터를 딥러닝하여 1차 아동 표정 인식 알고리즘을 구축하는 단계;를 포함하는 것을 특징으로 하는 디지털 환경 내 아동학대 검출 방법.

#### 청구항 4

제3 항에 있어서,

상기 (b1) 단계에서,

상기 1차 아동먹방 인식 알고리즘은,

상기 아동 얼굴 동영상 및 상기 성인 얼굴 동영상을 1초당 1장의 이미지로 슬라이싱하여 다수의 아동 얼굴 슬라이싱 이미지 및 다수의 성인 얼굴 슬라이싱 이미지를 수집하는 단계;

상기 다수의 아동 얼굴 슬라이싱 이미지 및 상기 다수의 성인 얼굴 슬라이싱 이미지에서 아동 얼굴 및 성인 얼굴을 각각 검출하는 단계;

상기 아동 얼굴 및 상기 성인 얼굴이 각각 검출된 결과를 기반으로 상기 다수의 아동 얼굴 동영상 및 상기 다수의 성인 얼굴 동영상 내의 얼굴을 각각 아동 및 성인으로 분류하는 단계; 및

상기 다수의 아동 얼굴 동영상 중 유해음식 또는 무해음식을 먹는 아동이 노출되는 아동 얼굴 동영상을 인식하는 단계;를 포함하는 것을 특징으로 하는 디지털 환경 내 아동학대 검출 방법.

## 청구항 5

제3 항에 있어서,

상기 (b1) 단계에서,

상기 1차 아동먹방 인식 알고리즘은 상기 아동먹방 학습데이터 중 아동이 출연하는 비율이 30% 이상인 아동먹방 학습데이터를 추출하고, 상기 추출된 아동먹방 학습데이터에서 아동과 음식이 함께 노출되는 비율이 10% 이상이면 아동학대가 발생하는 동영상으로 분류하는 것을 특징으로 하는 디지털 환경 내 아동학대 검출 방법.

## 청구항 6

제4 항에 있어서,

상기 (b2) 단계에서,

상기 1차 아동먹방 인식 알고리즘이 수행된 이후,

상기 1차 유해음식 인식 알고리즘은 유해음식 또는 무해음식을 먹는 아동이 노출되는 아동 얼굴 동영상 중 상기 유해음식을 먹는 아동이 노출되는 아동 얼굴 동영상을 인식하는 단계;를 포함하는 것을 특징으로 하는 디지털 환경 내 아동학대 검출 방법.

## 청구항 7

제5 항에 있어서,

상기 (b2) 단계에서,

상기 1차 유해음식 인식 알고리즘은 상기 분류된 아동먹방 학습데이터에서 유해음식이 등장하는 비율이 0.5% 이상이면 상기 분류된 동영상을 아동학대가 발생하는 동영상으로 판단하는 것을 특징으로 하는 디지털 환경 내 아동학대 검출 방법.

## 청구항 8

제3 항에 있어서,

상기 아동의 부정적 정서표현은 아동이 우는 얼굴의 표정, 아동이 찡그리는 얼굴의 표정, 아동이 놀라서 눈과 입을 벌리는 얼굴의 표정 중 적어도 어느 하나를 포함하고,

상기 (b3) 단계에서,

상기 1차 아동 표정 인식 알고리즘은,

상기 아동 얼굴 동영상 및 상기 성인 얼굴 동영상을 1초당 1장의 이미지로 슬라이싱하여 다수의 아동 얼굴 슬라이싱 이미지 및 다수의 성인 얼굴 슬라이싱 이미지를 수집하는 단계;

상기 다수의 아동 얼굴 슬라이싱 이미지 및 상기 다수의 성인 얼굴 슬라이싱 이미지에서 아동 얼굴 및 성인 얼굴을 각각 검출하는 단계;

상기 아동 얼굴 및 상기 성인 얼굴이 각각 검출된 결과를 기반으로 상기 다수의 아동 얼굴 동영상 및 상기 다수의 성인 얼굴 동영상 내의 얼굴을 각각 아동 및 성인으로 분류하는 단계;

상기 다수의 아동 얼굴 동영상 중 상기 아동의 부정적 정서표현이 노출되는 아동 얼굴 동영상을 아동학대가 발생하는 동영상으로 판단하는 것을 특징으로 하는 디지털 환경 내 아동학대 검출 방법.

## 청구항 9

제3 항에 있어서,

상기 (b3) 단계에서,

상기 1차 아동 표정 인식 알고리즘은 상기 아동 표정 학습데이터 중 아동이 출연하는 비율이 30% 이상인 아동 표정 학습데이터에서 아동먹방 학습데이터를 제외한 후 상기 아동의 부정적 정서표현이 노출되는 비율이 1% 이상이면 아동학대가 발생하는 동영상으로 판단하여 검출하는 것을 특징으로 하는 디지털 환경 내 아동학대 검출 방법.

## 청구항 10

제1 항에 있어서,

상기 (d) 단계에서,

상기 테스트 데이터는 152개의 기설정된 아동학대 정답데이터 및 구독자수의 상위 6개 채널에서 수집한 482개의 구독자 테스트데이터를 포함하는 것을 특징으로 하는 디지털 환경 내 아동학대 검출 방법.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 디지털 환경 내 아동학대 검출 방법에 관한 것으로, 보다 상세하게는 온라인 동영상 공유 플랫폼에서 노출되는 다수의 동영상 중에서 아동학대가 발생하는 동영상을 검출하는 디지털 환경 내 아동학대 검출 방법에 관한 것이다.

### 배경 기술

[0003] 정보통신기술의 발달과 스마트폰 보급률의 증가는 아동의 놀이 공간을 골목과 놀이터에서 디지털 환경으로 확장시키고 있다. 그중에서 유튜브가 아동에게 미치는 영향력은 날로 증가하고 있는데, 한국언론진흥재단이 발표한 2020 어린이 미디어 이용 조사 보고서에서 아동의 온라인 동영상 플랫폼 이용률은 78.7%이며, 그중에서 유튜브 이용률은 94.8%에 달하는 것으로 나타났다(한국언론진흥재단, 2020).

[0004] 아동의 유튜브 이용률 증가는 단순히 아동을 시청자로 제한하지 않고 아동을 직접 출연시켜 거대한 수익모델을 창출하는 결과로 이어지고 있다(박지혜, 2019). 인기 유튜브 채널 '보람튜브'의 출연아동이 2019년 기준 월 24억 이상의 광고 수익을 벌어들여 강남의 고액 빌딩을 구매한 사실이 알려지면서 아동 콘텐츠로 고수익을 올리려는 '애테크(아이+재테크)'가 유행처럼 번지기도 했다(신연수, 2019.7.26.).

[0005] 그러나 이에 대한 우려도 적지 않다. 유튜브 출연아동 콘텐츠의 높은 조회 수, 구독자 수가 수익과 연결되자 보호자 및 제작자들이 아동을 대상으로 자극적인 동영상을 제작하는 사례가 빈번하게 발생하면서 미디어 속 아동 인권 침해가 심각해지고 있기 때문이다(관계부처합동, 2020; 김지수, 윤석민, 2019; 배상률, 2016).

- [0006] 실제로 아동 출연자에게 몰래 부모의 지갑에서 돈을 훔치게 하거나 차도에서 장난감 자동차를 타게 한 동영상, 아동이 먹기에는 위험한 대왕문어를 먹도록 한 경우 아동학대로 판결된 바 있으며(이현지, 2018.7.30.) 아동이 출연하는 유튜브 동영상 40개 채널 4,690개를 직접 모니터링한 강희주, 정익중의 연구에서(2020) 출연아동에 대한 신체적 학대와 정서적 학대, 방임의 아동학대가 발생하고 있는 동영상은 152개로 전체 영상의 3.24%를 차지하는 것으로 나타났다. 이 수치는 2018년 아동학대 현황 보고서(중앙아동보호전문기관, 2019)에서 보고된 아동학대 발견율인 2.98%보다 높은 수치이다.
- [0007] 유튜브에 업로드한 동영상은 국내를 넘어 전 세계를 대상으로 불특정 다수의 사람에게 순식간에 공유될 수 있다는 것을 충분히 인식하고 있음에도 불구하고 유튜브 내에서 이렇게 높은 아동 학대율이 나타난 것은 우리 사회가 아동에 대해 얼마나 낮은 인식 수준을 가지고 있는지, 아동학대에 대해 얼마나 무감각한지(강희주, 정익중, 2020), 유튜브를 포함한 디지털 환경 내에서 아동을 보호하는 장치가 얼마나 미비한지를 보여준다.
- [0008] 이러한 상황에 대비해 유엔 아동 권리위원회는 일반논평 제25호를 통해 디지털 환경 속에서 아동 권리가 침해되지 않도록 국가의 적극적인 역할을 권고하고 있다(United Nations Committee on the Rights of the Child, 2021).
- [0009] 그러나, 기하급수적으로 증가하고 있는 아동 콘텐츠의 시장의 확장성과 수익성과 비교했을 때 출연아동에 대한 법률적 근거와 보호조치는 상대적으로 매우 미흡한 실정이다. 정부가 「제2차 아동정책기본계획」에 키즈 유튜브 등 '일하는 아동에 대한 권리 보호 강화' 과제를 포함하여 유튜브 출연아동 보호의 필요성을 인식하고(정부 부처 합동, 2020), '인터넷 개인 방송 출연아동 청소년 보호 지침(방송통신위원회, 2020)을 마련하여 출연아동 보호를 위한 대안을 마련하고자 노력하고 있지만, 모든 것이 권고 사항으로만 그쳐있어 부모와 제작자가 이를 지키지 않는 경우 제재할 수가 없다(강희주, 정익중, 2020).
- [0010] 방송통신위원회와 여성가족부에서 운영하는 아동·청소년 유해 영상 모니터링의 경우도 청소년 시청등급(19금) 표시, 청소년 유해 영상에 대한 점검 및 청소년 보호 조치 미이행 사업자 대상 등에 대한 제재에 국한되어 있어 유튜브 등 디지털 플랫폼에 출연하는 아동에게 가해지는 아동학대에 대한 국가의 모니터링과 제재는 실질적으로 이루어지지 못하고 있다. 출연아동 보호를 위한 모니터링 인력이 확보된다고 해도 2017년 기준으로 1분당 400시간 이상 분량의 영상, 다시 말해 하루에 올라오는 유튜브 신규 동영상을 한사람이 하루 종일 시청한다고 가정했을 때 66년이 걸리는 영상 분량을(백승구, 2017) 인간의 힘만으로 검열하기는 거의 불가능에 가깝다(강희주, 정익중, 2020; 백민제, 2021).
- [0011] 오프라인에서 발생하는 아동학대 위기아동 발굴을 위해서는 빅데이터 기반의 'e아동행복지원시스템'이 존재하고, 신고와 현장출동, 응급조치와 임시조치, 피해아동 보호명령 혹은 아동보호사건 송치, 피해아동 및 학대행위자를 포함한 가족 서비스 제공 등 아동학대 업무처리 매뉴얼(법무부, 2016)이 존재하는 반면에, 유튜브를 포함한 디지털 환경 내에서 발생하는 아동학대는 발견 단계조차 별도의 지침이 마련되어 있지 않다. 구글의 기술혁신의 목표가 지금까지 경험한 어떤 것보다 10배 더 위대하고, 더 나으며, 더 빨라야 한다는 철학을 기반으로 실용성을 극대화하고 있는 현실에 비해(Schulz, 2016) 스스로의 힘만으로는 급속한 변화를 온전히 감당해 내기 어려운 아동을 위한(정익중, 오정수, 2021) 디지털 환경 내 아동보호체계 구축은 얼마만큼 준비되어 있는지 자문해야 한다.
- [0012] 2017년 1월에 열린 다보스 세계경제포럼에서는 4차산업혁명이 초래할 다양한 사회적 위험에 대응하기 위한 방안으로 기존 사회안전망의 혁신적인 변화 전략 수립을 위한 인간 중심의 지능정보사회 실현을 강조한 바 있다(최현수, 오미애, 2017). 이는 우리 사회에 깊숙하게 침투한 4차 산업혁명의 변화에 대응해야 하는 사회복지 실천 방법의 지향점과 매우 유사하다고 볼 수 있는데, 인간에 대한 착취를 외면하고 기술 진보만을 기반으로 무한하게 성장한다면(한상진, 2017) 인간은 본연의 가치로부터 소외된 채 건강과 안녕을 위협받아야(Fitzpartirck, 2014, 한상진, 2017, 재인용) 지속 가능한 성장을 불가능하게 만들 뿐만 아니라 인간성의 존속까지도 어렵게 만들 수 있기 때문이다(이소영, 2012).
- [0013] 미래 사회에서의 사회복지 존립 이유가 여전히 인간의 존엄과 가치, 사회정의를 실현하는 것이라면(Zastrow, 2013) 미래의 사회복지 실천은 새로운 과학적 지식과 기술을 혁신적으로 설계하여 인권 사각지대를 선제적으로 예측하여 사회안전망을 공고하게 구축하는 일이 되어야 한다(김미옥, 최혜지, 정익중, 민소영, 2017). 이러한 사회복지학계의 노력은 근 미래에 대다수의 노동이 인공지능(artificial intelligence, AI)으로 대체될 것이라는 우려 속에서도 전문성을 인정받아 그 역할과 기능을 존속할 수 있는 전제조건이 될 것이다(김미옥 외, 2017). 그러나 현재까지 인간의 생명이나 복리를 위해 인공지능을 개발한 연구의 대부분이 의료계에서 이루어지고 있을 뿐(김지율, 2020; 윤명성, 2021; 윤영아, 2022; 정수민, 2020; 정태석, 2021; 조현성, 2019) 미래 사

회복지 실천을 위한 인공지능 개발을 위한 연구는 거의 이루어지지 못하고 있다.

[0014] 이에 따라 디지털 환경, 특히 아동들이 가장 많이 시청하고 있는 동영상 플랫폼인 유튜브에서 발생하는 아동학대 현황을 파악함과 동시에, 디지털 환경 내에서의 아동학대를 신속하게 발견하는 알고리즘을 개발하여 그 성능을 평가하는 기술이 필요한 실정이다.

[0016] (특허문헌 1) 한국 등록특허 제10-1790587호(2017.10.20.)

## 발명의 내용

### 해결하려는 과제

[0018] 상기와 같은 문제를 해결하기 위한 본 발명의 목적은 1차 아동 먹방 인식 알고리즘, 1차 유해음식 인식 알고리즘 및 1차 아동 표정 인식 알고리즘의 정확도가 90% 이상을 보일 때까지 학습데이터 및 변별 학습데이터를 딥러닝하고, 이에 아동학대요소를 기준값으로 적용하여 2차 아동 먹방 인식 알고리즘, 2차 유해음식 인식 알고리즘 및 2차 아동 표정 인식 알고리즘을 구축하여 아동학대 검출모델을 구축하며, 아동학대 검출모델에 테스트 데이터를 적용시켜 알고리즘 개발 전 실시된 내용분석 결과와 아동학대 검출모델로부터 도출되는 혼동행렬 통계값을 비교 및 분석한 유효성 내용평가를 실시함으로써 아동학대 검출모델을 통하여 다수의 동영상 중 아동학대가 발생하는 동영상을 검출하고, 아동학대 검출모델로부터 도출되는 결과의 정확성 및 신뢰성을 향상시킬 수 있는 디지털 환경 내 아동학대 검출 방법을 제공하는 것이다.

[0020] 본 발명이 이루고자 하는 기술적 과제는 이상에서 언급한 기술적 과제로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

### 과제의 해결 수단

[0022] 상기와 같은 목적을 달성하기 위한 본 발명의 구성은 (a) 학습데이터 및 변별 학습데이터를 준비하는 단계; (b) 상기 학습데이터 및 상기 변별 학습데이터 기반 딥러닝하여 1차 아동 먹방 인식 알고리즘, 1차 유해음식 인식 알고리즘 및 1차 아동 표정 인식 알고리즘을 구축하는 단계; (c) 상기 1차 아동 먹방 인식 알고리즘, 상기 1차 유해음식 인식 알고리즘 및 상기 1차 아동 표정 인식 알고리즘에 아동학대로 분류하는 아동학대요소를 기준값으로 적용하여 아동학대 검출모델을 구축하는 단계; (d) 테스트 데이터를 준비하는 단계; 및 (e) 상기 아동학대 검출모델에 상기 테스트 데이터를 적용하여 상기 아동학대 검출모델로부터 도출되는 결과에 대한 성능평가를 수행하는 단계;를 포함하는 것을 특징으로 하는 디지털 환경 내 아동학대 검출 방법을 제공한다.

[0023] 본 발명의 실시예에 있어서, 상기 (a) 단계에서, 상기 학습데이터는 디지털 환경에 노출되는 다수의 동영상 중 아동이 노출되는 아동 얼굴 동영상 및 성인이 노출되는 성인 얼굴 동영상을 포함하는 아동먹방 학습데이터, 상기 다수의 동영상 중 유해음식이 노출되는 유해음식 동영상 및 무해음식이 노출되는 무해음식 동영상을 포함하는 유해음식 학습데이터 및 상기 다수의 동영상 중 아동의 부정적 정서표현이 노출되는 부정적 정서표현 동영상을 포함하는 아동 표정 학습데이터를 포함하고, 상기 변별 학습데이터는 상기 다수의 동영상 중 상기 아동먹방 학습데이터를 제외한 아동먹방 변별 학습데이터, 상기 다수의 동영상 중 상기 유해음식 학습데이터를 제외한 유해음식 변별 학습데이터 및 상기 다수의 동영상 중 상기 아동 표정 학습데이터를 제외한 아동 표정 변별 학습데이터를 포함하는 것을 특징으로 할 수 있다.

[0024] 본 발명의 실시예에 있어서, 상기 (b) 단계는, (b1) 상기 아동먹방 학습데이터 및 상기 아동먹방 변별 학습데이터를 딥러닝하여 1차 아동먹방 인식 알고리즘을 구축하는 단계; (b2) 상기 유해음식 학습데이터 및 상기 유해음식 변별 학습데이터를 딥러닝하여 1차 유해음식 인식 알고리즘을 구축하는 단계; 및 (b3) 상기 아동 표정 학습데이터 및 상기 아동 표정 변별 학습데이터를 딥러닝하여 1차 아동 표정 인식 알고리즘을 구축하는 단계;를 포함하는 것을 특징으로 할 수 있다.

- [0025] 본 발명의 실시예에 있어서, 상기 (b1) 단계에서, 상기 1차 아동먹방 인식 알고리즘은, 상기 아동 얼굴 동영상 및 상기 성인 얼굴 동영상을 1초당 1장의 이미지로 슬라이싱하여 다수의 아동 얼굴 슬라이싱 이미지 및 다수의 성인 얼굴 슬라이싱 이미지를 수집하는 단계; 상기 다수의 아동 얼굴 슬라이싱 이미지 및 상기 다수의 성인 얼굴 슬라이싱 이미지에서 아동 얼굴 및 성인 얼굴을 각각 검출하는 단계; 상기 아동 얼굴 및 상기 성인 얼굴이 각각 검출된 결과를 기반으로 상기 다수의 아동 얼굴 동영상 및 상기 다수의 성인 얼굴 동영상 내의 얼굴을 각각 아동 및 성인으로 분류하는 단계; 및 상기 다수의 아동 얼굴 동영상 중 유해음식 또는 무해음식을 먹는 아동이 노출되는 아동 얼굴 동영상을 인식하는 단계;를 포함하는 것을 특징으로 할 수 있다.
- [0026] 본 발명의 실시예에 있어서, 상기 (b1) 단계에서, 상기 1차 아동먹방 인식 알고리즘은 상기 아동먹방 학습데이터 중 아동이 출연하는 비율이 30% 이상인 아동먹방 학습데이터를 추출하고, 상기 추출된 아동먹방 학습데이터에서 아동과 음식이 함께 노출되는 비율이 10% 이상이면 아동학대가 발생하는 동영상으로 분류하는 것을 특징으로 할 수 있다.
- [0027] 본 발명의 실시예에 있어서, 상기 (b2) 단계에서, 상기 1차 아동먹방 인식 알고리즘이 수행된 이후, 상기 1차 유해음식 인식 알고리즘은 유해음식 또는 무해음식을 먹는 아동이 노출되는 아동 얼굴 동영상 중 상기 유해음식을 먹는 아동이 노출되는 아동 얼굴 동영상을 인식하는 단계;를 포함하는 것을 특징으로 할 수 있다.
- [0028] 본 발명의 실시예에 있어서, 상기 (b2) 단계에서, 상기 1차 유해음식 인식 알고리즘은 상기 분류된 아동먹방 학습데이터에서 유해음식이 등장하는 비율이 0.5% 이상이면 상기 분류된 동영상을 아동학대가 발생하는 동영상으로 판단하는 것을 특징으로 할 수 있다.
- [0029] 본 발명의 실시예에 있어서, 상기 아동의 부정적 정서표현은 아동이 우는 얼굴의 표정, 아동이 찡그리는 얼굴의 표정, 아동이 놀라서 눈과 입을 벌리는 얼굴의 표정 중 적어도 어느 하나를 포함하고, 상기 (b3) 단계에서, 상기 1차 아동 표정 인식 알고리즘은, 상기 아동 얼굴 동영상 및 상기 성인 얼굴 동영상을 1초당 1장의 이미지로 슬라이싱하여 다수의 아동 얼굴 슬라이싱 이미지 및 다수의 성인 얼굴 슬라이싱 이미지를 수집하는 단계; 상기 다수의 아동 얼굴 슬라이싱 이미지 및 상기 다수의 성인 얼굴 슬라이싱 이미지에서 아동 얼굴 및 성인 얼굴을 각각 검출하는 단계; 상기 아동 얼굴 및 상기 성인 얼굴이 각각 검출된 결과를 기반으로 상기 다수의 아동 얼굴 동영상 및 상기 다수의 성인 얼굴 동영상 내의 얼굴을 각각 아동 및 성인으로 분류하는 단계; 상기 다수의 아동 얼굴 동영상 중 상기 아동의 부정적 정서표현이 노출되는 아동 얼굴 동영상을 아동학대가 발생하는 동영상으로 판단하는 것을 특징으로 할 수 있다.
- [0030] 본 발명의 실시예에 있어서, 상기 (b3) 단계에서, 상기 1차 아동 표정 인식 알고리즘은 상기 아동 표정 학습데이터 중 아동이 출연하는 비율이 30% 이상인 아동 표정 학습데이터에서 아동먹방 학습데이터를 제외한 후 상기 아동의 부정적 정서표현이 노출되는 비율이 1% 이상이면 아동학대가 발생하는 동영상으로 판단하여 검출하는 것을 특징으로 할 수 있다.
- [0031] 본 발명의 실시예에 있어서, 상기 (d) 단계에서, 상기 테스트 데이터는 152개의 기설정된 아동학대 정답데이터 및 구독자수의 상위 6개 채널에서 수집한 482개의 구독자 테스트데이터를 포함하는 것을 특징으로 할 수 있다.

### 발명의 효과

- [0033] 상기와 같은 구성에 따르는 본 발명의 효과는, 1차 아동 먹방 인식 알고리즘, 1차 유해음식 인식 알고리즘 및 1차 아동 표정 인식 알고리즘의 정확도가 90% 이상을 보일 때까지 학습데이터 및 변별 학습데이터를 딥러닝하고, 이에 아동학대요소를 기준값으로 적용하여 2차 아동 먹방 인식 알고리즘, 2차 유해음식 인식 알고리즘 및 2차 아동 표정 인식 알고리즘을 구축하여 아동학대 검출모델을 구축하며, 아동학대 검출모델에 테스트 데이터를 적용시켜 알고리즘 개발 전 실시된 내용분석 결과와 아동학대 검출모델로부터 도출되는 혼동행렬 통계값을 비교 및 분석한 유효성 내용평가를 실시함으로써 아동학대 검출모델을 통하여 다수의 동영상 중 아동학대가 발생하는 동영상을 검출하고, 아동학대 검출모델로부터 도출되는 결과의 정확성 및 신뢰성을 향상시킬 수 있다.
- [0035] 본 발명의 효과는 상기한 효과로 한정되는 것은 아니며, 본 발명의 상세한 설명 또는 특허청구범위에 기재된 발명의 구성으로부터 추론 가능한 모든 효과를 포함하는 것으로 이해되어야 한다.

## 도면의 간단한 설명

[0037]

- 도 1은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법을 나타낸 순서도이다.
- 도 2는 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동먹방 학습데이터를 수집하는 과정을 나타낸 도면이다.
- 도 3은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동먹방 변별 학습데이터를 수집하는 과정을 나타낸 도면이다.
- 도 4는 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동먹방 인식 알고리즘 및 유해음식 인식 알고리즘을 나타낸 도면이다.
- 도 5는 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동 표정 인식 알고리즘을 나타낸 도면이다.
- 도 6은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 훈련셋(train set)을 이용한 딥러닝을 통하여 아동먹방 인식 알고리즘, 유해음식 인식 알고리즘 및 아동 표정 인식 알고리즘의 정확도를 향상시키기 위한 데이터셋을 분리 및 구분한 것을 나타낸 도면이다.
- 도 7은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 딥러닝을 위하여 학습데이터 중 일 반음식이 노출되는 동영상상을 슬라이싱한 것을 나타낸 도면이다.
- 도 8은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 딥러닝을 위하여 학습데이터 중 유 해음식이 노출되는 동영상상을 슬라이싱한 것을 나타낸 도면이다.
- 도 9는 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 YOLOv5(pytorch) 데이터셋 변환 과 정을 나타낸 도면이다.
- 도 10은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동 얼굴 및 성인 얼굴을 탐지하 는 것을 나타낸 도면이다.
- 도 11은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동 얼굴 및 성인 얼굴의 탐지를 통해 아동학대 발생 가능성이 큰 먹방장면을 분류하는 과정을 나타낸 도면이다.
- 도 12는 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동 얼굴, 성인 얼굴 탐지 및 먹 방장면 분류 결과를 나타낸 도면이다.
- 도 13은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동 표정을 분류하는 개발 과정을 나타낸 도면이다.
- 도 14는 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동/성인 구분 및 아동 표정 분류 결과를 나타낸 도면이다.
- 도 15는 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 훈련셋(train set)을 이용한 1차 성능검사 결과를 나타낸 도면이다.
- 도 16은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 훈련셋(train set)을 이용하여 아 동 및 성인 구분에 대한 2차 성능검사 결과를 나타낸 도면이다.
- 도 17은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 훈련셋(train set)을 이용하여 무 해음식 및 유해음식 구분에 대한 2차 성능검사 결과를 나타낸 도면이다.
- 도 18은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 훈련셋(train set)을 이용하여 아 동의 부정적 정서표현 및 그 밖의 정서표현 구분에 대한 구분에 대한 2차 성능검사 결과를 나타낸 도면이다.

## 발명을 실시하기 위한 구체적인 내용

[0038]

이하에서는 첨부한 도면을 참조하여 본 발명을 설명하기로 한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 따라서 여기에서 설명하는 실시예로 한정되는 것은 아니다. 그리고 도면에서 본 발명을 명확 하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유

사한 도면 부호를 붙였다.

- [0039] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결(접속, 접촉, 결합)"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 부재를 사이에 두고 "간접적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 구비할 수 있다는 것을 의미한다.
- [0040] 본 명세서에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0041] 이하 첨부된 도면을 참고하여 본 발명의 실시예를 상세히 설명하기로 한다.
- [0043] 도 1은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법을 나타낸 순서도이다.
- [0044] 도 1을 참조하면, 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법은 (a) 학습데이터 및 변별 학습데이터를 준비하는 단계(S100), (b) 학습데이터 및 변별 학습데이터 기반 딥러닝하여 1차 아동 먹방 인식 알고리즘, 1차 유해음식 인식 알고리즘 및 1차 아동 표정 인식 알고리즘을 구축하는 단계(S200), (c) 1차 아동 먹방 인식 알고리즘, 1차 유해음식 인식 알고리즘 및 1차 아동 표정 인식 알고리즘에 아동학대로 분류하는 아동학대요소를 기준값으로 적용하여 아동학대 검출모델을 구축하는 단계(S300), (d) 테스트 데이터를 준비하는 단계(S400) 및 (e) 아동학대 검출모델에 테스트 데이터를 적용하여 아동학대 검출모델로부터 도출되는 결과에 대한 성능평가를 수행하는 단계(S500)를 포함한다.
- [0045] 도 2는 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동먹방 학습데이터를 수집하는 과정을 나타낸 도면이다. 도 3은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동먹방 변별 학습데이터를 수집하는 과정을 나타낸 도면이다.
- [0046] 도 2 및 도 3을 참조하면, 상기 (a) 단계에서, 학습데이터는 디지털 환경에 노출되는 다수의 동영상 중 아동이 노출되는 아동 얼굴 동영상 및 성인이 노출되는 성인 얼굴 동영상을 포함하는 아동먹방 학습데이터, 다수의 동영상 중 유해음식이 노출되는 유해음식 동영상 및 무해음식이 노출되는 무해음식 동영상을 포함하는 유해음식 학습데이터 및 다수의 동영상 중 아동의 부정적 정서표현이 노출되는 부정적 정서표현 동영상을 포함하는 아동 표정 학습데이터를 포함한다.
- [0047] 구체적으로 학습데이터는 '유튜브 출연아동에게 발생하는 아동학대 장면'을 구체적으로 정의하기 위해 '유튜브 아동학대 내용분석 결과'를 참조하여(강희주, 정익중, 2020) 아동학대가 발생한 152개의 영상을 재모니터링하면서 먹방과 상황설정, 일상, 몰카 콘텐츠에서 발생한 아동학대 영상들이 공통적으로 지닌 이미지 특성을 기반으로 한다.
- [0048] 디지털 환경 내 먹방 콘텐츠에서는 출연아동에게 혐오스러울 수 있는 음식인 산낙지와 문어, 성인이 먹기에도 매운 떡볶이 등 자극적인 음식을 장시간 섭취하게 하고, 아동이 먹기를 고통스러워 해도 음식섭취를 중단시키지 않고 조회 수와 구독자 수를 얻기 위해 촬영을 지속시키는 상황에서 아동학대가 발생하였다(강희주, 정익중, 2020).
- [0049] 이에 따라 상기한 내용을 근거로 아동학대 상황에서의 이미지 특성을 분석했을 때 아동학대가 발생하는 먹방 동영상에서 출연아동은 성인이 세팅해 놓은 식탁에 가만히 앉아서 먹기만 하는 공통적인 모습을 보였다(강희주 외, 2021).
- [0050] 따라서, '아동이 카메라 정면에 앉아 음식을 섭취함과 동시에, 아동의 음식섭취 영상이 전체 영상 시간의 상당한 시간을 차지할 만큼 아동에게 장시간 음식섭취 시키는 영상'을 아동학대가 발생할 가능성이 높다고 정의하였으며, 이와 관련된 다수의 동영상을 아동먹방 학습데이터(training data)로 선정하였다.
- [0051] 또한, 아동먹방 학습데이터의 이미지 특징은 '아동과 음식이 함께 있음(아동+음식)', '아동 앞에 식탁 혹은 도마가 있음(아동+식탁, 아동+도마)', '아동 앞에 음식, 식탁, 도마가 함께 있는 영상이 전체 영상에서 상당한 비중을 차지함'으로 정의하였다.

- [0052] 이에 따른 아동먹방 학습데이터는 도 2에 도시된 수집 과정을 통해 수집된다.
- [0053] 또한, 유해음식 학습데이터는 성인이 섭취하기에도 매운맛을 가진 특정 브랜드의 떡볶이와 볶음면, 김치 등이 노출되는 동영상, 아동과 테이블(도마) 위에 올려진 음식 이미지에 '유해음식'이 함께 노출되는 동영상을 포함한다.
- [0054] 또한, 무해음식 학습데이터는 성인이 섭취하기에도 매운맛을 가진 특정 브랜드의 떡볶이와 볶음면, 김치 등을 제외한 일반적인 음식이 노출되는 동영상을 포함한다.
- [0055] 또한, 아동 표정 학습데이터는 몰카, 상황설정, 일상 콘텐츠가 노출되는 동영상 중 아동의 부정적 정서표현이 노출되는 동영상을 포함한다.
- [0056] 여기서, 아동의 부정적 정서표현은 아동이 우는 얼굴의 표정, 아동이 찡그리는 얼굴의 표정, 아동이 놀라서 눈과 입을 벌리는 얼굴의 표정 중 적어도 어느 하나를 포함한다.
- [0057] 한편, 학습데이터와 구별하기 위한 변별 학습데이터는 다수의 동영상 중 아동먹방 학습데이터를 제외한 아동먹방 변별 학습데이터, 다수의 동영상 중 유해음식 학습데이터를 제외한 유해음식 변별 학습데이터 및 다수의 동영상 중 아동 표정 학습데이터를 제외한 아동 표정 학습데이터를 포함한다.
- [0058] 구체적으로 아동먹방 변별 학습데이터는 다수의 동영상 중 '성인이 음식을 섭취하는 먹방 이미지', '아동이 음식이 아닌 놀잇감을 카메라 정면에 두고 놀이하는 이미지', '아동 없이 음식만 있는 이미지', '음식 없이 아동만 있는 이미지' 등을 포함하며, 이러한 아동먹방 변별 학습데이터를 수집하는 과정이 도 3에 도시된다.
- [0059] 추가적으로 아동먹방 변별 학습데이터는 아동의 출연 없이 제작된 요리 콘텐츠, 식품 광고 목적의 음식 콘텐츠에도 음식이 등장하므로 아동학대 영상과의 변별도를 높이기 위해 '아동과 음식이 따로 나오는 장면', '음식이 조리되는 과정을 보여주는 이미지', '아동 없이 용기 안에 음식만 들어있는 이미지'의 경우도 아동학대가 발생하지 않는 동영상을 포함할 수 있다.
- [0060] 또한, 유해음식 변별 학습데이터는 유해하지 않는 음식이 노출되는 동영상을 포함한다.
- [0061] 또한, 아동 표정 변별 학습데이터는 긍정적 정서표현이 노출되는 동영상을 포함한다.
- [0062] 다음, 상기 (b) 단계는, (b1) 아동먹방 학습데이터 및 아동먹방 변별 학습데이터를 딥러닝하여 1차 아동먹방 인식 알고리즘을 구축하는 단계, (b2) 유해음식 학습데이터 및 유해음식 변별 학습데이터를 딥러닝하여 1차 유해음식 인식 알고리즘을 구축하는 단계 및 (b3) 아동 표정 학습데이터 및 아동 표정 변별 학습데이터를 딥러닝하여 1차 아동 표정 인식 알고리즘을 구축하는 단계를 포함한다.
- [0063] 도 4는 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동먹방 인식 알고리즘 및 유해음식 인식 알고리즘을 나타낸 도면이다. 도 5는 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동 표정 인식 알고리즘을 나타낸 도면이다. 도 6은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 훈련셋(train set)을 이용한 딥러닝을 통하여 아동먹방 인식 알고리즘, 유해음식 인식 알고리즘 및 아동 표정 인식 알고리즘의 정확도를 향상시키기 위한 데이터셋을 분리 및 구분한 것을 나타낸 도면이다. 도 7은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 딥러닝을 위하여 학습데이터 중 일반음식이 노출되는 동영상을 슬라이싱한 것을 나타낸 도면이다. 도 8은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 딥러닝을 위하여 학습데이터 중 유해음식이 노출되는 동영상을 슬라이싱한 것을 나타낸 도면이다. 도 9는 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 YoloV5(pytorch) 데이터셋 변환 과정을 나타낸 도면이다. 도 10은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동 얼굴 및 성인 얼굴을 탐지하는 것을 나타낸 도면이다. 도 11은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동 얼굴 및 성인 얼굴의 탐지를 통해 아동학대 발생 가능성이 큰 먹방장면을 분류하는 과정을 나타낸 도면이다. 도 12는 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동 얼굴, 성인 얼굴 탐지 및 먹방장면 분류 결과를 나타낸 도면이다. 도 13은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동 표정을 분류하는 개발 과정을 나타낸 도면이다. 도 14는 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 아동/성인 구분 및 아동 표정 분류 결과를 나타낸 도면이다.
- [0064] 구체적으로 상기 (b1) 단계에서, 1차 아동먹방 인식 알고리즘은, 도 4에 도시된 바와 같이 아동 얼굴 동영상 및 성인 얼굴 동영상을 1초당 1장의 이미지로 슬라이싱하여 다수의 아동 얼굴 슬라이싱 이미지 및 다수의 성인 얼굴 슬라이싱 이미지를 수집하는 단계(도 7, 도 8 참조), 다수의 아동 얼굴 슬라이싱 이미지 및 다수의 성인 얼굴

굴 슬라이싱 이미지에서 아동 얼굴 및 성인 얼굴을 각각 검출하는 단계(pre-trained 모델 활용), 아동 얼굴 및 성인 얼굴이 각각 검출된 결과를 기반으로 다수의 아동 얼굴 동영상 및 다수의 성인 얼굴 동영상 내의 얼굴을 각각 아동 및 성인으로 분류하는 단계(classification 모델 개발)(도 10 내지 도 12 참조) 및 다수의 아동 얼굴 동영상 중 유해음식 또는 무해음식을 먹는 아동이 노출되는 아동 얼굴 동영상을 인식하는 단계(yolov5 모델 개발)(도 11, 도 12 참조)를 포함한다.

[0065] 또한, 상기 (b1) 단계에서, 1차 아동먹방 인식 알고리즘은 아동먹방 학습데이터 중 아동이 출연하는 비율이 30% 이상인 아동먹방 학습데이터를 추출하고, 상기 추출된 아동먹방 학습데이터에서 아동과 음식이 함께 노출되는 비율이 10% 이상이면 아동학대가 발생하는 동영상으로 분류한다.

[0066] 다음, 도 4를 참조하면, 상기 (b2) 단계에서, 1차 아동먹방 인식 알고리즘이 수행된 이후, 1차 유해음식 인식 알고리즘은 유해음식 또는 무해음식을 먹는 아동이 노출되는 아동 얼굴 동영상 중 유해음식을 먹는 아동이 노출되는 아동 얼굴 동영상을 인식하는 단계(도 11, 도 12 참조)를 포함한다.

[0067] 또한, 상기 (b2) 단계에서, 1차 유해음식 인식 알고리즘은 분류된 아동먹방 학습데이터에서 유해음식이 등장하는 비율이 0.5% 이상이면 상기 분류된 동영상을 아동학대가 발생하는 동영상으로 판단한다.

[0068] 다음, 도 5를 참조하면, 상기 (b3) 단계에서, 1차 아동 표정 인식 알고리즘은, 아동 얼굴 동영상 및 성인 얼굴 동영상을 1초당 1장의 이미지로 슬라이싱하여 다수의 아동 얼굴 슬라이싱 이미지 및 다수의 성인 얼굴 슬라이싱 이미지를 수집하는 단계(도 10 참조), 다수의 아동 얼굴 슬라이싱 이미지 및 다수의 성인 얼굴 슬라이싱 이미지에서 아동 얼굴 및 성인 얼굴을 각각 검출하는 단계(pre-trained 모델 활용), 아동 얼굴 및 성인 얼굴이 각각 검출된 결과를 기반으로 다수의 아동 얼굴 동영상 및 다수의 성인 얼굴 동영상 내의 얼굴을 각각 아동 및 성인으로 분류하는 단계(classification 모델 개발)(도 13, 도 14 참조) 및 다수의 아동 얼굴 동영상 중 아동의 부정적 정서표현이 노출되는 아동 얼굴 동영상을 아동학대가 발생하는 동영상으로 판단하는 단계(도 14 참조)를 포함한다.

[0069] 또한, 상기 (b3) 단계에서, 1차 아동 표정 인식 알고리즘은 아동 표정 학습데이터 중 아동이 출연하는 비율이 30% 이상인 아동 표정 학습데이터에서 아동먹방 학습데이터를 제외한 후 아동의 부정적 정서표현이 노출되는 비율이 1% 이상이면 아동학대가 발생하는 동영상으로 판단하여 검출한다.

[0070] 다음, 상기 (c) 단계에서는 1차 아동 먹방 인식 알고리즘, 1차 유해음식 인식 알고리즘 및 1차 아동 표정 인식 알고리즘에 아동학대로 분류하는 아동학대요소를 기준값으로 적용하여 구축된 2차 아동 먹방 인식 알고리즘, 2차 유해음식 인식 알고리즘 및 2차 아동 표정 인식 알고리즘을 기반으로 아동학대 검출모델을 구축한다.

[0071] 이를 위해 '유튜브 아동학대 검출 알고리즘(YouTube Child Abuse Detection Algorithm)' 에 아동학대가 발생한 150개(22)의 영상에 대한 예비테스트를 세 차례 실시하였다.

[0072] child, child+food, child+spicy\_disgusting, child+negative facial expression에 각각 0.3, 0.1, 0.005, 0.01의 비율을 적용해 본 결과, 아동학대가 발생한 33개의 정답 값을 '비학대'로 인식하는 것을 확인하였다.

[0073] 아동학대가 발생한 정답 값 33개 영상에 대한 아동학대 검색이 실패한 원인을 확인하기 위해 33개의 동영상에 대한 재모니터링을 실시한 결과, 출연아동에게 아동의 선호 혹은 발달상황과 맞지 않는 고가의 물건을 판매하게 하거나 이벤트를 진행하게 하여 아동에게 어른의 경제적 선호를 노골적으로 드러나게 시키는 영상, 슬라이드를 하면서 일방적으로 친구를 비난하고 무시하는 언어표현을 하게 하는 영상, 위험한 챌린지 놀이를 시키면서 보호자가 적극적인 안전조치를 취하지 않아 아동의 안전보다는 촬영을 우선하는 영상, 자정이 넘은 시간 촬영을 한 영상, 아동에게 상처가 될 수 있는 악플을 아동 앞에서 노골적으로 읽음으로써 정서적인 상처를 주는 영상, 아동에게 성인과 같이 화장이나 네일케어를 하도록 유도하여 아동을 성인과 의 경계를 모호하게 하여 성적 대상화 하는 등의 영상을 아동학대 영상으로 인식하지 못하는 것으로 확인되었다.

[0074] 3세 아동에게 열차려를 시키는 영상의 경우, 아동의 얼굴이 바닥을 향하고 있어 검출되지 못하였다. 이와 같이 150개 정답 값 중에서 비학대로 인정한 33개의 동영상 모두 애초에 딥러닝이 불가능한 '맥락상 벌어진 학대'로 확인되어 'child', 'child+food', 'child+spicy\_disgusting', 'child+negative facial expression'에 각각 설정한 비율인 0.3, 0.1, 0.005, 0.01 값을 최종 알고리즘 기준값으로 설정하는 것이 가장 적합한 것으로 판단되었고, 이에 대한 내용은 [표 1]과 같다.

표 1

[0075]

|                                   |                                     |             |
|-----------------------------------|-------------------------------------|-------------|
| 기준값                               | child 비율                            | 0.3(30%)    |
|                                   | Child+food 비율                       | 0.1(10%)    |
|                                   | Child+spicy_disgusting 비율           | 0.005(0.5%) |
|                                   | Child+negative facial expression 비율 | 0.01(1%)    |
| 아동학대 검색 실패 영상 33개 / 아동학대 정답값 150개 |                                     |             |

[0076]

다음, 상기 (d) 단계에서, 테스트 데이터는 152개의 기설정된 아동학대 정답데이터 및 구독자수의 상위 6개 채널에서 수집한 482개의 구독자 테스트데이터를 포함한다.

[0077]

본 발명에서는 테스트 데이터를 상기 (d) 단계에서 준비하는 것으로 설명하였으나, 1차 아동먹방 인식 알고리즘, 1차 유해음식 인식 알고리즘 및 1차 아동 표정 인식 알고리즘이 구축되기 전에 수행될 수도 있다.

[0078]

위와 같은 테스트 데이터는 1차 아동먹방 인식 알고리즘, 1차 유해음식 인식 알고리즘 및 1차 아동 표정 인식 알고리즘이 구축되기 전에 내용분석에 이용될 수 있다.

[0079]

여기서, 내용분석은 내용분석은 문헌이나 영상 내용의 특성을 체계적으로 기술하면서 현상과 원인, 결과, 영향을 체계적으로 추론하는 분석법으로 방대한 자료를 요약, 분류, 분석하는 데 유용하며, 1차 자료에 대한 접근성이 낮아 조사자가 기록 문서 이외의 증거를 사용할 수 없을 때 활용된다.

[0080]

또한, 자료의 양이 조사자 개인의 조사 역량을 초과할 때 유용하다(Holsti, 1969; 이상호, 2010에서 재인용). 알고리즘 개발 전, 테스트데이터 634개의 동영상에 대한 내용분석은 다음과 같이 진행되었다.

[0081]

첫째, 강희주, 정익중(2020)이 아동보호 전문기관의 아동학대 판정 기준이 되는 아동복지법 제3조 제7호와 세이브더칠드런(2019)에서 제시한 '아이가 행복한 유튜브 촬영 가이드라인'을 참조하여 만든 '유튜브 출연아동 아동학대 내용분석 결과'를 참조하여 테스트 데이터 634개의 영상에 대한 내용분석을 실시하였다.

[0082]

둘째, 동영상은 먹방, 일상, 상황설정, 몰카, 이벤트, 놀이, 요리 콘텐츠로 분류(23)되었으며, 해당 영상에서 아동학대가 발생하면 1로 코딩하여 비정형화되어 있는 동영상을 양적인 데이터로 변환하였다. 강희주, 정익중(2020) 연구의 내용분석 기준과 결과에 포함되지 않는 새로운 아동학대나 권리침해 상황이 발생한 경우 채널별, 콘텐츠 유형별로 유목화하여 상황과 맥락, 출연자들의 행동과 말, 그 이유를 구체적으로 전사하였다.

[0083]

셋째, 출연아동 보호를 위한 댓글 제한 정책을 준수하지 않아 아동의 신체나 외모를 비하하는 댓글, 욕설 등이 출연아동에게 노출되어 정서적 학대가 발생하는 경우, 댓글과 출연아동의 대댓글 등을 그대로 전사하였으며, 관련 내용이 출연 아동에게 어떤 부정적인 영향을 미칠 수 있는지 분석하여 기술하였다.

[0084]

넷째, 아동학대 및 권리침해 여부가 모호한 영상의 경우 사회복지학과 교수와 최종적으로 의견을 일치시키는 과정을 거쳐 신뢰도를 최종적으로 1로 맞추었다.

[0085]

위와 같은 내용분석된 결과는 상기 (f) 단계에 사용된다.

[0087]

나아가, 아동출연 유튜브 아동학대 객체 검출 모델을 만들고 성능을 평가하기 위하여 도 6에 도시된 바와 같이 데이터셋(data set)을 분리하여 훈련 데이터셋(train data set)에 대한 자체 성능 검사를 실시할 수 있으며, 데이터셋(data set)의 분리 및 과정은 다음과 같다.

[0088]

예측이 필요한 테스트데이터를 모델에 적용시키기 위해서는 우선 모델을 학습시킬 데이터가 필요한데, 이때 사용되는 데이터가 훈련 데이터(train data)이며 학습된 모델을 가지고 최종 결과 예측에 쓰이는 데이터는 테스트 데이터(test data)이다. 그러나 훈련 데이터(train data)로 모델을 학습시키고 곧바로 테스트 데이터(test data)를 예측하게 되면 예측값이 과적합(Overfitting)된 값인지 아닌지 알 수 없으므로 훈련 데이터(train data)에서 일부 제공한 데이터(data)에 학습된 모델을 중간 검증한다. 여기서 사용되는 데이터가 검증 데이터(validation data)이다.

[0089]

도 6을 참조하면, 데이터셋(data set)의 분리 과정은 다음과 같다.

[0090]

1 단계 : 훈련 데이터(train data)와 테스트 데이터(test data)로 분리한다.

- [0091] 2 단계 : 훈련 데이터(train data)를 훈련(train)과 검증(validation)으로 분리한다.
- [0092] 3 단계 : 훈련(train) 데이터로 모델을 만들고 밸리테이션(validation)으로 검증한다.
- [0093] 상기한 단계를 만족 시 해당 모델을 훈련(train)과 밸리테이션(validation) 데이터를 합쳐서 학습을 시킨 후 테스트 데이터(test data)를 넣어 예측값을 확인한다.
- [0094] 본 발명에서는 훈련 데이터셋(train data set)으로 성인먹방 영상 37개, 아동먹방 동영상 124개, 아동 출연 동영상 129개, 아동의 부정적 정서표현이 들어있는 동영상 52개, 총 342개의 동영상을 수집하였다. 훈련 데이터셋(train data set)을 위해 수집한 영상들은 아동학대가 발생한 152개의 정답값 동영상과 중복되지 않게 수집하였다. 1초당 한 개의 이미지를 슬라이싱한 결과, 아동 얼굴 이미지는 3,234장, 성인 얼굴은 5,985장, 표정 구분을 위한 아동의 부정적 정서표현 얼굴 이미지는 2,339장, 그 외 표정 이미지는 5,750장을 수집하였다. 일반음식은 4,803장, 유해음식은 5,750장을 수집하였다. 훈련 데이터(train data)는 슬라이싱한 이미지들의 카테고리에서 train:validation=7:3 비율로 나누었다.
- [0095] 도 15는 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 훈련셋(train set)을 이용한 1차 성능검사 결과를 나타낸 도면이다.
- [0096] 훈련셋(train set) 자체 성능검사를 두 차례 실시하였다. 1차 자체 성능검사는 2차 자체 성능검사서 모아진 성인먹방 37개, 아동먹방 124개, 아동출연 영상 129개, 아동의 부정적 정서가 담긴 표정 영상 52개 총 342개의 영상이 모두 수집되기 전에 실시하여 슬라이싱된 이미지 개수가 2차 자체 성능검사서 모아진 데이터보다 적었다.
- [0097] 본 발명에서 주석(annotation) 데이터의 수집은 자체 개발한 주석 툴(annotation tool)을 사용하였으며, 딥러닝 학습을 위한 시스템은 네개의 Tesla K80 24GB (NVIDIA, CA, USA) 그래픽 처리 장치, Xeon E5-260v5 (Intel, CA, USA) CPU 및 128GB RAM으로 구성하였다. 딥러닝 학습은 Ubuntu 18.04 운영체제에서 Python3.8.13 및 Pytorch1.9.1. 프레임 워크를 통해 수행되었으며, 데이터에 대한 영상처리는 OpenCV 4.5.5 라이브러리를 사용하였다.
- [0098] 도 15를 참조하면, 1차 자체 성능검사 결과, 아동을 제대로 인식한 'child'의 경우 94.4%, 성인 얼굴 'grownup'은 94.3%의 정확도를 보였다. 아동 얼굴과 성인 얼굴을 구분할 수 없는 'not\_defined'정확도는 72.6%였다. 아동의 부정적 정서를 상징하는 표정 데이터인 'negative facial expression'은 62.5%로 상대적으로 낮은 정확도를 보였으며, 표정 구분이 어려운 데이터인 'not\_defined'는 0%를 보였다. 1차 검증의 경우 수집되고 학습된 동영상의 수가 적어 accuracy 값이 낮게 나온 것으로 추측된다.
- [0099] 도 16은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 훈련셋(train set)을 이용하여 아동 및 성인 구분에 대한 2차 성능검사 결과를 나타낸 도면이다.
- [0100] 도 16을 참조하면, 1차 자체 성능검사 이후 학습데이터에 신규 이미지를 지속 추가한 결과, 두 번째 성능검사에서는 단위 분류모델인 '아동/성인', '일반/유해음식', '부정적 정서표정/그외 정서표정'에 대해 수집한 데이터 모두 90% 이상의 정확도를 보였다.
- [0101] 구체적으로 훈련셋(train data)에서 아동 얼굴인 'child'를 제대로 인식한 경우는 총 3,234장 중 3,056장으로 94.5%의 정확도를 보였다. 성인 얼굴 5,985장 중 성인으로 제대로 인식한 'grownup'은 5,806장으로 97.0%의 정확도를 보였으며, 아동과 성인을 구분하기 어려운 'not\_defined'는 전체 947장 중 759장으로 80.1%의 정확도를 보였다.
- [0102] 검증(validation) 데이터에서 아동 얼굴 데이터 741장 중 아동을 'child'로 제대로 인식한 경우는 719장으로 97.1%의 정확도를 보였고, 성인 데이터 1,470장 중 'grownup'으로 인식한 경우는 1,434장으로 97.6%의 정확도를 보였다.
- [0103] 'not\_defined'는 전체 200장 중에서 172장을 제대로 인식하여 86.0%의 정확도를 보였다.
- [0104] 확인 결과, 아동/성인 구분에서 'not\_defined'로 분류된 데이터들은 아동의 얼굴 전체가 카메라에 잡히지 않았거나 얼굴을 흔들어 이목구비가 명확하지 않은 경우, 고개를 숙인 경우, 아동과 유사한 얼굴을 가진 애니메이션이 등장하는 경우로 분류되었다.
- [0105] 도 17은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 훈련셋(train set)을 이용하여 무

해음식 및 유해음식 구분에 대한 2차 성능검사 결과를 나타낸 도면이다.

- [0106] 도 17을 참조하면, 일반/유해음식 분류지도 학습 결과, train data에서 일반음식 4,802장 중 일반음식 'general'을 제대로 분류한 경우는 4,576장으로 95.3%의 정확도를 보였으며, 유해음식 'spicy\_disgusting'을 인식한 경우는 전체 2,793장 중에서 2,747장으로 98.4%의 정확도를 보였다. 음식이 아닌 것들인 'no food'의 경우 2,293장 중에서 2,234장을 인식해 97.4%의 정확도를 보였다.
- [0107] 검증(validation) 데이터에서 일반음식 1,155장을 'general'로 제대로 인식한 경우는 1,103장으로 95.5%의 정확도를 보였고, 유해음식 696장 중 'spicy\_disgusting'으로 잘 인식한 경우는 671장으로 96.4%의 정확도를 보였다. 'no food'는 전체 565장 중에서 521장을 인식하여 92.2%의 정확도를 보였다. 확인 결과 'no food'로 인식된 경우는 장난감을 포함한 주변 사물들이었다.
- [0108] 도 18은 본 발명의 일실시예에 따른 디지털 환경 내 아동학대 검출 방법에서 훈련셋(train set)을 이용하여 아동의 부정적 정서표현 및 그 밖의 정서표현 구분에 대한 구분에 대한 2차 성능검사 결과를 나타낸 도면이다.
- [0109] 도 18을 참조하면, 아동 표정 train data 2,339장 중에서 'negative facial expression'을 제대로 인식한 경우는 2,246장으로 96.0%의 정확도를 보였으며, 아동의 울음, 찡그림이 없는 표정 사진인 'etc' 데이터를 제대로 인식한 경우는 5,750장 중 5,280장으로 91.8%의 정확도를 보였다. 표정 인식이 불가능한 'not\_defined'는 396장에서 302장을 인식해 76.3%의 정확도를 보였다.
- [0110] 검증데이터(validation data)는 537장의 'negative facial expression' 데이터 중에서 505장을 인식해 94%의 정확도를 보였고, 1,263장의 'etc' 데이터를 잘 인식한 경우는 1,217장으로 96.4%의 정확도를 보였다. 'not\_defined'로 분류한 경우는 전체 92장 중에서 59장으로 64.1%였다. 아동 표정 구분에서 not\_defined로 분류된 데이터들은 아래와 같이 아동의 얼굴 전체가 나오지 않거나 얼굴을 흔들어 이목구비가 명확하지 않은 경우, 옆을 보고 있거나 고개를 숙인 경우, 아동이 아닌 아동과 유사한 얼굴을 가진 애니메이션이 등장하는 경우로 분류되었다.
- [0112] 상기한 훈련셋(train set) 자체검증 2차 결과에 따라 영상을 분석한 결과는 다음과 같다.
- [0113] 훈련(train), 밸리데이션(validation) 테스트에서 '성인먹방'의 평균 아동 출연 비율은 각각 2.8%, 3.0%였지만 '아동 출연'에서 아동의 비율은 41.6%, 48.8%, '아동 먹방'에서 아동 출연 비율은 85.7%, 80.8% 수준을 보였다.
- [0114] '아동 출연' 영상20)에서 음식 비율은 train, validation 테스트에서 2.4%, 1.0%인 반면, '아동 먹방'에서 아동과 음식이 함께 등장하는 비율은 65.1%, 40.3%를 보여 아동/성인 얼굴 분류, 아동+음식의 분류지도 학습이 적절히 이루어진 것을 확인하였다. '아동 먹방'에서 음식이 등장하는 비율이 '아동출연' 영상보다 더 빈번하지만, 음식 이외에 등장하는 아동 주변 배경을 알고리즘이 자주 인식하고, 등장하는 음식의 양이 적거나 사탕, 레몬 빵 등 아동이 먹는 음식의 크기 자체가 작은 경우, 알고리즘이 이를 인식하지 못하는 경우가 있었다. 이로 인해 '아동 먹방'에서 알고리즘이 음식을 인식하는 비율이 기대치보다 낮게 나오는 것으로 추정되었다.
- [0115] '아동 먹방' 영상에서 부정적 정서표현(negative facial expression)이 등장하는 평균 비율이 훈련(train), 밸리데이션(validation) 테스트에서 각각 0.9%, 3.3%이고, '아동 출연' 영상에 등장하는 부정적 정서표현(negative facial expression) 평균 비율이 1.2%, 1.4%인데 반해, 일부러 아동을 울리거나 화나게 하는 몰카에서 발생하는 '아동학대'21) 영상의 경우 평균 비율이 27%, 17.5%로 나타나 아동 표정에 대한 학습도 적절히 이루어진 것으로 확인되었다.
- [0116] 훈련(train)/밸리데이션 데이터(validation data) 자체검증 결과에 따른 성인먹방과 아동출연 영상, 아동먹방, 아동학대 영상의 최소값과 최대값, 평균값은 아래의 [표 2] 및 [표 3]과 같다.

표 2

| 아동학대요소           |                            | 최소값   | 최대값   | 평균    |
|------------------|----------------------------|-------|-------|-------|
| 성<br>인<br>먹<br>방 | child                      | 0.000 | 0.756 | 0.028 |
|                  | Negative facial expression | 0.000 | 0.000 | 0.000 |
|                  | child+food                 | 0.000 | 0.630 | 0.020 |
|                  | child+spicy_disgusting     | 0.012 | 0.012 | 0.003 |

|                  |                            |       |       |       |
|------------------|----------------------------|-------|-------|-------|
| 아<br>동<br>출<br>연 | child                      | 0.000 | 1.000 | 0.416 |
|                  | Negative facial expression | 0.000 | 0.219 | 0.012 |
|                  | child+food                 | 0.000 | 0.844 | 0.024 |
|                  | child+spicy_disgusting     | 0.000 | 0.011 | 0.000 |
| 아<br>동<br>먹<br>방 | child                      | 0.257 | 1.00  | 0.857 |
|                  | Negative facial expression | 0.000 | 0.066 | 0.009 |
|                  | child+food                 | 0.000 | 0.998 | 0.651 |
|                  | child+spicy_disgusting     | 0.000 | 0.988 | 0.171 |
| 아<br>동<br>학<br>대 | child                      | 0.114 | 1.000 | 0.748 |
|                  | Negative facial expression | 0.000 | 1.000 | 0.270 |
|                  | child+food                 | 0.000 | 0.003 | 0.000 |
|                  | child+spicy_disgusting     | 0.000 | 0.000 | 0.000 |

표 3

|                  |                            |       |       |       |
|------------------|----------------------------|-------|-------|-------|
| 아동학대요소           |                            | 최소값   | 최대값   | 평균    |
| 성<br>인<br>먹<br>방 | child                      | 0.004 | 0.092 | 0.030 |
|                  | Negative facial expression | 0.012 | 0.012 | 0.002 |
|                  | child+food                 | 0.002 | 0.028 | 0.012 |
|                  | child+spicy_disgusting     | 0.000 | 0.008 | 0.003 |
| 아<br>동<br>출<br>연 | child                      | 0.150 | 0.990 | 0.488 |
|                  | Negative facial expression | 0.000 | 0.144 | 0.014 |
|                  | child+food                 | 0.000 | 0.100 | 0.010 |
|                  | child+spicy_disgusting     | 0.000 | 0.008 | 0.001 |
| 아<br>동<br>먹<br>방 | child                      | 0.456 | 0.986 | 0.808 |
|                  | Negative facial expression | 0.001 | 0.280 | 0.033 |
|                  | child+food                 | 0.000 | 0.962 | 0.403 |
|                  | child+spicy_disgusting     | 0.000 | 0.527 | 0.086 |
| 아<br>동<br>학<br>대 | child                      | 0.281 | 1.000 | 0.648 |
|                  | Negative facial expression | 0.000 | 0.550 | 0.175 |
|                  | child+food                 | 0.000 | 0.000 | 0.000 |
|                  | child+spicy_disgusting     | 0.000 | 0.000 | 0.000 |

마지막으로 상기 (f) 단계에서는 아동학대 검출모델에 테스트 데이터를 적용하여 아동학대 검출모델로부터 도출되는 결과에 대한 성능평가를 수행한다.

이를 위해 상기 (f) 단계에서는 앞서 진행한 알고리즘 개발 전 내용분석 결과와 알고리즘 유효성 평가에 활용되는 혼동행렬(Confusion Matrix) 통계값을 비교·분석하면서 진행되며, 이와 관련된 혼동행렬(Confusion Matrix)은 [표 4]와 같다.

표 4

|                  |                                        |                                               |                                               |
|------------------|----------------------------------------|-----------------------------------------------|-----------------------------------------------|
|                  |                                        | 진단결과                                          |                                               |
| 실<br>제<br>상<br>태 |                                        | ○<br>(Positive)                               | X<br>(Negative)                               |
|                  | ○<br>(Positive)<br>실제로 아동학대가 발생한<br>영상 | True Positive(TP)<br>진양성<br>;아동학대(0)-아동학대(0)  | False Negative(FN)<br>위음성<br>;아동학대(0)-아동학대(X) |
|                  | X(Negative)<br>실제로 아동학대가 아닌<br>영상      | False Positive(FP)<br>위양성<br>;아동학대(X)-아동학대(0) | True Negative(TN)<br>진음성<br>;아동학대(X)-아동학대(X)  |

아동학대 유효성 평가를 위한 혼동행렬은 알고리즘의 예측이 실제 아동학대 동영상과 일치하는지 측정하는 True

positives(TP), 아동학대가 아닌 동영상을 아동학대라고 예측하여 측정하는 False Positives(FP), 실제 아동학대 동영상인데 아닌 것으로 예측한 결과인 False negative(FN)를 통계학적으로 분석한다.

- [0123] 여기서 민감도(Sensitivity)는  $TP/(TP+FN)$  계산 값으로 실제 아동학대가 발생한 동영상을 아동학대가 발생하였다고 잘 판단한 비율(True Positive Rate)을 의미한다.
- [0124] 특이도(Specificity)는  $TN/(TN+FP)$  계산 값으로 아동학대가 발생하지 않은 동영상을 아동학대가 발생하지 않았다고(True Negative Rate) 잘 판단한 비율을 말한다.
- [0125] 정밀도(Precision)는  $TP/(TP+FP)$  값으로 아동학대 영상이 있다고 판단한 영상 중 실제 아동학대가 발생한 영상이 차지하는 비율이다. 재현율(Recall)은 민감도(Sensitivity)와 동일한  $TP/(TP+FN)$  값으로 실제 아동학대가 발생한 동영상을 아동학대가 발생했다고 잘 판단한 비율(True Positive Rate)이다.
- [0126] 정확도(Accuracy)는  $(TP+TN)/ALL$  값으로 아동학대 여부에 따른 전체 케이스에서 정답만을 정확하게 맞춘 비율로 범주의 분포가 균형을 이룰 때 효과적인 평가지표이며, F1 score은 정확도(Precision)과 재현율(Recall)의 조화평균으로 정확도와 재현율에 동일한 가중치를 부여하여 내는 평균값으로 정확도와 재현율이 어느 한쪽으로 치우치지 않는 수치를 나타낼 때 상대적으로 높은 값을 가진다.
- [0127] 알고리즘 개발 전에 테스트데이터 634개를 대상으로 실시한 내용분석 결과와 알고리즘 개발 후의 혼동행렬 통계값을 비교·분석한 유효성 내용평가 절차는 다음과 같다.
- [0128] 첫째, 실제로는 아동학대가 발생했음에도 알고리즘이 이를 잘 인지하지 못한 영상을 찾아 그 원인을 분석하였다. (실제 아동학대○ - 알고리즘 인식 아동학대×).
- [0129] 실제 아동학대가 발생했는데 알고리즘이 비학대 영상으로 분류한 경우, '아동학대 영상 분류 폴더'에 들어가 알고리즘이 아동 얼굴과 음식/유해음식 분류, 아동의 표정 분류를 내용분석과 동일하게 했는지 직접 확인하였다. 잘못 분류한 이미지가 발견되면 '아동 얼굴 인식 오류(아동의 부정적 정서표정 얼굴인데 알고리즘이 인지하지 못한 경우)', '유해음식 인식 오류(유해음식인데 일반음식으로 분류한 경우)'로 명명하여 관련 이미지들을 따로 정리하였으며, 향후 알고리즘의 민감도를 높이기 위해 추가적인 학습이 필요한 경우도 함께 분석하여 정리하였다.
- [0130] 이외에도 민감도(Sensitivity)가 가장 높게 혹은 낮게 나온 콘텐츠 종류를 확인하고, 동일하게 '아동학대 영상 분류 폴더'에 들어가 알고리즘의 분류 과정을 확인하여 해당 콘텐츠의 민감도를 높이기 위한 보완 사항을 정리하였다. 아동학대가 발생했으나 알고리즘에게 학습시킨 음식이나 아동 표정 등의 구분만으로는 아동학대 여부를 분류할 수 없는 영상이 발견되면 그 이유도 분석하여 정리하였다.
- [0131] 둘째, 알고리즘이 아동학대가 발생한 영상이라고 분류했으나 실제로 아동학대가 발생하지 않은 영상을 찾아 그 원인을 분석하였다.(알고리즘 인식 아동학대○- 실제 아동학대×). '아동학대 영상 분류 폴더'에 들어가 아동학대가 발생하지 않았는데 아동학대 영상이라고 분류한 영상들의 아동얼굴, 음식/유해음식, 아동의 표정 분류 내용을 확인하였다. 알고리즘이 아동학대가 발생할 가능성이 높은 이미지대로 잘 분류했다면, 성능평가 이전의 내용분석 내용을 참조하여 왜 해당 영상들이 아동학대 요소를 가지고 있음에도 비학대 영상인지 분석하여 정리하였다.
- [0132] 일반음식인데 유해음식으로, 아동의 얼굴이 웃는 표정 혹은 웃으면서 찡그리는 표정인데 'negative facial expression'으로 잘못 인식해 아동학대 영상으로 분류한 경우에는 '아동 얼굴 인식 오류(부정적 정서 표정 얼굴이 아닌데 negative facial expression으로 인정한 경우)', '유해음식 인식 오류(일반음식인데 유해음식으로 분류한 경우)'로 명명하여 관련 이미지들을 따로 정리하였다.
- [0133] 추가로 특이도가 가장 낮게 나온 콘텐츠 종류를 확인하고, 해당 콘텐츠에서 아동학대가 발생하지 않았는데 아동학대로 분류된 공통적인 이유가 있는지 확인하여 향후 알고리즘의 특이도를 높이기 위해 필요한 보완 사항을 정리하였다.
- [0134] 셋째, 실제로 아동학대가 발생하였고, 알고리즘이 동일하게 아동학대로 인식한 동영상을 비교·분석하였다.(실제 아동학대○ - 알고리즘 인식 아동학대 ○) 결과값이 동일함에도 비교·분석한 이유는 알고리즘과 전문가가 동일하게 아동학대로 분류했더라도 아동학대 인지 및 분류 과정이 어떻게 동일하고 다른지를 이해해야 향후 민감도가 높은 알고리즘에 대해서도 관리감독이 가능하기 때문이다. 민감도가 가장 높은 콘텐츠를 중심으로 성능평가 전 내용분석 결과와 알고리즘의 분류과정, 즉, 아동얼굴, 음식/유해음식, 아동의 표정 분류 내용을 확인하

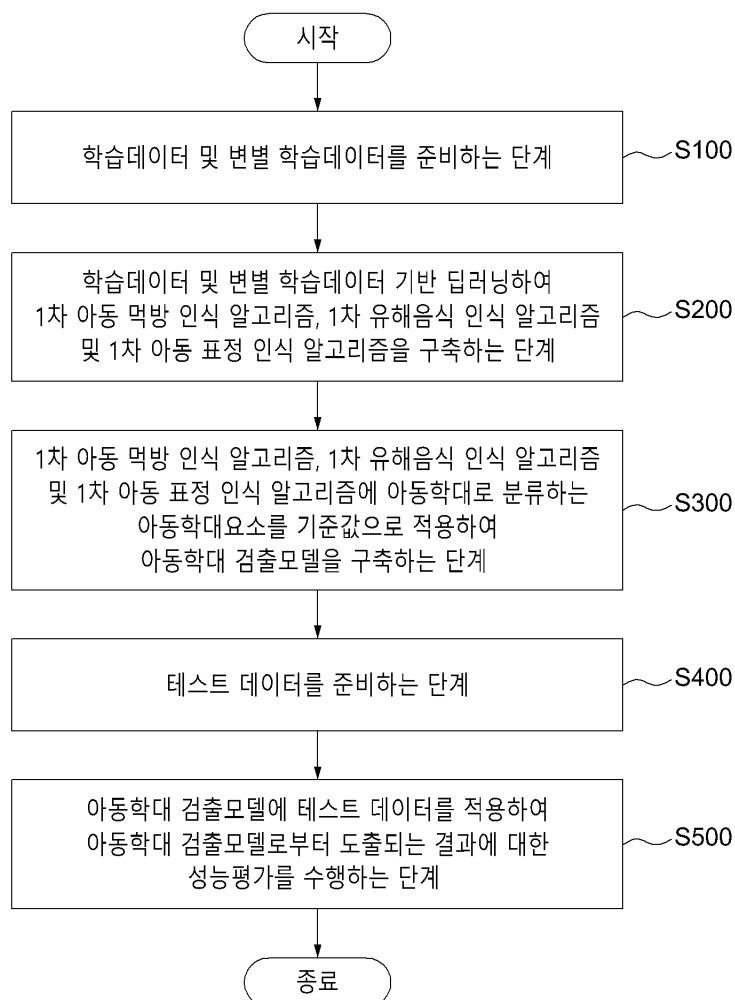
였다.

- [0136] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.
- [0137] 본 발명의 범위는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

## 부호의 설명

### 도면

#### 도면1



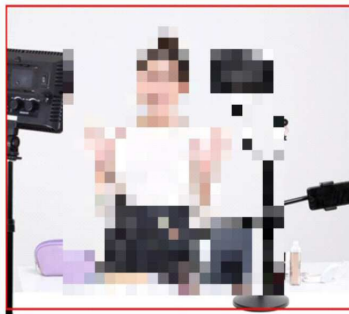
## 도면2



개별 음식이 아니라, 사람 앞에 펼쳐진  
음식영역 기준으로 인식



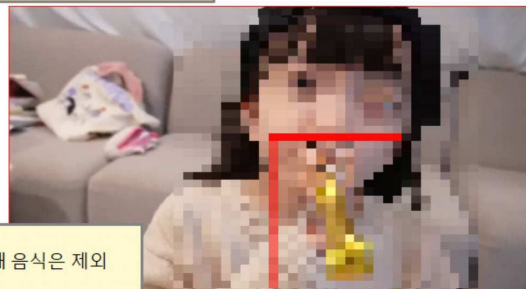
## 도면3



집이나 차안에서 하나씩 들고 먹는  
음식은 제외



마트나 편의점 진열대 음식은 제외



도면4



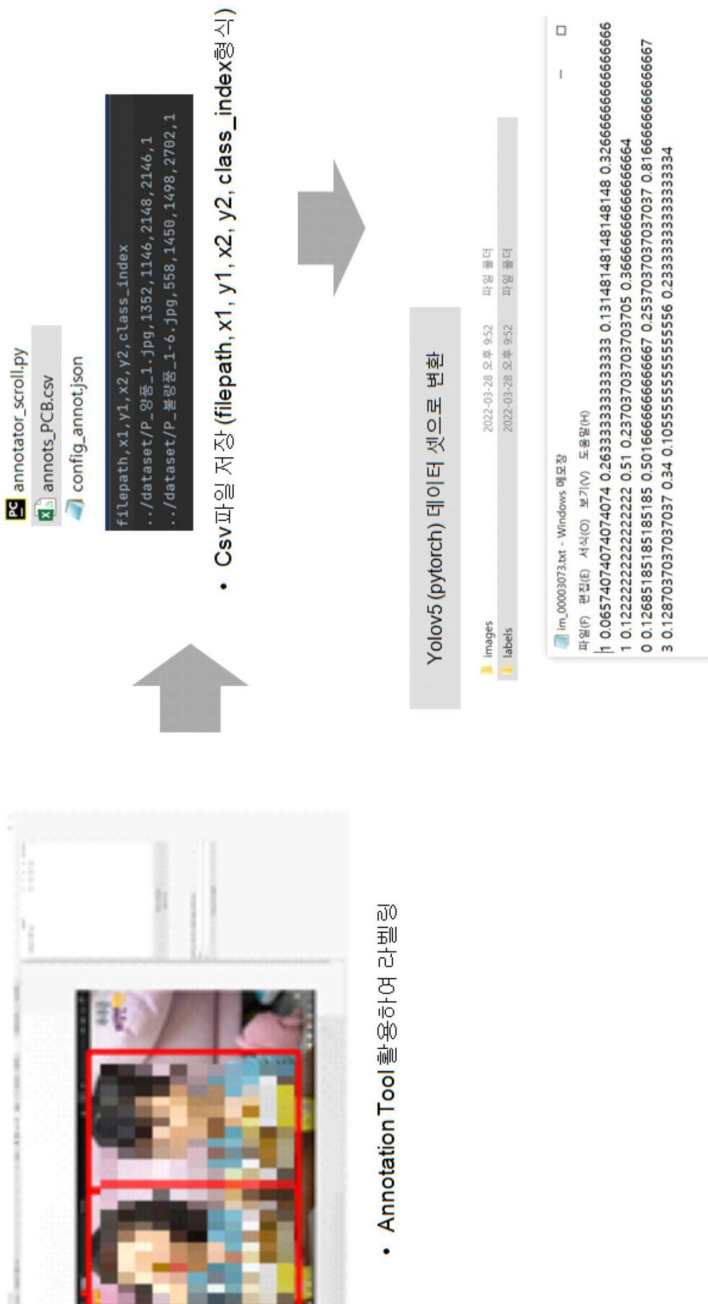
도면5



도면6



도면7



도면8

아동 / 성인 구분



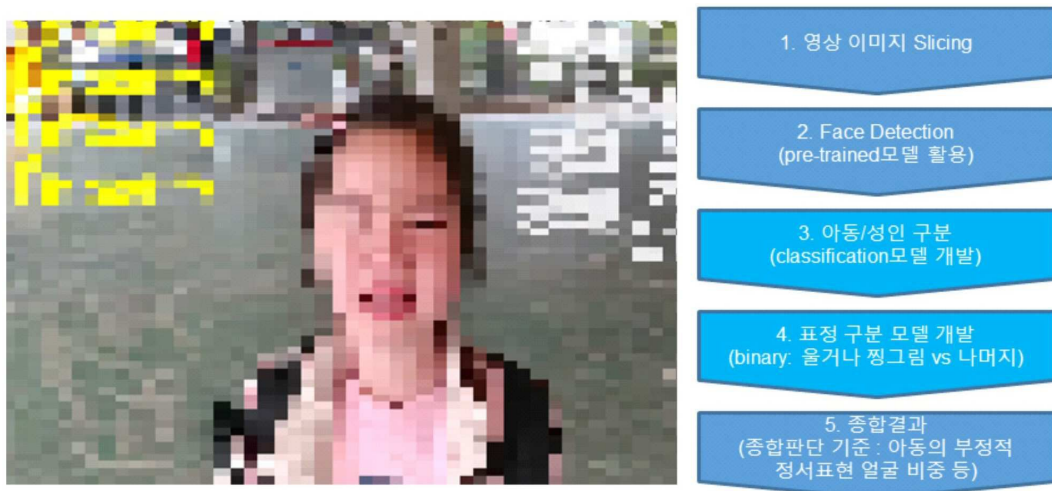
도면9



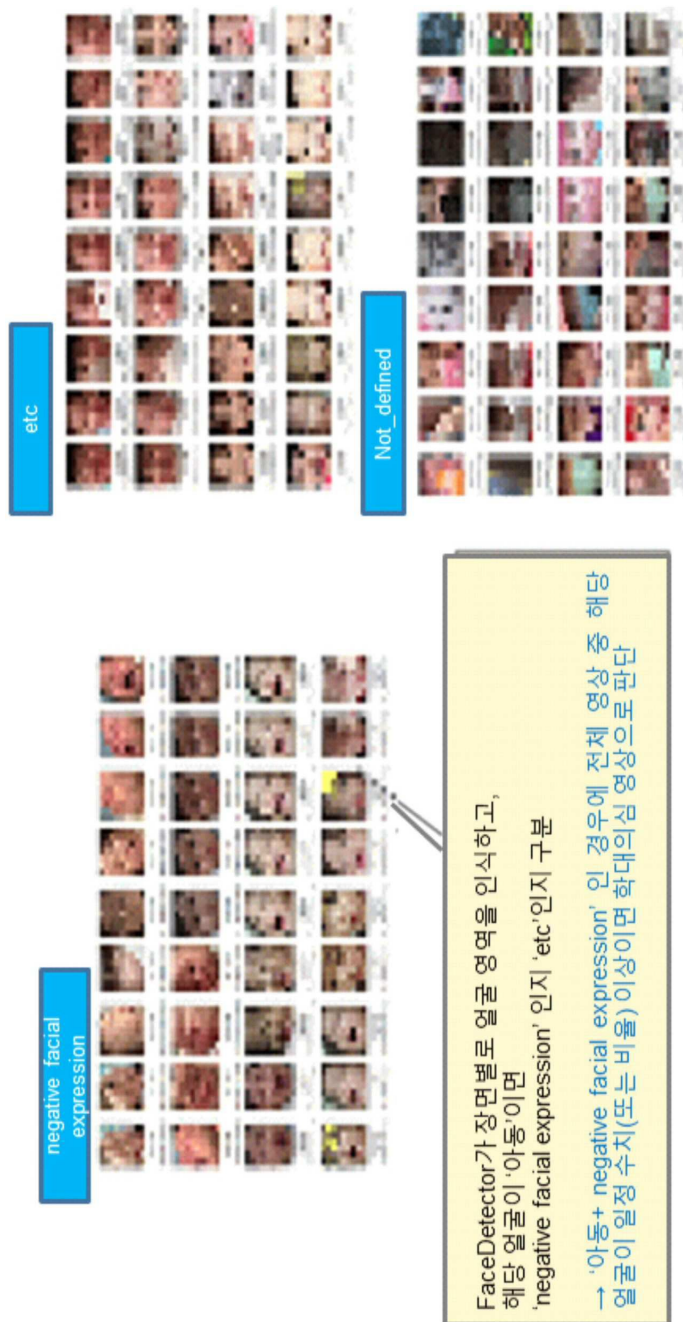
도면10



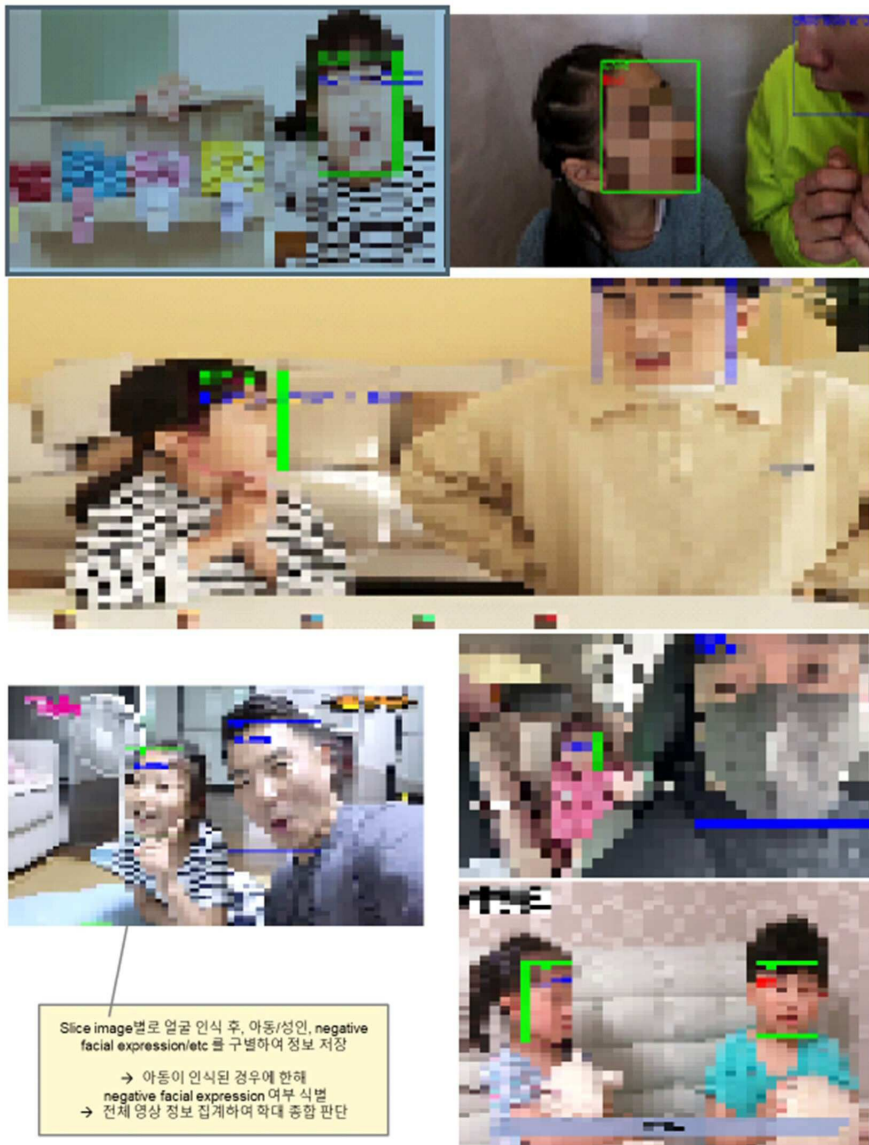
도면11



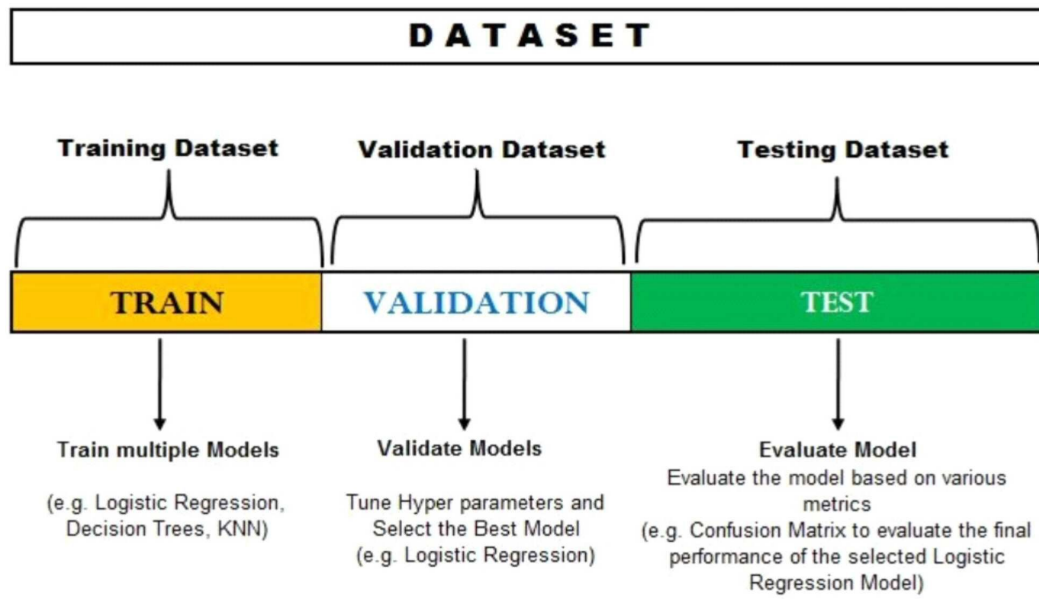
도면12



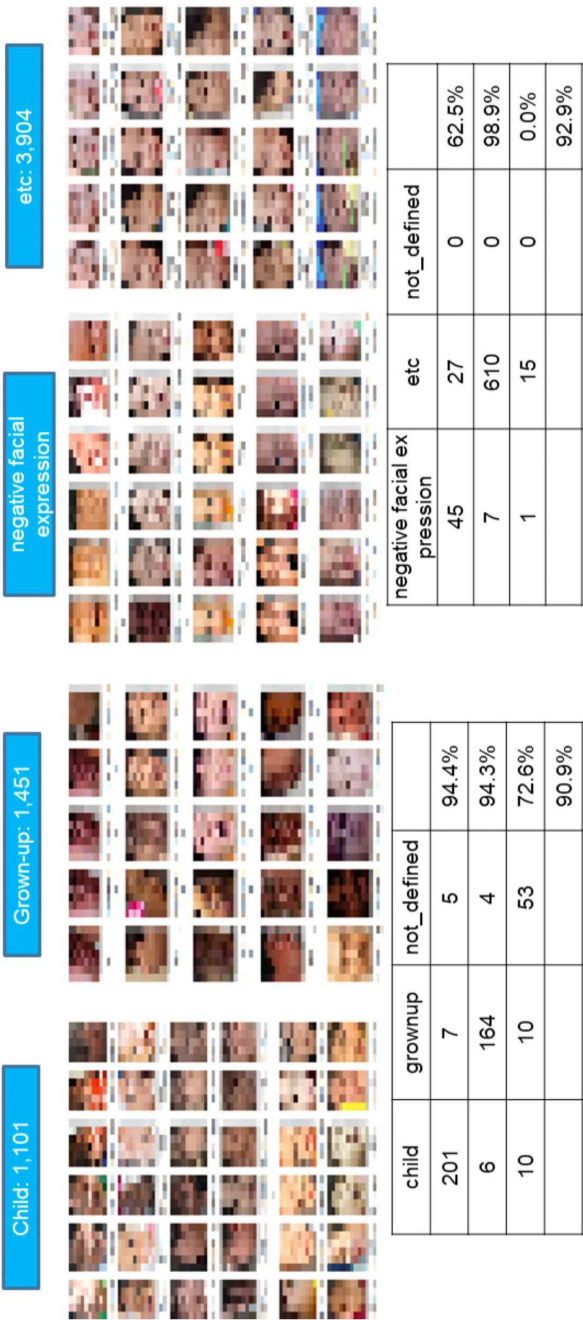
도면13



도면14



도면15



도면16



도면17



도면18

