

상업용 즉석 촬영기기에서의 음란 콘텐츠 촬영 방지를 위한 실시간 이미지 분석 모델 개발

송휘[†], 강경태^{*}

[†]한양대학교 인공지능융합학과

^{*}한양대학교 인공지능학과

e-mail: {[†]songhweel, ^{*}ktkang}@hanyang.ac.kr

Development of a real-time image analysis model to prevent filming of obscene content in commercial instant shooting devices

Hwee Song[†], Kyungtae Kang^{*}

[†]Dept. of Applied Artificial Intelligence, Hanyang University

^{*}Dept. of Artificial Intelligence, Hanyang University

요약

최근 몇 년 간 즉석 촬영기기의 보급이 급속도로 확대되면서 공공 장소에서의 부적절한 이미지 촬영이 증가하고 있다. 이는 개인의 사생활 침해, 불법 촬영과 같은 법적 문제 뿐만 아니라 음란물 유통과 같은 윤리적 문제를 포함한다. 특히, 무분별한 이미지 공유는 청소년과 어린이에게 미치는 부정적 영향이 크며, 이는 사회적 안정과 도덕적 가치를 위협하고 있다. 이러한 문제에 대응하기 위해 본 연구에서는 TensorFlow 기반의 컨볼루션 신경망(CNN)을 이용한 실시간 이미지 분석 시스템을 제안한다. 제안된 시스템은 촬영기기에서 촬영된 이미지 중 음란 콘텐츠를 자동으로 탐지하고 차단하여 기업들이 기술적 솔루션을 통해 소비자의 신뢰를 확보하고 기업 이미지를 개선할 수 있도록 지원하며, 공공 장소에서의 부적절한 행동을 감지하고 억제함으로써 사회적 기준과 안전을 유지하는 데 기여할 수 있다.

I. 서론

최근 몇 년 간 즉석 촬영기기의 보급이 급속도로 확대되면서 공공 장소에서의 부적절한 이미지 촬영이 증가하고 있다. 이는 개인의 사생활 침해, 불법 촬영과 같은 법적 문제 뿐만 아니라 음란물 유통과 같은 윤리적 문제를 포함한다. 특히, 무분별한 이미지 공유는 청소년과 어린이에게 미치는 부정적 영향이 크며, 이는 사회적 안정과 도덕적 가치를 위협하고 있다. 또한 대부분의 무인 스튜디오에서는 모바일로 촬영된 사진이 QR 코드를 통해 전송되는 시스템이 도입되어 있다. 그러나 음란 콘텐츠와 같은 민감한 데이터가 서버에 저장되는 것은 정보 유출의 위험을 초래할 수 있다.

본 연구는 이러한 문제에 대응하기 위해 TensorFlow 를 기반의 컨볼루션 신경망(CNN)을 이용한 실시간 이미지 분석 시스템을 제안한다. 제안된 시스템의 CNN 아키텍처는 다양한 컨볼루션 및 풀링

레이어를 통해 이미지에서 고차원의 특징을 추출하고, 완전 연결 레이어와 드롭아웃을 사용하여 이를 처리하며, 최종적으로 이진 분류를 수행하여 즉석 촬영기기에서 촬영된 이미지 중 음란 콘텐츠를 자동으로 탐지하고 차단한다. 또한 이미지 처리는 로컬에서 이루어져 데이터의 외부 전송을 최소화함으로써, 서버에 저장된 데이터의 유출 방지에 기여한다.

본 연구는 총 3,496 개의 이미지 [1, 2]를 이용해 모델을 학습하였으며, 이를 통해 0.9188 의 정확도, 0.2116 의 손실, 그리고 0.9906 의 ROC-AUC 점수를 달성하였다. 이는 제안된 시스템이 음란 콘텐츠를 효과적으로 탐지할 수 있음을 시사하며 이 기술의 도입은 기업들이 기술적 솔루션을 통해 소비자의 신뢰를 확보하고 기업 이미지를 개선할 수 있도록 지원하며, 공공 장소에서의 부적절한 행동을 감지하고 억제함으로써 사회적 기준과 안전을 유지하는 데 기여할 수 있다.

[†] 바이오인공지능융합전공

^{*} 교신저자

표 1. 모델 요약

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 148, 148, 32)	896
max_pooling2d (MaxPooling2D)	(None, 74, 74, 32)	0
conv2d_1 (Conv2D)	(None, 72, 72, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 36, 36, 64)	0
conv2d_2 (Conv2D)	(None, 34, 34, 128)	73,856
max_pooling2d_2 (MaxPooling2D)	(None, 17, 17, 128)	0
flatten (Flatten)	(None, 36992)	0
dense (Dense)	(None, 512)	18,940,416
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 1)	513

II. 이미지 처리 모델

2.1. 데이터 준비 및 전처리

연구에 사용된 이미지 데이터는 음란 이미지와 비. 음란 이미지로 구분되며, 각각 1,782 개와 1,714 개로 총 3,496 개의 이미지가 수집되었다. 음란 이미지 데이터는 github nsfw_data_source_urls [1] repository 의 url 데이터를 크롤링하여 수집하였으며, 비음란 이미지 데이터는 kaggle 의 AISegment.com - Matting Human Datasets [2] 데이터셋을 사용하였다. 수집된 데이터는 무결성 검사를 거쳐 부적절하거나 손상된 파일을 제거하였다.

2.2. 모델 아키텍처

본 연구에서 제안된 모델은 컨볼루션 신경망(CNN) 을 기반으로 하며, 모델은 다음과 같은 여러 계층으로 구성된다:

- 입력 계층: 이 계층은 $150 \times 150 \times 3$ $150 \times 150 \times 3$ 크기의 이미지를 입력으로 받는다.
- 컨볼루션 계층: 첫 번째 컨볼루션 계층은 3×3 필터를 사용하여 32 개의 특성 맵을 생성한다. 활성화 함수로는 ReLU(Rectified Linear Unit) [3]가 사용된다. 이후 더 많은 특징을 추출하고 복잡한 특징을 학습할 수 있도록 3×3 필터를 사용하는 두 번째와 세 번째 컨볼루션 계층에서 각각 64 개와 128 개의 특성 맵을 생성한다.
- 풀링 계층: 각 컨볼루션 계층 다음에는 2×2 2×2 최대 풀링 계층이 배치되어 특성 맵의 크기를 줄이면서 주요 특징은 유지한다. 이는 모델의 계산 부담을 줄이고, 과적합을 방지하는 데 도움을 준다.
- 플랫텐 계층: 컨볼루션 및 풀링 계층을 거친 후, 특성 맵은 일렬로 평탄화되어 완전 연결 계층의 입력으로 사용된다.
- 완전 연결 계층: 512 개의 유닛을 가지는 완전 연결 계층이 하나 존재하며, 여기에는 50%의 드롭아웃이 적용되어 과적합을 방지한다.
- 출력 계층: 마지막으로, sigmoid 활성화 함수를 사용하는 하나의 유닛으로 구성된 출력 계층이 있어, 이미지가 음란 콘텐츠를 포함하는지 여부를 이진 분류한다.

모델은 Keras [4]의 Sequential 모델 클래스를 사용하여 구현되었다. 이는 각 계층을 순차적으로 쌓아 올리는 방식으로 간단하고 직관적인 모델 설계를 가능하게 한다. 각 계층의 매개변수는 실험을 통해 최적화되었으며, 모델 컴파일 시에는 Adam 최적화 알고리즘을 사용하고, 손실 함수로는 이진 교차 엔트로피를 사용한다.

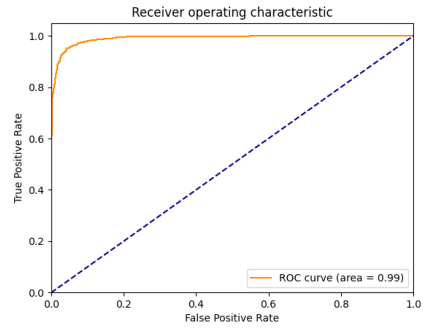


그림 1. ROC 곡선 시각화

III. 모델 평가

모델은 최종적으로 약 91.88%의 정확도와 0.2116의 손실 값을 달성하였다. 검증 데이터에 대한 ROC-AUC 점수는 0.9906으로, 모델이 음란 콘텐츠를 감지하는 데 매우 높은 정확성을 보임을 시사한다.

IV. 결론

본 연구에서는 공공 장소에서의 부적절한 이미지 촬영 문제에 대응하기 위해 컨볼루션 신경망(CNN) 기반의 이미지 분석 시스템을 제안하였다. 제안된 시스템은 약 93.00%의 높은 정확도와 0.9906의 ROC-AUC 점수를 달성하였으며, 이는 모델이 실제 환경에서도 높은 성능을 발휘할 수 있음을 시사한다. 본 시스템의 도입은 기업들이 기술적 솔루션을 통해 소비자의 신뢰를 확보할 수 있고 기업 이미지를 개선할 수 있도록 지원하며, 공공 장소에서의 부적절한 행동을 감지하고 억제함으로써 사회적 기준과 안전을 유지하는 데 기여할 수 있다.

ACKNOWLEDGMENT

이 논문은 2024 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.RS-2022-00155885, 인공지능융합혁신인재양성(한양대학교 ERICA))과 2024 년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임(P0012744, 2024 년 산업혁신인재성장지원사업)

참고 문헌

- [1] E. Bazarov, "nsfw_data_source_urls," GitHub, 2021. [Online]. Available: https://github.com/EBazarov/nsfw_data_source_urls. [Accessed: May 3, 2024].
- [2] L. Mih, "aisgmentcom-matting-human-datasets," Kaggle, 2020. [Online]. Available: <https://www.kaggle.com/datasets/laurentmih/aisgmentcom-matting-human-datasets>. [Accessed: May 3, 2024].
- [3] A. F. Agarap, "Deep learning using rectified linear units (relu)," arXiv preprint arXiv:1803.08375, 2018.
- [4] F. Chollet et al., "Keras," GitHub, 2015. [Online]. Available: <https://github.com/fchollet/keras>. [Accessed: May 3, 2024].