딥페이크 히트맵 분석 보고서

1. 분석 개요

- 모델명: MobileNetV3-Small
- 모델 유형: 외국인 전용 모델
- 분석 일시: 2025-11-07 09:10:19
- 예측 결과: Fake (52.49%)
- 딥페이크 확률: 18.80%

2. Grad-CAM 시각화

3. LangChain 기반 AI 해석

Grad-CAM 히트맵은 딥러닝 모델이 특정 클래스를 예측하는 데 있어 어떤 시각적 영역에 집중하고 있는지를 시각적으로 표현해주는 도구입니다. 주어진 예측 결과에 기반하여, 모델이 "Fake"라고 판단한 근거와 그 시각적 해석을 다음과 같이 설명할 수 있습니다.

Grad-CAM 히트맵 해석

1. **붉은색 영역 분석**:
    - **합성 흔적**: 붉은색 영역이 주로 얼굴의 경계, 턱선, 눈 주위에 집중되어 있을 경우, 이는 합성된 이미지의 자리잡기가 부자연스럽거나 경계가 흐리게 처리되어 있는 흔적을 나타낼 수 있습니다. 예를 들어, 합성된 피부와 자연 피부의 경계가 분명하게 드러난다면, 이 부분에 대한 주목이 높아질 수 있습니다.

- **피부 질감**: 만약 붉은색이 피부의 특정 부분에 나타난다면, 모공의 세밀한 표현 부족이나 불균형한 색조, 또는 비현실적인 매끄러움이 있을 수 있습니다. 피부 질감의 비자연스러운 처리(예: 너무 매끈하거나 반전된 광택)는 종종 딥페이크의 징후입니다.
- **조명 왜곡**: 조명이 불균형하게 나타나는 경우에도 붉은색 영역이 강조될 수 있습니다. 예를 들어, 얼굴의 한쪽 면은 자연광처럼 보이지만 다른 한쪽은 디지털 조명으로 보이는 경우, 이는 모델이 두 영역 간의 조명 차이를 감지한 것으로 해석될 수 있습니다. 조명의 비대칭은 관찰자에게 불편한 시각적 경험을 유발할 수 있습니다.

2. **모델의 판단 근거**:
- 모델의 예측 결과인 52.49%는 모델이 해당 이미지가 진짜일 확률보다 약간 더 높은(가령, 50% 기준) 확률로 딥페이크라고 판단했음을 나타냅니다. 이는 불확실성의 요인 중 하나로, 모델의 판단이 중요하지만 절대적이지 않다는 것을 시사합니다.
- 18.80%의 딥페이크 확률은 모델이 해당 이미지의 합성 가능성을 더 명확히 뒷받침하는 요소로 작용하며, 이는 이미지의 일반적인 특성이 아닌 특정 시각적 이상 사항 때문에 유도된 것으로 해석될 수 있습니다.

신뢰도와 한계점

- **신뢰도**:
  - 모델의 Grad-CAM 해석은 시각적으로 해석 가능하다는 점에서 의미가 있으며, 특정 부분에 대한 모델의 주의 집중을 명확히 보여줍니다. 따라서 전문가가 이 정보들을 기반으로 결론을 도출할 때, 유용한 기반 자료로 활용될 수 있습니다.
  - 피부 질감, 조명 등의 요소에 대해 전문가가 추가 정보를 제공할 수 있다면, 모델의 판단에 대한 신뢰도는 더욱 높아질 것입니다.

- **한계점**:
  - 모델이 특정한 시각적 패턴에 과도하게 의존할 경우, 새로운 형태의 딥페이크나 변형된 이미지에서 오탐지할 가능성이 있습니다. 실제로, 지나치게 자연스러운 합성 이미지는 모델이 '진짜'로 잘못 판단할 수 있으며, 반대로 잘못된 합성 이미지도 진짜로 판단할 수 있습니다.
  - 또한, Grad-CAM의 해석 결과는 주관적인 평가에 영향을 받을 수 있습니다. 이를 통한 근거가 절대적이지 않으므로, 다른 평가 방법이나 알고리즘을 통해 보완해야 합니다.

심층 결과
- **기타 심층 분석**: 이미지의 얼굴 인식, 비율, 포즈와 같은 추가적인 시각적 특성을 분석함으로써 모델이 판단한 ISP (Image Structure Prior) 요소들을 고려해 볼 수 있습니다. 이는 훈련 데이터셋에 따라 다르게 영향을 받을 수 있으며, 종합적인 분석이 필요합니다.

이와 같은 방식으로 Grad-CAM의 결과를 해석하고 각 시각적 요소의 의미를 규명함으로써, 딥페이크 탐지

분야에서의 신뢰성과 정확성을 높일 수 있습니다.

4. 결론 및 권장 조치

본 분석은 Grad-CAM 시각 주목도를 중심으로 진행되었습니다.
AI의 결과는 참고용으로 사용해야 하며, 법적 판단이나 공식 증거로 사용되지 않습니다.
결과의 신뢰도를 높이기 위해 다양한 이미지 소스로 교차 검증을 권장합니다.