

# 딥페이크 히트맵 분석 보고서

## 1. 분석 개요

- 모델명: MobileNetV3-Small
- 모델 유형: 한국인 전용 모델
- 분석 일시: 2025-11-06 16:54:04
- 예측 결과: Real (48.54%)
- 딥페이크 확률: 13.30%

## 2. Grad-CAM 시각화



## 3. LangChain 기반 AI 해석

Grad-CAM(Gradient-weighted Class Activation Mapping)은 딥러닝 모델의 예측을 해석하기 위해 사용되는 기법으로, 특정 클래스에 대한 네트워크의 관심 영역을 시각화합니다. 주어진 정보에 기반하여 Grad-CAM 히트맵을 해석해 보겠습니다.

### 모델의 판단 근거

### 1. \*\*붉은색 영역 분석\*\*:

- \*\*합성 흔적\*\*: 붉은색으로 표시된 영역은 모델이 딥페이크 여부를 판단하기 위해 중요하다고 생각하는 이미지의 특정 부분입니다. 이 영역에서는 얼굴 윤곽선이 자연스럽지 않거나, 턱선의 일관성이 결여된 경우가 많습니다. 예를 들어, 자연스러운 얼굴에서는 턱선이 부드럽고 일정하게 이어지지만, 딥페이크

합성에서는 경계가 불분명하거나 불규칙해보일 수 있습니다.

- **피부 질감**: 피부 질감의 불일치도 딥페이크 탐지에 중요한 요소입니다. 붉은색 히트맵이 피부의 특정 부분에 집중된다면, 그 부분에서 피부의 질감이 비정상적이거나 고르지 않을 수 있습니다. 노이즈가 많거나, 평면적이고 반사된 것처럼 보이며, 이는 일반적인 피부 조명과 다른 특성을 나타낼 수 있습니다.

- **조명 왜곡**: 조명은 얼굴의 자연스러움을 정의하는 중요한 요소입니다. 붉은 영역이 얼굴의 특정 부분에서 강하게 나타난다면, 그 부분의 조명이 비대칭적이거나 어색하게 처리되었음을 나타낼 수 있습니다. 특히, 조명의 방향이 인물의 다른 부분과 일치하지 않거나, 그림자와 하이라이트가 어색한 경우에는 딥페이크일 가능성성이 높아집니다.

### ### 전문가의 신뢰도와 한계점

- **신뢰도**: Grad-CAM을 사용한 분석은 특정 이미지의 중요한 시각적 특성을 파악하는 데 유용하지만, 완벽하지는 않습니다. 만약 붉은색 영역이 실질적으로 합성된 부분이나 불일치가 있는 부분과 일치한다면, 모델의 예측 결과에 대한 신뢰도가 높아질 수 있습니다. 이는 특정 패턴에 대해 통계적으로 학습한 모델이 해당 패턴을 인간 전문가가 인지하는 방식과 유사하게 인식하고 있다는 것을 나타냅니다.

- **한계점**: 그러나 Grad-CAM은 항상 정확한 해석을 제공하지 않을 수 있으며, 여러 요인에 의해 왜곡될 수 있습니다. 예를 들어, 모델이 학습한 데이터 세트에 따라 특정 시각적 특징이 과도하게 강조될 수 있고, 이는 잘못된 판단으로 이어질 수 있습니다. 또한, 딥페이크 기술이 진화하고 있는 상황에서 모델이 새로운 유형의 합성 이미지를 정확하게 탐지하지 못할 가능성이 존재합니다.

### ### 심층 결과

- **주목할 점**: 모델이 각 영역에 얼마나 민감한지를 파악하는 것도 중요한데, 특정 얼굴 표정이나 각도에서도 수치가 크게 달라질 수 있어, 단순히 예측 확률만을 가지고 판단하기에는 한계가 있습니다. 예를 들어, 모델이 잘못된 신뢰를 가질 수 있는 경향이 있다고 알려진 특정 상황(예: 비슷한 얼굴을 가진 비정상적인 표정)에서 신뢰도를 낮출 필요가 있습니다.

- **추가 고려사항**: 딥페이크 탐지의 마지막 단계에서는 시각적, 행동적 특성과 같은 다른 요소들도 종합적으로 고려해야 합니다. 예를 들어 시각 정보를 보완하기 위해 기계 학습 기법과 전통적인 영상 분석 기법을 접목하는 것이 효과적일 수 있습니다.

결론적으로, Grad-CAM은 딥페이크 탐지의 한 요소로서 중요한 역할을 하지만, 다른 다양한 요소들과 함께 신중하게 해석하고, 한계점을 인식하며 활용하는 것이 필요합니다.

## 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.