범죄 현장의 CCTV 속 '흐릿한' 눈·코·입...
인공지능 모델로 한층 '또렷이' 인식한다 !

- 고해상도 이미지로 학습한 정보를 저해상도 인식 AI 모델에 전달해 얼굴 특징 인식
- 세계 3대 컴퓨터 비전 학회(S급) <유럽 컴퓨터 비전 학술대회(ECCV)>서 10.23. 발표

▲ 이규빈 교수 연구팀_(왼쪽 하단부터 시계 반대 방향으로) 융합기술학제학부 통합과정 이주순, 통합과정 신성호, 이규빈 교수, 통합과정 이준석, 박사과정 유연국

지스트(광주과학기술원, 총장 김기선) 연구진이 **인공지능(AI) 딥러닝 기술을 적용**해 **저화질 영상에서도 눈·코·입 등 사람 얼굴의 주요 특징을 잘 인식할 수 있는 기술**을 개발했다.

저화질 화면에서는 얼굴 인식 성능이 급격하게 저하돼 수사기관이나 보안업체 등이 대상자의 신원 확인에 어려움을 겪는 일이 많은데, 이번 성과가 활용되면 **저화질 CCTV 영상 등에서의 얼굴 인식 성능이 개선될 것으로 기대**된다.

얼굴 인식 분야에서 딥러닝 기술은 99% 이상의 정확도를 보이며 범죄 현장 등 보안 분야에서 활발히 활용되고 있다.

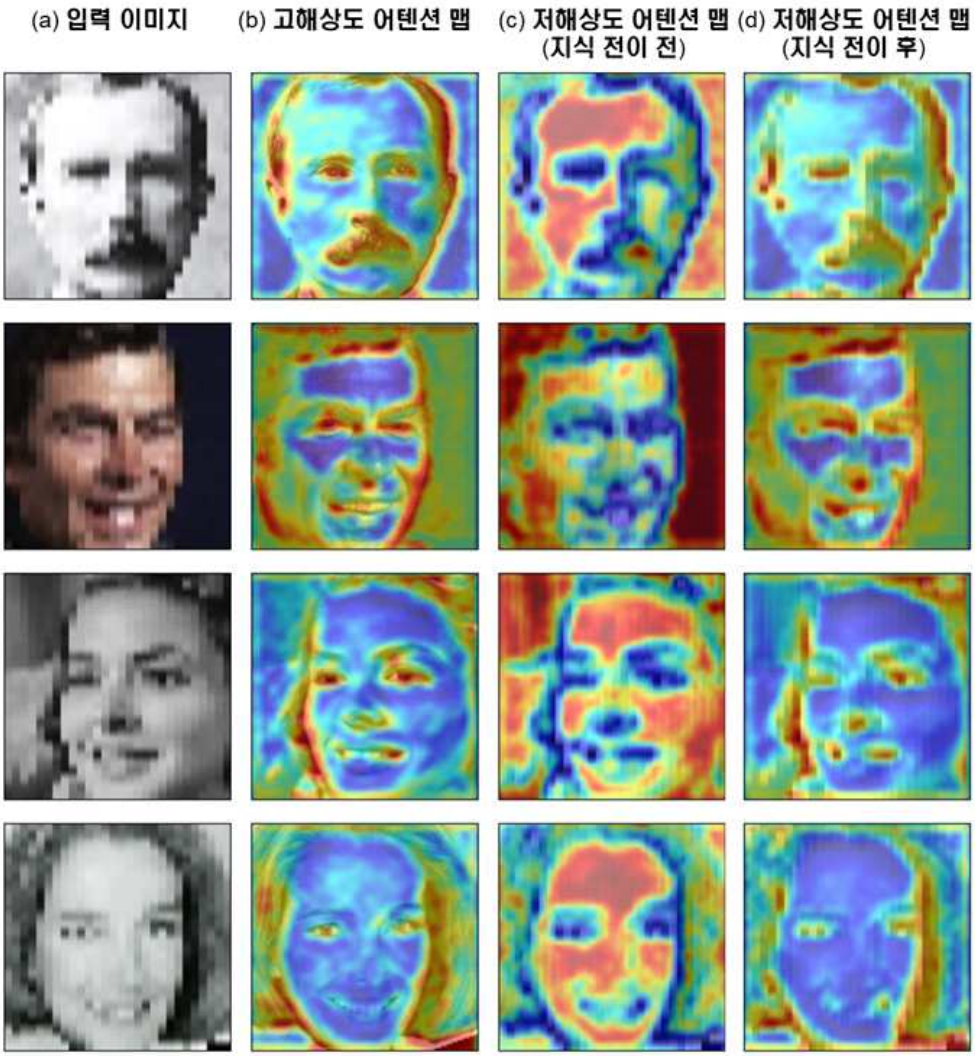
하지만 기존의 얼굴 인식 딥러닝 기술은 24×24 픽셀 수준의 저해상도 이미지에서는 사람 얼굴의 주요 특징을 담고 있는 이목구비 영역이 아닌, 피부와 같이 해상도에 영향을 적게 받는 부위에 집중해 얼굴 인식을 수행하며, 고해상도(평균 112×112 픽셀) 얼굴 이미지들을 학습한 기존 인식 모델의 경우도 저해상도(평균 24×24 픽셀) 이미지들로 구성된 TinyFace 벤치마크*에서 평가했을 때 인식 정확도가 30% 수준에 그치는 실정이다.

* **TinyFace 벤치마크**: 딥러닝 모델들의 공정한 비교를 위해 동일한 데이터 셋으로 평가 환경을 구성해둔 것을 벤치마크라고 하며, TinyFace 데이터 셋은 저해상도 얼굴 이미지의 인식 성능을 비교하기 위한 대표적인 벤치마크임. 다양한 자세, 조명, 배경 환경에서 5,139명의 사람들로부터 촬영된 169,493장의 저해상도 얼굴 이미지를 포함하고 있음.

또한, 이미지의 해상도가 높더라도 멀리서 촬영되거나 다수의 사람이 한꺼번에 촬영될 경우 각각의 얼굴 정보는 적은 픽셀로 구성돼 낮은 인식 성능을 보인다. 상용화된 얼굴 인식 제품들이 가까운 거리에서 한 명씩 촬영한 후 인식을 수행하는 이유이다.

지스트 융합기술학제학부 이규빈 교수 연구팀은 어텐션 맵*을 이용해 **고해상도 얼굴 이미지에서 학습한 정보를 저해상도 얼굴 이미지 인식 모델에 전달해 성능을 향상**시킬 수 있는 방법을 개발했다.

* **어텐션 맵(attention map)**: 딥러닝 네트워크가 인식을 수행할 때 입력 이미지의 어떤 영역이 인식 결과에 많은 영향을 주었는지 시각화 한 정보. 0에서 1범위의 활성값을 보이며, 1에 가까울수록 결과에 영향을 많이 주는 영역임.



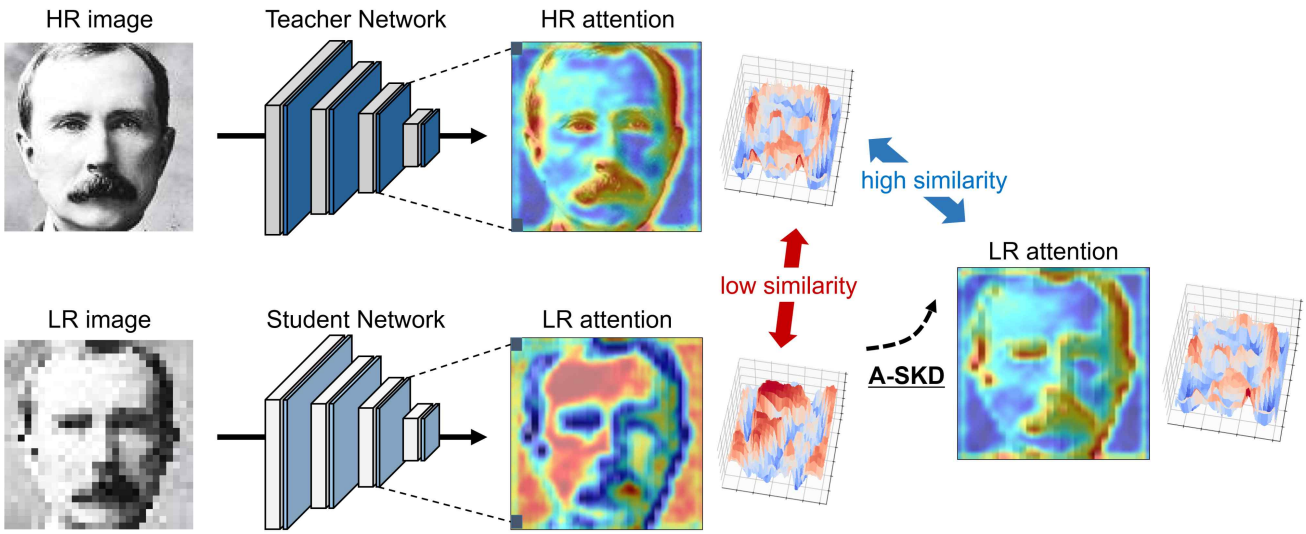
▲ **고해상도 네트워크와 저해상도 네트워크에서 추출된 어텐션 맵 예시**. (a) 입력 이미지, (b) 고해상도 네트워크에서 추출된 어텐션 맵, (c) 지식 전이 기법이 적용되지 않은 저해상도 네트워크에서 출력된 어텐션 맵, (d) 제안하는 어텐션 맵 전이 기법이 적용된 후의 저해상도 네트워크에서 추출된 어텐션 맵. **어텐션 맵 전이 기법이 적용된 후, 저해상도 네트워크에서 추출된 어텐션 맵이 고해상도 네트워크에서 추출한 어텐션 맵과 유사한 경향을 보이는 것을 확인하였다.**

높은 인식 성능을 보이는 고해상도 네트워크는 어텐션 맵에서 주요 얼굴 영역이 1에 가까운 높은 활성 값을 나타내지만, 저해상도 네트워크의 어텐션 맵은 해당 영역에서 0에 가까운 낮은 활성 값을 나타낸다.

연구팀은 저해상도 네트워크의 어텐션 맵이 고해상도 네트워크의 어텐션 맵과 유사해지도록 학습하는 손실 함수를 추가하여, 저해상도 네트워크가 얼굴 인식에 도움이 되는 주요 영역에 초점을 맞출 수 있도록 유도했다.

연구팀은 CASIA-WebFace 데이터 셋으로 훈련하여 인 TinyFace 공인 벤치마크에서 저해상도(평균 24×24 픽셀) 이미지로 얼굴 인식 성능을 평가했을 때 기존 세계 최고 수준의 인식 성능인 45.49%보다 **5% 향상된 47.91%의 인식 정확도**를 얻으며, **현재까지 가장 높은 수준의 성능을 달성**했다.

얼굴 인식 외에도 저해상도 이미지에서 물체의 종류를 분류하는 과업에서도 연구팀이 제안한 방법이 높은 성능 향상을 거두어 다양한 컴퓨터 비전 태스크에서 적용될 수 있음을 확인했다.



▲ **저해상도 얼굴 인식 성능 향상을 위한 어텐션 맵 전이 기법 개요**. 해상도가 높은 이미지를 학습한 인식 모델은 높은 인식 정확도를 보이며, 이때 추출된 어텐션 맵은 눈, 코, 수염과 같이 사람을 구분 짓는 주요 영역에서 활성화된다. 하지만, 저해상도 이미지를 학습한 모델은 고해상도 네트워크와 달리 피부와 같이 해상도 변화에 둔감한 영역에서 높은 활성도를 보이며 이는 성능 저하의 원인이 된다. **이번 연구에서는 저해상도 네트워크에서 추출된 어텐션 맵이 고해상도 네트워크에서 추출된 어텐션 맵과 유사해지도록 학습하는 손실 함수를 제안**하여, 저해상도 네트워크도 얼굴 인식에 중요한 영역에 집중하여 인식을 할 수 있도록 했다.

이규빈 교수는 "이번 연구 성과가 응용되면 범죄 해결의 중요 단서를 제공하는 CCTV로 멀리서 촬영된 사람의 얼굴 특징을 정확하게 인식할 수 있게 됐다"면서 "연구팀이 제안한 '어텐션 맵 전이 기법'은 얼굴 인식 외에도 물체 영역 검출, 종류 분류 등 컴퓨터 비전의 다양한 과업에서 핵심 기술로 활용될 것으로 기대된다"고 말했다.

이번 연구는 **과학기술정보통신부의 클라우드 로봇복합인공지능 핵심기술개발사업**, **산업통상자원부의 에너지기술개발사업**, **한국전자통신연구원 연구운영지원사업**의 지원을 받아 수행되었으며, 연구에 사용된 코드는 깃허브에서 오픈소스로 이용할 수 있다. (https://github.com/gist-ailab/teaching-where-to-look)

이번 논문은 세계 3대 컴퓨터 비전 학회인 <유럽 컴퓨터 비전 학술대회(ECCV: European Conference on Computer Vision) 2022>에서 10월 23일(일) 발표됐다.

논문 및 저자 내용

1. 발표 예정 학회 / 논문명

- 학회명 : European Conference on Computer Vision (ECCV)

  * 세계 3대 컴퓨터 비전 학회로 (2022년 기준 h5-index=186 in Computer Vision and Pattern Recognition, Top 3), 인공지능/컴퓨터비전 분야 최우수 학술대회, 한국정보과학회 기준 최우수 학술대회(S급)

- 논문명 : Teaching Where to Look: Attention Similarity Knowledge Distillation for Low Resolution Face Recognition

- 저자 정보 : 신성호 박사과정생(제1저자, 융합기술학제학부), 이주순 박사과정생, 이준석 박사과정생, 유연국 박사과정생, 이규빈 교수(교신저자, 융합기술학제학부)