'Blurry' eyes, nose, mouth in CCTV crime scene... Recognized more 'clearly' with an artificial intelligence model!

- Recognize facial features by passing information learned from high-resolution images to low-resolution recognition AI models
- Presented at world's top 3 computer vision conferences (Class S) <European Computer Vision Conference (ECCV)> 10.23

▲ Professor Kyoobin Lee's research team (counterclockwise from the lower left): Integrated student Joosoon Lee, integrated student Sungho Shin, Professor Kyoobin Lee, integrated student Junseok Lee, and Ph.D. student Yeonguk Yu

GIST (Gwangju Institute of Science and Technology, President Kiseon Kim) researchers have developed a technology that can recognize the main features of a human face, such as eyes, nose, and mouth, even in low-resolution images by applying artificial intelligence (AI) deep learning technology.

In low-resolution screens, face recognition performance deteriorates sharply, so investigative agencies and security companies often have difficulties in identifying the subject. If this achievement is utilized, it is expected that the face recognition performance in low-quality CCTV images will be improved.

In the face recognition field, deep learning technology has an accuracy of over 99% and is being actively used in security fields such as crime scenes.

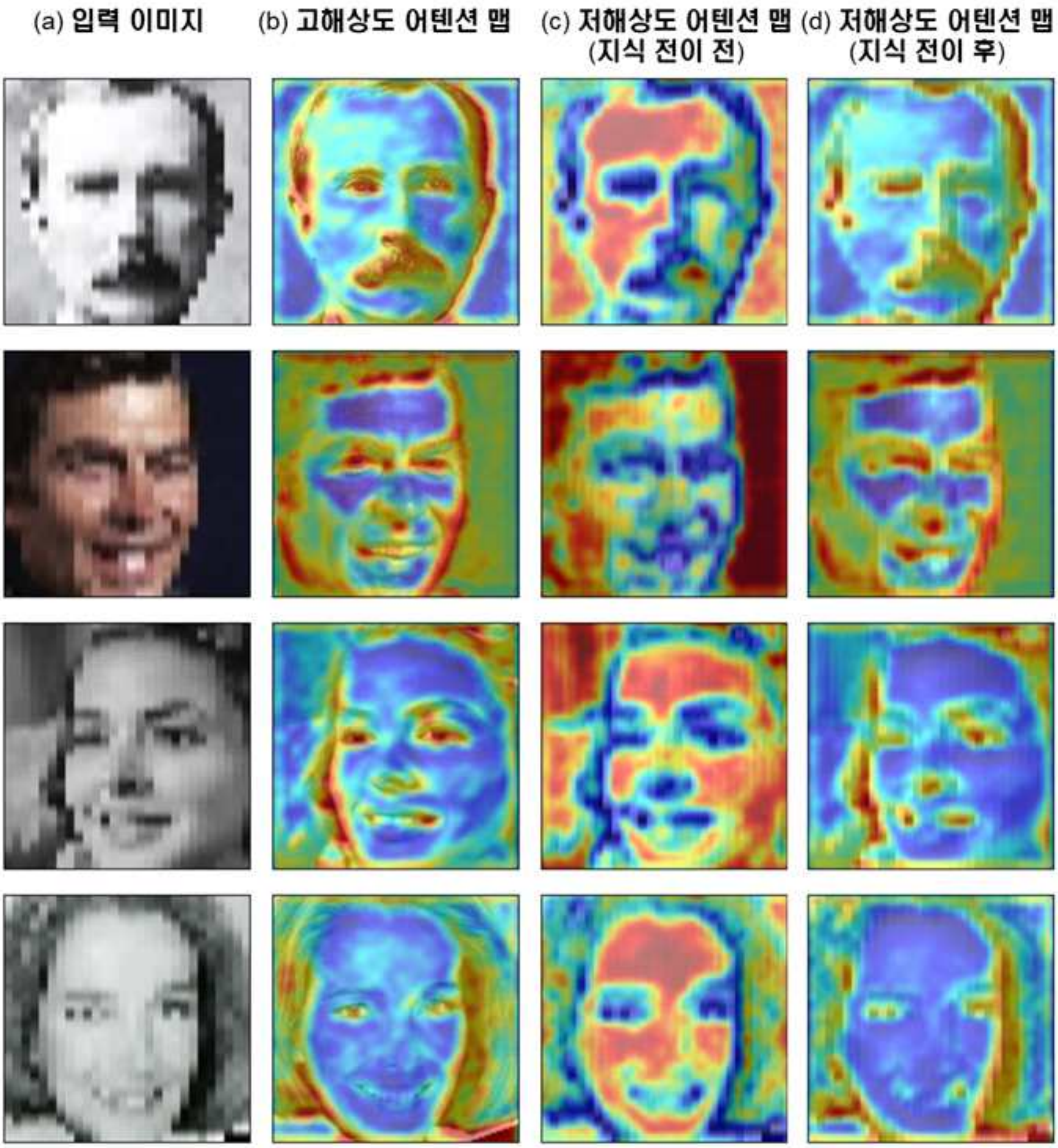
However, the existing deep learning technology for face recognition performs face recognition by focusing on areas that are less affected by resolution, such as skin, rather than the feature area containing the main features of a human face in a low-resolution 24×24 pixel level image. Even in the case of the existing recognition model trained on high-resolution (average 112×112 pixels) face images, the recognition accuracy is only 30% when evaluated in the TinyFace benchmark* composed of low-resolution (average 24×24 pixels) images.

* TinyFace benchmark: For fair comparison of deep learning models, an evaluation environment configured with the same data set is called a benchmark, and the TinyFace data set is a representative benchmark for comparing the recognition performance of low-resolution face images. It contains 169,493 low-resolution face images taken from 5,139 people in various poses, lighting, and background environments.

In addition, even if the resolution of the image is high, when shooting from a distance or shooting multiple people at once, each face information is composed of fewer pixels and shows low recognition performance. This is why commercialized face recognition products perform recognition after taking pictures of each person at a close distance.

A research team led by GIST School of Integrated Technology Professor Kyoobin Lee developed a method to improve performance by transferring information learned from high-resolution face images to a low-resolution face image recognition model using the attention map*.

* attention map: information that visualizes which areas of the input image have a great influence on the recognition results when the deep learning network performs recognition. It shows an active value in the range of 0 to 1, and the closer to 1, the more influential the result.



▲ Examples of attention maps extracted from high-resolution and low-resolution networks. (a) the input image, (b) the attention map extracted from the high-resolution network, (c) the attention map output from the low-resolution network to which the knowledge transfer technique is not applied, (d) the low-resolution network after the proposed attention map transfer technique is applied attention map extracted from. After the attention map transition technique was applied. It was confirmed that the attention map extracted from the low-resolution network showed a similar tendency to the attention map extracted from the high-resolution network.

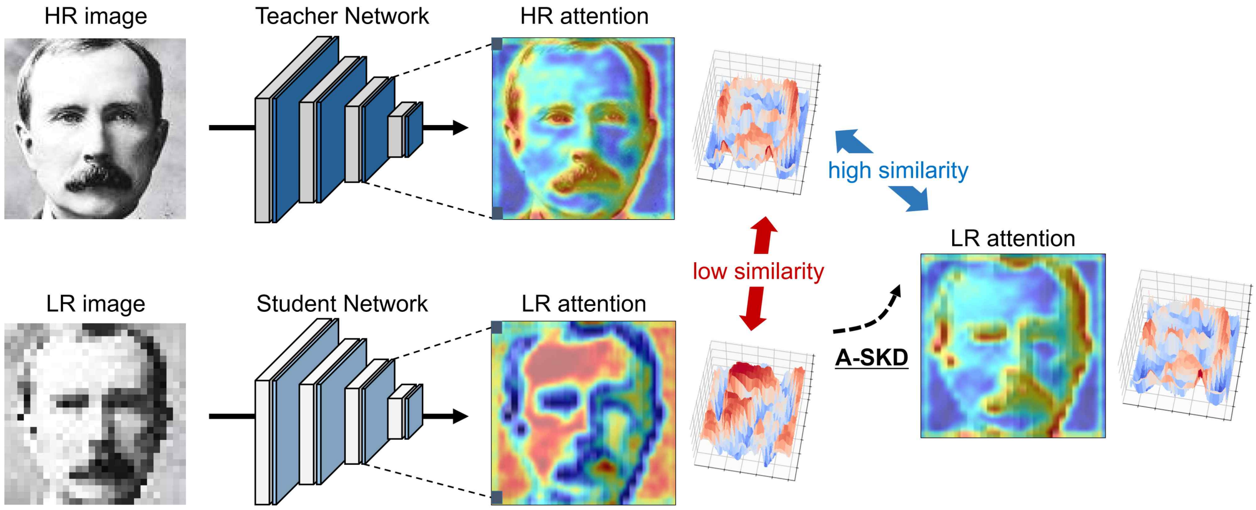
A high-resolution network with high recognition performance shows a high activity value close to 1 in the main face region in the attention map, but a low activity value close to 0 in the attention map of a low-resolution network.

The research team added a loss function that trains the attention map of the low-resolution network to be similar to the attention map of the high-resolution network, so that the low-resolution network can focus on key areas that are helpful for face recognition.

When the research team trained on the CASIA-WebFace dataset and evaluated the face recognition performance with low-resolution (average 24 × 24 pixels) images in the

TinyFace certified benchmark, 47.91% improved by 5% compared to the existing world-leading recognition performance of 45.49%. of recognition accuracy, and achieved the highest level of performance so far.

In addition to face recognition, it was confirmed that the method proposed by the research team achieved high performance improvement in the task of classifying types of objects in low-resolution images, so that it can be applied in various computer vision tasks.



▲ Overview of Attention Map Transition Techniques for Improving Low-Resolution Face Recognition Performance. A recognition model that has learned a high-resolution image shows high recognition accuracy, and the extracted attention map is activated in the main areas that distinguish people, such as eyes, nose, and whiskers. However, unlike the high-resolution network, the model trained on low-resolution images shows high activity in areas that are insensitive to resolution changes, such as skin, and this causes performance degradation. This study proposed a loss function that learns the attention map extracted from the low-resolution network to be similar to the attention map extracted from the high-resolution network.

Professor Kyoobin Lee said, "If the results of this research are applied, it will be possible to accurately recognize the facial features of people filmed from a distance with CCTV, which provides important clues in solving crimes. The 'attention map transfer technique' proposed by the research team is expected to be used as a core technology in various tasks of computer vision, such as object area detection and type classification, in addition to face recognition."

This research was carried out with the support of the Cloud Robot Complex Artificial Intelligence Core Technology Development Project of the Ministry of Science and ICT, the Energy Technology Development Project of the Ministry of Trade, Industry and Energy, and the Research Operation Support Project of the Electronics and Telecommunications Research Institute. It is available as open source on Hub. (https://github.com/gist-ilab/teaching-where-to-look)

This paper was presented on October 23 (Sun) at the <European Conference on Computer Vision (ECCV) 2022>, one of the world's top three computer vision conferences.

GIST
Since 1993