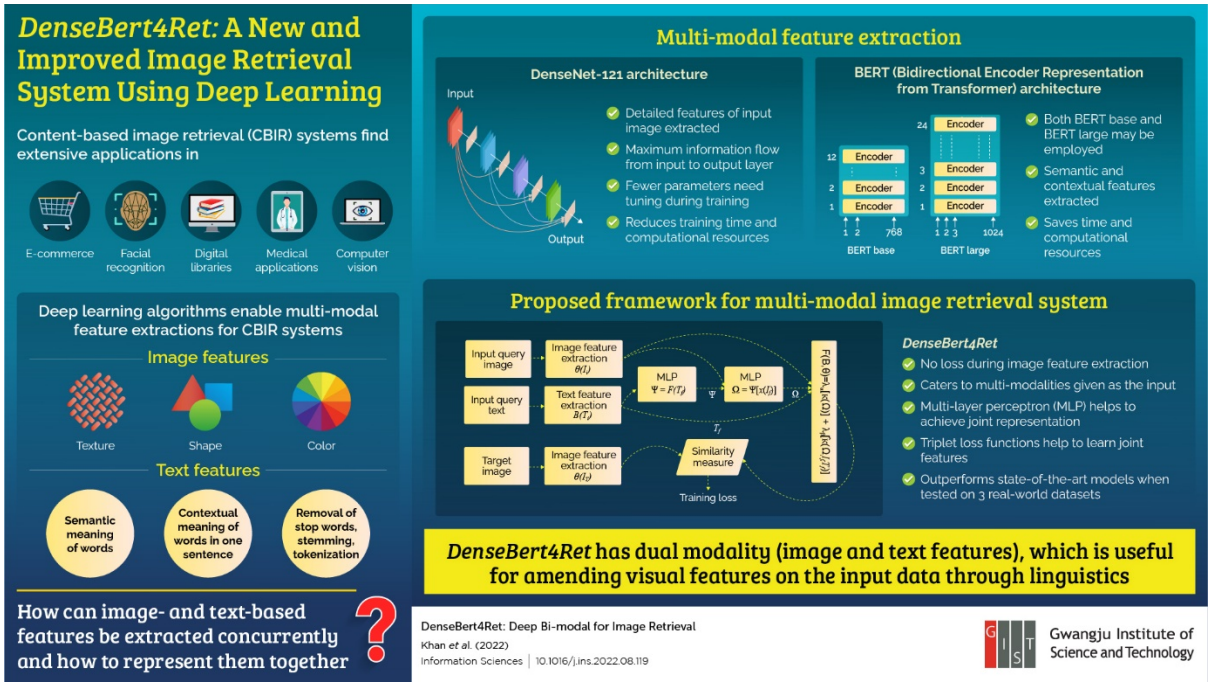
PRESS RELEASE

A Novel Multi-Modal Image Retrieval System from Researchers at the Gwangju Institute of Science and Technology

Researchers from Korea develop a new image retrieval system using deep learning algorithms

Sifting through large amounts of data available on the internet is a difficult task. Recently, researchers from the Gwangju Institute of Science and Technology in Korea developed a bi-modal image retrieval system that takes both image and text as the input query to extract the desired image from a database.



Title: A deep learning-aided multi-modal image retrieval system
Caption: Researchers from the Gwangju Institute of Science and Technology in Korea, have developed a new image retrieval system called *DenseBert4Ret*, which uses deep learning for image and text feature extraction from a dual-mode input query, with potential applications in e-commerce, computer vision, and medicine.

Image credit: Moongu Jeon from Gwangju Institute of Science and Technology, Korea
License type: Original content
Usage restrictions: Cannot be reused without permission

With the amount of information on the internet increasing by the minute, retrieving data from it is like trying to find a needle in a haystack. Content-based image retrieval (CBIR) systems are capable of retrieving desired images based on the user's input from an extensive database. These systems are used in e-commerce, face recognition, medical applications, and computer vision. There are two ways in which CBIR systems work: text-based and image-based. One of the ways in which CBIR gets a boost is by using deep learning (DL) algorithms.

DL algorithms enable the use of multi-modal feature extraction, meaning that both image and text features can be used to retrieve the desired image. Even though scientists have tried to develop multi-modal feature extraction, it remains an open problem.

To this end, researchers from the Gwangju Institute of Science and Technology have developed *DenseBert4Ret*, an image retrieval system using DL algorithms. The study, led by Prof. Moongu Jeon and Ph.D. student Zafran Khan, was made available online on September 14, 2022, and published in Volume 612 of *Information Sciences*. *"In our day-to-day lives, we often scour the internet to look for things such as clothes, research papers, news article, etc. When these queries come into our mind, they can be in the form of both images and textual descriptions. Moreover, at times we may wish to amend our visual perceptions through textual descriptions. Thus, retrieval systems should also accept queries as both texts and images,"* says Prof. Jeon, explaining the team's motivation behind the study.

The proposed model used both image and text as the input query. For extracting the image features from the input, the team used a deep neural network model known as *DenseNet-121*. This architecture allowed for the maximum flow of information from the input to the output layer and only needed tuning of very few parameters during training. *DenseNet-121* was combined with the bidirectional encoder representation from transformer (BERT) architecture for extracting semantic and contextual features from the text input. The combination of these two architectures reduced training time and computational requirements and formed the proposed model, *DenseBert4Ret*.

The team then used Fashion200k, MIT-states, and FashionIQ, three real-world datasets, to train and compare the proposed system's performance against the state-of-the-art systems. They found that *DenseBert4Ret* showed no loss during image feature extraction and outperformed the state-of-the-art models. The proposed model successfully catered for multi-modalities that were given as the input with the multi-layer perceptron and triple loss function helping to learn the joint features.

"Our model can be used anywhere where there is an online inventory and images need to be retrieved. Additionally, the user can make changes to the query image and retrieve the amended image from the inventory," concludes Prof. Jeon.

Here's hoping to see the *DenseBert4Ret* system in application in our everyday-use search engines soon!

Reference

Title of original paper:	DenseBert4Ret: Deep bi-modal for image retrieval

Journal: *Information Sciences*

DOI: 10.1016/j.ins.2022.08.119

Affiliations: [1]School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology (GIST)
[2]National University of Science and Technology (NUST), Pakistan
[3]Department of Computer Engineering, Dankook University

*Corresponding author's email: mgjeon@gist.ac.kr

About the Gwangju Institute of Science and Technology (GIST)

The Gwangju Institute of Science and Technology (GIST) is a research-oriented university situated in Gwangju, South Korea. Founded in 1993, GIST has become one of the most prestigious schools in South Korea. The university aims to create a strong research environment to spur advancements in science and technology and to promote collaboration between international and domestic research programs. With its motto of "A Proud Creator of Future Science and Technology," GIST has consistently received one of the highest university rankings in Korea.

Website: http://www.gist.ac.kr/

About the author

Professor Moongu Jeon received his Ph.D. in Scientific Computation from the University of Minnesota in 2001. He is currently a professor in the School of Electrical Engineering and Computer Science at the Gwangju University of Science and Technology in South Korea. His research group works in the fields of artificial intelligence, machine learning, computer vision, and natural language processing. Various groups under his supervision are working in the fields of autonomous driving, visual surveillance, sign language generation, social media analysis, and information retrieval. He has 309 publications credited to him and 5,154 citations to his name.