

A novel multi-modal image retrieval system

November 8 2022

DenseBert4Ret: A New and Improved Image Retrieval System Using Deep Learning

Content-based image retrieval (CBIR) systems find extensive applications in

Deep learning algorithms enable multi-modal feature extractions for CBIR systems

Image features

Text features

How can image- and text-based features be extracted concurrently and how to represent them together?

Multi-modal feature extraction

DenseNet-121 architecture

- ✓ Detailed features of input image extracted
- ✓ Maximum information flow from input to output layer
- ✓ Fewer parameters need tuning during training
- ✓ Reduces training time and computational resources

BERT (Bidirectional Encoder Representation from Transformer) architecture

- ✓ Both BERT base and BERT large may be employed
- ✓ Semantic and contextual features extracted
- ✓ Saves time and computational resources

Proposed framework for multi-modal image retrieval system

DenseBert4Ret

- ✓ No loss during image feature extraction
- ✓ Caters to multi-modalities given as the input
- ✓ Multi-layer perceptron (MLP) helps to achieve joint representation
- ✓ Triplet loss functions help to learn joint features
- ✓ Outperforms state-of-the-art models when tested on 3 real-world datasets

DenseBert4Ret has dual modality (image and text features), which is useful for amending visual features on the input data through linguistics

DenseBert4Ret: Deep Bi-modal for Image Retrieval
 Khan et al. (2022)
 Information Sciences | 10.1016/j.ins.2022.08.119

Researchers from 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. Credit: Moongu Jeon from Gwangju Institute of Science and Technology, Korea

With the amount of information on the internet increasing by the minute, and retrieving meaningful data from it is sometimes 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 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 published in *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 had 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 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.

More information: Zafran Khan et al, DenseBert4Ret: Deep bi-modal for image retrieval, *Information Sciences* (2022). [DOI: 10.1016/j.ins.2022.08.119](https://doi.org/10.1016/j.ins.2022.08.119)

Provided by GIST (Gwangju Institute of Science and Technology)

Citation: A novel multi-modal image retrieval system (2022, November 8) retrieved 26 April 2024 from <https://techxplore.com/news/2022-11-multi-modal-image.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.