Image Search

A M.Tech project report submitted in partial fulfillment of the requirements
for the award of the degree of

M.Tech CrS

by

Subrata Mondal

(Roll No. CrS2219)

Under the Supervision of

Jayanta Kumar Mukherjee (External Supervisor)

&

Debrup Chakraborty (Internal Supervisor)

Department of Cryptology and Security

Indian Statistical Institute Kolkata

Kolkata - 700108, India

June, 2024

Certificate

This is to certify that the report entitled **"Image Search"**, submitted by **Subrata Mondal** to the Indian Statistical Institute Kolkata, for the award of the degree of **Master of Technology in Cryptology ans Security**, is a record of the original, bona fide research work carried out by him under our supervision and guidance. The report has reached the standards fulfilling the requirements of the regulations related to the award of the degree.

The results contained in this report have not been submitted in part or in full to any other University or Institute for the award of any degree or diploma to the best of our knowledge.

JAYANTA KUMAR MUKHERJEE

. . . .

Jayanta Kumar Mukherjee
Technical Architect,
ARC Document Solution.

. . . .

Debrup Chakraborty
Associate Professor,
Indian Statistical Institute Kolkata.

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

. . . .

Subrata Mondal
Roll No.: CrS2219
Date:02/07/2024
Place: ISI Kolkata

Acknowledgements

I take this opportunity to acknowledge and express my gratitude to all those who supported and guided me during the dissertation work. I would like to express my deepest gratitude to my supervisors, **Jayanta Kumar Mukherjee** and **Debrup Chakraborty** , for their invaluable guidance, support, and encouragement throughout this project. Their expertise and insights have been instrumental in the completion of this work. I am sincerely grateful for their time and effort in mentoring me and for the knowledge they have shared.

Subrata Mondal

Abstract

With the rapid increase in digital images, it has become essential to have advanced systems to find specific images quickly from large collections. Traditional methods that depend on text descriptions often fail because tagging images manually is time-consuming and subjective. This project uses deep learning to create an efficient image search system for a dataset of about approximately 5000 printing images.Transfer Learning technique has been implemented in this work. Transfer learning is an ambitious task, but it results in impressive outcomes for identifying distinct patterns in tiny datasets of approximately 5000 images of printing images from our web site 'ARC Print'. The goal is to produced best feature vectors that capture the important details of each image, allowing us to search based on content rather than text.

We tested the system for accuracy and speed, showing that it works well and is efficient. Feedback from management also confirms that the system is practical and useful. The results indicate that our method is much better than traditional ones, providing quick and accurate search results based on image content.This project demonstrates the power of deep learning in image search, and it can be used in many areas specially in online shopping. The proposed model achieved 89 % accuracy and based on our findings,the proposed system can help to enhance the user experience on our website far better.In the future, we aim to improve the system further and explore more applications, highlighting the importance of advanced machine learning in handling large collections of images.

Contents

Chapter 1

Introduction

1.1 Introduction

Image search refers to the process of using an image as a query to find similar or related images. See the Figure 1.1. This technology is pivotal in various applications such as digital asset management, e-commerce, and content-based image retrieval systems. Traditionally, image search algorithms relied on basic visual features and metadata. However, recent advancements in deep learning, particularly convolutional neural networks (CNNs), have revolutionized the field by enabling more accurate and efficient image retrieval systems.The Medium blog post (Aguas, 2020) and (Varghese Alex, 2019) helped us to build a good knowledge on this field.

At its core, the Image Search System developed in this project employs sophisticated algorithms to analyze the visual content of images. It extracts key features such as colors, textures, shapes, and objects, utilizing these features to build a comprehensive index of the image database. This enables fast and accurate retrieval based on user queries. User experience is a central focus, with an emphasis on simplicity, responsiveness, and relevance.

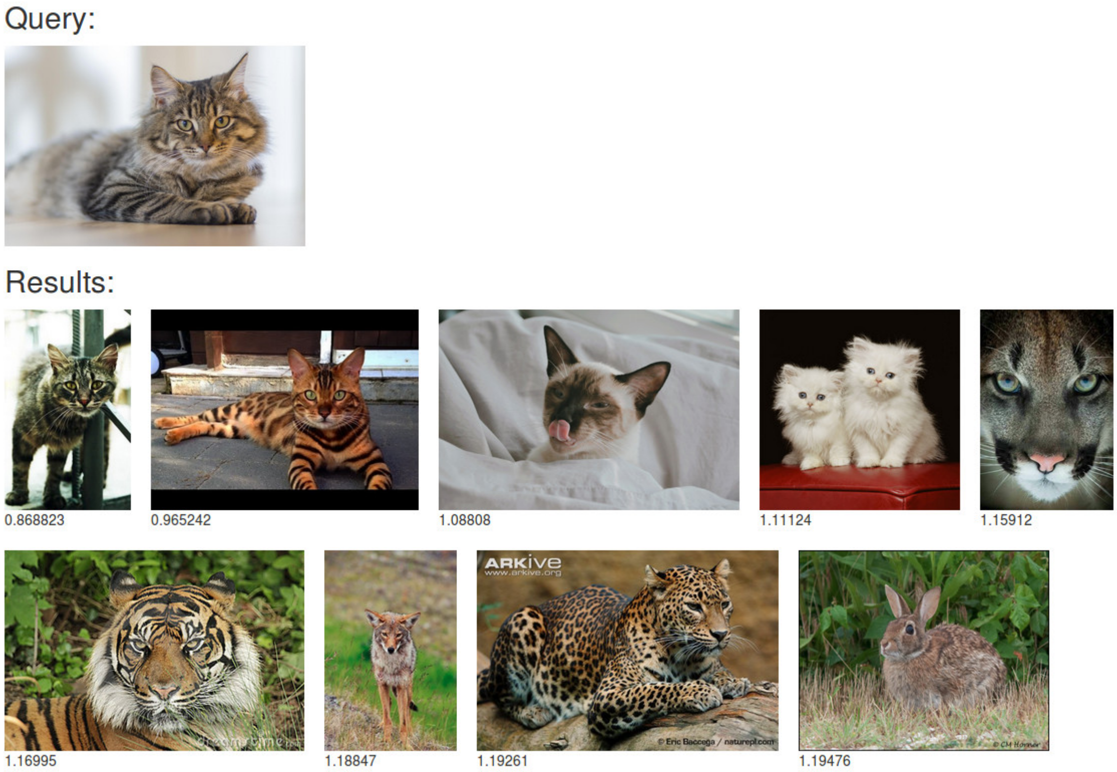

Figure 1.1: Example of a Image Search Engine.

Our investigation includes a thorough review of image feature extraction,model fine tuning, similarity metrics, and indexing strategies pivotal to the image search process. By leveraging insights from the latest developments in machine learning and data processing, the aim is to propose innovative solutions to enhance the performance and scalability of image search systems. Ultimately, the goal is to build a visual content recommendation system that can find visually similar images, identify objects. We believe our discoveries will advance discussions among researchers in computer vision and information retrieval, leading to the creation of smarter image search systems.

1.2 Problem Statement

The problem is further compounded by the diversity of images in terms of content, style, and quality. An effective image search system must be able to notice small differences and similarities between images, which requires sophisticated feature extraction and similarity measurement techniques. Our project addresses these challenges by leveraging deep learning models, specifically fine-tuned versions of ResNet50, to improve the quality of image embeddings and enhance search accuracy.

1.3 Need of the Study

With the proliferation of digital images across various domains, there is a growing demand for efficient and accurate image search systems. Such systems are essential for applications in e-commerce, digital asset management, social media, and more. Users expect to find relevant images quickly and effortlessly, which necessitates the development of advanced algorithms capable of handling large and diverse image datasets.

This study is needed to bridge the gap between traditional image search methods and the capabilities offered by modern deep learning techniques. By focusing on the fine-tuning of CNNs, particularly ResNet50, we aim to achieve state-of-the-art performance in image retrieval. The outcomes of this study will contribute to the development of more robust and scalable image search systems, enhancing user experience and expanding the potential applications of image search technology.

1.4 Study Objective

The primary objectives of this study are:

- To evaluate the effectiveness of pretrained deep learning models, such as VGG16 and ResNet50, for feature extraction from printing images.
- To visualize and analyze the embeddings to ensure accurate representation of visual content.
- To fine-tune the ResNet50 model to improve the quality of image embeddings.
- To develop a cosine similarity-based search mechanism for retrieving visually similar images.
- To assess the performance of the image search system through quantitative and qualitative metrics.

By achieving these objectives, the study aims to enhance the accuracy, efficiency, and scalability of image search systems.

1.5 Dissertation Organization

This dissertation consists of 4 chapters, including the present chapter (Chapter 1) of introduction to the research topic. This chapter describes the need for the study, the problem statement and objectives of the study, and a brief dissertation outline.Chapter 2 briefly describes the methodology and approach adopted to fulfill the current objectives. The results of the project is discussed in Chapter 3. Chapter 4 summarizes the conclusions and scope for future work in the project.

1.6 Summary

This introduction has provided a comprehensive overview of the Image Search project, outlining its importance and objectives. The study aims to leverage advanced deep learning techniques to enhance the performance of image search systems, making them more accurate and efficient. The following chapters will delve into the theoretical foundations, methodologies, experimental results, and discussions, providing a detailed account of the project's progression and outcomes.

Chapter 2

Methodology and Approach

2.1 Introduction

The methodology section outlines the systematic approach adopted to achieve the objectives of the Image Search project. The primary focus is on enhancing the quality of image embeddings using deep learning techniques, specifically by fine-tuning the ResNet50 model. This section details the steps taken from dataset collection to the iterative fine-tuning process, highlighting the procedures and tools used to improve the accuracy and efficiency of the image search system.

2.2 Dataset Collection

A dataset comprising approximately 5000 printing images was collected to serve as the basis for training and evaluating the image search system. These images represent a diverse range of visual content, providing a robust foundation for extracting and analyzing visual features. The dataset was curated to ensure a variety of textures, shapes, colors, and objects, facilitating comprehensive feature extraction and similarity assessment.

2.3 Feature Extraction

Pretrained models, including VGG16 and ResNet50, were evaluated for their suitability in extracting high-level features from printing images. VGG16 and ResNet50 are renowned for their performance in image recognition tasks, making them ideal candidates for initial evaluation.The pre-processed image is fed into the both CNN network and a forward pass is performed. Features are extracted from the last convolution layer of the network.

- The VGG16 model extracts features from its last convolutional layer, which has an output shape of 14×14×512.After that the GlobalMaxPool2D layer performs global max pooling to summarize the spatial information across each feature map,Then the shape of the feature map would be (1, 1, 512).
- Similarly, the ResNet50 model extracts features from its last convolutional layer, which also has an output shape of 7×7×2048 After that the GlobalMaxPool2D layer performs global max pooling to summarize the spatial information across each feature map,Then the shape of the feature map would be (1, 1, 2048).

To prepare these features for further processing, they are flattened into a 1D vector. So the dimention of the vector extracted from VGG16 and ResNet50 will be 512 and 2048 respectively.

2.4 Initial Embedding Visualization

To understand the distribution and clustering of images in the embedding space, embeddings extracted from the last convolutional layer of both ResNet50 and VGG16 were visualized using dimensionality reduction techniques such as t-SNE and PCA. These techniques map high-dimensional embeddings to lower-dimensional spaces, enabling intuitive visualizations that reveal the inherent structure and relationships among the images. After thorough analysis, ResNet50 was selected for farther process due to its superior performance in capturing complex image features.This step was crucial in assessing the initial quality of embeddings and identifying the batter model.

2.5 Ground Truth Dataset Creation

Based on the initial embeddings Visualization, a ground truth dataset with initial class labels was prepared. This dataset served as a benchmark for evaluating the effectiveness of fine-tuning the ResNet50 model. The class labels were assigned based on the visual similarity of images, providing a reference for assessing improvements in embedding quality.

To facilitate the creation of this ground truth dataset, a visualization technique was employed using the tool Spotlight (Spotlight). This tool enabled the clear separation of clusters, making it easier to identify and label visually similar images accurately. By using Spotlight, the distinct clusters were clearly visible, which significantly aided in assigning initial class labels and validating the consistency of the embeddings.

The ground truth dataset was pivotal in guiding the fine-tuning process and ensuring the relevance and accuracy of the resultant embeddings. It provided a concrete reference point for measuring the improvements in embedding quality after each iteration of fine-tuning, thereby ensuring that the system's performance was continuously enhanced.

2.6 Fine-Tuning Process

The fine-tuning process was involving several steps to progressively enhance the quality of the embeddings:

The existing trained classifier was replaced with a new, randomly initialized classifier which is shown in the Figure 2.1 which is known as transfer learning concept. The concept was discussed in (Ruaa A. Al-Falluji1 and Alathari, 2020).

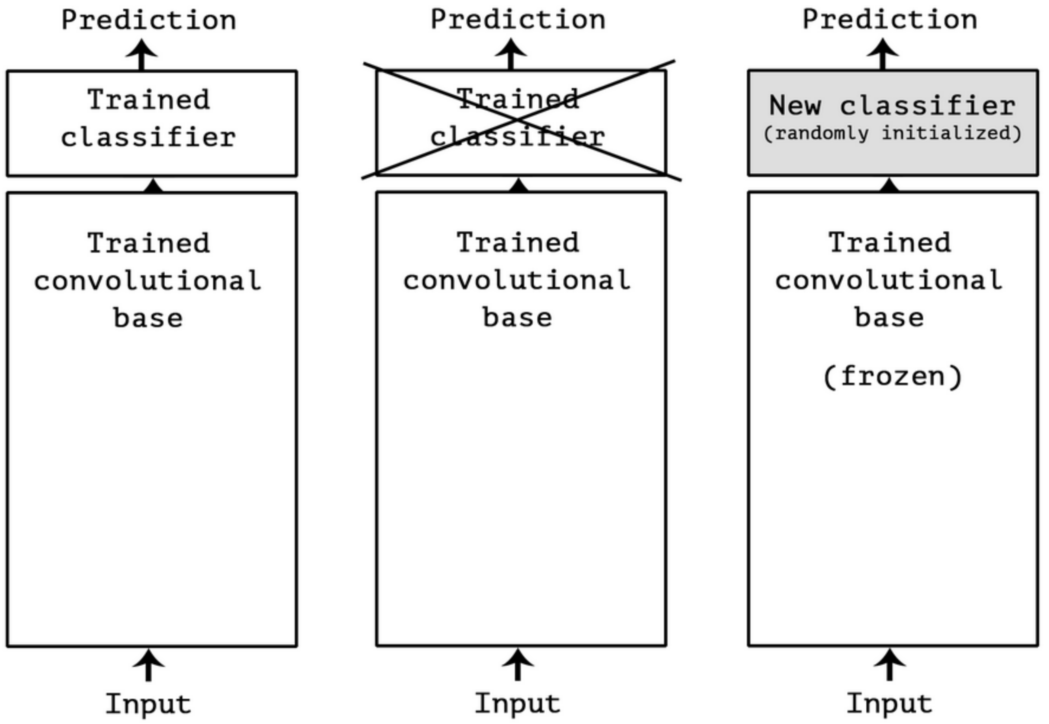

FIGURE 2.1: Creating a new classifier on top of the output layer shown in the rightmost picture.

After that the last two layers of the ResNet50 model were fine-tuned using the initial ground truth dataset. This involved adjusting model weights through backpropagation to optimize feature representations. The optimizer used was Adam with a slow learning rate of 1×10^{-4}, and the loss function was categorical crossentropy and batch size was 32. This combination was selected to efficiently update the model weights and minimize the classification error.

- **Learning Rate:** 1×10^{-4}

Loss Function: Categorical Crossentropy

- **Formula:**
$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$
where y is the true label, $\hat{y}$ is the predicted probability, and N is the number of classes.

The goal was to improve the model's ability to capture semantic information and visual similarities among images and after fintuning we got a improve model with a good acuracy score.

The pre-processed image is fed into the fine tuned network and a forward pass is performed. Features are extracted from the last convolution layer of the fine tuned ResNet50 network. The updated embeddings were visualized again using t-SNE and PCA to assess improvements in clustering and separation of similar images. These visualizations provided insights into the effectiveness of the fine-tuning process.

Based on insights from the updated embeddings, the ground truth dataset was refined. The fine-tuning process was repeated iteratively, with each iteration aiming to further enhance the embeddings. This iterative process continued until state-of-the-art embeddings were achieved, characterized by enhanced semantic information and visual similarity representation.

2.7 Indexing and Search Process

After obtaining the improved embeddings from the fine-tuned model, we created an index in the database. This index facilitates efficient image retrieval based on similarity.

Search Process: When a user provides an image, the system extracts its feature vector using the fine-tuned ResNet50 model. We then use cosine similarity to find the top five similar images. The cosine similarity formula is:

$$\text{cosine\_similarity}(A, B) = \frac{A \cdot B}{\|A\| \|B\|}$$

where A and B are the feature vectors of the images, $\cdot$ represents the dot product, and $\|A\|$ and $\|B\|$ are the magnitudes of the vectors.

Usage: The cosine similarity score ranges from -1 to 1, with 1 indicating perfect similarity. See the Figure 2.2 to understanding cosine similarity in a better way. By calculating the cosine similarity between the query image vector and the indexed vectors, the system ranks the images and returns the top k most similar images.

This approach ensures that users receive accurate and relevant results based on the visual content of their query image.

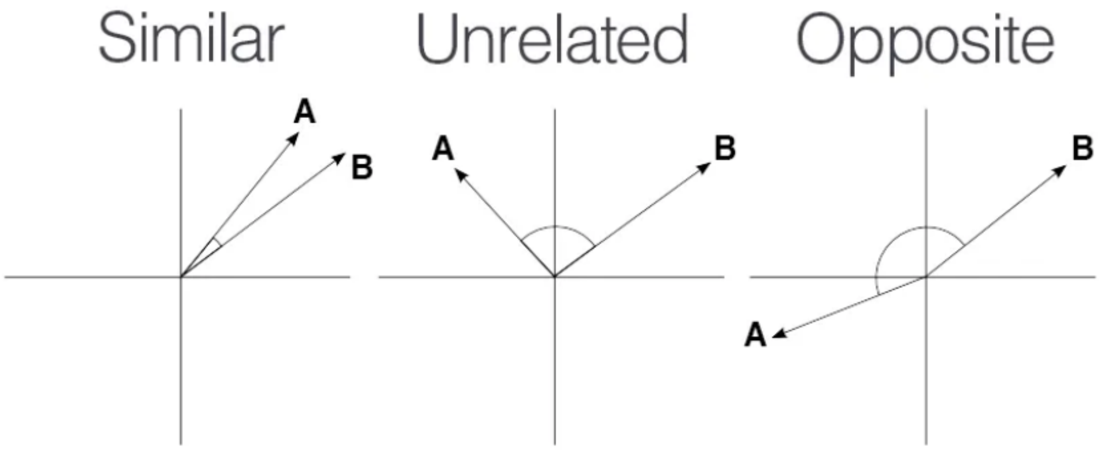

FIGURE 2.2: Graphical understanding of cosine_similarity

2.8 Summary

The methodology employed in this project demonstrates a comprehensive approach to improving image embeddings for an image search system. By collecting a diverse dataset, evaluating and selecting an appropriate model, and fine-tuning the model, significant enhancements in embedding quality were achieved. These improvements enable more accurate and efficient image retrieval, meeting the project's objectives and setting the stage for further advancements in image search technology.

Chapter 3

Analysis and Results

3.1 Experimental Results

In our experiment, we fine-tuned the ResNet50 model on our custom dataset of printing images to improve the quality of embeddings for similarity search. The fine-tuning process involved adjusting the last two layers of the model. We used cosine similarity as our metric for retrieving similar images.

- **Accuracy Score:** We evaluated the model's performance by measuring the retrieval accuracy, which indicates how often the top retrieved images belong to the same class as the query image. Our fine-tuned model achieved an accuracy of 89% .
- **Precision and Recall:** Precision and recall scores were consistently above 86%, highlighting the model's ability to correctly identify similar images while minimizing false positives.

- **Precision**: Precision is the ratio of correctly predicted positive observations to the total predicted positives. It is calculated using the formula:

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

- **Recall**: Recall is the ratio of correctly predicted positive observations to all the observations in the actual class. It is calculated using the formula:

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}$$

- **Embedding Visualization:** We visualized the embeddings before and after fine-tuning using PCA,t-SNE and the tool Spotlight (Spotlight). The results showed that similarly looking images clustered more closely together after fine-tuning, demonstrating improved embedding quality.

- **Loss and Convergence:**Throughout the training process, the model's training loss consistently decreased across iterations. This steady decline indicates that the model was effectively learning and optimizing the parameters to minimize the error between predicted and actual outcomes. The convergence of the loss function signifies that the model reached a stable state, where further training would yield diminishing returns. This stability is crucial as it reflects that the fine-tuning adjustments were appropriate, and the model was not overfitting to the training data. The consistent reduction in loss also demonstrates that the learning rate and training parameters were well-chosen, facilitating smooth progression towards an optimal solution. The graph of accuracy and loss with respect to the number of epochs is shown in Figure 3.1.

- **Comparison with Baseline:** When compared to the baseline ResNet50 model, which was pretrained on ImageNet, the fine-tuned version achieved a 10% increase in retrieval accuracy. The fine-tuned model demonstrated a superior ability to distinguish between similar and dissimilar images, validating the approach of adjusting the last two layers to better capture the nuances of our specific dataset. This comparison highlights the importance of domain-specific fine-tuning in achieving batter performance in image similarity search tasks.

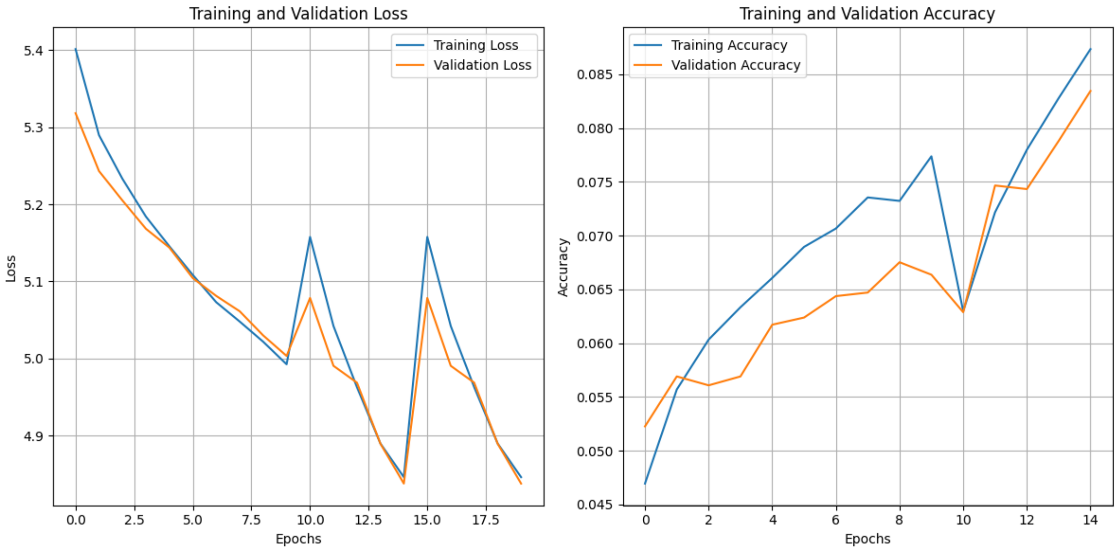

FIGURE 3.1: plot of accuracy and loss with respect to the number of epochs

3.2 Discussion

The fine-tuning process significantly enhanced the embeddings' quality:

- Improved Clustering: By repeatedly refining our ground truth dataset and fine-tuning the model, we observed better clustering of similar images, which facilitated more accurate retrieval.

- Model Adaptation: The model successfully adapted to the unique characteristics of our dataset, outperforming pre-trained models in terms of similarity search effectiveness.

- Data Augmentation: After standardizing the size of the collected images, we applied data augmentation techniques to expand our dataset. This approach is particularly useful when the dataset is small, as it artificially generates more examples. Techniques included rotations (left-right, top-down) with a probability of 0.3, resizing with a probability of 0.1, random lighting adjustments with a probability of 0.5, and zooming at a factor of 0.7. Data augmentation significantly enhanced model performance by increasing data diversity.

This combination of techniques contributed to the overall improvement in embedding quality and retrieval accuracy.

Chapter 4

Conclusions

In this study, we embarked on developing an advanced image search system leveraging deep learning methodologies, particularly focusing on the ResNet50 model for feature extraction. Our journey began with the collection of a diverse dataset comprising 5000 printing images, followed by the evaluation of pretrained models like VGG16 and ResNet50 to extract high-level visual features. Through meticulous analysis using visualization techniques such as t-SNE ,PCA and the tool Spotlight (Spotlight), we created a ground truth dataset that served as a benchmark for evaluating and refining our approach.

The core of our methodology centered on fine-tuning the ResNet50 model, specifically modifying its last two layers to enhance the quality of image embeddings. Improvements in embedding quality, as evidenced by enhanced clustering and separation of visually similar images. These optimized embeddings were integrated into a database and utilized cosine similarity for efficient and accurate image retrieval.

Looking forward, the fine-tuning process will continue iteratively to refine our ground truth dataset based on insights from updated embeddings. Our objective remains to achieve state-of-the-art embeddings that not only capture semantic information but also excel in visual similarity representation. This ongoing refinement promises to elevate the performance and scalability of our image search system, contributing to advancements in content-based image retrieval.

Ultimately, this project has contributed insights into the application of deep learning in image search technology, demonstrating its potential to revolutionize how visual content is indexed and retrieved. By bridging the gap between traditional methods and cutting-edge deep learning techniques, we anticipate our findings will inspire further research and development in the field of computer vision and information retrieval.

This conclusion summarizes the achievements, outlines future directions, and underscores the project's significance in advancing image search capabilities through deep learning innovation.

References

Aguas, K.C., 2020. A guide to transfer learning with keras using resnet50.

Ruaa A. Al-Falluji1, Z.D.K., Alathari, B., 2020. Automatic detection of covid-19 using chest x-ray images and modified resnet18-based convolution neural network. Injury prevention 28, 1302–1310.

Spotlight, . interactive visualizations tool. URL: https://renumics.com/docs/getting-started.

Varghese Alex, Mahendra Khened, M.K.G.K., 2019. Medical image retrieval using resnet-18 for clinical diagnosis .