Media Engineering and Technology Faculty
German University in Cairo

2D Arabic-based Dataset of Hand-Drawn Strokes

Bachelor Thesis

Author: Omar Hamdi Ebid

Supervisors: Dr.Wael Zakaria

Dr.Ahmed Abdelfattah

Submission Date: 19 May, 2024

This is to certify that:

(i) the thesis comprises only my original work toward the Bachelor Degree

(ii) due acknowledgement has been made in the text to all other material used

Omar Hamdi Ebid
19 May, 2024

Acknowledgments

First and foremost, I would like to thank Allah for giving me the strength, knowledge, ability and opportunity to complete this work satisfactorily. Without his blessings, this achievement would not have been possible. I would like to express my sincere gratitude to my supervisor, Dr. Wael Zakaria, for his continuous support, patience, motivation and immense knowledge throughout my bachelor's thesis. His guidance helped in laying the foundation of this research and thesis. My deep and sincere gratitude to my family for their continuous and unparalleled love, help and support. I am forever indebted to my parents for giving me the opportunities and experiences that have made me who I am. This journey would not have been possible if not for them, and I dedicate this milestone to them. Finally, it is a pleasure to thank my friends and colleagues who helped me in this thesis.

Abstract

With Sketch-Based Image Retrieval (SBIR) systems, users can use free-hand sketches as queries to find images in digital libraries. By bridging the gap between user intention and visual content, these technologies allow users to engage with digital picture databases in a more intuitive and natural way. The goal of this thesis is to develop and improve a deep-learning-powered SBIR system. Specifically, we will integrate an interactive sketching interface with a backend that uses a YOLOv8 neural network model to analyse and classify sketches.

A custom paint application that allows users to draw sketches, which are instantly analysed by the trained model to predict and display likely image categories, is the methodology used in two experiments to assess the system's performance, specifically its ability to generalise from training data to new, unseen sketches. Tracking metrics like loss and validation accuracy, the first experiment evaluated the training dynamics, and the second experiment tested the accuracy of the model on a different test dataset.

The algorithm does well at classifying sketches, as evidenced by the results, which show an accuracy of 89.94 percent for the top 5. Nonetheless, the accuracy of the top-1 was 54.79 percent, suggesting that the model's precision might be enhanced. The fields of interactive design and machine learning have greatly benefited greatly from the incorporation of real-time feedback and the visualisation of drawing sequences.

This study shows how machine learning may be incorporated into creative processes and emphasises the value of user-centred design in the development of successful SBIR systems. In order to better improve user engagement and system performance, the findings propose options for future research, especially in the areas of model optimisation and user interface enhancements.

Contents

Chapter 1

Introduction

1.1 Motivation and Objectives

FREE-HAND sketching is a universal communication and art modality that transcends barriers to link human societies. It has been used since ancient times,comes naturally to children before writing, and transcends language barriers. Sketches can convey many words or even concepts that are hard to convey in words. Free-hand sketch can be illustrative, despite its highly concise and abstract nature, making it useful in various scenarios such as communication and design. Therefore, free-hand sketching has been widely studied in computer vision and pattern recognition. In the realm of digital artistry, the ability to seamlessly capture and interpret hand-drawn sketches holds immense significance. These sketches, often characterized by their organic and expressive nature, serve as the foundation for countless artistic creations. To harness the potential of these forms effectively, it becomes essential to develop systems capable of efficiently identifying and utilizing hand-drawn sketches [11].

At the core of this lies the need to establish a seamless and normalized database of hand-drawn sketches. Such a database serves as a vital resource for training and testing recognition systems, leading the way for advancements in various fields, including digital art, machine learning, and human-computer interaction. Understanding this crucial need, the primary aim of this project is to construct an intelligent system designed to facilitate the collection of hand-free drawings.

1.2 Research Questions

A sketch-based image retrieval (SBIR) system must be developed, hence it is critical to investigate and improve the tools that promote interaction and improve data processing. In order to further the use of interactive systems and deep learning technologies in practical settings, this thesis seeks to answer two essential research questions:

Is it feasible to create an application that can be used to create and test datasets that produce free-hand sketches? The study demonstrates that it is possible to create a tool that can produce free-hand sketches, which may be used for dataset testing and data production. In addition to facilitating the creation of sketches, the interactive sketching interface created for this thesis incorporates vital features, including precise stroke analysis and real-time predictive feedback. These characteristics are necessary to produce extensive datasets that record a broad range of hand-drawn patterns, which are then used to train and improve deep learning models.

How can deep learning models for image retrieval be trained and tested using the generated sketches? The use of generated sketches improves the training of deep learning models by offering a wealth of diverse human sketching styles, which is essential for creating reliable recognition systems. They provide a useful gauge of how models function with actual data, which is crucial for verifying and assessing model accuracy. With the use of feedback from real-world user interactions, this real-time application enables models to repeatedly adapt and develop, improving both user experience and model correctness.

The results of this study show that creating a tool for creating free-hand sketches is not only feasible but also greatly advances the development of SBIR systems. The drawings made are essential to creating a variety of datasets that support deep learning models that are more precise and flexible, improving the usefulness and efficiency of image retrieval systems.

1.3 Outline

Chapter 1: Introduction

The issue of sketch-based image retrieval and its significance in the fields of computer vision and human-computer interaction are introduced in this chapter, which also sets the stage for the thesis. It lays out the primary research questions that direct the investigation, details the goals of the thesis, and provides a brief explanation of the methodology used. In order to aid the reader in following the upcoming chapters, this part also offers a summary of the thesis framework.

Chapter 2: Background

A thorough analysis of the pertinent literature is provided in this chapter. It addresses the state-of-the-art methods and tools for sketch-based image retrieval systems, emphasising the use of deep learning models for picture categorization and retrieval. The chapter examines numerous obstacles and developments in the field of user interaction designs for sketching interfaces, as well as reviewing earlier research in the area. The inadequacies in current technology that this thesis seeks to fill are highlighted in this survey of the literature.

Chapter 3: Methodology

The methodologies and processes used to create and assess the sketch-based image retrieval system are described in detail in this chapter. It explains the preparation of the dataset, including the methods used to gather and handle the sketches. The chapter also describes the training procedure, the integration of the sketching interface with real-time feedback mechanisms, and the technical details of the machine learning models (YOLOv8) that were utilised in the design and implementation of the retrieval system. Understanding the experimental design and the analytical methods used in the study requires reading this section.

Chapter 4: Results and Limitations

The results of the experiments carried out for the study are presented in the fourth chapter. It examines the effectiveness of the sketch-based image retrieval system and talks about the model's dependability, efficiency, and correctness in light of the data gathered. This chapter also critically evaluates the existing system's shortcomings, going into possible biases, variations in sketch interpretation, and technical difficulties that arose during the study. Key findings are illustrated with detailed visuals like tables and graphs.

Chapter 5: Conclusion and Future Work

The research findings are summarised in the last chapter, which also makes judgements regarding the system's efficacy and implications for the field of sketch-based picture retrieval. It explores the ramifications of the results and considers how the research issues have been approached. This chapter also suggests directions for future research, offering solutions to the problems found and new features or approaches that may be investigated to improve the system's functionality even more.

Chapter 2

Background

2.1 Introduction to Sketches

Sketches are simplified visual representations designed to convey ideas, concepts, objects, or scenes through quick, freehand strokes. They often exclude details to focus attention on vital features and conveniently communicate visual ideas. In the digital age, sketching has evolved due to the rise of digital drawing tablets and styluses. These tools copy traditional sketching techniques, allowing artists to draw directly into software applications. Pereira and Dias (2017) deduced that sketch-based interfaces connect the differences between visual communication and computer-aided design, providing greater creative freedom when compared to conventional interfaces. They also noted that these interfaces ease more intuitive engagement, which can lead to more creative solutions and exploratory designs [7].

2.1.1 Types of Sketches and Their Application

Sketches differ widely in purpose, especially in design and computer science. Rapid prototyping through sketches enables designers to brainstorm and clear ideas quickly before moving to more formal designs. In computer science, Lee and LaViola Jr (2011) evaluated the effectiveness of sketch-based interfaces for early conceptual 3D model design. They highlighted how sketches allow for creative exploration before developing formal 3D models. They deduced that designers could rehearse quickly and explore new concepts by combining basic shapes like circles and rectangles to represent objects [4]. These sketch-based interfaces rely on machine learning algorithms to recognize these simple shapes and classify them correctly. As a result, sketching can be used effectively for both prototyping and conceptualizing complex 3D models.

2.1.2 Freehand Sketch Recognition Systems

The style and abstraction of freehand sketches vary greatly, making them special problems for recognition algorithms. Because sketches differ greatly from photographic images, traditional pattern recognition techniques have difficulties in effectively interpreting them. In their evaluation of systems for recognising freehand sketches, Xu et al. (2022) outlined how deep learning models—more specifically, Convolutional Neural Networks (CNNs)—can use feature extraction to recognise the subtleties of freehand sketches. The significance of attention mechanisms and Recurrent Neural Networks (RNNs) in improving sketch categorization was highlighted by them, since these methods have the ability to prioritise crucial aspects while eliminating irrelevant data [11]. In order to enable models to recognise objects across a variety of styles and abstraction levels, recognition systems often involve feature extraction, form categorization, and semantic comprehension.

2.1.3 Advances and Challenges in Sketch Recognition

Cultural heterogeneity is one of the main obstacles to developing complete sketch recognition systems. Zhang et al. (2019) noted that sketches of everyday items, such as houses or animals, vary greatly between geographical areas, which makes it challenging for recognition algorithms to understand them reliably. For example, drawings of a house in Southeast Asia or sub-Saharan Africa may not resemble those in Europe at all [15]. The complexities of Arabic-speaking countries' cursive writing and ligatures provide particular difficulties for the classification and segmentation of strokes. Because the script is cursive, models must comprehend regionally specific semantic nuances and contextual forms. Style differences make recognition much more difficult, so in order to increase model accuracy, culturally relevant datasets are required. The abstraction and diversity of sketches present another difficulty. Sketches frequently lack the exact features necessary for recognition models to quickly identify certain objects. Because of this abstraction, models need to be trained to handle different levels of abstraction across drawings while still identifying important elements. Creating and annotating high-quality sketch collections is also essential to building successful generalizable recognition models. Deep learning developments, particularly in transformer-based models and attention mechanisms, have the potential to improve sketch identification systems; nonetheless, further study is needed to fully address the particular difficulties this area presents.

2.2 Deep Learning

Systems for sketch detection and retrieval have seen a substantial transformation in capabilities because of deep learning. In particular, the introduction of Deep Convolutional Neural Networks (DCNNs) and other advanced deep learning approaches has increased the efficiency and accuracy of sketch interpretation and classification. For precise recognition and retrieval applications, these networks excel in extracting rich feature representations from the sparse and abstract data typical of sketches.

The creation of techniques that span multiple domains and allow for the retrieval of detailed natural images from huge databases using sketching is a noteworthy accomplishment in this discipline. As an example, DCNNs can be trained with both image and sketch data, which improves the model's comprehension and interpretation of sketches as well as its performance in cross-domain retrieval scenarios. The networks are able to acquire strong, discriminative features because of this dual training strategy, which greatly enhances retrieval operations [10].

Furthermore, by teaching neural networks to map sketches and images into a shared feature space, deep learning frameworks that optimize sketch-based image retrieval have been presented. Reducing the domain gap between the sketch and the picture data is made possible by this mapping, which makes retrieval more precise and effective. In order to improve feature extraction and comparison capabilities and achieve better retrieval performance, these frameworks frequently use innovative designs that combine several neural network models [5].

The field of sketch recognition and retrieval has benefited from the development of techniques that dynamically learn and adapt to the differences in sketch style and detail, in addition to technological advancements. By utilising advanced neural network architectures that are capable of adapting to the distinct features of sketches, these techniques enhance the system's capacity to generalise from training data to real-world scenarios [14].

All things considered, the incorporation of deep learning methods has resulted in notable progress in the identification and retrieval of sketches, underscoring the potential of these technologies to completely change the way digital systems perceive and engage with hand-drawn images.

2.2.1 Deep Learning Techniques in Sketch Recognition

Utilising deep learning technologies has proven to be quite effective in the field of sketch recognition, adjusting to the inherent difficulties that the sparse and abstract characteristics of freehand sketches provide. Combining texture and form features using a Sequential Dual Deep Learning framework is an example of a sophisticated method. According to Jia et al. (2017), this technique greatly improves the recognition process by dynamically combining these unique qualities to accommodate the substantial heterogeneity present in sketches created by various people [3]. Specialised neural network topologies such as Sketch-a-Net, which provides significant gains over human performance, demonstrate further refinement. This model, designed by Yu et al. (2016), uses special network configurations and data augmentation procedures that are intended to capture the fundamental abstract qualities of sketches. It is specifically optimised for sketch data rather than photographic images [13].

Furthermore, promising results have been observed in the adaptation of traditional deep neural networks, which were trained on large picture datasets like ImageNet, to sketch recognition. Yang and Hospedales (2015) show how the constraints of standard models applied to sketches can be addressed by re-engineering these networks to concentrate on sketch-centric properties. This entails adjustments that improve the network's capacity to handle the stylized and less detailed shapes typical of sketches, hence improving performance on benchmarks tailored specifically for sketches [12].

Furthermore, the use of transfer learning strategies has revolutionised this field. Sert and Boyaci (2019) show that learned image attributes can be greatly leveraged for sketch detection by fine-tuning pre-trained convolutional neural networks on sketch datasets. Compared to methods that rely on manually created features, our approach not only produces higher accuracy rates but also allows for superior model generalisation across different drawing styles. The combination of these profound learning techniques signifies a noteworthy advancement in computational systems' capacity to decipher and scrutinise hand-drawn illustrations with an unparalleled degree of accuracy and adaptability [8].

All things considered, these breakthroughs highlight the versatility and resilience of deep learning techniques for sketch identification, mirroring continuous improvements that consistently push the limits of what robots can comprehend and engage with when it comes to artistic content created by humans.

2.2.2 Advancements in Hybrid CNN-RNN Architectures and YOLOv8 for Enhanced Image Recognition and Object Detection

By combining the unique strengths of both architectures to create potent hybrid models that tackle challenging problems in a range of applications, the convergence of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) represents a revolutionary advancement in the field of image recognition and analysis. When it comes to applications requiring strong visual recognition capabilities, CNNs are essential because of their reputation for processing visual input by efficiently identifying spatial details and hierarchical patterns in images. In order to provide a more nuanced knowledge of temporal or sequential dynamics—which is important in situations where context over time is critical—RNNs supplement this by being very good at managing sequences. This is because of their ability to retain history.

Hu et al.'s (2018) development of specialised models for gesture recognition serves as an example of this integration. Through the use of their novel attention-based hybrid CNN-RNN architecture, meaningful features are extracted from raw electromyography (sEMG) data by CNNs, and these features are then gradually interpreted by RNNs through sequential processing, which improves the model's ability to recognise complex gestures. This method raises the bar for gesture detection system accuracy and efficiency while also expanding the use of deep learning in human-computer interaction [2].

Advances in models like YOLOv8, the newest of a series of 'You Only Look Once' architectures, have simultaneously revolutionised the field of object identification. These models are ideal for real-time processing because they strike a good balance between speed and accuracy, making them ideal for applications that need quick responses, like traffic control and surveillance. According to Subhahan et al. (2023), YOLOv8 performs better than its predecessors in terms of object detection and recognition, with improved accuracy and faster processing times. Because of this, it works especially well in settings like automatic licence plate recognition systems, where security and efficiency depend on quick and accurate recognition [9].

The ongoing development of these technologies exemplifies an important trend in deep learning research: the combination of several neural network architectures and iterative model improvement results in complex systems that can handle an increasing range of challenging tasks. With every new development, artificial intelligence's potential applications across a wider range of sectors—from automation to healthcare and security—are expanded, pushing the envelope of what is possible. This dynamic interaction between technology and application is driving further advancements that may eventually allow robots to perform sequential decision-making tasks and picture identification on par with or even better than humans.

2.2.3 Sketch-based Image Retrieval

Sketch-Based Image Retrieval (SBIR) is a field that has been transformed by deep learning. In SBIR, sketches are used as queries to find visually related photos in large databases. This method is essential for content-based picture retrieval and digital asset management since it makes working with huge image collections easier and more approachable for users. SBIR systems can now successfully bridge the gap between the detailed information of images and the abstract nature of sketches by using advanced neural networks.

Lei et al. (2017), for example, improved SBIR performance by image-aided cross-domain learning, which uses deep learning models to learn discriminative features supplemented with real picture data, hence greatly improving the link between linked image-sketch pairs [5]. Furthermore, Wang et al. (2016) showed that deep convolutional neural networks (DCNNs) are very well suited for cross-domain image identification, efficiently retrieving natural images from big datasets through the use of drawings. According to Wang et al. (2016), their method uses both image and sketch data to train CNNs in tandem, and even basic Euclidean distance measurements on learned features can significantly outperform conventional SBIR methods [10].

These developments demonstrate how deep learning is revolutionising SBIR systems by enabling more accurate and effective retrieval procedures, which are becoming more and more important in a variety of digital content management applications.

Chapter 3

Methodology

This section describes the approach used to apply sophisticated deep learning techniques to Sketch-Based Image Retrieval (SBIR) systems in order to enhance their capabilities. Specifically, we highlight the usage of the YOLOv8 model. The setup, datasets, and testing tools used in this methodology chapter will be covered in detail, as well as the evaluation criteria used to measure the retrieval systems' performance. The goal of the research is to improve the accuracy and efficiency of SBIR systems by the integration of these components, thereby facilitating more efficient applications in content-based image retrieval and digital asset management.

3.1 Graphical User Interface

The development of the interactive user interface for the sketch-based image retrieval system began with the creation of a custom paint application designed to facilitate user interaction through an intuitive drawing interface. The application is equipped with essential tools such as a pencil for freehand drawing and an eraser, along a color palette that allows users to select any desired color. These features are seamlessly integrated to ensure that users can effortlessly create and modify their sketches. The initial framework of the paint application was adapted and expanded from an open-source project found on GitHub, authored by Vishesh Pandey. This foundational code served as a robust starting point for further customization and integration with the image retrieval system (Pandey, 2021) [6].

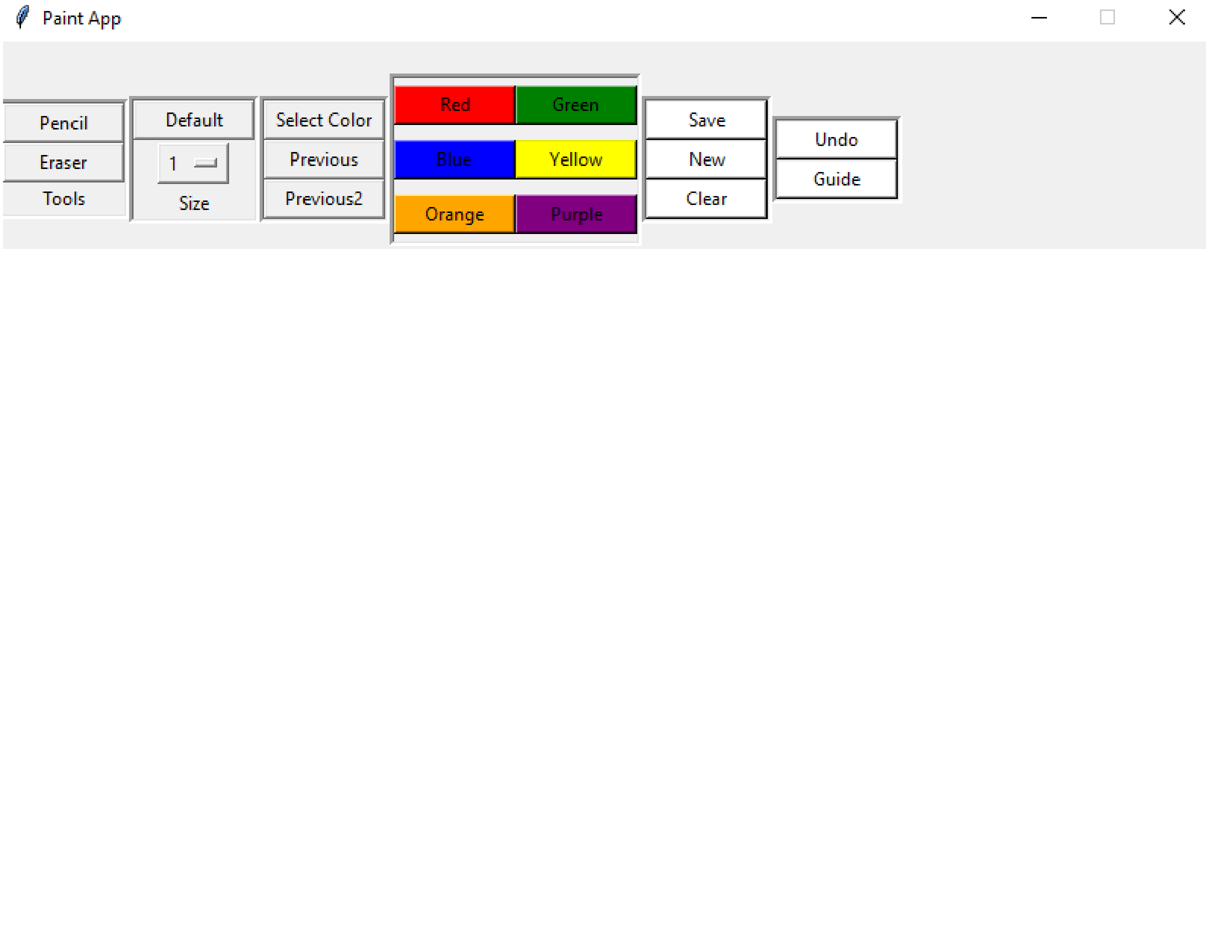

Figure 3.1: Paint App

An example of the paint application interface is shown in 3.1 . This interface includes various tools designed to enhance the user experience. The primary tools available are the pencil for sketching and the eraser for modifying drawings. Users can select from a range of colors using the color palette, which features basic colors such as red, green, blue, yellow, orange, and purple, ensuring a broad spectrum for creative expression. Additionally, the interface includes buttons for saving the current sketch, creating a new sketch, clearing the canvas, and undoing the last action.

The Guide button allows users to open a dataset file containing reference images, which can be displayed beside the drawing canvas to aid in sketching. This feature is particularly useful for users who require visual aids while drawing. The interface is designed to be user-friendly, with clearly labelled buttons and an intuitive layout that minimizes the learning curve, allowing users to focus on their creative process.

This graphical user interface forms the core of the sketch-based image retrieval system, providing an accessible and efficient platform for users to create and manage their sketches, which are then used for image retrieval tasks.

3.2 User Interface and Data Capture Enhancements

Following on from the creation of the user interface, the second section of the technique concentrates on the collection of data and the training of models, both of which are essential elements in improving the sketch-based picture retrieval system. The TU Berlin collection provided the dataset for this study, which included 20,000 distinct sketches evenly split over 250 object categories [1]. To expedite the creation and early testing stages, the scope of the initial tests was reduced, notably employing sketches from the 'dog' and 'cat' categories (TU Berlin Sketch Dataset, 2021).YOLOv8n-cls.pt, a pre-trained model well suited for classifying photos within these particular categories, was used to perform the model training. In order to standardise the input size and maximise processing efficiency, the photos were shrunk to 64x64 pixels. With 20 epochs of training, there were enough iterations to fine-tune the model parameters on the drawing data. This method made sure the model could efficiently pick out characteristics that distinguished one category from another, which is necessary for precise retrieval in the SBIR system. The durability of the retrieval system is supported by this meticulous setup of data preparation and model training, which guarantees that it operates at peak efficiency in real-world situations requiring prompt and accurate sketch-based searches.

Adding more features to the paint application to improve user involvement and integration with the underlying deep learning model was the third stage of developing the sketch-based image retrieval system. One important new feature that was added was the Save button, which let users save their completed sketches straight to their device. This feature makes it easier for users to process the photographs later on within the system, in addition to allowing users to save their work. Likewise, a brand-new module called Predict was created. This part is essential to the retrieval system since it predicts the content of the saved sketches by processing them. After the Predict module receives the file path of the stored image, it applies the learned YOLOv8 model to determine the likelihood that the sketch depicts a dog or a cat. This feature creates a seamless link between the user's artistic input and the machine learning component of the project by fusing advanced image classification with drawing skills. With this functionality integrated, the system can now intelligently engage with user-generated material and instantly provide feedback on the categorization results based on predictions from the deep learning model.

In the fourth stage of enhancing the sketch-based image retrieval system, significant modifications were made to the paint application to support detailed data capture and improve the user experience. The Save button functionality was expanded to not only save the sketches created by the users but also to generate a companion CSV file in the same directory. This CSV file, named stroke data file, records detailed information about each stroke made by the user, including the x and y coordinates, stroke number, start and end time of the stroke, and the duration of each stroke. This data is initially stored temporarily in a stroke data file and then written to a new CSV file that accompanies each saved image, providing a rich dataset for analyzing drawing behaviours and patterns.

Additionally, modifications were made to the "Clear" button, enhancing its functionality to not only clear the drawing canvas but also to reset the data within the stroke data file, ensuring that each new sketch starts with a clean data slate. An "Undo" button was also introduced, allowing users to remove the last stroke from the canvas. Correspondingly, this action also deletes the last stroke's data from the stroke data file, maintaining the accuracy of stroke data in sync with the visual content on the canvas.

Furthermore, a Guide button was implemented to augment the drawing experience by enabling users to open and view any image from a dataset containing 20,000 distinct sketches evenly distributed across 250 object categories. This feature allows users to select a reference image from the dataset to display alongside the paint application window, assisting them in crafting their sketches with a visual guide, thereby enhancing the fidelity and detail of the user-generated sketches.

Finally, real-time predictive feedback was added to the paint application during user drawing as part of an upgrade to the sketch-based image retrieval system. The program takes a snapshot of the drawing as it is right now every time a user adds a stroke to the canvas. The previously trained YOLOv8 model is then applied to this screenshot, predicting the sketch's most likely category by calculating the total number of strokes. The prediction is dynamically shown in the canvas's lower right corner, giving the user continual and instant feedback. This feature enables for the iterative improvement of sketches based on real-time AI assistance, which not only improves the interactive experience but also directly integrates machine learning insights into the creative process. In the area of sketch-based image retrieval, this predictive approach guarantees that users can edit and modify their drawings in response to the recommendations provided by the system, thereby bridging the gap between human creativity and machine intelligence.

3.3 Stroke Replay Visualization Tool

A new feature was created as the methodology's final improvement to see the drawing process from stroke data that was kept. This feature was added to a different project module, which asked the user to input the file path of a CSV file containing stroke data via an interactive canvas interface. When you enter the file path and use the Start Plotting command, the application reads the coordinates from the CSV file and starts to replicate each stroke on the canvas one at a time.

With a purposeful three-second break in between each stroke, the drawing of each stroke is animated to match the original time and sequence of the strokes. This delay gives viewers the opportunity to clearly see the steps taken and methods applied in each drawing, giving them insight into the different ways that different people approach sketching different objects. This feature is very helpful for comprehending drawing dynamics and can be used by researchers examining the cognitive processes involved in drawing or by artists displaying artistic skills for educational purposes.

article graphicx array longtable float placeins

Chapter 4

Results & Limitations

4.1 Dataset Description

The dataset for the study was meticulously constructed to analyze the capabilities and performance of the sketch-based image retrieval system. The data collection involved eight participants, each tasked with creating sketches of nine different objects, resulting in a total of 72 unique sketches. Participants were instructed to draw each object with strokes that were both intentional and meaningful, adhering to the principles of gestalt continuity. This approach ensures that the sketches represent realistic scenarios where strokes are used to convey significant parts of each object, rather than just outlining shapes.

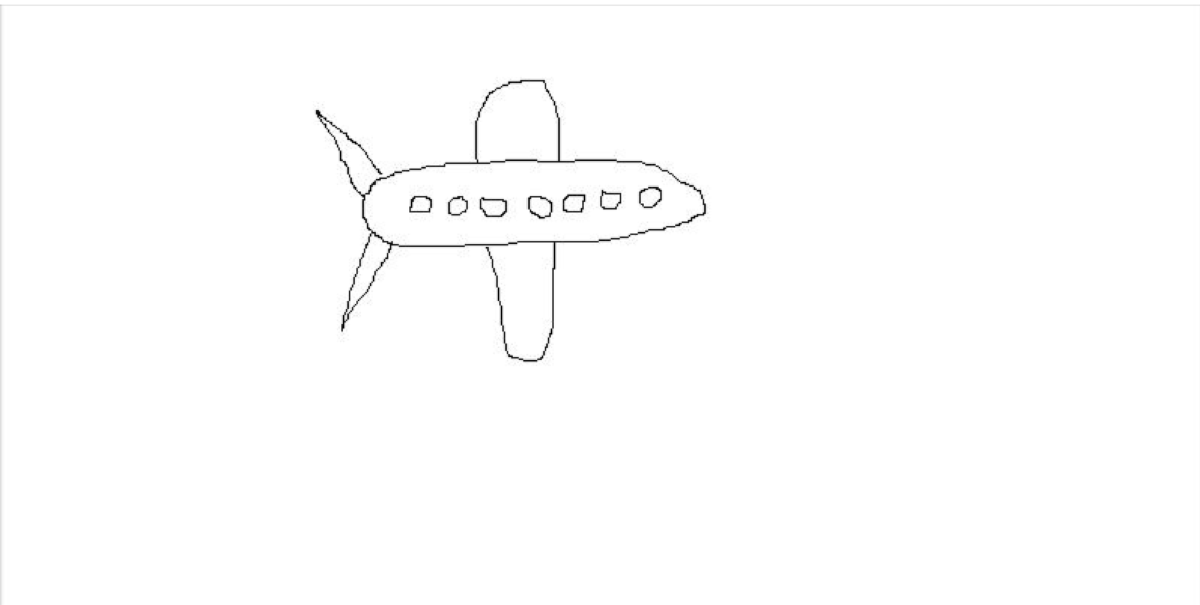

Figure 4.1: Airplane Dataset Example

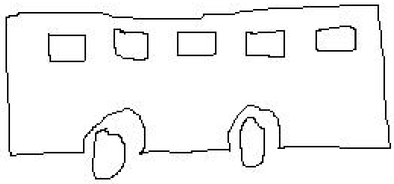

Figure 4.2: Bus Dataset Example

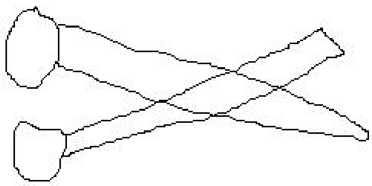

Figure 4.3: Scissors Dataset Example

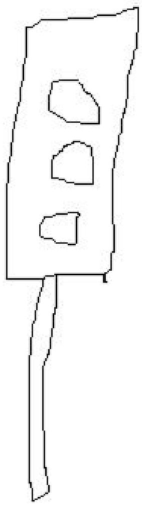

Figure 4.4: Traffic Light Dataset Example

For example, included in the dataset are sketches such as an airplane, a bus, a pair of scissors, and a traffic light. These sketches demonstrate the variety of objects participants were asked to depict and the range of complexity involved in their representations. Each sketch, such as the airplane in figure 4.1, the bus in figure 4.2, the scissors figure in 4.3, and the traffic light figure in 4.4, is characterized by its simplistic yet clear lines, highlighting how participants captured essential features with minimal strokes. Such examples showcase the balance each participant maintained between simplicity and detail, crucial for training the system to recognize diverse sketch styles accurately.

Every item drawing had to strike a compromise between simplicity and detail. For example, participants could design a rectangle as a single, continuous curve or as a combination of several distinct stroke sections. This variety in drawing approach is crucial for assessing the system's ability to identify and comprehend designs made up of several stroke methods.

The structured dataset created from these sketches was divided into Train and Val subfolders. This separation was essential to the training process as it allowed the model to learn from one collection of images termed train and test its predictions and new-found knowledge on another set of images termed val. The YOLOv8 model, which was trained for 20 training epochs and was designed to process photos reduced to 64x64 pixels, necessitates that the dataset be organized in a specific way.

This dataset is the main component of the experiment; it evaluates how well the paint program and the backend model identify and retrieve images from user-generated sketches.

4.1.1 Analysis of Model Training Performance

An important part of this research's first experiment was assessing the machine learning model's training dynamics. In order to accomplish this, a specialised script called plot metrics was created in order to evaluate and display the performance metrics accumulated throughout the model training stage. Examining the efficacy and efficiency of the model's learning process over time requires the use of this script.

Training and validation data are automatically extracted by the plot metrics script from a CSV file created during the model training process. In addition to recording top-1 accuracy for the validation set at each training epoch, this file also logs comprehensive records of training and validation losses, giving a comprehensive summary of the model's performance during training.

Using Python charting and data manipulation modules, the script generates two primary visual representations:

Visualization of Accuracy: The validation accuracy over the epochs is displayed as a percentage in the first plot in figure 4.5. This figure is important because it shows how the model can generalise to new data, which is essential for using it in real-world situations.

Loss Visualization: The training and validation losses are shown as a function of the number of epochs in the figure 4.6. The convergence or divergence of these metrics on this graph indicates whether the model is overfitting or underfitting.

In addition to providing insight into the model's learning trajectory, these visualisations help decision-makers make well-informed choices on how best to modify the training parameters in order to maximise overall performance. By means of methodical visual examination, perceptions of the model's behaviour during training are obtained, augmenting comprehension of its functionalities and constraints.

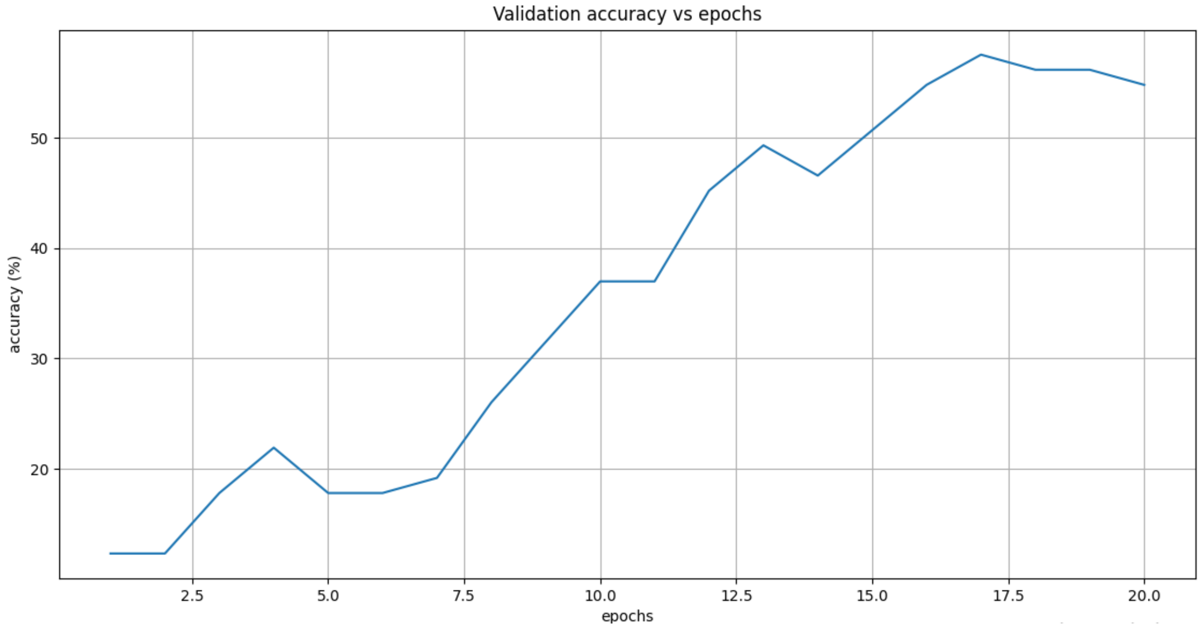

Figure 4.5: Validation Accuracy

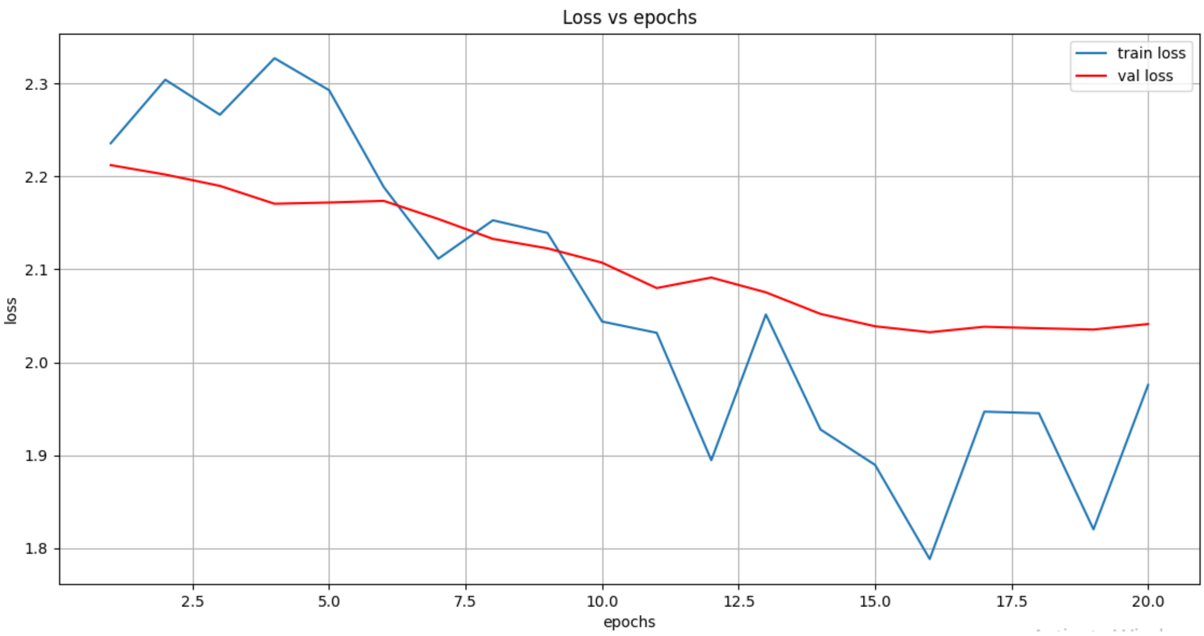

Figure 4.6: Loss vs Epochs

4.1.2 Model Evaluation on Test Data

The trained model was subjected to a critical evaluation in the next step of the study in order to gauge its efficacy and accuracy in practical application settings. A new function that was added to the software's predictive component, to measure the model's performance on an independent test dataset, made this evaluation easier.

The evaluatemodelontestset function automates the process of feeding test data into the model and extracting metrics for evaluation. This function is essential because it gives an unbiased assessment of the model's correctness, which is a critical sign of its dependability and effectiveness in real-world applications. In order to ensure that the assessment is accurate, the photos are processed at a resolution of 64x64 pixels, which is consistent with the model's training settings, ensuring that the assessment conditions are aligned with the training setup.

This experiment's main objective was to:

Validate the Model's Generalization: We can evaluate the model's ability to generalise from the training set to new unseen data by putting it to the test on data that it hasn't seen during training.

Put Performance Metrics in Numbers: This function gathers performance metrics, which provide a quantitative basis to assess the efficacy of the model. These metrics include accuracy, among other statistical measurements. Before the system is put into practice, this thorough evaluation procedure helps find possible areas for improvement and validates the model's capabilities, guaranteeing that the system is stable and dependable.

Table 4.1: Model Training Performance Metrics

Epoch	**metrics/accuracy_top1**	**metrics/accuracy_top5**
1	0.13239	0.56164
2	0.13239	0.57534
3	0.17808	0.57534
4	0.21918	0.65753
5	0.17808	0.64834
6	0.17808	0.69863
7	0.19178	0.75342
8	0.26027	0.75342
9	0.31507	0.80822
10	0.36986	0.82192
11	0.36986	0.82192
12	0.45205	0.86301
13	0.46575	0.83562
14	0.50685	0.86301
15	0.54795	0.87671
16	0.54795	0.87671
17	0.56164	0.89041
18	0.56164	0.89041
19	0.54795	0.89041
20	0.54795	0.89041

4.2 Results Analysis and Discussion

In the first experiment, the machine learning model's performance was assessed in terms of validation accuracy and training dynamics. The results of the training procedure, as shown in the accuracy and loss curves, are covered in this section

Validation Accuracy figure 4.5: The validation accuracy of the model showed a progressive improvement across the epochs. Initially, the accuracy started at around 20 percent, which is indicative of a model beginning to learn from a relatively complex dataset. As the epochs advanced, there was a consistent upward trend in accuracy, reaching a peak of approximately 50 percent by the 20th epoch. This trend suggests that the model was effectively learning and adapting to the nuances of the dataset, albeit with room for further optimization to achieve higher accuracy.

Figure 4.6 shows the training and validation losses. The training and validation loss curves provide information about the generalisation capacity and learning consistency of the model. A model that is successful in minimising the error between its predictions and the actual data will show a steady drop in the training loss. On the other hand, there were some variations in the validation loss; these were especially noticeable between epochs 12 and 15, when a sizable increase was recorded. This may point to problems like overfitting or an abnormality in the validation data that prevented the model from generalising for a while.

The analysis of the training and validation losses, along with the accuracy, provides a comprehensive view of the model's performance. The increasing validation accuracy coupled with decreasing training loss confirms that the model was learning effectively. However, the volatility observed in the validation loss highlights potential areas for improvement. Addressing these could involve techniques such as parameter tuning, the introduction of regularization methods to combat overfitting, or revising the data augmentation strategies to enhance the model's robustness.

The results from this experiment lay a foundational understanding of the model's behavior in a controlled training environment and set the stage for subsequent experiments aimed at refining the model's architecture and training regimen to better handle the complexities of the dataset.

The purpose of the second experiment was to evaluate the model's performance with fresh, untested data. This was a crucial stage in assessing the trained machine learning model's capacity for generalisation. The experiment produced a confusion matrix and comprehensive metrics that shed light on the model's accuracy and capacity for accurate image classification.

Metrics for Model Performance:

Top-1 Accuracy: At 54.79 percent, the model's Top-1 accuracy was attained. For well over half of the test set, this statistic shows that the model accurately identified the sketch's category on its first attempt.

Top-5 Accuracy: With a Top-5 accuracy of 89.94 percent, it can be inferred that approximately 90 percent of the time, the correct category was among the top five predictions. Although more work is required to improve the accuracy of the model's initial estimations, the high accuracy of the Top-5 indicates the model's potential reliability.

Assessment Findings:

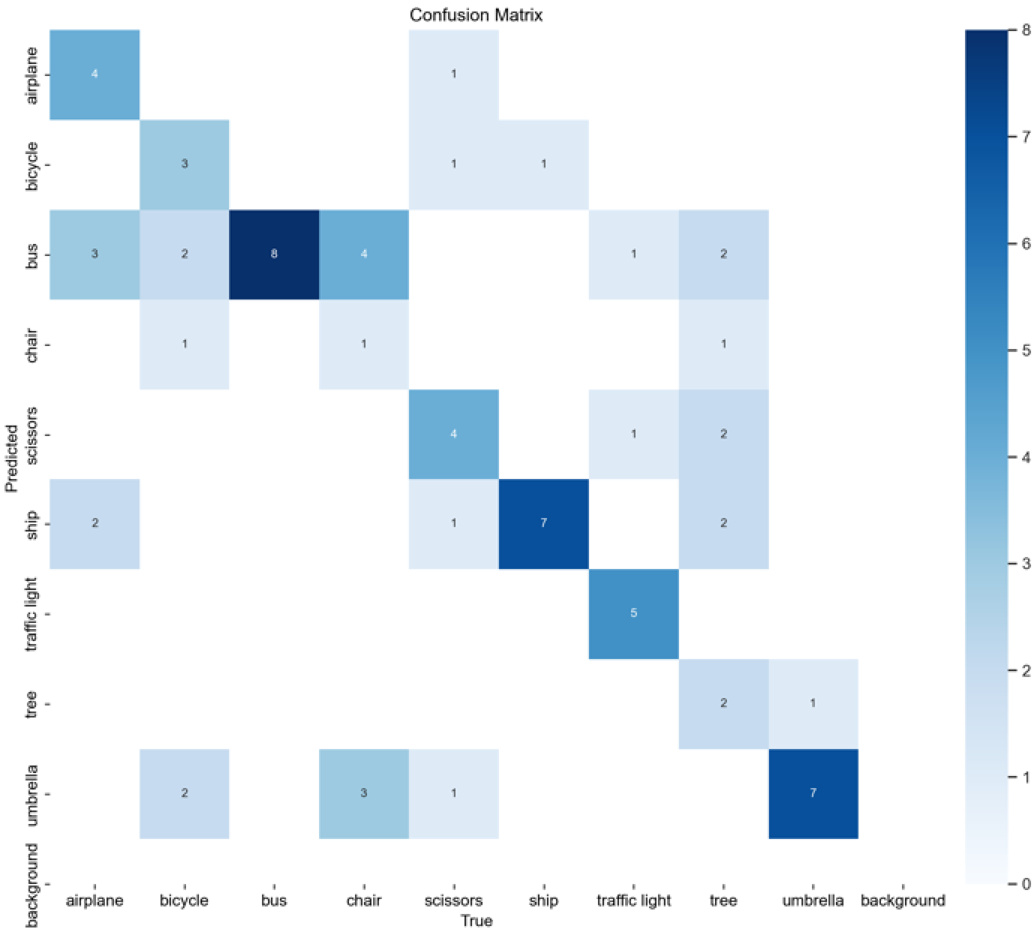

Figure 4.7: Model Accuracy

During the experiment, a confusion matrix was created to fully comprehend the categorization performance of the model. This matrix is essential for identifying which categories are more commonly confused, thereby aiding in model improvement or training adjustments.The confusion matrix in figure 4.7 reveals several key insights:The model correctly identified 8 instances of the bus category, which is the highest number of correct predictions among all categories. Airplane sketches were correctly classified 4 times but were occasionally confused with the bicycle and bus categories. The bicycle category had 3 correct classifications but also showed confusion with bus and scissors.The model identified scissors correctly 4 times. However, this category also showed some confusion with ship and bus sketches.Traffic light sketches were identified correctly 5 times, with minor confusion observed with other categories. Umbrella sketches were correctly classified 7 times, indicating strong performance in this category.The ship category showed 7 correct identifications but was occasionally confused with scissors and bus. This detailed analysis of the confusion matrix helps pinpoint specific areas where the model's performance could be improved, such as reducing the confusion between similar categories. Understanding these nuances is crucial for refining the model and enhancing its overall accuracy.

Efficiency and Inference Speed: The system's efficiency is demonstrated by the model's inference speed, which is 7.5 ms per image and appropriate for real-time applications. The model's practical usefulness for implementation in interactive applications such as the sketch-based image retrieval system is highlighted by its rapid processing capabilities and the fact that no extra time is needed for pretreatment or post processing for each image.

The experiment's outcomes highlight the model's abilities to categorise sketches from a variety of categories with a respectable degree of accuracy. To improve the precision of the model, more optimisation is necessary, as seen by the moderate Top-1 accuracy. Increasing the size of the training dataset, putting more effective data augmentation approaches into practice, or improving the model architecture are some possible strategies.

The high Top-5 accuracy indicates that even though the model might not always correctly identify the category at first, it probably makes the right suggestion in one of its later tries. This knowledge is useful for applications that allow the user to choose from a variety of alternatives.

Chapter 5

Conclusion & Future Work

5.1 Conclusion

The development and operational testing of a sketch-based image retrieval (SBIR) system that makes use of cutting-edge deep learning techniques to improve user interactions with digital image databases have been successfully explored in this thesis. One way to effectively bridge the gap between machine interpretation skills and user-generated sketches is to integrate a YOLOv8 neural network model-powered, user-friendly sketching interface with real-time prediction feedback.

The study undertaken yielded significant findings about the training dynamics of deep learning models tailored for drawing recognition. The trials showed that the algorithm can achieve a respectable degree of accuracy, especially when it comes to correctly classifying sketches within the top-5 predictions, achieving an accuracy of 89.94 percent. This result highlights how well technology can help people navigate and retrieve digital images based on their sketching.

Furthermore, the addition of real-time feedback to the sketching interface has greatly improved the system's usability, resulting in a more interactive and useful experience for users. By offering instantaneous visual feedback in response to input, this feature not only improves the user experience but also advances both theoretical and applied knowledge of interactive systems. Replaying drawing sequences also provides insightful information about the drawing process that can be useful for teaching and for deciphering user behaviour in interactive systems.

In conclusion, this thesis shows how deep learning can be incorporated into real-time applications that react to user-generated material, making a significant contribution to the domains of interactive systems and artificial intelligence. The effective setup and assessment of this SBIR system demonstrate how these technologies can be used to improve digital interactions and make digital image content more accessible.

5.2 Future Work

This thesis's work has established a strong basis for the creation and application of a deep-learning-based sketch-based image retrieval system. Still, there are a number of directions that could be pursued in the future to improve the system's functionality and broaden its uses.

Future work should focus on utilising the saved datasets, which contain precise timestamps for every stroke. Rich temporal data from these datasets can be utilised to conduct a more thorough analysis of the drawing process. This temporal data can be used in future studies to create models that can recognise static sketches and comprehend the time and order of strokes. This might result in the development of more complex sketch identification algorithms that take drawing dynamics into account, which could increase accuracy and provide information about user behaviour and drawing patterns.

Furthermore, integrating more advanced neural network architectures and experimenting with various deep learning techniques could yield improvements in the system's performance. For instance, exploring Recurrent Neural Networks (RNNs) or transformer models might enhance the system's ability to process and interpret temporal data effectively.

Increasing the number of participants and diversifying the items in the dataset can aid in the development of a more robust and generalised model. This would entail making sure that the dataset reflects a diversity of drawing styles and levels of complexity in addition to gathering new sketches.

Further work may also concentrate on enhancing the sketching application's user interface. More sophisticated drawing tools, adaptable functionality, and improved device and application integration are possible enhancements. This would increase the system's adaptability and user-friendliness and promote wider acceptance and utilisation.

Lastly, the creation of real-time collaborative sketching features that allow numerous users to draw and communicate at the same time may create new opportunities for applications in both education and the workplace. By fostering teamwork and collaborative creation, these qualities would increase the usefulness of the sketch-based picture retrieval system.

To summarise, further research endeavours will leverage the comprehensive datasets gathered, investigate sophisticated deep learning methodologies, broaden the dataset, enhance the user interface, and create collaborative functionalities. These initiatives will help to improve the sketch-based image retrieval system over time, increasing its accuracy, adaptability, and user-friendliness.

Appendix

List of Figures

Bibliography

[1] Timothy Eitz. Classify sketch project, 2024. Accessed: 2024-05-17.

[2] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, and W.-d. Geng. A novel attention-based hybrid cnn-rnn architecture for semg-based gesture recognition. *PLoS ONE*, 13, 2018.

[3] Q. Jia, M. Yu, X. Fan, and H. Li. Sequential dual deep learning with shape and texture features for sketch recognition. *ArXiv*, abs/1708.02716, 2017.

[4] J. Lee and J. J. LaViola Jr. An initial evaluation of a sketch-based interface for early conceptual design of 3d models. *Computers Graphics*, 35(3):594–609, 2011.

[5] J. Lei, K. Zheng, H. Zhang, X. Cao, N. Ling, and Y. Hou. Sketch based image retrieval via image-aided cross domain learning. In *2017 IEEE International Conference on Image Processing (ICIP)*, 2017.

[6] Vishesh Pandey. Paint app using tkinter, 2021. Accessed: 2024-05-17.

[7] Alfredo Pereira and João Dias. Sketch interfaces: bridging the gap between sketch recognition and visualization. *Journal of Visual Languages & Computing*, 44:72–87, 2017.

[8] M. Sert and E. Boyaci. Sketch recognition using transfer learning. *Multimedia Tools and Applications*, 78:17095–17112, 2019.

[9] D. A. Subhahan, S. R. Divya, U. K. Sree, T. Kiriti, and Y. Sarthik. An efficient and robust alpr model using yolov8 and lprnet. In *2023 International Conference on Recent Advances in Information Technology for Sustainable Development (ICRAIS)*, 2023.

[10] X. Wang, X. Duan, and X. Bai. Deep sketch feature for cross-domain image retrieval. *Neurocomputing*, 207:387–397, 2016.

[11] Peng Xu, Timothy M. Hospedales, Qian Yin, Yi-Zhe Song, Tao Xiang, and Li Wang. Deep learning for free-hand sketch: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[12] Y. Yang and T. M. Hospedales. Deep neural networks for sketch recognition. *ArXiv*, abs/1501.07873, 2015.

[13] Q. Yu, Y. Yang, F. Liu, Y-Z. Song, T. Xiang, and T. M. Hospedales. Sketch-a-net: A deep neural network that beats humans. *International Journal of Computer Vision*, 122:411–425, 2016.

[14] H. Zhang, P. She, Y. Liu, J. Gan, X. Cao, and H. Foroosh. Learning structural representations via dynamic object landmarks discovery for sketch recognition and retrieval. *IEEE Transactions on Image Processing*, 28:4486–4499, 2019.

[15] X. Zhang, X. Li, Y. Liu, and F. Feng. A survey on freehand sketch recognition and retrieval. *Image and Vision Computing*, 89:67–87, 2019.