Electronics and Computer Science
Faculty of Physical Sciences and Engineering
University of Southampton

Daniel-Dragos Braghis

12 December 2023

Conversational Image Search:
A sketch-based approach

Project supervisor: Haiming Liu

A project progress report submitted for the award of
BSc Computer Science

Statement of Originality

- I have read and understood the ECS Academic Integrity information and the University's Academic Integrity Guidance for Students.
- I am aware that failure to act in accordance with the Regulations Governing Academic Integrity may lead to the imposition of penalties which, for the most serious cases, may include termination of programme.
- I consent to the University copying and distributing any or all of my work in any form and using third parties (who may be based outside the EU/EEA) to verify whether my work contains plagiarised material, and for quality assurance purposes.

You must change the statements in the boxes if you do not agree with them.

We expect you to acknowledge all sources of information (e.g. ideas, algorithms, data) using citations. You must also put quotation marks around any sections of text that you have copied without paraphrasing. If any figures or tables have been taken or modified from another source, you must explain this in the caption and cite the original source.

I have acknowledged all sources, and identified any content taken from elsewhere.

If you have used any code (e.g. open-source code), reference designs, or similar resources that have been produced by anyone else, you must list them in the box below. In the report, you must explain what was used and how it relates to the work you have done.

I have not used any resources produced by anyone else.

You can consult with module teaching staff/demonstrators, but you should not show anyone else your work (this includes uploading your work to publicly-accessible repositories e.g. Github, unless expressly permitted by the module leader), or help them to do theirs. For individual assignments, we expect you to work on your own. For group assignments, we expect that you work only with your allocated group. You must get permission in writing from the module teaching staff before you seek outside assistance, e.g. a proofreading service, and declare it here.

I did all the work myself, or with my allocated group, and have not helped anyone else.

We expect that you have not fabricated, modified or distorted any data, evidence, references, experimental results, or other material used or presented in the report. You must clearly describe your experiments and how the results were obtained, and include all data, source code and/or designs (either in the report, or submitted as a separate file) so that your results could be reproduced.

The material in the report is genuine, and I have included all my data/code/designs.

We expect that you have not previously submitted any part of this work for another assessment. You must get permission in writing from the module teaching staff before re-using any of your previously submitted work for this assessment.

I have not submitted any part of this work for another assessment.

If your work involved research/studies (including surveys) on human participants, their cells or data, or on animals, you must have been granted ethical approval before the work was carried out, and any experiments must have followed these requirements. You must give details of this in the report, and list the ethical approval reference number(s) in the box below.

My work did not involve human participants, their cells or data, or animals.

ECS Statement of Originality Template, updated August 2018, Alex Weddell aiofficer@ecs.soton.ac.uk

Abstract

This paper introduces DoodleShop, a forward-thinking conversational image search assistant designed for searching online products that emphasises visual diversity that can't be adequately expressed through words. Such limitations of the traditional keyword-based searches are overcome through the conversational search approach, particularly in the sphere of image retrieval.

The proposed modular architecture integrates a state-of-the-art Language Model with the latest Stable Diffusion technologies in the image generation fields to offer users more accessible and accurate image search. Preliminary results demonstrate enhanced outcomes even in an untuned state significantly overperforming traditional keywords-based image search methods and promise a textless environment for groups with communication disabilities.

To qualitatively assess these improvements, a user study will be conducted to compare user satisfaction and attention metrics when using text-only, mixed, and sketch-only environments for search. DoodleShop aims to showcase the feasibility of sketch as a conversational search medium by seamlessly integrating various components and providing a framework for designing loosely coupled ML-powered agents for image retrieval.

Table of Contents

1. Background and Literature

1.1. Conversational Image Search

Conversational search has steadily garnered attention as the next step from traditional keyword-based search in information retrieval (IR) (Adewumi et al., 2022). A fundamental challenge of using keywords is the emergence of a knowledge gap during the information-gathering stage between users and machines, leading to the formulation of vague one or two-word queries in over 50% of instances (Spink et al., 2001). This discrepancy arises due to the users lacking domain expertise to provide relevant keywords, coupled with search systems often operating with incomplete metadata during the IR process.

Conversational search methods strive to offload the keyword generation process to a machine learning model by identifying user intent through natural language conversations and then leveraging domain expertise "to help users clarify their needs by asking appropriate questions from the users directly (Zhang et al., 2018, p.1)." Following this phase, the Feedback First and After strategies are implemented for result assessment in which the user either provides feedback on the search query before receiving results or the search space is explored collaboratively with the users, ideally employing both approaches (Aliannejadi et al., 2021).

Nevertheless, in instances when the search medium is switched to images, additional complexities arise for conversational search. Image data often lacks comprehensive embedding to describe the image and frequently omits important details of the scene or object represented (Bertasius et al., 2015). Linking the user intent to the image features is a complex task contingent upon outside factors such as the user's English proficiency or contextual ambiguity (Keyvan et al., 2022). Most research focuses on tackling the "conversational" facet of the problem, leaving insufficient attention directed toward image retrieval (Nie et al., 2021). The LARCH approach emerges as a viable solution that splits the task into query representation learning, multi-form knowledge modelling, and image representation learning to then fuse the three to estimate search result ranking scores for candidates (Nie et al., 2021).

Furthermore, conversational search lacks robust datasets, limiting the objective comparison of implementations. Consequently, it relegates the practical analysis of conversational models to qualitative user studies (Al-Thani, 2023).

1.2. Conditional Control in Text-to-Image

Sketch-Based Image Retrieval (SBIR) is a relatively nascent domain within IR that found practicality in the problem of retrieving images with high detail (FG-SBIR) as the domain gap is substantial with few methods attempting to bridge it, as highlighted by (Zhang et al., 2022). Their proposed solution involves a Dual-GAN trained on pseudo sketches generated from the image in conjunction with hand-drawn ones, but it is noteworthy that their approach still relies on human annotation during training. Another promising avenue is leveraging deep neural networks, as demonstrated in Mohian et al.'s (2022) work on retrieving UI screens.

While augmenting the image data with a textual prompt and using encoders in conjunction with a GPT achieved notable performance compared to zero-shot models (Dey 2019), it faces challenges with complex text prompts and sketches with scale mismatch (Sangkloy et al., 2022).

The effectiveness of this approach is further underscored by evidence suggesting that sketch SBIR benefits from "feature disentanglement to optionally combine multiple modalities, and (ii) support both discriminative and generative tasks (Chowdhury, 2023, p.4)." This acknowledgment highlights the complexity of challenges faced by SBIR when not only considering the integration of modalities but also satisfying both the requirements of discriminative and generative tasks. It is essential to address the interplay between image representations, textual prompts, and specific demands of SBIR applications for a more comprehensive and effective approach to image retrieval in this rapidly evolving domain.

1.3. SignWriting

Sign Language and SignWriting are two interrelated means of communication developed to assist deaf and hard-of-hearing individuals. Signwriting, unlike popular American Sign Language transcription systems, is a comprehensive writing system that has become an important part of "the cultural and linguistic expression of deaf people (Kato, 2008, p.113)." An important feature of the adoption of this writing system in the deaf community is the integration with the local writing system to seamlessly translate between glyphs and text. Despite creators making symbols computer-vision friendly and demonstrating effective classifiers for them (Stiehl et al., 2015), recent research evaluating performance with contemporary models is lacking. Moreover, the training set is yet to be extended to the entire symbol dataset. This presents a future research opportunity to enhance the accessibility and usability of SignWriting as a medium of human-computer interaction, particularly in the context of Large Language Models like ChatGPT.

2. Proposed Approach

In response to the aforementioned challenges, a sample problem was put forward. The objective is to develop a conversational image search assistant tailored to help users find shopping products online. The product categories of interest include items that exhibit significant visual diversity such as furniture, pottery, and clothing with features effortlessly expressed through sketches. Conversely, products characterized by banality and lacking intricacy such as food items, stationary, posters, jewellery, or objects bearing textual elements, were excluded from consideration. Notably, this exclusion does not imply the incapacity of the proposed solution to produce satisfactory search outcomes, although optimisation for such products is not the focus. Additionally, a secondary challenge involves accommodating alternative communication with the assistant beyond conventional typed text messages.

The proposed solution is named DoodleShop for the user's convenience with the following architecture:

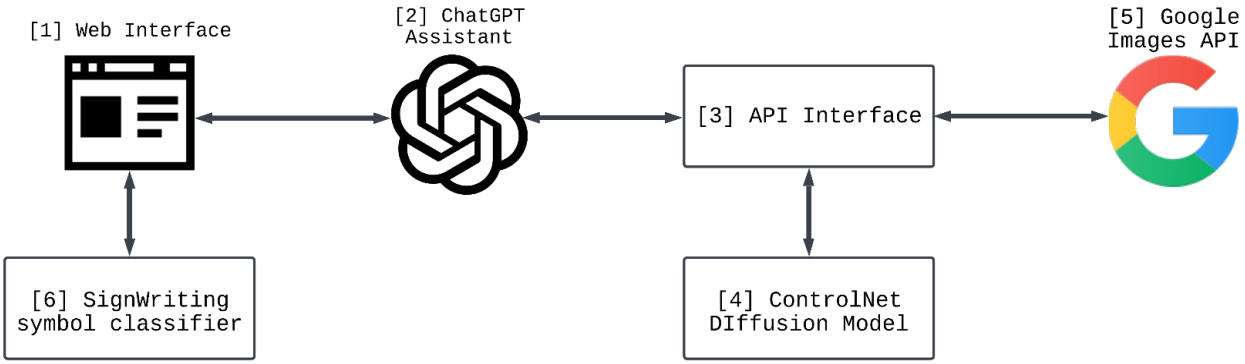


Figure 1. Proposed DoodleShop high-level architecture.

As depicted in Figure 1, users engage with the system through a web interface [1] composed of a conventional chat thread where an Assistant [2] states its role in guiding the user through the product search process. The assistant proceeds to construct a profile of the product by posing relevant questions to the users, eventually prompting them to provide a hand-drawn sketch of the desired product. This step would be made optional in a commercial version of the application. The Assistant then invokes an API endpoint, supplying both the prompt and sketch - an approach inspired by Shoib et al.'s (2023) Image and Text CBFIR and the TASK-former architecture (Sangkloy, 2022). The API [3] calls a Diffusion Model [4] to generate an image of the product based on the user data supplied, subsequently interfacing with the Google Images API [5] to search for similar images. The Google API search result is then sent downstream to the web interface as a chat message notifying the user of the search query completion. Finally, this process iterates upon the user's request for refinement of the result, or the chat thread is closed if the user expresses satisfaction.

To diversify communication beyond text, the web interface offers the option to switch to an annotated sketch-only mode utilizing SignWriting symbols. The web interface calls a symbol classifier [6] to detect symbols based on hand-drawn sketches. Notably, symbols are subsequently converted to text for communication with the Assistant, and responses are transmitted back to the classifier for conversion to images before being displayed to the user.

Note that the following architecture is relevant for any other applications outside the online product search example used for demonstration and all the components can be reused for any tasks involving conversational image search. Moreover, such modular design permits upgrading and swapping the chatbot, image generator, image search engine, and symbol interpreter according to the use case. A notable advantage over the majority of tightly coupled systems in use today.

2.1. Web Interface

The web interface orchestrates user interactions with the system, featuring a home page with a menu option for each of the three types of user interaction the later study aims to explore and compare: Text-only search with no sketch input as the baseline for existing chatbots (like Siri, Alexa, and others), Mixed input with text and sketch, and Sketch-only input using SignWriting symbols instead of text.

Figure 2. Mode options menu adapted from Yu, 2023

Each of these options implements a classic web chat interface between a bot and the user, as illustrated in Figure 3.

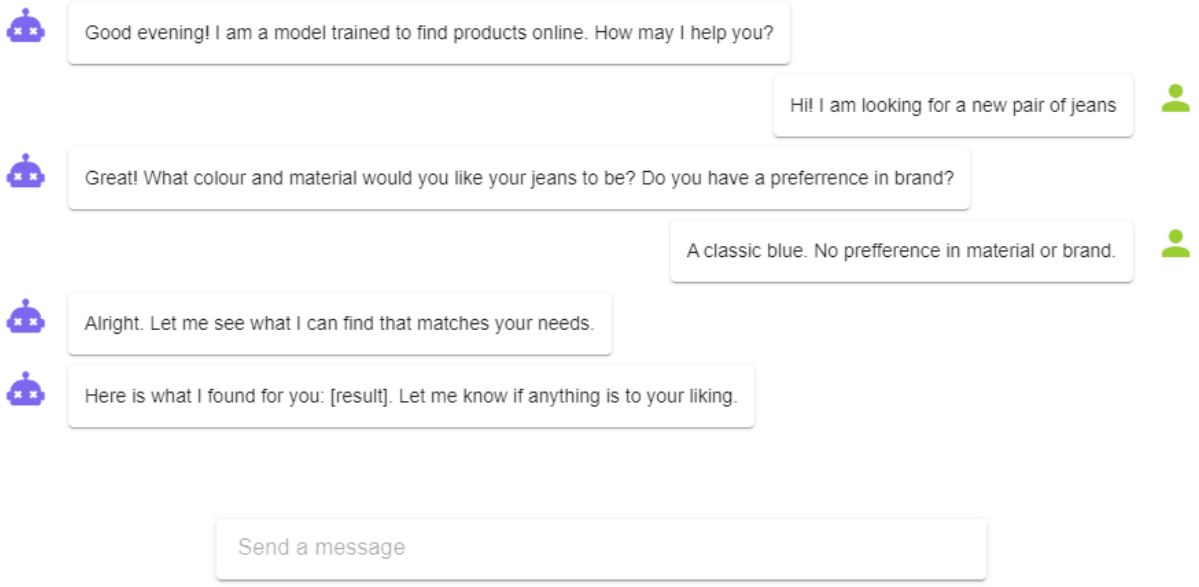


Figure 3. Web chat interface adapted from Yu, 2023

The Mixed and Sketch Only interfaces will incorporate a draw canvas pop-up to provide a sketch submission.

2.2. ChatGPT Assistant

The Assistants API, released by OpenAI, introduces new features to complement the capabilities of ChatGPT language model (LM). An instance started through the API functions as an individual Agent with specified behaviour and tools, extending the capabilities of the LM. For our purposes, this entails the ability to make function calls for image generation and online search - features originally missing from the conversational model.

An instance is initialised with a set of instructions, including the LM to use, available functions, and other tools like a code interpreter or files to use for retrieval if needed. For illustrative purposes, the DoodleShop Assistant received the following instructions before engaging in a conversation soliciting its help to find a blue t-shirt with white stars:

You are a shop assistant helping users find products that best resemble what they are looking for. Start by asking users for details about the product they are looking for to build a 5 to 15-word prompt describing the product. Continue asking questions to get enough details for the prompt. Respond with a JSON with fields for the product name (shorter than the prompt), prompt (5 to 15 words), material, colour, brand and details. Make the details field a JSON array for any details that are not included in the rest of the JSON. Put open-closed braces for fields that have no value.

A sample result of a conversation using the gpt-4-1106-preview model:

```json
{
  "product name": "Star-Side Navy Crew T-Shirt",
  "prompt": "Navy blue cotton crew neck t-shirt with white stars",
  "material": "Cotton",
  "color": "Navy Blue",
  "brand": {},
  "details": [
    "Crew neck style",
    "Short sleeves",
    "Small white stars on the right side"
  ]
}
```

Figure 4. Sample JSON result following a conversation.

This indicative result showcases the ability of the LM to collect data through conversations and structure it in the format requested. In the final implementation, the Assistant will be instructed to make a function call through the API, supplying the relevant data in the specified format in Figure 4, including images in the payload when necessary. The Assistant then waits on the API response before formatting and forwarding it to the web interface.

2.3. API Interface

Given the limitation of ChatGPT (at the time of writing) in accessing information on the internet outside the large corpus of data it was trained on, particularly image databases like Google's, there is the need for a function invocation when the actual image generation and search are done. To facilitate this, an API interface manages the processing of the information received from the Assistant to call a ControlNet service for image generation and then supplies the generated image for a request to the Google Images API. It is important to consider that both the ControlNet and Google APIs could be provided directly to the Assistant but the model is still in active development and previous GPT versions suffered from "hallucinations" (Bang, 2023) which would add unnecessary variance to the search outcomes. The API layer, with well-defined behaviour, is preferred to minimise error rates and perform validation and data processing.

The API manages the following operations:

1. Validate Assistant requests prompting the generation of a new response if the formatting is incorrect.
2. Construct and dispatch a request for the ControlNet diffusion model if a sketch is supplied.
3. Perform a Google search through the API, with either a prompt or generated image, depending on the type of search.
4. Validate and format the response to the Assistant.

The cyclical process repeats when users provide follow-up instructions on how to modify the search parameters. The result of the previous cycle is then reused to produce a line "sketch" and resupplied to the model with the modified prompt.

2.4. ControlNet Diffusion Model

In the context of Diffusion Models it was demonstrated that "line images (human-drawn sketches) are successfully converted to colour images in different styles, where the target style can be intuitively controlled by text guidance (Maungmaung et al., 2023, p.4)." ControlNets, as image-to-image augmentations of pre-trained Stable Diffusion Models, enable such conversions (Zhang et al., 2023). Given an input, they generate an image with the same underlying structure as the original. The modifications from the original are guided by a prompt. For DoodleShop a ControlNet trained on human-made sketches is used to generate an image to search by.

The official demo software available includes endpoints to communicate with a Diffusion model running a ControlNet, requiring minimal modifications to interface with the API.

To exemplify the capabilities of ControlNets, a hand-drawn sketch for the "Navy blue cotton crew neck t-shirt with white stars" prompt was supplied to the latest Stable Diffusion model *v1-5-pruned-emaonly* (Rombach et al., 2022) running the latest *control_sd15_scribble*

ControlNet (Zhang et al., 2023). The “invert” preprocessor (for white background and black lines) was applied to the sketch in Figure 5 before generation. No additional instructions or negative prompts were added to the original prompt.

Figure 5. Sketch representation of the prompt

Following up to 30 second generation times for five 512x512 images on an NVIDIA GTX 1660Ti graphics card, the model outputted the following results:

Figure 6. Sketch ControlNet generated image results for
“Navy blue cotton crew neck t-shirt with white stars”

For comparison, sample results for the same settings but without ControlNet enabled (no sketch supplied):

Figure 7. Text to image Stable Diffusion 1.5 generated image results for
“Navy blue cotton crew neck t-shirt with white stars”

As observed, even in an untuned state, ControlNet yields qualitatively superior results for the prompt supplied, emphasising critical features like position and number of stars on the t-shirt during image generation. Additionally, the batch generated with ControlNet contains fewer human “mannequins” even without a negative prompt to filter out human subjects. Such behaviour is desired in general, even if permitted in the case of a t-shirt. Further fine-tuning may improve the results for both text and sketch-based generation. More sketch-based samples are available in Appendix A.

2.5. Google API

The Google API supports both prompt and image-based search, offering reverse image search through visual analysis using a neural network or keywords-based search through prompts. It was important to evaluate the performance and quality of search results when supplying generated images with imperfections instead of real-world images of objects. As presented in Appendix B, the image search using the generated image consistently returns relevant search results with various generated images. Notably, the higher quality of this type of search method is evident when juxtaposed with classic image search using the prompt, as demonstrated in Appendix C, where filtering a sample of at least 15 images as a user is needed to achieve the quality of results of the top 4 searches using our approach.

It is imperative to acknowledge that cropped images of the object are occasionally generated. These have a strong adverse effect on the quality of the image search as seen in this example:

[1] *Generated Image*

[2] *Google Image Search Result*

An exemplar case of a cropped drawer [1] was interpreted as a leg, leading to the respective search result [2], underscored the challenges posed by such occurrences, emphasizing the need for rigorous validation in the ControlNet image generator.

In the final implementation, a Python wrapper for the API will be used to communicate between the API interface and Google Search.

2.6. SignWriting Symbol Classifier

To address the complexities of textual conversation and search for minority groups with disabilities, visual communication through SignWriting symbols was deemed appropriate. Symbols are optimised for image recognition algorithms and optimal performance was observed using a Convolutional Neural Network (Stiehl, 2015). However, new experiments with current classification models are needed before deciding on a model for the final implementation.

The classifier API will communicate with the web interface to identify symbols from the database based on hand-drawn sketches and text.

3. Project Management

3.1. Work to Date

The conceptualisation and system design drew inspiration from the work of Yu and Xia (2023). A significant amount of time was dedicated to familiarising myself with their work in the domain of conversational image search and onboarding the system they have built to address the challenges of conversational search. As a result, ChatGPT and a web interface were chosen for the conversational part of the project.

A comprehensive literature review was performed both before and during the System Design to identify the most promising technologies to be used. Following this research, Signwriting was selected as the symbol dataset as an alternative to text for its design optimised for classification models and ControlNet was overlaid on the Diffusion model to generate images from sketches.

A proof of concept as described in the previous section was extensively tested to validate the feasibility of the approach. All components have compatible API interfaces for data exchange.

An ERGO application is being submitted for approval to conduct a study comparing different search methods, with a focus on understanding user engagement variations between text-based search and search methods employing sketches.

3.2. Planned Work

While individual validation of each component had been performed, the next development phase involves connecting each component to an automated system. Emphasis will be placed on connecting the API Interface and Web Interface with the other services that already provide interfaces. After the text-only and mixed search methods are implemented, additional work will be done to develop a sketch-only interface that uses the Signwriting symbol classifier.

Following system development and testing, a study will be conducted to qualitatively compare the three search methods and assess the sketch-based approach in the final report. The experimental setup and design will be similar to Artemi (2021), with a specific focus on assessing the user attention metrics when using the sketch-based approach.

3.3. Project Gantt Chart

For a detailed breakdown, including dates and risk assessment, please refer to Appendix D.

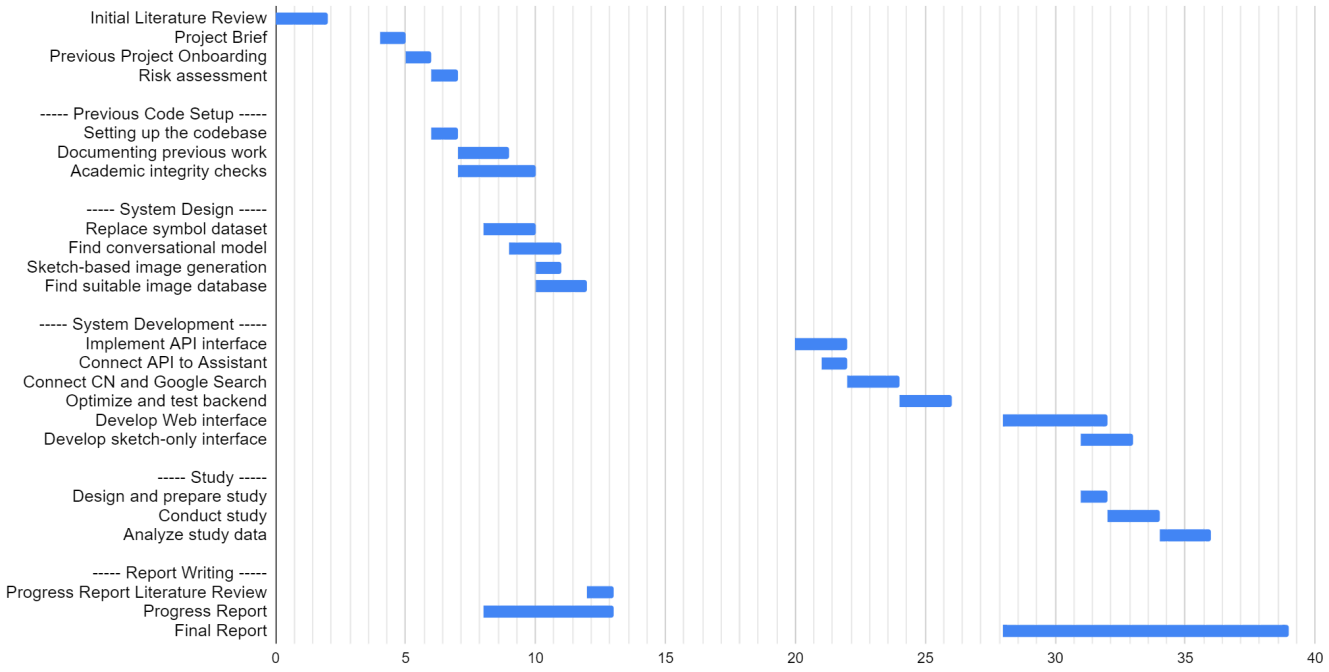


Figure 8. Project Gantt chart

3.4. Costs

A budget allocation of up to 100£ has been reserved for ChatGPT tokens. To date, 1£ has been spent on model validation. Usage will increase through the project implementation but it is expected to remain well below the allocated budget. Noteworthy is the absence of associated costs for the other components of the system, as they will be hosted on the local machine.

4. References

YU, Lei, 2023, Frontend development of an Image-based conversational search system, A dissertation of MSc Software Engineering at the University of Southampton

XIA, Tianyu, 2023, Image-based conversational search, A project of BSc Computer Science at the University of Southampton

ADEWUMI, T., LIWICKI, F. & LIWICKI, M. 2022. State-of-the-Art in Open-Domain Conversational AI: A Survey. Information, 13, 298.

ALIANNEJADI, M., AZZOPARDI, L., ZAMANI, H., KANOULAS, E., THOMAS, P. & CRASWELL, N. 2021. Analysing Mixed Initiatives and Search Strategies during Conversational Search. Proceedings of the 30th ACM International Conference on Information & Knowledge Management. Virtual Event, Queensland, Australia: Association for Computing Machinery.

AL-THANI, H. 2023. Open-Domain Conversational Search: Addressing Challenges and Limitations Using Reformulation and Data Augmentation. Ph.D., Hamad Bin Khalifa University (Qatar).

ARTEMI, M. & LIU, H. 2021. A User Study on User Attention for an Interactive Content-based Image Search System.

BANG, Y.; CAHYAWIJAYA, S.; LEE, N.; DAI, W.; SU, D.; WILIE, B.; LOVENIA, H.; JI, Z.; YU, T.; CHUNG, W.; DO, Q. V.; XU, Y. & FUNG, P. (2023), 'A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity', cite arxiv:2302.04023 Comment: 52 pages.

CHOWDHURY, P. N., BHUNIA, A. K., SAIN, A., KOLEY, S., XIANG, T. & SONG, Y.-Z. 2023. SceneTrilogy: On Human Scene-Sketch and its Complementarity with Photo and Text. arXiv pre-print server.

DEY, S., RIBA, P., DUTTA, A., LLADOS, J. L. & SONG, Y.-Z. Doodle to Search: Practical Zero-Shot Sketch-Based Image Retrieval. 2019 2019. IEEE.

KATO, M. 2008. A study of notation and sign writing systems for the deaf. Intercultural Communication Studies, 17, 97-114.

KEYVAN, K. & HUANG, J. X. 2022. How to Approach Ambiguous Queries in Conversational Search: A Survey of Techniques, Approaches, Tools, and Challenges. ACM Comput. Surv., 55, Article 129.

MAUNGMAUNG, A., SHING, M., MITSUI, K., SAWADA, K. & OKURA, F. 2023. Text-Guided Scene Sketch-to-Photo Synthesis. arXiv pre-print server.

MOHIAN, S. & CSALLNER, C. 2022. PSDoodle. 2022 2022. ACM.

MUHAMMAD SHOIB, A., SUMMAIRA, J., WANG, C. & JABBAR, A. 2023. Methods and advancement of content-based fashion image retrieval: A Review. arXiv pre-print server.

NIE, L., JIAO, F., WANG, W., WANG, Y. & TIAN, Q. Conversational Image Search. IEEE Transactions on Image Processing, 2021.

ROMBACH, R., BLATTMANN, A., LORENZ, D., ESSER, P. & OMMER, B. High-resolution image synthesis with latent diffusion models. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022. 10684-10695.

SANGKLOY, P., JITKRITTUM, W., YANG, D. & HAYS, J. 2022. A Sketch is Worth a Thousand Words: Image Retrieval with Text and Sketch. Springer Nature Switzerland.

SPINK, A., WOLFRAM, D., JANSEN, M. B. J. & SARACEVIC, T. 2001. Searching the web: The public and their queries. Journal of the American Society for Information Science and Technology, 52, 226-234.

STIEHL, D., ADDAMS, L., OLIVEIRA, L. S., GUIMARÃES, C. & BRITTO, A. S. Towards a SignWriting recognition system. 2015 13th International Conference on Document Analysis and Recognition (ICDAR), 23-26 Aug. 2015 2015. 26-30.

ZHANG, L. & AGRAWALA, M. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. arXiv pre-print server.

ZHANG, X., SHEN, M., LI, X. & FENG, F. 2022. A deformable CNN-based triplet model for fine-grained sketch-based image retrieval.

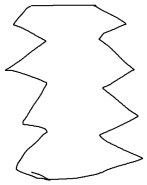
ZHANG, Y., CHEN, X., AI, Q., YANG, L. & CROFT, W. 2018. Towards Conversational Search and Recommendation: System Ask, User Respond.

BERTASIUS, G., SHI, J. & TORRESANI, L. 2015. High-for-Low and Low-for-High: Efficient Boundary Detection From Deep Object Features and its Applications to High-Level Vision. Proceedings of the IEEE International Conference on Computer Vision (ICCV).

Appendix A

Sketch and a sample of 4 images generated using ControlNet 1.1.419 and Stable Diffusion 1.5 with random seed.

Prompt: “Modern abstract white ceramic flower vase”



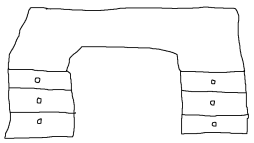


Prompt: “Classic wooden chair woven base”





Prompt: “Hardwood oak desk, drawers”





Appendix B

Generated image given as input to Google Images search and top 4 image search results based on that image.
(Search Region: UK, Search Date: 26/11/2023).

Prompt: “Navy blue cotton crew neck t-shirt with white stars”

Prompt: “Hardwood oak desk, drawers”

Appendix C

Top 15 Google Image search results using only the prompt.
(Search Region: UK, Search Date: 05/12/2023)

Query: “Navy blue cotton crew neck t-shirt with white stars”

Query: “Hardwood oak desk, drawers”

Appendix D

Datailed project Gantt chart table with dates.

Tasks	Start	Due
Initial Literature Review	9/9/2023	9/23/2023
Project Brief	10/9/2023	10/13/2023
Previous Project Onboarding	10/16/2023	10/22/2023
Risk assessment	10/23/2023	10/30/2023
----- Previous Code Setup -----		
Setting up the codebase	10/23/2023	10/30/2023
Documenting previous work	10/27/2023	11/12/2023
Academic integrity checks	10/30/2023	11/19/2023
----- System Design -----		
Replace symbol dataset	11/3/2023	11/19/2023
Find conversational model	11/10/2023	11/24/2023
Sketch-based image generation	11/18/2023	11/26/2023
Find suitable image database	11/19/2023	12/3/2023
----- System Development -----		
Implement API interface	1/29/2024	2/11/2024
Connect API to Assistant	2/4/2024	2/8/2024
Connect CN and Google Search	2/12/2024	2/25/2024
Optimise and test backend	2/26/2024	3/10/2024
Develop Web interface	3/23/2024	4/21/2024
Develop sketch-only interface	4/10/2024	4/21/2024
----- Study -----		
Design and prepare study	4/15/2024	4/21/2024
Conduct study	4/22/2024	5/5/2024
Analyze study data	5/6/2024	5/17/2024
----- Report Writing -----		
Progress Report Literature Review	12/1/2023	12/8/2023
Progress Report	11/5/2023	12/12/2023
Final Report	3/25/2024	6/9/2024

Risk Assessment.

Risk	Impact (1-5)	Probability (1-5)	Mitigation
Technical Challenges	4	3	Keep close attention to the project timeline and overestimate the time needed for solution implementation. Maintain open communication with supervisor and seek expert advice if needed.
User Engagement Study Approval	2	2	Ensure all ethical guidelines are followed when submitting the ERGO application. Consult the supervisor before submitting the application. Address any concerns promptly to expedite the approval process.
Symbol Classifier Inaccuracy	3	4	Train the symbol classifier using new and improved classification models. As a fallback, implement the system without symbol classification.
Budget Overruns	2	1	Regularly track the budget spent to date and set up alerts for spikes in token usage.