

**Helwan University**

**Faculty of Computers and Artificial Intelligence**

**Computer Science Department**

**Interior Design**

A graduation project dissertation by:

Ahmed Gamal Ali Elsayed ( 202000024 )

Hasnaa Muhammed Abd El-Fattah Marei ( 202000274 )

Shaden Khaled Sayed Hafez ( 202000421 )

Omar Atef Wagdy Attia ( 202000601 )

Mahmoud Osama Nabawy Ashour ( 202000839 )

Nada Mohamed Soliman Mohamed ( 202000994 )

Submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Computers & Artificial Intelligence, Helwan University

Supervised by:

**Dr. Soha Ahmed Ehssan**

June 2024

**Acknowledgments**

We would like to show our gratefulness for all the people who gave us their hand of help.

First and foremost, we would to express our deepest gratitude to our Supervisor **Dr. Soha Ahmed Ehssan** for her support, outstanding guidance and encouragement throughout our graduation project and her continuous support and help since the beginning of our project.

We are also so thankful for our beloved faculty,

**Faculty of Computer Science & Artificial Intelligence – Helwan University,**

For providing us a suitable environment for learning, practicing & flourishing in our career.

**Abstract**

Good interior design plays a significant role in creating comfortable living or working environments. A well-designed space can greatly enhance overall comfort and satisfaction of its occupants.

It is not easy to imagine your interior design without actually visualizing it with your own eyes, so we created an interior design Multi-Modal Web Application that helps users to visualize the interiors of an empty space, like Augmented Reality which helps you to imagine how furniture items will look in the space. It helps them make interiors that are reflective of their personality and taste. It also helps businesses to increase customer engagement.

**Chapter 1: Introduction**

In this chapter, we will give you a brief overview of the project, we will talk about the project’s objectives and purposes and how we managed to achieve them, what is the project scope, what were the constraints and how we overcame them.

* 1. **Overview:**
* In this project we will cover the difficulty for users to visualize their interior design items in their rooms.
* The goal is to build a search engine for interior design that allows users to search for images and designs based on visual and textual inputs.
  1. **Problem Statement:**
* Existing interior design search engines are limited in their ability to simultaneously leverage visual features from images and semantic understanding from textual descriptions. This limitation often results in suboptimal search experiences, where users struggle to find designs or items that match their specific requirements and preferences.
  1. **Scope and Objectives:**

1. Helping users to visualize their interiors clearly and offering cost-effective alternative way to design users’ space, as they don’t have to go out looking for a professional home designer for each time, they want to decorate an interior space.
2. Offering immense personalization options to anyone looking to decorate their interiors, as users have the option to choose various colors, and sizes of furniture so that it fits your own unique taste
   1. **Work Methodology:**

Search Engine Models:

1. SBERT (Sentence-BERT) model is a variant of the BERT (Bidirectional Encoder Representations from Transformers) model specifically designed to produce semantically meaningful sentence embeddings. SBERT was introduced to address the limitations of BERT in generating fixed-size vector representations of sentences that can be used effectively in tasks like semantic similarity, clustering, and information retrieval.
2. RoBERTa (Robustly Optimized BERT Pretraining Approach) is an advanced version of the BERT model, By integrating RoBERTa with SBERT, we can leverage the robustness and improved performance of RoBERTa for generating high-quality sentence embeddings.
3. YOLOv8 (You Only Look Once version 8) is a state-of-the-art object detection model that builds on the previous versions of the YOLO series. YOLO models are designed for real-time object detection, which means they can identify and locate objects within an image quickly and efficiently.
4. ConvNeXt v1 is a type of deep learning architecture specifically designed for computer vision tasks.
5. Feature Fusion features of SBERT with RoBERTa and YOLOv8 with ConvNextv1 and fusing them which is a method of merging the different extracted features.
6. Similarity for images & text by cosine similarity which measures the similarity between two vectors of an inner product space. It is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in roughly the same direction. It is often used to measure document similarity in text analysis.
7. Output: Images related to the input from the user based on the similarity between the input (text & image) and our dataset and this is done using the above 6 steps.

**Chapter 2: Related work**

In this chapter we well talk about our related paper, its abstract, used technologies & proposed model.

**Related paper:** <https://arxiv.org/pdf/1707.06907>

* 1. **Abstract:**
* In this paper, the proposed multi-modal search engine for interior design combines visual and textual queries. The goal of the engine is to retrieve interior objects, eg: furniture or wall clocks, that share visual and aesthetic similarities with the query.
  1. **Literature Survey:**

**2.2.1 Textual Search:**

* First methods proposed to address textual information retrieval have been based on

token counts, ex: **Bag-of-Words** or **TF-IDF**

* **Drawback:** The scalability of those methods is very limited and when using such representations long sequences (documents) tend to have similar token distributions which results in lower discriminative power of the representation and lower retrieval precision.
* To avoid those problems is to apply a **SVD decomposition** of the token cooccurrence matrix and, hence, reduce the dimensionality of a representation vector.
* **Drawback:** SVD assumes that the relationships between features are linear, which can limit its ability to capture complex non-linear relationships.
* To handle those shortcomings, a new type of representation called word2vec has been proposed “Continuous Bag of Words (**CBOW**) and **Skip-Grams**”.
* They allow the token representation to be learned based on its local context
* **Drawback:** The CBOW model averages the context of a word and it is a unidirectional model, which means that it only considers the context words in the forward direction. Skip-Grams model tends to have higher memory requirements.

**2.2.2 Visual Search:**

* Traditionally, image-based search methods drew their inspiration from textual retrieval systems.
* By using **K-Means Clustering** method in the space of local feature descriptors, such as SIFT, they are able to mimic textual word entities with the so-called visual words.
* **Video Google** was one of the first visual search engines that relied on this concept that checks for geometrical correctness of initial query and eliminates the results that are not geometrically plausible.
* **R-MAC** technique is an image representation based on the outputs of CNNs that could be computed in a fixed layout of spatial regions.
* **Siamese networks** are proved successful when applied to content-based image retrieval
* **Drawback:**

They do not take into account the contextual and stylistic similarity of the retrieved objects, which yields their application to the problem of infeasibility of interior design items retrieval.

* 1. **Analysis of Related Work:**
* YOLO9000 was used as an object detection model integrated with RESNET-50 that they will be able to get higher performance on overcoming the previous drawbacks.
* **Drawback:** Accuracy was very low (48%) for detecting the objects in the images.

**Chapter 3: Proposed Solution**

In this chapter we will show the architecture of our Model.

**3.1 Proposed Model:**

* In this section, we present the pipeline of our multi-modal Style Search Engine. As an input, it takes two types of query information: an image of an interior, e.g. a picture of dining room, and a textual query used to specify search criteria, e.g. black chair.
* Then, an object detection algorithm is run on the uploaded picture to detect objects of classes of interest such as chairs, tables or sofas. Once the objects are detected, their regions of interest are extracted as picture patches and submitted to visual search method.
* Simultaneously, the engine retrieves the results for a textual query. With all visual and textual matches retrieved, our blending algorithm ranks them depending on the similarity in the respective features spaces and serves the resulting list of stylistically and aesthetically similar objects. Below, we describe each part of the engine in more details.

1. **Visual search:**

* We have integrated YOLOv8 with ConvNeXt v1 to provide precise coordinates for the bounding boxes that encapsulate each detected object by YOLOv8 with the advanced convolutional network design of ConvNeXt. Our object detection method leverages the state-of-the-art YOLOv8 model, which is based on CSPDarknet53 and features 53 convolutional layers. This enables YOLOv8 to detect multiple furniture classes and their bounding boxes accurately.
* These bounding boxes are then used to generate Regions of Interest (ROIs) in the images, on which visual searches are performed. We have optimized YOLOv8's parameters, particularly the detection confidence threshold, setting it to 0.7. In cases where overlapping bounding boxes are returned, the one with the highest confidence score is selected.
* After extracting the ROIs, we input both the ROI and the whole image into ConvNeXt v1, with a weight ratio of 4:1 of ROI image. This approach ensures that the model focuses on the ROI while still considering the entire image's features.
* Finally, we normalize the extracted output vectors so that their L2 norm equals 1. Using this representation, we search for similar images within the dataset.

1. **Textual Search:**
2. **Fusion Method:**

**3.2 Dataset:**

In order to evaluate our proposed Style Search Engine, we collected our dataset images from Google.

* It contains many interior items like chairs & beds.
* It contains some room scenes like bedrooms and living rooms.
* There is a textual description for each room scene image and description for each item in the dataset.

To our knowledge, no such dataset is publicly available. Hence, we collected our own dataset by recursively scrapping many websites on Google, we were able to download 2900 images of room scenes and individual items without a description, so we had to write the description for each image.

In addition, we also divided the images into 14 categories based on the room class and items (living room, bedroom, chairs, clock).

**Chapter 4: Results and Discussion**

In this chapter we will include the results of our model and will discuss the process to get these results.

**Results:**

Graphs of accuracy results of the proposed model

Inputs and outputs by Images ( snapshots from code of the input and output )

**Discussion:**

comparison before and after integration ( snapshots of accuracy graph before and after integ )

graph of the each model used ( sbert,Roberta,convnext &yolo v8 )

**Chapter 5: Application**

In this chapter we will show our **Web Application** and will also mention the technologies we used.

**5.1 Application:**

Snapshots of the frontend ( input and outputs )

Image generation ( user experience ) ( snapshots input and the generated output )

**5.2 Software Tools and Technologies:**

**Visual Studio Code IDE:**

* It is a free source-code editor made by Microsoft, its Features include supporting for debugging, syntax highlighting, intelligent code completion, code refactoring and embedded Git, and it supports web languages.
* It is a creative launching pad that you can use to edit, debug, and build code, and then publish an app

**Google Colab:**

* Colab is a hosted Jupyter Notebook service that requires no setup to use and provides free access to computing resources, including GPUs and TPUs.
* Colab is especially well suited to machine learning, data science, and education.

**Python:**

* It best fits machine learning due to its independent platform and its popularity in the programming community.

**HTML:**

* Hyper Text Markup Language (HTML) is the standard markup language for documents designed to be displayed in a web browser. It defines the content and structure of web content.

**CSS:**

* CSS (Cascading Style Sheets) is used to style and layout web pages.

**JavaScript:**

* JavaScript (JS) is a cross-platform, object-oriented programming language used by developers to make web pages interactive.

**Flask Python:**

* Flask is a popular web framework for Python that is used for building web applications and APIs. It makes it simple to create RESTful APIs with its flexible routing system and request handling with make it easier to make

**Chapter 6: Conclusion**

This chapter will include a conclusion about our project, our recommendations to enhance the project and our future plans to improve our application.

**6.1 Conclusion:**

**6.2 Recommendations:**

1. Tending to use Google Colab Pro to get more access to higher RAM runtimes and more powerful GPUs and TPUs compared to the free version of Colab.
2. By adding more variety to our dataset.

**6.3 Future Plans:**

1) Search by voice note.

2) Mobile Application to be more user friendly.

3) Provide translators for more languages.

4) By adding history search for users.