International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 12 | December - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Eye Frame Recommender System using Convolutional Neural Networks

Mr.Dhruv Sachdeva, Mr.Abhyudaya Bhardwaj
*Maharaja Agrasen Institute of Technology

***Abstract*—In the dynamic field of retail analytics, this project represents a groundbreaking foray into online eyewear shopping. At its core, it aims to create a sophisticated Image Recommender System powered by Convolutional Neural Networks (CNNs). This innovation is driven by a growing demand for intuitive and visually-driven product searches, challenging traditional text-based search methods. The project recognizes the evolving consumer preference for immersive and visually engaging shopping experiences, prompting a shift away from conventional search paradigms. The Image Recommender System, therefore, becomes a pivotal solution, breaking free from the limitations of text-based searches and aligning with the modern shopper's incli- nation towards visual stimuli. It's a strategic response to make eyewear discovery more intuitive and personalized, ushering in a transformative era in online shopping.**

***Index Terms*—Visual Search, Retail Analytics,CNN.**

I. INTRODUCTION

Nowadays when it comes to retail data analytics we found a need of continuous advancement for addressing the ever evolving needs and expectations of consumers, using images can address this issue much more effectively than text based queries. This project is an application of image based product search in our case eye frames which can help us to simplify the product search process.

A. Motivation

There are numerous instances when consumers struggle to recall the name of a product or articulate its features for search purposes. Visual product search, as a solution, allows users to take a picture of the product and search for it, eliminating the need for textual descriptions. The motivation behind this work is to develop a model capable of finding the most similar images to a given eye frame image, catering to scenarios where users seek to reorder a product or find visually similar alternatives. A survey conducted with 1,000 consumers highlighted the inefficiency of traditional text-based keyword queries, with 74% expressing the need for more visual, image-based, and intuitive search functionalities. The project aligns with the growing trend of consumers engaging in "image-based window shopping" on search engines like Google, emphasizing the importance of optimizing product images for visual search.

B. Context and Significance

The retail industry is undergoing a transformative shift, with consumers now seeking immersive and visually intuitive experiences. Recent studies reveal a significant dissatisfaction among online shoppers with conventional text-based search functionalities. The limitations of describing nuanced product preferences through words often hinder the exploration of diverse product offerings, leading to a call for more visually-oriented solutions.

C. The Rise of Visual Search

This research is grounded in the recognition of a paradigm shift towards visual search mechanisms. With the ubiquity of smartphones equipped with high-quality cameras, consumers are increasingly capturing and searching for products using images. This shift is not merely a trend but a fundamental change in how users prefer to engage with online platforms. Industry leaders, including Google, have responded by prioritizing and enhancing visual search functionalities, setting the tone for a new era in online discovery.

D. Eye Frames as a Focal Point

Focusing on the specific domain of eye frames, this research acknowledges the challenges faced by consumers in articulating their preferences. The proposed Image Recommender System caters to scenarios where users may have encountered a visually appealing eye frame but struggle to describe it in textual terms. By harnessing the power of CNNs, the system aims to seamlessly translate visual intent into effective product discovery, revolutionizing how users find and engage with eye frames online.

E. Evolution in Retail Analytics

The project aligns with the broader evolution in retail analytics, where enhancing the customer experience is a strategic imperative. Offering a seamless and visually intuitive product discovery process has emerged as a key differentiator for retailers in an increasingly competitive market. As consumers demand more personalized and visually compelling shopping experiences, the role of Image Recommender Systems transcends being a mere feature; it becomes a pivotal component in shaping the future of online retail.

F. Expected Impact

The anticipated outcome of this research is a sophisticated and adaptive Image Recommender System for Eye Frames. By amalgamating advanced CNN-based image analysis with user-centric feedback loops, the system not only aims to meet user expectations but also endeavors to anticipate and exceed them. The iterative nature of the system ensures continual improvement, aligning with the ever-evolving preferences of discerning consumers navigating the vast digital marketplace.

II. Related Work and Articles

A. Machine Learning Method for Cosmetic Product Recognition: A Visual Searching Approach

A relevant study by A. S. Umer, Partha Pratim Mohanta, Ranjeet Kumar Rout, and Hari Mohan Pandey focuses on a machine learning method for cosmetic product recognition using a visual searching approach. The proposed system recognizes cosmetic products based on a processed database containing image samples of forty different cosmetic items. The system analyzes customer preferences for products and brands, making it relevant to our eye frame recommender system. The implementation involves preprocessing, feature extraction, and classification using various machine learning methods, including Logistic Regression, Linear Support Vector Machine, Adaptive k-Nearest Neighbor, Artificial Neural Network, and Decision Tree classifiers.

B. Image Classification for E-Commerce — Part I

An article on Towards Data Science explores the application of image classification in solving business problems, specifically in the context of a giant online marketplace like Indiamart. The goal is to categorize products into macro categories for effective listing. The article provides insights into how neural networks can be trained to identify the micro category of a product using its images.

C. Building a Reverse Image Search Engine: Understanding Embeddings

This resource details the process of building a reverse image search engine, emphasizing steps like feature extraction and similarity search on datasets such as Caltech101 and Caltech256.This research paper provides huge amount of insights when it comes to scaling searches to large dataset which helps making our system much more sophisticated, robust, optimized and accurate. The use of ANN algorithms and libraries, including Annoy, NGT, and Faiss, is explored, along with insights into fine-tuning models for improved accuracy.

III. Dataset Description and Preprocessing

A. Dataset Description

The dataset comprises 5570 eye frames, each associated with attributes such as product name, product IDs, frame shape, parent category, and URLs of the eye frame images. The parent category includes three classes: Eye Frame, Sunglasses, and Non-Power Reading, while the frame shape has four classes: Rectangle, Wayfarer, Aviator, and Oval.

Attributes in the Dataset:

- **Product Name:** Descriptive names for each eye frame, aiding in cataloging and identification.
- **Product IDs:** Unique identifiers assigned to each eye frame, facilitating efficient referencing.
- **itemFrame Shape:** Categorization of eye frames into distinctive shapes such as Rectangle, Wayfarer, Aviator, and Oval.

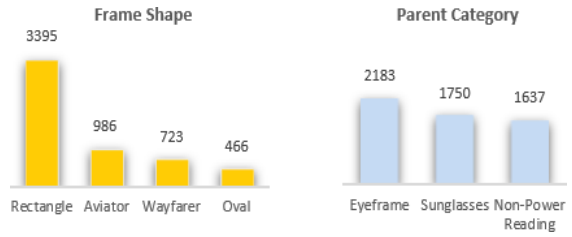

Fig. 1. Bar plot desribing the dataset according to categories.

- **itemParent Category:** A broader classification encompassing categories like Eye Frame, Sunglasses, and Non-Power Reading.
- **Image URLs:** Links pointing to the image files of the respective eye frames, forming the visual essence of the dataset.

Fig. 2. Figure representing pre processing actions.

Preprocessing Steps The preprocessing phase is a meticulous process aimed at preparing the dataset for effective training of the CNN. The steps involved are as follows:

- **Image Retrieval:** The first step involves retrieving the images corresponding to each eye frame from the provided URLs. This ensures that the dataset is complete with the necessary visual data for analysis.
- **Data Cleansing:** To address inconsistencies and ensure data integrity, a comprehensive data cleansing process is undertaken. This involves identifying and handling missing values, correcting errors, and standardizing formats. Ensuring a clean dataset is crucial for the effectiveness of subsequent analyses.
- **Label Encoding:** Categorical variables such as Frame Shape and Parent Category are encoded numerically to facilitate model training. Label encoding assigns a unique numerical label to each distinct category, allowing the model to interpret and learn from these categorical features.
- **Image Resizing:** Standardizing the dimensions of all images is essential for model compatibility. The images are resized to a uniform dimension of (224, 224, 3). This step is crucial to ensure consistency in input size when utilizing the VGG16 model for feature extraction.
- **Feature Extraction:** The dataset is split into training and validation sets. The VGG16 model, a pre-trained convolutional neural network, is employed for feature extraction. The model processes each image to generate a feature array of size (7, 7, 512). These features capture essential information about the visual characteristics of each eye frame and serve as inputs for subsequent stages

of the recommender system.

By comprehensively addressing data retrieval, cleansing, encoding, resizing, and feature extraction, the preprocessing steps ensure that the dataset is well-prepared for training the CNN model and subsequent stages of the eye frame recommender system.

IV. Methodology

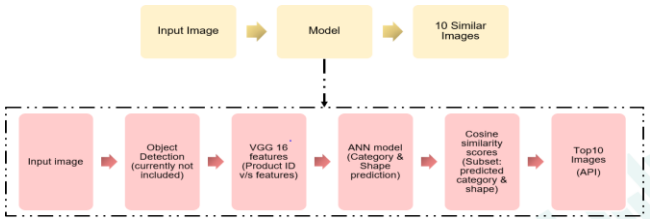

Fig. 3. scheme of progress during execution.

A. CNN Training and Feature Extraction

The Convolutional Neural Network (CNN) serves as the cornerstone of our image recommender system, bringing the power of deep learning to the task of eye frame recognition. Specifically, we employ the VGG16 architecture, renowned for its effectiveness in image classification tasks.

The architecture consists of 16 layers, including 13 convolutional layers and 3 fully connected layers. Each convolutional layer is followed by a Rectified Linear Unit (ReLU) activa- tion function, introducing non-linearity to the model. Max- pooling layers are strategically inserted to downsample spatial dimensions, enabling the network to capture hierarchical features.VGG16 Architecture Details are:

The VGG16 model is pretrained on a large dataset, allowing it to extract rich and hierarchical features from images. During the training phase, the model learns to recognize intricate patterns, textures, and shapes within the eye frame images, culminating in a robust feature representation for each frame.

B. Image Recommendation Process

Feature Generation

The process begins with the generation of features for a user-uploaded eye frame image. The image undergoes prepro- cessing, including resizing to (224, 224, 3), to match the input dimensions expected by the VGG16 model. The model then extracts high-level features from the image.

$$F_{\text{upload}} = \text{tf.keras.applications.VGG16(preprocess\_input}(I_{\text{upload}}))$$

The resulting feature array F_{upload} encapsulates the salientvisual characteristics of the user-uploaded eye frame.

Artificial Neural Network (ANN) Prediction

The feature array F_{upload} serves as the input to an Artificial Neural Network (ANN) responsible for predicting two essential attributes: frame shape and frame category. The architecture of the ANN includes dense layers with rectifiedlinear unit (ReLU) activations to introduce non-linearity.

$$P_{\text{shape}}, P_{\text{category}} = \text{ANN}(F_{\text{upload}})$$

where P_{shape} and P_{category} denote the predicted probabilities forframe shape and category, respectively.

Subset Selection

Once the shape and category predictions are obtained, a subset of eye frames from the main dataset that match the predicted attributes is selected. This subset represents frames that are visually similar in shape and category to the user- uploaded eye frame.

Similarity Score Calculation

Pairwise Cosine Similarity is employed to quantify the visual similarity between the user-uploaded eye frame and frames in the selected subset. This metric ensures that the recommendation is based on the visual features extracted by the VGG16 model.

$$\text{Cosine Similarity}(A, B) = \frac{\text{tf.tensordot}(A, B, \text{axes} = 1)}{\|A\| \times \|B\|}$$

Cosine similarity is a metric used to determine the similarity between two vectors in a multidimensional space. In the context of re vectors of the user-uploaded eye frame and the frames in the selected subset from the main dataset.

Understanding Cosine Similarity

- In our scenario, each eye frame is represented as a feature vector, encapsulating the essential visual characteristics extracted by the VGG16 model during the training phase. These feature vectors serve as a numerical representation of the visual content of the images.
- Cosine similarity is derived from the geometric interpretation of vectors. Imagine the feature vectors as arrows in a multidimensional space, where each dimension corresponds to a specific feature. The cosine of the angle between these vectors provides a measure of their similarity.
- The cosine similarity yields a value between -1 and 1. A cosine similarity of 1 indicates perfect alignment, implying that the vectors are identical. On the other hand, a similarity of -1 suggests complete misalignment, while a value of 0 implies orthogonality or no similarity.
- In the context of our recommender system, cosine similarity is employed to quantify how closely the feature vector of the user-uploaded eye frame aligns with the feature vectors of frames in the selected subset. The higher the cosine similarity, the more visually similar the frames are, indicating a stronger recommendation.

In conclusion, cosine similarity provides an intuitive and robust measure of similarity between vectors, making it a valuable tool in our image recommender system for eye frames.

TOP 10 RECOMMENDATIONS

The frames in the subset are ranked based on their similarity scores, and the top 10 frames with the highest scores are recommended to the user. This step ensures that the recommended frames not only share common attributes with the user-uploaded frame but also exhibit high visual resemblance.

By seamlessly integrating CNN-based feature extraction with an ANN for attribute prediction and a robust recommendation pipeline, our methodology leverages the strengths of deep learning to provide users with personalized and visually compelling eye frame recommendations.

V. RESULTS

A. Convolutional Neural Network (CNN) Training

The success of our Image Recommender System hinges on the effective training of the Convolutional Neural Network (CNN), specifically the VGG16 model. The training process involves optimizing the model's parameters to accurately capture the intricate features of eye frames.

The following details highlight the outcomes of the CNN training:

- **Optimizer and Learning Rate:** The Adam optimizer is employed with a learning rate of 0.01. This choice aims to strike a balance between rapid convergence and fine-tuning to achieve optimal performance. The learning rate plays a pivotal role in determining the step size during parameter updates, influencing the speed and stability of the training process.
- **Loss Function:** Sparse categorical entropy is utilized as the loss function during training. This choice aligns with the multi-class classification nature of the task, where each eye frame belongs to specific categories of both shape and category.
- **Training Epochs:** The CNN is trained for 250 epochs, ensuring that the model iteratively refines its understanding of the dataset. This iterative process allows the model to discern patterns and features, enhancing its ability to make accurate predictions.

B. Evaluation Metrics

To comprehensively assess the performance of the Image Recommender System, multiple evaluation metrics are employed. The primary metrics include accuracy, categorical crossentropy loss, and precision.

- **Category Prediction Metrics:** For the prediction of eye frame categories, the CNN achieves an accuracy of 97.2%. This metric reflects the percentage of correctly predicted categories out of the total predictions. A low categorical crossentropy loss of 0.0845 further underscores the robustness of the category prediction model.
- **Shape Prediction Metrics:** Shape prediction exhibits a noteworthy accuracy of 90.4%, emphasizing the model's proficiency in recognizing diverse shapes of eye frames. The associated loss of 0.2700 signifies the model's ability to minimize errors during shape prediction.

C. Training Plots

Visual representations of the training process provide valuable insights into the convergence and stability of the CNN. The following plots depict the training accuracy and loss over the 250 epochs:

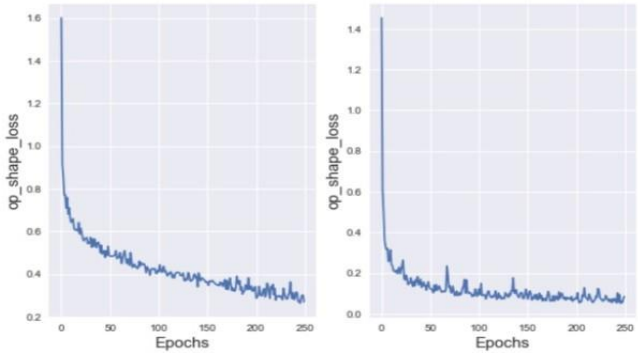

Fig. 4. Plot Showing Training Loss.

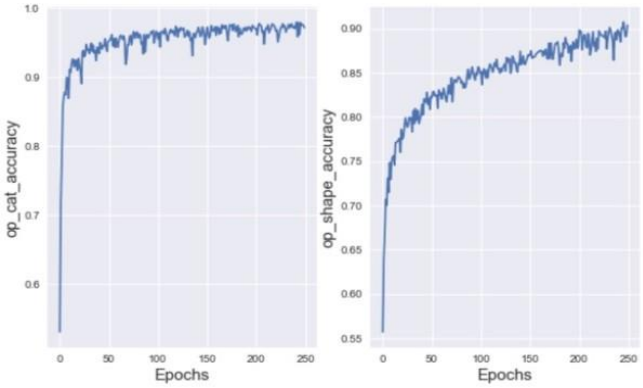

Fig. 5. Plot showing Training Accuracy.

- **Training Accuracy Plot:** The training accuracy plot showcases the gradual increase in accuracy over epochs. This upward trend signifies that the model is effectively learning and adapting to the nuances of the eye frame dataset.
- **Training Loss Plot:** The training loss plot exhibits a steady decline, indicating that the model is minimizing errors during the training process. A lower loss signifies that the predictions are aligning closely with the ground truth labels.

VI. EXPECTED FINAL OUTCOME

The project aims to implement an advanced Image Recommender System for eye frames, combining CNN-based feature extraction and ANN-based predictions. The model is designed to recommend the top 10 eye frames visually similar to the uploaded image. Future work involves experimenting with larger datasets and fine-tuning the model to achieve state-of-the-art performance.

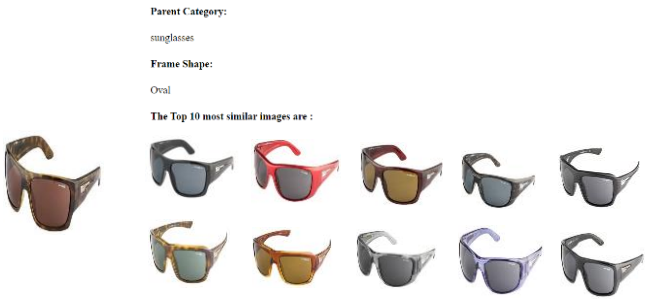

Fig. 6. Final Output from Project.

VII. ACKNOWLEDGEMENT

Thanks to our mentor Dr.Vibhor Sharma, advisors, and educators, for their guidance, encouragement, and invaluable insights that shaped the direction and execution of this research. The developers and contributors of open-source frameworks, libraries, and tools that were instrumental in implementing and refining the image captioning system. Our peers and colleagues for their continuous support, discussions, and constructive feedback, which immensely con- tributed to the evolution of this project. Their collective con- tributions and unwavering support have been indispensable in the journey of conceptualization, development, and realization of this image captioning endeavor. This endeavor stands as a culmination of concerted efforts and support from various indi- viduals and resources that have contributed significantly to its realization. The academic community, whose extensive research and publications in the fields of computer vision and natural language processing served as a guiding light throughout this project.

REFERENCES

[1] Lowe, David G. ”Distinctive image features from scale-invariant keypoints.” International Journal of Computer Vision60.2 (2004): 91-110.
[2] Viola, Paul, and Michael Jones. ”Rapid object detection using a boosted cascade of simple features.” Computer Vision and Pattern Recognition, 2001.
[3] Bartolini, Ilaria, Paolo Ciaccia, and Marco Patella. ”WIND- SURF: A region-based image retrieval system.” Technical Report CSITE-011-00, CSITE–CNR,2000.
[4] Zagoruyko, Sergey, and Nikos Komodakis. ”Learning to compare image patches via convolutional neural networks.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
[5] Agrawal, Pulkit, Ross Girshick, and Jitendra Malik. ”Ana- lyzing the performance of multilayer neural networks for ob- ject recognition.” European Conference on Computer Vision. Springer International Publishing.Wang, J. “Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication),” IEEE J. Quantum Electron., submitted for publication.
[7] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2016). Region-based convolutional networks for accurate object detection and segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(1), 142–158.
[8]Gao Y., Zhang H., Chen X., & Li P. (2019). ”Visual Fashion Recommendation with Image Content and Social Context.” IEEE Transactions on Multimedia, 21(4), 966-977.
[9]Wang Z., Zhang H., Zhang Z., Zhang W., & Wang H. (2020). ”Deep Style: Transfer Style in Fashion Recommendation with Visual Attention.” Information Sciences, 506.
[10]Li, J., Zhao, Y., & Lu , Y. (2019). ”Hybrid Neural Recommender System for Personalized Fashion Recommendation.” Information Processing & Management, 56(5), 1605-1622.
[11]Simo-Serra, E., Trigeorgis, G., Moreno-Noguer, F., & Urtasun, R. (2016). ”Neuroaesthetics in Fashion: Modeling the Perception of Fashionability.” Computer Vision and Image Understanding, 149, 49-59.
[12] McAuley, J., Targett, C., Shi, Q., & van den Hengel, (2015). ”Image-based Recommendations on Styles and Sub- stitutes.” SIGIR, 43(2).
[13] He, R., & McAuley, J. (2016). ”VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback.” AAAI, 30(1).
[14]Salakhutdinov, R., & Hinton, G. (2007). ”Semantic Hashing.” International Journal of Approximate Reasoning, 50(7), 969-978.
[15]Zheng, Y , Zhang, H , & Cui, P (2018). ”A Neural Influenced Collaborative Filtering Model for Fashion Recommendation.” Knowledge-Based Systems, 139, 3-13. [16]Wang, H., Wang, N., Yeung, D. Y., & Shi (2016). ”Col-laborative Deep Learning for Recommender Systems.” KDD.