A VISUAL SEARCH MODEL FOR AN E-COMMERCE PLATFORM

[1]Dr. OKWU, HACHIKARU NGOZI

Department of Computer Science, Rivers State University, Nkpolu Port Harcourt, Nigeria
okwuhachikaru@yahoo.com

[2]Kilakime, Warikaramu-ere, Bridget.
Department of Computer Science, Bayelsa State Polytechnic, Aleibiri, Nigeria
kilakimekaramu@gmail.com

Abstract

Online shoppers are faced with difficulties such as searching or finding the right keywords to use in search of a particular product needed or searching through a large database of products to find an exact type or the type of product needed which thereby consumes their resources and time. Researchers over time have solved these problems by creating a visual or reverse image search to ensure that users do not need to have the right keyword to search for a product which also saves time and resources by narrowing the database of the product to the particular class of product needed by the user thereby predicting product close to the user's wants. Their work suffered some setbacks because errors were found due to some misclassification and computational problems. This research work presents a platform that solves these problems by making use of a Convolutional Neural Network which is a Deep-Learning technique to enable the system to learn and predict from the pattern, shape, and colour of the product inputted by the user and also computational problems by using a DenseNet architecture. A considerable amount of dataset was used to train the model. Object-oriented analysis and design (OO-AD) methodology was used to develop the system. The programming language used for this research is Java.

Keyword: E-Commerce, Visual Search, Machine Learning, Neural Network, Deep Learning, Convolutional Neural Network.

Introduction

The tremendous growth of E-commerce over the last few years is highly due to advancements in artificial intelligence technology. This advancement of artificial intelligence in E-commerce has been beneficial in creating a more personalized customer experience, this is achieved by checking and analysing customer's behaviour before, during, and after purchase. It also analyses the products a customer needs and is likely to buy, it predicts and recommends what the customer wants from

its past purchase. Artificial intelligence such as Chatbots can be used to find the best price of an item for the customer based on the type of product the customer likes. It retargets potential customers by following up with customers effectively and also keeps accurate track of progress. Chatbots can also remarket products by navigating customers to available and new products. It creates an efficient sales process which is achieved through the use of natural language processing (voice recognition app) like Siri to answer customer queries, solve problems, and even identify new opportunities for sales. It creates customer-centric search which allows businesses to develop a customer-centric experience through advanced image recognition. It uses machine learning to automatically tag and visually search content by labelling features of the image.

Visual search is the act of thoroughly looking for something in a visual environment that is cluttered. The item being searched is called a target while other items are distractors. Its search engine is designed to search for images or information through the input of an image. According to (Hagen et al, 2000) keywords are dominant for finding products for retailers and users and it is unsatisfying. He also said that search is the only mechanism used to navigate through large data without any guarantee of getting closer to what is desired, therefore search is not enough (Hagen et al, 2000). Visual search is necessary because words are not sufficient enough to express what the customers want (the human brain can process images faster than words). It is also important because it reduces resource consumption, and time complexity in retrieving information about items due to difficulty in getting a precise description of the product desired especially products without a brand name. Visual search is made possible with the invention of artificial intelligence and machine learning. It is modelled in the same manner as the sensory organ EYE, which is used to view images and then sends information to the brain which then interprets this information and sends back results. The eye is the most important sense organ out of the five sense organs because it is the main source of information. The visual search, therefore, uses machine learning which is a computer program to learns based on the inputs provided to it using experience and then predicts what the image is, based on its feature and labelled data given as an output.

Statement of the Problem

The problem facing many cases of Visual Search is the problem of misclassification, where most images are not properly grouped into different classes, such classes include; class one based on the type of images such as trousers, skirts, shirts, gowns, ankle boots, bags, etc and class two based on the pattern, shape, and colour.

Aim and Objectives of the Study

This study is aimed at developing a model that simplifies the shopping process of customers through the use of image-based search as opposed to text-based searches.

The specific objectives used to achieve the aim are as follows:

1. To develop an improved visual search model.
2. To train the images using a Log Loss Function
3. To implement the model using Java programming language and Convolutional Neural Network (CNN).

Significance of the Study

This research is significant to the academic community and the society. Academically, this research will aid researchers in the field to further develop visual search technology, by identifying different areas where this technology will be useful and implement it**.** In society, this technology will aid in improving the growth of E-commerce whereby users can take and upload photos of the kind of commodity they like and get what they want.

Scope of the Study

The study of visual search and recommendation is broad in its scope with new research being made daily to improve on this fast-growing technology. This study is being applied in different areas of specification as it is being used in Google Lens, camfind, and so on. (Shankar et al, 2017) created a visual search engine using Deep-Learning to enable users to search for products based on their shape, colour, and pattern with the use of an image. They achieved this using VISNet which is a CNN architecture to achieve its result. This study covers a scope of visual search and recommendation using DenseNet architecture of Convolutional neural networks (CNN). It reduces error, it has fewer parameters and it is the latest architecture used in visual recognition.

Visual Search Technology

Visual search is the use of an image as input in the search space (query) to get results related to the image inputted, unlike the keyword search. The keyword-based search process can, however, be especially challenging for inexperienced searchers. Studies have shown that keyword-based queries significantly limit the expressiveness of users and, therefore, degrade the effectiveness of search (Murdock *et al*, 2007). Using of keywords is inadequate which often forces users to formulate queries using a language they are not native to (Clough and Eleta). This inadequacy it may take users a considerable amount of time and effort to discover the right set of keywords through a trial-and-error process. The way to overcome this is to allow users to use the picture of the target interface as a visual query. E-commerce major sales are driven by fashion and lifestyles (clothes, footwear, bags, and accessories). Users buying decisions are primarily influenced by the product's visual appearance. With visual search technologies, people can hunt online for bargains using pictures of clothing, handbags, shoes, or other items they might desire. Examples of visual search engines are; Google, Google Goggles, Google Lens, Google Images, Bing Visual Search, and Pinterest

Machine-learning

Machine-learning technology powers many aspects of modern society: from web searches to content filtering on social net-works to recommendations on e-commerce websites, and it is increasingly present in consumer products such as cameras and smartphones. Machine-learning systems are used to identify objects in images, transcribe speech into text, match new items posts, or products with users' interests, and select relevant search results. Increasingly, these applications use a class of techniques called deep learning.

Conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, constructing a pattern-recognition or machine-learning system required careful engineering and considerable domain expertise to design a feature extractor that transformed the raw data (such as the pixel values of an image) into a suitable internal representation or feature vector from which the learning subsystem, often a classifier, could detect or classify patterns in the input.

Deep-Learning

Deep learning has evolved from three waves, which are; cybernetics (1940s-1960s), connectionism (1980s-1990s), and deep learning starting from 2006. These models originated from the human brain which is how learning happens or could happen in the brain, as a result of this one of the names of deep learning is artificial neural networks (ANNs). Neural networks used in machine learning have been used to understand the brain's functions but are not generally designed to be a realistic model of biological function (Hinton and Shallice, 1991). Deep learning has been motivated by two main ideas. One is that the brain is proof by example that intelligent behavior is possible and a path to build intelligence reverses the computational principles behind the brain and duplicates its functionality. The second is that it is interesting to understand the brain and the principles underlying human intelligence so machine learning can be used to solve problems. (Bengio and LeCun, 2007; Delalleau and Bengio,2011; Pascanu *et al*, 2014; Montufar *et al*, 2014). The third wave began with a focus on new unsupervised learning techniques and the ability of deep models to generalize well from small datasets, However, today, there is more interest in much older supervised learning algorithms and the ability of deep models to leverage large labeled datasets.

Convolutional Networks

Convolutional networks are convolutional neural networks, ConvNet, and CNNs. It was first introduced and used by Kunihiko Fukushima. Fukushima designed neural networks with multiple pooling and convolutional layers. An artificial neural network was developed called Neocognitron, which used a hierarchical multilayered design in 1979 by Fukushima. This design allowed the computer to “learn” to recognize visual patterns. The networks resembled modern versions but were trained with a reinforcement strategy of recurring activation in multiple layers, which gained strength over time. Additionally, Fukushima’s design allowed important features to be adjusted manually by increasing the “weight” of certain connections.

Neocognitron concepts are still being used. The use of top-down connections and new learning methods have allowed various neural networks to be realized. When more than one pattern is presented at the same time, the Selective Attention Model can separate and recognize individual patterns by shifting its attention from one to the other. (The same process many of us use when multitasking). A modern Neocognitron can not only identify patterns with missing information (for example, an incomplete number 5) but can also complete the image by adding the missing information. This could be described as “inference.”

Convolutional networks processes data that has a known grid-like topology (LeCun, 1989). They are designed to process data that

comes in the form of multiple arrays, many data modalities are in the form of multiple arrays: 1D for signals and sequences, including language; 2D for images or audio spectrograms; and 3D for video or volumetric images. It is a network that uses a mathematical operation called convolution and linear operation. Convolutional networks are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers. It is an example of neuroscientific principles that influenced deep learning.

Related Works

Amazon Flow, Google, and Bing Similar Images are examples of widely used visual search and recommendation systems. Various works (Jing and Baluja, 2008; Frome *et al*, 2015) have been proposed to improve the ranking of the image search results using visual features. In recent years, research has also focused on domain-specific image retrieval systems such as fashion recommendation (Yamaguchi *et al*, 2015; Liu *et al*, 2016; Simo-Serra and Ishikawa, 2016), product recommendation (Liu *et al*, 2012; Kiapour *et al*, 2015) and discovery of food images (Aizawa and Ogawa, 2015). Over the past few years, convnet architectures such as AlexNet (Krizhevsky *et al*,2012), GoogLeNet (Szegedy *et al*, 2016), VGG, and ResNet have continuously pushed the state-of-the-art on large-scale image classification challenges. Though trained for classification in one particular domain, the visual features extracted by these models perform well when transferred to other classification tasks (Razavian *et al*, 2014), as well as related localization tasks like object detection and semantic segmentation (Long *et al*, 2014). Based on image features in retrieval settings on web-scale services, the part-based model (Felzonszwalb *et al*, 2010) approach was well-studied for detection, but recently deep learning methods have become more prevalent, with applications such as face detection (Taigman *et al*,2014), street number detection (Good fellow *et al*, 2013), and text detection (Jaderberg *et al*, 2014). Recent research focuses on application of detection architectures such as Faster RCNN (Ren *et al*, 2015), YOLO (Redmon *et al*, 2016) and the Single Shot Detector (Liu *et al*, 2016). Recent work has also demonstrated the effectiveness of learning embeddings or distance functions directly from ranking labels such as relative comparisons (Schroff *et al*, 2015; Jing *et al*, 2013) and variations of pairwise comparisons (Bell and Bala, 2015; Song *et al*, 2016), or using Bilinear features (Gao *et al*, 2015) for fine-grained recognition.

Material and Methods

Analysis of the Proposed System

The proposed system makes use of a Convolutional neural network (CNN) which involves two steps; Log loss function and DenseNet architecture. There are various implementation architectures for CNN. The goal of using DenseNet architecture is to improve the parameter effect (parameter efficiency), and also because of its high computational efficiency. In Densenet architecture, every layer is connected to every other layer in the network, each of the layers receives signals from all preceding layers which therefore has a direct route for the information backward. This model benefits in feedforward propagation where a task can get a low-level feature activation as

well as a high-level feature activation. In classification, the low-level feature activation determines the edges while the high-level feature activation determines large-scale features such as the presence of a face. In backward propagation having all layers connected allows us to quickly send gradients to their respective places in the network easily figure 1.1: shows the architecture of the proposed system. In DenseNet explicitly different information can be added to the network and that information is preserved. Their layers are very narrow adding only a small set of the feature map to the collective knowledge of the network, keeping the remaining feature map unchanged and the final classifiers make a decision based on the feature maps in the network. Each layer has direct access to the gradient from the loss function and original input signal leading to implicit deep supervision. It has a regularizing effect which reduces over-fitting on the task with smaller training sizes.

Each layer can be thinner and can obtain a more compact model. Each layer generates k feature maps. X0 is concatenated to X1 to generate the next output, it keeps concatenating until the end of the program. The concatenating feature maps learned by different layers increases variations in the input of the subsequent layers and improves efficiency. Each layer adds k feature maps of its own to this state. The growth rate determines how much new information each layer contributes to the global state. The global state once written can be accessed from everywhere within the network.

The proposed system utilizes the more efficient method of ranking by using the Log loss function. Individual layers receive additional supervision from the Log loss function through the shorter connection. The classification attached to every hidden layer enforces the intermediate layers to learn discriminative features. The loss and gradient of DenseNet are substantially less complicated as the same loss function is shared between all layers. Also, the proposed system brings to the fore a recommendation system that enables the users of an e-commerce site to not only identify their item of choice through visual search but also get recommendations based on the images searched for. Besides utilizing a convolutional neural network algorithm for image recognition, the proposed system also implements an item-to-item recommendation using a model-based recommendation algorithm.

Material and Methods

Analysis of the Proposed System

The proposed system makes use of a Convolutional neural network (CNN) which involves two steps; Log loss function and DenseNet architecture. There are various implementation architecture for CNN. The goal of using DenseNet architecture is to improve the parameter effect (parameter efficiency), and also because of its high computational efficiency. Densenet architecture, every layer is connected to every other layer in the network, each of the layer receives signals from all preceding layers which therefore has a direct route for the information backward. This model benefits in feedforward propagation where a task is being able to get a low-level feature activation as well as a high-level feature activation. In classification, the low-level feature activation determines the edges while the high-level feature activation determines large-scale features such as the presence of a face. In backward propagation having all layers connected allows us to quickly send

gradients to their respective places in the network easily figure 1.1: shows the architecture of the proposed system. In DenseNet explicitly different information can be added to the network and that information is preserved. Their layers are very narrow adding only a small set of the feature map to the collective knowledge of the network, keeping the remaining feature map unchanged and the final classifiers make a decision based on all feature maps in the network. Each layer has direct access to the gradient from the loss function and original input signal leading to implicit deep supervision. It has a regularizing effect which reduces overfitting on the task with smaller training sizes.

Each layer can be thinner and can obtain a more compact model. Each layer generates k feature maps. X0 is concatenated to X1 to generate the next output, it keeps concatenating until the end of the program. The concatenating feature maps learned by different layers increases variations in the input of the subsequent layers and improves efficiency. Each layer adds k feature maps of its own to this state. The growth rate determines how much new information each layer contributes to the global state. The global state once written can be accessed from everywhere within the network.

The proposed system utilizes a more efficient method of ranking by using the Log loss function. Individual layers receive additional supervision from the Log loss function through the shorter connection. The classification attached to every hidden layer enforces the intermediate layers to learn discriminative features. The loss and gradient of DenseNet are substantially less complicated as the same loss function is shared between all layers. Also, the proposed system brings to the fore a recommendation system that enables the users of an e-commerce site to not only identify their item of choice through visual search but also get recommendations based on the images searched for. Besides utilizing a convolutional neural network algorithm for image recognition, the proposed system also implements an item-to-item recommendation using a model-based recommendation algorithm.

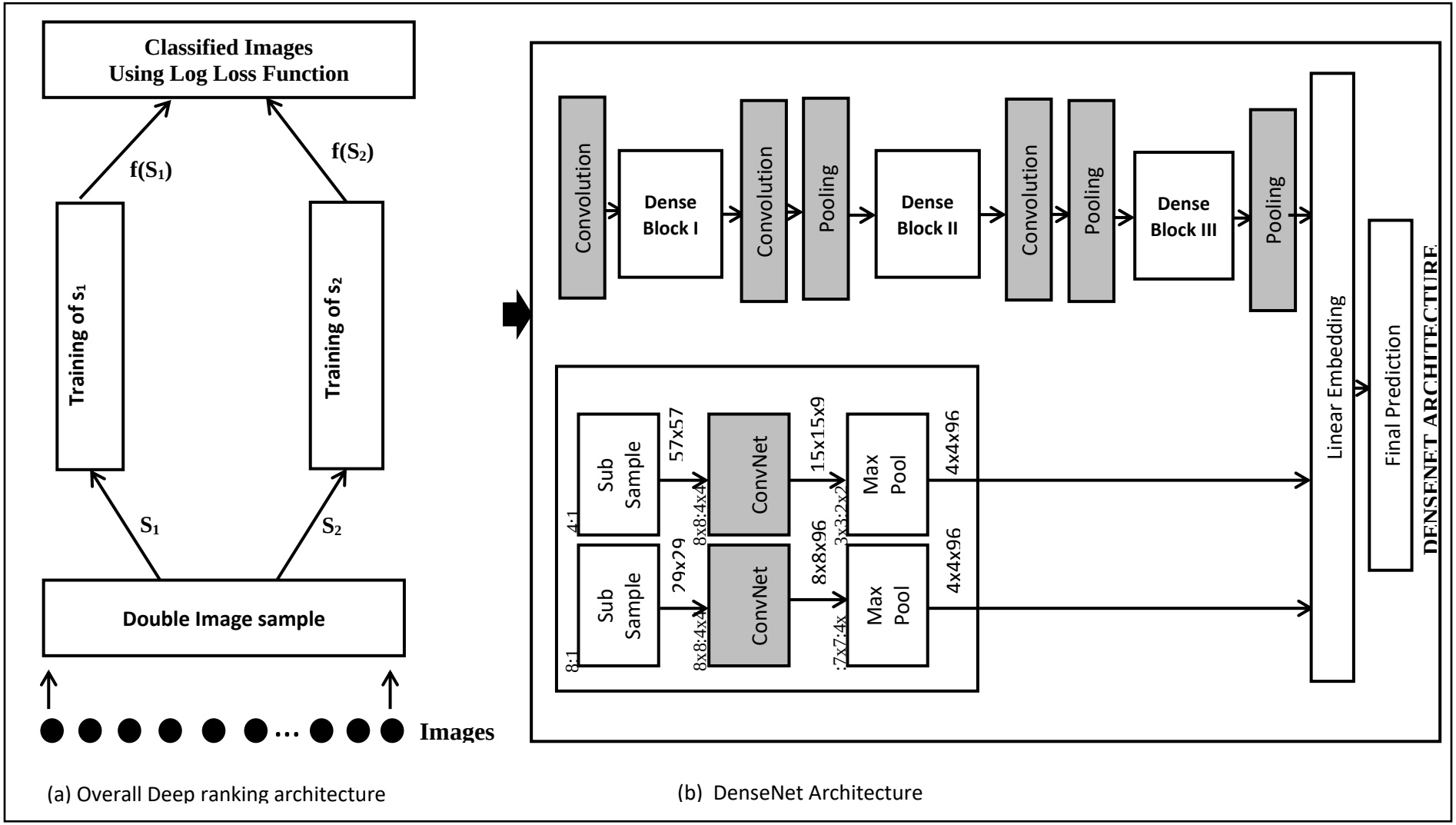

Figure 1.1: Architecture of the Proposed System

Advantages of the Proposed System

The advantages of the proposed system include the following

1. The proposed system has a strong gradient flow thereby error signal can be easily propagated to other layers more directly.
2. The proposed system has implicit deep supervision thereby making other layers get more direct supervision.
3. The proposed system improves the flow of information and gradient throughout the network which makes them easy to train.
4. They alleviate the vanishing gradient problem.
5. It encourages feature reuse and substantially reduces the number of parameters. Therefore, reduces its parameters and is computationally efficient.
6. The proposed system maintains low complexity features.

Methodology

The methodology employed by the proposed system is known as an Object-Oriented Analysis and Design (OOAD) methodology. Object Oriented Analysis and Design (OOAD) methodology is efficient because of its known characteristics of breaking down a large system activity into subclasses and modules thereby making implementation of the system fast and easy. The task of an Object-Oriented Analysis and Design (OOAD) methodology involves; the recognition of significant objects in a class design and the simplification of the classes by disintegration into sub-classes, carrying out software processes on this classes and the re-application of software processes on the identified objects.

The procedures for Object Oriented Analysis and Design (OOAD) methodology as stated above are necessary in almost all Software designs that utilize an object-oriented software pattern and must be implemented recursively and consistently to arrive at the set goal.

Use Case Diagram

These interactions as described in the diagram shown in Figure 1.2. below explains how

The user: captures or selects images from his directory and searches the image, from the searched results the user selects an image from the recommended images displayed, adds it to cart pays for the product, checks out, and rates our site. The recommender algorithm computes the image inputted by the user with the images trained to get similar results from the product available, it also rates the product based on what the user selected. The admin confirms the order, processes shipping and assigns a delivery agent. The admin updates new products remove old products, and puts some products on sales. The recommender algorithm computes the image inputted by the user with the images trained to get similar results from the product available, it also rates the product based on what the user selected. The admin confirms the order, processes shipping and assigns a delivery agent. The admin updates new products remove old products, and puts some products on sales.

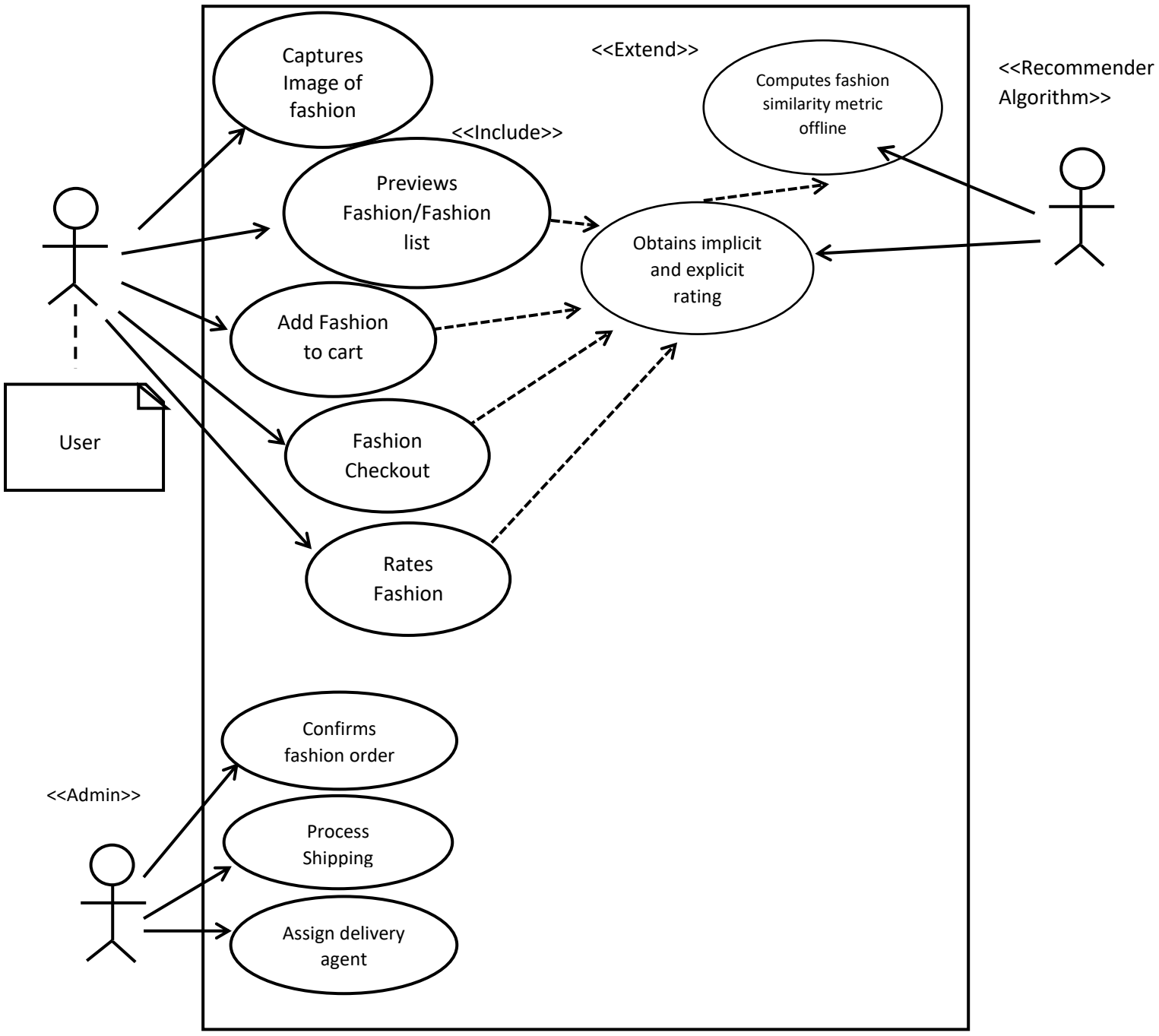

Figure 1.2. Use Case Diagrams of the Proposed System

Dataset for the proposed system

The dataset of the visual search for the e-commerce recommendation system was not easily and openly accessible due to the sensitivity of the data and the copyright attached to online images. The dataset used in the proposed system was obtained from the Google dataset repository. Tables 1.1 and 1.2 illustrate the dataset attributes

Table 1.1. Attributes of Fashion Item Ratings Data

Attribute No.	Attribute Name	Attribute Description
1	User-ID	Numerical
2.	Fashion Rating	Numerical

User ID is the primary key, fashion rating is the field for saving rating.

Table 1.2. Attributes of Fashion Item Data

Attribute No.	Attribute Name	Attribute Description
1	Product ID	Numerical
2.	Product-Title	Qualitative
5.	Publisher	Qualitative
7.	Image-URLs-L	Qualitative

Product ID is a primary key, Product-Title is the product name, Publisher is the name of the publisher and Image-URLs-L is used to save the image link.

Choice of Programming Language

To implement this system, the choice of programming language was considered. The Node is an open-source cross-platform JavaScript run-time environment that executes JavaScript code outside the browser. JS was used primarily for client-side scripting, in which scripts written in JavaScript are embedded in a webpage's HTML and run client-side by a JS engine in the user's web browser. Node JS lets developers use JS to write Command Line Tools and for server-side scripting running scripts to produce dynamic web page content before the page is sent to the user's web browser. Node JS represents a "JavaScript everywhere" paradigm unifying web application development around a single programming language rather than different languages for server-side and client-side scripts. Java codes itself can run on all platforms that support Java without the need for recompilation. Java applications are typically compiled to bytecode so they can run on any Java Visual Machine (JVM) regardless of the computer architecture. Java is capable of handling databases and can run on a network environment, it's a machine-independent language that can run in any operating system without modifying the codes, it can be used for writing Application programs, and it is simple, secure, portable, object-oriented, robust, multithreaded, architecture-neutral, interpreted, has high performance, distributed and dynamic. Node JS is used for building image search and PHP is used for building the interface. The IDE used for this work is Visual Studio Codes. IDE is a software application that provides comprehensive facilities to computer programmers for software development. My MySQL is used also, MySQL is an open-source relational database management system based on Structured Query Language. It's by far the most popular database management system for small-to-medium-sized projects**.**

Hypertext Markup Language (HTML) is also used, HTML represents the standard markup language for creating web pages and web applications. With Cascading Style Sheets (CSS) and JavaScript, it forms a triad of cornerstone technologies for the World Wide Web. Web browsers receive HTML documents from a web server or local storage and extract the documents into multimedia web pages. Additional use of Cascading Style Sheets (CSS) which is a style sheet language used for describing the presentation of a document written in a markup language like HTML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript. CSS is designed to enable the separation of presentation and content, including layout, colours, and fonts. This separation can improve content accessibility, provide more flexibility and control in the specification of presentation characteristics, enable multiple web pages to share formatting by specifying the relevant CSS in a separate .CSS file, and reduce complexity and repetition in the structural content.

Discussion of Results

This is where screenshots of the web application and admin pages are shown.

Figure 1.3. Welcome Page

This form is the first page that displays on the entry into the software as a display for what the thesis is all about.

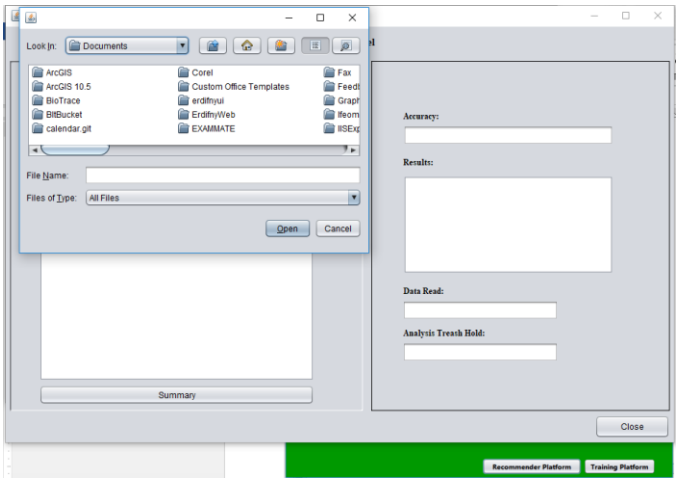

Figure 1.4. Training Form Data

This form allows for selection of the training dataset folder, where the data for training will be iterated over.

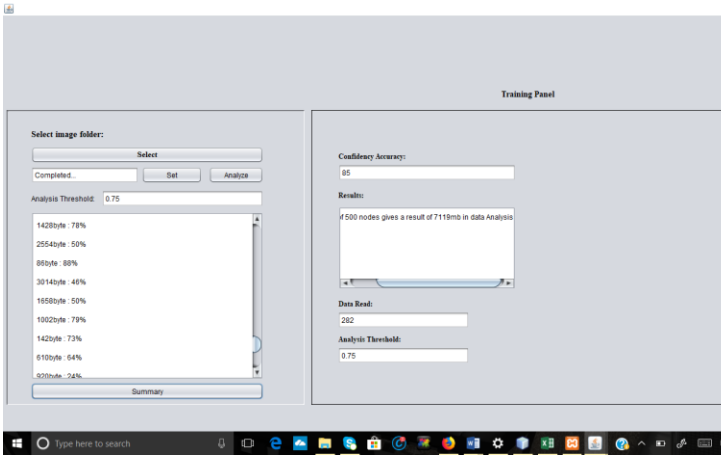

Figure 1.5. Training Result Page

This page displays the results of the analysis and training after an image folder has been selected and set for training.

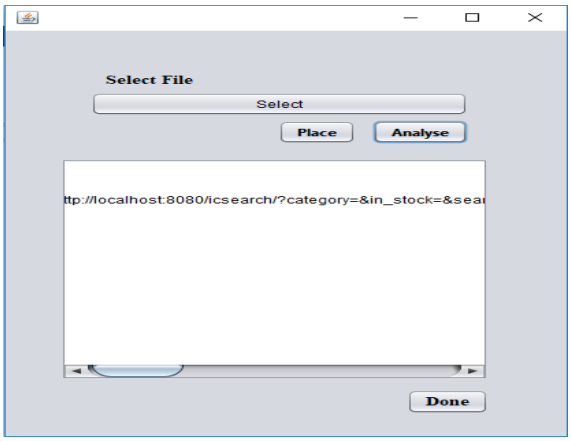

Figure 1.6. Recommender Form

This form is used as a recommender to the localhost-based ecommerce system, using the trained CNN dataset as a base to recommend the possible product being searched by the user.

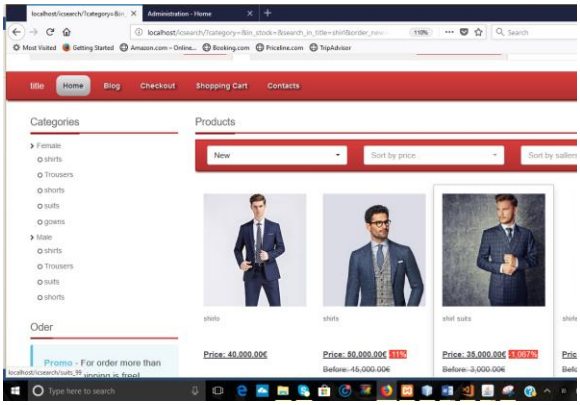

Figure 1.7. Recommendation Results

This is a display of the result from the local based ecommerce system, where the recommendation is gotten from the initial platform and its recommended product result.

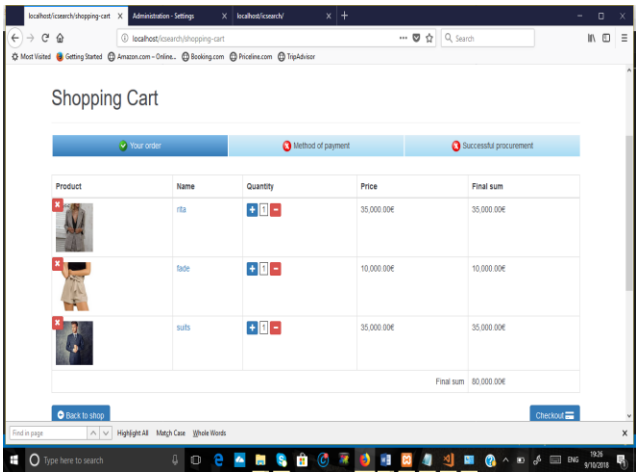

Figure 1.8. shows Shopping Cart Page.

In this page items are saved while shopping is still taking place or are saved for future shopping. Your order, method of payment, successful procurement, back to shop and check out are all found in this page.

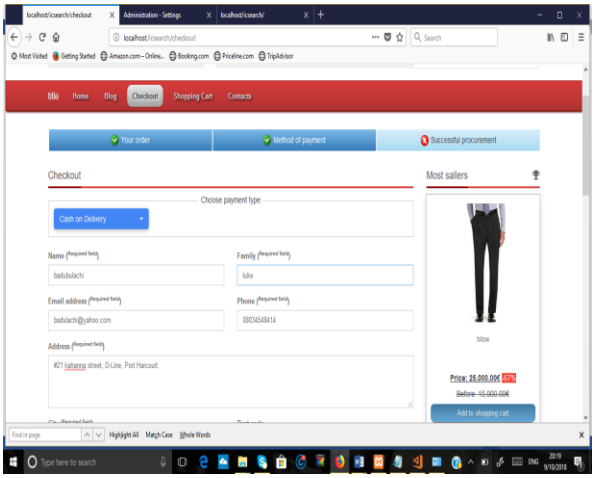

Figure 1.9. shows Checkout Page.

This page shows the method of payment, details of the user, remarks, discount codes, most sellers, the selected product, back to shop and make an order.

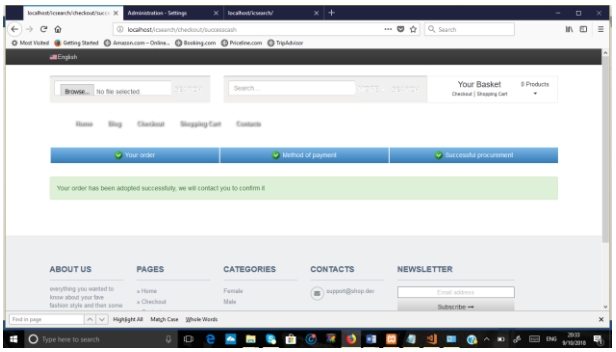

Figure 1.10. shows Successful Procurement Page.

Conclusion

This research reach obtained a better result using DenseNet which make computation more effective, reduces parameters, has strong gradient flow thereby error signal can be easily propagate to other layers more directly. The proposed system has implicit deep supervision thereby making other layers get more direct supervision.

Reference

Aizawa P. and Ogawa M. (2015). "Foodlog: Multimedia tool for healthcare applications". IEEE MultiMedia, 22(2):4-8.

Bengio, Y. and LeCun, Y. (2007). *Scaling learning algorithms towards AI*. In Large Scale Kernel Machines, 19.

Clough, P. and Eleta, V. *Investigating Language Skills and Field of Knowledge on Multilingual Information Access in Digital Libraries*. International Journal of Digital Library Systems (IJDLS), 1(1), 89–103.

Delalleau, O. and Bengio, Y. (2011). *Shallow vs. deep sum-product networks.* In NIPS. 19, 556

Hagen, P., Manning H. and Paul, Y. (2000) *Must Search Stink?*The Forrester Report. June.

Hinton, G. and Shallice, T. (1991). Lesioning an attractor network: investigations of acquired dyslexia. *Psychological review*, 98(1), 74.

Murdock, V., Kelly, D., Croft, W., Belkin, N., and Yuan, X. (2007). *Identifying and Improving Retrieval for Procedural Questions, Information Processing and Management*, 43, 1, 181–203.

Pascanu, R., Gülçehre, Ç., Cho, K., and Bengio, Y. (2014). *How to construct deep recurrent neural networks*. In ICLR'2014. 19, 199, 265, 398, 399, 400, 412, 462.

Shankar Devashish, Narumanchi Sujay, Ananya H.A., Kompalli Pramod and Krishnendu Chaudhury. (2017). *Deep leaning based large scale visual recommendation and search for E-Commerce*. http://arxiv:1703.02344v1. CS-CV.

Yamaguchi K., Kiapour M., Ortiz L., and Berg T. (2015) Retrieving similar styles to parse clothing. IEEE Trans. Pattern Anal. Mach. Intell., 37(5):1028-1040.

Ren S., He K., Girshick R., and Sun J. (2015) Faster R-CNN: Towards

real-time object detection with region proposal networks. In Neural Information Processing Systems (NIPS).