UMONS
Université de Mons

Faculté Polytechnique

Master thesis: AI-based Analysis of Luxury Brands Images on Social Media

Master 2 - Signals Systems and Bio-engineering

GeorgesTSOLAKIS

POLYTECH MONS

Under the supervision of M. Bernard GOSSELIN, MMe Virginie VANDENBULCKE and M. Kevin EL HADDAD

2021

pôle hainuyer

Contents

Chapter 1

Abstract

Social media are among the main communication channels in today's society. From regular conversations and messages to advertising campaign's, social media became an important tool to master due to its popularity and reach. This study focuses specifically on the analysis of images used by luxury brands on the social media platform Instagram. In this work, with the help of deep learning and signal processing techniques, we expose similarities and differences to the content generated by creative directors from different brands. While doing this, we also created an automated process for analyzing social media content focusing on Instagram, and images in general. The results will help understand better the promotional methods these brands use for their products and the way they affect their audience.

Chapter 2

Introduction

In today's age, social media play a fundamental role in our society. From communication between people to organizing and running a business, social media and their use is necessary. This is why their understanding plays a crucial role [6]. This study focuses on the analysis of luxury brands on the social media platform Instagram. The use of Instagram for brands has seen an exponential rise over the last years [3]. By promoting their products to a vast amount of clients, from every corner on the world, brands can see their market share evolve in a dimension never experienced before. For this reason, the understanding of how brands approach this task is interesting. The purpose of this research is to find common or different elements used by the creative directors of each brand. In addition, the purpose of this study is to create an automated process for analyzing social media content focusing on Instagram, or more generally images. The results can help us understand better how these brands promote their products and they influence their audience. Furthermore, using the metadata from the data, such as likes and comments, in the future a prediction system can be foreseen in order to see which luxury brand succeeded to engage more people in their accounts. For demonstrative purposes the use case of this study is limited to the social media platform of Instagram. Even though all the analysis is based on this platform, with a minimal number of changes the same algorithms, networks and tools can be used to perform analysis on a different platform and on a different sector.

This paper is structured in the following way. First, in the section 4 we expose the reasons that inspired this study. We present why analysing luxury brands on social media is important and why their understanding will be beneficial. Then, in section 5 we mention the previous studies made on this topic and we focus on the improvements that this work will provide. We also present the different approaches, in comparison to previous studies, that will be implemented. After that, in the section 6 we detail the whole process of getting the data, the type of data, and where we store them. We also provide some quick statistics to inform the reviewer about the time it takes to get the data necessary for this study. In the section 7 we present the selected features deemed interesting for this study. For each feature we will mention the network architecture selected, the way that this network was trained, the data used and how the evaluation was made. Each network was evaluated independently so that we are sure that we can use its results for analysis later on. In addition, in section 8 we analyse the results we get from the networks presented in section 7. We draw conclusions for each network and we also perform a more generalized analysis by using a Kmeans algorithm to take under account all the features together. In the next section 9 we expose our conclusions made for this study and finally in section 10, we suggest improvements that can take this study a step further.

Chapter 3

Acknowledgements

I wish to thank MMe Vandenbulcke Virginie for proposing this subject for my master thesis and M. Gosselin Bernard for giving me the opportunity and access to resources to go through with it. In addition, i wish to thank M. El Haddad Kevin for making this project a reality and for his continuous support throughout this study. It is whole-heartedly appreciated that their great advice was proved monumental towards the success of this study.

Chapter 4

Motivations and Contributions

This study is partially inspired by the paper of Mercanti-Guérin and Christel de Lassus [16]. In this paper the authors try to find similarities and differences between the content posted by luxury brands on Instagram. Their analysis remains on the surface by just classifying the thematic of the photos. We wish to improve this analysis by selecting more features that will make the differences or similarities of luxury brands on Instagram more apparent. We wish to interpret the results of each future such as age and gender appearing in the photos but also combine the selected features to draw more precise conclusions regarding the tendencies used by artistic creators of luxury brands on Instagram. The authors of [16] manually select their data and draw conclusions by analysing them. This is also a point we wish to improve. By providing an automated and scalable method of data harvesting, that can be also used in sectors different that luxury brands, we wish to draw conclusions that can be generalized more easily since we make use of a larger amount of information.

Chapter 5

Previous work

Social media has been the topic of a lot of research papers during the last years. Their use and effects, either in communication, marketing or other sectors have been studied in depth in various papers (ie [6], [3]).The paper [6] presents the need for social media advertising. It states that in today's age it is impossible to consider a marketing strategy without the use of social media. The second study [3] also emphasizes on the fact that brands who have invested in social media marketing has seen their products receive more exposure and sales. The recent development in deep learning and machine learning algorithms and the increase of cheaper computing power opened the door for a new kind of analysis. Image and text analysis can be very interesting methods of analysing luxury brands. Through the use of photos, tags and descriptions brands try to get the attention of a certain target group and promote their products. Being able to analyse in depth the elements of the photos, texts etc gives us a clearer understanding of the methods they endorse. At the moment of writing this paper, there are, to the best of our knowledge, very few studies that try to perform a complete in depth automated study of luxury brands on social media.

5.0.1 Deep learning luxury brands analysis

Our study is also inspired by the work of Yu-I Ha [12]. The paper [12] uses a balanced method of employing deep learning algorithms but also manual labelling of data. They proceed to the extraction of image features such as the information regarding the type of image (Body shot, selfie, face shot). On top of that they gather stats on how many images the product is uniquely present and the category of the posted image (street fashion, haute couture, designer fashion). By exploiting these features and using the number of likes in each image they are able to make some conclusions on what attracts more feedback from customers on social media. The study comes to the conclusion that product-only images make up the majority of fashion conversation in terms of volume, body snaps and face images that portray fashion items more naturally tend to receive a larger number of likes and comments by the audience. Even though this paper [12] studies also the interactions the users have with the content (likes and comments), these information are out of the scope of the analysis presented on our study. Finally, they collect a data-set of 24,752 labeled images on fashion conversations, containing visual and textual cues, available for the research community. This further evolves the progress of building an automated tool for classifying fashion information. Our project tries to combine the main ideas of the 2 studies [12] and [17] and serve as a complete automated analysis of Luxury brands on Instagram. We try to find common points and differences between brands like in the paper of Christel de Lassus, Maria Mercanti-Guérin [17] by exploiting features similar to the paper [12]. Our study uses machine learning algorithms to achieve this task. This focus on the use of machine learning algorithms makes it so that we need to examine the state of the art on machine learning based systems. For example in regards to the age-gender detection we are inspired by the paper [10] and [2] and we use a CNN network. CNN networks use filters that extract features from an image. By making a set of distinguishable features the are capable of predicting the age interval of a person. In regards to the object detection, inspired by the paper [4] we make use of the YOLOV4 network.

Chapter 6

Data

This chapter explains the whole process of getting the data, storing them into a corresponding database and also explaining the type of information we are able to extract using the Instagram API.

6.1 Data Harvesting

Previous research topics that focused on a similar subject downloaded manually a set of images from Instagram and analyzed them [17]. The goal of this project is to create an automated process capable of getting a vast amount of data. Since we will make use of machine learning networks we need a large amount of data in order to produce accurate results. By creating an automated process we will facilitate the collection and processing of large amount of data for our study, giving us more analysis opportunities and generally more accurate results. For this reason we are using an API capable of getting automatically images from Instagram. It is also worth noting that even though this paper focuses on luxury brands, it is possible for anyone interested to use the same data harvesting method and apply it on other fields than luxury brands.

6.1.1 Instagram API

There are many available packages that can automatically get data from Instagram. For the purpose of this study the Instaloader API was used [1]. There are a lot of alternatives available online but it is worth noting that the official support for the Instagram API ended on June 2020. For this reason before the final decision on the use of a specific package, it is important to check that all available options from this package are still supported and fully functioning. Regarding the use of Instaloader, there are mainly two ways that we can get data. The first way is to specify a tag and get all the data related to it. This is an easy way to obtain a good amount of data. The downside of this approach is that we also get unrelated data, or data not created by the official brand account. The second approach of getting data, requires more user input. The user needs to create a list of luxury brand names and use them as input on Instaloader. The API then goes to each username provided and downloads the number of pictures or videos requested by the user. Using this option we are sure to get the data we need directly from the luxury brand page. It is though more time consuming.

[1] `https://instaloader.github.io/`

6.1.2 Metadata

Along with the images, which is the main data type interesting for this project, we also collect different metadata for each downloaded image. This includes geo-location, time of posting, likes, comments, hashtags and post description. Even though these data are not used for the analysis part presented in section 8, in the future they can be used with other data types to study user interaction (likes, comments, etc) with the brand content. The data can also be used to build a forecasting and automatic analysis system of the user behavior.

6.1.3 Instagram API limitations

Almost every social media platform reduces the size of the images after uploading them. This is a quick and easy way for social media platforms to save bandwidth. For this reason downloading a single image with the Instaloader is not time consuming. By this size reduction though, the resolution of the image decreases and this leads to some information loss. The only time that we need to consider is the time that the API connects to Instagram and performs a tag search or goes through the username list provided by the user. The main time consuming restriction comes from the distributed denial-of-service (DDoS) protection that almost every website has. In more detail, for every image that Instaloader API needs to download, a request to Instagram's web server needs to be send. If we send too many requests in a short period of time the search and downloading gets stopped. To overcome this restriction the following method was implemented. In the code written for the use of Instaloader API we wait in between downloads of batches of images a certain amount of minutes (in our case 2). This step improves the time we need to get more data (since our connection is not getting blocked) but it also introduces a new delay, 2 minutes in our case. So even if in the global time scale we reduce the needed time, this option isn't optimal. For this reason, a second tactic was implemented, the use of proxy servers. We specify in our code that for every 10 downloads the connection needs to change to another SSID (another Wi-fi network). By doing this, we can download more images without having to wait and without having our connection blocked. Of course, this option also has its limitation which is the amount of proxy servers (different Wi-fi networks) that we can use. At the end we use both strategies in the following way. We download from a single connection the right amount of images to get our connection almost reaching the limit of available requests and then we switch to another proxy server. At the end of all the proxy servers we wait for 2 minutes. In this time, our first proxy server has the right to send again requests. We repeat this process until we reach the desired amount of collected data. By using the method explained before, we are able to get 700 images in almost 2 hours. We selected 7 luxury brands that we use as input for Instaloader. We make use of 3 proxy servers that we switch every time that we reach the requests limit.

6.2 Databases

After performing the data harvesting operation, a decision needs to be made on how to store our data. Storing the data in the right way is as important as getting them. We need to be able to automatically store the data that we get from the Instagram API into the database. The different systems created for the analysis need to be able to access this database to get the data. After extracting features from the machine-learning systems, the results need to be stores back into the database to a corresponding field. In order to have an efficient automatic data harvesting system, we need to be able to gather, store and access the data automatically and efficiently. Indeed a slow or inefficient storage system will induce a limitation in the data collection. Similarly a slow and inefficient access to the data will limit the later processing steps. Machine learning tasks are time consuming and computationally expensive. We wish to not introduce further delays by having a slow and limited access to the data.

6.2.1 MongoDB Vs SQL

There are mainly two options for storing elements in a database. The first one is creating a relational database (SQL) and the second is a non relational database (NoSQL), or in our case MongoDB. A choice was made to use the latter. MongoDB has the advantages of being easily scalable and capable of handling unstructured data. Instead of the table based storing method of SQL, MongoDB has the ability to store data in a document form, similar to Json [11]. This is a very interesting element for this study cause we need to be able to dynamically add fields in the database for every new system created. Another useful element that explains why this database type was preferred its the ability to access data by providing a specific key. This made the access and research of data incredibly comprehensive and fast. The only limitation of MongoDB is that each stored individual element (document) is limited to 16mb. This restriction doesn't affect this study, since instead of saving the actual image on the database, we save a direct link to it and we download it if necessary. This also makes our database easily transferable since there is no actually information stored on a local computer. Regarding the architecture herself, we opted for a simplistic one. Each of the images that we get by the Instagram API is considered as a separate document into the database. If we have 1000 images, we then have 1000 documents stored into the database. Each document contains specific fields in the form of a JSON file. We have the following fields inside each document.

- Username (Name of luxury brand)
- Source (Social media name)
- Image URL (Direct link to the image)
- Metadata
- Systems

It is worth mentioning that the username, source and Image URL fields contain a single value while the fields metadata and systems are more complicated. The content of metadata is discussed in a previous subsection 6.1.2. Regarding the systems field, the main idea is that for each system created we are gonna place inside the system field the name of the system and its results. For example, if we have a color detection network we name inside systems a field color detection and we place blue as its result.

Chapter 7

Machine Learning-based Feature Extraction

In this section the systems used for the extraction of features are presented. During this study we are interested in the following features, Age-Gender, Principal image color, Facial expressions, Visual attention, Object detection, Object color, Image Captioning and Optical character recognition. Different models were implemented for each feature. For every model, its architecture, data, training and evaluation is presented. The features selected for this study are not the same as the ones in previous studies. We selected the features based on what we perceived as interesting for the analysis and also on the available open-source systems as it was required by my supervisor. After the evaluation of each system and the interpretation of its results we keep the systems providing accurate results that we can take under consideration for analysis.

7.1 Age-gender classification

A very important information for social media analysis is if a person is present on the photo or not. Some brands make use of models in the majority of their photos while others focus simply on the product with no other distraction. If there is a person present, it is important to know their gender and also their age. It is interesting to be able to detect this information and thus get an insight look on how brands tend to attract their customers. There are many marketing techniques executed by brands. Some of them focus solely on their products while others play with human emotions and desires to make a product desirable. Luxury brands can use a person with a gender identical to the gender of their target group. By doing this, they allow the customer to connect, inspire to become like the model and eventually getting the advertised product. Another usual tactic is the use of different genders between the model in the photo and the target group. By doing this, brands are able to pass the unconscious message that getting the product will make the customer desirable for the other gender.
Another useful information is the age of the person presented in the advertisement. The most common trend in brands is to use people in their photos with a similar age as their target group. This allows customers to connect more easily to the model and make the product more relatable to them. For this reason, and in order to have a solid foundation for the rest of this study we deemed important to perform age and gender detection in our data.
In order to achieve the task mentioned before we made use of the following network [14]. The full model can be found here [1]

7.1.1 Architecture

The implemented model has the following architecture. It is a convolutional neural (CNN) type network with multiple layers. CNN type networks have seen a wide use in deep learning applications. They are used in image and video related applications due to their advantage of being

[1] `https://github.com/GilLevi/AgeGenderDeepLearning`

able to remain invariant to geometrical transformations and learn features that get more and more complicated and detailed. This fact makes them powerful feature extractors thanks to their convolutional layers. On top of that, they combine the extracted features and aggregate them in a non linear fashion to predict the output. Due to this fact, convolutional neural networks are widely used as classifiers thanks to their fully connected layers.
The exact architecture of the network used in this study can be seen in the figure 7.1

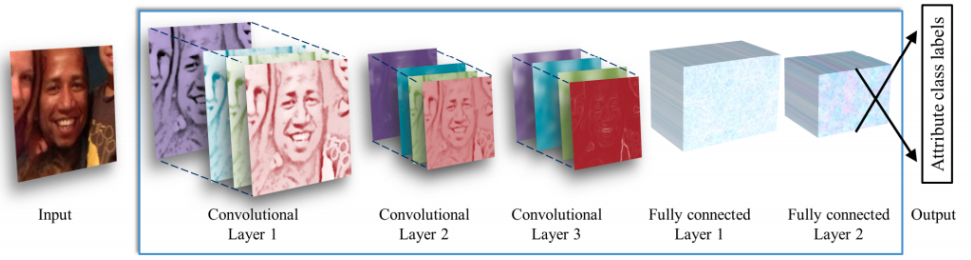


Figure 7.1: Age-Gender model architecture [14]

The system is made up of 3 convolutional layers followed by 2 fully-connected layers. This CNN network is quite small in size. This was made in order to reduce the 2 main problem that come with age and gender classification. Deep learning algorithms require a lot of data in order to be precise and to avoid over fitting. Age and gender information are most of the times private. For this reason is it a complicated task to get a hold of a big data-set containing these information. For these reasons, the network architecture remains quite simple. The network first uses the dlib [2] library in order to perform face detection. After the face is detected then the network performs age and gender classification. At the end we obtain probabilities for each of the classes for age and gender respectively and we select as prediction the class with the highest probability. It is worth noting that the system is capable of detecting many faces from one image and thus performing age and gender detection for many people at once.

The detailed layer by layer analysis of the network is given in the annexes. 11.1

7.1.2 Data

The network responsible for age-gender classification was trained with the Adience data. This data-set includes 26K images of 2,284 subjects. The Adience set consists of images automatically uploaded to Flickr from smart-phone devices.
A breakdown of the data-set can be seen at the image below.

	0-2	4-6	8-13	15-20	25-32	38-43	48-53	60-	Total
Male	745	928	934	734	2308	1294	392	442	8192
Female	682	1234	1360	919	2589	1056	433	427	9411
Both	1427	2162	2294	1653	4897	2350	825	869	19487

Figure 7.2: Adience data-set

As seen by 7.2 the data-set is separated in 8 classes. These classes contain the age intervals that the system is able to predict. In our specific case we use this network to predict on which interval the person in a photo from a luxury brand belongs to. Regarding the gender, there are three classes male, female or both. The first step of the system is to perform face detection. The

[2] http://dlib.net/

face detection library used is capable on detection multiple faces on an image. For this reason the third class "both" was created. If two people of opposite gender are present on the image the systems output "both".

7.1.3 Training

Aside from a non-complex network architecture, during the training phase two additional methods were used to limit over fitting. First, the application of two dropout learning layers of 0.5 % ratio. By introducing a dropout layer into the network we force the network to ignore some neurons during the training phase. This has as result to normalize the knowledge through the different neurons. A second method was used during the train of the age-gender network. Data augmentation, by taking a random crop of 227 $\times$ 227 pixels from the 256 $\times$ 256 input image and randomly mirror it in each forward-backward training pass. Finally, the training itself was performed using stochastic gradient decent with a batch size of fifty images. The number of epochs is set to 10.000. The pre-trained model provided by the authors of this network was used during this study.

7.1.4 Evaluation

After training the network and in order to evaluate its performance, 1000 samples from the database that have not been used during training were used to create predictions and evaluate them. As we can see in the following tables the task of gender prediction is easier than the task of age prediction. This is fairly simple to understand since the necessary features to distinguish between male and female are simpler than the ones needed to predict the correct age of a person. On top of that, if we have a set of 1000 images all of them can be used to train the system for gender classification. On the opposite, we will have approximately 100 images per age interval if we wish that our system predicts 10 different age groups. We can see the results of the evaluation for gender and age in the tables 7.1 , 7.2 respectively.

Gender prediction	Accuracy
Tested on 1000 images	77.8 %

Table 7.1: Gender evaluation results

Age prediction	Accuracy
Tested on 1000 images	49.5 %

Table 7.2: Age evaluation results

Even though the model was evaluated by its authors it is important to evaluate the model on our own data-set. This necessity is mainly due to 2 reasons. Firstly, the goal of this project is to find similarities and differences between the 7 selected luxury brands. For this, we need to have accurate information about age, gender etc in order for our conclusions to be precise. For this reason, we evaluate the model on our data-set before using it. Second reason for this evaluation is that our images from Instagram are quite unique. Models take extreme poses and the images are too complicated. From products, to models posing and sometimes even text present on the image, we need to make sure that the selected model is capable of producing accurate results.
In order to perform the evaluation, we take 100 images and annotate them. This annotation was made in 2 ways. For every image that we get using the Instaloader API we also get the metadata. If the name of the model in the photo is present we perform a search to find his/her age and gender. For the models for which the names are not indicated we manually annotate the perceived age and gender of the model. In our case this was done by one person. This mixture of subjective data (perceived age) and more objective ones (age collected from online sources such as Wikipedia) decreases the consistency of the evaluation of the system. This is not the optimal way to perform

this evaluation and its a point that can be improved in the future. Age prediction in our selected system is divided into age intervals and thus reducing the risk of errors during annotations (since we need to predict an interval and not a precise number). Regarding the gender, this has a low risk of not being accurate in a manual annotation. Future work will focus on improving these steps and make the annotation method more consistent.

The tables 7.3, 7.4 show the evaluation results for gender and age in our data-set.

It is worth noting that before we make an estimation for the age and gender of each person in the picture we need to detect a face. For this reason, the evaluation of the face detection is as important as the evaluation of the age and gender prediction. Even if the age and gender prediction is 100% accurate the system won't perform well if only a low percentage of faces is detected. There are many ways we could have perform this evaluation. The method followed for the age and gender evaluation of manually labelling images and checking them with the predictions of the network is a time consuming method. For this reason, to validate the accuracy of the face detection we followed another tactic. On a later section we present the use of YOLOV4 in order to perform object detection on our data-set. YOLOV4 is also considered state of the art for face recognition. For YOLOV4 a face is also an object presented in its classes capable of predicting. In our system, we can count the faces detected on our data-set since if a face isn't detected the system cannot proceed to the age prediction and thus produces an error. By simply summing the number of these errors and subtracting them from the total number of photos we obtain an approximate number of detected faces. Then we run the object detection network and we keep only the prediction for the class that corresponds to the face of a person. We cross check the results and we see that both of the systems detect a very similar amount of faces. Of course this is not a strict accuracy evaluation but it gives us an approximate idea of what we can expect. For reference, both the systems detected a face in close to 400 pictures.

Gender prediction	Accuracy
Tested on 350 images	69.36 %

Table 7.3: Gender evaluation results

Age prediction	Accuracy
Tested on 350 images	46.82 %

Table 7.4: Age evaluation results

7.1.5 Interpretation

As we can we by the comparison of the tables 7.3 - 7.1 and 7.4 - 7.2 the accuracy of the network is lower than the evaluation made by the creator of the network. This corresponds with our expectation since as explained before our data-set is more complicated than a simple face image. Even though the accuracy is lower we feel like the results are capable of being taken into account for our analysis. For this reason this system was used during the final analysis.

7.2 Color detection

Another interesting element that can be used for analysis of luxury brand photos is the colors appearing on the image. It is important to know how brands try to get the attention of their customers. Some of them make use of vivid colors while others prefer black and white or other variations.

7.2.1 Architecture

For this task we decide to make use of KMeans clustering. The Kmeans clustering algorithms tries to split a data-set into a number of fixed K clusters. The first step is to find the center of each cluster called centroid. Centroids are data points in the given data-set. They are picked initially at random so that each centroid is unique. They are then used to train a classifier that will classify the data and give a set of random clusters. After this step each centroid is set to the mean value of the cluster it defines. The process of classification and centroid correction is repeated until the values of the centroids are stabilized. In the final repetition of the process, the centroids are used to cluster the input data into classes.

7.2.2 Data

In our specific use case the Kmeans algorithm clusters the pixels of every image. Given an example image of NXM dimensions, we obtain N*M pixel values. Each pixel is characterized by its RGB value. We are clustering pixels characterized by their RGB values. Pixels that are in the same cluster have similar colour. Kmeans algorithm requires to pre define the number of clusters. We can either define them manually or apply another algorithm responsible of finding the optimal number of clusters. For this study, as a first step we chose 3 clusters, in order to get the 3 main colors used in the pictures. Later in a following section a more detailed color detection on certain parts of the image (object and visual attentive area) will be presented.

7.2.3 Training

Since we actually cluster the RGB pixel values, the training of a network is not necessary. Kmeans is an unsupervised method that doesn't required labeled data and the training of a network in order to make predictions.

7.2.4 Evaluation

Since we actually cluster the pixel values, there is no evaluation part of the system. No predictions are made, we get the actual values and we assign them to clusters.

7.2.5 Interpretation

At the end of the clustering process we have groups of pixels with similar values. After every image we use the library Webcolors available in python, to get the name of the color that corresponds to the RGB value. Then, we cluster the color information of each luxury brand and we look at the percentage of each color. The downside of this method is that not every RGB value combination is named. When Webcolors cannot assign an exact name to the RGB value, it assigns the closest color cluster name. This fact reduces the accuracy of our results.

7.3 Facial expression detection

The visual attentive area in photos on fashion magazines is the focus of many existing studies. Most of these studies concluded that men and female models focus on different body parts. It is demonstrated that men focus more on their face while women more on their body [8]. Another interesting conclusion is the fact that women tend to express less serious emotions in fashion photo shoots while men go for more serious looks. Since such variations exist, it is interesting to see

what facial expressions models for luxury brands, on online platforms, use the most. By obtaining this information, we get one step closer to understanding how luxury brands try to attract their customers and the feelings they want to provoke.

In order to achieve the task mentioned before we implement the following network called MicroExpNet [7]. This network is widely used and one of its best advantages is its small size and increased speed. The model is trained for 8 different classes. Those classes are the following: neutral, anger, contempt, disgust, fear, happy, sadness, surprise. These 8 facial expressions constitute the 7 main expressions plus the neutral expression [13].

7.3.1 Architecture

The main architecture of the network is based on the InceptionV3 architecture. InceptionV3 has a proven record of success on classification tasks [19]. The architecture of MicroExpNet is hybrid. Except for the InceptionV3 network we also get the following separated architecture, two convolutional layers (conv1, conv2) and two fully-connected layers (fc1, fc2). We use rectified linear units (ReLU) as activation functions. There are max-pooling layers after each convolution layer. The small size of MicroExpNet is achieve by implementing the popular student-teacher compression method. In our case InceptionV3 plays the role of the teacher network and the simpler architecture of 2 convolutional and 2 fully connected layers the role of the student. Teacher and student compression (TSC) is a widely used form of compression in Deep learning applications. The main idea is reduce the time and resources needed for the training by training a less expensive student model to mimic a more complicated teacher model, while most of the accuracy is maintained.

7.3.2 Data

Regarding the data used, MicroExpNet is trained with the CK+ benchmark database for facial recognition. The data-set contains 327 image sequences with eight labels that correspond to the 8 classes mentioned before. The total number of images is 1574 and they are spitted as shown in table 7.5

Anger	Contempt	Disgust	Fear	Happy	Sad	Surprise	Neutral
135	54	177	75	207	84	249	593

Table 7.5: Data split for facial expressions

7.3.3 Training

For the training itself the InceptionV3 network originally trained on ImageNet was fine tuned using the CK+ data-set. The number of epochs for the training is set to 3000 and the table 7.6 shows the classification performances. It is worth noting that the compression method of Teach-Student is applied and this is why in the performance table we obtain two different results.

Model	Performance on CK+
TeacherExpNet	97.6%
MicroExpNet	84.8%

Table 7.6: Data split for facial expressions

For the purpose of this study the pre-trained model created by the authors of the network was used.

7.3.4 Evaluation

One of the main challenges during this study was facial expression detection. Models in photos pose with unique facial expressions that aren't very common in most databases. On top of that most of the trained networks for facial expressions are focusing on front facing photos. This is not the case for most of the photos in our database from Instagram. Also, the images in our database are quite complex, having text, objects and persons appearing at the same time. For this reason, it is very important to evaluate the MicroExpNet in our database. Regarding the evaluation in our data-set, we need to have a reference to compare the predictions of the network. For this reason, we annotated 100 images from our data-set by assign a facial expression class to each of the 100 images. As mentioned before the differences between the common data-sets and our unique Instagram data-set (complex images, non facing frontal images, unique facial expressions) make so that the networks accuracy for this study is limited to 19.6%.

7.3.5 Interpretation

Facial expression detection is based on the position of the eyes, the spread of the lips etc. To classify an expression as a smile, the lips should be spread a certain distance, the eyes narrower than normal etc. If we do not have these information from our images the task of detecting an expression becomes very complicated. We can see at the figures 7.3 that our data-set isn't optimal for facial expression detection. For this reason, the facial expression detection results are not used to draw conclusions about luxury brands.

7.4 Visual attention network

Luxury brands and marketing campaigns in general employ many techniques to get the attention of their target group. By placing objects at a certain part of the image, creating subconscious messages to customers and finally with the use of models, they try to get their target to focus on the message they want to send. To be able to have information on where the artistic director of each luxury brand wants us to focus, we implement a visual attention network.

7.4.1 Architecture

With the arrival of Deep Neural Networks (DNN) many studies have been made on Deep Neural Networks saliency. The implementation of these type of models can help us gather information on many fields such as image segmentation, robotics and visual marketing. However, despite the success of DNN type networks, they present some disadvantages such as not being able to take into account low-level features (pixel intensity, color, etc) and also that they are not generic enough to adapt to images different from the training data. Another approach to visual attention is the use of Deep features models. These models do not need training. However, for the time being, they do not have the accuracy of DNN networks.
The study chosen for this particular project of analysing luxury brands proposes a model (DeepRare2019) that mixes both of the pre-mentioned models [15].

In this model a convolutional neural network (CNN) is used to extract a complete set of features (from low level to high level features). We can see in figure 7.4 the architecture of the first layer of the network. The process shown is repeated for all layers of the CNN model (VGG16). After the features are extracted, on each feature map, the data rarity is computed. The rarity function used is the R function 7.1 of the histogram of each feature map on a few bins. Once the occurrence probability p(i) for the pixels in all the bins is calculated, we obtain the rarity histogram R. We then use the histogram of a feature to find this feature in an image projecting each histogram value

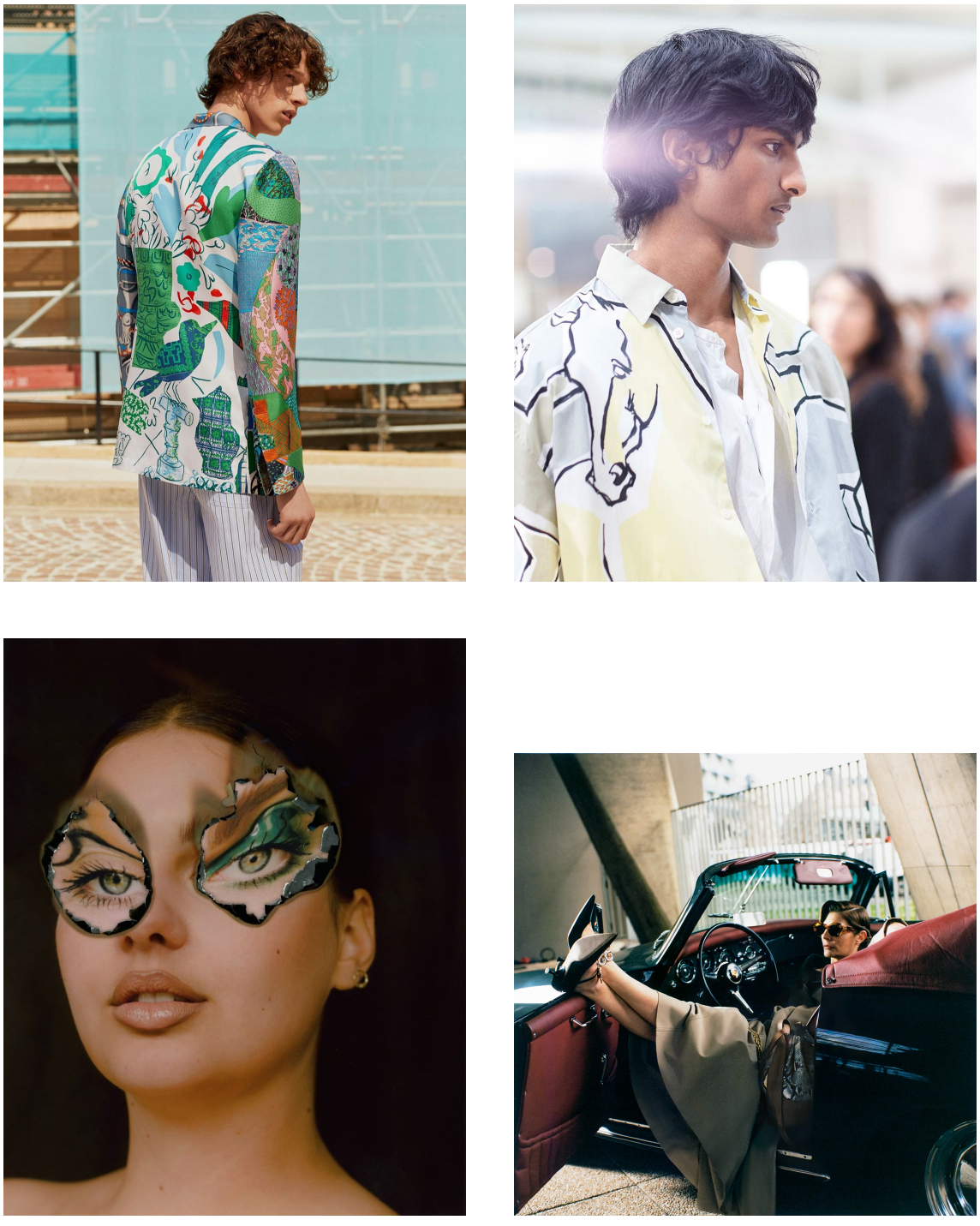


Figure 7.3: Difficult to detect facial expressions images

on the corresponding pixel in an image. This will highlight pixels in the feature map which are rare compared to the other pixels in the feature map.

$$R(i) = -log(p(i)) \tag{7.1}$$

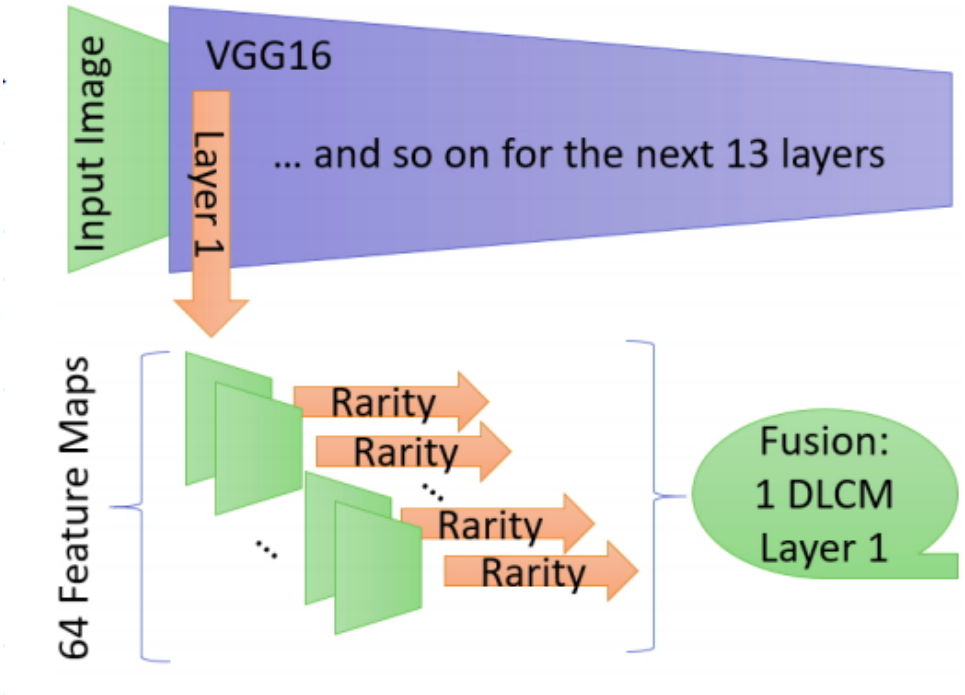


Figure 7.4: Processing for Layer 1. This processing is iterated for all 13 convolutional layers from a VGG16 network. Figure taken from paper [15]

7.4.2 Data

The data used for the training of this network is the ImageNet data-set. ImageNet is a visual database for use in the image recognition sector. It contains more the 14 million images that have been annotated by humans and indicate objects belonging to more that 20.000 categories

7.4.3 Training

The VGG16 network is trained on the ImageNet and the Deep feature model needs no training.

7.4.4 Evaluation

The creators of DeepRare2019 used 3 data-sets (MIT1003, O^3, P^3 to evaluate their creation. We can see the results 7.5

As with most of the networks in this study, we do not remain on the evaluation that the author of the network made but we evaluate the models on our own data-set. Since this data-set presents unique features and not so common images, we need to be sure that the system produces accurate results before we take into account its results. Validating the results of a visual attention network

	AUCJ	AUCB	CC	KL	NSS	SIM
SAL	0.83	-	**0.51**	**1.12**	**1.84**	**0.41**
DR	***0.86***	***0.85***	*0.48*	*1.25*	*1.58*	*0.36*
MLNet	0.82	-	0.46	1.36	1.64	0.35
DFeat	0.86	0.83	0.44	1.41	-	-
eDN	0.86	0.84	0.41	1.54	-	-
GBVS	0.83	0.81	0.42	1.3	-	-
RARE	0.75	0.77	0.38	1.41	-	-
BMS	0.75	0.77	0.36	1.45	-	-
AWS	0.71	0.74	0.32	1.54	-	-

(a) Evaluation on MIT1003 data-set

Model	MSR_t	MSR_b
MLNet	0.96	0.91
SALICON	0.90	1.26
DR	***1.06***	***0.89***

(b) Evaluation on O^3 data-set

Model	Avg. # fix.	% found
MLNet	42.00	0.44
SALICON	49.37	0.65
DR	***16.34***	***0.87***

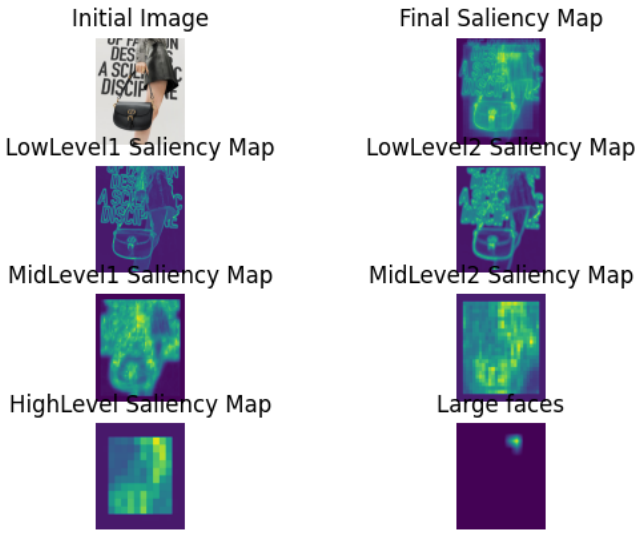
(c) Evaluation on P^3 data-set

Figure 7.5: Evaluation on multiple data-sets

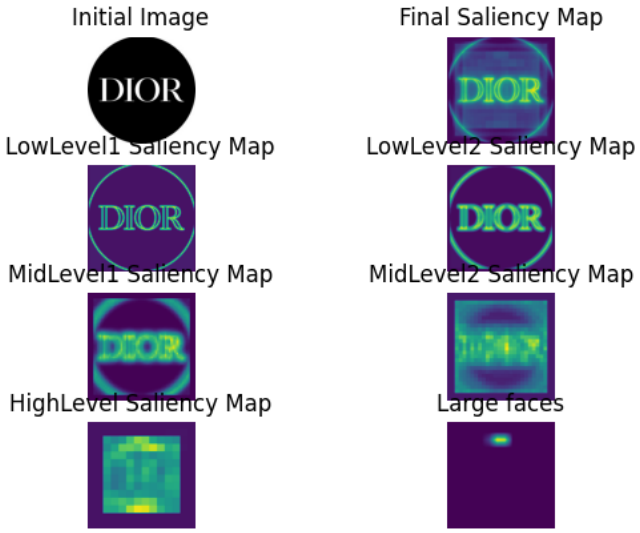
is objective, at least when its done by only one annotator. The validation process followed here is the following, we annotate 50 images after we execute the DeepRare2019 system on them. We visualize the output and we give a label to each results of the system. The labels range from bad to good to medium. We can see on the table 7.7 that the system performed very well. Examples on what the system produces and that corresponds to an accurate visual attention results are given in the figure 7.6.

Good	Medium	Bad
45.71%	37.14%	14.28%

Table 7.7: Evaluation of visual attention on Instagram data-set



(a)



(b)

Figure 7.6: Example of good visual attention results

7.5 Object detection network

One very important element in our analysis is having information regarding the objects existing in the image. Just as if people are present or not, it is important to know if luxury brands implement objects in their pictures to attract the attention of their customers. Having an expensive car in the picture can give the impression that acquiring a product can lead to financial success. Brands try to pass subconscious messages through the objects they present in their photos. Research papers show to correlation between customers reaction and marketing campaigns concepts using objects [20]. For these reasons, an object detection network is used to detect if there is an object present

in the image and if there is, the type of object it is.

7.5.1 Architecture

For many years detection algorithms used classifiers to detect where they should apply the model to the image and they considered the detected object as the region of the image with the highest probability. This process is slow and needs important computational power. The birth of YOLO networks changed the object detection approach and reduced the necessary computational power. For the purpose of this project the YOLOV4 network is used. The newest version YOLOV5 has been released a few months prior to the writing of this paper. Due to some controversies about the accuracy and legitimacy of YOLOV5 a choice to use YOLOV4, which is considered state of the art, is made. In the figure 7.7 we can see the typical structure of a YOLO type network. The purpose of the different versions of YOLO is to improve the networks architecture used, the loss function, the data augmentation techniques etc. Notably, YOLOV4 consists of: Head: YOLO, SPP,PAN, Backbone: CSPDarknet53, Activation function: Leaky-ReLU.

A full description of the process of creating the YOLOV4 can be found in the official paper [5].

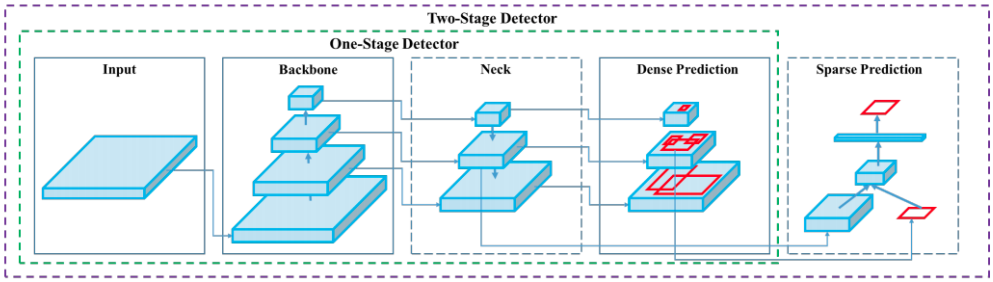


Figure 7.7: Object detector architecture [5]

7.5.2 Data

YOLOV4 is trained on the COCO data-set which contains 80 different classes. Those classes represent the number of objects the network is capable to predict from our images. The objects that the network is capable of detecting can be found in Annexes (11.2 and 11.3). We use an open source implementation of YOLOV4 [3] so that we can run YOLOV4 without training it ourselves.

YOLOV4 uses ImageNet (14 million annotated images) as its data-set and augmentation techniques are implemented in order to increase the number of data. The following methods are used to augment the data: CutMix, Mosaic data augmentation.

7.5.3 Training

As for the training itself, the default hyper-parameters are as follows: the training epochs is 8.000 the batch size and the mini-batch size are 128 and 32, respectively. The lost function is a Leaky-ReLu
Since YOLOV4 is a very complex network the reader is welcome to review the full paper [5]

[3] `https://github.com/AlexeyAB/darknet#pre-trained-models`

7.5.4 Evaluation

Regarding the evaluation of YOLOV4, for the moment it is consider state-of-the art for various reasons. The 2 great advantages of YOLOV4 is speed and accuracy. We can see in the figure 7.8 that YOLOV4 performs in general better than most object detectors.

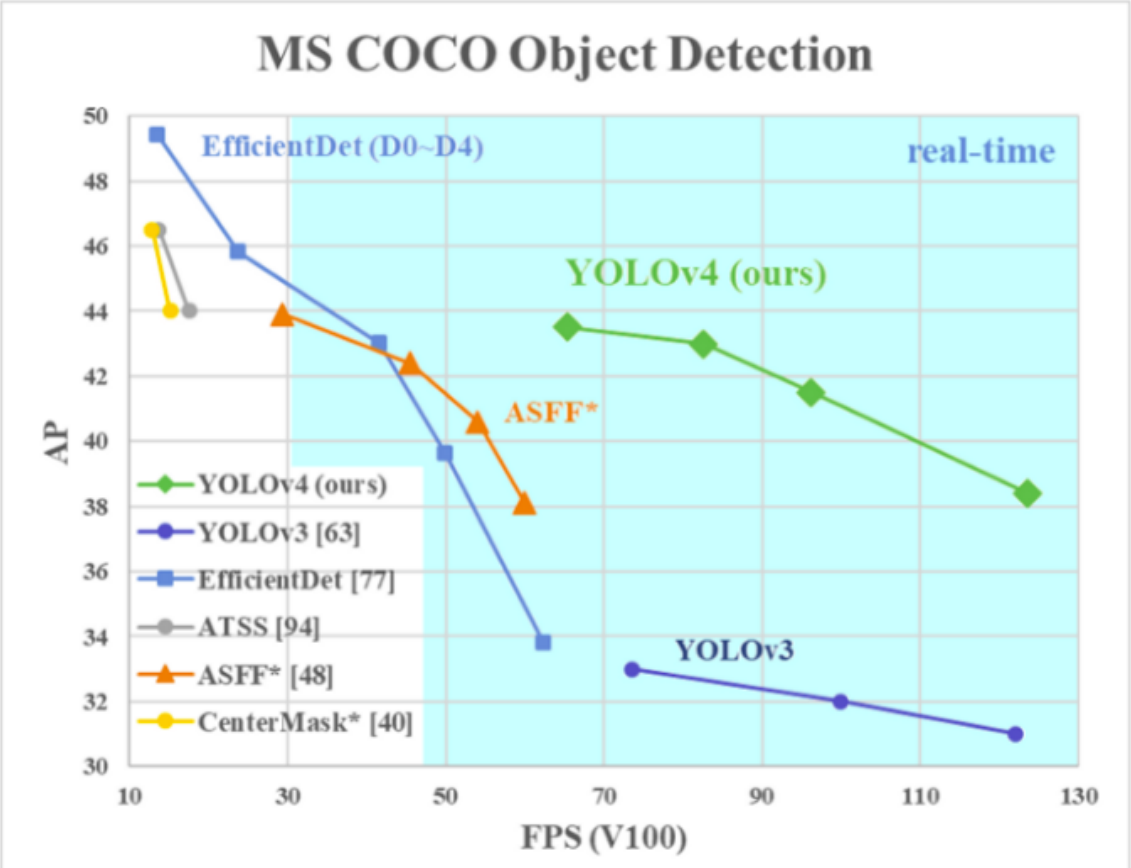


Figure 7.8: YOLOV4 accuracy and comparison [5]

It is worth mentioning that YOLOV4 is affected by the image input size. The general tendency is that higher resolution images will provide better results. We can see a comparison in the table [4] 7.8

Detection	320x320	416x416	512x512
YOLOV4%	48.58%	56.92%	61.71 %

Table 7.8: Evaluation of image captioning

In order to be produce more precise conclusions we check the accuracy of the object detection model in our own data-set. For this purpose we chose 50 images and we annotated the objects

[4] https://medium.com/analytics-vidhya/introduction-to-yolov4-object-detection-fcba8bb72449

present in the image. Even though this is still an annotating task made by only one person, we do not expect it to be inaccurate. Objects are not similar to facial expressions, they more easily distinguishable and they can be easily perceived correctly. The data-set consists of images that have a variable size. The minimum size of an image in our data-set is 320x320 while the maximum size is 1080x1350. The percentage of this 2 dimensions is 11.2% and 88.8% respectively. After using the network to predict 50 results for the validation data-set we compare it with our manual annotations. We get an accuracy of 55.24%. It worth noting that the evaluation metric used here presents the accurate object detection in each image. This means that is measures both detecting an object and also classifying it correctly to a class. While the object detection of YOLOV4 is very high the accurate prediction of the object it is still not optimal as this point.

7.5.5 Interpretation

As we can see in the section 7.5.4 even though YOLOV4 is considered state of the art for object detection, its accuracy is limited. We deemed that the objects presented in Instagram photos are an important element for our analysis. For this reason the results of the object detection network were used for the final analysis.

7.6 Image captioning network

In the majority of cases, Instagram posts have tags in their post description. These tags range from describing the product, mentioning the brand name and sometimes the name of the model. Apart from this, a useful information for this study can be the automatic description of what is happening in the picture. We can obtain like this the general concept of the image and have an idea on the main idea behind luxury brands photos.

7.6.1 Architecture

The network we use in this study is based on the InceptionV3 for extracting features from the data. On top of that a gated recurrent unit (GRU) is used to produce the captions. The pre-trained model InceptionV3 is used by using the weights of ImageNet. The GRU network is then created. This network contains 3 CuDNNGRU layers and a dense layer. The activation function of the GRU network is a linear function.

7.6.2 Data

The data-set used for this network is the MS-COCO data-set. COCO data-set contains 330K images that belong to 80 classes.

7.6.3 Training

In regards to the training itself, the loss function is set to sparse crossentropy and the number of epochs to 25. The batch size is set to 512. As for the optimizer used during the training, the creator selected a RMSprop function with a learning rate (lr) of 1e-3.

7.6.4 Evaluation

It is complicated to have an accurate view if the network produces good captions or not. To get the best possible idea about the accuracy of the network we need multiple people producing manual captions for each image. After that, the people responsible for the annotations need to cross check between them if they agree on the captions or not. Finally, the comparison of the networks captions

and the manual annotations will give us the accuracy of the network. This described process is similar to the process needed to get the accuracy of the facial expression network. For evaluation purposes we annotated 50 samples from our data-set and we verified the predictions of the network into 3 levels, good, medium, bad. We can see in the table 7.9 the results of this evaluation.

Good	Medium	Bad
5%	21%	74%

Table 7.9: Evaluation of image captioning

7.6.5 Interpretation

As we can see on the table 7.9 the image captioning network performs quite poorly on our data-set. Most of the captions produced were very simplistic and did not capture the concept behind the image. This is probably due to our complicated data-set. For this reason, in order to not draw inaccurate conclusions, this network is not taken into account for our final analysis.

7.7 Optical Character Recognition

Another interesting idea for feature extraction is Optical Character Recognition (OCR). Having information about the characters presented in the luxury brand photos (text, numbers) can be another element useful for the analysis we wish to achieve.
For the purpose mentioned before we made use of Tesseract. Tesseract is a free google optical character recognition engine capable of recognizing characters in more than 100 languages. It is considered for the moment the state of the art in OCR engines. We can easily install Tesseract from the pre build provided packages [5]. After installing the Tesseract engine we try to detect characters in our data-set. As expected, the engine is not capable of detecting characters in the majority of our images. This probably due to 2 reasons. First, OCR engines are created in order to mainly detect characters in specific scenarios (bank cheques, book scanning etc). Our colorful images, filled with objects and people make the detection more challenging. For this reason the use of this tool was dropped.

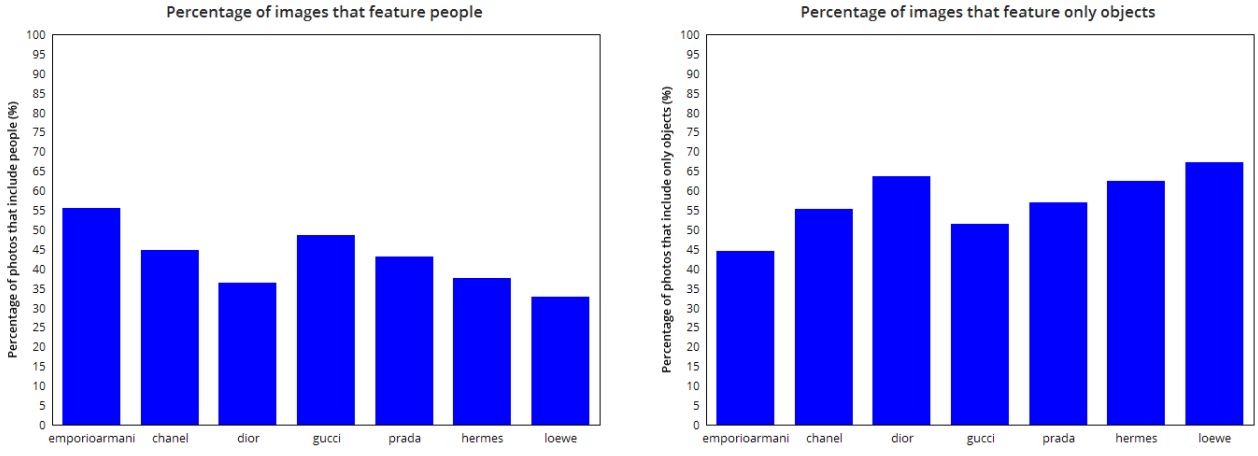
[5] `https://github.com/tesseract-ocr/tesseract`

Chapter 8

Luxury Brand Content Analysis

8.1 Humans and objects present in photos

A very crucial information for our analysis is to be able to see the percentage of people in a photo for each brand. Some brands prefer to only advertise their product with no other distraction while others try to inspire their target group by using female or male models. We present the following graphs in order to begin our analysis. First, the graph of if a person is present or not and second the percentage of pictures that show only objects. 8.1. We can see in the graph 8.1 that from the seven indicated brands Dior, Hermes and Loewe have the lowest percent of people appearing in the photos. This finding is also verified by the fact that in the graph 8.1 we can see that these 3 brands (Dior, Hermes, Loewe) make use of objects in a percentage higher than other brands. There are many reasons why brands can chose to just advertise their product with no model included in the photo. Brands sometimes, prefer to focus exclusively on the product they are selling with no distraction. On top of that, some brands produce products that don't necessarily need a model for advertisement. A good example of that is Loewe, a brand that has its biggest market share in handbags. Handbags differ from dresses, jewellery etc that need a model in order to be advertised. On the opposite spectrum emporioarmani holds a massive market share in the clothing sector. For this reason they need models to advertise their clothes more than anyone, just like Gucci. An interesting case is Chanel. As it can be seen by the graphs 8.1 they follow a very balanced strategy of using models and objects in their photo. After a closer look in the data that we got from Instaloader API regarding Chanel, we can see the this is clearly the case. This brand prefers most of the time to post a picture with the product they want to advertise using a model and also a separate advertisement with a closer view of the product. This strategy can be seen in the figure 8.2. Finally, it worth noting that most of the brands follow the same tendency of somewhat balanced use of models and objects in the photos. The slight variations in the percentages can be possibly explained due to their different market shares in various sectors.



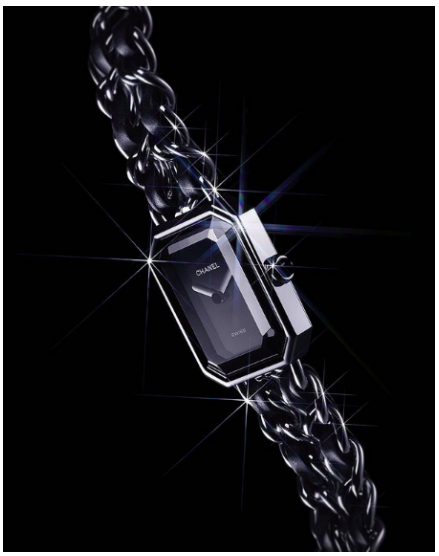
H

Figure 8.1

(a)

(b)

(c)



(d)

Figure 8.2: Examples of Chanel marketing strategy

8.1.1 Age and Gender distribution in general and per brand

In marketing campaigns, specially today, nothing is chosen in random. Artistic directors have to think every element they include in their photos otherwise the repercussions can be destructive for the brand image and also equity. The gender chosen for a model present in a image and also the age aren't random. It is then important to have the relative information about these choices. We can see in the respective figures 8.3 the different age intervals brands prefer for advertising purposes. The age interval preferred by every brand can be explained by various factors. Mainly, brands tend to prefer models with an age similar to their target group. This choice makes the target group easier to connect with the model and makes the product more desirable. As we can observe in the respective figures 8.3 most of the brands prefer models aged 20-30 or 30-40 years old. This is logical since this target group represents the biggest market group for luxury brands. It is worth noting that some brands target people older than 40 years old and some younger than 10 years old. Brands like Gucci, EmporioArmani include kids in the photo shoots while others such as Hermes and Prada include models with an age older than 40 years old. Brands like Hermes and Prada represent companies that still promote a classic style of clothing and not only following the clothing trends like Gucci and EmporioArmani. This is why Prada and Hermes focus also on people aged 40-50, cause they represent an important market share for them. On the opposite, brands such as EmporioArmani and Gucci try to monetize different trends and styles so the are more appealing for the youth.

Another important aspect is the gender. Two main approaches can be seen. The first one is to use a model with the same gender as the target group in order to make the product more relatable. The second approach can be the opposite, trying to pass the unconscious message that acquiring the product will make the user more attractive to the opposite gender. Of course as for the previous elements, the products that each brands specializes, if they have a collection for both genders plays a role on the model they use on their photographs. Gender distribution is one of the elements where we actually see some important differences between brands. In only 2 out of 7 brands (Hermes and Gucci) we notice a bigger percentage of males in comparison to females. In all the others brands the female gender is more represented. This can be explained by the fact that most luxury brands focus on women products. Women tend to be more attractive to luxury products and spend a bigger percentage of their annual income to buy products such handbags, make-up, dresses, etc. For this reason it is normal for brands to target more the female gender since it leads to a greater brand equity. On top of that, some brands simply don't advertise their male collection as much as they do their female collection, or they simply don't have a male collection. An extreme case of these facts can be seen on the gender graph corresponding to the brand Chanel 8.4. We can clearly see that this historic brand, inspirational for many women, uses female models in 90% of their photos. While Chanel has a man's collection, kh the market share it represents isn't worth investing a lot of marketing resources.

8.1.2 Dark skin or light skin color

For those luxury brands that make use of models for their photographs, it is important to know the racial profile of the models. After excluding the gray photos from our data-set we make use of the color detection network explained in the section of color detection. In order to find the skin color, we first obtain the facial region and extract it from the original image. To perform the face recognition detection we use the dlib library [1]. After this step, we execute them color detection algorithm and we detect the most used color. To make the difference between different shades of black and white we manually set thresholds and classify different shades of white and black as simple dark skin and light skin color. To verify this approach we select a sample of 20 images and verify our results by cross examining the networks prediction and seeing the actual image.
We can see in the figure 8.5 the dark or light face color percentages of each selected luxury brand. We can observe that there is an important bias in favor light face color from every brand. Some brands such as Chanel and Emporioarmani use almost 100% light skin color models. The situation is a little better for brands such as Gucci and Dior but even here the percentage of dark skin face models is only 35% and 28%. On a global scale luxury brands make use of 82% light skin color

[1] http://dlib.net/

models and 18% of dark skin models. These results correlate with different articles [18] also making this bias apparent. This bias towards light skin models has been dominant in the luxury brands industry for years. For indicative purposes in 2017 the marketing expenses for TV, radio, internet were 75$ billion dollars. From that amount only 2.24$ billion was towards a dark skin audience. We can then see that the results from the figure8.5 correspond with these facts.

8.2 Colors

Regarding the color analysis of our data-set, the following strategy was chosen. We analyze the 3 main colors used in the photos. We detect the 3 most used colors in every picture and their percentages. Of course each photo contains a vast amount of different colors but by getting the 3 most used ones we can have an opinion about the total hue and color of the image. Some brands prefer the classic gray, black and white style while others experiment with more vivid colors. In a general approach we can assimilate the 3 most used colors to different parts of the image. In most of the cases the biggest part of the images and thus the most used color is the background. This is why in the respective figure 8.6 we notice that in most of the cases the most used color is black, gray or in less frequent cases white. We also perform and present the data of other color analysis in a later subsection. More precisely, we perform the color analysis of the most attentive area of the image and also the color analysis of the object detected in each image. We can see in the figure 11.4 some examples from different brands. In order to have an easier representation of the results shown in the figure 8.6 we can see in the example images for Chanel and Prada that they use more black and dark gray. On the opposite, brands like Loewe and Hermes prefer more light colours for their backgrounds. Emporioarmani maintains their grayish look throughout most of their photos. These images are representative examples of the tendencies for each brand artistic director.

8.3 Object detection stats

Another interesting element that artistic directors of luxury brands use are the objects appearing in the photos. Since nothing is random on the published photographs it is important to have information on what these objects are. It is important to make the distinction between objects appearing in the picture for decorating purposes and objects that represent the product itself. For example in many brands (Dior, Prada, Gucci) we can detect that they posses a high percentage of handbags in their photos. This is easily explainable by the fact that these brands represent dominant players in the leather handbag industry. We can also see that many brands tend to use the same objects to decorate their photos and create a sophisticated look. These objects such as ties represent an important percentage in the total objects detected. Other objects such as books and bottles are implemented into the photographs to create a mysterious and appealing look. It is worth noting that in order for photos to stand out brands also make use of unique objects that don't represent a general tendency. These objects don't represent a repetitive trend and in some cases can only be used in one photo and never again. This category of objects is represented in the figure 8.7 under the label "others". During the object detection many objects can be detected at the same time. Many objects can be present during a photo so we can have a combination of the objects shown in the respective graphs.

8.4 Object detection colors

As we can see in the figure 8.8 luxury brands make use of various colours for their objects. These colours vary depending on the color of the products each brand produces. The main colours are black, white, red, blue, brown. Two particular brands, Hermes and Loewe differ from other brands

cause the have many brown objects presented in their photographs. This can be explained by the fact that these brands are two of the biggest players in leather bags. Since leather bags are usually brown colored it is logical than in our analysis of their data we find many brown colored objects. Specifically for Hermes, orange is also a dominant color since the logo of the brand is orange. For other brands black, silver and white seem to be dominant colors. This result seems logical since perfumes, jewellery are other luxury objects that these brands produce are mainly black and silvered colored. In smallest percentages we can detect red, gold, blue. Brands might use these colors in their products, and thus we detect them in their objects in the pictures, in order to make the product more appealing to the eye.

8.5 Visual attention colors

Like in the object color distribution we perform the same color analysis on the most visual attentive area. As we can see in the figure 8.9 the main color in the visual attentive areas are black, white-gray, red-orange and brown. We observe this tendency in most of the brands. Luxury brands such as Hermes and Loewe that have an important market cap in the luxury leather bags have the highest percentages of brown. Other brands such as Dior and Prada prefer more light colors (White, Light-Gray) in comparison to Chanel and Emporioarmani. To obtain these colors we used the network described in section 7.4. We modified the code in order to obtain the visual attentive area and select a part of the image that has the most importance, according to the network. We set a threshold manually, in order to keep the pixels of the image that satisfy this condition. After we keep this attentive part of the image we execute the color detection algorithm described in section 7.2. As for the color distribution itself we use the python library webcolors to get the color names from the RGB values.

8.6 Object color and visual attention color overlapping

As we can see in the sections 8.4 and 8.5 the color distribution that we get for each brand and also the global color distribution are similar. This is mainly due to the fact that the objects presented in the pictures, in which most of the cases are also the advertised products, are situated in the most visual attentive area. We notice that in visual attention color distribution 8.9 we have a higher percentage of black and white color than in the object color distribution 8.8. This can be due to the fact that in the most visual attentive area we also have the models featured in the photo. The color detection algorithm detects the skin color, light or dark skin and assigns it to a color. In general terms, we see that the colors from visual attentive area and the colors from object detection overlap quite a lot.

8.7 Kmeans clustering : Age Gender Main color

After getting all the necessary information as explained in the previous sections we can draw specific conclusions for each characteristic. By observing the figures of each category (age, gender, color) the general tendencies of brands become apparent. An interesting idea will be to group all these information together and see the similarities and differences. Instead of drawing conclusions for each criteria separately, as for example that Chanel uses 90% female models, we can by applying a clustering algorithm create 7 groups that correspond to the number of luxury brands selected for this study and see the differences and similarities of each brand. In an ideal case that each director would use totally unique characteristic features for the photos each brand would have 100% occupancy of its cluster. Since this is not the case, each cluster will be occupied by a certain percentage of each luxury brand cluster. For clustering purposes we use as first approach the

information we gathered from color detection, age detection and gender detection. Then we add more features in order to make the clusters more discriminatory. It is worth noting that for the color detection we introduce to the clustering algorithm the raw data (RGB values) and not color names. Since the color names given to RGB values are most of the times an approximation that the webcolors python library does, we give the raw RGB values in order to not introduce inaccuracies into the clustering algorithm. In addition, we represent the male gender as "10", the female gender as "01" and the class both genders as "00". We apply a Kmeans algorithm and we obtain the following results shown in table 8.1

Cluster/Brand	Prada	Loewe	Hermes	Gucci	Dior	Chanel	Emporioarmani
Cluster Prada	13.1%	13.94%	9.01%	**23.77%**	14.75%	16.39%	9.0%
Cluster Loewe	16.12%	15.32%	8.87%	13.7%	**18.54%**	13.7%	13.78%
Cluster Hermes	15.1%	14.28%	5.88%	15.12%	**27.7%**	10.9%	14.2%
Cluster Gucci	13.11%	**16.39%**	12.29%	18.6%	14.75%	10.65%	13.93%
Cluster Dior	9.1%	18.8%	10.74%	14.8%	**23.14%**	10.74%	11.5%
Cluster Chanel	**17.64%**	14.28%	5.85%	20.1%	15.12%	15.12%	11.7%
Cluster Emporioarmani	16.6%	13.3%	6.66%	**17.5%**	15.83%	15%	15%

Table 8.1: Clustering results

As we can see in the table 8.1 the features extracted from the data are not discriminatory enough. Most brands use similar characteristics in their photos. On top of that, this first study of luxury brands uses machine learning algorithms that are quite generic. For example even though YOLOV4 detection is considered state of the art, the prediction accuracy is only 50%. The fact that we aren't using tailor made systems for our case and also the state of current machine learning algorithms introduces inaccuracies in our predictions. This results to the cluster of each brand being made up by similar percentages of each brand. The results we get by clustering correlate with the individual conclusions we made during the age, gender, skin color, color, object detection. As we can see in each respective figure of these sections luxury brands use similar objects, similar ages, skin color models and colors in their photographs. It is then very unlikely to expect to have discriminatory clusters for each brand.

8.8 Kmeans clustering with more feautures

Even though at this point we expect non-discriminatory clusters for illustrative purposes we cluster all our features. These features include age, gender, skin color, main image color, object and visual attentive area color. We can see the results of this clustering in the table 8.2

Cluster/Brand	Prada	Loewe	Hermes	Gucci	Dior	Chanel	Emporioarmani
Cluster Prada	17.7%	14.4%	5.08%	**20.33%**	15.25%	15.2%	11.8%
Cluster Loewe	**17.21%**	13.11%	6.55%	16.4%	15.57%	16.39%	14.75%
Cluster Hermes	13%	15.7%	8.9%	**23.57%**	14.63%	16.26%	8.93%
Cluster Gucci	13.22%	15.7%	12.4%	**20.6%**	14.9%	9.1%	14.04%
Cluster Dior	9.1%	19%	11.57%	14.87%	**23.14%**	10.74%	11.57%
Cluster Chanel	15.7%	15.07%	9%	12.39%	**19%**	14.1%	14%
Cluster Emporioarmani	14%	15.07%	5.78%	12.4%	**27.7%**	10.74%	14.1%

Table 8.2: Clustering results with all the features

We can observe in the table 8.2 that in comparison to the table 8.1 we get slightly higher percentages of each brand for its respective cluster. In the cluster Prada the brand holds now 17.7% in contrast to 13.1% of before. Hermes from 5.88% passes to 8.9%, Gucci passes to 20.6% from 18.6%. The percentage for Dior remains the same, while the percentages for Emporioarmani and Loewe decrease. We can say that in general this increase is due to the addition of more features. These new features slightly help distinguish one brand from another and thus increasing their percentages in their respective cluster. Nevertheless, as before the features are not discriminatory enough to be able to have a clear segregation between luxury brands.

8.9 Clustering with 3 clusters

The main idea was to create separate clusters for each brand. In an ideal situation where every artistic director would use totally different distinguishable features each of the 7 clusters in the tables 8.1 and 8.2 would have been populated by a high percentage of a single brand. As we can see in the sections 8.7 and 8.8 this is not possible. For this reason we try the same Kmeans algorithm in 3 clusters. We once again can observe in the tables 8.3 the same results as previously. Each cluster is populated by similar percentages of each brand. This leads us once again to the conclusion that our features are not discriminatory enough.

Cluster/Brand	Prada	Loewe	Hermes	Gucci	Dior	Chanel	Emporioarmani
Cluster 1	13.52%	16.01%	8.54%	19.2%	**18.5%**	13.16%	11.3%
Cluster 2	14.48%	14.84%	7.42%	15.54%	**21.2%**	10.65%	14.13%
Cluster 3	14.84%	15.54%	9.54%	**16.9%**	15.9%	14.1%	13%

Table 8.3: Clustering in 3 clusters

8.10 Clustering with 5 clusters

By trying the same Kmeans algorithm in 5 clusters 8.4 we obtain the same conclusion as in section 8.9

Cluster/Brand	Prada	Loewe	Hermes	Gucci	Dior	Chanel	Emporioarmani
Cluster 1	16.86%	13.95%	7.55%	14.5%	**17.44%**	14.53%	15.11%
Cluster 2	15.38%	15.3%	5.91%	15.38%	**24.85%**	10.65%	12.42%
Cluster 3	14.37%	14.37%	8.9%	**20.9%**	16.76%	11.37%	13.17%
Cluster 4	12.3%	16.4%	11.17%	**17.6%**	14.7%	12.35%	15.29%
Cluster 5	12.42%	17.15%	8.87%	17.75%	**18.93%**	17.15%	7.62%

Table 8.4: Clustering in 5 clusters

8.11 Principal component analysis

During this study the method of principal component analysis was considered. Principal component analysis (PCA) is used to perform dimensionality reduction. The goal is to reduce the dimension of the data from a large number of variables to a smaller set of variables that still contains most of the information. After performing the standardization of the data, in order for all the data to have the same significance, we proceeded to calculating the covariance matrix. Finally, we compute the eigenvectors and eigenvalues from the covariance matrix. In our specific case the different dimensions are the feautures we extracted from the data (gender, age, color etc). We noticed that our eigen values were similar. We came to the conclusion that none of our features are discriminatory enough. For this reason and in order to have more precise results we did not exclude any features. This led us to not be able to have graphs that represent the clustering points since we have more than 3 dimensions. We believe that this is a justified trade-off since it won't introduce more inaccuracies to our results.

8.12 Interpretation

From all the clustering results shown in tables 8.1,8.2,8.3,8.4 we get the same conclusion, the features selected for this study aren't distinguishable enough to let us segregate the selected luxury brands. We remained into the use of Kmeans algorithm for 2 reasons. First, instead of changing the clustering algorithm to SVM, KNN etc we preferred to increased the number of features as it is visible in section 8.8. Second, Kmeans algorithm is the baseline and a wide used and well respected algorithm. Taken under account the results from the PCA analysis 8.11 we believe that the change of clustering algorithm would not have led to different results. These results can be verified also by looking at the individual conclusions regarding each feature separately. As we saw in previous sections luxury brands use almost identical age, gender, colors, objects etc. It is then normal to be hard to distinguish one brand from another.

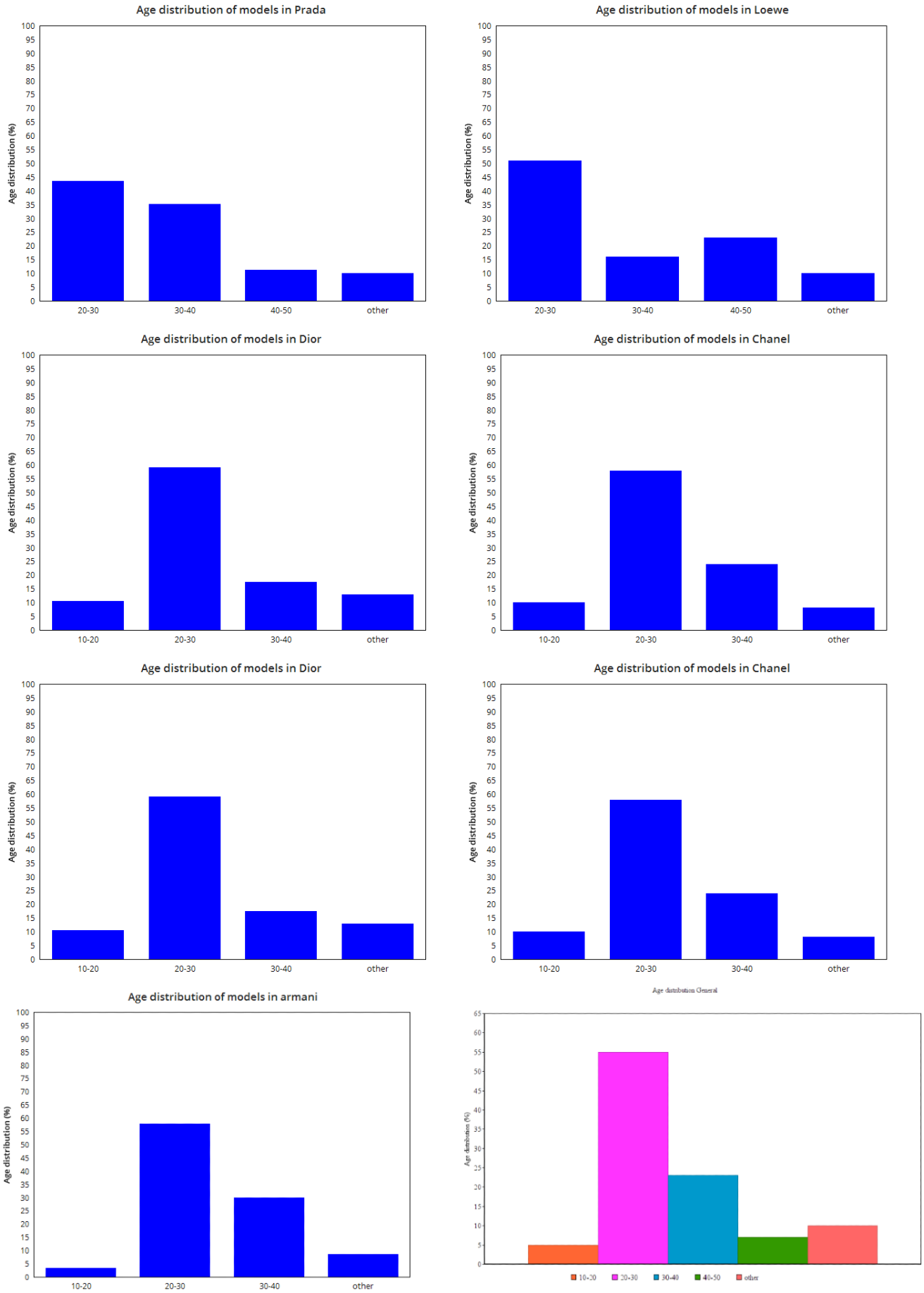


Figure 8.3: Age distribution

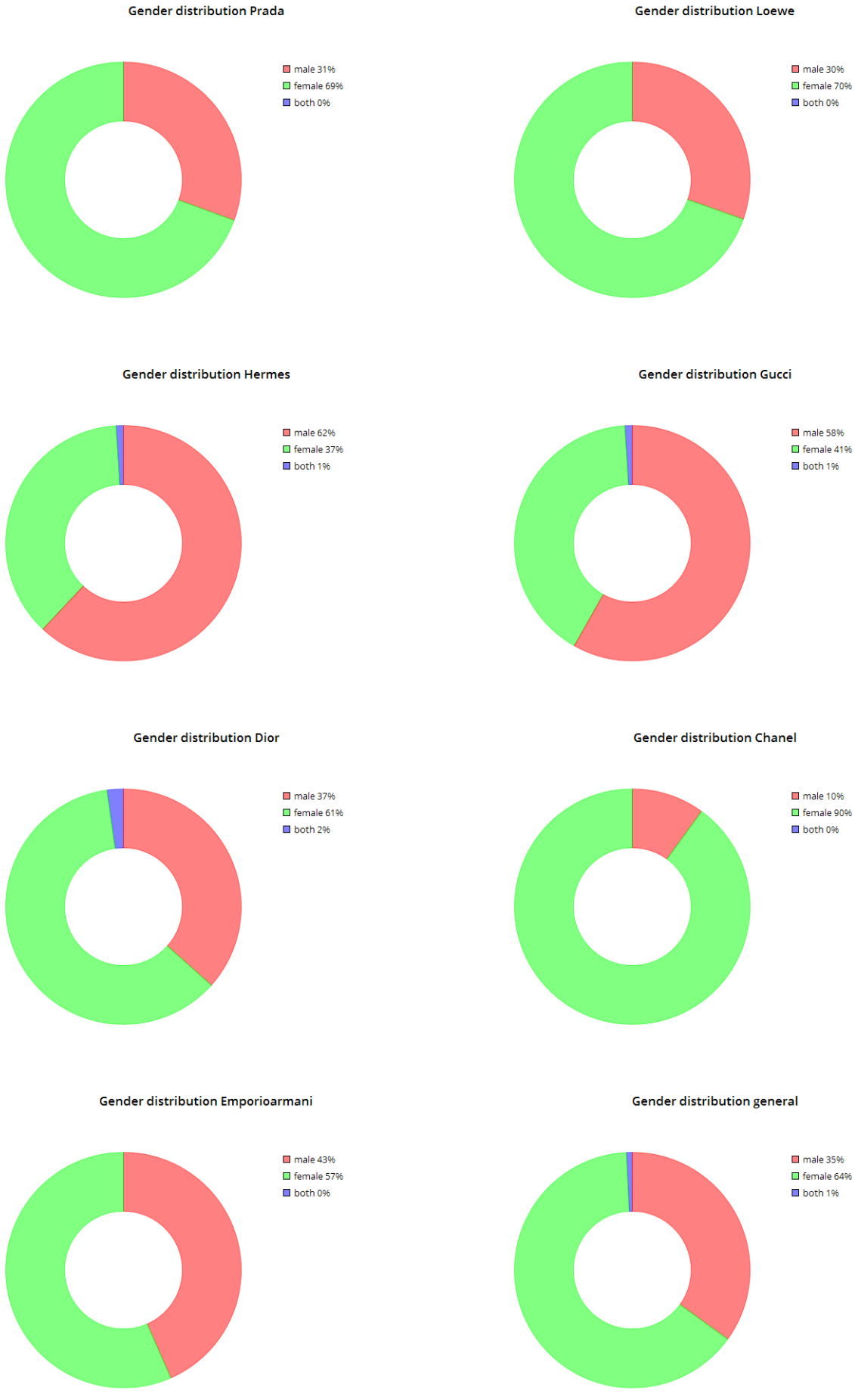
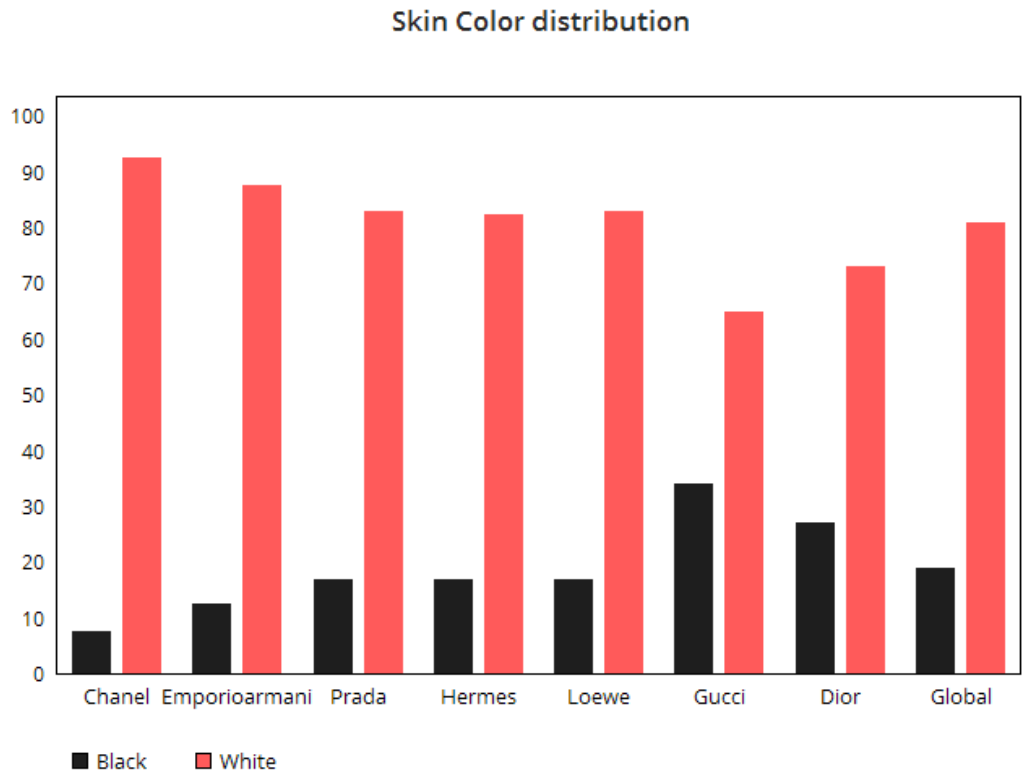


Figure 8.4: Gender distribution



H

Figure 8.5: Percentage of dark or light face color

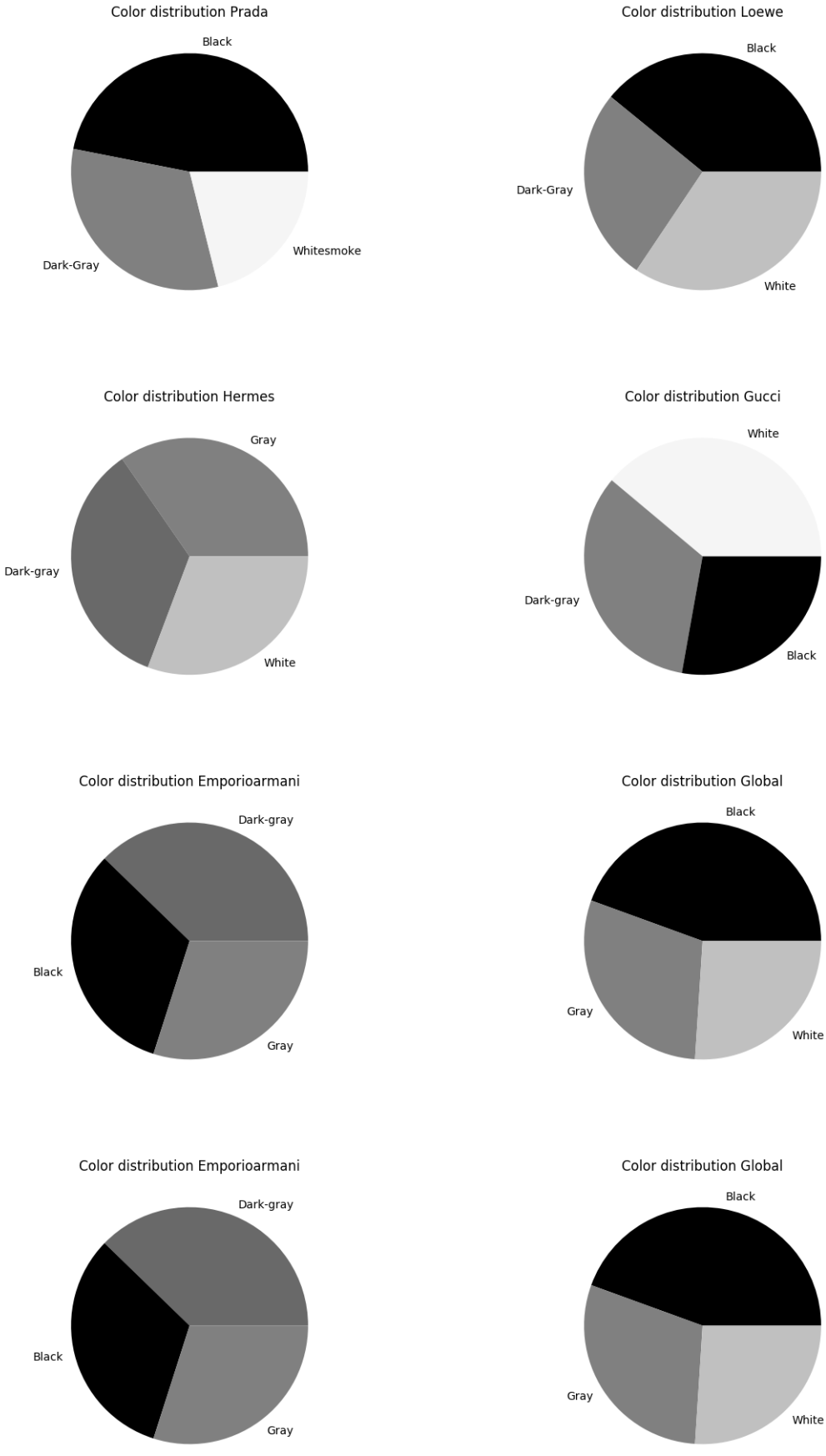


Figure 8.6: Color distribution

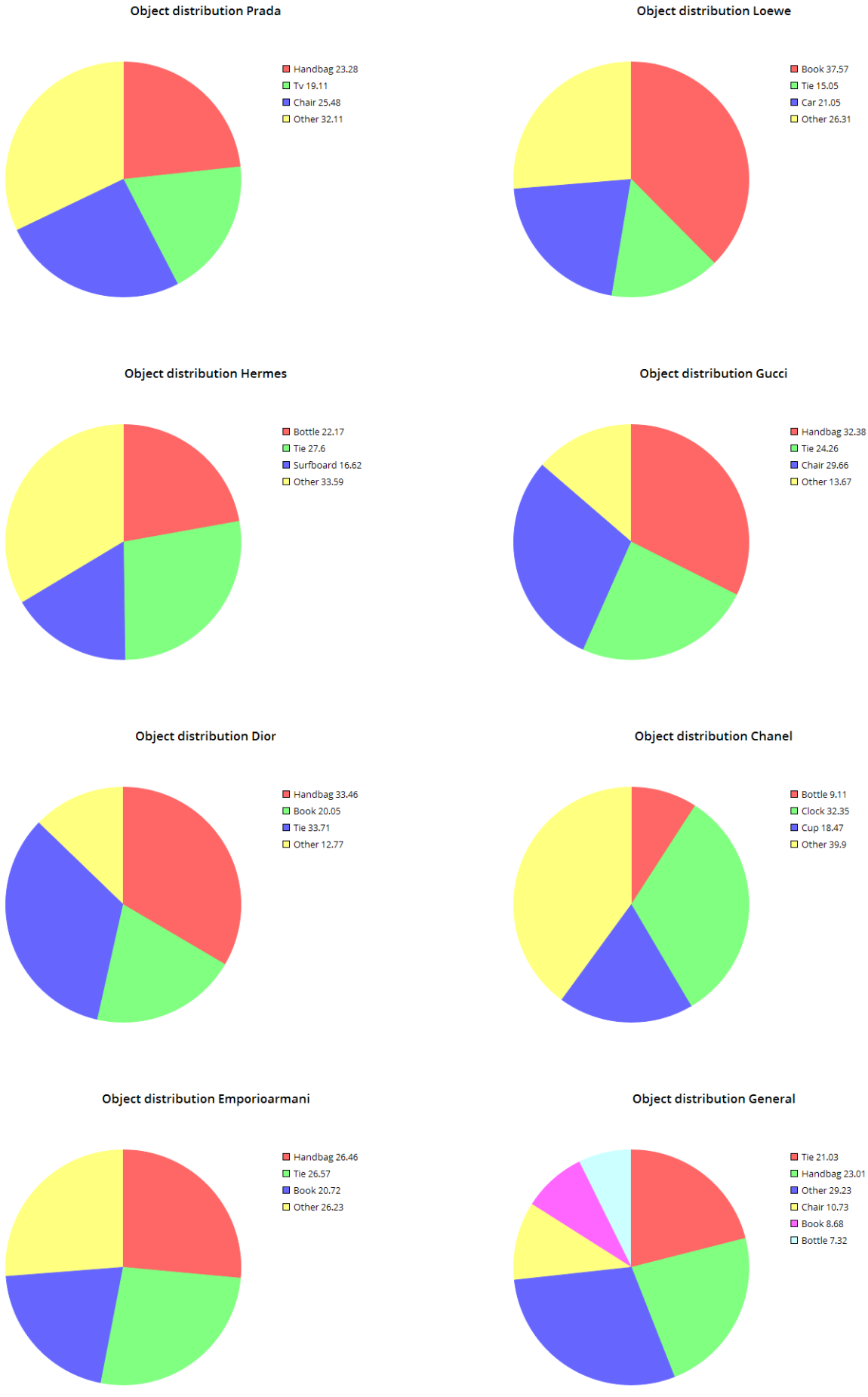


Figure 8.7: Object distribution

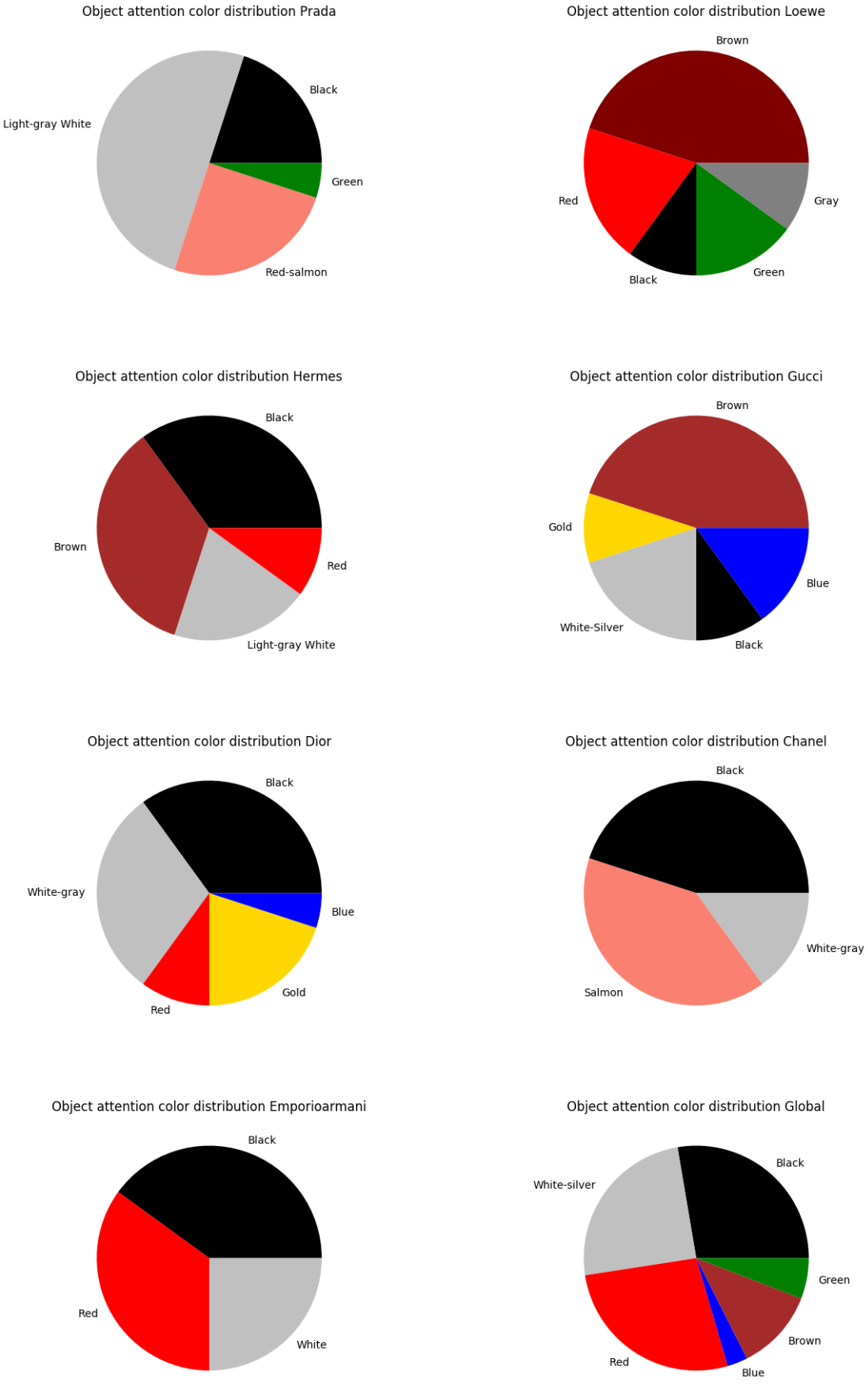


Figure 8.8: Object color distribution

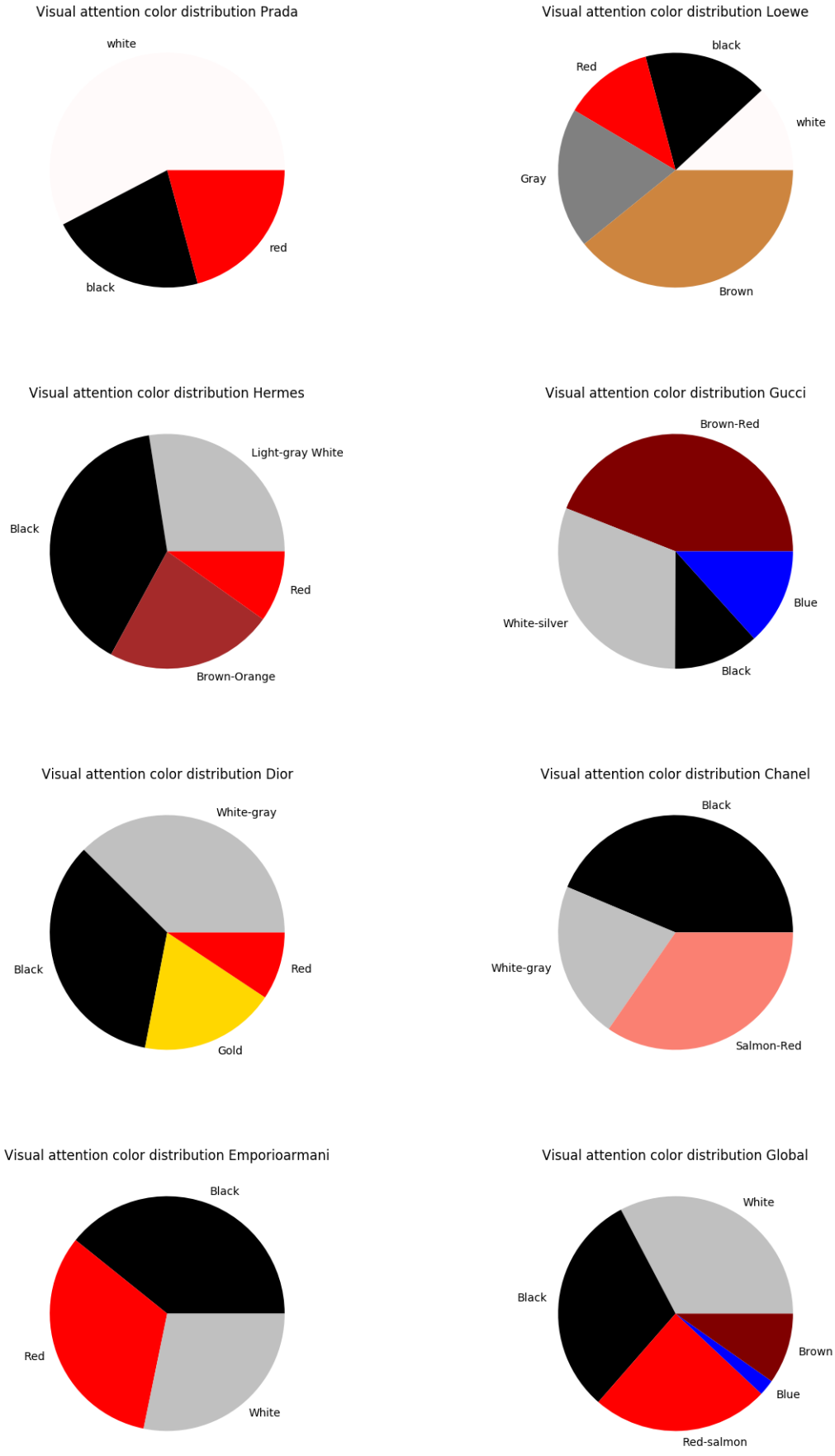


Figure 8.9: Visual attention color distribution

Chapter 9

Conclusion

This study is a first analysis of luxury brands appearing on the social media platform Instagram. The selected luxury brands are Prada, Loewe, Hermes, Gucci, Dior, Chanel, Emporioarmani. After creating a MongoDB data-base with the images extracted using the instaloader API we performed analysis on the data (photos) and extracted features. These features include age and gender prediction, skin color, most used color in the picture, object detection, object color and visual attention of the image. The emotion detection and also image caption and optical character recognition were tried but due to low accuracy weren't taken into account for the conclusions. For every feature we observed that brands follow similar techniques. Most of them use the same aged models, with the same skin color and even the same objects/object color. Of course there are some outliers such us that Chanel that uses 90% female models or that Dior uses a lot of gold color in their photos. These outliers make so that a certain product painted gold is more appealing to the eye and to make the advertised product stand up. These outliers hold a very small percentage in comparison to the general tendencies. As result we cannot distinguish one brand from another. We can see in the tables 8.1 and 8.2 that since the features extracted from our data are not distinguishable every luxury brand cluster is populated with similar percentages of the totality of brands. This seems logical since brands compete with each other. Luxury brands employ artistic directors that are professionals in the marketing sector. Apart from the individual creativity of each artistic director most of them follow the same methods since they have similar knowledge and skills. This is one of the reasons why these brands have similar features. Another reason is that since these brands compete for the same market it is normal to follow similar strategies and tendencies with outliers to differ slightly from each other. This seems to be a safe marketing method for established brands. In conclusion, we observe that artistic directors in luxury brands seem to follow similar techniques for the posted photographs on Instagram.

Chapter 10

Future Work

10.1 Metadata analysis and prediction

During the data harvesting using Instaloader, apart from the photos which is the main data type of this study, we also get the metadata related to every image. The metadata we get include geo-location, time of posting, likes, comments, hashtags and post description. It can be imagined that a prediction system can be created using the metadata in question. We can create a prediction network that will map the features extracted from the photos and the user involvement (likes, comments, sharing). In this way, we can have an insight on how the users will interact with an Instagram post before even creating it! The literature provides us with many already made prediction systems, that even though applied to different sectors, they can be re-trained for our purpose [21]. By doing so, we can obtain the following hypothetical results. Let's imagine that we create an Instagram post containing a female character and a dark gray background. By looking at the features extracted in this study, we see that most luxury brands perform a similar concept for their photos. We can then look at the metadata and see how did the users interacted with similar posts. If users preferred and liked this type of post more, the prediction system needs to be able to predict that using this concept for our photo, we will get a certain amount of likes and a certain amount of comments (positive/negative). This can be a very important tool for artistic directors or brand managers. By using a prediction system they can avoid negative interactions and negative comments from their clients. Also posts that might have negative impact to the brand name can be avoided by seeing their influence before making them publicly visible. This last point is extremely important because as demonstrated by various papers, brand image and brand name are very closely related to brand equity [9].

10.2 Twitter API and text analysis

At the time of writing, there are several social media platforms. Even though their main concept and functioning is quite similar, different social media platforms specialize in different forms of communication and expression by their users. For example while Instagram users express themselves by uploading images and videos, twitter users express their opinions by text (even though they can also post images). We can then, imagine performing the same kind of analysis we did in this paper but instead of applying it to images we can apply it to text. Different systems can be foreseen, emotion text analysis [1], length of text analysis, type of language used (formal or informal), language used etc. This analysis can help us increase the accuracy of our conclusions about luxury brands and have a deeper understanding on the way they approach their customers. Twitter provides access to their official API and we have the possibility to create stats about the likes, comments and general involvement of users with luxury brands. At least for the time being, Instagram is the number one platform for luxury brands and advertisement. It is then interesting to see if luxury brands try to evolve their marketing plan by expanding to more platforms such as twitter.

10.3 Facial expression detection

Up to the moment of doing this study, most of the facial expression networks available focus on front facing images. This might be interesting for using facial recognition for security purposes (accessing buildings, validating personnel etc) but poses a problem for our data-set. As explained before, the majority of the images on our data-set are not front facing. On top of that, part of the face of the models is covered either by a hat or the use of heavy make up. These facts make so that the facial expression networks available are not capable of providing accurate results. We can imagine an improvement of this study on the future, with the creation of our own facial recognition network that won't solely focus on front facing images. This task will give us interesting results since it is important to know the emotions and expressions used by models during marketing campaign. In order to complete this task a large data-set is needed and manual annotation of the expression of each image. This is why a network like this isn't created on the scope of this study and it can be an improvement point for the future.

10.4 Object detection

Object detection is still a very complicated task to perform. The variety of objects makes so that for now, object detection algorithms aren't capable of producing a very high accuracy. This is visible in our results. We get an accuracy of 55.24%. Taking under account that for the moment YOLOV4 is the state of the art for object detection networks we cannot hope for an immediate improvement if we use another network. The only imaginable solution is the creation of an object detection network appropriate for luxury brands. We can create from zero or re-train a YOLO network with data specific to our use case. For this we need to gather a large amount of data and manually label them. After this we need to re-train are re-evaluate the network. We hope that this approach will give better results for the use case presented in this study. Taking under account the time consuming task of gather data and re-training a network this improvement is not in the scope of this project.

10.5 Annotations

Many of the networks presented during this study cannot be evaluated by standard metrics. Image captioning, visual attention networks, etc produce results that are subjective. A good caption or an important area of the picture for a person can be totally wrong and unimportant for someone else. For this reason, to evaluate the systems with a better precision, the annotations made for the images we use to test each system should be annotated by multiple people. We need a minimum of 2 people making the same annotations. After they are done annotating the images they should cross examine both of their results to see the percentage of annotations that they agree on. By implementing this method we can obtain a more precise base value that we can evaluate the networks we wish to execute.

10.6 Low accuracy systems

As explained in previous sections not all the systems that we tried were taken into account for the final analysis. The reason is that three of the systems implemented during this study, Optical character recognition, Facial expression detection and image captioning performed poorly in our data-set. Since we deemed that these features might be interesting to analyze, it can be a future point for improvement to built machine learning networks capable of producing reliable results. Taken under account the time restrictions of this study and the complicated task of building and training machine learning networks this improvement is not in the scope of this study.

Chapter 11

Annexes

Figure 11.1: Detailed architecture of the age and gender network

```
{0: u'__background__',
                          1: u'person',
                          2: u'bicycle',
                          3: u'car',
                          4: u'motorcycle',
                          5: u'airplane',
                          6: u'bus',
                          7: u'train',
                          8: u'truck',
                          9: u'boat',
                          10: u'traffic light',
                          11: u'fire hydrant',
                          12: u'stop sign',
                          13: u'parking meter',
                          14: u'bench',
                          15: u'bird',
                          16: u'cat',
                          17: u'dog',
                          18: u'horse',
                          19: u'sheep',
                          20: u'cow',
                          21: u'elephant',
                          22: u'bear',
                          23: u'zebra',
                          24: u'giraffe',
                          25: u'backpack',
                          26: u'umbrella',
                          27: u'handbag',
                          28: u'tie',
                          29: u'suitcase',
                          30: u'frisbee',
                          31: u'skis',
                          32: u'snowboard',
                          33: u'sports ball',
                          34: u'kite',
                          35: u'baseball bat',
                          36: u'baseball glove',
                          37: u'skateboard',
                          38: u'surfboard',
                          39: u'tennis racket',
                          40: u'bottle',
                          41: u'wine glass',
                          42: u'cup',
                          43: u'fork',
                          44: u'knife',
                          45: u'spoon',
                          46: u'bowl',
                          47: u'banana',
```

https://gist.github.com/AruniRC/7b3dadd004da04c80198557db5da4bda

Figure 11.2: Classes that can be predicted

```
48: u'apple',
49: u'sandwich',
50: u'orange',
51: u'broccoli',
52: u'carrot',
53: u'hot dog',
54: u'pizza',
55: u'donut',
56: u'cake',
57: u'chair',
58: u'couch',
59: u'potted plant',
60: u'bed',
61: u'dining table',
62: u'toilet',
63: u'tv',
64: u'laptop',
65: u'mouse',
66: u'remote',
67: u'keyboard',
68: u'cell phone',
69: u'microwave',
70: u'oven',
71: u'toaster',
72: u'sink',
73: u'refrigerator',
74: u'book',
75: u'clock',
76: u'vase',
77: u'scissors',
78: u'teddy bear',
79: u'hair drier',
80: u'toothbrush'}
```

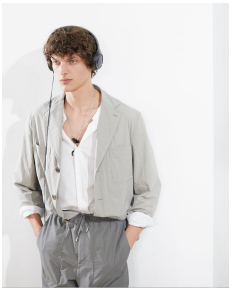

Figure 11.3: Classes that can be predicted

(a) Chanel photo example

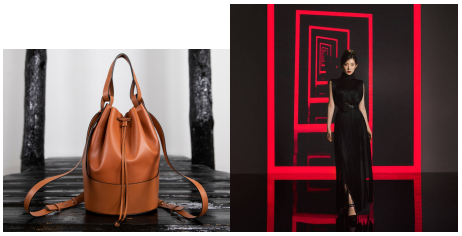
(b) Dior photo example

(c) Emporioarmani photo example

(d) Gucci photo example



(e) Hermes photo example



(f) Loewe photo example

(g) Prada photo example

Figure 11.4: Example images

Bibliography

[1] Saima Aman and Stan Szpakowicz. "Identifying Expressions of Emotion in Text". In: *Text, Speech and Dialogue*. Ed. by Václav Matoušek and Pavel Mautner. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 196–205. ISBN: 978-3-540-74628-7.

[2] M. F. Aydogdu, V. Celik, and M. F. Demirci. "Comparison of Three Different CNN Architectures for Age Classification". In: *2017 IEEE 11th International Conference on Semantic Computing (ICSC)*. 2017, pp. 372–377. DOI: 10.1109/ICSC.2017.61.

[3] Abu Bashar, Irshad Ahmad, and Mohammad Wasiq. "EFFECTIVENESS OF SOCIAL MEDIA AS A MARKETING TOOL: AN EMPIRICAL STUDY". In: *International Journal of Marketing, Financial Services Management Research* 1 (Dec. 2012).

[4] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-yuan Liao. *YOLOv4: Optimal Speed and Accuracy of Object Detection*. Apr. 2020.

[5] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. *YOLOv4: Optimal Speed and Accuracy of Object Detection*. 2020. arXiv: 2004.10934 [cs.CV].

[6] Dhani Chaubey, Sajjad Husain, and Ali Ghufran. "RELEVANCE OF SOCIAL MEDIA IN MARKETING AND ADVERTISING". In: *Splint International Journal of Professionals* 3 (July 2016), pp. 21–28.

[7] Ilke Cugu, Eren Sener, and Emre Akbas. "MicroExpNet: An Extremely Small and Fast Model For Expression Recognition From Face Images". In: *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE. 2019, pp. 1–6.

[8] David Dodd et al. "Face-Ism and Facial Expressions of Women in Magazine Photos". In: *The Psychological Record* 39 (July 1989), pp. 325–331. DOI: 10.1007/BF03395884.

[9] James B. Faircloth, Louis M. Capella, and Bruce L. Alford. "The Effect of Brand Attitude and Brand Image on Brand Equity". In: *Journal of Marketing Theory and Practice* 9.3 (2001), pp. 61–75. DOI: 10.1080/10696679.2001.11501897. eprint: https://doi.org/10.1080/10696679.2001.11501897. URL: https://doi.org/10.1080/10696679.2001.11501897.

[10] Furkan Gurpinar et al. "Kernel ELM and CNN Based Facial Age Estimation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. June 2016.

[11] C. Győrödi et al. "A comparative study: MongoDB vs. MySQL". In: *2015 13th International Conference on Engineering of Modern Electric Systems (EMES)*. 2015, pp. 1–6. DOI: 10.1109/EMES.2015.7158433.

[12] Yu-I Ha et al. *Fashion Conversation Data on Instagram*. 2017. arXiv: 1704.04137 [stat.ML].

[13] D. Keltner et al. "Facial expression of emotion". In: *Handbook of Affective Sciences. Series in Affective Science* 17 (Jan. 2003), pp. 415–432.

[14] Gil Levi and Tal Hassner. "Age and Gender Classification Using Convolutional Neural Networks". In: *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) workshops*. June 2015. URL: %5Curl%7Bhttps://osnathassner.github.io/talhassner/projects/cnn_agegender%7D.

[15] Mancas Matei, Phutphalla Kong, and Gosselin Bernard. *Visual Attention: Deep Rare Features*. May 2020.

[16] Maria mercanti-guérin and Christel Lassus. "DÉFENDRE LE POSITIONNEMENT LUXE D'UNE MARQUE GRÂCE À SON CHAMPION DE MARQUE : LE RÔLE CLÉ DU DIRECTEUR ARTISTIQUE SUR INSTAGRAM". In: Oct. 2019.

[17] Maria mercanti-guérin and Christel Lassus. "DÉFENDRE LE POSITIONNEMENT LUXE D'UNE MARQUE GRÂCE À SON CHAMPION DE MARQUE : LE RÔLE CLÉ DU DIRECTEUR ARTISTIQUE SUR INSTAGRAM". In: Oct. 2019.

[18] Jennifer Millard and Peter Grant. "The Stereotypes of Black and White Women in Fashion Magazine Photographs: The Pose of the Model and the Impression She Creates". In: *Sex Roles* 54 (Nov. 2006), pp. 659–673. DOI: 10.1007/s11199-006-9032-0.

[19] Olga Russakovsky et al. "ImageNet Large Scale Visual Recognition Challenge". In: *International Journal of Computer Vision* 115 (Sept. 2014). DOI: 10.1007/s11263-015-0816-y.

[20] Jyrki Suomala et al. "Neuromarketing: Understanding Customers' Subconscious Responses to Marketing". In: *Technology Innovation Management Review* 2 (Dec. 2012), pp. 12–21. ISSN: 1927-0321. DOI: http://doi.org/10.22215/timreview/634. URL: http://timreview.ca/article/634.

[21] Ming Wen et al. "Deep-Learning-Based Drug–Target Interaction Prediction". In: *Journal of Proteome Research* 16.4 (2017). PMID: 28264154, pp. 1401–1409. DOI: 10.1021/acs.jproteome.6b00618. eprint: https://doi.org/10.1021/acs.jproteome.6b00618. URL: https://doi.org/10.1021/acs.jproteome.6b00618.