UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft
Faculty for Computer Science, Electrical Engineering and Mathematics
Department of Computer Science
Research Group Responsible AI for Biometrics

Master's Thesis

Submitted to the Responsible AI for Biometrics Research Group
in Partial Fulfilment of the Requirements for the Degree of

Master of Science

Exploring the effect of dropouts on face image quality — assesment using stochastic embedding robustness (SER-FIQ)

by
Krupa Tharahunise Krishnappa

Thesis Supervisor:
Dr.-Ing. Philipp Terhörst

Paderborn, May 18, 2024

Erklärung

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen worden ist. Alle Ausführungen, die wörtlich oder sinngemäß übernommen worden sind, sind als solche gekennzeichnet.

Ort, Datum	Unterschrift

Abstract.
Biometrics has been an interest of research for long, where face recognition(FR) has been a major focus in today's authentication systems. Face image quality assessment (FIQA) has been immensely beneficial in enhancing the performance of the FR systems. Where stochastic embedding robustness (SER) has served as a baseline for many works to benchmark more innovative techniques. Despite the progressions made with stochastic embedding robustness face image quality (SER-FIQ) algorithm, there is a need for further exploration. Our research delves into the exploration of dropout regularization, a critical aspect that has not been explored much. In our exploration, we investigate the use of dropout regularization. Specifically, different configurations of dropout regularization was employed thoroughly throughout the experiments. We fill this gap of dropout exploration by employing two pretrained models, one with dropouts and another without dropouts. Each model serves as a foundation for two on-top models, where we train one with dropouts and another without dropouts.
Our comprehensive exploration of these models across various use cases shows that employing the dropout regularization in both the pretrained model and the on-top model probably gives optimal results observed in error vs reject curves. Our research highlights the superior performance of the pretrained model with dropouts. To elaborate on our investigation, we conducted experiments based on two assumptions with the SER-FIQ algorithm namely assigning the lowest quality scores to the least quality images and vice versa. The results reveals that an inherent ambiguity may exist between the two assumptions, which was observed due to an appearance of similar trends in the error vs reject curves in both the assumptions. To unravel this ambiguity and to address unexpected results, we analyze each use case, identify associated problems, and propose potential solutions for future work. This analysis is presented with coherence and brevity, offering insights into the challenges and avenues for improvement in the realm of FR and FIQA.

"In today's world, every scientific innovation leads to a brighter future. Every challenge in a research gets us one step closer to a world where technology becomes more capable of understanding a human face. Through this thesis, we promise to commit towards an iterative progress with every challenge we face for the betterment of the society, and to ensure that the journey of innovation continues to further unveil the new dimensions of human potential through technological development".

"I express my sincere heartfelt gratitude towards my supervisors who have supported me rigorously and unconditionally throughout my thesis !" Have fun reading the thesis !

Contents

1 Introduction

Biometrics play a crucial role in shaping the foundation for the modern day authentication and security systems. Biometrics also serves as a groundwork for understanding the intricacies of our experiments. By establishing a better understanding of the principles of the biometric systems in detail, we provide a structured way for the exploration of our experiments and the future innovations.

1.1 Biometrics

Biometrics is an human identification science [WJMM05] that involves certifying an individual's identity according to their physical, biological or behavioral qualities[JFR07], [RDR19], [BRA+09], [WJMM05]. The goal of biometrics is identity authentication[BRA+09], [JFR07] where biometrics is an automated authentication that is performed by a machine or a digital computer system[WJMM05]. Passwords, PIN [Wea06] codes and ID-card techniques were the conventional authentication methods used in olden days[BRA+09], [JFR07]. These methods are stable. But an individual may forget such data or they maybe lost or attacked by an external person easily[BRA+09], [JFR07].

Recent year's biometric modalities include: fingerprint, ear, face, hand geometry, iris, voice etc[JFR07], [BRA+09]. These modalities cannot be lost or forgotten, hence they provide more stability in a biometric system. Figure 1.1 illustrates various possible examples of biometric modalities that are used for identifying an individual in the recent years[JFR07], [BRA+09]. Out of all these biometric modalities the commonly used ones are: Face, Fingerprints and iris[JFR07]. Several different applications incorporate biometric systems in today's world where identity management is very important. In such applications basic services are typically based on the accurate determination of an individuals identity[JFR07], [BRA+09]. Primary objectives of an identity management system is to restrict the fraudsters from accessing the private data of the individuals, like for example in areas like in an airport security check, secure attendance systems, banking applications, confidential home security systems etc[JFR07], [BRA+09], [WJMM05],[RDR19] . Each of these applications use biometric authentication in several ways.

Basically, irrespective of any biometric modality, a biometric system works similarly by capturing a biometric trait of a person and then cross verifying it with the various biometric data that is already available in its database[JFR07], [RDR19]. This is a similar pattern observed in

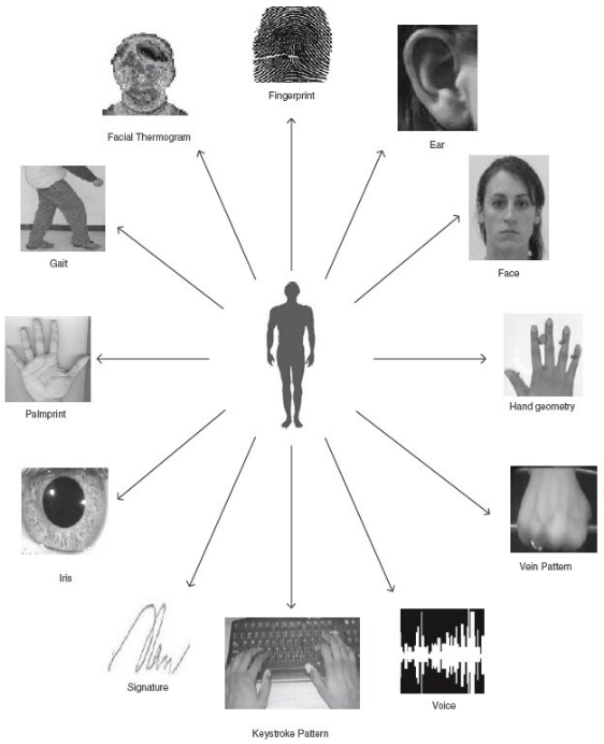

Figure 1.1: Biometric traits: (a.)physical includes iris, face, finger print and hand geometry , (b.) Behavioral includes signature, gait and key stroke dynamics[JFR07]

every biometric system out there in the world [JFR07], [BRA$^+$09]. Therefore it is important to develop efficient biometric systems with good quality data and at the same time monitoring the systems[BRA$^+$09]. According to A.K Jain et al. [JFR07], stages of a biometric system (operation modes) is viewed as an integration of three phases: an enrollment phase, a verification phase and an identification phase[JFR07]. In order to understand the operation modes of a biometric system, it is necessary to understand the basic modules of a biometric systems. Therefore the following section delves into explaining the basic modules of a biometric system, following which stages of a biometric system will be explained.

1.1.1 Basic modules of a biometric system

A biometric system is basically a pattern recognition system .i.e. it recognises patterns unique to each individual[BRA$^+$09],[JFR07]. Lets consider face recognition in the context of this thesis because the research in the present thesis is fully based on face recognition. A face recognition system receives face images sensed by a sensor(in an FR system), from an individual[JFR07]. As a consequence, it extracts important features from the input face image, these extracted features are consequently compared against the features stored in a reference database[JFR07]. After comparison, the intended action is performed which is based on the results of the comparison[JFR07]. Therefore a biometric system is viewed as an integration of three basic modules, namely a pre-processing module, template extraction module and a template matching module[JFR07]. Each of these modules are described below.

A **preprocessing module** P acquires a raw face image from an individual using a suitable biometric reader or a scanner[JFR07]. This module processes the acquired image such that the important facial features can be extracted[JFR07]. The aforementioned module has three sub

modules namely "face detector and preprocessing module", "face quality assessment module" and "the presentation attack detection module"[JFR07]. Firstly the **face detector module** detects the face in the obtained raw image[JFR07]. To make sure that all the faces are of same alignment and size, the detected face image is scaled, cropped, rotated and translated[JFR07]. Secondly, **face quality assessment(FIQA) module** estimates the face image quality (FIQ) using a Face image quality assessment algorithm(FIQAA)[SRH+22],[JFR07]. An FIQAA has a standard procedure to estimate the quality of a face image, so that only high quality face images are passed into the FR system[SRH+22]. A good quality estimation not only improves the performance of an FR system but also reduces the errors that can occur in the future[SRH+22],[JFR07]. In other words, when we reject lowest quality face images, the recognition system tends to perform better making less to no errors in recognising a genuine user or an imposter user[SRH+22]. Thirdly, the **presentation attack detection module** makes sure that the image captured is from a genuine user or an imposter user to prevent incorrect decisions by employing different kind of presentation attacks[JFR07].

A **template extraction module** T receives the preprocessed face image from the preprocessing module P[JFR07]. Then it extracts significant facial features from the detected face image and outputs a corresponding face template[JFR07], [PR04]. The obtained corresponding face template is created in such a way that the template is unique for each face that is detected[JFR07], [PR04]. The facial features extracted in this module are in such a way that it locates the position and orientation of landmarks such as eyes, nose and mouth, along with the distances between them[JFR07], [PR04], , [BRA+09]. An illustration of the extraction of facial features can be seen in figure 1.2. The shape structure, orientation of a face is also captured[JFR07], [PR04], [BRA+09]. All these facial features are converted into a template. The template extraction happens in various ways through Convolutional neural networks(CNNs), deep convolutional neural networks(DCNN's) etc[JFR07], [PR04], [BRA+09]. To use these networks machine learning models are trained and tested, to make sure the performance of the models are good enough[JFR07], [PR04], [BRA+09], [ZCPR03].

A **template matching module** M, operates in such a way that it takes in the template created from the template extraction module T, and compares it with the templates stored in the reference database of the biometric system[JFR07]. The comparison of both templates is done with a similarity function (e.g. cosine similarity-1.2 [SWY75]) resulting in a comparison score s[JFR07].

$$\text{Cosine Similarity}(\mathbf{A}, \mathbf{B}) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \cdot \|\mathbf{B}\|} \tag{1.1}$$

$$= \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}} \tag{1.2}$$

Where A and B are two vectors[SWY75], [JFR07], [JL11]. Where cosine similarity calculates how similar the given two vectors A and B are[SWY75], [JFR07], [JL11]. Applying a threshold t on the comparison score s results in a genuine or imposter decision[JFR07], [JL11].

$$\text{Decision} = \begin{cases} \text{Genuine} & \text{if Cosine Similarity} \geq \text{Threshold t} \\ \text{imposter} & \text{if Cosine Similarity} < \text{Threshold t} \end{cases} \tag{1.3}$$

The following section explains how does the stages of a biometric system work.

Figure 1.2: Recognition of the features of a face with a bounding box and detecting the landmarks on the face like eyes.[JFR07], [BRA+09],[KS12]

1.1.2 Stages of a biometric system - Operation modes

An **enrollment phase** has a human machine interface where it is a biometric sensor that reads or scans a biological biometric trait[JFR07]. Assume we are dealing with a face recognition system. The human face gets enrolled through the biometric sensor, where a biographic data like a PIN code or a password is needed to enroll into a database[JFR07], [JL11], [ZCPR03]. The face is then preprocessed, where a face is detected, then quality of the image is assessed, then features are extracted from the face using a deep face recognition network, finally a unique template is extracted[MWHN18], [JL11], [JFR07]. The final extracted template is usually in the machine readable format[MWHN18], [JL11], [JFR07]. At last, the enrolled face image is passed into a database[JL11], [JFR07].

A **verification phase** is employed to confirm the identity a person claims to have. This unit verifies if the identity of a person is genuine or not[JL11], [JFR07]. For example: a face image of "Krupa" is enrolled in the enrollment unit. Here the verification unit checks if the enrolled person is Krupa or not. Hence the face is again preprocessed where the presence of a face is detected, the quality of the image is assessed, then features are extracted from the face using a deep face recognition network, finally a unique template is extracted[JL11], [JFR07], [MWHN18], [TYRW14]. The final extracted template is usually in the machine readable format[JL11], [JFR07]. The final extracted template is compared with the templates stored in the reference database[JL11], [JFR07]. In this comparison, a cosine similarity or a euclidean distance is calculated[JL11], [JFR07], [ZCPR03]. Usually there are many more parameters associated in this comparison, but a focus on cosine similarity is more relevant, since it is commonly used. If the images are more similar, .i.e. if the similarity score of the compared templates if very high, for example: assume we obtained a similarity score of 0.80. Then authentication is provided as a genuine match by the matching module[JL11], [JFR07], [ZCPR03]. Whereas if the similarity score is 0.06, then authentication is provided as an imposter match by the matching module[JL11], [JFR07], [ZCPR03], , where it verifies an enrolled user.

An **identification phase** assigns an identity to an unknown subject[JL11], [JFR07]. Even in this unit the image is preprocessed, face is detected and quality is assessed[JL11], [JFR07]. Then a template is extracted from the deep face recognition network to obtain a final face template[JL11], [JFR07]. The obtained face template is matched with various face templates stored in the reference database[JL11], [JFR07]. Cosine similarity between the input face template and various other face templates is computed[JL11], [JFR07]. Again here a face template that has higher similarity score is pointed out and the respective identity is given to the input face image[JL11], [JFR07]. In this case for example if the similarity score is somewhere around 0.82 .i.e. greater than 0.5 , is considered as a genuine match because the reference template in the database becomes more similar[JL11], [JFR07]. In case if the score comes out to be 0.24, it is claimed as an imposter match, where authentication is not provided by the matching module[JL11], [JFR07]. So the output of this module might have a genuine user or an imposter

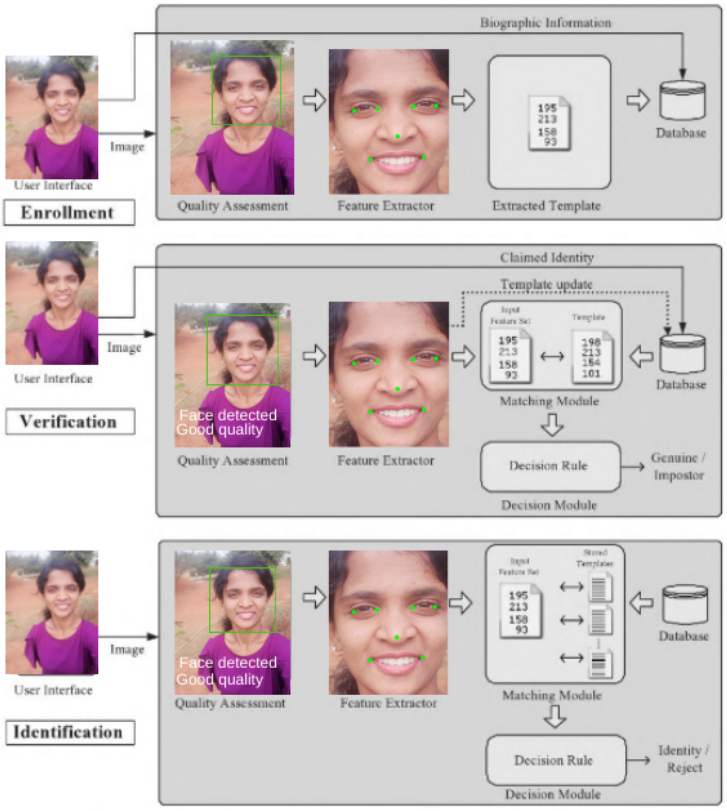

Figure 1.3: Operation modes of a Biometric system with an enrollment phase where the face image is enrolled. In the verification phase the face is verified if the user is genuine or not. In the identification phase, an identity is assigned to an unknown subject.[JFR07]

user[JL11], [JFR07], where it identifies and assigns an identity to the enrolled user. A detailed illustration of the operations of a biometric system is provided in figure 1.3.

Overall, an enrollment phase enrolls an individual, a verification phase verifies an individual and an identification phase identifies an individual. In order to understand the biometric systems in depth, it becomes crucial to understand the characteristics of a biometric system. Hence the following section is all about the characteristics of a biometric system.

1.1.3 Characteristics of a biometric system

Different biometric modalities used, have their own unique characteristics that play an important role in casting the biometric systems functionality.[JFR07], [JL11], [BRA+09], [Kom04]. Since the present thesis is primarily based on face recognition, lets focus on facial characteristics in the context of face recognition.

Attributing to the characteristics of a biometric system in face recognition, there exists various inherent characteristics [vdHvGP13]. First being, the concept of **uniqueness**. There are a crores of individuals in this world, even then each individual possess a unique face with distinctive facial characteristics[JFR07], [vdHvGP13]. These distinctive facial characteristics are the key to building a face recognition system that has a characteristic of uniqueness[JFR07], [vdHvGP13]. Second being **Universality**, is another universal trait that must be common among all the individuals in this world because a biometric system has to be designed in such a way that it covers the largest possible ratio of the population[JFR07], [vdHvGP13]. In the context of FR

systems, facial features are visually universal[JFR07], [vdHvGP13]. A human face is a nearly a unique trait in every individual out there in the world, thus making face recognition systems applicable to a wide range of population irrespective of the age, gender or ethnicity[JFR07], [vdHvGP13]. Related problems in universality may include the presence, absence or a degraded quality of a certain biometric trait in each individual[JFR07], [vdHvGP13].

Third being **Permanence**, is yet another important characteristic for the solidity of biometric systems. This ensures that the trait chosen is proportionately stable and consistent over time[JFR07], [vdHvGP13]. In face recognition systems, the continuity of facial features all through the life of a human being, enables the system to be accurate over extended time periods[JFR07], [vdHvGP13]. One of the major challenges here is the fact that as people age, the appearance of the face changes[JFR07], [vdHvGP13]. Fourth being **Performance** is a central characteristic particularly in case of face recognition systems. An FR system aims to maximize its performance and minimize the computational work load[JFR07], [vdHvGP13]. In other words this means, the system's ability to acquire high accuracy, fast processing and minimal false acceptance and false rejection rates is of cardinal importance[vdHvGP13]. Achieving such a level of performance should be a constant focus of research and development efforts, as it directly impacts the real-world applicability of facial recognition systems.

Fifth being **collectability** is the simplicity with which facial data can be measured or captured[JFR07], [vdHvGP13]. Such a characteristic plays a vital role in employing biometric systems. In the context of FR, it becomes easier to be more user friendly, thus simplifying the face enrollment process with an excellent user experience[JFR07], [vdHvGP13]. Fifth being **Acceptability** is a regard that becomes extremely crucial especially in facial recognition. Social and cultural factors should be undoubtedly considered[JFR07], [vdHvGP13]. Addressing privacy concerns and individual objections is crucial to sustain trust among users and communities[JFR07], [vdHvGP13]. In other words it means that, the convenience of the users is taken into account and the usability among the population is maximized[JFR07], [vdHvGP13]. Lastly **circumvention** is one of the critical characteristic that is very important in biometric systems. A biometric system is designed in such a way that collecting and replicating a fake biometric sample/template becomes very hard[JFR07], [vdHvGP13]. Especially in the context of FR systems, the facial features like length of the eyes, mouth, nose, face ,distance between each feature, shape of the face, dimensions of the face and many other features are considered to make it hard to replicate a face template[JFR07], [vdHvGP13]. All the above explained characteristics are measured in the scales of high, low or medium[JFR07], [vdHvGP13].

In addition to the basics of biometric systems, it is also necessary to understand the evaluation prospects of the biometric systems to comprehend the present thesis. Hence the following section follows with the evaluation prospects of biometric systems.

1.1.4 Evaluation of biometric systems

In the field of biometrics, assessing the strengths, performance and drawbacks is of vital importance. Evaluation becomes a foundation of this endeavour, where we dig deep into complex errors that can manifest at different stages of a biometric system operation[JFR07]. After digging deep into the errors, we use different types of evaluation curves to evaluate the biometric system[JFR07]. Errors in biometric systems can be categorised into four different categories each demanding its own inspection, namely **errors in the acquiring process**, **errors during verification** and **errors during identification**[JFR07]. These errors are then plotted on curves like, Receiver Operating Characteristic (ROC) Curve, Cumulative Match Characteristic (CMC) Curve etc[JFR07]. A detailed exploration of these error categories and respective important

performance curves, which are necessary to comprehend the present thesis are presented below.

Errors during verification

The process of biometric verification particularly in the scenarios of a face recognition system is a perfect verification between two face templates that is necessary to validate the claimed identity[JFR07]. A perfect verification in a biometric system is not possible because of several uncertainties that can occur[JFR07]. Firstly imperfect sensing conditions, where a biometric sensor might be very old and thus not capable of sensing a face properly[JFR07]. Various capturing devices and technologies are present to capture a face, therefore we cannot predict which uncertainty can occur at which point of time or space[JFR07]. Secondly, alterations in the individuals biometric characteristic. When an individual ages , the appearance and looks of his/her face changes, hence making it difficult for an FR system to do a perfect match[JFR07].

Thirdly, changes in the ambient conditions, such as inconsistent illumination levels. In other words when you consider a face image of a specific person, the lightning conditions of the image may vary from situation to situation, which indirectly affects verification[JFR07]. Lastly, variations in user-sensor interaction, such as different head poses and different face expressions[JFR07]. Different poses, different expressions or different angles of capturing a face image, indirectly affects verification[JFR07]. We can also imagine the presence of face occlusions, that may lead to imperfect sensing of face images, thus generating different face templates, which indirectly also affects face verification in a biometric system[JFR07].

When evaluating biometric systems during verification, there are some ISO standards that are followed in the standard research field of biometrics[JFR07]. These are **False non-match rate (FNMR)** and **False match rate (FMR)**[JFR07]. Where FNMR is a proportion of genuine attempt samples that are falsely declared not to match the template of the same characteristic from the same user supplying the sample[JFR07]. In other words, FNMR is an error rate that indicates how frequently a system fails to recognise a genuine user, thus leading to falsely rejecting a genuine user[JFR07]. For example, consider a company staff face recognition system. When a company employee comes in for login, the biometric system does not log in a genuine company employee.

FMR is a proportion of zero-effort imposter attempt samples that are falsely declared to match the compared non-self template[JFR07]. In other words, FMR is an error rate that indicates how frequently a system incorrectly accepts a user as a genuine user, when they are an imposter user[JFR07]. For example, an external fraudster tries to login to a company premises and the biometric system accepts him/her as a genuine company employee. It is evident that for security reasons, its important to strike a balance between FNMR and FMR. The tradeoff between FNMR and FMR can be adjusted by setting a decision threshold for a biometric system[JFR07]. Figure 1.4, shows a probability distribution curve, with FNMR and FMR areas marked, with a threshold t that strikes a balance between them. The probability distribution equations for FNMR and FMR are provided in equations 1.4 and 1.5, which are obtained by integrating over the genuine and imposter distributions respectively[JFR07]. Where genuine and imposter distributions are in turn computed using the mean and standard deviations of genuine and imposter scores, as shown in below equations[JFR07]. FNMR and FMR are algorithmic-level evaluation errors of "verification" performance[JFR07].

$$FNMR = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi\sigma^2_{\text{genuine}}}} \cdot \exp\left(-\frac{(x-\mu_{\text{genuine}})^2}{2\sigma^2_{\text{genuine}}}\right) dx \tag{1.4}$$

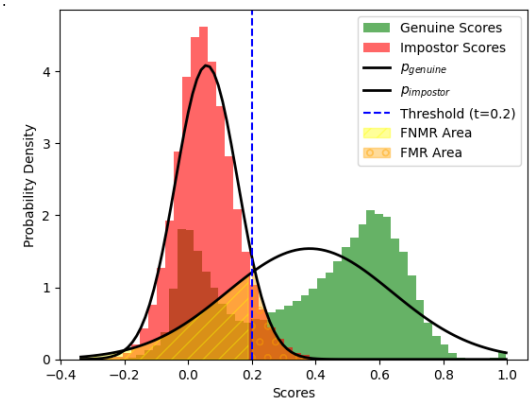

Figure 1.4: Distribution curve for genuine and imposter scores with threshold t to adjust the tradeoff.[JFR07]

Where:

μ_{genuine} : Mean of the genuine scores.
σ_{genuine} : Standard deviation of the genuine scores.

$$FMR = \int_t^{\infty} \frac{1}{\sqrt{2\pi\sigma^2_{\text{imposter}}}} \cdot \exp\left(-\frac{(x-\mu_{\text{imposter}})^2}{2\sigma^2_{\text{imposter}}}\right) dx \tag{1.5}$$

Where:

μ_{imposter} : Mean of the imposter scores.
σ_{imposter} : Standard deviation of the imposter scores.

However in simple terms FNMR and FMR are given by equations 1.6 and 1.7.

$$FNMR = \frac{\text{Number of False Non-Match Instances}}{\text{Number of Genuine Instances}} \tag{1.6}$$

$$FMR = \frac{\text{Number of False Match Instances}}{\text{Number of imposter Instances}} \tag{1.7}$$

Another ISO standard that is used in the evaluation of biometric systems is a **receiver operating curve (ROC) curve**[JFR07]. This curve is used to report the trade offs between FNMR and FMR values at different or every threshold value[JFR07]. Such a curve is extremely beneficial for comparison in almost every application area. We need the ROC curves to show a wide range of possible operating points[JFR07]. A visual representation of an ROC curve is illustrated in 1.5.

When **FMRs are relatively high**, they are useful in convenience focused applications, where usability is the primary goal[JFR07]. For example in personal electronic gadgets[JFR07]. The range of FMRs in such gadgets are about $FMR < 10^{(} - 2)$ or $FMR < 0.01$[JFR07]. This means that the gadgets accept a small number of imposter attempts (e.g., 1% or less) for the sake of user convenience[JFR07]. Whereas, when **FMRs are relatively low**, they are useful

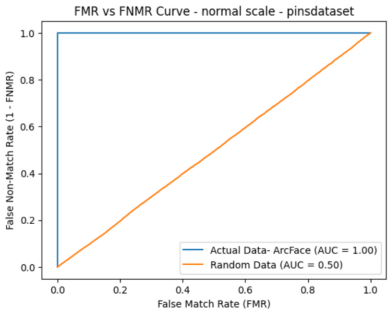

Figure 1.5: Receiver operating curve distribution.[JFR07]

in security focused applications[JFR07]. Security is the primary goal[JFR07]. For example in national border control areas, the range of FMRs in such areas are about $FMR < 10^{(} - 3)$ or $FMR < 0.001$, to minimize the risk of imposter access and prioritize high security[JFR07]. This means that the security areas accept really small or almost no imposter attempts for the sake of security of the native country[JFR07].

In most of the biometric systems it is evident that for the verification FNMR @ 10^{-x} FMR or FRR @ 10^{-x} FAR are maintained as per the specifc application[JFR07]. Where the value of x varies from one application to another[JFR07]. There is another popularly used error rate .i.e. **Equal Error Rate(EER)** which is a popular measurement in the evalution of verification in biometric systems[JFR07]. EER is well known as a single-value indicator of the verification performance and equals the FMR at the decision threshold t where FMR and FNMR are the same[JFR07]. In layman's terms, EER is a single easy to understand value point that summarizes the performance of the whole system[JFR07]. In easy words, an EER technically means, it is a point where the system makes an equal number of false matches and false non-matches[JFR07]. Hence EER, as a point serves as a summary of the system's performance, providing a simple and clear measure of accuracy[JFR07]. Figure 1.6 illustrates an ROC curve which plots FNMR vs FMR and gives the EER line and marks the EER point.

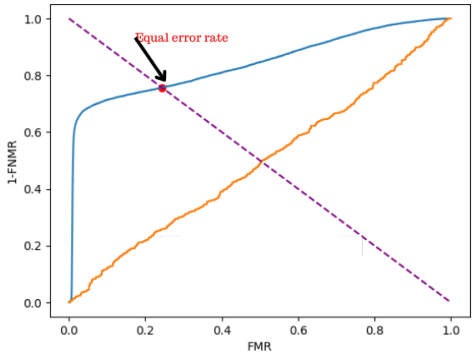

Figure 1.6: Receiver operating curve distribution with Equal Error Rate marked.

Overall, biometrics is a branch of science that is applicable almost in every aspect in this world ranging from border control to every security purpose. Hence performing efficient research in the field of biometrics becomes really important. Although biometrics has various traits, the primary focus of the present thesis will be specifically on face recognition. Where main focus will be on face image quality assessment (as explained in the enrollment unit). Hence the following section follows with face recognition.

2

Related work

Face recognition(FR) has been an interest of research for at least three decades now[AOBTA20]. This intriguing interest is due to the simplicity provided by the FR systems in almost every security authentication in this world[AOBTA20], [ZCPR03], [WYWL20]. Such a security authentication is used by every organisation. The ease of use of the face images in the FR systems makes them a choice of every organisation[JL11], [ZCPR03]. Due to a large variety of commercial and legal requests from various organisations from all around the world, FR systems have become quiet popular[AOBTA20]. Face images in such FR systems are the biometric modalities that are widely accepted because of their inherent characteristics[AOBTA20].

Firstly, face images are a very realistic and a natural feature that every human being possess[AOBTA20], [ZCPR03]. This makes them a very relevant feature for the purposes of identification and authentication[AOBTA20]. Secondly, they are non-intrusive, which means that the face images can be obtained without any physical contact[AOBTA20]. Individuals feel more relaxed when face images are used as the biometric identifiers, because a face recognition device collects data in a friendly manner[AOBTA20]. Thirdly, collection of the data requires less cooperation. Which means that the face images are collected without any assistance and without the active participation of the user[TKD$^+$20],[AOBTA20].

Today, face recognition has advanced to such a level that it can easily recognise a person who he/she claims to be[AOBTA20]. Such a great advancement dates back to 2011. Since 2011, the process of face recognition has accelerated[KSH],[MWHN18] due to deep learning[DCWZ16], which is a part of machine learning that is built open the artificial neural networks[HKW11], [Yeg09]. From 2011, there has been significant progress in overcoming the challenges in face recognition systems due to the incorporation of convolutional neural networks(CNNs), which are a part of deep learning. where these networks learn better when trained with millions of face images[WYWL20], [MWHN18],[OCP$^+$18]. These CNNs have played an important role in enhancing the accuracy and robustness of the FR systems, where they have been impressively instrumental in achieving great heights in the research of face recognition[WYWL20], [MWHN18],[OCP$^+$18]. Despite the growing interest of research in FR and its advantages, there exists a variety of challenges in recognising faces. According to various research works, some of these challenges are unconstrained scenarios like different head poses, different illuminations, different age groups, different facial expression related variations and different ethniciies[TKE20], [ZCPR03], [TKD$^+$20]. In each scenario, the facial features vary and their analysis becomes challenging with distinct factors[TKE20].

2.1 Face recognition

In order to overcome the current challenges, it is very important to know the recent advancements in the domain of face recognition. Very recently in 2024, a good step was taken by Kabir et. al for the aid of deaf community in Bangladesh(Bangla sign language - BdSL)[KMHS24]. Where the people with hearing impairments need automatic systems to detect their sign languages like **face expressions**, body movements, hand shapes[KMHS24]. **They used max voting ensemble technique to combine the state of the art deep neural networks like Xception, InceptionV3, DenseNet121, ResNet50, and MobileNetV2**[KMHS24].They experimented on two different datasets of the deaf people in Bangladesh(BdSL-49 and BdSL-38)[KMHS24]. Where each model gives a unique prediction and the class that is predicted by the maximum number of models is considered as a final prediction(this is called max voting)[KMHS24]. In case if two or more classes receive an equal number of votes, the class that has the highest confidence score is selected[KMHS24]. The final ensemble prediction becomes an aggregate of the majority vote across different models[KMHS24]. They proved that this method acquires a good accuracy of around 96%[KMHS24].

Yet another research work unveiled by Tan et. al [TXS24] proposed an **Informative and Discriminative Semantic Features Learning (IDSFL) network for the facial expression recognition(FER)**[TXS24]. This network consists of a multi channel feature(MCF) modulator with gabor filters to capture diverse information[TXS24]. They also added a specific emotion aware(SEA) module, to focus on each facial expression category[TXS24]. They experimented on datasets like RAF-DB, FERPlus and AffectNet-7 and proved a robust performance when compared to other FER methods[TXS24]. In another research by Lu et. al , they proposed a **long short-term perception network(LSTPNet) for the FER problem**[LJF$^+$24]. This network consists of a temporal channel excitation(attention modelling) and a long short term temporal transformer(LSTformer)[LJF$^+$24]. **This network was detailed with CNN units, attention modules and units of LSTMs(Long short term memory)[an example or recurrent neural networks] to capture diverse features in facial expressions**[LJF$^+$24]. While another research by Xiao et. al, **integrated CNNs and graph neural networks(GNNs) to solve the FER problem**[XWX23].

An important step was taken by Ramos et. al [RC24] **to recognise face images of individuals in uncontrolled scenarios**[RC24]. They combined different biometric recognition methods of the ear recognition and face recognition[RC24]. They leveraged a dataset with the fusion of the information of face and ear images called a VGGFace-Ear dataset[RC24]. With a model for the generalization of the ear images called VGGEar model[RC24]. They fused face recognition and ear recognition in a common recognition pipeline. **When they fused the information of the face images and ear images, it was proved that the recognition rates increase upto 82%**[RC24]. In a yet another important work by Huang et. al, they volunteered a **new dataset termed as Webface-OCC[HWW$^+$21]. This dataset was created by combining a variety of face occlusions, it consists of 804,704 face images of 10,575 different identities**[HWW$^+$21].

In the paper [ZL19] by Zhi et. al, **a face recognition model was introduced, which was based on support vector machines(SVMs), genetic algorithm and principal component analysis(PCA)**[ZL19]. Where the SVMs were used for classification, genetic algorithm to optimise the search strategy and PCA to reduce the feature dimensions of the face images [ZL19]. They also proved that their model reaches an accuracy of upto 99%[ZL19]. In yet another paper [WD20] by Wang et. al, **they addressed the problem of different ethnicities in face recognition[WD20]. They proposed a reinforcement learning**

based raced balance network[WD20]. This network leverages deep Q-learning to approximate the Q-function by guiding the agent to perform explicit tasks that are assigned to it[WD20].

In the research work [MMRY24] by Maharani et. al, they proposed a strategy to overcome the problems of face occlusions, blurred images and faces that turn away from the camera[MMRY24]. **They integrated CNN units with the LSTM units to optimise the process of face recognition[MMRY24]. Also to reduce the computational load and to track the human faces in uncontrolled scenarios, they incorporated Q-learning**[MMRY24]. Where Q-learning is a reinforcement algorithm for making optimal decisions[MMRY24]. Their system enhances tracking efficiency and accuracy, which solves major challenges in face recognition[MMRY24]. In yet another research by Yengikand et. al [YAN24], they integrated CNNs and attention mechanism to detect different emotions of a human face. In a similar way there are many other research works that have leveraged attention based transformer models to address the current challenges in FR [GDXZ18], [ZHSZ20], [VSP+17], [CYA20], [GLQZ24]. In this way, face recognition has been advancing explicitly since two decades by the integration of the interdisciplinary aspects right from the advancements in machine learning to the advancements in neural networks. It is crucial to address face image quality assessment(FIQA), since it acts as a baseline research for enhancing the performance of face recognition systems.

2.2 Face image quality assessment

Face image quality assessment(FIQA) is where we assign a value to a face image sample to decide its quality[TKD+20], [SRH+22]. The research in FIQA has been significant. Very recently in the paper [LDL24] by Liu et. al, they addressed the problems in the process of face frontalization[LDL24]. **Face frontalization without glasses based on perceptual quality and pixel-level quality assessment(FF-PPQA)**[LDL24]. Where face frontalization is a process of creating realistic frontal views of the face images captured in different poses[LDL24]. Especially in the side poses of face images wearing glasses, they remove the glasses in such images[LDL24]. After removing the glasses, they design a perceptual and pixel level FIQA modules, so that the performance of the process of face frontalization is improved[LDL24].

Prior to FF-PQA, in the paper [ZK24] by Zhang et. al introduced an FIQA strategy based on MTCNN AND FaceNet [ZK24]. **Firstly, they employed FIQA before face detection is done and only the image that reaches the threshold can be inserted into the model** [ZK24]. Secondly, an additional mechanism is introduced to deal with the problem of partial face occlusion[ZK24]. Both these methods enhances the performance of the FR model and iteratively adapts to the face occlusion problems [ZK24]. In yet another paper, by Tie Liu et. al, they introduced a **transformer based quality assessment method(TransFQA).** Where they introduced a face components guided network(FT-NET) using progressive attention mechanism. Another network .i.e. **distortion specific prediction network (DP-NET) was introduced for an accurate quality score prediction**.

In the research work[OCZ+21] by Zhao et. al, they proposed an FIQA method called **Similarity Distribution Distance for Face Image Quality Assessment (SDD-FIQA)**k[OCZ+21]. Where SDD-FIQA calculates the Wasserstein distance between the inter class and intra class similarity distributionsk[OCZ+21]. By using this calculated Wasserstein distance the method generates quality pseudo labelsk[OCZ+21]. With these labels, a regression network is trained for the quality predictions[OCZ+21]. In another paper [BFK+23] by Boutros et. al**(CR-FIQA), an FIQA method that learns the relative classifiability was proposed**[BFK+23]. This classifiability is based on a measure .i.e. based on how the feature repersentation of a face image

is represented in an angular space w.r.t its class center and the center of the nearest negative class[BFK+23].

In the paper [CY22] by Chen et. al, an FIQA method .i.e. learning to rank (L2R) algorithm and a vision transformer .i.e. called L2RT-FIQA was proposed[CY22]. It has three parts relative quality labels, L2R framework and a transformer model[CY22]. Here the quality labels for the samples are generated based on the normalised inter class and intra class angular distance[CY22]. Then the L2R model helps in a better generalization to the quality order than the quality value[CY22]. In another paper [SLH+23] by Shaolin et. al, a quite large dataset was developed .i.e. called **generic face image quality assessment(GFIQA-20K) dataset[SLH+23]. This dataset was explicitly developed for the purpose of FIQ prediction**[SLH+23]. It contains 20,000 face images from several individuals in different scenarios with generative prior information[SLH+23].

In the paper[KBR23] by Wassim et. al, they proposed a **robust sclera based segmentation for Face image quality assessment**. Their segmentation was a process that showed that the sclera pixels **become invariant to the different skin tones, age and ethnicity when statistically analysed**. In the paper [TKD+20] by Dr. Ing - Philipp Terhorst et al, they proposed a strategy called stochastic embedding robustness - face image quality(SER-FIQ) where the face image quality is computed by passing a sample through a neural network by employing dropouts. **Stochastic embeddings are obtained in each iteration to compute the quality score by determining the variations between the stochastic embeddings obtained for a sample.**SER-FIQ is the foundation of our experiments and let us discuss it in detail in the coming chapter. SER-FIQ also has been leveraged to benchmark against different recent methods and algorithms.

Overall, irrespective of every research work that has been performed in either face recognition or face image quality assessment, it is evident that continuous improvement is necessary.

2.3 Motivation

After reviewing the research works performed, it was clear that an in depth exploration of different configurations of dropout regularization was not performed by leveraging SER-FIQ algorithm[TKD+20]. Which might otherwise provide more deeper insights. Since face image quality assessment is the major area of concern in face recognition, to improve the performance of the system, further research is required. Where SER-FIQ[TKD+20] has been proven as one of the state of the art algorithm in FIQA. SER-FIQ[TKD+20] still possess limited explorations. Hence to fulfill this limited exploration by figuring out the research gap, the thesis proceeds forward with experiments.

Keeping SER-FIQ[TKD+20] as a starting point, an in depth investigation is provided on, how were the experiments performed to explore the effect of dropout regularization on FIQA using the SER-FIQ algorithm[TKD+20]. Therefore six different cases were framed namely, Case 1 consists of a Pretrained model with dropouts, Case 2 consists of a Pretrained model with dropouts and an on-top model trained without dropouts, Case 3 consists of a Pretrained model with dropouts and an on-top model trained with dropouts, Case 4 consists of a Pretrained model without dropouts, Case 5 consists of a Pretrained model without dropouts and an on-top model trained without dropouts, Case 6 consists of a Pretrained model without dropouts and an on-top model trained with dropouts. All these six different cases are considered for the intended experiments (refer chapter 4).

However in the paper [TKD+20], it was concluded that the image samples that have high stochas-

tic embedding variations (lowest quality scores) are regarded as low quality images. While the image samples that have low stochastic embedding variations (High quality scores) are regarded as high quality samples. In spite of this conclusion, we plan to investigate on a broader range of experiments. Where, we plan to analyse these experiments with two different assumptions. Firstly, assigning the lowest quality scores to the lowest quality images and secondly assigning highest quality scores to the lowest quality images. Such an indepth exploration is needed to obtain multiple perspectives and unbiased results.

The objective of the thesis is to perform experiments and obtain valuable insights on how does incorporating dropout regularization in the training of different use cases impact the performance evaluation of a face recognition system in different ways. Considering the experiments and their observations, the later chapters and sections describe how things can be made more dynamic for the future research. Collaborating with different scientific areas in FIQA should be done, so that we will be able to achieve better face recognition systems in the near future. This thesis seeks the heed of the individuals working and studying in the field of biometrics and automation domains. The main motivation behind this thesis is to investigate how does varying the dropout regularization in the neural network models might impact the face image quality assessment. The aspiration is to provide useful context and implications on how things could be improved. Specifically in terms of the experimental setup, modelling and performance. The thesis further provides valuable insights that could be carried out in the near future to address the challenges that we faced.

2.4 Structure of the thesis

Following the introduction chapter, the present chapter 2 provides related work from the last five years related to face recognition and face image quality analysis. The chapter 3 provides a detailed theoretical background that is necessary to comprehend our experiments. The next chapter 4 provides the methodology that we adopted in our experiments. Where it includes all the details ranging from the dataset preparation to the testing phase. The chapter 5 provides a detailed illustration of significant information that comprehensively analyses our results ranging from our results, findings to the future work that can be performed. The thesis finally concludes in the chapter 6. We provide our best efforts to perform the experiments and analyse them. Further our commitment for a continuous progress in the research has been loyally maintained.

3

Theoretical background

The theoretical background needed to comprehend the present thesis is really huge. It starts from the basics of biometrics, which we already saw in the introduction section. The present chapter introduces face recognition, face detection to other theoretical details which are necessary to comprehend our thesis. The decision to start the theoretical background from introducing face recognition attributes to its important role in today's world. Where face recognition might act as a basic theory that needs to be understood to establish an authentication system in most of the areas. By establishing a strong theoretical background, we aim to provide a way for understanding our thesis.

3.1 Face recognition

Face recognition (FR) is a biometric technology that plays a pivotal role in modern identification and verification systems[ZCPR03]. Basically FR is an automated technology that recognises an individual from digital face images or video frames[LMLP20]. It involves the analysis of an individual's facial features to determine their identity[ZCPR03], [LMLP20], [ZF14]. FR systems capture and process a range of facial patterns, including the spatial relationships between key features like the eyes, nose, and overall facial contours[MGMLG03], [PR04]. The applications of FR systems are diverse, spanning from security access control to convenient smartphone unlocking and social media photo tagging[JL11]. These applications of FR are spread worldwide. Table 3.1 shows a variety of applications used in different domains[JL11], [ZCPR03]. Such systems function by comparing a captured facial image with a database of known faces to ascertain a person's identity[TCZZ06], [JL11]. In recent years, FR technology has gained significant prominence due to its convenience and the potential it holds for enhancing security measures[ZF14], [JL11].

As the realm of Face Recognition (FR) continues to progress, the imperative to address the inherent challenges and limitations within the field becomes increasingly evident[JL11]. One notable challenge is the limited accuracy of current FR systems in unconstrained environments[JL11]. This limitation hinders their ability to perform efficient face recognition effectively under unconstrained environments[JL11], [JFR07], [Ter21], [ZZLQ16], [TIH$^+$23]. In such unconstrained environments, factors such as lightning, facial expressions and occlusions significantly affect the performance of FR systems leading to high error rates[JL11], [JFR07], [Ter21], [ZZLQ16], [TIH$^+$23]. Moreover, the increased use of FR systems in almost every aspect of this world in

Areas	Specific applications
Entertainment	Video game, virtual reality, training programs
	Human-robot-interaction, human-computer-interaction
Smart cards	Drivers' licenses, entitlement programs
	Immigration, national ID, passports, voter registration
	Welfare fraud
Information security	TV Parental control, personal device logon, desktop logon
	Application security, database security, file encryption
	Intranet security, internet access, medical records
	Secure trading terminals
Law enforcement and surveillance	Advanced video surveillance, CCTV control
	Portal control, postevent analysis
	Shoplifting, suspect tracking and investigation

Table 3.1: Important applications of face recognition [ZCPR03]

border control, law enforcement, security purposes, makes it very important in today's world. Additionally, confirming robustness of FR systems in adversarial attacks becomes very crucial. It poses a reasonable challenge, hence pushing the boundaries of the ongoing research. Since face images are a major part of face recognition systems, its important to explore more about face images[TKD+20].

Face images

Face images are one of the most widely used biometric modalities[TKD+20]. These face images are generally ubiquitous and acquirable in unconstrained environments[TKD+20], [JL11] . Acquiring these face images does not need an active user participation[TKD+20], [JL11], [JFR07]. Face image modalities in biometrics provides very high performance especially under constrained environments for several reasons[TKD+20]. First reason being non-intrusiveness, where face recognition does not need any active user participation[TKD+20], [JL11]. This reason makes it suitable for the scenarios where individuals might not be willing to provide their biometric data such as fingerprint scans or iris scans[TKD+20], [JL11]. Second reason being non contact scenarios, where face recognition can be done from a large distance[TKD+20], [JL11], [JFR07], [Ter21]. This is essential in scenarios where contact based biometric methods may not be feasible or hygienic, such as during pandemics[TKD+20], [JL11]. There are several other reasons.

High performance especially under constrained environments has been evidently visible with "super human performance" since 2014 with DeepFace framework [TYRW14]. There are many potential applications where face images are the commonly used modalities such as in financial sectors, public security systems like in automated border control, surveillance and forensics[JFR07], [JL11]. Face images are crucial in FR and we need to understand how did face recognition evolve overtime. Due to their good rate of acceptance, its becomes more crucial to push the research boundaries in assessing the face image quality, which inturn will contribute to a more robust FR systems[TKD+20], [SRH+22]. While face recognition is deeply attached to face images, its important to know about the evolution of face recognition.

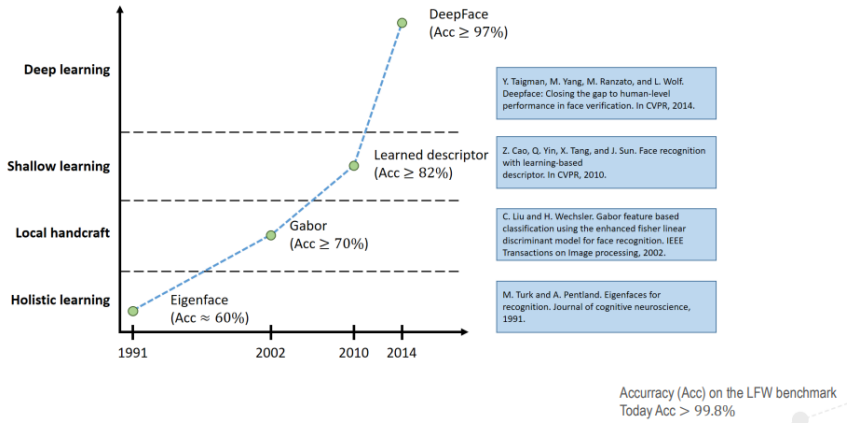

Figure 3.1: Evolution of face recognition[TYRW14], [CYTS10], [LW02], [TP91]

3.1.1 Evolution of face recognition

Evolution of face recognition started significantly since 1991 with the holistic learning approach of eigen faces by M.Turk et. al in the research study [TP91]. This holistic learning approach provided an accuracy of approximately upto sixty percent. Later Liu et. al [LW02], introduced a local handcraft method. Where they unveiled a Gabor-Fisher classifier(GFC) for FR[LW02]. This classifier is robust to the variations in facial expressions and illuminations[LW02]. It applies enhanced Fisher linear discriminant model (EFM)[LW02]. The GFC method was tested on various subjects on the FERET dataset[LW02], where the images have wide variety of illuminations and facial expressions[LW02]. This acquired an accuracy of about seventy percent. Yet later, a shallow learning approach was introduced by Z.Cao et. al [CYTS10], which reached an accuracy of about eighty four percentage[CYTS10]. Finally in 2014, Taigman et. al [TYRW14] introduced a deep learning approach which reached an accuracy of more than ninety seven percentage [TYRW14]. Thus, face recognition has been evolving from time to time where researchers have been continuously working on challenges. We need to be rest assured that the current challenges can also be tackled by the extensive research that is being carried out in the domain of biometrics. A grapical representation of the evolution of face recognition is illustrated in figure 3.1.

Throughout the evolution of face recognition, one significant factor that affects the performance of face recognition systems is the type of face images that are being fed into the biometric system[TKD$^+$20], [JL11]. Further we notice that illumination levels, facial expressions and face occlusions are few factors that classify various types of face images[TKD$^+$20], [ZCPR03]. Then we come down to capture types of face images because these factors come into picture only while capturing face images. While illumination levels are camera based, which is attached to the face recognition system, other factors like facial expressions and face occlusions are external factors that depends on the individual[ZCPR03], [JL11]. Hence while focusing on the capture types of the face recognition system, the factor that we need to address is illumination levels[ZCPR03], [JL11]. Therefore, light is a factor that we need to address. There are mainly three capture types based on the illumination of the light in the face images, which are **ultra violet (UV) radiation**[SSG$^+$18], **visible (VIS) light**[CFB05], [KB03] and **infrared radiation (IR)**[CFB05], [KB03].

Section 1.1.1 has a preprocessing module, in which the first component is a face detector module. It is one of the most important one in face recognition. Hence the following section describes face detection in detail.

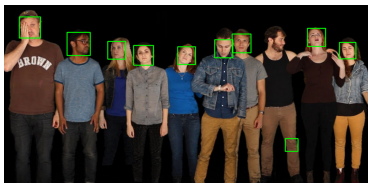

Figure 3.2: MTCNN for multiple people

Figure 3.3: Viola-Jones algorithm

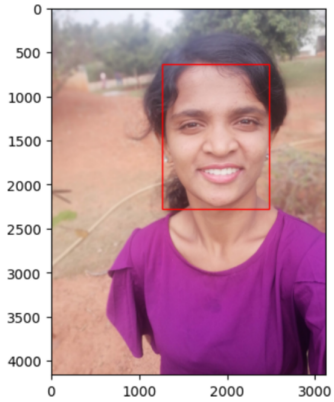

Figure 3.4: MTCNN for single person

Figure 3.5: Face detection

3.1.2 Face detection

Face detection is performed in the preprocessing module of the biometric system, where the face is detected in an input image, then the face is scaled, rotated, translated and cropped to ensure a consistent alignment between all faces[JL11], [JFR07]. In other words, face detection is a scientific discipline of biometrics, which is used to identify and locate a face in an image or a video stream, but does not go into the specifics of whose face it is[JL11], [JFR07].

More precisely, during the preprocessing step, the goals of face detection can be put forward as finding a bounding box that includes the face and to minimize the details that are not related to the identity. For example imagine a picture in which a person is standing in front of a movie theatre with his car. In this example, the face detection algorithm tries to minimize the details such as the movie theatre and his car, while detecting only his face. The face detection further simplifies in such a way that if an image has multiple people in it, it tries to detect multiple identity's faces, which further simplifies the comparison. After all these activities the face is scaled, cropped and aligned. Figure 3.4(single person face detection) shows an illustration of face detection, where a face is detected on a graphical scale. Whereas 3.2 and 3.3 shows face detection of multiple people in a group.

Its primary purpose is to determine whether a face is present or absent in the input image[JL11], [JFR07]. Which typically involves identifying the face bounding boxes without identifying the person in the face[ZZLQ16]. Face detection is commonly used in the auto focus of cameras, counting the number of people in a crowd and for basic surveillance[ZZLQ16], [JL11]. Face detection is usually done with algorithms that detects eyes, nose and mouth. These algorithms are designed to work efficiently to detect the presence of the face[ZZLQ16]. These algorithms prioritize the speed and efficacy in identifying the presence of faces in visual data and may sometimes tolerate few false positives and false negatives[ZZLQ16]. Some examples of these algorithms are: Retina face([DGV+20], [DGZ+19]), deepface([SWH18], [FSL15]), Multi-task Cascaded Convolutional Networks(MTCNN)[ZZLQ16], Viola Jones([CDRP18], [Wan14], [LM19]) etc.

MTCNN is the most commonly used face detection methodologies, which is an overall solution in the preprocessing step because it can simultaneously detect and align faces[ZZLQ16]. It exploits the inherent correlation between detection and alignment to boost their performance with multi-task learning[ZZLQ16]. In layman's terms, MTCNN makes use of the relationship between detection and alignment to improve the performance of the detection algorithm with doing

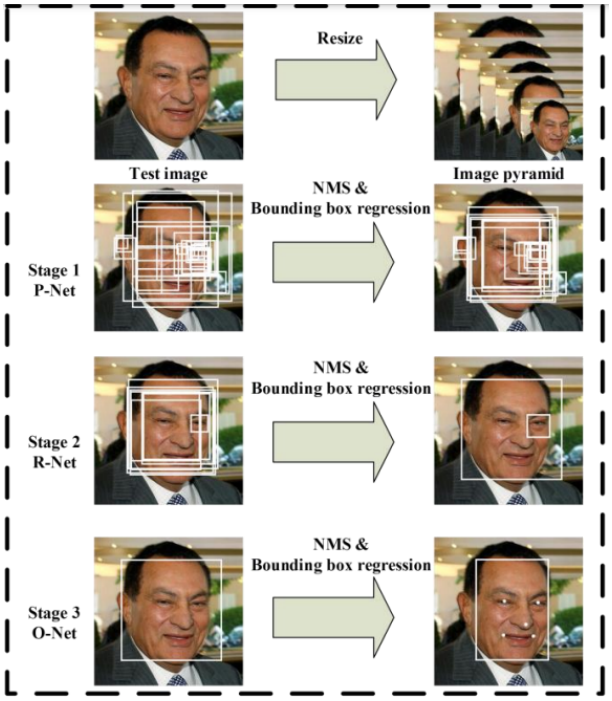

Figure 3.6: Pipeline of the cascaded network of MTCNN with its three different stages[ZZLQ16].

multiple tasks at a time[ZZLQ16]. These multiple tasks include Face/NoFace classification with a Cross-entropy loss function[LDL20] [WMZT20], Bounding box regression with a Euclidean loss function[CB17] [WMZT20] and Facial landmark localization with Euclidean loss function[CB17] [WMZT20].

MTCNN is a cascaded network architecture that is designed to efficiently detect faces in images[ZZLQ16]. Before passing an image into MTCNN, an image is resized. MTCNN basically has three stages, where each stage performs a specific task in the face detection process[ZZLQ16]. The first stage is Proposal network(P-Net)[ZZLQ16]. The main purpose of this stage is to generate candidate bounding boxes for potential faces[ZZLQ16]. Where, fully convolutional network is used to scan the input image and produce a set of bounding box proposals[ZZLQ16]. These proposals are scored for their likelihood of containing a face and are refined for better alignment[ZZLQ16]. The second stage is Refinement Network (R-Net)[ZZLQ16]. The second stage takes in the bounding box proposals generated by the first stage and then refines them[ZZLQ16]. R-Net then calculates the accuracy of each bounding box it has received from P-Net and then further narrows down the samples of bounding boxes to reduce the false positives[ZZLQ16]. This stage is very crucial to improve the peformance of the algorithm. The third stage is output network(O-Net)[ZZLQ16]. The third stage refines the candidates of bounding boxes from R-Net once more, and then provides the facial landmark localization[ZZLQ16]. This stage determines whether a face exists or not within the bounding box, so it also locates facial landmarks such as eyes , nose and mouth [ZZLQ16]. Figure 3.6 shows a detailed illustration of these three stages from the paper [ZZLQ16]. Further figure 3.7 shows a network level architectural representation of different stages of MTCNN.

3.1.3 Embedding extraction

Tn the context of face recognition, an embedding is a form of a matrix representation of a face image[SKP15], [DGXZ19a]. They are referred to as vector representations that cap-

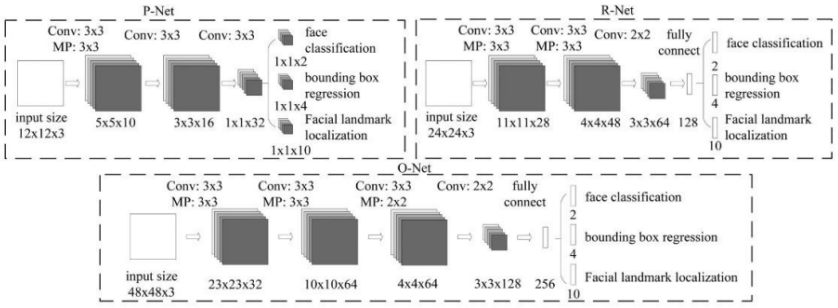

Figure 3.7: The architectural representation of P-Net, R-Net, and O-Net,"MP"-max pooling and "Conv"-convolution.[ZZLQ16].

ture important facial features in a compact and meaningful manner(Also known as a feature vector)[TYRW14]. The main objective of the process of embedding extraction is to map each face image into a mathematical space where important facial features are preserved[SKP15]. Once face images are converted into face embeddings, it becomes easy for the models and algorithms to perform a specific task assigned[SKP15], [DGXZ19a], [TYRW14]. For example in verifying an individual's face or an identification of an individuals face from a database[JFR07], [DGXZ19a]. Sometimes these embeddings are robust to the variations in face images and obtaining such robust embeddings is challenging in few scenarios[DGXZ19a]. Like when the face is too much occluded, different facial expressions, different age groups, different ethinicity. These factors make it challenging for a model to predict a robust embedding[SKP15], [DGXZ19a], [TYRW14]. These embeddings can be obtained through rigorously training the models through different architectures and loss functions[DGXZ19a].

Deep Face Recognition Models

In the modern era there has been a huge revolution in face recognition with the usage of deep learning(DL)[DL is explained in fundamentals section] in face recognition[FSL15] [SWH18] [TYRW14] [MWHN18] [PVZ15] [SO20]. Where, in these deep face recognition models, face recognition is a **"zero-shot representation learning task"**[MWHN18]. A zero shot representation learning task basically means that once a model is trained with a single input image, it is capable of learning all the intricate facial features, which need not be retrained to recognise new individuals[MWHN18]. Since for most applications it is not possible to include candidate faces during training, deep face models serve as a best choice in such scenarios[MWHN18]. For example : Imagine a database where huge population is involved, such scenarios are not scalable to train the model for each enrolment/deletion[MWHN18]. Such an excellent recognition has been possible due to the presence of the deep neural networks present in the deepface models[MWHN18].

Most tasks in deepface are performed with **transfer learning**[WKW16] [TS10] [ZQD$^+$20a] [MWHN18] , meaning that the network training is based on a closed pool of subjects and is then used as a feature extractor on unseen faces[MWHN18]. More precisely, in machine learning we have this amazing technique called transfer learning[MWHN18] [WKW16] [TS10]. In which, initially a model is trained for a particular task, subsequently the same model is repurposed for another related task[WKW16] [TS10]. In the context of face recognition, it means that a model is trained on a specific dataset(closed pool of subjects) to perform an explicit face recognition task[MWHN18]. Where a closed pool of subjects is the predefined group of people in a specific dataset, where the deepface model learns to recognise their face images during the

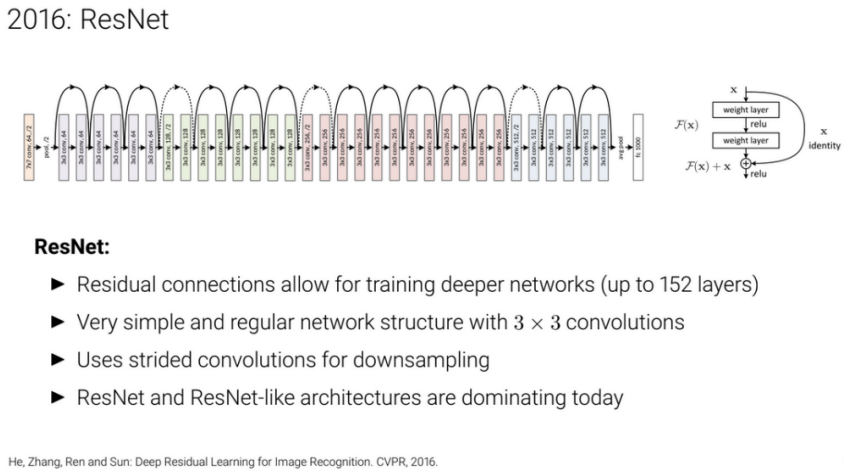

Figure 3.8: Resnet architecture.[HZRS16].

training phase[MWHN18]. Following the training phase, the neural network of the model is used as a feature extractor on the unseen images[MWHN18] [TKD+20]. This feature extractor captures/extracts intricate facial features from the unseen images[MWHN18] [TKD+20]. Thus the trained model is applied on unseen face images, which were not available in the training phase[MWHN18]. Here the learned features from the training phase are used to extract features and recognise the unseen faces[MWHN18]. In such a way a generalizability is possible since human faces share a similar shape and texture[MWHN18] [TKD+20] [TIH+23] [Ter21].

Generally, deep face recognition solutions mainly differ in three aspects. Firstly, the **utilized network architecture that is trained for the task of recognising faces**[MWHN18]. The architectures used in face recognition are mostly similar to that of the object detection. Few of them are AlexNet[KSH], VGGNet[SZ14], ResNet[HZRS16], iResNet[DLZS21] etc. A pictorial representation of the most dominant and commonly used architecture is shown in figure 3.8. Secondly, the **loss function that guides the network training**[MWHN18]. Thirdly, **the utilized training data** that reflects the inter- and intra-subject variations and thus, builds the fundamentals of the training stage[MWHN18].

Embedding extraction aims to minimize intra-identity variations and to maximize inter-identity variations[MWHN18], [VV21], [PVZ15], [ZCPR03]. There are two main approaches to learning face embedding using deep neural networks. First and foremost is the training of the networks to directly learn the face embeddings, which is otherwise called **direct representation learning(DRL)**[SKP15] , [ALS+16]. In this learning strategy, the deep face models are trained to directly learn face embeddings without the involvement of the training phase[SKP15]. In such a scenario, the face images are widely mapped onto a continuous vector space[SKP15] [ALS+16]. Here, the distance between the vectors illuminates the similarities and dissimilarities in the facial features[SKP15] [ALS+16]. The most common loss function that guides the neural network in DRL is the **triplet loss function**[HBL17] [Ge18] [ALS+16]. In a triplet loss, a selection of triplets of face images is involved, an anchor(A), a positive(P) and a negative(N)[HBL17] [ALS+16]. Where a network is trained with triplets of the form a,p,n, a (anchor) and p (positive) are samples that belong to the same identity 'i' and n (negative) is a sample that belongs to another identity 'j'[HBL17] [ALS+16]. The goal is to minimize the distance between the anchor and positive, such that it is smaller than the distance between the anchor and any other negative sample of any other identity[SKP15] [ALS+16]. This lead to the triplet loss function[HBL17] :

$$\mathcal{L}_{\text{Triplet}} = \frac{1}{N}\sum_{i}^{N}\max\left\{0, \|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha\right\} \text{[HBL17]} \tag{3.1}$$

This equation guides the network training to minimize intra-subject variations as well as maximize the separation between different identities[HBL17]. One major challenge here is that this training procedure is not suitable for large datasets [Ge18]. Since the number of possible triplet pairs grow exponentially and thus, the selection of suitable (semi-hard) triplets become difficult[HBL17] [Ge18].

Moving on to the second approach, is to train the neural networks using multi-class classifiers which is otherwise called **Classification learning** on the training identities[MWHN18] [TYRW14] [SO20] [DGXZ19b] [LWYY16]. In this learning strategy, the deep face models are trained on a dataset with necessary hyperparameters with multiple classes[MWHN18] [LWYY16] . The model is compiled and trained with a desired batch size and the desired number of epochs[MWHN18]. Here, after training the model, the last layer is removed to use the second-last layer to extract the facial features for predictions. The most common loss function that guides the neural network in classification learning is the **Softmax loss function**[MWHN18] [JL11] [DGXZ19b] [WWZ+18] [LWYY16].

Softmax-based approaches aim at classifying a closed-set of identities during training and utilizes a previous layer as a feature extractor for unseen faces [LWYY16] [MWHN18] [JL11]. The equation of a traditional softmax loss function is given by the below equation[LWYY16] [MWHN18] [DGXZ19b] [WWZ+18]:

$$\mathcal{L}_{\text{Softmax}} = -\frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{N} e^{W_j^T x_i + b_j}}\right) \text{[LWYY16]} \tag{3.2}$$

This equation combines a softmax activation on the classification layer with a standard cross-entropy loss[LWYY16] [Ter21] [TIH+23] [MWHN18] [PVZ15] where $x_i \in R^d$ refers to the template of the i-th of N training samples that belong to subject y_i, $W_j \in R^d$ denotes the j-th column of the weight matrix $W \in R^{d\times n}$ with n equals the number of training identities. , $b_j \in R^n$ defines the bias term.

Softmax loss does not explicitly minimize intra -subject variations.This issue was tackled by another loss function called as **Center loss**[QS17]. Center loss tackled this issue by minimizing intra-subjects distances between samples x_i and their corresponding class-centroids C_{yi} that determines the class center of the deep features[QS17]. This results in the center loss, which is represented by the equation below[QS17]:

$$\begin{aligned} \mathcal{L}_{\text{Centerloss}} &= \mathcal{L}_{\text{Softmax}} + \frac{\lambda}{2}\mathcal{L}_{\text{Center}} \\ \mathcal{L}_{\text{Center}} &= \frac{1}{2}\sum_{i=1}^{N}\|x_i - c_{y_i}\|_2^2 \end{aligned} \quad \text{[QS17]} \tag{3.3}$$

Other methods are directly based on softmax loss[LWY+17] [DGXZ19b] [WWZ+18]. For simplicity,

1. The bias terms can be fixed to $b_j = 0$.
2. The individual weights can be normalized $\|W_j\| = 1$.

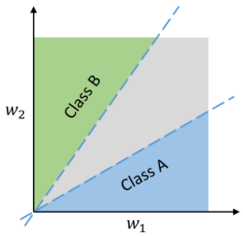

Figure 3.9: Transformation of softmax loss.[LWYY16].

3. The embedding $\|x_i\|$ can be rescaled to $\|x_i\| = r$.

By computing all the above values in the softmax loss function leads to the following transformation :

$$W_j^T x_i + b_i \overset{b_i=0}{=} \|W_j\| \|x_i\| \cos(\theta_j) \overset{\|W_i\|=1}{=} r \cos(\theta_j)$$

Where θ_j is the angle between weight W_j and the feature vector x_i, the embeddings are only dependent on the angle (angular margin)and the embeddings are distributed on a hypersphere with radius r. This leads to the SphereFace[LWY+17] loss. A graphical representation of this transformation is shown in the figure 3.9.

SphereFace[LWY+17] loss is represented by the equation below:

$$\mathcal{L}_{\text{Sphereface}} = -\frac{1}{N}\sum_{i=1}^{N} \log\left(\frac{e^{r\cos(\theta_{y_i})}}{e^{r\cos(\theta_{y_i})} + \sum_{j=1, j\neq y_i}^{N} e^{r\cos(\theta_{y_i})}}\right) \text{[LWY+17]} \tag{3.4}$$

SphereFace[LWY+17] loss introduces the idea of angular features which aims to learn angularly discriminative features.

Adding a cosine margin penalty to the same function leads to **CosFace**[WWZ+18]. Where a more higher generalization due to the added margin principle is seen[WWZ+18]. Shifting the margin to the angular level leads to **ArcFace**[DGXZ19b]. As the representations are distributed around each representation center on the hypersphere of radius r, adding this additive angular margin penalty simultaneously improves the inter-subject separability and the intra-subject compactness[DGXZ19b]. Arcface generally has more distinct features with a more stable training[DGXZ19b]. The equations of cosface and arcface are as below[DGXZ19b] [WWZ+18]:

$$\begin{aligned}
\mathcal{L}_{\text{CosFace}} &= -\frac{1}{N}\sum_{i=1}^{N} \log\left(\frac{e^{r\cos(\theta_{y_i})-m}}{e^{r\cos(\theta_{y_i})-m} + \sum_{j=1, j\neq y_i}^{N} e^{r\cos(\theta_{y_i})}}\right) \\
\mathcal{L}_{\text{ArcFace}} &= -\frac{1}{N}\sum_{i=1}^{N} \log\left(\frac{e^{r\cos(\theta_{y_i}+m)}}{e^{r\cos(\theta_{y_i}+m)} + \sum_{j=1, j\neq y_i}^{N} e^{r\cos(\theta_{y_i})}}\right)
\end{aligned} \quad \text{[DGXZ19b][WWZ+18]} \tag{3.5}$$

The decision boundaries for class A and class B for all the softmax bases loss functions we have seen until now are shown in figure 3.10. Where Arcface is bench marked as the best loss function which provides the best performance[DGXZ19b]. It is also important to address a loss function that is used in our experiments to train the on-top models. In every use case, there was one choice of the loss function to compile the on-top models .i.e. sparse categorical crossentropy (SCCE)[TCERPCU23]. SCCE intelligently handles multiple classes, it

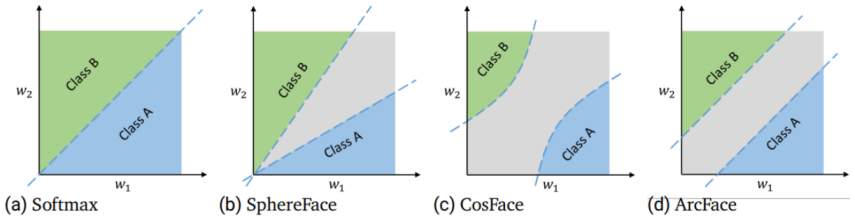

Figure 3.10: Softmax-based losses–Decision boundaries [DGXZ19b], [LWY+17] [WWZ+18] [LWYY16]

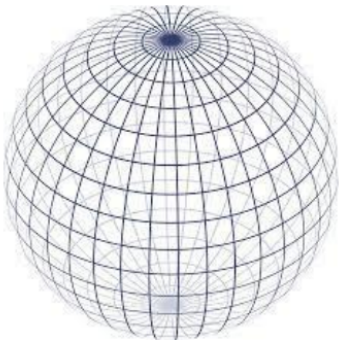

Figure 3.11: Sphere

is compatible with dropout regularization and guides the neural network for an efficient feature extraction[TCERPCU23]. It maximizes the likelihood of the correct identity to learn exact embeddings that pertains to the intricate features of that identity[TCERPCU23]. It's formula can be derived from the softmax loss function which we saw in this chapter before in the equation 3.2[TCERPCU23]. Where the idea of the same softmax loss function can be extended to an idea of multiple classes[TCERPCU23].

$$\text{SCCE} = -\sum_{i}^{N} y_i \cdot \log\left(\frac{e^{f_i}}{\sum_j^N e^{f_j}}\right) \text{[TCERPCU23]} \tag{3.6}$$

Where N is the total number of classes in the classification problem, i is the index representing the current class, y_i is an indicator function that takes the value 1 if the true class label is i and 0 otherwise[TCERPCU23]. In the context of this loss function, it is sparse, meaning it is one-hot encoded, i.e., $y_i = 1$ for the true class and $y_j = 0$ for non-true class ($j \neq i$)[TCERPCU23]. f_i is the logit (pre-softmax) value corresponding to class i[TCERPCU23]. Where, $\sum_j^N e^{f_j}$ is the sum of the exponential values of all logit scores across all classes[TCERPCU23]. This term is in the denominator of the fraction and represents the normalization factor ensuring that the probabilities sum to 1 after applying the softmax[TCERPCU23].

Feature comparison : In practice, face embeddings x are often scaled to unit-length

$$||x|| = 1$$

This simplifies the cosine-similarity calculation to a simple dot-product calculation[ZCPR03] [MWHN18].

$$\text{cosine similarity } = S_C(A, B) := \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} = A.B$$

Embeddings are represented as points on the surface of an n-dimensional sphere [LWY+17] [QS17] [DGXZ19b] [WWZ+18]. While every aspect of deep face recognition models is important. It is important to know about the models that were used in our experiments.

The first model, **ArcFace model(Model pretrained with dropouts)** was picked from the DeepFace library[SO20]. It is a model which was built as a regular FR algorithm that is responsible for representations[SO20]. This model is purely based on ResNet34 architecture 3.8 which has an input size of (112, 112, 3) and the output of this model is a 512 dimensional vector. This model is pretrained with 162 layers including dropouts. Whose training is purely guided by Additive Angular Marginal Loss[DGXZ19b], explained in chapter 3.1.3. The training of this model started with training a deep convolutional neural network(DCNN)[MWHN18] [DGXZ19b] which is guided by ArcFace loss function[DGXZ19b]. Where the normalised features and the normalised weights are passed through the additive angular margin penalty to obtain the intricate features , finally passing them through the softmax function to get the cross entropy loss[DGXZ19b]. The original pseudo code for ArcFace with its training architecture is presented in figure 3.13 [DGXZ19b]. **The main objective of choosing ArcFace model** was because many scientific researches have proved the accuracy that, this model provides because of its loss function[MWHN18], [DGXZ19b].

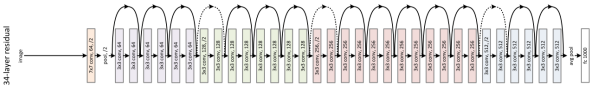

Figure 3.12: ResNet34 Architecture

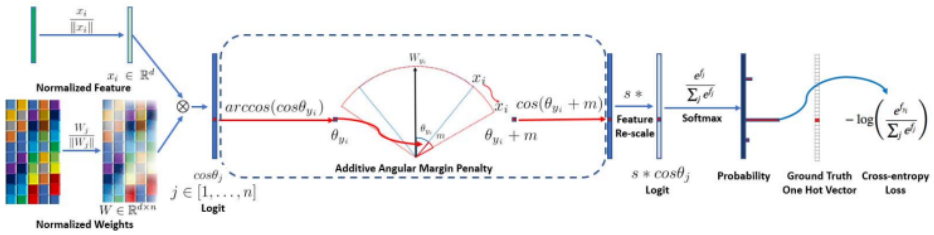

Figure 2. Training a DCNN for face recognition supervised by the ArcFace loss. Based on the feature x_i and weight W normalisation, we get the $\cos\theta_j$ (logit) for each class as $W_j^T x_i$. We calculate the $arccos\theta_{y_i}$ and get the angle between the feature x_i and the ground truth weight W_{y_i}. In fact, W_j provides a kind of centre for each class. Then, we add an angular margin penalty m on the target (ground truth) angle θ_{y_i}. After that, we calculate $\cos(\theta_{y_i} + m)$ and multiply all logits by the feature scale s. The logits then go through the softmax function and contribute to the cross entropy loss.

Algorithm 1 The Pseudo-code of ArcFace on MxNet

Input: Feature Scale s, Margin Parameter m in Eq. 3, Class Number n, Ground-Truth ID gt.

```
x = mx.symbol.L2Normalization (x, mode = 'instance')
W = mx.symbol.L2Normalization (W, mode = 'instance')
fc7 = mx.sym.FullyConnected (data = x, weight = W, no_bias = True, num_hidden = n)
original_target_logit = mx.sym.pick (fc7, gt, axis = 1)
theta = mx.sym.arccos (original_target_logit)
marginal_target_logit = mx.sym.cos (theta + m)
one_hot = mx.sym.one_hot (gt, depth = n, on_value = 1.0, off_value = 0.0)
fc7 = fc7 + mx.sym.broadcast_mul (one_hot, mx.sym.expand_dims (marginal_target_logit - original_target_logit, 1))
fc7 = fc7 * s
```

Output: Class-wise affinity score $fc7$.

Figure 3.13: Pseudo code for ArcFace algorithm along with its training architecture[DGXZ19b]

The second model, **OpenFace Model(Model pretrained without dropouts)** was also picked from the DeepFace library[SO20], but was from the original paper [ALS+16] by Amos et. al. released in October 2015 under the Apache 2.0 license. It is available at: http://cmusatyalab.github.io/openface/ [ALS+16] and it has 165 layers. Where OpenFace's

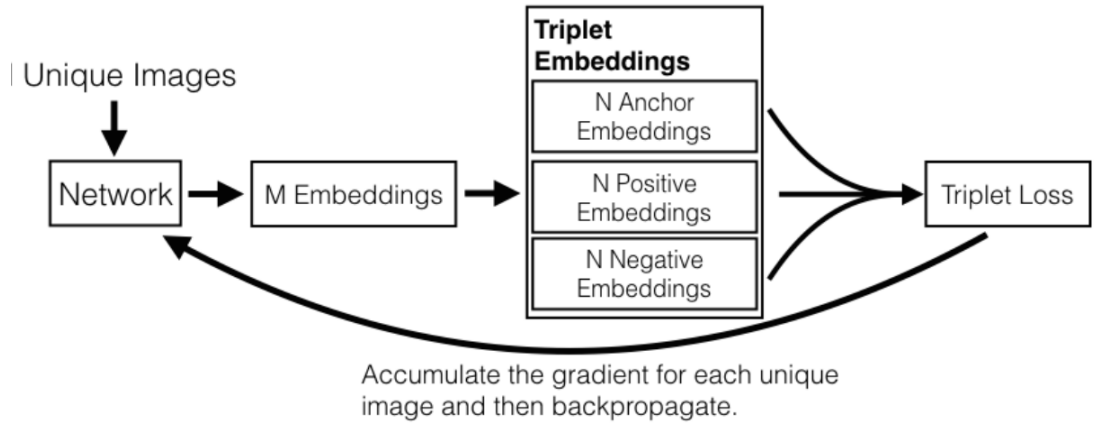

Figure 6: OpenFace's end-to-end network training flow.

Figure 3.14: OpenFace network training flow with an improved triplet loss function, where the gradients of the loss function are backpropagated, to train the OpenFace model.[ALS+16]

end-to-end network training flow starts with M-Unique images passed into the network, then M-embeddings are extracted from the network which are then divided into N-Anchor embeddings, N-Positive embeddings and N-negative embeddings to obtain the triplet loss, then the gradients are back propogated[ALS+16]. In other words, unique images are mapped from the network to the triplets[ALS+16]. The resulting gradient of the triplet loss through the network is backpropagated back through mapping to the M-unique images[ALS+16]. There are mini batches created, where they sample X images per person from Y people in the dataset and M = XY images are passed through the network in a single forward pass to obtain M embeddings[ALS+16].All anchor positive pairs were considered to have $N = Q \begin{pmatrix} P \\ 2 \end{pmatrix}$ triplets. Then the triplet loss is computed and its derivate is mapped back to the original image in a backward network propagation[ALS+16]. Finally, if the negative image is found within the range of α^2-margin for a given anchor-positive pair, the image is not used further[ALS+16]. A detailed visual illustration of OpenFace's network training flow is shown in figure 3.14. **The main objective of choosing OpenFace model** was with an assumption that the improved triplet loss function could give some desirable results.

Current face recognition systems perform highly accurate under controlled/constrained conditions[HA15], [AOBTA20]. But under unconstrained environments, the FR system has to deal with large variabilities[TKD+20]. Hence leading to a degraded recognition performance[TKD+20]. The degraded recognition performance can be due to many factors namely Image acquisition conditions, Factors of the face(face expressions, face occlusions), Model bias etc[TKD+20], [ZCPR03]. To avoid the degraded recognition performance, there is a need to detect the face images that are less suitable for face recognition before any comparison[TKD+20], [SRH+22]. Can we really do this? We can reject the low quality images during enrolment and reduce the number of errors a system will make[TKD+20], [TIH+23]. The increasing need for a more secure and accurate FR systems has activated a continuous growing interest in Face Image Quality assessment (FIQA) [SRH+22] as a vital scope within face recognition.

3.2 Face image quality

The performance of biometric recognition is driven by the quality of its samples[SRH+22], [TKD+20]. Where **Biometric sample quality** is defined as an estimate that determines the utility of a sample for recognition(**ISO/IEC TR 29794-5:2010**). **Face Image Quality(FIQ)** assesses the quality of a face-image, aiming to quantify its suitability for face recognition[SRH+22]. Some advantages of sample quality are that they lead to a more robust enrolment, more secure negative identification systems and enables quality-based fusion approaches. Consider an experiment where we tend to compare high quality images and noisy images against each other. When we Report (1.)the similarity between the pairs of an original image and its respective degraded image, and (2.)the similarity between degraded images of different identities. The reported similarities indicate how well are the degraded images and high quality images scored.

The assessment of face image quality (FIQA) involves taking facial images as input and generating an estimation of their image quality[SRH+22]. FIQA is a significant part of the FR systems for processing face images[SCDR22], [TKD+20].The conventional face image quality is solely focused on perceptual quality, but there are a variety of things that affects the quality like the capture, the environment, background and lightening [SCDR22], [TKD+20]. Input face images with low quality will undoubtedly lead to poor performance [SCDR22], [TKD+20]. Further improving the quality of facial images helps in improving the system performance.

Where the process of FIQA involves an automated approach .i.e. called a Face image quality assessment algorithm (FIQAA)[SRH+22]. Each FIQAA consists of a quality score (QS)[SRH+22], [TKD+20]. The FIQA process takes in a face image, preprocess it with face detection algorithms, after which the quality assessment is performed and finally a quality score(QS) is calculated [SRH+22], [TKD+20]. After calculating the QS, the FIQAA classifies the face images into different quality categories, such as low and high quality, or assigns a specific quality label to each image. Figure 3.15 shows a detailed illustration of the FIQA process. Parallely figure 3.16 shows different qualities of face images of a single identity.

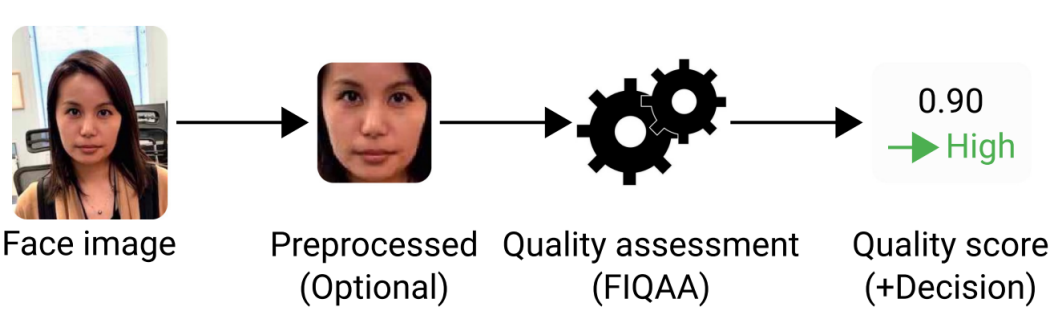

Figure 3.15: FIQAA process where a raw face image is preprocessed and is passed through an algorithm that assigns a quality score to the face image, where the decision is made if the image is of low quality or high quality.[SRH+22].

There are a lot of advancements done on FIQA and FIQAA. In this way the research on FIQA has been really intensive, where researchers have been developing more sophisticated approaches (FIQAAs) to enhance the FIQA process so that the accuracy, robustness and the efficiency improves [SRH+22]. Some key aspects of the research conducted in FIQA would mostly include privacy and security considerations, algorithmic development, modelling architectures and performance of the models [SRH+22]. Despite of all its advantages, FIQA has its own drawbacks like impediments in handling image variability (pose, illumination, occlusions), subjectivity in quality perception, limited generalization to new scenarios, biases from training datasets etc.

Figure 3.16: Different face images of a specific identity. The face image quality degrades from left to right, as different poses, illuminations and face expressions are introduced.[SRH+22], Taken from the reference [GGNH14]

[SRH+22], [TKD+20].

3.2.1 Evaluation of Face image quality

The problem during the evaluation of the face image quality is that ground truth information about the face images is not available[TKD+20], [Ter21], [GT07]. Which means that there is no specific information regarding the quality of the face images in a dataset[TKD+20]. The solution to evaluate FIQ is to consider high quality samples and these high quality samples should lead to improved recognition performance[Ter21]. The main idea here is to iteratively remove low quality samples from the dataset, with this the recognition error should iteratively decrease[TKD+20], [GT07]. To perform this task, we need to plot an **error versus reject curve**[TKD+20]. On the Y-axis, we consider the recognition error and on the X-axis we consider the percentage of the neglected data (of the worst estimated quality)[TKD+20], [GT07]. When interpreting these curves, the assessment method with a faster falling curves better estimates the quality of face images in a dataset[TKD+20]. In other words, as we start rejecting the low quality images iteratively, the error rate on the Y-axis decreases rapidly[TKD+20], [GT07]. A visual illustration of the error vs reject curve is shown in figure 3.20 from the SER-FIQ[TKD+20] paper.

What Affects Biometric Sample Quality?

There are various factors that affect the biometric sample quality. Firstly, **Model-Specific FIQA approaches** generally result in low-quality values for images with high error rates. In other words, the FIQA approaches that are designed with a specific recognition model results in low quality values of the face images and thus leading to higher error rates through the process. Secondly, **head pose**[SRH+22]. The images with different head poses in different directions can affect the sample quality significantly[SRH+22], [BVS14]. For example: If there are various face images of the same person 3.17, but each image varies with different head poses. Then the recognition model tends to detect intricate features from the images, thus generating different face embeddings, which indirectly affects the biometric face quality. Figure 3.17 shows a collection of various images of the same identity with different qualities and different head poses[BVS14]. Different head poses introduce variability in the facial features. Where there is an exclusive information loss because it FIQA algorithms are trained to recognise the facial landmarks with their spatial relationships. When there are head poses like side ways of backwards, there is a loss of the facial information, thus leading to inaccurate face image qualities[BVS14], [SRH+22].

Thirdly, ethnicity. With varying ethnicity all across the world with different facial features and skin tones, there is a significant effect on the biometric sample quality[KBR23], [TGBB20]. Different ethnicities have different facial features and skin texture[KBR23], [TGBB20]. The FR

Figure 3.17: A collection of different quality images of the same identity[BVS14]

models that are not trained with this diference will face variations in accuracies across different ethnic groups. Also, different ethnicities possess different different different facial adornments and different hairstyles, which will affect the biometric sample quality significantly[TGBB20], [SRH+22]. Similarly there can be several ethnic factors listed. Fourthly, age. As an individual ages, his/her physical features tend to change overtime, which also significantly affects the biometric sample quality[SRH+22]. Lastly, face occlusions like face masks, eye glasses etc[SRH+22],[TKD+20]. Face occlusions play a major role in affecting the biometric sample quality, because the recognition of facial features become difficult when an individual wears a face mask or some other thing[TIH+23], [SRH+22].

3.3 Stochatic embedding robustness - face image quality(SER-FIQ)

SER-FIQ is a face image quality assessment method, which was introduced by Dr.-Ing. Philipp Terhörst et. al in the paper [TKD+20]. To implement this method, a face image is passed through a model that is trained with dropouts, to get different stochastic embeddings [TKD+20]. This is possible due to the presence of dropouts in the model [TKD+20]. To ensure that the method should be applicable on every model, they proposed to build an ontop model, on the base model [TKD+20]. Here they call the base model as the SER-FIR(same model) and the model that is built upon the same model as the SER-FIQ(on-top model) [TKD+20]. Then after training the on-top model, testing is done by passing the face images in the dataset, to compute the quality scores for each image [TKD+20]. Subsequently after computing the quality scores, error versus reject curves are plotted [TKD+20]. Here, the face image that produces small variations in the stochastic embeddings is regarded to demonstrate high robustness, and thus high image quality [TKD+20]. Whereas, the face image that produces higher variations is regarded as a low robustness, thus is is considered as a low quality [TKD+20]. The visual working principle of this method is presented in figure 3.18.

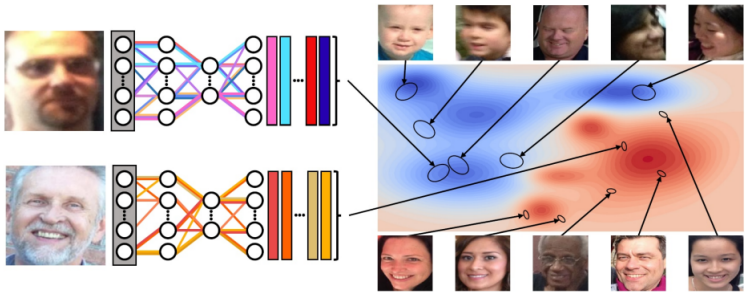

Figure 3.18: SER-FIQ working principle - Visualization of the proposed FIQA technique in the paper. Where stochastic embeddings are generated for each image sample. The sample that has greater variations in its stochastic embeddings is regarded as a low quality sample. While the sample that has low variations in its stochastic embeddings is regarded as a high quality sample[TKD+20].

3.3.1 Algorithm

In the algorithm of the SER-FIQ method, the exact procedure for the computation of the quality scores is described [TKD+20]. The algorithm takes in the input as a preprocessed input image I and a Neural Network model M [TKD+20]. The function SER takes in the parameters SER(I, M, m=100), Where I is the input preprocessed image, M is the neaural network model and m=100 is the number of stochastic forward passes to obtain different embeddings [TKD+20]. Further, variable X is intialised as an empty list to store the embeddings from each pass [TKD+20]. Where each pass produces unique embeddings 3.19. Then the loop of the number of stochastic forward passes begins [TKD+20]. Now dropouts are applied and predictions are made from the neural network model by passing through the input image. Each pass gives a unique embedding. In each iteration, the embeddings are collected in the list X [TKD+20]. Finally, this list X is used to compute the quality score q [TKD+20]. The pseudocode for this algorithm is presented in algorithm 1. The equation to compute the quality score is presented inside the algorithm[TKD+20], rather it is separately explained for an in depth understanding.

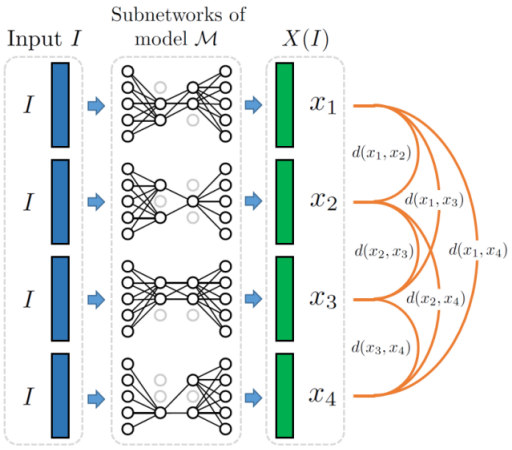

Figure 3.19: [TKD+20].

Algorithm 1: Stochastic Embedding Robustness (SER)

Input: Preprocessed input image I, NN-model M
Output: Quality value Q for input image I

Procedure `SER`$(I, M, m = 100)$
  $X \leftarrow$ empty list;
  for $i \leftarrow 1$ **to** m **do**
    $x_i \leftarrow M.pred(I, \text{dropout} = \text{True})$;
    X $\leftarrow$ X.add(x_i);
  end
  $Q \leftarrow$ `ComputeQuality(`X`)` ; `// Compute the quality value`
  return Q;
Function `ComputeQuality(`X`)`
  $q \leftarrow 2\sigma\left(-\frac{2}{m^2}\sum_{i<j} d\left(x_i, x_j\right)\right)$;
  return q;

The equation 3.7, has a parameter $d\left(x_i, x_j\right)$ which computes mean pairwise distances between all the embeddings stored in the list $X(I) = \{x_s\}_{s\in\{1,2,\ldots,m\}}$ by using the euclidean distances[TKD$^+$20]. Therefore they define the face quality metric of an image I as the sigmoid of the negative mean euclidean distances between all stochastic embedding pairs in $X(I)$[TKD$^+$20]. A greater variations in the stochastic embeddings in $X(I)$ indicates low robustness, thus the value of the quality metric q becomes low indicating a low quality image[TKD$^+$20]. Whereas smaller variations in the stochastic embeddings in $X(I)$ indicates high robustness, thus the value of the quality metric q becomes high indicating a high quality image[TKD$^+$20].

$$ComputeQuality(X) = 2\sigma\left(-\frac{2}{m^2}\sum_{i<j} d\left(x_i, x_j\right)\right), [\text{TKD}^+20] \tag{3.7}$$

The final evaluation of the performance of a face recognition system for the SER-FIQ algorithm is performed using the error versus reject curve[TKD$^+$20] as described above. Where low quality images are iteratively rejected to compute the error rate at each iteration[GT07], [TKD$^+$20]. As we reject low quality images, the performance of the face recognition system improves[GT07]. This improvement of the performance can be inferred from the error versus reject curve, when the curve decreases in each iteration[TKD$^+$20]. A decreasing curve indicates a good face recognition performance, while an increasing curve indicates that the face recognition performance is not as expected[GT07]. Few plots from the original paper are illustrated to provide an overview of how the performance of an ideal face recognition system looks like in figure 3.20.

3.4 Deep learning and convolutional neural networks

Machine learning is a broader field that has techniques and algorithms, which makes decisions or predictions based on the data[WR17]. Deep learning is a subset of machine learning, where it can automatically learn hierarchical representations of the data[WR17]. It basically makes use of the neural networks with multiple layers for solving very complex problems[DCWZ16]. Such neural networks were devised based on the neural patterns of a human brain(Artificial neural networks)[WR17], [DCWZ16]. They are made up of several interconnected nodes where the depth of these networks makes them more effective[WR17], [DCWZ16]. Especially convolutional neural networks(CNNs) have been immensely beneficial[WR17], [DCWZ16], [ON15].

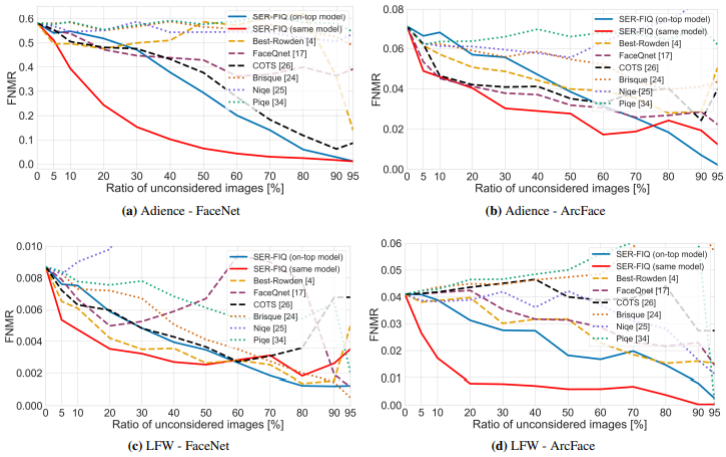

Figure 3.20: Evaluation of the face recognition system with face image quality assessment (Error versus reject curves)[TKD+20].

3.4.1 Convolutional neural networks

The convolutional neural networks(CNNs) self optimise through the process of learning, where each neuron receives an input and performs operations on the input[ON15]. They perform a **'convolution operation'** on the input data by using certain filters, to extract an output data[ON15]. An architecture of a CNN has five basic elements namely **an input layer, a convolutional layer, a pooling layer and a fully connected layer.**[ON15]

Assume we have an input image. An **input layer** is intended to hold the pixel values of this input image[ON15]. **A convolutional layer** determines what will be the output of the neurons, which are connected to the primary regions of the input image[ON15]. Each neuron in the convolutional layer has a kernel(filter) and this kernel convolves throughout the input image performing scalar product operations[ON15]. The output of each operation is determined by performing a scalar product between the weights of the neurons and the corresponding regions of the input image[ON15]. For a specified position (i,j) of an image, the scalar product is computed mathematically as follows [ON15]:

$$(I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) \cdot K(m,n)$$

where I is the input image, K is the filter, and m and n are the indices of the filter. After obtaining this output, a non linearity is introduced by passing this output into an activation function(reLU - Rectified Linear Unit). The final output is a two dimensional vector which is the result of activation of a neuron in response to the convolution[ON15]. The **pooling layer** further reduces the dimensionality and the number of parameters from the output of the convolutional layer[ON15]. However, this pooling layer retains the most important features[ON15]. After the convolutional layer and pooling layer, comes the **fully connected layer**[ON15]. This fully connected layer considers the high dimensional feature vectors from the previous layers and flattens into a one dimensional vector[ON15]. Every neuron in this layer is connected to every neuron in the previous layer so that the network can learn all the complex features and give final predictions[ON15]. A simple CNN architecture that has five simple layers is illustrated in figure 3.21.

These convolutional neural networks capture the spatial dimensions[ON15]. Hence these CNNs are used to train large datasets with different hyperparameters(parameters used for training), so

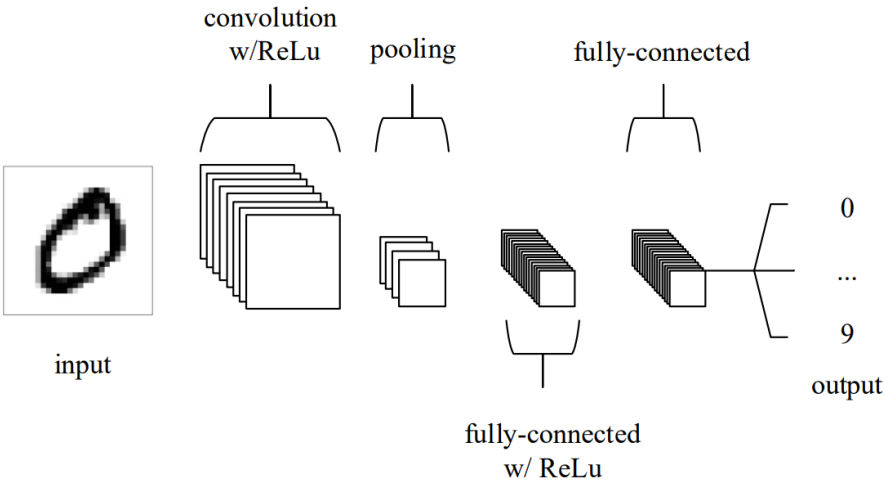

Figure 3.21: CNN architecture with five simple layers[ON15].

that a model is formed[ON15]. These models are further used for predictions on different datasets to extract intricate features, so that advanced functions can be performed[ON15]. Like for example: face image quality assessment can be performed by extracting feature embeddings by using the models that are already trained. This is how SER-FIQ algorithm was implemented. As we know that the CNNs capture the spatial data, there are another type of neural networks that capture the temporal data which are called recurrent neural networks(RNNs)[MJ01]. In other words temporal data means, the RNNs capture the pattern of change of data over time[MJ01]. Examples of such networks are LSTMs(Long short term memory)[MJ01]. However, RNNs are not included in our experiments, that is the reason RNNs are explained very shortly here. But they are needed to explore the future research.

3.4.2 Dropout regularization

As described in the previous section, a model is formed by using the neural networks like CNNs[SHK+14], [Sri13]. Such models sometimes learn the features from the training dataset too well that, they tend to memorize those features, that the models finds it entirely difficult to generalise to the unseen data[SHK+14], [Sri13]. This surely increases the training accuracy upto 99.99%, but the validation accuracy abruptly drops around 0.1 or less than 50% . Such a scenario is called **Over fitting.** Another scenario analogous to over fitting exists, .i.e. under fitting. In case of **under fitting**, the model becomes too simple to learn the complex features. Here the training accuracy and the validation accuracy, both reduce with a high training error. In such cases of varied training performance of deep learning models, dropout regularization is used. Basically, dropout regularization is a technique where we drop few nodes of specific layers in a neural network with a probability(p). For example if the value of $p = 0.5$, then we drop 50% of the nodes in a specific layer. If the value is $p = 0.2$, we drop 20% of the nodes in a layer. Commonly the values of the parameter p is chosen between 0.2 and 0.5. Nevertheless, it is necessary to experiment to find an optimal value for a specific task. A real time illustration of dropouts is shown in figure 3.22.

In this way, we can reduce over fitting and make the model to generalize to the unseen features in a better manner. In simple words, when we use dropouts, the model regularizes better to the unseen data. Where we retain a good training accuracy, as well as a good validation accuracy with less training error. Dropout regularization is used in the real time to maintain a balance

$$model = Sequential() \tag{3.8}$$
$$\#...ConvolutionalandPoolingLayers.. \tag{3.9}$$
$$model.add(Flatten()) \tag{3.10}$$
$$model.add(Dense(512, activation =' relu')) \tag{3.11}$$
$$model.add(Dropout(0.5))Dropoutlayer \tag{3.12}$$
$$withadropoutrateof0.5 \tag{3.13}$$
$$\#...OutputLayer... \tag{3.14}$$
$$\tag{3.15}$$

Figure 3.22: Implementation of dropouts

between over fitting and under fitting. Where it prevents the model from becoming overly specific to the training data, while having enough complexity to capture complex features. These dropouts are added to a neural network to the fully connected layers after the convolutional and pooling layers. These dropouts alter the whole training dynamics by randomly deactivating the neurons and introduces a noise during the process of training. This forces the neural network to learn more important features and generalize better to the unseen data. While the model generalization gets improved, the training time gets increased because there are fewer nodes after dropping out certain nodes in dropout regularization. Nevertheless, this extra time is justified by the improved generalization.

To further improve the model's performance after dropout regularization, we can adopt more techniques. Firstly, utilizing **ensemble learning.** Ensemble learning is a strategy in which we train multiple models/subnetworks and during inference the predictions of all these models are averaged to further enhance the performance of a model. In such a scenario, we can train multiple subnetworks/models by employing dropouts on each model, then averaging them during the inference. This methodology can further enhance the model generalization, contributing to an effective feature extraction. Secondly, combining other regularization techniques with dropout regularization. Like for example combining L1/L2 regularization can provide further better results. As we already know that the dropout regularization technique is quiet versatile with CNNs, it can also be applies to other neural networks. For example dropouts can also be used in RNNs for a better generalization, but the concept of application differs. Note that a clear indication of the dropouts can be seen in figure 3.19, where few nodes are switched off in between the network.

3.5 Dataset characteristics

A dataset may have several characteristics like size, dimensions, labelling, noise, ethical considerations and class imbalance. However one characteristic that is necessary to understand our experiments is class imbalance[THKG20]. Where the dataset may have huge imbalances with number of samples per identity with significantly more samples from some classes than others[THKG20]. For example : imagine four identities and the number of samples per identity is as below[THKG20]:

1. Identity A has 400 face images.
2. Identity B has 500 face images.

3. Identity C has 450 face images.
4. Identity D has 60 face images

In this scenario, Identity A,B and C are over represented. Whereas identity D is underrepresented[THKG20]. In such imbalanced datasets, the model gets biased towards the over represented classes in the process of training and does not learn the features from the under represented classes at all[THKG20]. In such cases data refinement techniques like resampling should be done, so that the class representations in a dataset becomes balanced[THKG20].

4 Methodology

The methodology chapter is structured to understand the experimental pipeline that was adopted in our experiments. Due to the lack of exploration of dropouts in FIQA using the SER-FIQ algorithm[TKD$^+$20], the focus is on closing this gap in the research. To close this gap, we define different model configurations of dropout regularization, so that an optimal configuration can be finalised. We also aim to investigate a wide variety of assumptions, so that unbiased results can be obtained. The goal of the present thesis is to setup experiments efficiently so that the effect of dropout regularization on FIQA is comprehensively analysed. Since the results are highly dependent on the experimental setup, this section transparently explains the proposed experimental setup with all the possible details. The chapter begins by introducing the dataset preparation and the motive behind our dataset choices. Subsequently the chapter delves into the details of the model selection and model training. The chapter concludes by the exploration of the testing phase, where the effectiveness of our approach is meticulously examined by the baseline performance evaluation. This evaluation is done with the help of the ROC curves.

4.1 Dataset preparation

Experiments were performed on two different datasets chosen on the variations in the quality of images in the datasets that were publicly available. Firstly, Labeled Faces in the Wild (LFW) dataset contains 12,966 face images from 5696 unique identities[HRBLM07]. Out of which, 6616 face images were utilized to train the on-top models(on-top models are explained in section 4.2). While the remaining 6350 face images were retained for testing[HRBLM07]. Further Youtube Faces dataset (YTF) which contained a total of 6,21,126 face images from 1595 unique identities[WHM11]. The total size of the dataset was 2.4 Gigabytes (GB). Out of 621126 face images, 9088 images from YTF were extracted randomly, thus clubbing with the 6350 face images from LFW dataset[WHM11], [HRBLM07]. Therefore the first dataset was manually prepared. **The objective of doing so, was to maintain wide variety of qualities in the first dataset, so that the exploration of dropouts on FIQA using SER-FIQ algorithm[TKD$^+$20] can be generalized.** The objective was entirely set to enhance the diversity of face image qualities to derive the robustness of SER-FIQ across different datasets and to have unbiased perspectives[TKD$^+$20]. This manually prepared dataset had a size of 15,438 face images, with varied image sizes, which was further passed through Multi-task Cascaded Convolutional Networks(MTCNN) framework[ZZLQ16] to detect faces and was stored separately for the testing purpose. It is important to note that even after applying MTCNN for

face detection, the image size of different face images varied a lot.

Secondly, another publicly available dataset was chosen. Pinsdataset[Bur20] which consists of 17,534 faces from 105 celebrities collected from Pinterest[Bur20]. **Not much research has been performed on this dataset till date as per the research documentation on google scholar website[Bur20].** This dataset had images of a high degree of image sharpness and image clarity with mostly high resolution face images[Bur20]. It consisted of already detected faces, hence there was no need to apply MTCNN[ZZLQ16] to detect faces again. The objective behind choosing this dataset for testing, was to experiment SER-FIQ only on specific high resolution images, which also had different face expressions and head poses[Bur20]. **Here the objective was entirely set to explore the effect of dropouts using SER-FIQ on the high image sharpness, high resolution images and imbalanced dataset characteristics.** Note that, dataset characteristics also contribute to the final observations.

4.2 Model selection and training

Model selection was based on how the models were pretrained with and without dropout regularization. The process of training was adopted for the on-top models built upon the selected models. Even in the on-top models we train different configurations with and without dropout regularization. The objective was to explore the effect of dropouts in the experimental setup on FIQA. This section exclusively explains the model selection, and training in a very transparent and detailed manner.

4.2.1 Model selection

The goal of the model selection was to select two base models that were already pretrained. **Two base models** were selected, 'where a pretrained model with dropouts was selected(ArcFace model[DGXZ19b], [SO21], [SO20])' and 'a pretrained model without dropouts was selected(OpenFace Model[ALS+16], [SO21], [SO20])'. The first model, **ArcFace model(Model pretrained with dropouts)** was picked from the DeepFace library[SO20], [SO21], [Ser21]. **The main objective of choosing ArcFace model** was because many scientific researches have proved the robust accuracy of this model provides because of its loss function[MWHN18], [DGXZ19b]. The second model, **OpenFace Model(Model pretrained without dropouts)** was also picked from the DeepFace library[SO20],[SO21], [Ser21] but was from the original paper [ALS+16] by Amos et. al. released in October 2015 under the Apache 2.0 license. **The main objective of choosing OpenFace model** was with an assumption that the improved triplet loss function could give some desirable results. Dropouts were the only criteria for the model selection. Both these models are introduced in the introduction chapter. Overall, these were the two models selected just based on a fact that whether the models were trained with dropouts or without dropouts.

4.2.2 Training

For the training purpose both the ArcFace model and the OpenFace model were used. Initially, embeddings were extracted from both the models and two on-top models were built upon each model, one on-top model with dropouts and another on-top model without dropouts. The architectures of the ontop models differ based on the pretrained model's input and output characteristics. In every case, there was one choice of the loss function to compile the on-top models .i.e. sparse categorical crossentropy(SCCE) loss function[TCERPCU23], [WMZT20].

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 512)               262656

 dense_1 (Dense)             (None, 128)               65664

 dense_2 (Dense)             (None, 512)               66048

 dense_3 (Dense)             (None, 512)               262656

 dense_4 (Dense)             (None, 5686)              2916918

=================================================================
Total params: 3573942 (13.63 MB)
Trainable params: 3573942 (13.63 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Figure 4.1: On-top model trained without dropouts on the ArcFace model. Input of the model is dependent on the 512 dimensional output vector from the ArcFace model.[TKD+20]

```
Model: "sequential_3"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense_15 (Dense)            (None, 512)               262656

 dropout_8 (Dropout)         (None, 512)               0

 dense_16 (Dense)            (None, 128)               65664

 dropout_9 (Dropout)         (None, 128)               0

 dense_17 (Dense)            (None, 512)               66048

 dropout_10 (Dropout)        (None, 512)               0

 dense_18 (Dense)            (None, 512)               262656

 dropout_11 (Dropout)        (None, 512)               0

 dense_19 (Dense)            (None, 5686)              2916918

=================================================================
Total params: 3573942 (13.63 MB)
Trainable params: 3573942 (13.63 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Figure 4.2: On-top model trained with dropouts on the ArcFace model. Input of the model is dependent on the 512 dimensional output vector from the ArcFace model.[TKD+20]

Initially we commenced our training process by using the pretrained model with dropouts .i.e. the **ArcFace model**. Embeddings were extracted from 6616 images that were intended for training, then an on-top model was trained without dropouts and another on-top model with dropouts. For the training of the**ontop model without dropouts**, following architecture was developed. The model had five layers each with dimensions of n_{emb} / 128 / 512 / n_{emb} / n_{ids} .

Each layer with a tanh activation function and the last layer with a softmax activation function. The loss function used to compile the model was sparse categorical crossentropy loss and an adam optimizer. The batch size of 1024 and 100 epochs. For the training of **ontop model with dropouts**, the architecture was similar, but dropout regularization (p = dropout rate) was applied to each layer n_{emb} (p=0.2) / 128(p=0.5) / 512(p=0.19999) / n_{emb}(p=0.22) / n_{ids}. While testing, the dropout rates were kept constant in all the layers.

Secondly, the training process was performed by using the pretrained **OpenFace model**[ALS+16]. Embeddings were extracted from the 6616 images that were intended for training, then an on-top model was trained without dropouts and another on-top model with dropouts. For the training of the**on-top model without dropouts**, following architecture was developed. The model had six layers each with dimensions of n_{emb} / 128 / 512 / n_{emb} / n_{emb} / n_{ids}. Each layer with a tanh activation function and the last layer with a softmax activation function. The loss function used to compile the model was sparse categorical crossentropy and an adam optimizer. The batch size of 32 and 100 epochs. For the training of the **on-top model with dropouts**, the architecture was similar, but dropout regularization (p = dropout rate) was applied only to the second last layer, where p=0.4. While testing, dropout regularization was applied to every layer[SHK+14], [Sri13]. [TKD+20] was the work referred while building the on-top models. While doing predictions from the on-top models, the last layer was removed and the last second layer was used to extract the features of the embeddings. Thus we are performing Transfer learning here.

```
Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 128)]             0

 dense (Dense)               (None, 128)               16512

 dense_1 (Dense)             (None, 512)               66048

 dense_2 (Dense)             (None, 128)               65664

 dense_3 (Dense)             (None, 128)               16512

 dense_4 (Dense)             (None, 5686)              733494

=================================================================
Total params: 898230 (3.43 MB)
Trainable params: 898230 (3.43 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Figure 4.3: On-top model trained without dropouts on the OpenFace model. Input of the model is dependent on the 128 dimensional output vector from the OpenFace model.[TKD+20]

```
Model: "model_9"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_13 (InputLayer)       [(None, 128)]             0

 dense_43 (Dense)            (None, 128)               16512

 dense_44 (Dense)            (None, 512)               66048

 dense_45 (Dense)            (None, 128)               65664

 dense_46 (Dense)            (None, 128)               16512

 dropout_18 (Dropout)        (None, 128)               0

 dense_47 (Dense)            (None, 5686)              733494

=================================================================
Total params: 898230 (3.43 MB)
Trainable params: 898230 (3.43 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Figure 4.4: On-top model trained with dropouts on the OpenFace model. Input of the model is dependent on the 128 dimensional output vector from the OpenFace model.[TKD+20]

4.3 Use cases and baseline performance evaluation

4.3.1 Use cases

Our experiments involved a systematic framing of the use cases to explore the effect of different configurations on the performance of the face image quality analysis. These cases were carefully structured to provide an in-depth understanding of how varying the dropout regularization in both pretrained models and on-top models would influence the overall recognition performance. The following text indicates the different use cases that were framed for a better understanding.

The initial use case (Case 1), involved the usage of pretrained model that is trained with dropouts. The objective of this case was to investigate the effects of the model on the feature extraction and the classification of face images. Here we decided to use the ArcFace model because of its exceptional robustness. The subsequent use case (Case 2), involved the utility of the same pretrained model that is trained with dropout regularization, but additionally with an on-top model trained without the usage of dropout regularization. This use case aimed to evaluate the model's ability to extract features and classify images in the absence of dropouts in the on-top model. The next use case (Case 3), involved a pretrained model trained with dropout regularization and an on-top model trained with dropout regularization. In this use case the aim was to investigate the overall performance when there is a dual application of dropouts in both the models.

The next three use cases involved the usage of the pretrained model that is trained without the use of dropout regularization. The following case (Case 4), involved the usage of the pretrained model that is not trained with dropout regularization. Here we decided to use the OpenFace model that was picked up from the DeepFace library. The objective was to compare its performance with other cases and examine if this use case could do better. The next use case (Case 5), involved the use of the pretrained that is not trained with dropout regularization, and additionally an on-top model that is not trained with dropout regularization. This use case was to

investigate the performance in the total absence of dropouts. In the subsequent case (Case 6), the pretrained model that is not trained with dropout regularization was used with an on-top model that is trained with dropout regularization. This use case explored the performance due to the usage of dropout regularization solely in the on-top model.

According to the quality score calculation in the SER-FIQ algorithm, we know that if there is a high variation in the stochastic embeddings, it means that the quality score will be low and vice versa[TKD+20]. Therefore each case was performed on both the datasets prepared with two assumptions. **Firstly, assuming that the lowest quality scores are assigned to the lowest quality images. Secondly, assuming that the highest quality scores are assigned to the lowest quality images. All the experiments were performed to get the unbiased results, to get deeper insights**. Figure 4.5 is referenced in the whole thesis wherever these assumptions are needed.

Assumption 1:{{Assuming that the lowest quality scores are assigned to the lowest quality images.}},
*Assumption*2 : {{Assuming that the highest quality scores are assigned to the lowest quality images.}}

Figure 4.5: Assumptions for the experiments. Where lowest quality score means that the variations between the stochastic embeddings of a sample is high. While highest quality score means that the variations between the stochastic embeddings of a sample is low.

4.3.2 Baseline Performance evaluation

The performance evaluation was performed after training the on-top models. For this evaluation, ROC curves were computed for every use case. Initially, embeddings were extracted by using all the use cases, then genuine and imposter scores were computed. Ratio of genuine to imposter scores was 1657:16162455. Even though we obtained optimal performance in ROC curves, during the computation of the error vs reject curves, **a huge sampling bias was adopted, which significantly affected our results**. ROC curves were computed for both the datasets in each use case. Then the same embeddings were passed through the on-top models to check if it works. Surprisingly, all the models probably worked by observing the ROC performance. But there were some unusual observations.

In case 3, the AUC score for dataset was 0.91. Whereas for dataset 2, the AUC score was approximately equal to 1. These unusual observations might be due to the imbalance between the bias and variance in case 3 and the different dataset characteristics. Also,While, in case 4, the AUC score for dataset 1 was 0.95. Whereas for dataset 2, the AUC score was approximately equal to 1. The unusual observations in case 4 might be due to the reason that the OpenFace model was a more generalised one and due to its loss function(Improved triplet loss function). The reasons for different observations in different datasets might be due to the fact that the first dataset had a variety of face image qualities and more anomalies. Whereas the second dataset had high resolution images with huge class imbalance. Figure 4.8, 4.11, 4.14, 4.17, 4.20 and 4.23 shows the ROC plots for all the use cases.

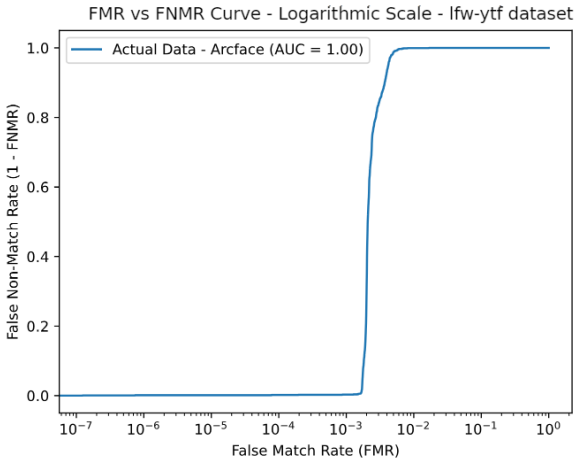

Figure 4.6

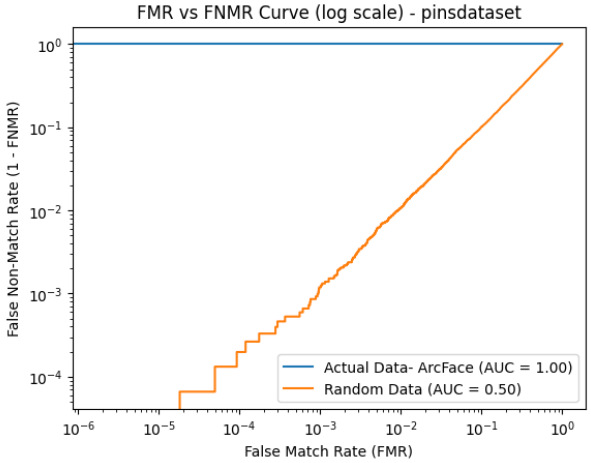

Figure 4.7

Figure 4.8: ROC curves for the ArcFace model(Case 1:Pretrained model with dropouts), for both datasets.

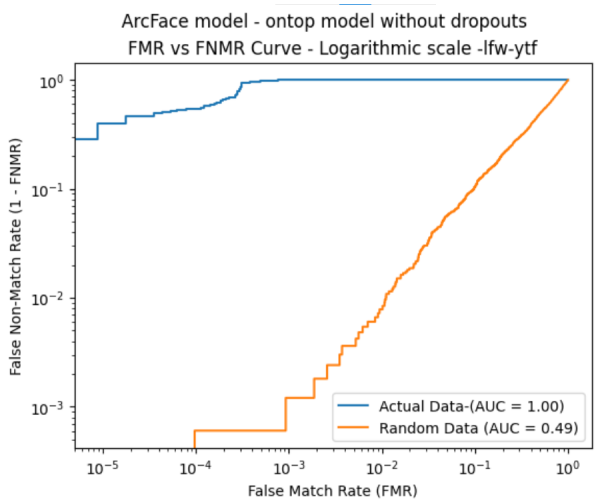

Figure 4.9

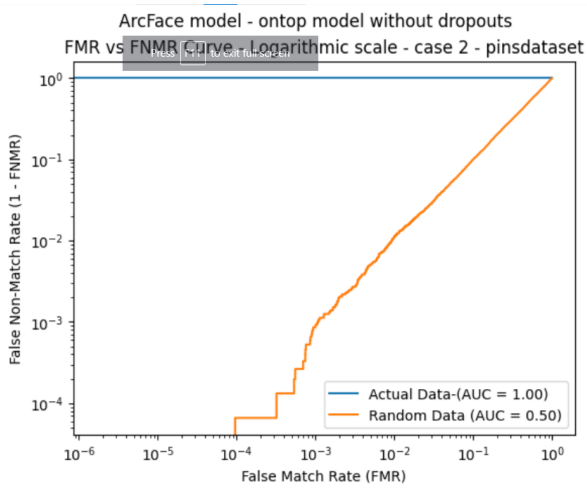

Figure 4.10

Figure 4.11: ROC curves for the ArcFace model with an on-top model trained without dropouts(Case 2:Pretrained model with dropouts - with an on-top model trained without dropouts) for both datasets.

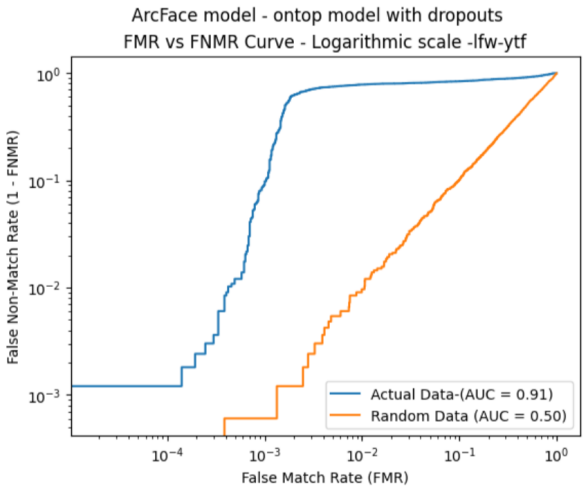

Figure 4.12

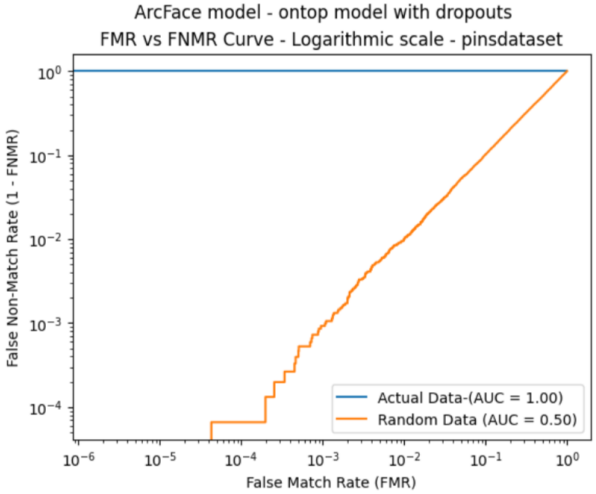

Figure 4.13

Figure 4.14: ROC curves for the ArcFace model with an on-top model trained with dropouts(Case 3:Pretrained model with dropouts - with an on-top model trained with dropouts), for both datasets.

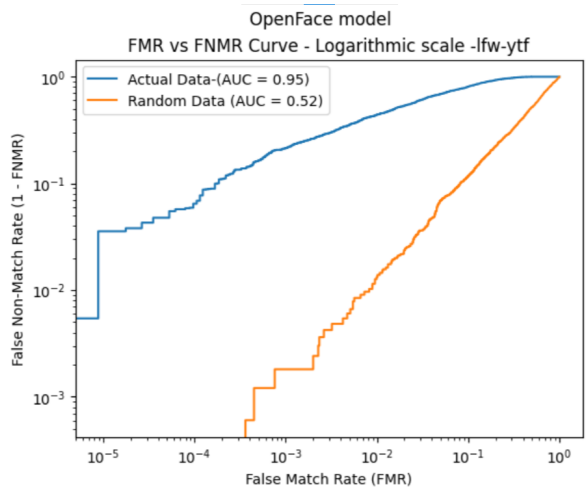

Figure 4.15

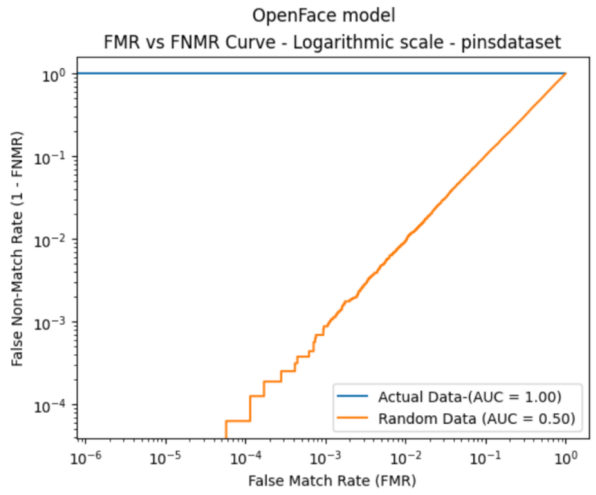

Figure 4.16

Figure 4.17: ROC curves for the OpenFace Model(Case 4:Pretrained model without dropouts), for both datasets.

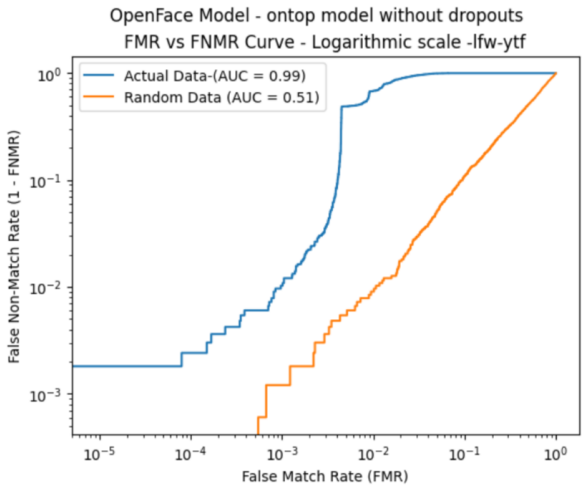

Figure 4.18

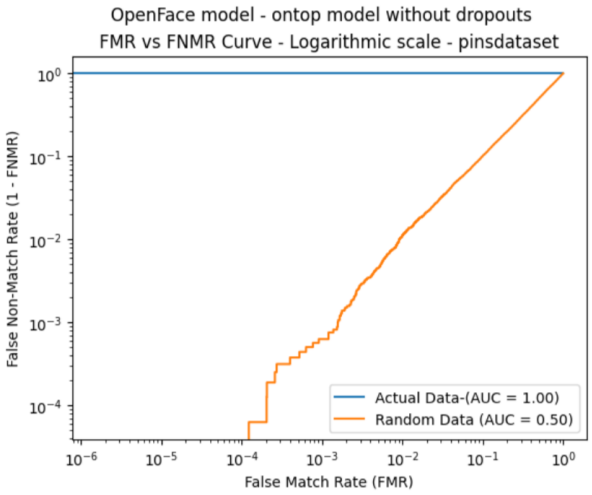

Figure 4.19

Figure 4.20: ROC curves for the OpenFace Model with an on-top model trained without dropouts(Case 5:Pretrained model without dropouts - on-top model trained without dropouts), for both datasets.

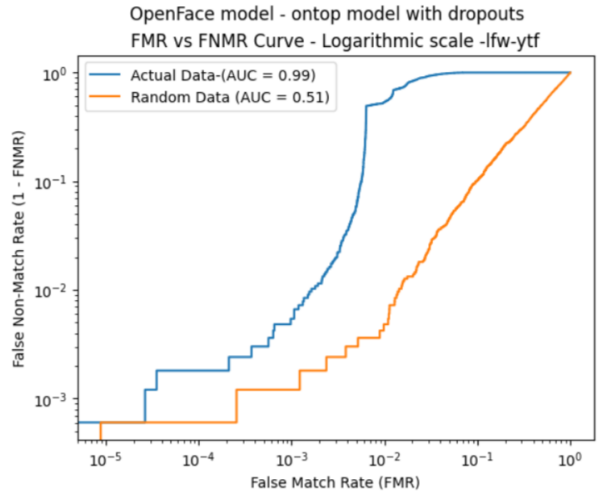

Figure 4.21

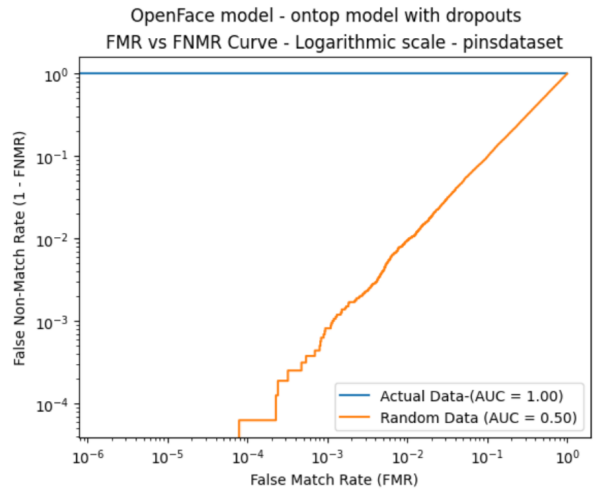

Figure 4.22

Figure 4.23: ROC curves for the OpenFace Model with an on-top model trained with dropouts(Case 6:Pretrained model without dropouts - on-top model trained with dropouts), for both datasets.

4.4 Testing

4.4.1 Quality scores computation

In the testing phase, all the six predefined use cases 4.3.1 were considered. Both the datasets were passed through the models in every use case. Note that the testing done on cross-dataset(dataset 1) with a mixture of two different datasets might probably provide valuable insights on how the models can generalise to unseen faces. Hence it is crucial to explore the process of testing for

different use cases with distinct datasets.

In the **initial case (Case 1)** with the pretrained model that is trained with dropouts, we have an ArcFace model trained with dropouts. Here the dropouts which were already present, were re-used for testing, where the last fourth layer had dropout regularization applied to it. We preprocessed the images from both the datasets to an input size of the model (112*112). We then passed the preprocessed images from the datasets through the model, while the dropout patterns were varied, to collect different stochastic embeddings in each forward pass. Subsequently, for each image in the dataset, a quality score was computed. In the **subsequent case (Case 2)**, the last layer from the on-top model was removed, so that we can extract the feature vectors(Embeddings) from the second last layer. Even though the dropout regularization was absent during the training of the on-top model, the dropouts were applied during the testing phase. A similar procedure was followed as in case 1. Where we varied the dropouts in the ArcFace model. Then passed the embedding obtained from the ArcFace model through the on-top model. In this way, we collected different stochastic embeddings from each forward pass. Subsequently, for each image in the dataset, a quality score was computed. A similar procedure was followed in the **next case (case 3)**, but the difference is that dropout regularization was present during the training phase and also during the testing phase, the dropouts were present.

In the next three cases the OpenFace model was used. In the **following case (Case 4)**, OpenFace model which was pretrained without dropouts was chosen[SO21],[ALS+16]. During the testing phase, dropouts were applied only to the last layer. We preprocessed the images from both the datasets to an input size of the model (96*96). The preprocessed images from the datasets were passed through the OpenFace model, while the dropout patterns were varied, to collect different stochastic embeddings in each pass. Subsequently, for each image in the dataset, a quality score was computed. In the **next cases (Case 5) and (Case 6)**, similar procedure was followed as in the previous cases. The difference is that, in the first three use cases we had an ArcFace model which had an output dimension of 512 for the feature vectors, while the extraction of the feature vectors from the OpenFace model had a dimension of 128. Thus every use case followed the SER-FIQ algorithm. Where we adopted the value of m (.i.e the number of forward passes) according to our system complexity and normalized the final scores. 1.

4.4.2 Error vs reject curves for our experiments.

Following the computation of the quality scores in the predefined use cases, error vs reject curves were computed to evaluate the performance of a face recognition system. Initially embeddings are loaded from the saved pickle dictionary format[Fas18]. In this dictionary, identity names are considered as keys, and the associated embeddings with quality scores are stores in the values. For example, assume that we have six samples of the identity named AJ cook, four samples of the identity named Yasmin and fifty samples of an identity named Rita. Where each face image sample in a dataset is associated with its identity name, embedding and the corresponding quality score. The stored dictionary format is illustrated in 4.24.

$$
\begin{aligned}
AJ\ \text{cook:}\{&\{\text{embedding1, quality\_score1}\},\\
&\{\text{embedding2, quality\_score2}\},\\
&\{\text{embedding3, quality\_score3}\},\\
&\{\text{embedding4, quality\_score4}\},\\
&\{\text{embedding5, quality\_score5}\},\\
&\{\text{embedding6, quality\_score6}\}\},\\
Yasmin : \{&\{\text{embedding1, quality\_score1}\},\\
&\{\text{embedding2, quality\_score2}\},\\
&\{\text{embedding3, quality\_score3}\},\\
&\{\text{embedding4, quality\_score4}\}\},\\
Rita : \{&\{\text{embedding1, quality\_score1}\},\\
&\{\text{embedding2, quality\_score2}\},\\
&\{\text{embedding3, quality\_score3}\},\\
&\{\text{embedding4, quality\_score4}\},\\
&\ldots,\\
&\{\text{embedding20, quality\_score20}\}\}
\end{aligned}
$$

Figure 4.24: Dictionary format of the embeddings.

Subsequent to loading the data, the quality scores are sorted. This sorting order matters a lot. **Due to these sorting orders, when lowest quality scores are assigned to the lowest quality images , we call this scenario as 'ascending quality plots' and when the highest quality scores are assigned to the lowest quality images it is called as 'descending quality plots'.** They are called ascending quality plots and descending quality plots because during the computation of the error vs reject curves we sort the quality scores in ascending order in the case of ascending quality plots and we sort the scores in the descending order in the latter case. Then by iteratively rejecting the lowest quality samples, fnmr values at different fmr thresholds $fnmr@0.01fmr$, $fnmr@0.001fmr$, $fnmr@0.0001fmr$ and $EER(euqalerrorrates)$ in each iteration were computed to plot the error rates against the ratio of unconsidered images.

Before we discuss the observations, it is crucial to know the system configuration. The system had a processor of Intel(R) Core(TM) i5-6200U CPU @2.30GHz-2.40GHz, installed RAM 8.00 GB (7.89 GB usable) and a RAM of 64-bit operating system, x64-based processor. **To adapt to the system's capability, the number of imposter scores were limited**. In brief, the number of imposter pairs are computed by considering the unique identities and for each identity, yet another identity is selected to generate an imposter pair. In this way, the system was not able to generate all the possible pairs, **so for each identity, a random subset of five samples from different identities were chosen from the prepared dataset to compute imposter scores**. This approach was adopted because it balances the computational efficiency of the system and the representation of the imposter scores. Hence these steps provide a comprehensive evaluation of the recognition performance, while maintaining the computational efficiency.

Succeeding the methodology, it is crucial to understand, discuss and analyse the observations. In the following section, the final results in the error vs reject curves are analysed. Parallel to

this analysis, related causes, intended actions and the implications are provided. Further only scientific reasons were provided in each case.

5 Discussion

According to the standards in FIQA, the error vs reject curves that decrease abruptly indicates an ideal face recognition performance[TKD$^+$20]. **Before delving into the discussion, it is important to revisit the work done under SER-FIQ[TKD$^+$20], which concluded that the images that exhibit large variations in the stochastic embeddings were regarded as low quality images, whereas the images that exhibit small variations in their stochastic embeddings were regarded as high quality images. Which basically means that the images that have low quality scores are regarded as low quality images and the images that have high quality scores are regarded as high quality images.** However, our experiments were performed to get broader perspectives on these conclusions, where all the results were not as expected.

While looking into our observations, it is important to note that in each use case for every dataset, the error versus reject curves were plotted with two different assumptions. Firstly, assuming that low quality scores are assigned to lowest quality images, thus the scores were sorted in the ascending order and then the curves were computed, hence called ascending quality plots. Secondly, assuming that the high quality scores are assigned to the lowest quality images, thus the scores were sorted in the descending order, hence called descending quality plots. In total four different curves were plotted namely, $fnmr@0.01fmr$ vs ratio of unconsidered images, $fnmr@0.001fmr$ vs ratio of unconsidered images, $fnmr@0.0001fmr$ vs ratio of unconsidered images, $Equalerrorrate$ vs ratio of unconsidered images. These four different curves were plotted for both ascending and descending quality plots, for every model configuration and for each dataset. The results of our experiments provided valuable insights into the performance of the different configurations of the models in different use cases.

This section provides a detailed discussion of our observations. Firstly the results are stated, illustrated, tabulated and interpreted. Then the findings are analysed, where we analyse the factors that might influence our results. Later we compare and contrast different use cases and concluded that when the dropout regularization, if incorporated in both the pretrained model and on-top model, the results become optimal. We also analysed the limitations and strengths of our experiments with the future work that can be done.

5.1 Results

Manually prepared dataset

There were few noteworthy results in case of the manually prepared dataset. When only the **ArcFace model** trained with dropout regularization was employed[SHK+14], [Sri13], there was an increasing trend in the error vs reject curves, **which was unusual**. But, when the same ArcFace model was employed with the on-top models trained with and without dropout regularization[SHK+14], [Sri13], the curves were surprisingly decreasing in the error vs reject plots.**These noteworthy results perfectly aligned with the standards outlined in the performance evaluation of the error vs reject curves in face image quality assessment(FIQA)**. When the **OpenFace model** trained without dropout regularization was employed, we observed decreasing trends in the error vs reject curves. When the on-top models trained with and without dropout regularization, were built upon this same OpenFace model[ALS+16], we observed the similar decreasing trends in the error vs reject curves. **Even these results perfectly aligned with the standard performance of the FIQA**.

Despite the consistency in the observed trends, a deviation was observed in the curve 'FNMR@0.0001FMR VS ratio of unconsidered images'. Where the FMR threshold becomes too low and the curves were not as expected. This indicates that the system becomes too sensitive in extreme conditions. **One major limitation was that there existed an ambiguity between the ascending and descending quality plots, which mirrored from the similar trends which were observed in both the cases**. For further insights into the ascending and descending quality plots, please refer to the figure 4.5. While plotting the curves we also observed that the dataset becomes really unstable when a small amount of data is left after rejecting more than 80% of the low quality data. The upcoming sections provides an in depth exploration behind the reasons for these results.

Pinsdataset

The results in case of the pins dataset were entirely unusual. When the ArcFace model trained with dropout regularization was solely utilized, we observed that every curve had an horizontal straight line. But when the same ArcFace model was employed with the on-top models with and without dropout regularization, the trends in the error vs reject curves were different. The curves 'FNMR@0.001 vs Ratio of Unconsidered Images' and 'FNMR@0.0001 vs Ratio of Unconsidered Images' showed 100% errors with their horizontal lines at 1.0. The other two curves 'FNMR@0.01 vs Ratio of Unconsidered Images' and 'Equal error rate vs Ratio of Unconsidered Images' become entirely random.

When the OpenFace model trained without dropout regularization[SHK+14], [Sri13] was solely employed, the results were similar. The curves 'FNMR@0.001 vs Ratio of Unconsidered Images' and 'FNMR@0.0001 vs Ratio of Unconsidered Images' showed 100% errors with their horizontal lines at 1.0. The other two curves 'FNMR@0.01 vs Ratio of Unconsidered Images' and 'Equal error rate vs Ratio of Unconsidered Images' become entirely random. But when the same OpenFace model[ALS+16] is employed with the on-top models trained with and without dropouts, the curves 'FNMR@0.001 vs Ratio of Unconsidered Images' and 'FNMR@0.0001 vs Ratio of Unconsidered Images' showed 100% errors with their horizontal lines at 1.0. The other two curves 'FNMR@0.01 vs Ratio of Unconsidered Images' and 'Equal error rate vs Ratio of Unconsidered Images' were generally decreasing. Despite the unusual results in this dataset, the ambiguity between the ascending and descending plots still persisted. **The similar trends in both ascending and descending quality plots clearly indicates that sorting the**

quality scores is not resonating with the actual distinction between the genuine and imposter pairs.

All the plots from our experiments are shown in figures casewise : case 1 : 5.2, 5.3, 5.4, 5.5, case 2 : 5.6, 5.7, 5.8, 5.9, Case 3 : 5.10, 5.11, 5.12, 5.13, Case 4 : 5.14, 5.15, 5.16, 5.17, 5.18, Case 5 : 5.19, 5.20, 5.21, Case 6 : 5.22, 5.23, 5.24, 5.25. Further we have documented all the results in the table 5.1 for the manually prepared dataset and in the table 5.2. Please refer to the methodology section and the tables for more details on the cases. Some samples with their quality scores are represented in the figure 5.1. Overall the results, offer valuable insights into the performances of different model configurations. Which emphasizes the importance of considering various factors in FIQA. While the results were stated and briefly interpreted in this section, it is further important to gain deeper insights on what, why and how behind these results.

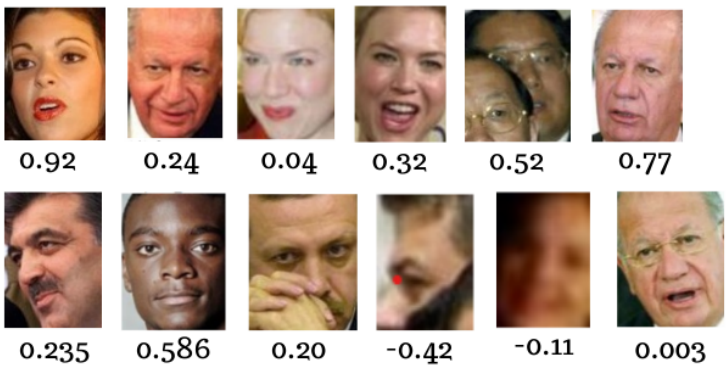

Figure 5.1: Quality scores of different samples. It is evident, how the quality score varies with face occlusions, age and different face expressions.[WHM11], [HRBLM07]

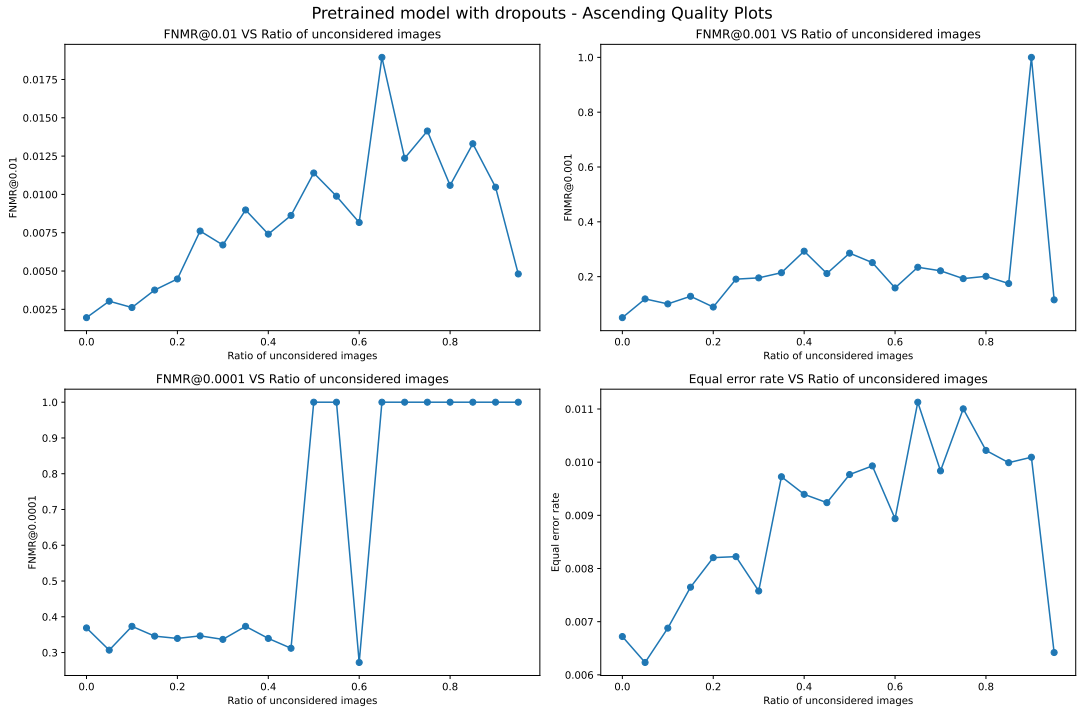

Figure 5.2: Case 1 plots for dataset 1(manually prepared dataset), with an assumption that lowest quality scores are assigned to lowest quality images.

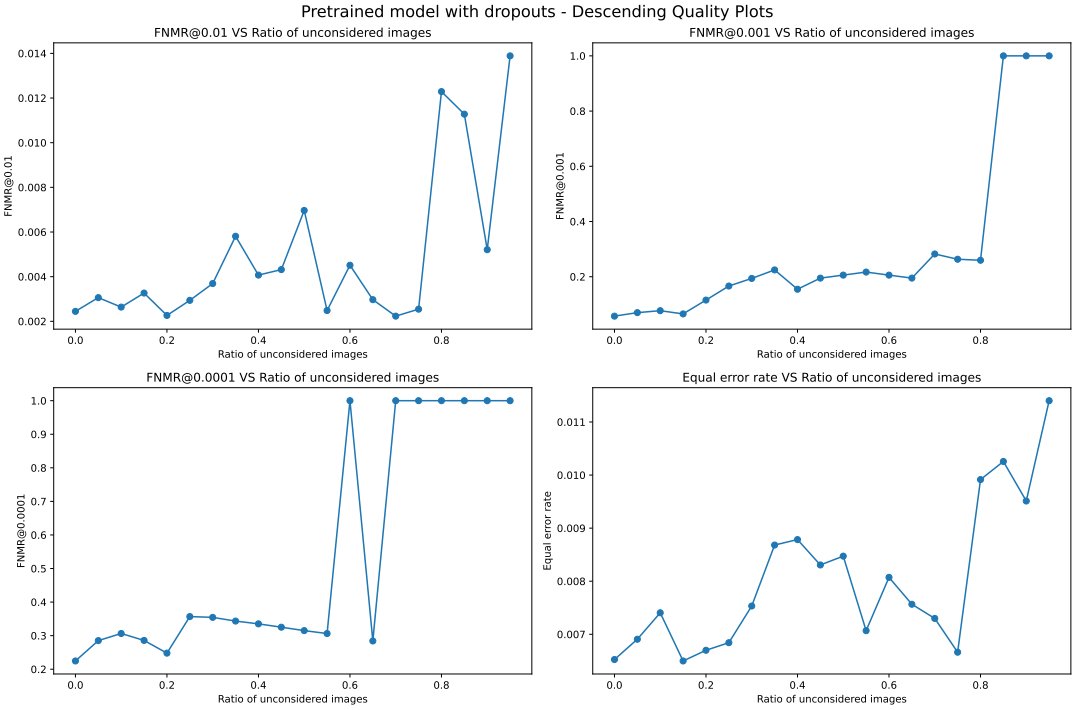

Figure 5.3: Case 1 plots for dataset 1(manually prepared dataset), with an assumption that highest quality scores are assigned to lowest quality images.

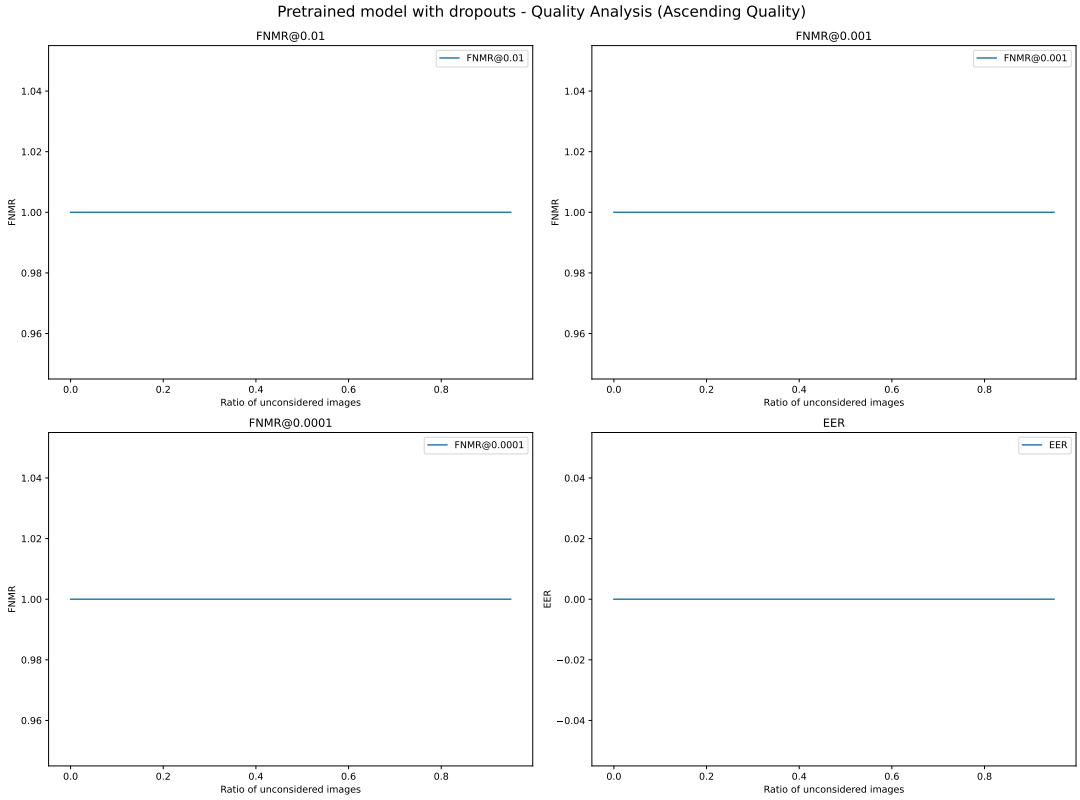

Figure 5.4: Case 1 plots for (pinsdataset)dataset 2, with an assumption that lowest quality scores are assigned to lowest quality images.

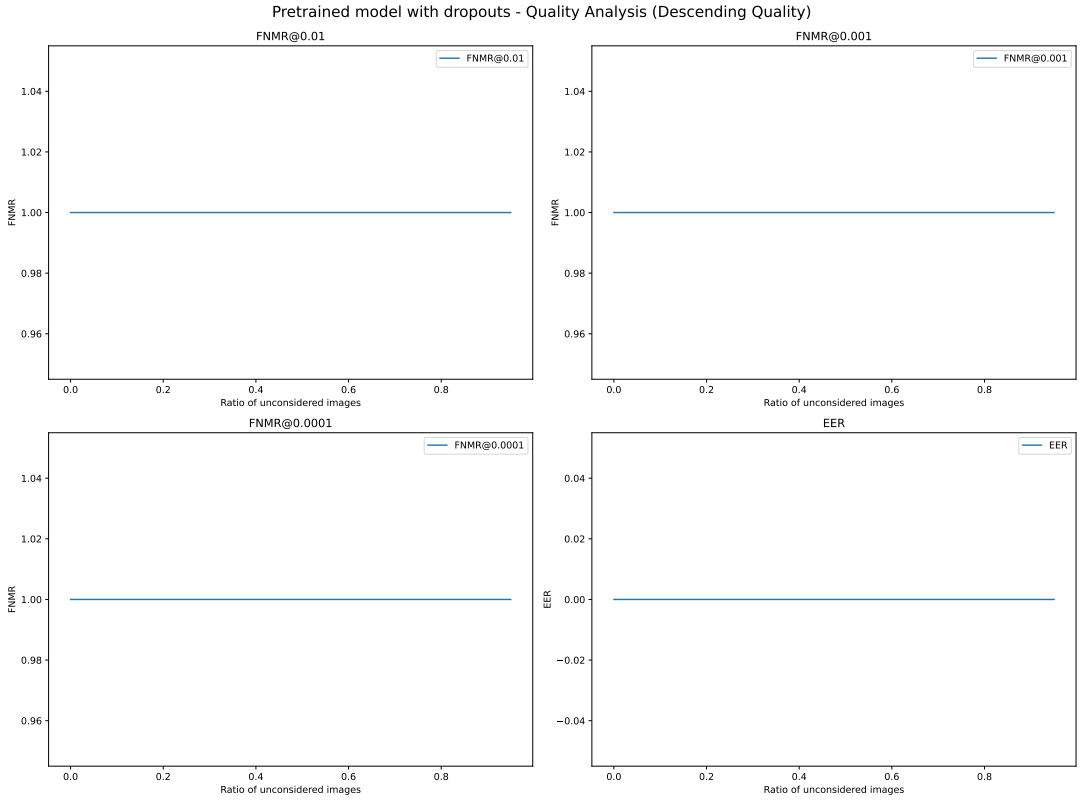

Figure 5.5: Case 1 plots for (pinsdataset)dataset 2, with an assumption that highest quality scores are assigned to lowest quality images.

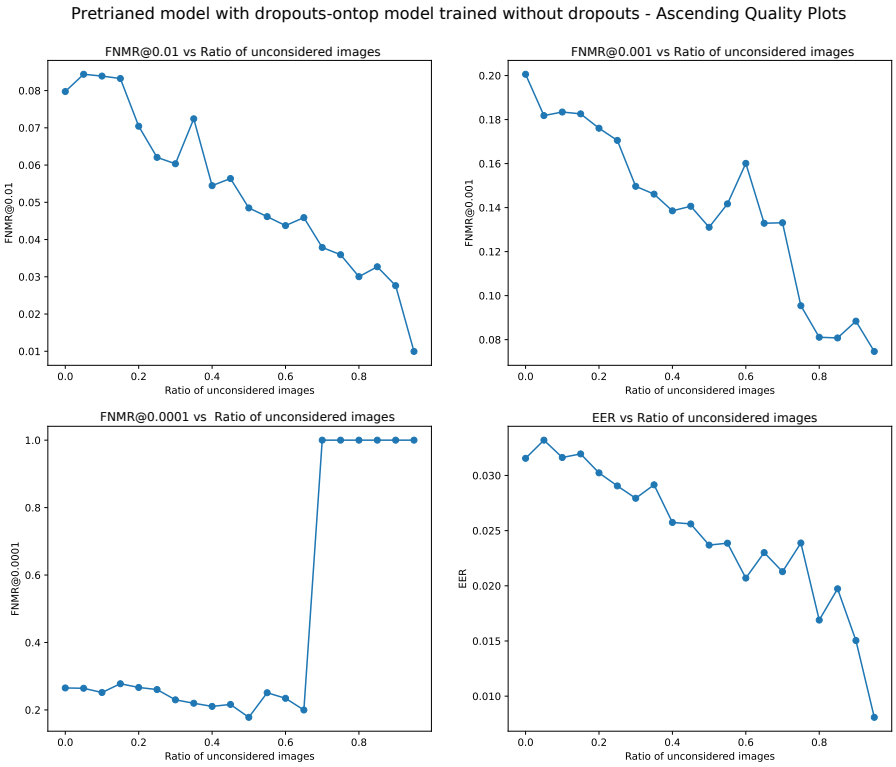

Figure 5.6: Case 2 plots for dataset 1(manually prepared dataset), with an assumption that lowest quality scores are assigned to lowest quality images.

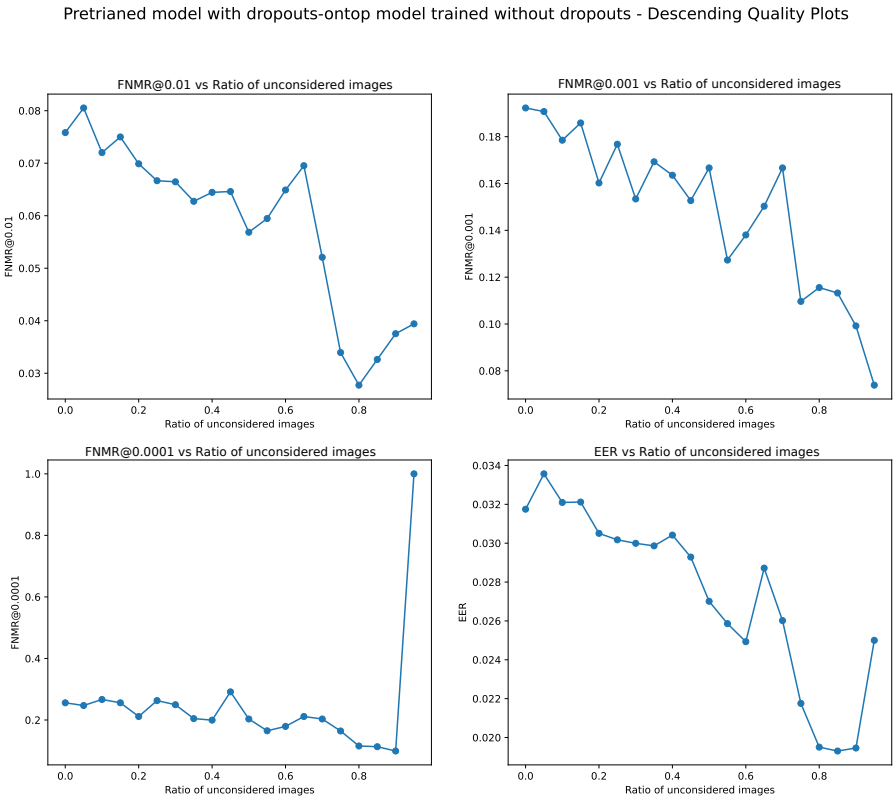

Figure 5.7: Case 2 plots for dataset 1(manually prepared dataset), with an assumption that highest quality scores are assigned to lowest quality images.

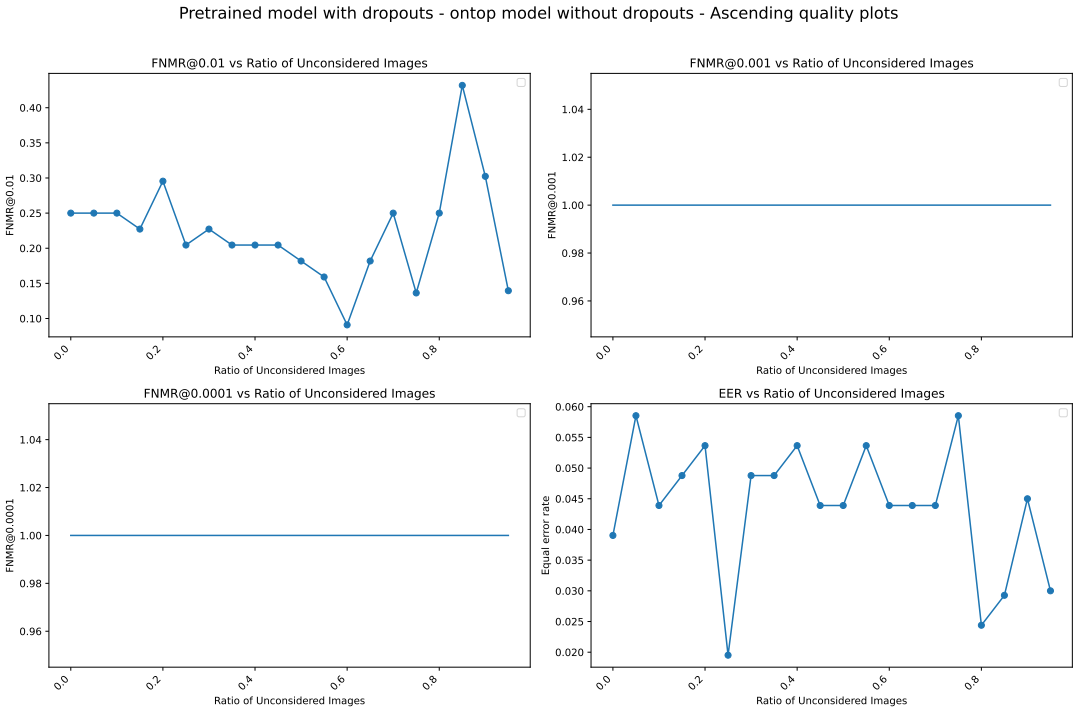

Figure 5.8: Case 2 plots for (pinsdataset)dataset 2, with an assumption that lowest quality scores are assigned to lowest quality images.

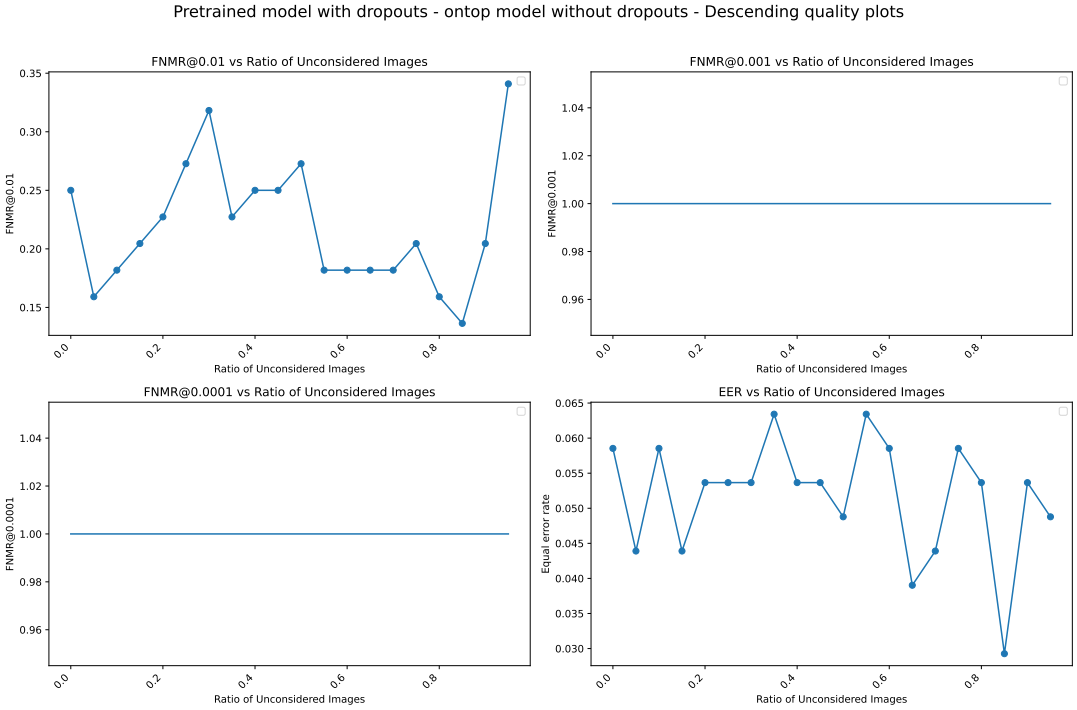

Figure 5.9: Case 2 plots for (pinsdataset)dataset 2, with an assumption that highest quality scores are assigned to lowest quality images.

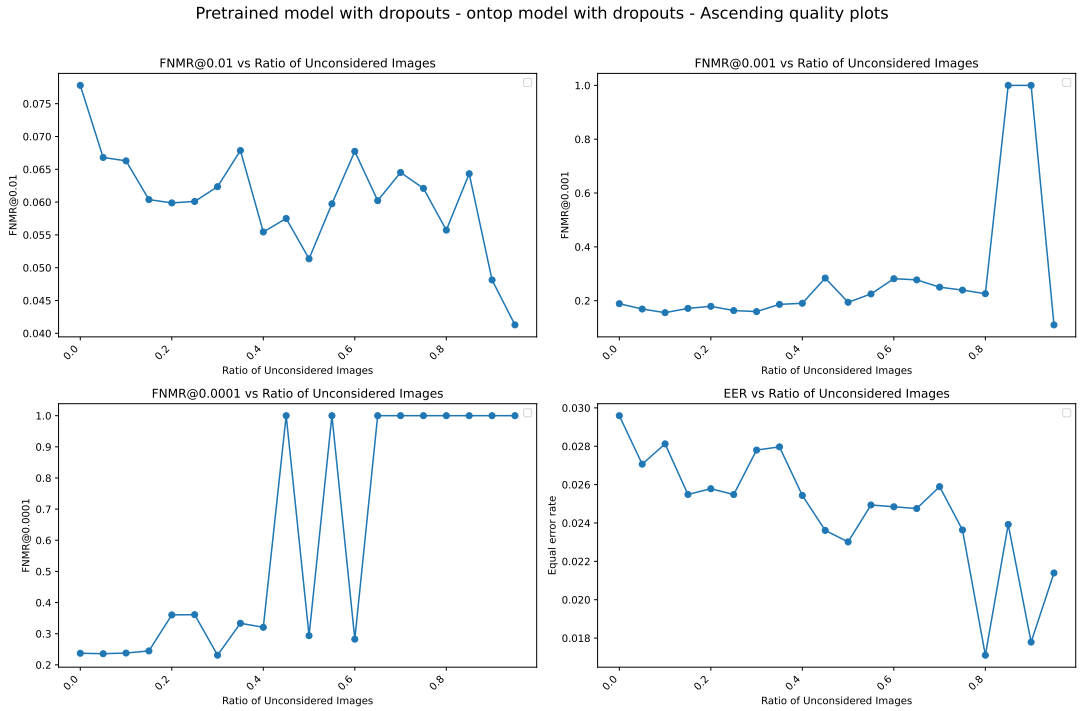

Figure 5.10: Case 3 plots for dataset 1(manually prepared dataset), with an assumption that lowest quality scores are assigned to lowest quality images.

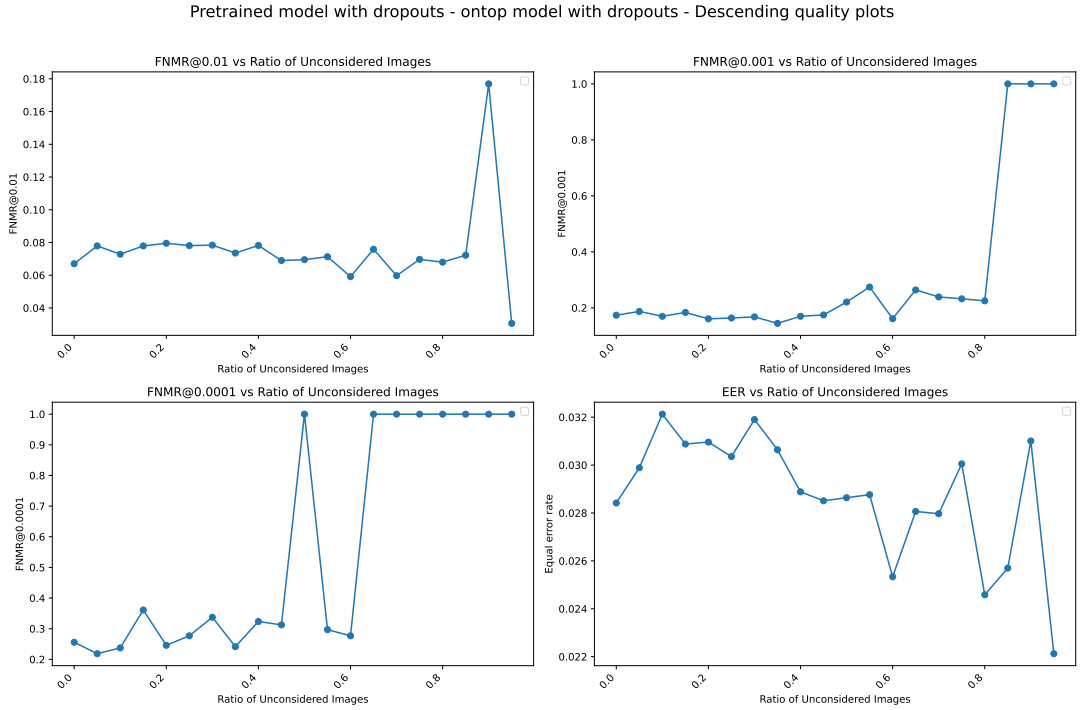

Figure 5.11: Case 3 plots for dataset 1(manually prepared dataset), with an assumption that highest quality scores are assigned to lowest quality images.

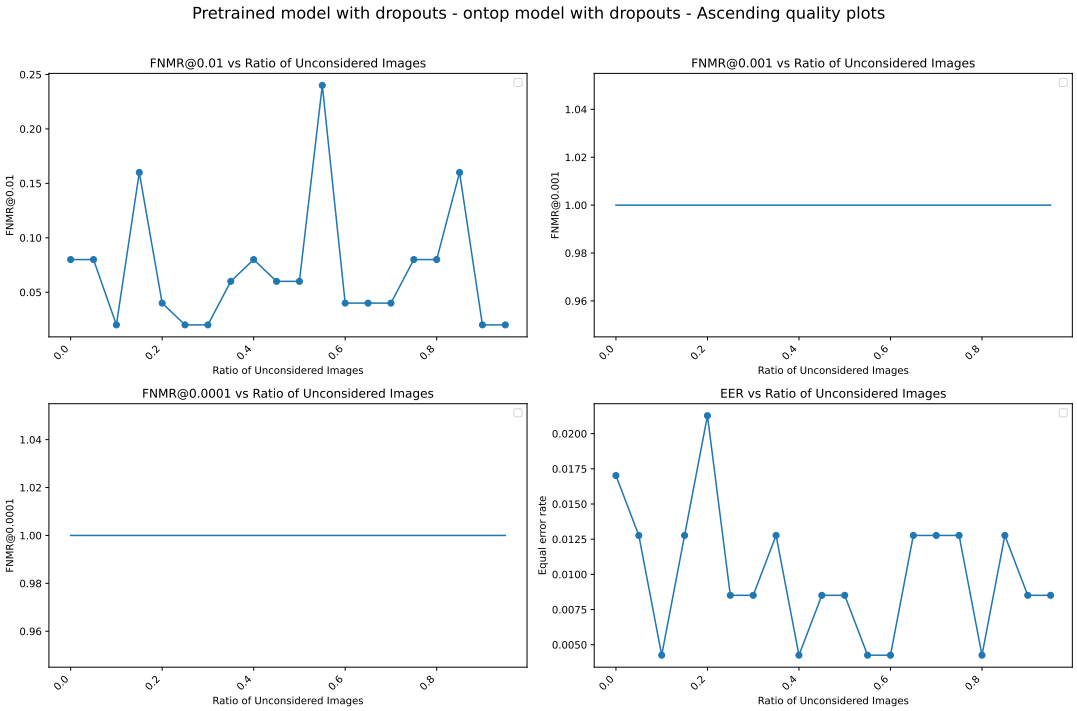

Figure 5.12: Case 3 plots for (pinsdataset)dataset 2, with an assumption that lowest quality scores are assigned to lowest quality images.

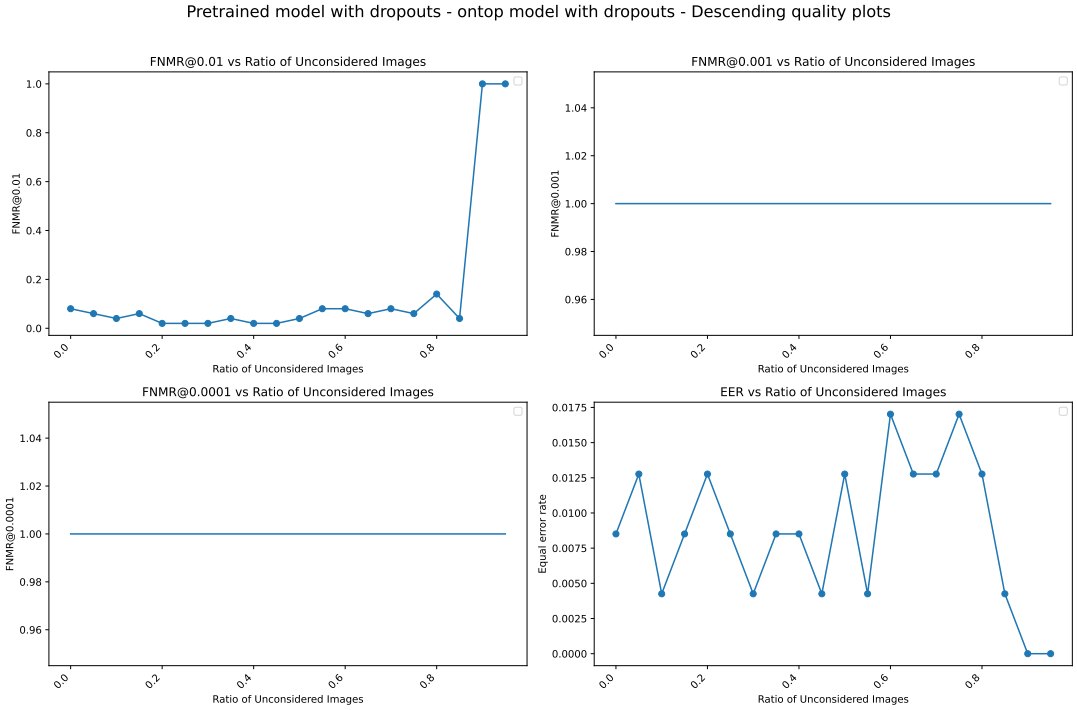

Figure 5.13: Case 3 plots for (pinsdataset)dataset 2, with an assumption that highest quality scores are assigned to lowest quality images.

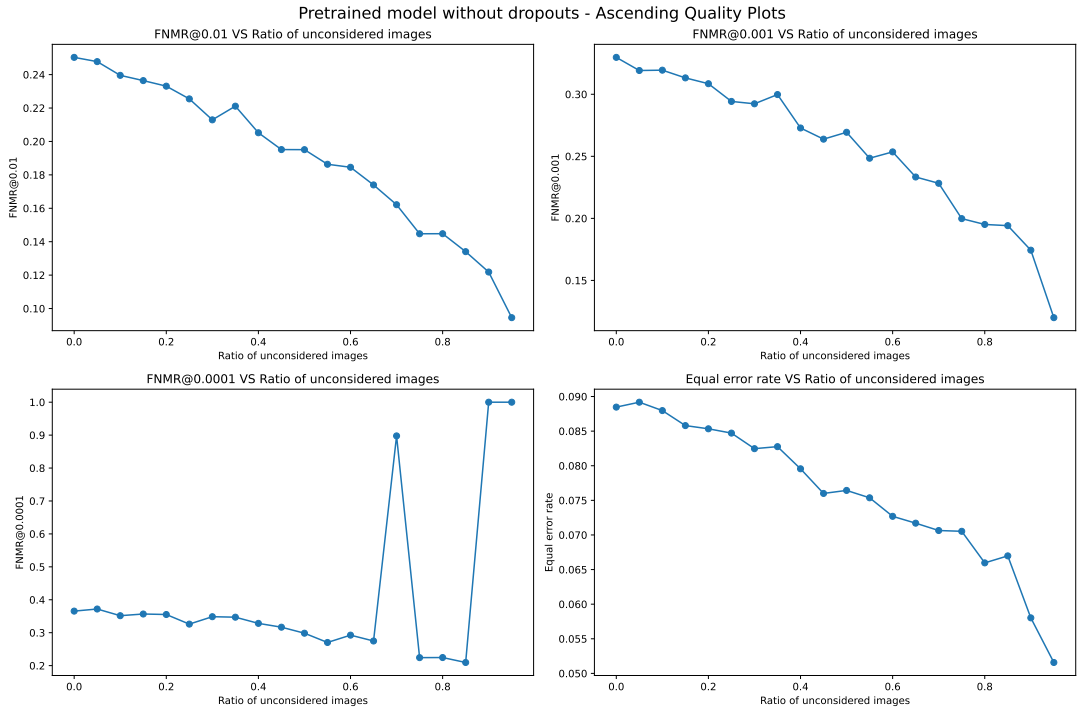

Figure 5.14: Case 4 plots for dataset 1(manually prepared dataset), with an assumption that lowest quality scores are assigned to lowest quality images.

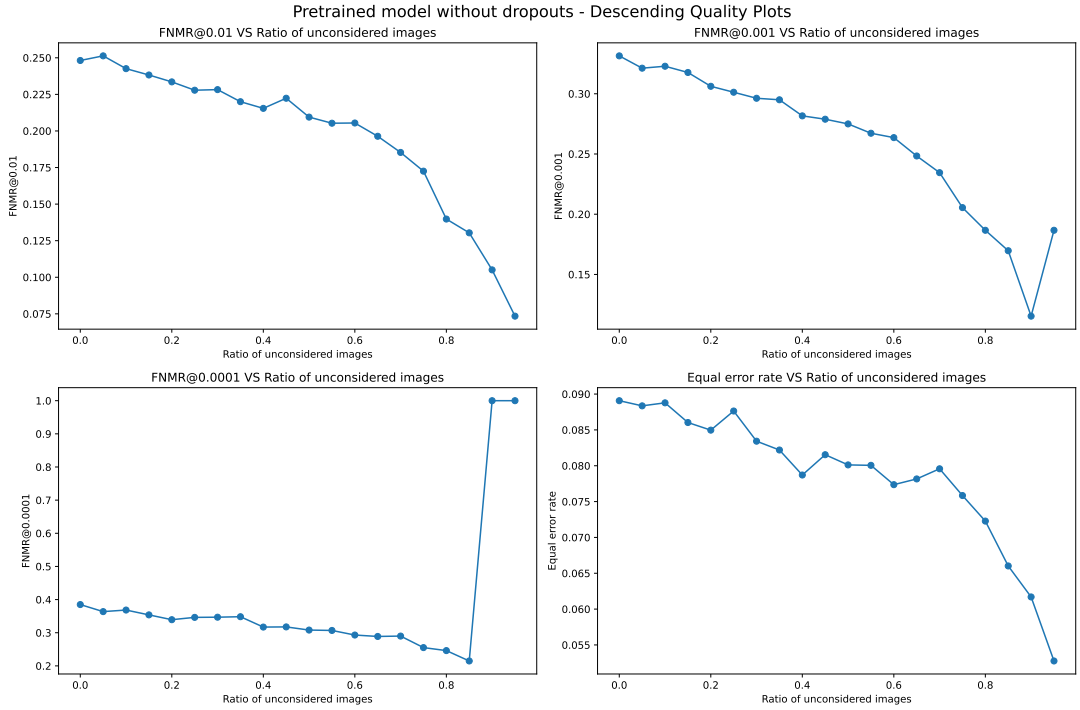

Figure 5.15: Case 4 plots for dataset 1(manually prepared dataset), with an assumption that highest quality scores are assigned to lowest quality images.

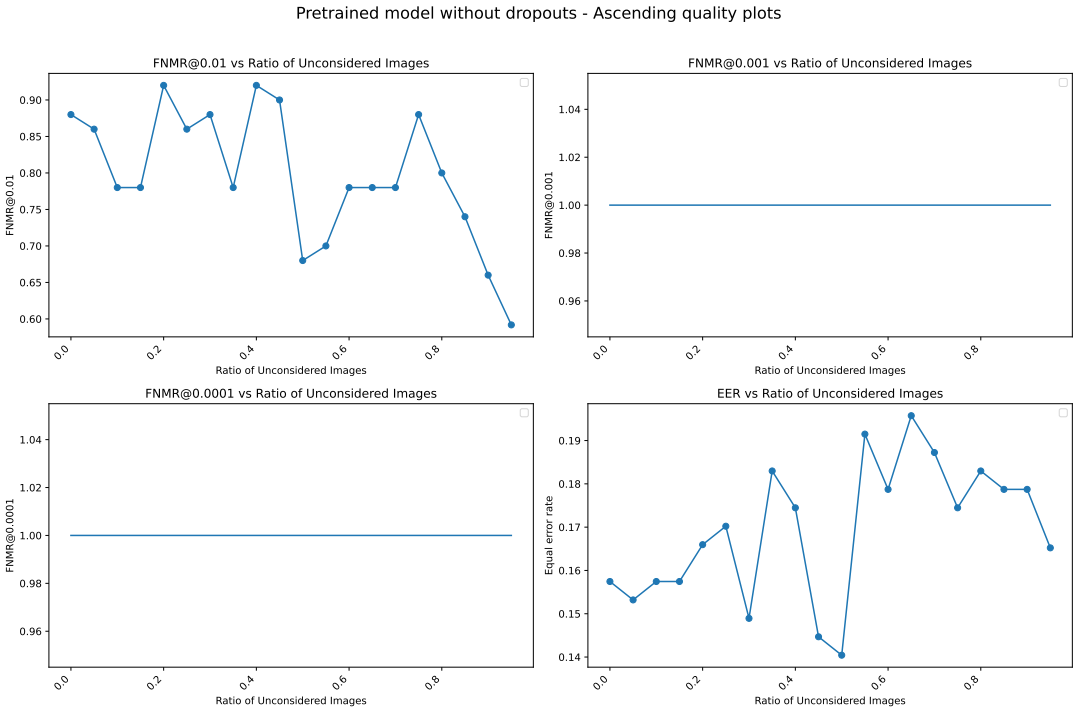

Figure 5.16: Case 4 plots for (pinsdataset)dataset 2, with an assumption that lowest quality scores are assigned to lowest quality images.

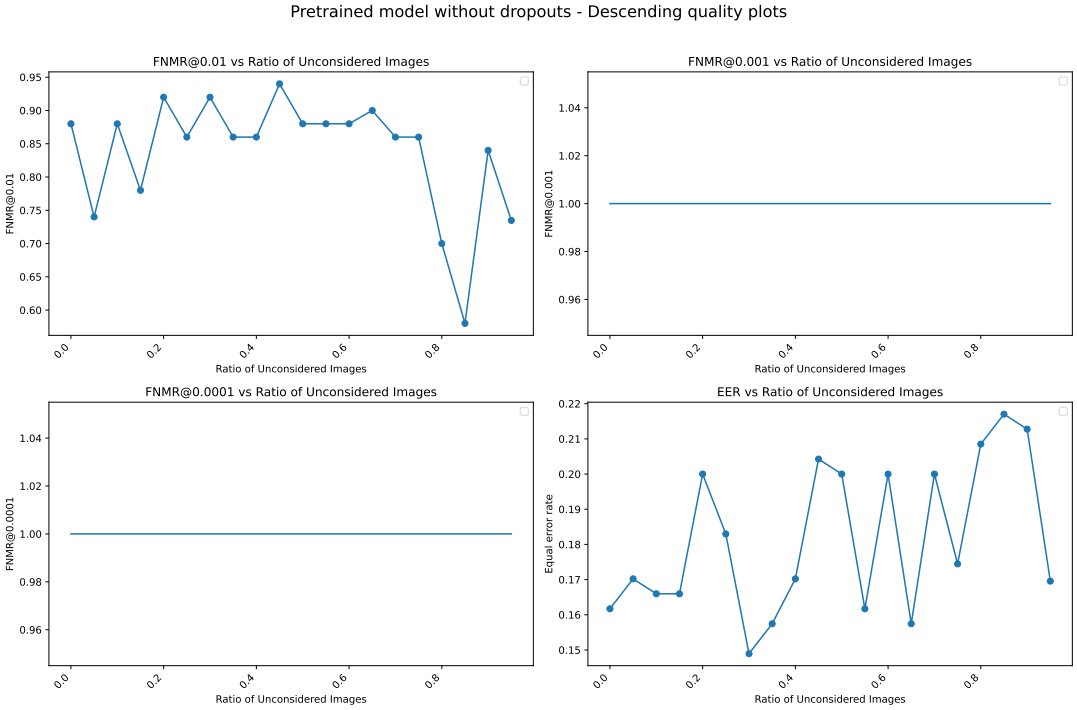

Figure 5.17: Case 4 plots for (pinsdataset)dataset 2, with an assumption that highest quality scores are assigned to lowest quality images.

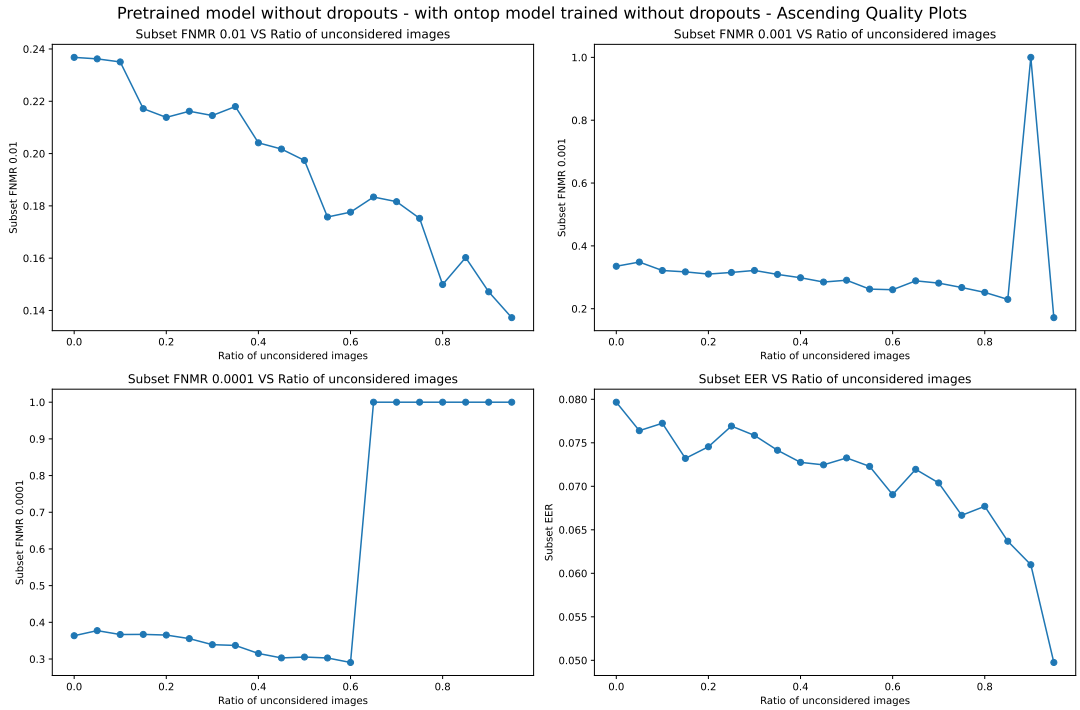

Figure 5.18: Case 5 plots for dataset 1(manually prepared dataset), with an assumption that lowest quality scores are assigned to lowest quality images.

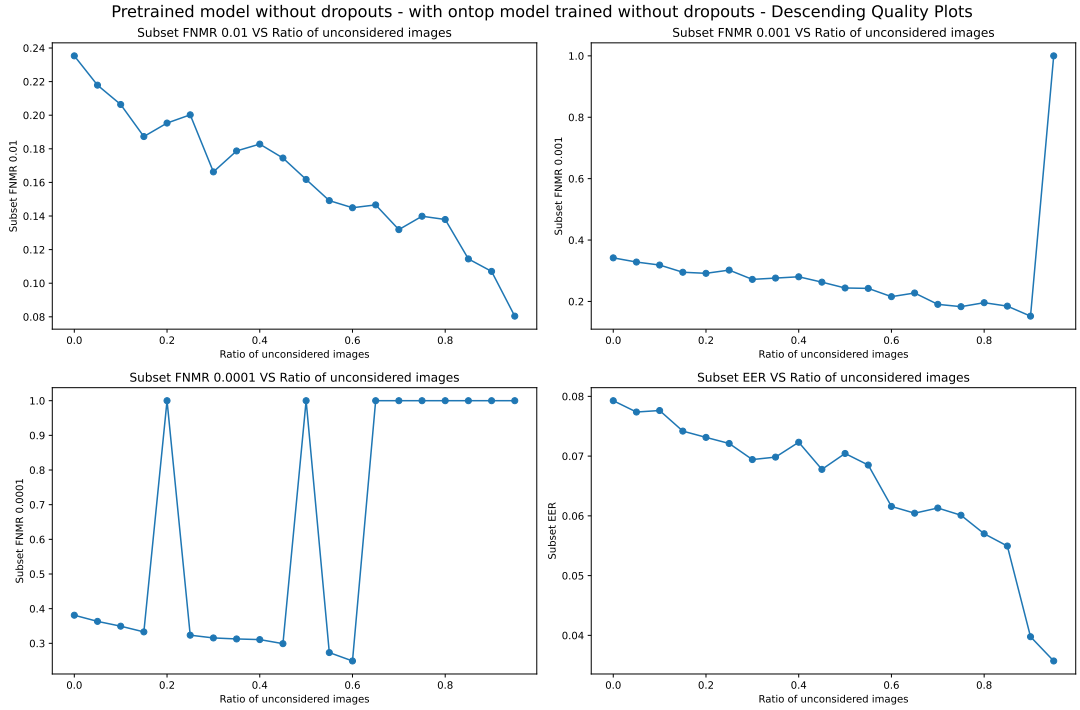

Figure 5.19: Case 5 plots for dataset 1(manually prepared dataset), with an assumption that highest quality scores are assigned to lowest quality images.

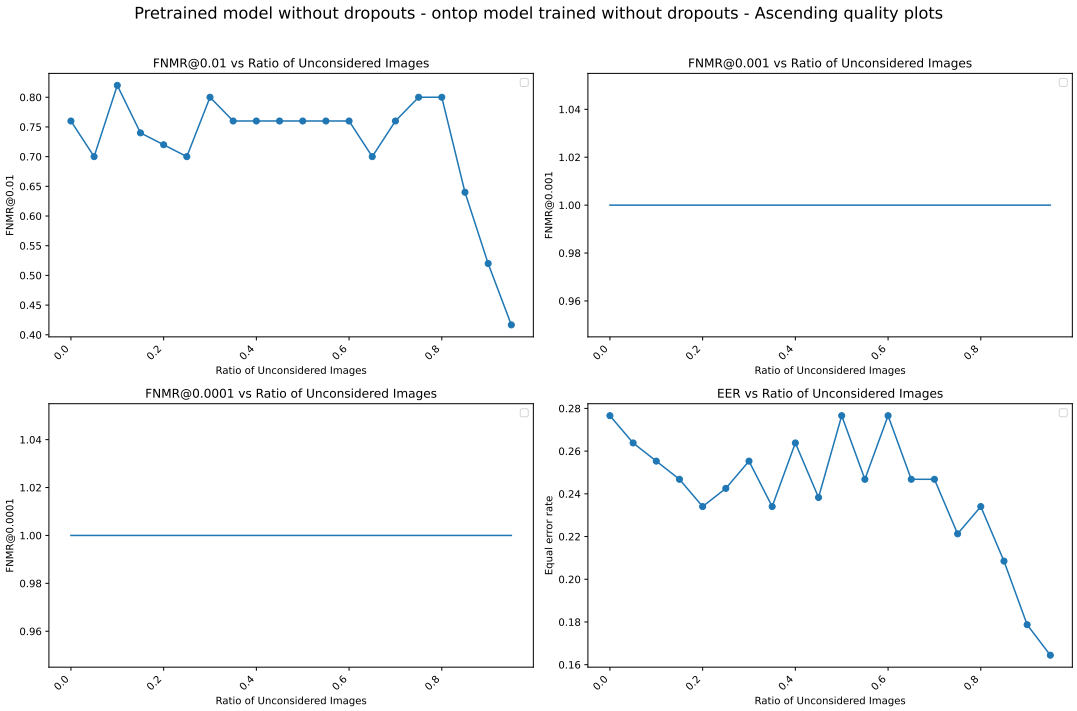

Figure 5.20: Case 5 plots for (pinsdataset)dataset 2, with an assumption that lowest quality scores are assigned to lowest quality images.

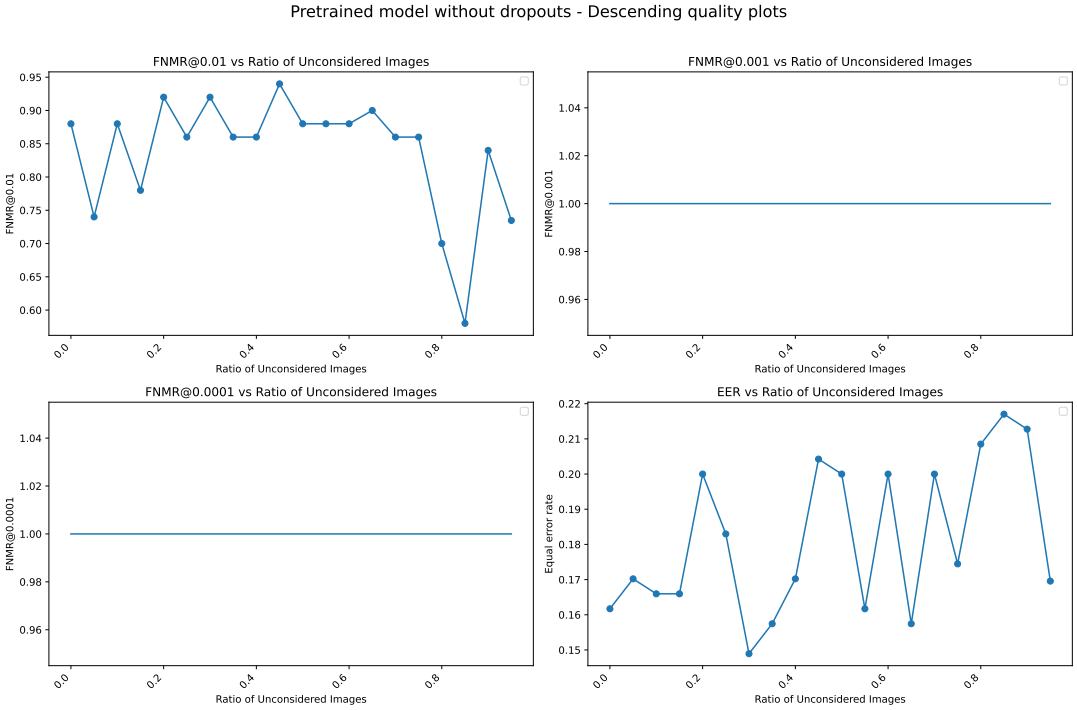

Figure 5.21: Case 5 plots for (pinsdataset)dataset 2, with an assumption that highest quality scores are assigned to lowest quality images.

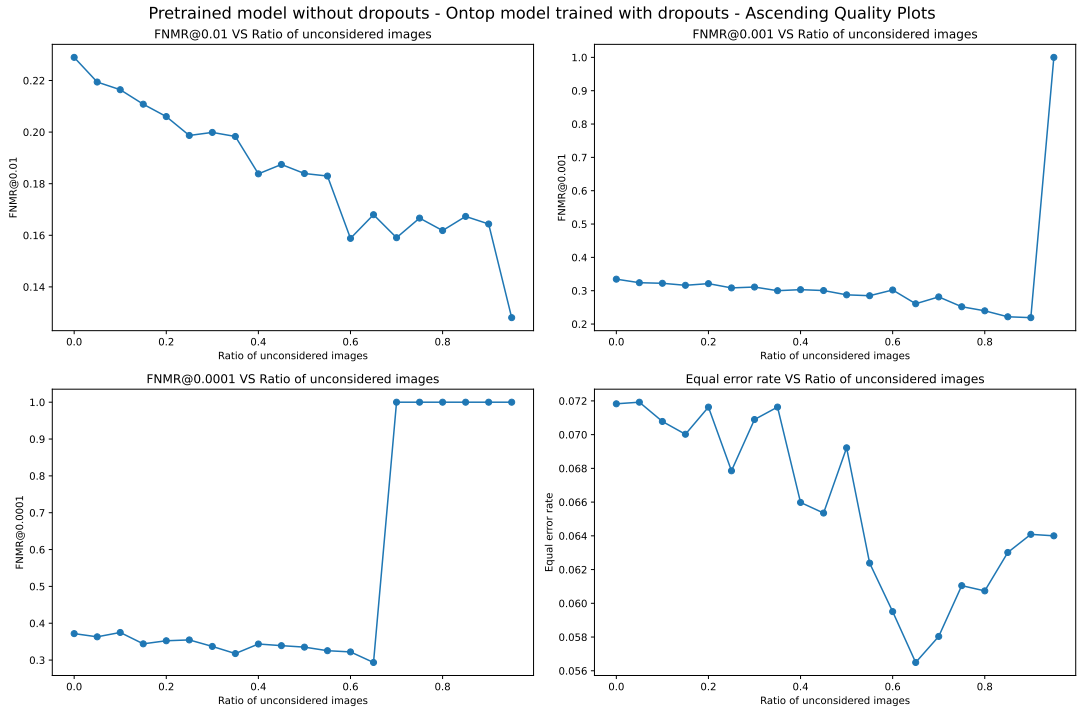

Figure 5.22: Case 6 plots for dataset 1(manually prepared dataset), with an assumption that lowest quality scores are assigned to lowest quality images.

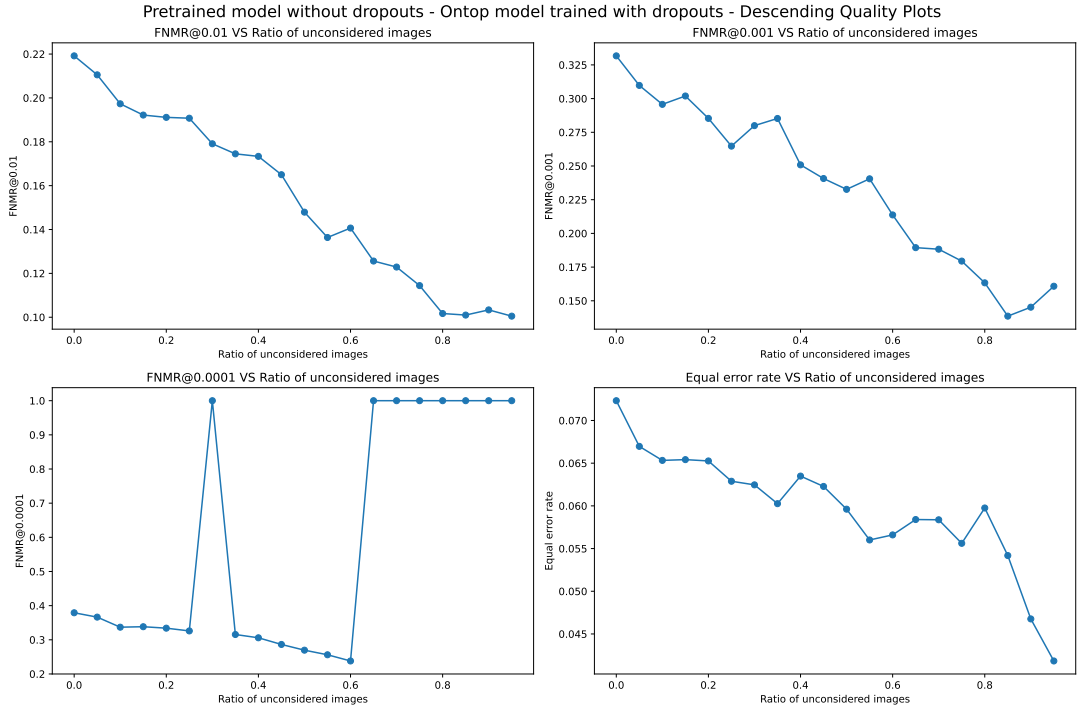

Figure 5.23: Case 6 plots for dataset 1(manually prepared dataset), with an assumption that highest quality scores are assigned to lowest quality images.

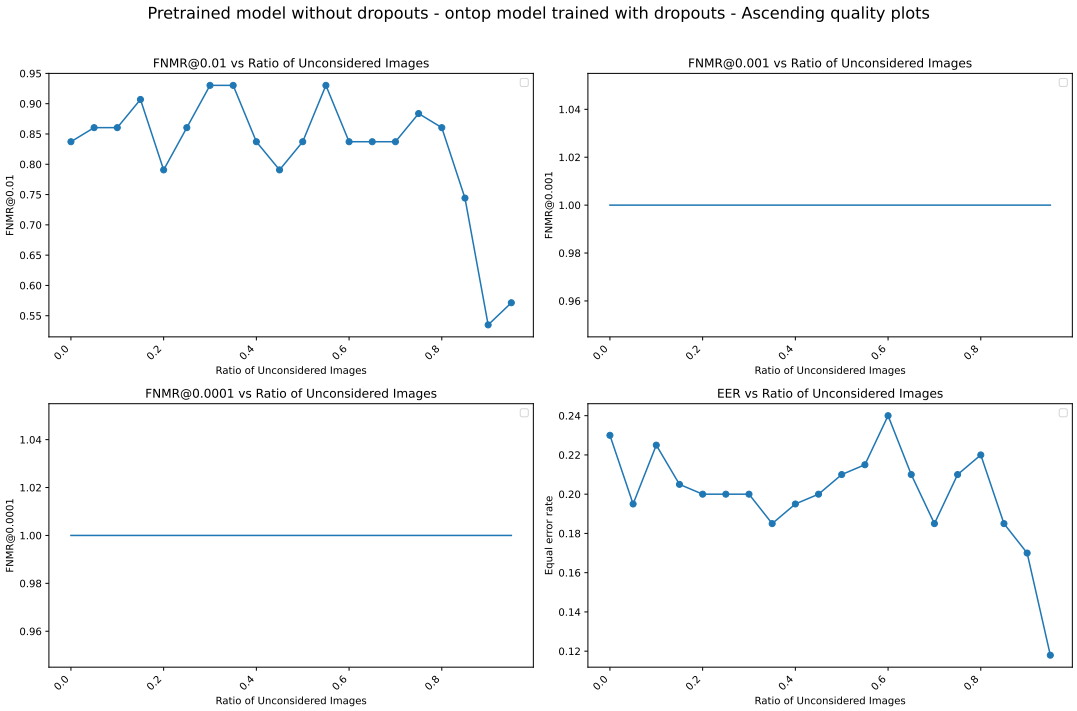

Figure 5.24: Case 6 plots for (pinsdataset)dataset 2, with an assumption that lowest quality scores are assigned to lowest quality images.

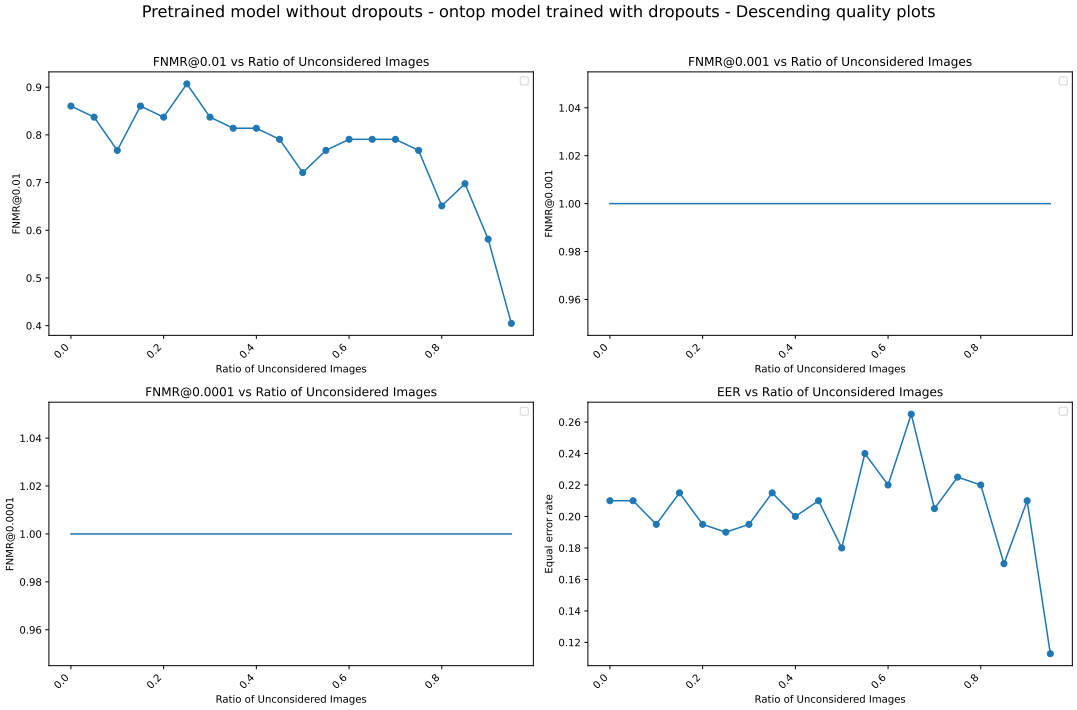

Figure 5.25: Case 6 plots for (pinsdataset)dataset 2, with an assumption that highest quality scores are assigned to lowest quality images.

				Manually prepared dataset (lfw-ytf)															
Case	Model specifications	Models used	Loss functions	AUC score obtained (Reciever operating curve)	Ascending (Error vs reject curves)								Descending (Error vs reject curves)							
					FNMR@ 0.01 FMR		FNMR@ 0.001 FMR		FNMR @0.0001 FMR		EER		FNMR @0.01 FMR		FNMR @0.001 FMR		FNMR @0.0001 FMR		EER	
					error range	trend	error range	trend	error range	trend	error range	trend	error range	trend	error range	trend	error range	trend	error range	trend
case 1	Pretrained model with dropouts	ArcFace model	ArcFace loss	1.0	0.0025 to 0.0175	Increasing	0.05 to 0.20	Increasing	0.3 to 1.0	Increasing	0.006 to 0.011	Increasing	0.002 to 0.014	Increasing	0.05 to 1.0	Increasing	0.2 to 1.0	Increasing	0.007 to 0.011	Increasing
case 2	Pretrained model with dropouts and an on-top model trained without dropouts	ArcFace model & on-top model trained without dropouts	On-top model trained with sparse categorical cross entropy	1.0	0.08 to 0.01	Decreasing	0.20 to 0.08	Decreasing	0.2 to 1.0	Increasing	0.030 to 0.010	Decreasing	0.08 to 0.03	Decreasing	0.18 to 0.08	Decreasing	0.3 to 0.1	Decreasing (dataset unstable at last)	0.034 to 0.020	Decreasing
case 3	Pretrained model with dropouts and an on-top model trained with dropouts	ArcFace model & on-top model trained with dropouts	On-top model trained with sparse categorical cross entropy	0.91	0.08 to 0.040	Decreasing	0.2 to 0.09	Decreasing (anomalies in between)	0.24 to 1.0	Increasing	0.030 to 0.018	Decreasing	0.065 to 0.02	Decreasing (anomalies in between)	0.2 to 1.0	Increasing	0.27 to 1.0	Increasing	0.032 to 0.022	Decreasing
case 4	Pretrained model without dropouts	OpenFace model	Improved triplet loss	0.95	0.245 to 0.05	Decreasing	0.40 to 0.09	Decreasing	0.38 to 1.0	Increasing	0.089 to 0.052	Decreasing	0.250 to 0.070	Decreasing	0.39 to 0.10	Decreasing	0.4 to 0.1	Decreasing (dataset unstable at last)	0.090 to 0.055	Decreasing
case 5	Pretrained model without dropout and an on-top model trained without dropouts	OpenFace model & on-top model trained without dropouts	On-top model trained with sparse categorical cross entropy	0.99	0.24 to 0.10	Decreasing	0.35 to 0.10	Decreasing (dataset unstable at last)	0.38 to 1.0	Increasing	0.080 to 0.050	Decreasing	0.24 to 0.08	Decreasing	0.35 to 0.09	Decreasing (dataset unstable at last)	0.4 to 1.0	Increasing	0.08 to 0.02	Decreasing
case 6	Pretrained model without dropouts and an on-top model trained with dropouts	OpenFace model & on-top model trained with dropouts	On-top model trained with sparse categorical cross entropy	0.99	0.24 to 0.10	Decreasing	0.32 to 0.2	Decreasing (unstable at last)	0.39 to 1.0	Increasing	0.072 to 0.056	Decreasing	0.22 to 0.10	Decreasing	0.325 to 0.150	Decreasing	0.40 to 1.0	Increasing	0.075 to 0.030	Decreasing

Table 5.1: Recorded observations for the manually prepared dataset (Reasons for the observations are described in the section **??** .) Where all the results were transparently recorded without any bias and the scientific reasons behind each scenario was analysed.

Pinsdataset																				
Case	Model specifications	Models used	Loss functions	AUC score obtained (Reciever operating curve)	Ascending (Error vs reject curves)								Descending (Error vs reject curves)							
					FNMR@ 0.01 FMR		FNMR @0.001 FMR		FNMR@ 0.0001 FMR		EER		FNMR @0.01 FMR		FNMR @0.001 FMR		FNMR @0.0001 FMR		EER	
					Error range	Trend	Error range	Trend	Error range	Trend	Error range	Trend	Error range	Trend	Error range	Trend	Error range	Trend	Error range	Trend
Case 1	Pretrained model with dropouts	ArcFace model	ArcFace loss	1.0	At 1.0	Horizontal straight line.	At 1.0	Horizontal straight line.	At 1.0	Horizontal straight line.	At 0.0	Horizontal straight line.	At 1.0	Horizontal straight line.	At 1.0	Horizontal straight line.	At 1.0	Horizontal straight line	At 0.0	Horizontal straight line
Case 2	Pretrained model with dropouts and an on-top model trained without dropouts	ArcFace model & an on-top model trained without dropouts.	On-top model trained with sparse categorical cross entropy	1.0	from 0.10 to 0.40.	Decreasing then increasing.	At 1.0.	Horizontal straight line.	At 1.0	Horizontal straight line.	from 0.065 to 0.030.	very random.	0.15 to 0.35.	very random	At 1.0	Horizontal straight line.	At 1.0	Horizontal straight line.	from 0.065 to 0.030	Decreasing
Case 3	Pretrained model with dropouts and an on-top model trained with dropouts	ArcFace model & an on-top model trained with dropouts.	On-top model trained with sparse categorical cross entropy	1.0	from 0.25 to 0.02	Decreasing.	At 1.0	Horizontal straight line.	At 1.0	Horizontal straight line.	from 0.0200 to 0.0050.	Decreasing	0.0 to 1.0	Increasing	At 1.0	Horizontal straight line	At 1.0	Horizontal straight line	from 0.0175 to 0.0000	Decreasing
Case 4	Pretrained model without dropouts	OpenFace model	Improve triplet loss	1.0	from 0.90 to 0.60	Decreasing	At 1.0	Horizontal Straight line	At 1.0	Horizontal straight line	from 0.14 to 0.19	Increasing	from 0.95 to 0.60	Decreasing	At 1.0	Horizontal straight line	At 1.0	Horizontal straight line	from 0.15 to 0.22	very random
Case 5	Pretrained model without dropouts and an on-top model trained without dropouts	OpenFace model & an on-top model trained without dropouts.	On-top model trained with sparse categorical cross entropy	1.0	from 0.80 to 0.40	Decreasing	At 1.0	Horizontal straight line	At 1.0	Horizontal straight line	from 0.28 to 0.16	Decreasing	0.650 to 0.850	very random	At 1.0	Horizontal straight line	At 1.0	Horizontal straight line	from 0.28 to 0.18	Decreasing
Case 6	Pretrained model without dropouts and an on-top model trained with dropouts.	OpenFace model & an on-top model trained with dropouts.	On-top model trained with sparse categorical cross entropy	1.0	from 0.95 to 0.55	Decreasing	At 1.0	Horizontal straight line	At 1.0	Horizontal straight line	from 0.240 to 0.12	Decreasing	0.9 to 0.4	Decreasing	At 1.0	Horizontal straight line	At 1.0	Horizontal straight line	from 0.26 to 0.12	Decreasing

Table 5.2: Recorded observations for the pinsdataset (Reasons for the observations are described in the section **??** .) Where all the results were transparently recorded without any bias and the scientific reasons behind each scenario was analysed. Two main reasons for the unusual results are the dataset imbalance and sampling bias.

5.2 Findings

This section states the factors that might be responsible for the results we observed in our experiments. The factors might range widely from the dataset preparation to the computation of the error vs reject curves.

5.2.1 Factors influencing the results

Anomalies in the datasets

The unusual results might probably attribute to the **presence of anomalies within the datasets, incorrect labelling, inefficient face detection by MTCNN[ZZLQ16], which in turn might result in the inaccuracies during the preprocessing phase**. When we applied the MTCNN algorithm[ZZLQ16], we found out that there were several anomalies present in the dataset, which might probably be the main cause of the unusual results. Figure **??** shows the presence of anomalies, inefficient face detection and inaccurate preprocessing(indicated by irregular sizes). Which might also be due to the **presence of different face expressions, head poses, face occlusions, really low quality images, inaccurate capturing of images and sometimes even the absence of a face in an image**[TKD+20]. Sometimes these anomalies were seen to have inaccurate labelling. Figure 5.26 illustrates few anomalies present in our experiments with their associated quality scores. There were at least three to four such anomalies spotted in every 100 images, in the manually prepared dataset. In this way there were a lot of instances where the face detection and other defects existed. Hence the problems lie in face detection, where unusual face sizes were seen, anomalies, unusual face detections, where some face images also had incorrect naming and further **preprocessing them finally had led to a lot of transformations in the face images**. Hence lies a lot of problems. All these problems contributed to the ambiguity observed between the ascending quality plots and descending quality plots and other unusual results in our experiments. Continuous investigation is necessary to overcome these problems.

Figure 5.26: Presence of anomalies with their associated quality scores in our experiments. Where, in some images, the face is absent. We saw a lot of such false positives and false negatives.[HRBLM07], [WHM11]

Sampling bias

Sampling bias might have significantly impacted our results, **particularly in limiting the number of imposter scores. The ratio of genuine to imposter scores was maintained**

as 1:5 due to the low system complexity. This sampling bias could have significantly impacted the performance in case of pinsdataset. Where there were horizontal lines in the curves, the graphs showed 100% FNMR errors and 0% Equal error rates, which means that the system is rejecting all the matches and becomes stringent[TKD+20]. Sampling bias might probably be the reason for the unusual observations when ArcFace model was solely employed in case 1[DGNZ19]. Additionally, sampling bias might probably be a reason for the ambiguity observed with the ascending and descending quality plots. Therefore , sampling bias might have played a huge role in every unusual result that we obtained. **It was observed that if the dataset has a good class distribution and wide variety of image qualities, the SER-FIQ algorithm might probably be robust to the sampling bias**. Careful experimentation with a good system complexity would mitigate the sampling bias.

Imbalanced datasets

A dataset imbalance occurs when few classes are over represented in a dataset and some classes are underrepresented in a dataset. In such cases the model that is being trained might get biased towards the over represented classes. The threshold tuning in the *roccurve*() might get affected due to such imbalances in the dataset[THKG20]. Such a condition might have happened with the pinsdataset[Bur20] where just 105 identities were spread across more than 17,000 face image samples. The class imbalance ratio of pinsdataset was found to be 0.3628691983122363(86:237). Imbalance in the class distribution might be the reason for the unusual results observed in case of pinsdataset. For deeper insights we also calculated the class imbalance ratio of the manually imbalanced dataset, which was found to be 530.0(530:1). The class distribution of both the datasets is represented in figures 5.27 and 5.28. This ratio is also hugely imbalanced, therefore the class imbalance[THKG20] might be a reason for the ambiguity observed between the ascending quality and descending quality plots. A face recognition system may encounter underrepresented samples when faced with the imbalanced datasets, it may lead to impaired robustness, thus making it difficult to evaluate the performance of the face recognition system. **Incorporating cross validation and resampling techniques would mitigate the class imbalance**[THKG20].

Impact of pretrained models

Irrespective of the different datasets used in our experiments, it was evident from the error vs reject curves that selecting the pretrained models plays an important role in the evaluation of the performance of the recognition system. Where the ArcFace model[DGNZ19] trained with dropout regularization(with its on-top models), provided results with a range of lower errors from almost 1% to 8% on the y-axes. While in the OpenFace model(With its on-top models), the errors were too high from 24% to more than 80%[ALS+16]. This indicates that the **pretrained models played an important role in our experiments**. Where selecting a pretrained model with dropout regularization might be better to deploy the SER-FIQ algorithm[TKD+20] for robust results[SHK+14], [Sri13].

Over fitting and under fitting

Imbalance between the over fitting and under fitting might have impacted the ascending quality and descending quality plots[Yin19], [Haw04]. Which means that there might have existed severe **imbalance between the over fitting and under fitting of the on-top models**. This imbalance was evident from the varying training accuracy and validation accuracy. Where case 6 had a training accuracy of 94.19% and validation accuracy of 8%. Case 5 had a Training

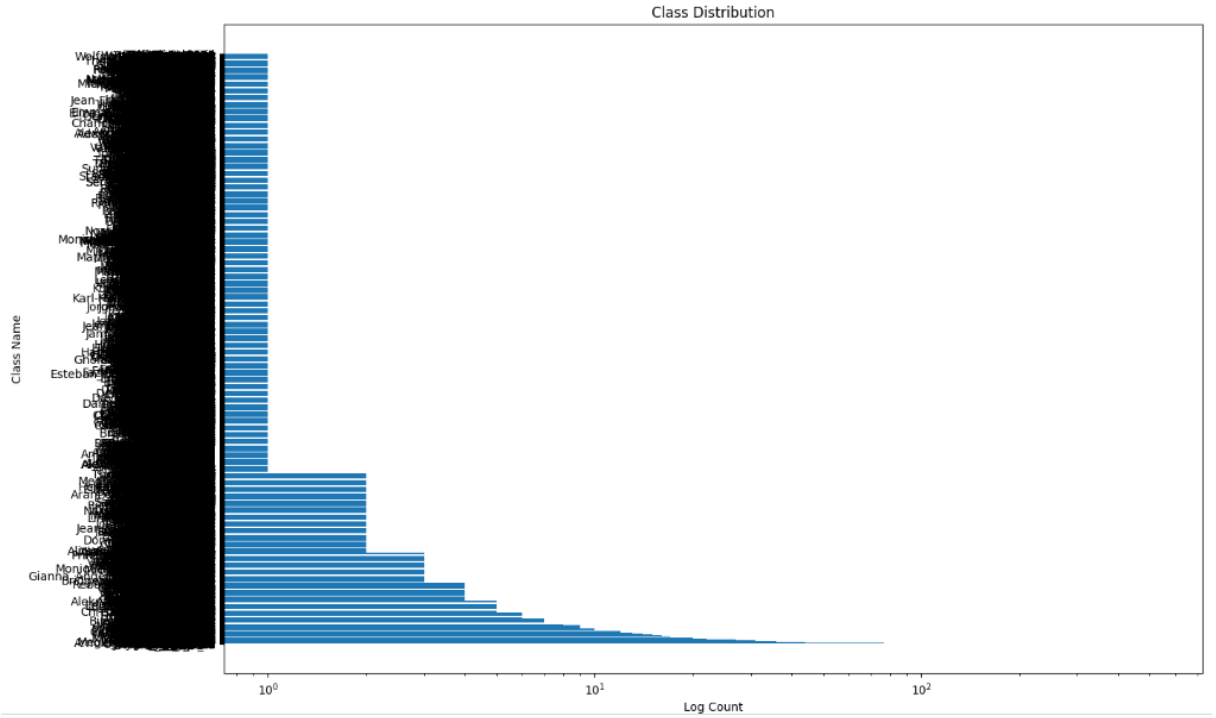

Figure 5.27: Class distribution of the manually prepared dataset. Class Imbalance ratio of the manually imbalanced dataset was found to be 530.0(530:1).[WHM11], [HRBLM07]

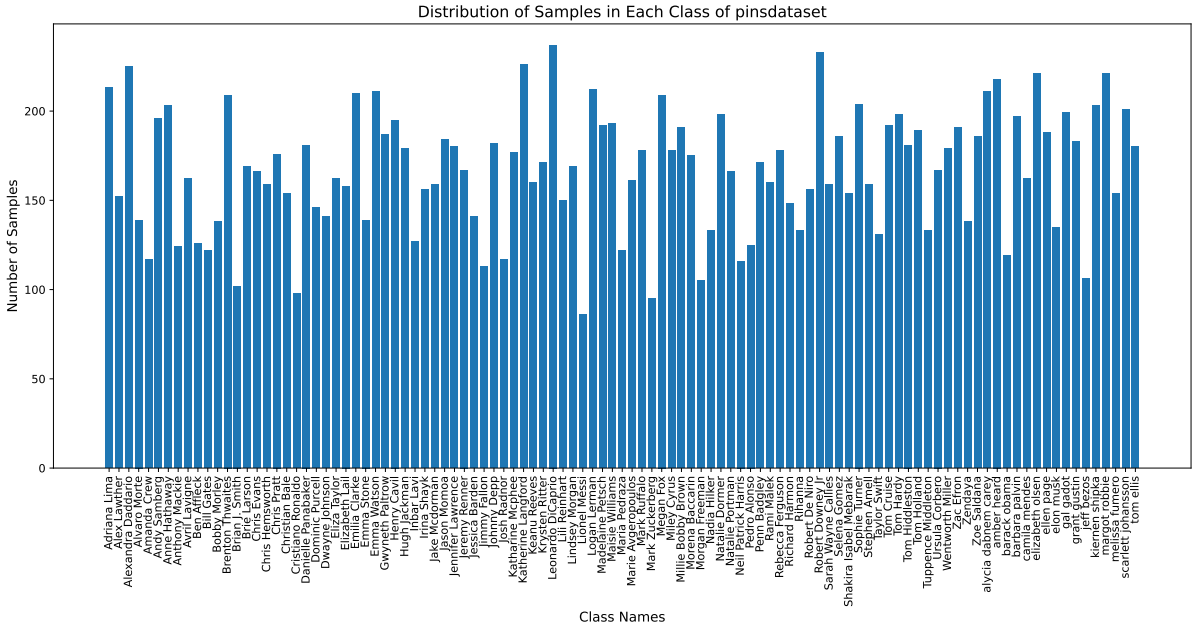

Figure 5.28: Class distribution of the pinsdataset. The class imbalance ratio of pinsdataset was found to be 0.3628691983122363(86:237)[Bur20].

accuracy : 99% and validation accuracy: 20%. Case 3 had a training accuracy of 92.79% with a validation accuracy of 82%. Case 2 had a training accuracy of 99.9999% and the validation accuracy of 20%. In each case, the imbalance was evident. This bias and variance might probably be the reason for the unusual results in each use case. **This over fitting[Yin19], [Haw04]**

and under fitting might be mitigated by the integration of dropout regularization with L1/L2 regularization and ensemble learning techniques[KSST12], [NJ18], [Die00].

Loss functions

In case of the ArcFace model, the loss function was the ArcFace loss[DGNZ19]. As explained in the introduction, **ArcFace loss** might serve as a good benchmark to provide a good classification and for a better feature extraction purpose. The factor that led to higher range of errors in case of the OpenFace model[ALS+16] might probably be the loss function .i.e. the OpenFace model uses an improved triplet loss function. This triplet function might have contributed to the unusual results because as discussed in the introduction, triplet loss is suitable for only small datasets[Ge18], [HBL17], [ALS+16]. This triplet loss function works around maximizing the distance between the anchor-positive pairs and minimizing the distance between anchor-negative pairs[ALS+16]. **Initially it was assumed that the improved triplet loss training in the OpenFace model with back propagation, would provide some desirable results**[ALS+16]. But the fact is that the range of errors were a little high, and the decreasing curves were as anticipated. While in case of the on-top models, we used sparse categorical cross entropy(SCCE) loss function[WMZT20]. This SCCE loss function is sensitive to the class imbalance and has many other disadvantages[WMZT20]. Hence the SCCE also might have contributed to the unusual results.

Other reasons for the ambiguity between the ascending and descending quality plots

Existence of this ambiguity clearly means that the sorting of quality scores does not resonate well with the generated genuine and imposter pairs. The order in which the quality scores are sorted did not impact the performance evaluation in our experiments, since it might probably depend on the model's capability to address different factors in a face image. Which might logically mean that the models might have not learnt the features well enough. Therefore lack of data augmentation might be a reason for the existence of this ambiguity. Architectural choices play a major role in such scenarios. Therefore meticulously experimenting with a proper model selection, training, hyper parameter choices and cross validation might probably solve the ambiguity between the ascending quality and descending quality plots. The causes of the same trends in both ascending and descending quality plots might attribute to every implementation detail that contributes in the computation of the overall performance

Other reasons for the unusual results with pins dataset.

In addition to the above factors like sampling bias, dataset imbalance, architectural choices, model selections and other factors, there might be other factors that could have influenced the unusual results in the pinsdataset. This pinsdataset includes wide variety of ethnicities and age groups. Different ethnicities, different age groups and different facial expressions might have contributed to the unusual results we observed in our experiments.

Other reasons

Other reasons for the overall unusual results, except the above factors might include the **hyper parameter choices and architectural choices of the on-top models**. Not being aware of the full training information of the pretrained models(ArcFace model and the OpenFace model) might be another factor. Adapting the architectural choices as per the latest research on CNNs might give better results. Another factor that influenced our unusual results might be the**use of insufficient training data for the on-top models.** The insufficient data used in training

these models, might have made the models to learn less features. Which might have been the reason for the models to not generalise better to the unseen features. Hence it was probably evident that, the use of insufficient training data might have been the reason for the unusual results observed in our experiments.

5.3 Comparative analysis

Comparative analysis is necessary to discuss which models did better, which case did better on which dataset. Hence doing so, this section will focus on a comprehensive comparative analysis, to derive the effect of dropouts on face image quality assessment(FIQA). Thus proving the robustness of SER-FIQ algorithm[TKD+20]. The tables 5.1 and 5.2 shows the recorded observations for manually prepared dataset and pinsdataset[Bur20] respectively. Each use case is characterized by specific model configurations, dataset preparation and training procedures. Here we aim to understand how can we compare different use cases and different datasets. To come to a conclusion on which use case might give better results when SER-FIQ is tested on and why? A detailed comparative analysis is performed in the table 5.3 to analyse strengths and weaknesses of each case for both datasets.

When comparing different cases, it is evident that all results were not as expected, but some results were as expected. When ArcFace model[DGNZ19] was solely employed in case 1, we observed increasing trends in the error vs reject curves. While in case 2 and case 3, the curves were decreasing. With the manually prepared dataset, the results were robust with low error ranges and decreasing curves. **This is because using dropout regularization[KSST12], [NJ18], [SHK+14], [Sri13] in both the base model and the on-top model might have mitigated over fitting[Yin19], [Haw04] and helped the models to generalise well enough to the unseen data. Compared to case 2 and case 1, case 3 was probably proved to be far better.** When we compare the plots of both the datasets, the manually prepared dataset was probably far more better than the pinsdataset[Bur20]. In-spite of having huge sampling bias, the manually prepared dataset might have done better. **The results clearly demonstrate that the dataset with a large variety of face images, qualities and other factors might perform better.** Further experiments were performed with a different base model that does not contain dropouts during training, this was done to obtain unbiased results throughout the experiments.

When we consider a pretrained model without dropouts, the error ranges were high with decreasing curves(Case 4 - OpenFace model[ALS+16]). The decreasing curves, basically means that the error rates decrease while we start rejecting lowest quality samples. Even though the dropouts were not used to train the OpenFace model, careful **incorporation of the dropouts during validation might have proved to provide desirable results**. Moreover, not using dropouts during training, might have led to a little high errors. These high errors might be e a result of the consequence that with the absence of dropouts during training, the model performs poorly as it finds it difficult to generalise to the unseen facial features. Therefore it is significant that SER-FIQ algorithm performs better when dropouts are used during training.

When we look at the further two cases(case 5 and case 6) when an on-top model is built upon the OpenFace model, without dropouts and with dropouts . Case 5 is probably better than case 4, case 6 is probably better than case 5, due to the utility of dropouts in the on-top model during the training process. This is because , in case 6, dropouts are used in the on-top model for training. The 'EER vs reject curve' can be monitored in each case for this comparison, where case 6 has a plot with lower errors when compared to case 5 and case 4. **Overall, the impact of dropouts might have been really huge.** When the dropouts were used during

training the pretrained model and the on-top model, the results might probably be fascinating. **Case 2 is probably better than case 1 and case 3 is probably better than case 2, focusing on the improvements observed due to the consistent incorporation of the dropouts in the training process. All in all, case 3 was probably found to be better than all cases, due to its low errors and decreasing curves.** It is clear that consistent incorporation of dropouts might lead to effective and reliable performance.

The manually prepared dataset might have probably done better than the pins-dataset, even if there was a huge sampling bias. The reason for this might attribute to the wide variety of image qualities and identities. But in each case, the curve 'FNMR@0.0001FMR vs ratio of unconsidered images' became more unstable because the FMR threshold was too low. Still the results of case 1 were probably not clear and the ambiguity between the ascending quality plots and descending quality plots was probably not resolved. This is one of the open question that arises from our experiments. Table 5.3 shows a detailed comparative analysis for different cases and different datasets. **When it comes to the impact of dropouts, case 3 might have performed better amongst all other cases.**

Case	Model specifications	Manually prepared dataset (5686 identities spread over 15,438 face images)						Pinsdataset (105 identities spread over more than 17,000 face images)						
		Strengths	Weakness (Refer last row for the assumptions)	Impact of dropouts during training	Impact of dropouts during validation	Reasons for unusual observations	Notes	Strengths	Weakness (Refer last row for assumptions.)	Impact of dropouts in ROC curve.	Impact of dropouts during validation	Reasons for unusual observations	Notes
Case 1	A Pretrained model with dropouts.	Low error range, Improved ROC performance.	Increasing curves, Ambiguity between the assumptions of different sorting orders not resolved	Dropouts used during training leads to low errors and better ROC performance.	Low error range and improved robustness.	- Inefficient face detection. - Presence of anomalies. - Incorrect labelling. - Dataset class imbalance. - All training details not available. - Sampling bias(limiting the number of imposter scores) - Lack of data augmentation.	- Cross validation is needed with careful model selection. - Mitigating sampling bias.	- High image resolutions, - Good dataset accuracy - Best ROC performance.	Horizontal straight line graphs. Ambiguity between the assumptions of different sorting orders not resolved.	Best ROC performance	Best ROC performance.	- Less number of identities spread over huge number of samples.(imbalanced dataset) - Sampling bias(limiting the number of imposter scores) - All training data not available. - Lack of data augmentation. - Ethnicities, different age groups, different face expression and different face occlusions.	- Cross-validation is necessary to select an optimal pretrained model. - Incorporate resampling techniques. - Mitigate sampling bias (compute all genuine and imposter pairs and scores.)
Case 2	A Pretrained model with dropouts and an on-top model trained without dropouts.	Low error range, decreasing curves and enhanced ROC performance.	Ambiguity between the assumptions of different sorting orders not resolved	Dropouts used in the base model leads to low errors with improved model generalization.	Improved robustness with enhanced effects seen through low error range and decreasing curves. Enhanced performance in the curve : 'equal error rate vs ratio of unconsidered images'	- Insufficient training data. - Overfitting. - Sampling bias. - Inefficient face detection. - Dataset class imbalance. - Incorrect labelling. - Hyperparameter sensitivity. - Lack of data augmentation. - Incorrect labelling.	- Use sufficient training data. - Cross validation is necessary. - Data augmentation is necessary.	- Improved robustness seen through a better curve 'Equal error rate vs ratio of unconsidered images' - Enhanced performance.	FNMR curves were still unusual. Ambiguity between the assumptions of different sorting orders not resolved.	Best ROC performance	- Improved robustness seen through a better curve 'Equal error rate vs ratio of unconsidered images'	- Insufficient training data. - Lack of data augmentation. - Hyperparameter sensitivity. - Overfitting.	- Using sufficient training data. - Mitigating sampling bias. - Cross validation is important.
Case 3	A Pretrained model with dropouts and an on-top model trained with dropouts.	Low error range, decreasing curves and enhanced ROC performance.	Ambiguity between the assumptions of different sorting orders not resolved	Dropouts used in the base model leads to low errors with enhanced model generalization due to the use of dropouts in the on-top model.	Improved robustness with enhanced effects seen through low error range and decreasing curves. Smarter performance in the curve : 'equal error rate vs ratio of unconsidered images'	- Insufficient training data. - Overfitting. - Sampling bias. - Incorrect labelling. - Inefficient face detection. - Imbalanced dataset - Incorrect labelling. - Hyperparameter sensitivity.	- Use sufficient training data -Cross validation is necessary. - Data augmentation is necessary. - Mitigating sampling bias.	Improved robustness seen through a better curve 'Equal error rate vs ratio of unconsidered images' - Smarter performance.	Very random FNMR curves. All FNMR curves were still unusual.	Best ROC performance.	- Smarter performance	- Insufficient training data. - Lack of data augmentation. - Hyperparameter sensitivity. - Overfitting. - Lack of cross validation.	- Using sufficient training data. - Mitigate overfitting. - Hyperparameter tuning. - Careful architectural choice. - Cross validation is necessary.
Case 4	A Pretrained model without dropouts.	Decreasing curves and reasonable ROC performance	High error range, Ambiguity between the assumptions of different sorting orders not resolved	Dropouts not used during training leads to high errors.	Improved robustness indicated by the decreasing curves. Better performing curve :'equal error rate vs ratio of unconsidered images'	- Generalized version of OpenFace model used. - All training details not available. - Presence of anomalies. - Incorrect labelling. - Inefficient face detection. - Hyperparameter sensitivity. -Imbalanced dataset	- Use an OpenFace model .i.e. more tailored to face recognition tasks. - Mitigating sampling bias.	FNMR@0.01FMR curve was decreasing.	- Other three plots were unusual. - High error ranges	Best ROC performance.	Due to sampling bias and imbalance in dataset,the graphs were unsual.	- Due to the use of more generalized version of the OpenFace model. - Sampling bias. - Imbalanced dataset.	- Resample the pinsdataset. - Cross validation is necessary. - Use a more tailored version of the OpenFace model .i.e. specific to the face recognition.
Case 5	A pretrained model without dropouts and an on-top model trained without dropouts.	Decreasing curves and enhanced ROC performance.	High error range, Ambiguity between the assumptions of different sorting orders not resolved	Dropouts not used in the base model during training leads to high errors, with improved generalization due to on-top model.	Improved robustness indicated by the decreasing curves. Enhanced performance in the curve :'equal error rate vs ratio of unconsidered images'	- Insufficient training data. - Overfitting. - Hyperparameter sensitivity. - Presence of anomalies. - Incorrect labelling. - Inefficient face detection.	- Use L1/L2 regularization - Use sufficient training data. - Hyperparameter tuning. - Data augmentation is necessary. - Mitigating sampling bias.	Improved robustness with 'FMR@0.01FMR vs ratio of unconsidered images' curve and the curve 'Equal error rate' were decreasing. Better curves than case 4.	- Other curves FNMR@0.001FMR threshold and FNMR@0.0001FMR were still straight lines. - High error ranges.	Best ROC performance.	(Better performance) Improved robustness with 'FMR@0.01FMR vs ratio of unconsidered images' curve and the curve 'Equal error rate' were decreasing.	- Overfitting. - Hyperparameter sensitivity. - Insufficient training data. - Inefficient face detection on the training data. - Presence of anomalies in the training data.	- Use different techniques to mitigate overfitting. - Mitigate sampling bias. - Use sufficient training data. - Anomaly detection.
Case 6	A pretrained model without dropouts and an on-top model trained with dropouts.	Decreasing curves and enhanced ROC performance.	High error range, Ambiguity between the assumptions of the different sorting orders not resolved.	Dropouts not used in the base model during training leads to high errors, with enhanced generalization due to the dropouts in the on-top model.	Improved robustness indicated by the decreasing curves. Smarter performance in the curve :'equal error rate vs ratio of unconsidered images'	- Insufficient training data - Overfitting - Presence of anomalies. - Incorrect labelling. - Inefficient face detection. - Hyperparameter sensitivity. - Sampling bias.	- Use L1/L2 regularization. - Use sufficient training data. - Hyperparameter tuning. - Data augmentation is necessary. - Sampling bias.	Improved robustness with 'FMR@0.01FMR vs ratio of unconsidered images' curve and the curve 'Equal error rate vs ratio of unconsidered images' were decreasing. Better curves than case 4 and case 5 (smarter performance)	- Other curves FNMR@0.001FMR threshold and FNMR@0.0001FMR were still straight lines. - High error ranges.	Best ROC performance.	(Better performance than case 5) Improved robustness with 'FMR@0.01FMR vs ratio of unconsidered images' curve and the curve 'Equal error rate' were decreasing.	- Overfitting. - Hyperparameter sensitivity. - Insufficient training data. - Inefficient face detection on the training data. - Presence of anomalies in the training data.	- Use L1/L2 regularization or other techniques to mitigate overfitting. - Hyperparameter tuning. - Using sufficient training data. - Anomaly detection.
Assumption 1:{Assuming that the lowest quality scores are assigned to the lowest quality images.} Assumption2 : {Assuming that the highest quality scores are assigned to the lowest quality images.}														

Table 5.3: Comparative analysis of both the datasets with different cases analysed with their strengths, weaknesses and respective notes if necessary.

5.4 Limitations

There were few limitations in our experiments. Firstly, we were **unable to explore the effect of different dropout rates on the performance due to the availability of low computational resources**. Future work could focus on experimenting the impact of different dropout rates. Secondly, the **analysis between the ascending and descending quality plots exhibited an ambiguity**. This ambiguity can be resolved by addressing the current challenges. Thirdly, **the findings were not generalised across wide variety of datasets**. Which clearly indicates that there was a limit in the experimental conditions. Fourthly, **there was a lack of cross validation and resampling techniques in our experiments**. Therefore continuous refinement is necessary by integrating different techniques to mitigate the challenges that we faced in our experiments.

Another limitation was that, our experiments **limited the computation of the number of imposter pairs(sampling bias), due to the low system complexity**. This sampling bias significantly impacted our results, especially with pinsdataset, where we saw 100% errors. Relevant steps have to be taken in the future to mitigate this sampling bias. Yet another constraint was that, **we did not address anomalies in the dataset**. Where we noticed that there were quiet a lot of anomalies in our experiments (refer to figure 5.26). Therefore future works can focus on addressing anomalies in the datasets. Lastly, **We did not use the improved version of the OpenFace model specifically tailored for the face recognition tasks[BZLM18].** Future experiments might include the improved versions of the OpenFace model.

5.5 Strengths

Major strength of our experiments was the **meticulous integration of diverse datasets**. This diverse integration of different datasets might have paved a way for a comprehensive evaluation of the SER-FIQ algorithm[TKD+20] across varying conditions and characteristics. Additionally **the selection of different pretrained models like ArcFace model and the OpenFace model might have provided a robust investigation of the impact of model architectures on FIQA**. Where we understood that even when triplet loss function is used with improved training conditions, it might provide reasonable results. **Our experimental setup incorporated wide variety of model configurations and dataset characteristics**. This approach might have provided valuable insights on the performance of the SER-FIQ algorithm[TKD+20]. Lastly, **the identification and discussion of various factors influencing our results such as anomalies, sampling bias and dataset imbalance might have demonstrated an analytical and a transparent approach**.

5.6 Consistency with the existing literature

Our work reflects a lot of scientific areas that might echo similar discussion in the existing literature. Firstly, building an on-top model upon the pretrained model, emphasizes the significance of choosing appropriate model architectures for an efficient feature extraction and classification. This particularly **might resonate with the discussion around the domain of transfer learning in the existing literature**[TS10], [WKW16], [ZQD+20b]. Secondly, the acknowledgement of anomalies, inaccurate labelling and inefficient face detection **resonates with the existing literature that is discussing the detection of anomalous face images and other challenges faces in face recognition as well as FIQA**[BRF18], [ZCPR03], [SRH+22].

Thirdly, discussing the imbalance between over fitting and under fitting in deep neural network models, strongly matches the literature that is discussing the impact of these factors on feature extraction and classification[Yin19], [Haw04], [MWHN18]. Lastly, recognition of the factors that may influence our results, strengths, limitations and implications for future work, **might perfectly resonate with the literature that is discussing comprehensive improvements in the domain of face recognition and face image quality assessment**[ZF14], [ZCPR03], [SRH+22],[MWHN18].

Although we pointed out several factors that resonate with the existing literature, there might be no other work that might probably have explored the effect of dropout regularization in different model configurations. Our results provide valuable insights into the existing research of face image quality assessment. Where we showcase how the SER-FIQ algorithm responds to different model configurations and different dataset characteristics.

5.7 Practical implications

Firstly, our results indicates the significance of the incorporation of dropout regularization, and balancing the model bias in the process of face image quality assessment. This indicates that the practitioners should include dropout regularization but with a good balance to manage the model bias. Secondly, we have given a practical guidance with our results on how model selection impacts the performance. Where every detail of the model should be cross validated with rigorous experiments. Thirdly, lack of resampling techniques in our experiments, highlights an area of improvement in the practical applications. Hence practitioners should consider resampling the datasets before performing the experiments. Lastly, practical implementations must be vigilant about the presence of anomalies, incorrect labelling and the efficiency of face detection.

5.8 Future work

The future work can address the factors that have influenced our unusual results. Anomalies, inaccurate labelling, inaccurate face detection and inaccurate preprocessing can be addressed first by removing the anomalies then using an improved version of the MTCNN algorithm[WC23], [ZZLQ16] for face detection. The sampling bias can be mitigated by computing more than 80% of the comparisons, which can balance the ratio of genuine to imposter scores. This might give more deeper insights into the performance of the FIQA through the error vs reject curves. Especially with pinsdataset, the number of imposter scores should not be compromised. This is because the **class distribution of pinsdataset demands computation of all genuine and imposter pairs**.

Imbalanced datasets can be addressed by incorporating the resampling techniques like under sampling, oversampling or using data augmentation techniques to introduce more data transformations into the training process[THKG20]. Using generative adversarial networks(GANs)[SC21] might be helpful here to generate more synthetic samples of the datasets. Additionally, retraining the model multiple times can avoid the model bias, which in turn can address the dataset class imbalance[THKG20]. We used sparse categorical cross entropy loss function[WMZT20] in the on-top models, but this loss function is really sensitive to the imbalanced dataset[WMZT20], [THKG20]. The future work can explore and experiment with different loss functions to get better results[WMZT20]. However it is important to note that effective experimentation and validation is necessary to form meaningful conclusions.

When selecting pretrained models in the future experiments, it is important to take more time

and explore wide variety of options. Perform cross validation first to assess the performance of the model, then choose a good pretrained model that generalizes well among different datasets. Also, one should always opt for a state of the art model for future experimentation. Like we used a more generalised version of the OpenFace model[ALS+16], future experiements can use a more improved version of the OpenFace model specifically tailored for the face recognition tasks[BZLM18], [SK18].

The overfitting and underfitting problems should be balanced very rigorously[Yin19], [Haw04]. Where the model may get biased towards learning the training data too well, which can make it difficult for the model to generalise well to the unseen data. Here, one can utilise ensemble methods[Die00] like bagging(Random forests) or boosting(Gradient boosting) techniques to reduce over fitting[Die00]. Train different sub networks with dropout regularization[SHK+14], [Sri13] and L1/L2 regularization[KSST12], [NJ18], combine different networks into one ensemble[Die00] to get a better prediction. Which can balance the over fitting and under fitting during the training process. Even though, use case 3 might have concluded that the consistent incorporation of dropout regularization in both the pretrained model and the on-top model is better. It might be even more better to balance the dropout regularization in both models to avoid the model bias during the training process. **This balance might be achieved by rigorously experimenting with different dropout rates(the probability value of p in dropout regularization)**.

The use of loss functions across different models should be done after extensive experimentation. Architectural choices must be done attentively, so that the model's ability to extract intricate feature vectors and to classify images is accurate. Meticulous hyper parameter tuning must be done the caters the needs of the experiments after a thorough cross verification. While training the on-top models, one must make sure that enough large datasets are used with extensive resampling. This must be done in a hope that the challenges of face recognition like face occlusions, different face expressions, ethnicities, different age groups and other challenges, can be resolved[TKD+20], [ZCPR03]. Further adapting the future experiments with the advanced architectures of CNNs might provide better results[MMRY24], [CPRW19], [LJZ+20]. Please refer to our tabulated results for an in depth understanding.

6

Conclusion

Different configurations of dropout regularization was not explored in previous works. In this work we explore the effect of different model configurations by using two pretrained models, one that is trained with dropout regularization and another that is trained without dropouts. Here we choose an ArcFace model that is trained with dropout regularization and an OpenFace model that is trained without dropout regularization. By building on-top models with and without dropout regularization, we investigated six use cases and assessed their performance on various datasets. By measuring the embedding variations from the random networks created by dropout regularization, we had two assumptions namely assigning the lowest quality scores to the lowest quality images and assigning the highest quality scores to the lowest quality images.

Our experiments revealed that incorporating dropout regularization in both the pretrained model and the on-top model during the process of training, might give optimal performance. We also observed that the pretrained model with dropout regularization might perform better when compared to the pretrained model without dropout regularization. Additionally we observed that, there might exist an ambiguity between the two assumptions that we made, which was observed by the similar trends in the error vs reject curves. We were unable to explore the effect of different dropout rates on the overall performance and we we did not incorporate cross-validation and resampling techniques. But we provided a detailed analysis of the factors that might influence our results including the presence of anomalies, sampling bias and much more to it.

Overall, our findings indicate the importance of different configurations of dropout regularization for an optimal face image quality assessment. The provided thorough analysis across different datasets and different models might have provided valuable insights for future research. It is expected that our commitment to a thorough and insightful exploration might serve valuable contributions in the advancements of face image quality assessment.

Bibliography

[ALS+16] Brandon Amos, Bartosz Ludwiczuk, Mahadev Satyanarayanan, et al. Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*, 6(2):20, 2016.

[AOBTA20] Insaf Adjabi, Abdeldjalil Ouahabi, Amir Benzaoui, and Abdelmalik Taleb-Ahmed. Past, present, and future of face recognition: A review. *Electronics*, 9(8):1188, 2020.

[BFK+23] Fadi Boutros, Meiling Fang, Marcel Klemt, Biying Fu, and Naser Damer. Cr-fiqa: face image quality assessment by learning sample relative classifiability. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5836–5845, 2023.

[BRA+09] Debnath Bhattacharyya, Rahul Ranjan, Farkhod Alisherov, Minkyu Choi, et al. Biometric authentication: A review. *International Journal of u-and e-Service, Science and Technology*, 2(3):13–28, 2009.

[BRF18] Anand Bhattad, Jason Rock, and David Forsyth. Detecting anomalous faces with'no peeking'autoencoders. *arXiv preprint arXiv:1802.05798*, 2018.

[Bur20] Burak. Pinterest face recognition dataset. `www.kaggle.com/datasets/hereisburak/pins-facerecognition`, 2020.

[BVS14] Samarth Bharadwaj, Mayank Vatsa, and Richa Singh. Biometric quality: a review of fingerprint, iris, and face. *EURASIP journal on Image and Video Processing*, 2014(1):1–28, 2014.

[BZLM18] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pages 59–66, 2018.

[CB17] Marco Cuturi and Mathieu Blondel. Soft-dtw: a differentiable loss function for time-series. In *International conference on machine learning*, pages 894–903. PMLR, 2017.

[CDRP18] Monali Nitin Chaudhari, Mrinal Deshmukh, Gayatri Ramrakhiani, and Rakshita Parvatikar. Face detection using viola jones algorithm and neural networks. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pages 1–6, 2018.

[CFB05] Xin Chen, Patrick J Flynn, and Kevin W Bowyer. Ir and visible light face

recognition. *Computer Vision and Image Understanding*, 99(3):332–358, 2005.

[CPRW19] Adele Chui, Anshuman Patnaik, Krishn Ramesh, and Linda Wang. Capsule networks and face recognition. *Lindawangg. github. io*, 2019.

[CY22] Zehao Chen and Hua Yang. L2rt-fiqa: Face image quality assessment via learning-to-rank transformer. In *International Forum on Digital TV and Wireless Multimedia Communications*, pages 270–285. Springer, 2022.

[CYA20] Shan Cao, Yuqian Yao, and Gaoyun An. E2-capsule neural networks for facial expression recognition using au-aware attention. *IET Image Processing*, 14(11):2417–2424, 2020.

[CYTS10] Zhimin Cao, Qi Yin, Xiaoou Tang, and Jian Sun. Face recognition with learning-based descriptor. In *2010 IEEE Computer society conference on computer vision and pattern recognition*, pages 2707–2714. IEEE, 2010.

[DCWZ16] Xuedan Du, Yinghao Cai, Shuo Wang, and Leijie Zhang. Overview of deep learning. In *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pages 159–164, 2016.

[DGNZ19] Jiankang Deng, Jia Guo, Xue Niannan, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *CVPR*, 2019.

[DGV+20] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5203–5212, 2020.

[DGXZ19a] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4685–4694, 2019.

[DGXZ19b] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019.

[DGZ+19] Jiankang Deng, Jia Guo, Yuxiang Zhou, Jinke Yu, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-stage dense face localisation in the wild. *arXiv preprint arXiv:1905.00641*, 2019.

[Die00] Thomas G Dietterich. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer, 2000.

[DLZS21] Ionut Cosmin Duta, Li Liu, Fan Zhu, and Ling Shao. Improved residual networks for image and video recognition. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 9415–9422, 2021.

[Fas18] Laurent Fasnacht. mmappickle: Python 3 module to store memory-mapped numpy array in pickle format. *Journal of Open Source Software*, 3(26):651, 2018.

[FSL15] Sachin Sudhakar Farfade, Mohammad J Saberian, and Li-Jia Li. Multi-view face detection using deep convolutional neural networks. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, pages 643–650,

2015.

[GDXZ18] Jia Guo, Jiankang Deng, Niannan Xue, and Stefanos Zafeiriou. Stacked dense u-nets with dual transformers for robust face alignment. In *BMVC*, 2018.

[Ge18] Weifeng Ge. Deep metric learning with hierarchical triplet loss. In *Proceedings of the European conference on computer vision (ECCV)*, pages 269–285, 2018.

[GGNH14] Patrick J Grother, Patrick J Grother, Mei Ngan, and K Hanaoka. *Face recognition vendor test (FRVT)*. US Department of Commerce, National Institute of Standards and Technology, 2014.

[GLQZ24] Weijun Gong, Zhiyao La, Yurong Qian, and Weihang Zhou. Hybrid attention-aware learning network for facial expression recognition in the wild. *Arabian Journal for Science and Engineering*, pages 1–15, 2024.

[GT07] Patrick Grother and Elham Tabassi. Performance of biometric quality measures. *IEEE transactions on pattern analysis and machine intelligence*, 29(4):531–543, 2007.

[HA15] M Hassaballah and Saleh Aly. Face recognition: challenges, achievements and future directions. *IET Computer Vision*, 9(4):614–626, 2015.

[Haw04] Douglas M Hawkins. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12, 2004.

[HBL17] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017.

[HKW11] Geoffrey E Hinton, Alex Krizhevsky, and Sida D Wang. Transforming auto-encoders. In *Artificial Neural Networks and Machine Learning–ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011, Proceedings, Part I 21*, pages 44–51. Springer, 2011.

[HRBLM07] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[HWW+21] Baojin Huang, Zhongyuan Wang, Guangcheng Wang, Kui Jiang, Kangli Zeng, Zhen Han, Xin Tian, and Yuhong Yang. When face recognition meets occlusion: A new benchmark. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4240–4244, 2021.

[HZRS16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[JFR07] Anil K Jain, Patrick Flynn, and Arun A Ross. *Handbook of biometrics*. Springer Science & Business Media, 2007.

[JL11] Anil K Jain and Stan Z Li. *Handbook of face recognition*, volume 1. Springer, 2011.

[KB03] XCPJF Kevin and W Bowyer. Visible-light and infrared face recognition. In *Workshop on Multimodal User Authentication*, volume 48. Citeseer, 2003.

[KBR23] Wassim Kabbani, Christoph Busch, and Kiran Raja. Robust sclera segmentation

for skin-tone agnostic face image quality assessment. In *2023 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–6, 2023.

[KMHS24] Md Humaun Kabir, Abu Saleh Musa Miah, Md Hadiuzzaman, and Jungpil Shin. Combining state-of-the-art pre-trained deep learning models: A noble approach for bangla sign language recognition using max voting ensemble. *Available at SSRN 4693354*, 2024.

[Kom04] Leonid Kompanets. Biometrics of asymmetrical face. In *International Conference on Biometric Authentication*, pages 67–73. Springer, 2004.

[KS12] Przemysław Kocjan and Khalid Saeed. Face recognition in unconstrained environment. In *Biometrics and kansei engineering*, pages 21–42. Springer, 2012.

[KSH] A Krizhevsky, I Sutskever, and G Hinton. Imagenet classification with deep convolutional networks. In *Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS)*, pages 1106–1114.

[KSST12] Sham M Kakade, Shai Shalev-Shwartz, and Ambuj Tewari. Regularization techniques for learning with matrices. *The Journal of Machine Learning Research*, 13(1):1865–1890, 2012.

[LDL20] Li Li, Miloš Doroslovački, and Murray H Loew. Approximating the gradient of cross-entropy loss function. *IEEE access*, 8:111626–111635, 2020.

[LDL24] Hao Liu, Xinyi Duan, and Jiuzhen Liang. Ff-ppqa: Face frontalization without glasses based on perceptual quality and pixel-level quality assessment. *Signal, Image and Video Processing*, pages 1–15, 2024.

[LJF+24] Chengcheng Lu, Yiben Jiang, Keren Fu, Qijun Zhao, and Hongyu Yang. Lstp-net: Long short-term perception network for dynamic facial expression recognition in the wild. *Image and Vision Computing*, page 104915, 2024.

[LJZ+20] Jing Li, Kan Jin, Dalin Zhou, Naoyuki Kubota, and Zhaojie Ju. Attention mechanism-based cnn for facial expression recognition. *Neurocomputing*, 411:340–350, 2020.

[LM19] Wen-Yao Lu and YANG Ming. Face detection based on viola-jones algorithm applying composite features. In *2019 International Conference on Robots & Intelligent System (ICRIS)*, pages 82–85. IEEE, 2019.

[LMLP20] Lixiang Li, Xiaohui Mu, Siying Li, and Haipeng Peng. A review of face recognition technology. *IEEE Access*, 8:139110–139120, 2020.

[LW02] Chengjun Liu and Harry Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image processing*, 11(4):467–476, 2002.

[LWY+17] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 212–220, 2017.

[LWYY16] Weiyang Liu, Yandong Wen, Zhiding Yu, and Meng Yang. Large-margin softmax loss for convolutional neural networks. *arXiv preprint arXiv:1612.02295*, 2016.

[MGMLG03] Catherine J Mondloch, Sybil Geldart, Daphne Maurer, and Richard Le Grand.

Developmental changes in face processing skills. *Journal of experimental child psychology*, 86(1):67–84, 2003.

[MJ01] Larry R Medsker and LC Jain. Recurrent neural networks. *Design and Applications*, 5(64-67):2, 2001.

[MMRY24] Devira Anggi Maharani, Carmadi Machbub, Pranoto Hidaya Rusmin, and Lenni Yulianti. Real-time human tracking using multi-features visual with cnn-lstm and q-learning. *IEEE Access*, pages 1–1, 2024.

[MWHN18] Iacopo Masi, Yue Wu, Tal Hassner, and Prem Natarajan. Deep face recognition: A survey. In *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 471–478, 2018.

[NJ18] Ismoilov Nusrat and Sung-Bong Jang. A comparison of regularization techniques in deep neural networks. *Symmetry*, 10(11):648, 2018.

[OCP+18] Alice J. O'Toole, Carlos D. Castillo, Connor J. Parde, Matthew Q. Hill, and Rama Chellappa. Face space representations in deep convolutional neural networks. *Trends in Cognitive Sciences*, 22(9):794–809, 2018.

[OCZ+21] Fu-Zhao Ou, Xingyu Chen, Ruixin Zhang, Yuge Huang, Shaoxin Li, Jilin Li, Yong Li, Liujuan Cao, and Yuan-Gen Wang. Sdd-fiqa: unsupervised face image quality assessment with similarity distribution distance. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7670–7679, 2021.

[ON15] Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

[PR04] M. Pantic and L.J.M. Rothkrantz. Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(3):1449–1461, 2004.

[PVZ15] Omkar Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. In *BMVC 2015-Proceedings of the British Machine Vision Conference 2015*. British Machine Vision Association, 2015.

[QS17] Ce Qi and Fei Su. Contrastive-center loss for deep neural networks. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 2851–2855, 2017.

[RC24] Solange Griselly Ramos Cooper. Multimodal unconstrained people recognition with face and ear images using deep learning. *SUNEDU*, 2024.

[RDR19] Ajita Rattani, Reza Derakhshani, and Arun Ross. Selfie biometrics. *Advances and Challenges. Cham: Springer Nature*, 2019.

[SC21] Divya Saxena and Jiannong Cao. Generative adversarial networks (gans) challenges, solutions, and future directions. *ACM Computing Surveys (CSUR)*, 54(3):1–42, 2021.

[SCDR22] Praneet Singh, Haoyu Chen, Edward J. Delp, and Amy R. Reibman. Evaluating image quality estimators for face matching. In *2022 IEEE 5th International Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 204–209, 2022.

[Ser21] Sefik Ilkin Serengil. Deepface: A python framework for face recognition. `https:`

//pypi.org/project/deepface/, 2021. Accessed: Month Day, Year.

[SHK+14] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[SK18] Kevin Santoso and Gede Putra Kusuma. Face recognition using modified openface. *Procedia Computer Science*, 135:510–517, 2018.

[SKP15] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[SLH+23] Shaolin Su, Hanhe Lin, Vlad Hosu, Oliver Wiedemann, Jinqiu Sun, Yu Zhu, Hantao Liu, Yanning Zhang, and Dietmar Saupe. Going the extra mile in face image quality assessment: A novel database and model. *IEEE Transactions on Multimedia*, pages 1–15, 2023.

[SO20] Sefik Ilkin Serengil and Alper Ozpinar. Lightface: A hybrid deep face recognition framework. In *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pages 23–27. IEEE, 2020.

[SO21] Sefik Ilkin Serengil and Alper Ozpinar. Hyperextended lightface: A facial attribute analysis framework. In *2021 International Conference on Engineering and Emerging Technologies (ICEET)*, pages 1–4. IEEE, 2021.

[SRH+22] Torsten Schlett, Christian Rathgeb, Olaf Henniger, Javier Galbally, Julian Fierrez, and Christoph Busch. Face image quality assessment: A literature survey. *ACM Comput. Surv.*, 54(10s), sep 2022.

[Sri13] Nitish Srivastava. Improving neural networks with dropout. *University of Toronto*, 182(566):7, 2013.

[SSG+18] Tiomotheos Samatzidis, Dirk Siegmund, Michael Goedde, Naser Damer, Andreas Braun, and Arjan Kuijper. The dark side of the face: Exploring the ultraviolet spectrum for face biometrics. In *2018 International Conference on Biometrics (ICB)*, pages 182–189, 2018.

[SWH18] Xudong Sun, Pengcheng Wu, and Steven CH Hoi. Face detection using deep learning: An improved faster rcnn approach. *Neurocomputing*, 299:42–50, 2018.

[SWY75] G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. *Commun. ACM*, 18(11):613–620, nov 1975.

[SZ14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[TCERPCU23] Juan Terven, Diana M Cordova-Esparza, Alfonzo Ramirez-Pedraza, and Edgar A Chavez-Urbiola. Loss functions and metrics in deep learning. a review. *arXiv preprint arXiv:2307.02694*, 2023.

[TCZZ06] Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, and Fuyan Zhang. Face recognition from a single image per person: A survey. *Pattern recognition*, 39(9):1725–1745, 2006.

[Ter21] Philipp Terhörst. Mitigating soft-biometric driven bias and privacy concerns in face recognition systems. 2021.

[TGBB20] J Touati, M Sc A Gudi, S Bhulai, and ER Belitser. A deep learning approach to face image quality assessment. 2020.

[THKG20] Fadi Thabtah, Suhel Hammoud, Firuz Kamalov, and Amanda Gonsalves. Data imbalance in classification: Experimental evaluation. *Information Sciences*, 513:429–441, 2020.

[TIH+23] Philipp Terhörst, Malte Ihlefeld, Marco Huber, Naser Damer, Florian Kirchbuchner, Kiran Raja, and Arjan Kuijper. Qmagface: Simple and accurate quality-aware face recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3484–3494, 2023.

[TKD+20] Philipp Terhorst, Jan Niklas Kolf, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Ser-fiq: Unsupervised estimation of face image quality based on stochastic embedding robustness. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5651–5660, 2020.

[TKE20] Murat Taskiran, Nihan Kahraman, and Cigdem Eroglu Erdem. Face recognition: Past, present and future (a review). *Digital Signal Processing*, 106:102809, 2020.

[TP91] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991.

[TS10] Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global, 2010.

[TXS24] Yumei Tan, Haiying Xia, and Shuxiang Song. Learning informative and discriminative semantic features for robust facial expression recognition. *Journal of Visual Communication and Image Representation*, page 104062, 2024.

[TYRW14] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1701–1708, 2014.

[vdHvGP13] Helen van de Haar, Darelle van Greunen, and Dalenca Pottas. The characteristics of a biometric. In *2013 Information Security for South Africa*, pages 1–8, 2013.

[VSP+17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[VV21] Edward Vendrow and Joshua Vendrow. Realistic face reconstruction from deep embeddings. In *NeurIPS 2021 Workshop Privacy in Machine Learning*, 2021.

[Wan14] Yi-Qing Wang. An analysis of the viola-jones face detection algorithm. *Image Processing On Line*, 4:128–148, 2014.

[WC23] Yanjun Wang and Xiaohui Cheng. An improved mtcnn face detection algorithm. In *Proceedings of the 2022 5th International Conference on E-Business, Information Management and Computer Science*, EBIMCS '22, page 116–120, New York, NY, USA, 2023. Association for Computing Machinery.

[WD20] Mei Wang and Weihong Deng. Mitigating bias in face recognition using

skewness-aware reinforcement learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9322–9331, 2020.

[Wea06] A.C. Weaver. Biometric authentication. *Computer*, 39(2):96–97, 2006.

[WHM11] Lior Wolf, Tal Hassner, and Itay Maoz. Face recognition in unconstrained videos with matched background similarity. In *CVPR 2011*, pages 529–534, 2011.

[WJMM05] James Wayman, Anil Jain, Davide Maltoni, and Dario Maio. An introduction to biometric authentication systems. In *Biometric systems: Technology, design and performance evaluation*, pages 1–20. Springer, 2005.

[WKW16] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3(1):1–40, 2016.

[WMZT20] Qi Wang, Yue Ma, Kun Zhao, and Yingjie Tian. A comprehensive survey of loss functions in machine learning. *Annals of Data Science*, pages 1–26, 2020.

[WR17] Haohan Wang and Bhiksha Raj. On the origin of deep learning. *arXiv preprint arXiv:1702.07800*, 2017.

[WWZ+18] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5265–5274, 2018.

[WYWL20] Di Wang, Hongzhi Yu, Ding Wang, and Guanyu Li. Face recognition system based on cnn. In *2020 International Conference on Computer Information and Big Data Applications (CIBDA)*, pages 470–473, 2020.

[XWX23] Huimin Xiao, Xinghao Wang, and Qiang Xing. Facial expression recognition in the wild based on convolutional neural network and graph convolutional network. In *Fourth International Conference on Signal Processing and Computer Science (SPCS 2023)*, volume 12970, pages 976–981. SPIE, 2023.

[YAN24] Amir Khani Yengikand, Mostafa Farrokhi Afsharyan, and Payam Nejati. Facial expression recognition based on separable convolution network and attention mechanism. 2024.

[Yeg09] Bayya Yegnanarayana. *Artificial neural networks*. PHI Learning Pvt. Ltd., 2009.

[Yin19] Xue Ying. An overview of overfitting and its solutions. In *Journal of physics: Conference series*, volume 1168, page 022022. IOP Publishing, 2019.

[ZCPR03] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, dec 2003.

[ZF14] Ghazi Mohammed Zafaruddin and H. S. Fadewar. Face recognition: A holistic approach review. In *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, pages 175–178, 2014.

[ZHSZ20] Haiping Zhu, Zhizhong Huang, Hongming Shan, and Junping Zhang. Look globally, age locally: Face aging with an attention mechanism. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1963–1967. IEEE, 2020.

[ZK24] Tao Zhang and Zewu Ke. Research on face recognition algorithm based on

facenet and coordinate attention. In *International Conference on Algorithm, Imaging Processing, and Machine Vision (AIPMV 2023)*, volume 12969, pages 143–147. SPIE, 2024.

[ZL19] Hui Zhi and Sanyang Liu. Face recognition based on genetic algorithm. *Journal of Visual Communication and Image Representation*, 58:495–502, 2019.

[ZQD+20a] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2020.

[ZQD+20b] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2020.

[ZZLQ16] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters*, 23(10):1499–1503, 2016.